

IBM Cognos Analytics
Version 11.1.0

Explorationen - Benutzerhandbuch



Inhaltsverzeichnis

Kapitel 1. Getting started with Explorations.....	1
Explorationen.....	1
Daten hochladen.....	1
Exploration von einem bestehenden Dashboard oder einer bestehenden Story starten	1
Neue Exploration über das Menü Neu starten.....	1
Neue Exploration aus einem Datenasset auf der Begrüßungsseite starten.....	2
Datenquelle hinzufügen.....	2
Kapitel 2. Exploring relationships.....	3
Beziehungen in Daten erkunden.....	3
Beziehungsdigramm öffnen.....	7
Kapitel 3. Visualizations.....	9
Visualisierungen.....	9
Karten im Navigationsfenster anzeigen	9
Einzelne Visualisierung erstellen.....	10
Visualisierung mithilfe der Suche in Datenfeldern erstellen.....	10
Zwei Visualisierungen vergleichen.....	11
Zwei Datenpunkte einer Visualisierung vergleichen.....	12
Erweiterte Datenanalyse.....	12
Anderen Visualisierungstyp auswählen.....	13
Insights in Visualisierungen.....	46
Korrelierte Insights auswählen.....	47
Empfohlene Visualisierungen auswählen.....	47
Zugehörige Visualisierungen auswählen.....	47
Kapitel 4. Forecasting.....	49
Prognose.....	49
Prognosefunktionen.....	49
Prognoseoptionen.....	50
Visualisierungstypen, die die Prognose unterstützen.....	53
Prognosedaten.....	53
Prognose - Statistikdetails.....	56
Prognosemodelle.....	59
Kapitel 5. Principles of advanced data analytics.....	63
Richtlinien für die erweiterte Datenanalyse.....	63
Datenaufbereitung.....	64
Datenaufbereitung für numerische Felder.....	64
Datenvorbereitung für kategoriale Felder.....	65
Datenaufbereitung für Zielfelder.....	65
Unidirektionale zentrale Einflussfaktoren.....	66
Bidirektionale zentrale Einflussfaktoren.....	67
Entscheidungsstruktur.....	68
Insights in Visualisierungen.....	70
Details zur natürlichen Sprache.....	79
Beziehungen.....	83
Unterschiede zwischen Cognos Analytics und Watson Analytics in Bezug auf erweiterte Daten- analyse.....	84

Kapitel 6. Assistent.....	91
Assistent.....	91
Funktionen des Assistenten.....	92

Kapitel 1. Getting started with Explorations

Explorationen

Explorationen sind flexible Arbeitsbereiche zum Erkunden, in denen Sie Daten erkennen und analysieren können. Sie können auch eine bestehende Visualisierung aus einem Dashboard oder einer Story erkunden. Erkennen Sie verdeckte Beziehungen und decken Sie Muster auf, die Ihre Daten in Insights verwandeln. Korrelierte Insights werden durch ein grünes Symbol mit einer Zahl auf der X-Achse, auf der Y-Achse oder dem Titel eines Diagramms dargestellt.

Explorationen starten

Zum Starten von Explorationen stehen mehrere Methoden zur Verfügung.

Daten hochladen

Laden Sie ein Datenasset zur Verwendung in Ihrer Exploration in Ihren Ordner **Eigener Inhalt** hoch.

Vorgehensweise

1. Klicken Sie auf das Symbol **Neu** und dann auf **Dateien hochladen**.



2. Navigieren Sie an die Speicherposition Ihres Datenassets und wählen Sie es aus.
Das Datenasset wird im Ordner **Eigener Inhalt** angezeigt.


Exploration von einem bestehenden Dashboard oder einer bestehenden Story starten

Wenn Sie an einem Dashboard oder einer Story arbeiten, können Sie eine Exploration direkt über eine Visualisierung erstellen oder bearbeiten.

Informationen zu diesem Vorgang

Führen Sie die folgenden Schritte aus, um eine Visualisierung in einer neuen Exploration zu öffnen oder zu einer bestehenden Exploration hinzuzufügen.


Vorgehensweise

1. Öffnen Sie ein bestehendes Dashboard oder eine bestehende Story.
2. Wählen Sie eine Visualisierung aus.
3. Klicken Sie in der Symbolleiste auf das Symbol **Explorationen** .
4. Wählen Sie die Option **Neue Exploration** oder **Hinzufügen zu vorhandenem Element** aus.

Neue Exploration über das Menü Neu starten

Sie können auf der Seite **Willkommen** über das Menü **Neu** eine neue Exploration starten.

Vorgehensweise


1. Klicken Sie auf **Neu**  und dann auf **Exploration**.
2. Wählen Sie eine Datenquelle aus und klicken Sie auf **Hinzufügen**.

Aus der von Ihnen ausgewählten Datenquelle wird eine Ausgangspunktseite erstellt.

Neue Exploration aus einem Datenasset auf der Begrüßungsseite starten

Sie können das **Aktionsmenü** zu einem kürzlich verwendeten Datenasset auf der Seite **Willkommen** auswählen.


Vorgehensweise

1. Klicken Sie auf der **Begrüßungsseite**, wenn das Datenasset, das Sie verwenden möchten, als Kachel im Bereich **Zuletzt verwendete Projekte** angezeigt wird, auf das Symbol **Aktionsmenü** .
2. Wählen Sie **Exploration erstellen** aus.

Datenquelle hinzufügen

Fügen Sie eine Datenquelle zu Ihrer Exploration hinzu, um deren Daten zu untersuchen.

Vorgehensweise

1. Klicken Sie im Teilfenster **Ausgewählte Quellen** auf das Symbol **Quelle hinzufügen** .
2. Wechseln Sie in den Ordner **Eigener Inhalt** oder **Teaminhalt** und wählen Sie die Datenquelle aus, die hinzugefügt werden soll. Klicken Sie auf **Hinzufügen**.
3. Erweitern Sie die Datenquelle im Fenster **Ausgewählte Quellen**, um die verfügbaren Elemente anzuzeigen.
4. Verwenden Sie die Ausgangspunktseite, um ein Beziehungsdigramm aus Ihren Daten zu generieren.

Kapitel 2. Exploring relationships

Beziehungen in Daten erkunden

Beim Erstellen einer Exploration können Sie von einer Datenquelle starten. Es wird eine Ausgangspunktseite angezeigt mit Vorschlägen für erste Schritte.

Sie können einen Spaltennamen eingeben, der in Ihrer Datenquelle angezeigt wird. Alternativ können Sie auf eine der vorgeschlagenen Spalten klicken, die vom System als interessant ermittelt werden. Wenn Sie nicht sicher sind, mit welcher Spalte Sie beginnen sollen, klicken Sie auf **Überspringen - beliebige Anzeige**. Daraufhin wird ein Beziehungsdiagramm angezeigt, das einige Vorschläge für Ausgangspunktvisualisierungen enthält.

Where do you want to start?

Every exploration includes a starting card.

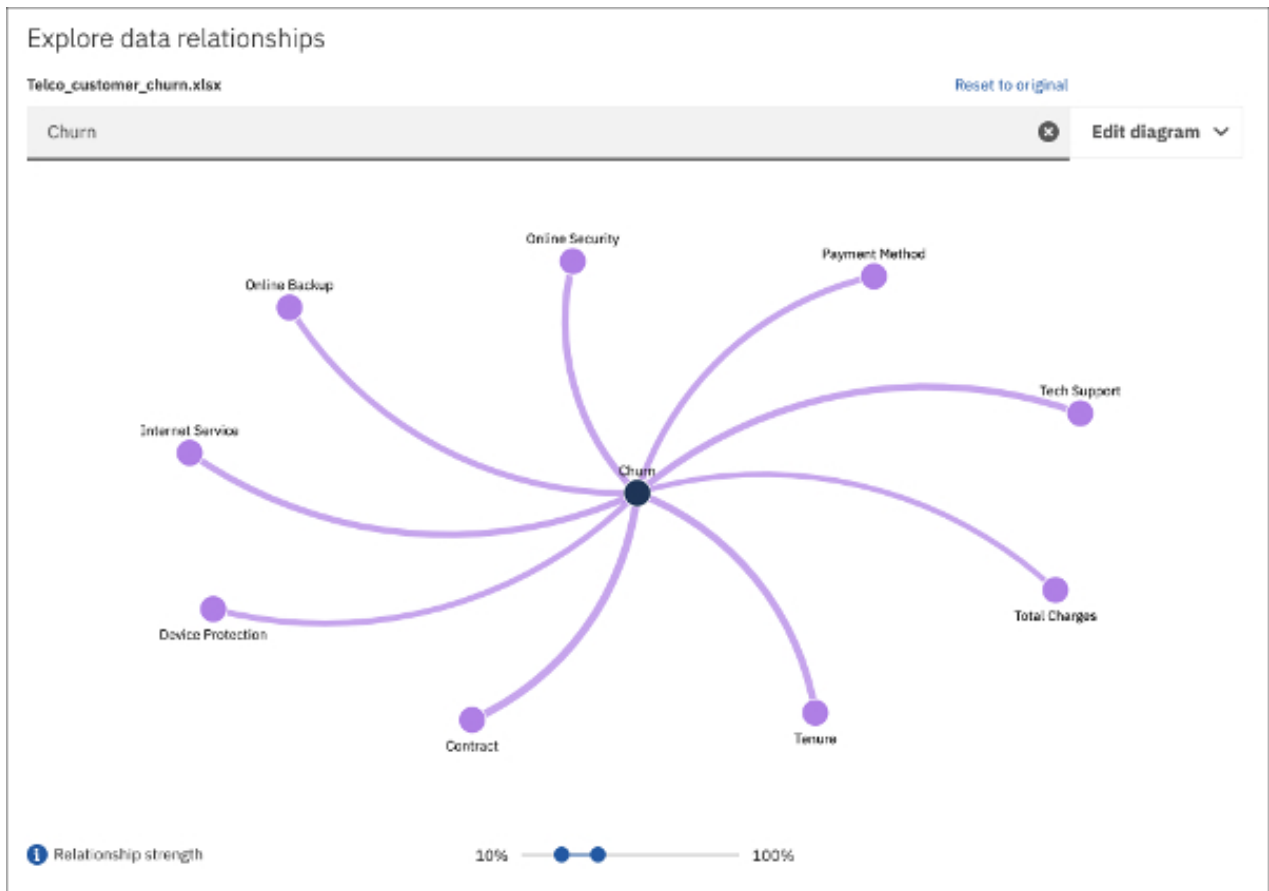
Start with any column in 'Telco_customer_churn.xlsx'. You can always change it later.

Try starting with one of these...

Churn Monthly Charges Total Charges Tenure

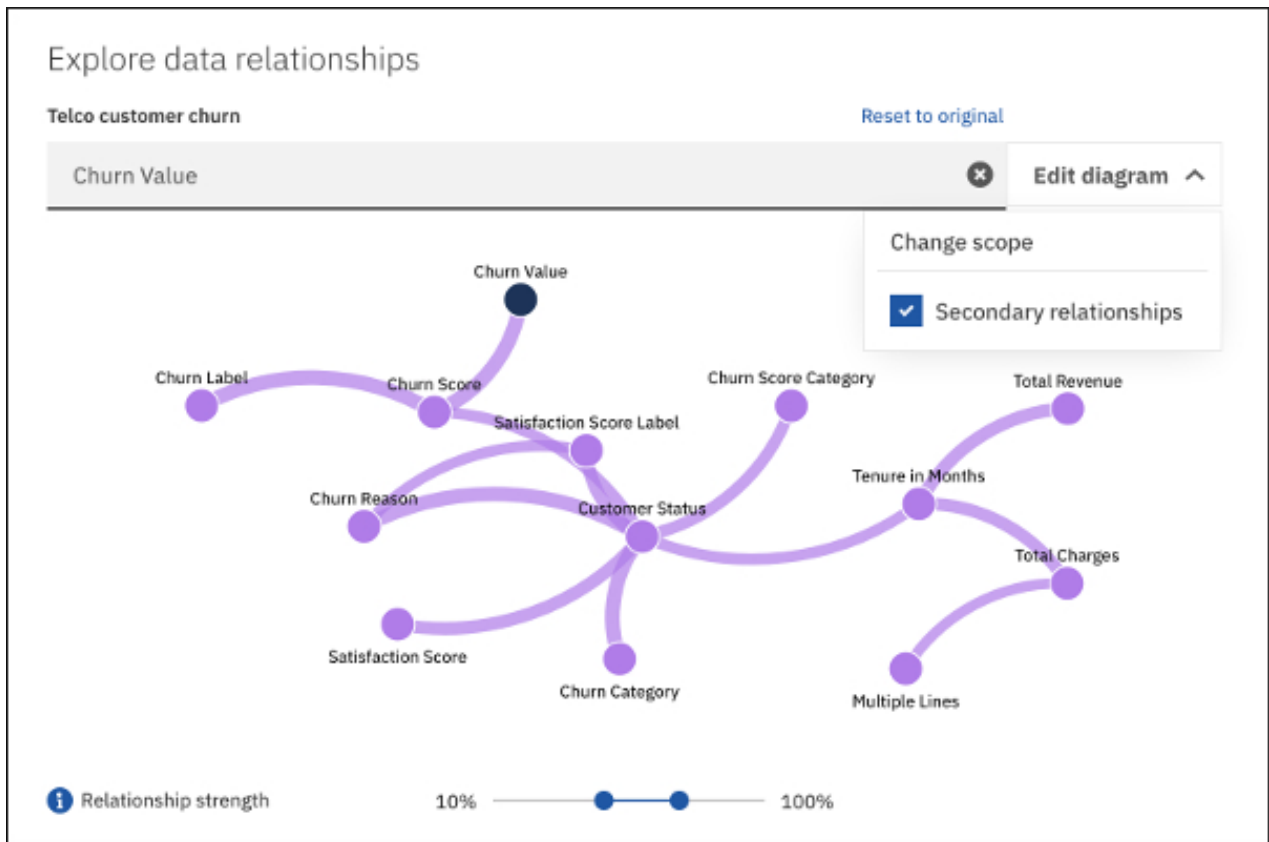
[Skip - show me anything](#)

Im Beziehungsdiagramm stellt die Spalte, mit der Sie beginnen, den primären Fokus dar und sie wird durch einen dunkelblauen Knoten dargestellt. Zugehörige Felder werden durch purpurrote Knoten dargestellt. Linien verbinden die Knoten und stellen Beziehungen dar. Die Breite der Linie weist auf die Stärke der Beziehung hin.



Die stärksten primären Beziehungen werden standardmäßig angezeigt und sind die direkten Beziehungen zwischen dem primären Fokus und den zugehörigen Feldern. Bei sekundären Beziehungen handelt es sich um die Beziehungen zwischen anderen Feldern, die direkt oder indirekt mit dem Ziel in Beziehung stehen.

Wenn sowohl die primären als auch die sekundären Beziehungen angezeigt werden sollen, wählen Sie das Kontrollkästchen **Sekundäre Beziehungen** unter **Diagramm bearbeiten** aus.



Im Beziehungsdiagramm werden diese Felder basierend auf einer statistischen Auswertung verwandter Elemente dargestellt. Das Beziehungsdiagramm ist keine Abbildung des Datenmodells. Das Modell kann jedoch ein Einflussfaktor bei der Analyse sein. Um die Leistung zu verbessern, wenn die Datenquelle viele Zeilen enthält, basiert die Analyse auf einer repräsentativen Stichprobe der gesamten Daten.

Sie können mit dem Beziehungsdiagramm interagieren, indem Sie einen Knoten auswählen, an dem Sie interessiert sind. In diesem Fall wird die Liste der vorgeschlagenen Ausgangspunktvisualisierungen rechts vom Diagramm entsprechend aktualisiert und enthält die von Ihnen ausgewählten Knoten. Zum Auswählen mehrerer Knoten können Sie Strg+Mausklick verwenden.

Klicken Sie auf **Auf Original zurücksetzen**, wenn Sie den Bereich und die Anzeige aller Felder im Beziehungsdiagramm auf die Standardeinstellung zurücksetzen möchten.

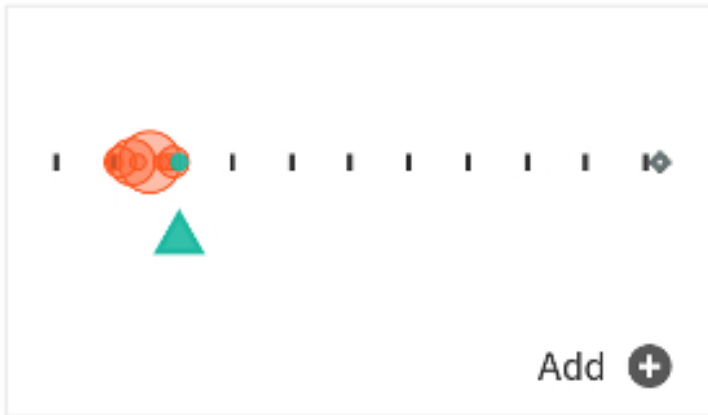
Vorgeschlagene Ausgangspunkte

Vorgeschlagene Ausgangspunktvisualisierungen werden als Piktogramme neben dem Beziehungsdiagramm angezeigt. Wählen Sie einzelne oder mehrere Knoten im Beziehungsdiagramm aus, um diese Visualisierungen zu erstellen.

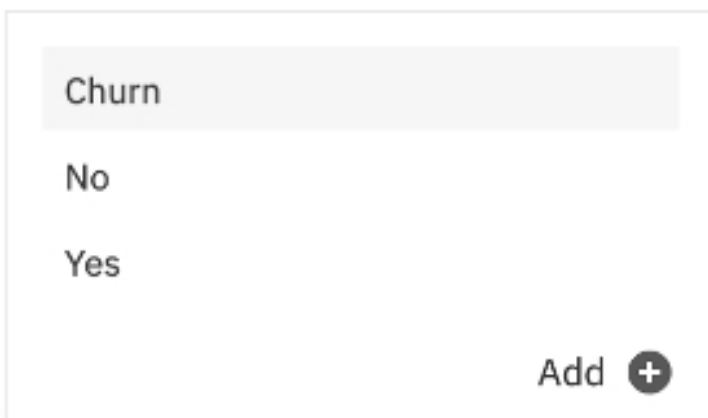
Klicken Sie auf eine Visualisierung, wenn Sie sie zu Ihrer Exploration hinzufügen und sie gleichzeitig

anzeigen möchten. Klicken Sie auf das Plusymbol  in der Ausgangspunktvisualisierung, um sie Ihrer Exploration hinzuzufügen und die aktuelle Ansicht beizubehalten.

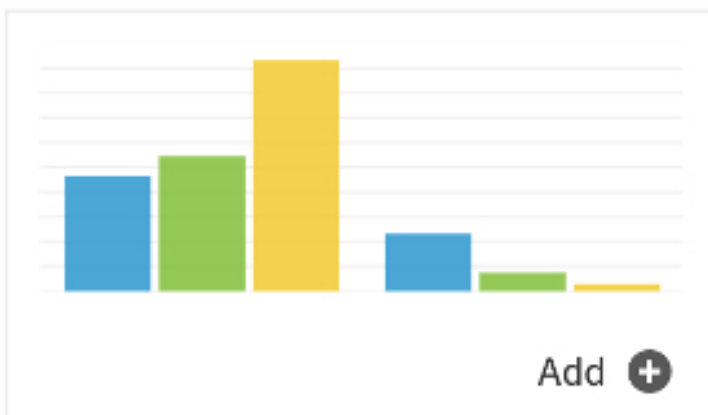
Explore visualizations related to 'Churn'



Churn



Churn



Tenure by Churn colored by Contract

Beziehungsdiagramm öffnen

Wenn Sie gerade eine Visualisierung anzeigen und zum Beziehungsdiagramm zurückkehren möchten, verwenden Sie die **Karte für Datenbeziehungen**, um zur Ausgangspunktansicht zurückzukehren.

Informationen zu diesem Vorgang

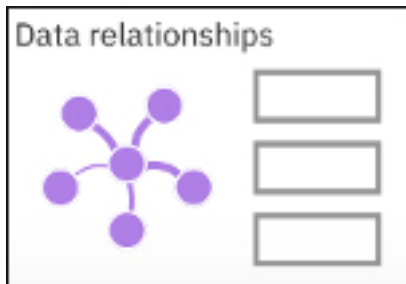
Führen Sie die folgenden Schritte aus, um zur Ausgangspunktansicht zurückzukehren, um ein Beziehungsdiagramm und die vorgeschlagenen Ausgangspunkte anzuzeigen.

Vorgehensweise

1. Klicken Sie auf das Symbol **Explorationen** im seitlichen Fenster.



2. Klicken Sie auf die **Karte für Datenbeziehungen**.



Tipp: Die **Karte für Datenbeziehungen** ist auch über das Menü **Neue Karte** in der Symbolleiste verfügbar.

Kapitel 3. Visualizations

Visualisierungen

Sie können den Visualisierungstyp ändern oder die Spalten, die in der Visualisierung verwendet werden, ändern.

Karten im Navigationsfenster anzeigen

Zeigen Sie Piktogramme Ihrer Visualisierungen, sogenannte Karten, im Navigationsfenster links der Hauptansicht an.

Informationen zu diesem Vorgang

Karten sind eine Sammlung von Visualisierungen in Ihrer Exploration. Verwenden Sie die Karten zum Öffnen Ihrer Visualisierungen, um Details anzuzeigen und sie unter Verwendung der Datenfelder zu ändern.

Vorgehensweise

1. Klicken Sie auf das Symbol **Explorationen** im seitlichen Fenster.



Ihre Karten werden hier aufgelistet.

2. Klicken Sie auf die Karte, die ein Visualisierungspiktogramm anzeigt, das Sie öffnen möchten.

Die Visualisierung wird in der Hauptansicht geöffnet.

3. Zeigen Sie den generierten Text unter der Registerkarte **Details** an oder fügen Sie weitere Datenelemente unter der Registerkarte **Felder** hinzu.

Wenn die Registerkarten **Details** und **Felder** nicht angezeigt werden, klicken Sie auf das Symbol für **Details und Felder anzeigen**.

Visualisierungsdetails anzeigen

Wenn Sie eine Visualisierung öffnen, wird sie im Hauptexplorationsbereich angezeigt. Im Explorationsfenster werden die Registerkarten **Details**, **Felder** und **Eigenschaften** angezeigt.

Wenn das Explorationsfenster nicht angezeigt wird, klicken Sie auf das Symbol **Details und Felder anzeigen**.

Visualisierungsdetails

In der Registerkarte **Details** wird Text angezeigt, der generiert wird, um Aspekte der in den Visualisierungen dargestellten Daten zu beschreiben. Diese Details sind beim Ansehen der Visualisierung nicht offensichtlich. Beispielsweise könnten die Details einen Durchschnitt der Werte über einen Zeitraum hinweg sichtbar machen.

Felder

In der Registerkarte **Felder** können Sie Spalten hinzufügen, um Visualisierungen zu erstellen und zu ändern. Fügen Sie für jedes obligatorische Feld eine Spalte hinzu.

Eigenschaften

Auf der Registerkarte **Eigenschaften** können Sie Eigenschaften ändern, die für Ihre Visualisierungen gelten.

Einzelne Visualisierung erstellen

Möglicherweise stellen Sie während der Arbeit mit Ihrer Exploration fest, dass Sie eine weitere Visualisierung benötigen.

Informationen zu diesem Vorgang

Führen Sie zum Erstellen einer einzelnen Visualisierung die folgenden Schritte aus.

Vorgehensweise

1. Wählen Sie in der Symbolleiste **Neue Karte** aus.
2. Wählen Sie die leere Karte **Einzel** aus.
3. Erweitern Sie im Fenster **Quellen** das Datenasset, das Sie verwenden möchten.

Wenn ein anderes Datenasset geöffnet ist, klicken Sie neben dem Namen des geöffneten Datenassets auf **Zurück**.



4. Führen Sie zum Erstellen einer neuen Visualisierung eine der folgenden Aktionen aus:
 - Ziehen Sie Datenelemente in den Bereich **Visualisierung erstellen** und legen Sie sie dort ab.
IBM® Cognos Analytics erstellt eine den Datenelementen entsprechende Visualisierung. Wenn Sie beispielsweise 'Jahr' oder 'Abteilung' hinzufügen, wird eine Tabelle erstellt. Ziehen Sie eine Kennzahl, beispielsweise 'Einnahmen', hinein, wird eine Balkendiagrammvisualisierung erstellt.
 - Klicken Sie auf **Typ auswählen** und wählen Sie einen Visualisierungstyp aus. Fügen Sie dann jedem Feld ein Datenelement hinzu.

Visualisierung mithilfe der Suche in Datenfeldern erstellen

Eine Visualisierung, bei der die Suche in Datenfeldern genutzt wird, kann ohne zeitliche Verzögerung erstellt werden.

Vorgehensweise

1. Klicken Sie auf das Symbol **Explorationen** im seitlichen Fenster.




Ihre Karten werden hier aufgelistet.


2. Klicken Sie auf die Karte, die ein Visualisierungspiktogramm anzeigt, das Sie öffnen möchten.
Die Visualisierung wird in der Hauptansicht geöffnet.
3. Klicken Sie auf die Registerkarte **Felder**.
Wenn die Registerkarte **Felder** nicht angezeigt wird, klicken Sie auf das Symbol für **Details und Felder anzeigen**.
4. Suchen Sie nach den gewünschten Daten. Klicken Sie auf die Daten, um das Datenfeld zu füllen.


Ergebnisse

Details **Fields** Properties



 Bars



City

 California Zip Website Visits Xlsx

 City

Length* Required field


 Blog Visits 

 Unique Visitors 


Click or drag data here


y-start

Click or drag data here

 Target

Click or drag data here

 Color

 Measures group (2)

Click or drag data here

Zwei Visualisierungen vergleichen



Sie können einen eigenen Vergleich erstellen, um die Daten zwischen zwei Visualisierungen zu analysieren. Alternativ können Sie mit einem empfohlenen Vergleich beginnen. In beiden Fällen wird eine Auswertung von Schlüsselinformationen und Unterschieden zwischen den beiden Visualisierungen generiert.

Informationen zu diesem Vorgang

Führen Sie zum Erstellen eines Vergleichs zwischen zwei Visualisierungen die folgenden Schritte aus.

Anmerkung: Wenn Sie eine neue Visualisierung erstellen, können Sie eine leere Vergleichskarte mit zwei Feldern für Visualisierungen auswählen.


Vorgehensweise

1. Klicken Sie im seitlichen Fenster auf das Symbol **Explorationen** .
Das Fenster **Karten** wird geöffnet.
2. Wählen Sie eine Karte für einen Vergleich aus.
Es wird eine Visualisierung angezeigt.
3. Klicken Sie in der Symbolleiste auf **Vergleichen**.
Daraufhin wird die Seite **Wie soll der Vergleich erfolgen?** angezeigt mit entsprechenden Anleitungen zum Erstellen eines eigenen Vergleichs bzw. zum Starten mit einer Empfehlung.
4. Klicken Sie auf das Symbol für **Karte hinzufügen**  in einem Kartenpiktogramm, um die Karte zur Liste von Karten im Navigationsfenster hinzuzufügen. Klicken Sie alternativ auf das Kartenpiktogramm, um die neue Karte hinzuzufügen und sie sofort anzuzeigen.
5. Ändern Sie optional die Daten in einer Visualisierung, um sie mit der anderen Visualisierung zu vergleichen.
 - a) Wählen Sie eine der beiden Visualisierungen aus.
 - b) Ändern Sie auf der Registerkarte **Felder** die Visualisierung auf eine der folgenden Weisen:
 - Filter entfernen.
 - Obere oder untere Werte anzeigen.
 - Datenelemente entfernen.
 - Im Teilfenster **Quellen** neue Datenelemente oder Filter hinzufügen. Oder die Datenfelder durchsuchen. Weitere Informationen finden Sie in [„Visualisierung mithilfe der Suche in Datenfeldern erstellen“](#) auf Seite 10.

Zwei Datenpunkte einer Visualisierung vergleichen

Sie können in einer bestehenden Visualisierung zwei Datenpunkte auswählen und die Daten vergleichen.

Vorgehensweise

1. Klicken Sie im seitlichen Fenster auf das Symbol **Explorationen** .
Das Fenster **Karten** wird geöffnet.
2. Wählen Sie im Fenster **Karten** die Karte aus, die ein Visualisierungspiktogramm anzeigt, das Sie öffnen möchten.
Die Visualisierung wird in der Hauptansicht geöffnet.
3. Wählen Sie in der Visualisierung zwei Datenpunkte aus.
4. Klicken Sie mit der rechten Maustaste und klicken Sie dann auf **Vergleichen anhand von**.
5. Geben Sie eine Spalte ein, um die beiden Datenpunkte zu vergleichen.
In einer Tabelle werden Informationen zum Vergleich der beiden Datenpunkte angezeigt.

Erweiterte Datenanalyse

IBMCognos Analytics ist ein Business-Intelligence-Tool zur Verwaltung und Analyse von Daten. Es enthält eine Reihe von Self-Service-Features, die es Benutzern ermöglichen, Daten vorzubereiten, zu erkunden und gemeinsam zu nutzen. Im Rahmen dieses Produktangebots umfasst Cognos Analytics eine Reihe von Prognose-, Beschreibungs- und Explorationsverfahren, die auch als 'Numeric Intelligence' bezeichnet werden. Cognos Analytics verwendet zahlreiche statistische Tests für die Datenanalyse. Es ist von zentraler Bedeutung, die Definitionen dieser Tests im Hinblick auf Cognos Analytics zu kennen.

Weitere Informationen finden Sie in der Veröffentlichung *IBM Cognos Analytics - Dashboards und Storys*.

Anderen Visualisierungstyp auswählen

Visualisierungen vermitteln Vergleiche, Beziehungen und Trends. Sie heben Zahlen hervor und gestalten sie transparenter. Berücksichtigen Sie beim Auswählen eines Visualisierungstyps, was Sie mit der Visualisierung veranschaulichen möchten und was die Zielgruppe der Visualisierung anspricht.

Vorbereitende Schritte

Weitere Informationen zu Visualisierungstypen finden Sie in der Dokumentation zur Visualisierung in der Veröffentlichung *IBM Cognos Analytics Dashboards und Storys - Benutzerhandbuch*.

Vorgehensweise

1. Wählen Sie im Fenster **Karten** die Karte aus, die für die Visualisierung steht, die Sie öffnen möchten.
2. Klicken Sie in der Symbolleiste auf das Symbol **Visualisierungstyp auswählen**.
3. Klicken Sie auf den Visualisierungstyp, den Sie verwenden möchten.

Beachten Sie, dass die einzelnen Visualisierungstypen Daten unterschiedlich vermitteln. Verwenden Sie beispielsweise eine Balken-, Säulen- oder Liniendiagrammvisualisierung zum Vergleichen einer Gruppe von Werten. Verwenden Sie eine Linien- oder Flächendiagrammvisualisierung zum Nachvollziehen von Beziehungen. Verwenden Sie eine Treemap- oder Kreisdiagrammvisualisierung, um die Teile eines Ganzen zu betrachten.

Flächendiagramm

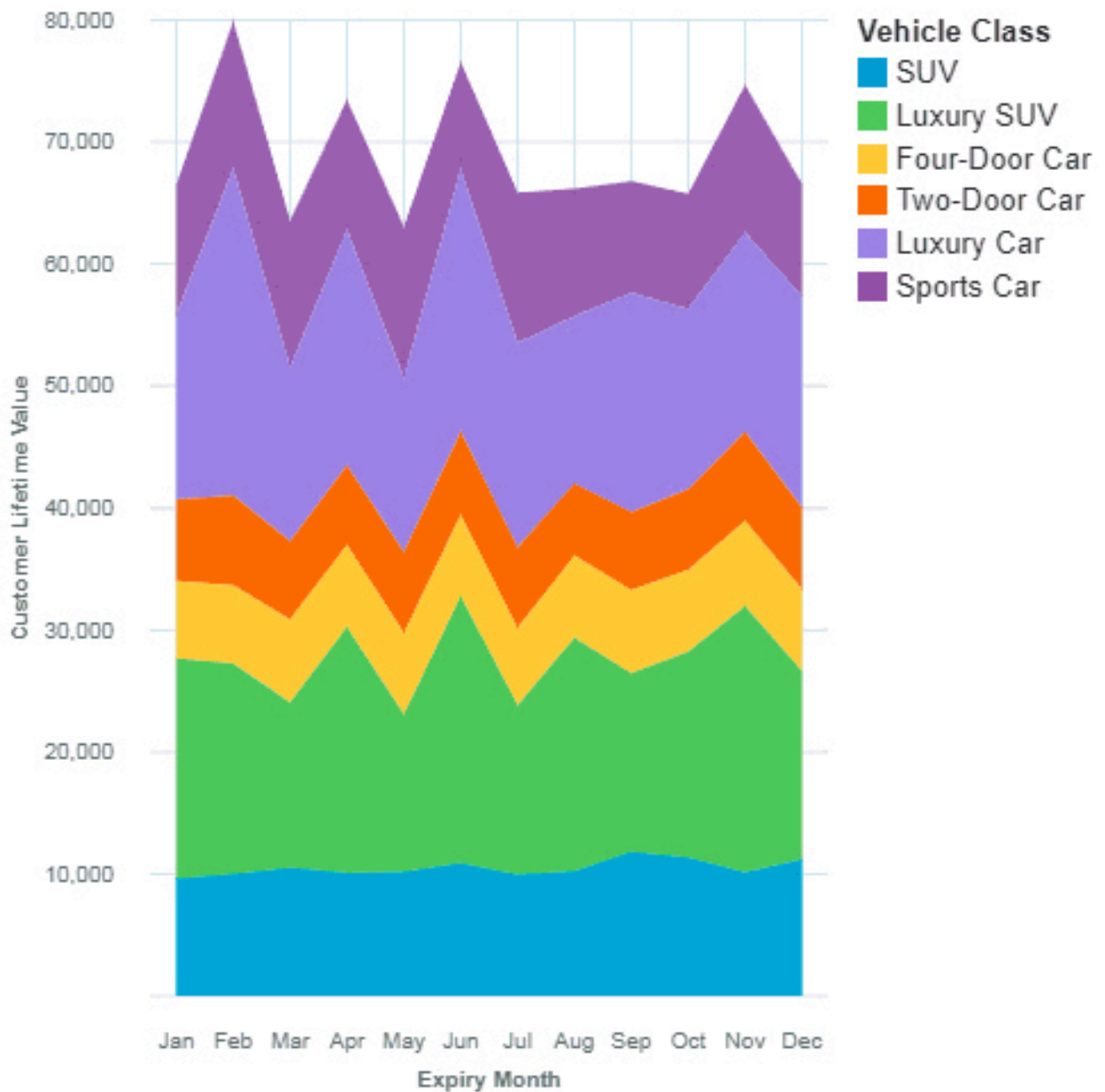
Verwenden Sie eine Flächendiagrammvisualisierung, um das Ausmaß von Veränderungen über einen Zeitraum hinweg hervorzuheben.

Flächendiagramme ähneln Liniendiagrammen, aber die Flächen unterhalb der Linien sind mit Farben oder Mustern gefüllt. Gestapelte Diagramme sind nützlich beim Vergleichen proportionaler Beiträge in einer Kategorie. Sie stellen den relativen Wert dar, den jede Datenserie zum Gesamtergebnis beiträgt.

Da in einer Flächendiagrammvisualisierung die Ergebnisse aller Spalten oder Elemente gestapelt werden, ist das Gesamtergebnis aller Ergebnisse leicht zu erkennen.

Eine Flächendiagrammvisualisierung eignet sich beispielsweise ausgezeichnet zum Darstellen der Einnahmen über einen bestimmten Zeitraum hinweg für verschiedene Produkte.

Diese Flächendiagrammvisualisierung zeigt den Kundenertragswert für jede Fahrzeugklasse pro Monat an. Da die Ergebnisse in der Flächendiagrammvisualisierung gestapelt werden, können Sie das Gesamtergebnis pro Monat sehen.



Die Flächendiagrammvisualisierung wurde erstellt, indem die folgenden Datenelemente aus der Anzeige 'Quellen' gezogen werden:

- Ziehen Sie den Typ **Ablaufmonat** in das Feld **X-Achse**.
- Ziehen Sie **Fahrzeugklasse** in das Feld **Farbe**.
- Ziehen Sie **Kundenertragswert** in das Feld **Y-Achse**.

Beispiele

Im Beispielbericht **Kundenertragswertanalyse** wird ein Beispiel für eine Visualisierung des Typs 'Wortwolke' angezeigt. Sie finden das Beispiel hier: **Teaminhalt > Beispiele > Berichte > Kundenertragswertanalyse**.

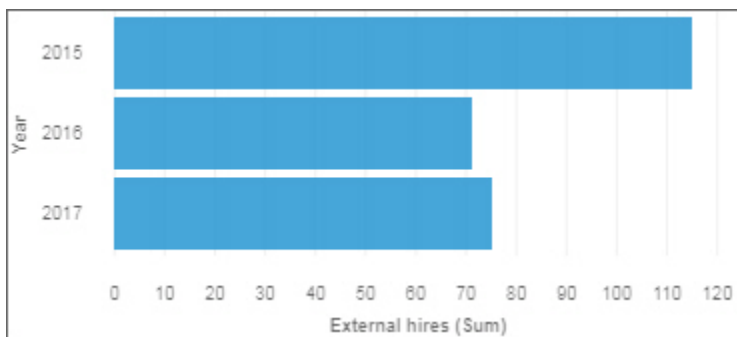
Wenn eines der Beispielobjekte fehlt, wenden Sie sich an Ihren Administrator.

Balkendiagramm

Verwenden Sie eine Balkendiagrammvisualisierung zum Vergleichen von Werten für eine oder mehrere Spalten, wie zum Beispiel den Umsatz für Produkte oder den Umsatz für Produkte pro Monat.

Balkendiagrammvisualisierungen verwenden horizontale Datenpunkte, die in Gruppen angeordnet sind, um einzelne Werte zu vergleichen. Sie können Balkendiagrammvisualisierungen zum Vergleichen von diskreten Daten oder zur Darstellung von Trends über einen bestimmten Zeitraum hinweg verwenden.

Einer Balkendiagrammvisualisierung können Veränderungen über einen bestimmten Zeitraum hinweg entnommen werden oder sie kann zum Vergleichen und zur Gegenüberstellung zweier oder mehrerer Spalten über einen bestimmten Zeitraum hinweg verwendet werden. Falls es so viele Balken gibt, dass die Beschriftung nicht mehr lesbar ist, filtern Sie die Daten, um sich auf ein Subset der Daten zu konzentrieren, oder verwenden Sie eine Treemap.



Verwenden Sie das Feld **Ziel** zur Darstellung von Kennzahlen, die mit einem Zielwert verglichen werden sollen.

Verwenden Sie das Feld **y-start**, um zu definieren, wo die Kennzahl beginnen muss.

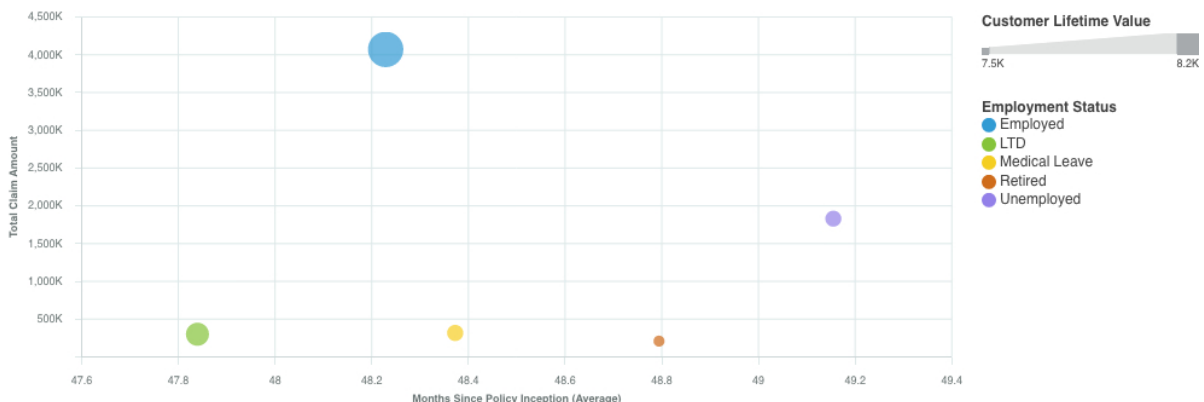
Blasendiagramm

Verwenden Sie eine Blasendiagrammvisualisierung, um Beziehungen zwischen Spalten darzustellen, die numerische Werte enthalten, wie beispielsweise Einnahmen und Gewinn.

Eine Blasendiagrammvisualisierung verwendet Datenpunkte und Blasen, um Kennzahlen an beliebigen Stellen entlang einer Skala darzustellen. Entlang jeder Achse wird jeweils eine Kennzahl dargestellt. Die Größe der Blase stellt eine dritte Kennzahl dar. Verwenden Sie Blasendiagrammvisualisierungen zur Darstellung von Finanzdaten oder anderen Daten mit zusammengehörigen Kennzahlenwerten.

Die Blasen haben verschiedene Größen und Farben. Die X-Achse stellt eine Kennzahl dar. Die Y-Achse stellt eine weitere Kennzahl dar, und die Größe der Blasen stellt die dritte Kennzahl dar. In dem unten gezeigten Beispiel wird die Farbe durch eine Kennung dargestellt.

Das gezeigte Beispiel stellt die Monate seit der Konzeption der Richtlinie dar.





Erstellen Sie die Blasendiagrammvisualisierung, indem Sie die folgenden Datenelemente aus **Kunden-**

analyse in die Anzeige **Quellen**  ziehen:

- Ziehen Sie **Monate seit Konzeption der Richtlinie** in das Feld **X-Achse**.
- Ziehen Sie **Gesamter Anspruchsbetrag** in das Feld **Y-Achse**.
- Ziehen Sie **Kundenertragswert** in das Feld **Größe**.
- Ziehen Sie **Beschäftigungsstatus** in das Datenfeld **Farbe**.

Sie können das Blasendiagramm anpassen. Führen Sie die folgenden Schritte aus, damit die X-Achse des Blasendiagramms beispielsweise so wie im Beispiel dargestellt wird:

1. Klicken Sie auf die Visualisierung, und klicken Sie dann in der Anzeige **Daten** auf das Datenelement **<Gesamter Anspruchsbetrag>**.
2. Klicken Sie auf .
3. Klicken Sie neben **Datenformat** auf  und legen Sie die folgenden Optionen fest:
 - **Formattyp:** Währung
 - **Währungssymbol:** K
 - **Position des Währungssymbols:** Ende
 - **Anzahl der Dezimalstellen:** 0
 - **Skalierung:** -3 (zeigt Werte in Tausenden an).
4. Klicken Sie auf **OK**.

Wenn Sie die Größe der Visualisierung ändern möchten, klicken Sie auf die Visualisierung und legen Sie in der Anzeige "Eigenschaften" die folgende Option fest:

- **Größe - Breite:** 700 px, **Höhe:** 300 px

Klicken Sie auf , um die Anzeige **Eigenschaften** zu schließen.

Beispiele

Beispiele für Visualisierungen finden Sie im Beispielbericht **Analyse des Kundenlebensdauerwerts**. Die Beispiele finden Sie hier: **Teaminhalt > Beispiele > Berichte > Analyse des Kundenlebensdauerwerts**.

Wenn eines der Beispielobjekte fehlt, wenden Sie sich an Ihren Administrator.

Bullet-Diagramm

Verwenden Sie Bullet-Diagramme zur Darstellung von Kennzahlen, die mit einem Zielwert verglichen werden sollen.

In einem Call-Center kann ein Bullet-Diagramm zur Messung von Metriken wie der Anzahl der Anrufe, der Dauer bis zur Annahme von Anrufen und dem Prozentsatz erfolgloser Anwahlversuche verwendet werden.

In der Fertigung kann ein Bullet-Diagramm zur Überwachung von Metriken wie der Anzahl von Mängeln und von ausgelieferten Bestellungen verwendet werden.

Im Fitnessbereich kann ein Bullet-Diagramm zur Messung von Metriken wie der Anzahl zurückgelegter Schritte und verbrannter Kalorien verwendet werden.

Bullet-Diagrammvisualisierungen vergleichen eine tatsächliche Kennzahl (das Bullet) mit einer angestrebten Kennzahl (dem Ziel). Bullet-Diagrammvisualisierungen stellen außerdem die miteinander verglichenen Kennzahlen in Beziehung zu farblichen Markierungen im Hintergrund, die qualitative Aspekte einbringen, wie beispielsweise gut, zufriedenstellend und schlecht. Auch mit Bullet-Diagrammvisualisierungen in geringerer Größe können Informationen noch effektiv dargestellt werden.

Eine Bullet-Diagrammvisualisierung beschäftigt sich mit einer einzigen primären Kennzahl. Zum Beispiel die im laufenden Jahr aufgelaufenen Einnahmen. Und sie vergleicht diese Kennzahl mit einer oder mehre-

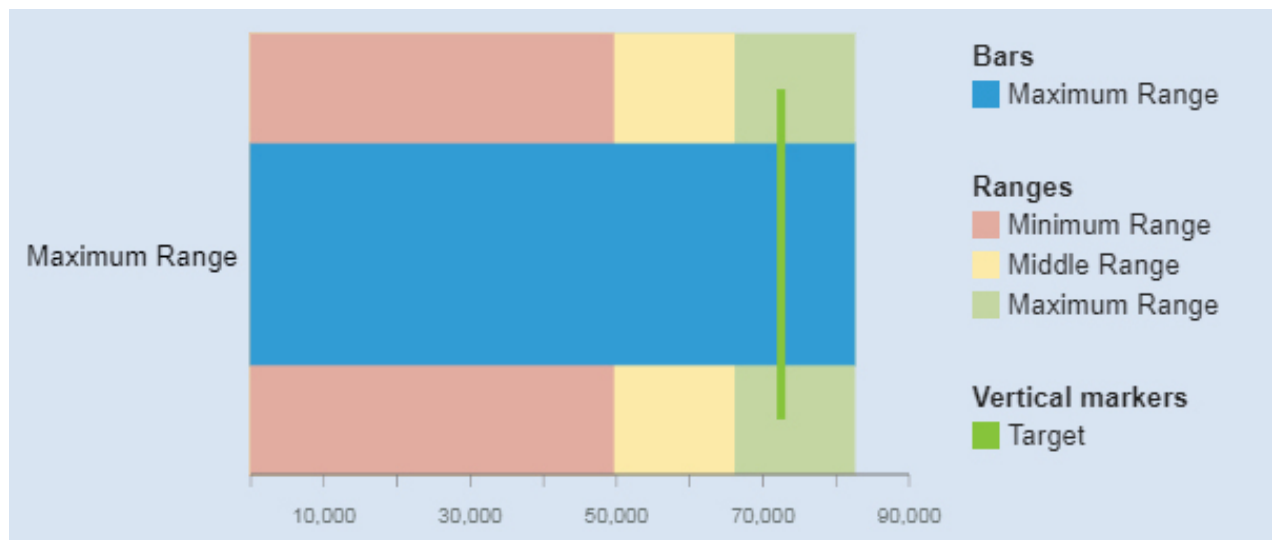
ren anderen Kennzahlen, um die Aussagekraft zu erhöhen. Zum Beispiel beim Vergleich mit einem Ziel. Die primäre Kennzahl wird im Kontext eines qualitativen Leistungsbereichs angezeigt, wie beispielsweise schlecht, zufriedenstellend und gut.

Wenn Sie eine Bullet-Diagrammvisualisierung auswählen, geben Sie die folgenden Felder an:

- Das Feld **Balken mit tatsächlichen Werten** gibt die tatsächliche Kennzahl an.
- Das Feld **Ziel** gibt die Zielkennzahl an.
- Das Feld **Bereich - Minimum** gibt den minimalen qualitativen Bereich an.
- Das Feld **Bereich - Mittel** gibt den mittleren qualitativen Bereich an.
- Das Feld **Bereich - Maximum** gibt den höheren qualitativen Bereich an.

Anmerkung: Drillthrough ist für Bullet-Diagrammvisualisierungen nicht verfügbar.

Stellen Sie sicher, dass die Bereiche Minimum, Mittel und Maximum sowohl für die tatsächliche als auch für die Zielkennzahl gelten.



Die Bullet-Diagrammvisualisierung wurde erstellt, indem die folgenden Datenelemente aus der Anzeige 'Quellen' gezogen werden:

- Ziehen Sie **Bereich - Minimum** in das Feld **Bereich - Minimum**.
- Ziehen Sie **Bereich - Minimum** in das Feld **Bereich - Minimum**.
- Ziehen Sie **Bereich - Maximum** in das Feld **Bereich - Maximum**.
- Ziehen Sie **Bereich - Maximum** in das Feld **Balken mit tatsächlichen Werten**.
- Ziehen Sie **Zielwert** in das Feld **Zielwert**.
- Ziehen Sie **Fahrzeugklasse** in das Feld **Zusätzliche Daten**.

Beispiele

Im Beispielbericht **Kundenertragswertanalyse** wird ein Beispiel für eine Bullet-Diagrammvisualisierung angezeigt. Sie finden das Beispiel hier: **Teaminhalt > Beispiele > Berichte > Kundenertragswertanalyse**.

Wenn eines der Beispielobjekte fehlt, wenden Sie sich an Ihren Administrator.

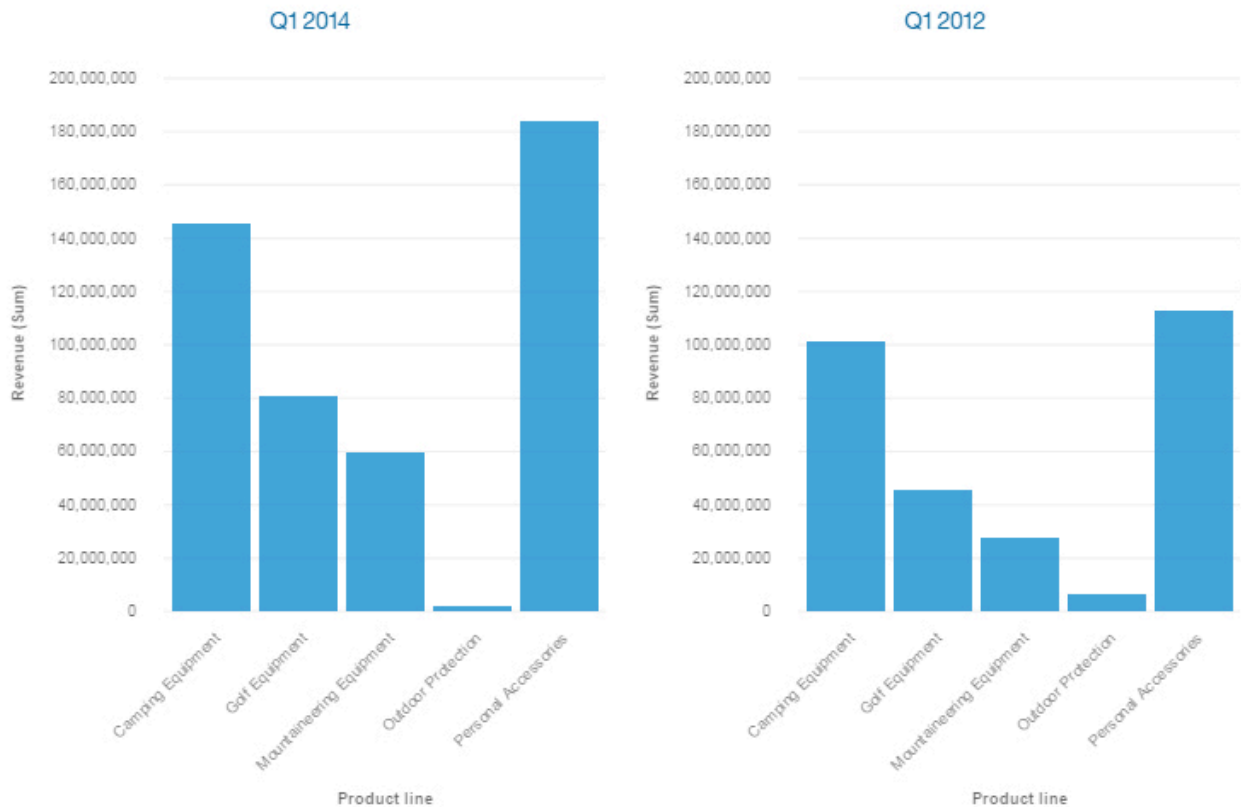
Säulendiagramm

Verwenden Sie eine Säulendiagrammvisualisierung zum Vergleichen von Werten für eine oder mehrere Spalten, wie zum Beispiel den Umsatz für Produkte oder den Umsatz für Produkte pro Monat.

Säulendiagrammvisualisierungen verwenden vertikale Datenpunkte, die in Gruppen angeordnet sind, um einzelne Werte zu vergleichen. Verwenden Sie Säulendiagrammvisualisierungen zum Vergleichen von diskreten Daten oder zur Darstellung von Trends über einen bestimmten Zeitraum hinweg.

Einer Säulendiagrammvisualisierung können Veränderungen über einen bestimmten Zeitraum hinweg entnommen werden oder sie kann zum Vergleichen und zur Gegenüberstellung zweier oder mehrerer Spalten über einen bestimmten Zeitraum hinweg verwendet werden. Falls es so viele Balken gibt, dass die Beschriftung nicht mehr lesbar ist, filtern Sie die Daten, um sich auf ein Subset der Daten zu konzentrieren, oder verwenden Sie eine Treemap.

Beispiel: Die Einnahmen für jede Produktlinie werden nach Quartal gruppiert, sodass die Leistung pro Quartal hervorgehoben wird.



Verwenden Sie das Feld **Ziel** zur Darstellung von Kennzahlen, die mit einem Zielwert verglichen werden sollen.

Verwenden Sie das Feld **y-start**, um zu definieren, wo die Kennzahl beginnen muss.

Kreuztabelle

Verwenden Sie eine Kreuztabelle zur Darstellung der Beziehungen zwischen drei oder mehr Spalten. In Kreuztabellen werden Daten in Zeilen und Spalten mit einer Zusammenfassung der Informationen an den Schnittpunkten dargestellt.

In dieser Kreuztabelle werden beispielsweise die Kursgebühren für jede Abteilung pro Organisation dargestellt.

Course cost	GO Asia Pacific corporate						
	Executive Offices	Finance	Sales (Corporate)	Procurement	Human Resources	Marketing	Summary
2014	10,500	29,750	10,500	18,250	11,000	15,500	95,500
2015	11,500	46,250	14,750	7,750	11,500	30,750	122,500
2016	12,500	43,000	33,250	5,000	14,500	22,500	130,750
2017	13,500	24,000	36,000	7,000	15,750	21,750	118,000
Summary	48,000	143,000	94,500	38,000	52,750	90,500	466,750

Ab Cognos Analytics Version 11.1.4 können Sie Daten aus der Anzeige **Ausgewählte Quellen** ziehen und Sie können Daten in eine Zeile/Spalte einfügen oder die Daten auf vorhandenen Daten ablegen, um diese zu ersetzen.

Data Player

Sie können einen Data Player verwenden, um eine Animation der Auswirkung eines Säulendiagramms auf die anderen Visualisierungen anzuzeigen.

Entscheidungsstruktur

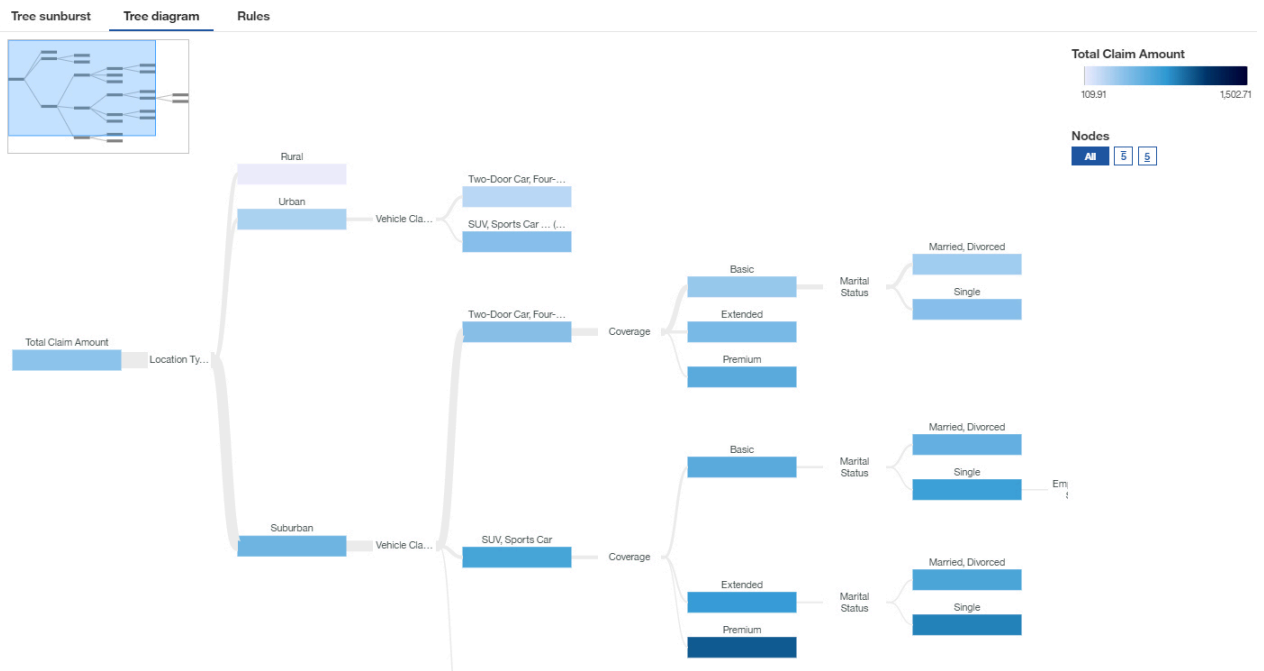
In einer Entscheidungsstruktur werden die Werte von Datensätzen in einer verbundenen Hierarchie von Rechtecken dargestellt.

Die Datensätze werden in Gruppen aufgeteilt, die als Knoten bezeichnet werden. Jeder Knoten enthält Datensätze, die sich bezüglich des Zielfelds statistisch ähnlich sind. Ein Knoten kann beispielsweise die Datensätze für Männer enthalten, die eine Ausbildung von mehr als 18 Jahren haben. Knoten können dann verwendet werden, den Feldwert eines Ziels zu prognostizieren. So könnte zum Beispiel der Knoten über Männer und Ausbildung zur Prognose des Gehalts verwendet werden.

Jeder Zweig in einer Entscheidungsstruktur entspricht einer Entscheidungsregel. Weitere Informationen zu Entscheidungsregeln finden Sie in [„Entscheidungsregeln anzeigen“](#) auf Seite 23.

Um die Leistung zu verbessern, basiert die Analyse aufgrund der Anzahl der Zeilen in der Datenquelle auf einer repräsentativen Stichprobe der gesamten Daten.

Eine Entscheidungsstrukturvisualisierung kann beispielsweise wie folgt aussehen:



Anmerkung: Filter werden für Entscheidungsstrukturvisualisierungen nicht unterstützt.

Weitere Informationen finden Sie in „Entscheidungsstrukturvisualisierung untersuchen“ auf Seite 20.

Entscheidungsstrukturvisualisierung untersuchen

Eine Entscheidungsstrukturvisualisierung wird verwendet, um darzustellen, wie die zugrunde liegenden Daten ein gewähltes Ziel prognostizieren, und hebt wesentliche Insights über die Entscheidungsstruktur hervor.

Informationen zu diesem Vorgang


Die Prognosestärke einer Entscheidungsstruktur bestimmt den Grad, zu dem die Entscheidungen, die durch die einzelnen in der Struktur angezeigten Zweige dargestellt werden, den Wert des Ziels prognostizieren.

Entscheidungsstrukturen haben nur ein einziges Ziel. Ist das Zielfeld der Entscheidungsstruktur stetig, heben die Anzeiger für wesentliche Insights ungewöhnlich hohe oder niedrige Gruppen hervor. Ist das Zielfeld der Entscheidungsstruktur kategorial, ist das wesentliche Insight der Modus des Knotens. Der Modus des Knotens ist/sind die am häufigsten auftretende(n) Kategorie(n) des Zielfelds innerhalb der Gruppe.

Um die Leistung zu verbessern, basiert die Analyse aufgrund der Anzahl der Zeilen in der Datenquelle auf einer repräsentativen Stichprobe der gesamten Daten.

Beim Prüfen einer Entscheidungsstruktur:

- Wenn Sie alle Einflussfaktoren sehen möchten, verwenden Sie entweder die Registerkarte **Verzeichnisstruktur - Diagramm** oder die Registerkarte **Regeln**.
- Wenn Sie sich auf die zentralen Einflussfaktoren konzentrieren möchten, verwenden Sie die Registerkarte **Verzeichnisstruktur - Sunburst-Diagramm**.

Wenn Sie zentrale Einflussfaktoren bearbeiten oder hinzufügen möchten, klicken Sie auf  im Zielfeld.

Insights sind unterschiedlich in Abhängigkeit vom Typ Ihres Ziels. Wenn Sie eine stetige Kennzahl prognostizieren, zum Beispiel Einkommen, Alter oder Gewinn, zeigt die Entscheidungsstruktur innerhalb des Knotens den Durchschnittswert des Ziels an unter den bisherigen Bedingungen innerhalb der Gruppe, die durch den Knoten dargestellt wird. Nehmen Sie als Beispiel eine Struktur, die das Einkommen prog-

nostiziert und einen Zweig für Geschlecht und dann Stadt hat. Wenn Sie dem Pfad von 'Männlich' nach 'Chicago' folgen, enthält der Knoten 'Chicago' als Wert das Durchschnittseinkommen von Männern in Chicago.

Vorgehensweise

1. Wenn Sie eine stetige Kennzahl haben, wird die Entscheidungsstruktur durch das folgende Beispiel erläutert.

Die Farbe gibt an, ob der Wert des Knotens mit hohen, mittleren oder niedrigen Werten des Ziels verbunden ist. Die Farbe des Knotens basiert auf dem Durchschnitt des Ziels für die Kennzahl. Je höher der Durchschnittswert des Ziels für einen Knoten ist, desto dunkler ist die Farbe.

Das nächste Beispiel ist die detaillierte Visualisierung für Restaurantausgaben. Das internationale Flughafenterminal ist ein starker Prädiktor für hohe Restaurantausgaben durch Geschäftsreisende. Unzureichende Sauberkeit am Flughafen ist ein Prädiktor für niedrige Ausgaben durch reisende Konferenz-/Tagungsteilnehmer.

Das Minifenster hilft Ihnen beim Durchblättern der Bereiche der Struktur. Das Minifenster ist besonders nützlich, wenn es viele Knoten gibt.

In diesem Beispiel sind die fünf höchsten Zielwerte durch eine Zahl gekennzeichnet. Sie können unter den folgenden Optionen wählen:

- Vollständige Struktur. Höchste oder niedrigste Werte werden nicht besonders gekennzeichnet.



- Fünf höchste Zielwerte. Die fünf höchsten Zielwerte werden angezeigt.

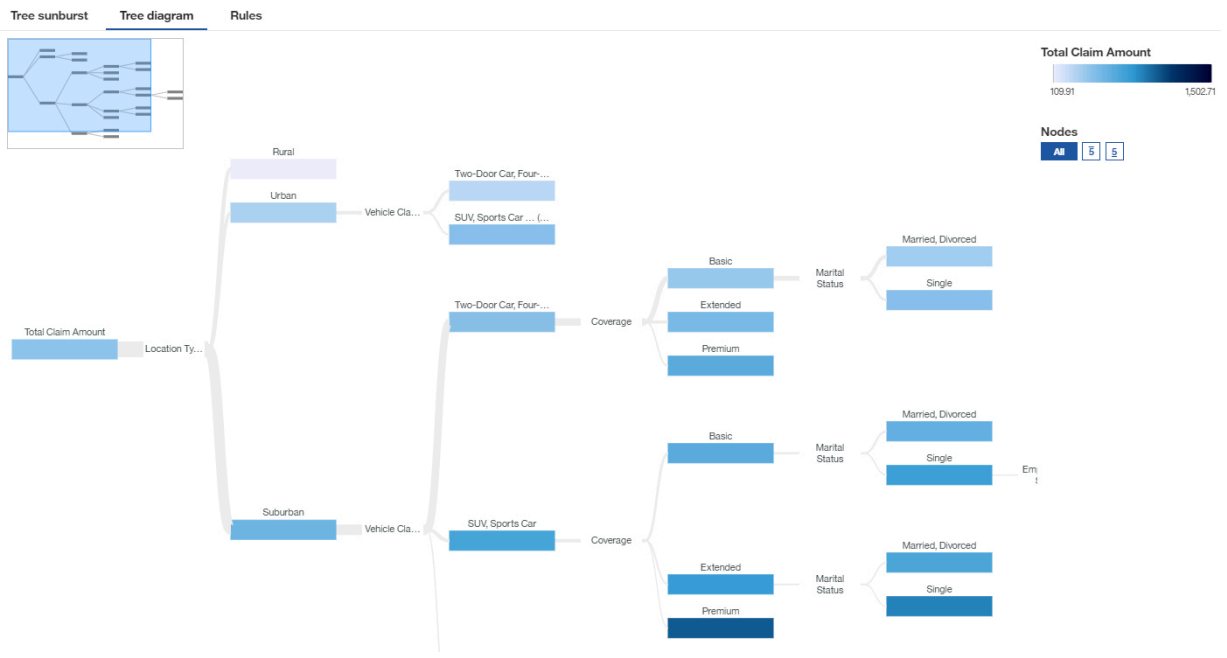


- Fünf niedrigste Zielwerte. Die fünf niedrigsten Zielwerte werden angezeigt.



Wenn Sie eine kategoriale Kennzahl haben, wählen Sie die Kategorie, für die Sie die fünf höchsten oder niedrigsten Ziele anzeigen möchten, über das Menü **Höchste 5 Knoten für:** bzw. **Niedrigste 5 Knoten für:** aus.

Falls Sie die Ansicht zu stark vergrößert haben, sind die obersten bzw. untersten fünf Knoten nicht sichtbar.

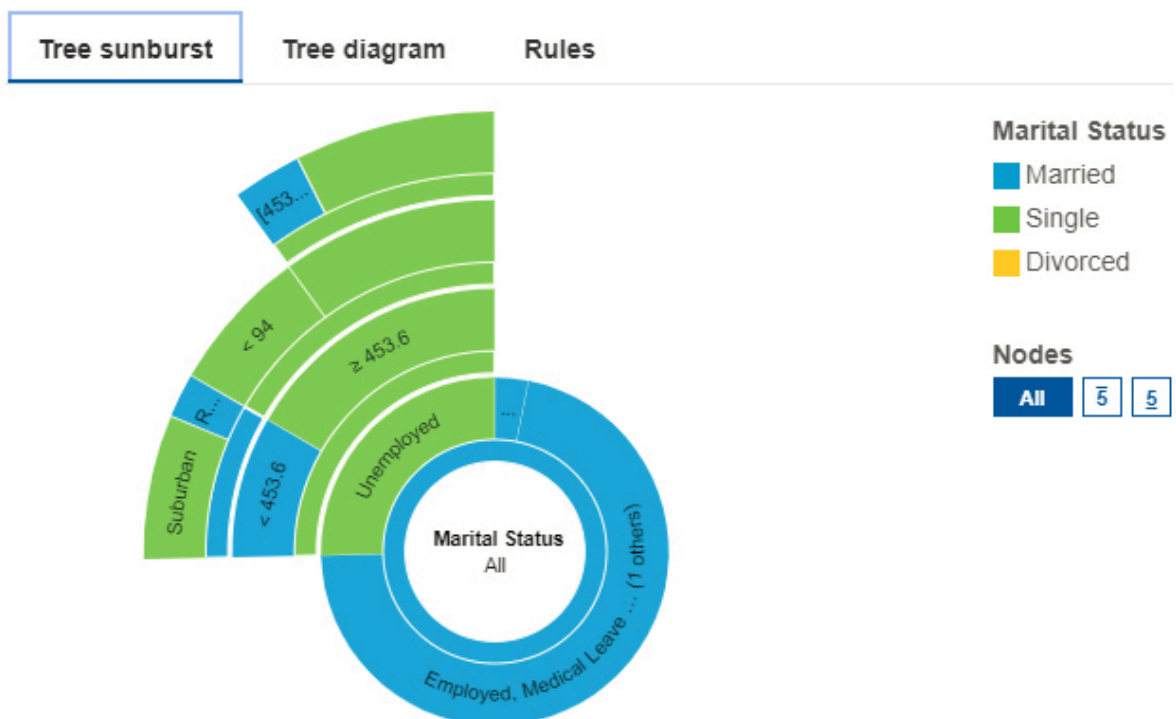


2. Wenn Sie eine kategoriale Kennzahl haben, wird die Entscheidungsstruktur durch das folgende Beispiel erläutert.

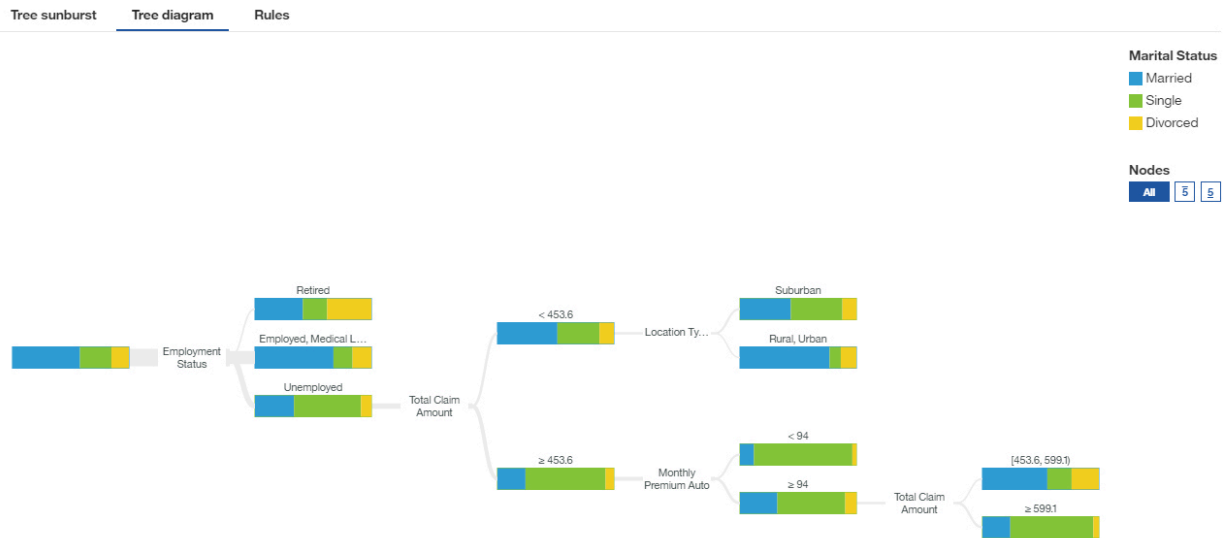
Die Farbe zeigt an, welche(r) Feldwert(e) am häufigsten vorkommt/vorkommen.

In der Registerkarte **Verzeichnisstruktur - Sunburst-Diagramm** können Sie sehen, dass, wenn die Kennzahlen innerhalb der Entscheidungsstruktur starke Prädiktoren für einen Zielwert oder Zielwerte sind, die Farben in dem betreffenden Knoten überwiegen. Die unbedeutenden Werte werden weggelassen.

Das nächste Beispiel ist die detaillierte Visualisierung des Familienstands in der Registerkarte **Verzeichnisstruktur - Sunburst-Diagramm**. Sie zeigt, dass Berufstätigkeit ein starker Prädiktor für das Verheiratetsein ist.



In der Registerkarte **Verzeichnisstruktur - Diagramm** stellen die Knoten grafisch die Verteilung der Personen nach Familienstand dar.



Entscheidungsregeln anzeigen

Eine Entscheidungsregel prognostiziert ein Ergebnis im Zielfeld. Das Anzeigen der Entscheidungsregeln hilft Ihnen festzustellen, welche Bedingungen voraussichtlich zu einem bestimmten Ergebnis führen.

Denken Sie beispielsweise an einige hypothetische Entscheidungsregeln, die die Abwanderung von Kunden vorhersagen könnten. Diese Regeln könnten Klassifikationen angeben, die auf den Bereichen Kundenalter und Anzahl früherer Reklamationen basieren. Anhand dieser Regeln könnten Sie möglicherweise feststellen, dass bei Kunden mit keiner oder 1 Reklamation und einem Alter über 50 eine erhöhte Abwanderungswahrscheinlichkeit vorliegt.

Jeder Zweig in einer Entscheidungsstruktur entspricht einer Entscheidungsregel.

Vorgehensweise

1. Tippen Sie in einer Entscheidungsstruktur **Regeln** an.
2. Überprüfen Sie die Entscheidungsregeln.
3. Tippen Sie **Verzeichnisstruktur - Diagramm** an, um zur Visualisierung zurückzukehren.

Analyse der Einflussfaktoren


Eine Visualisierung der Analyse der Einflussfaktoren zeigt Ihnen die wichtigsten Einflussfaktoren (oder Prädiktoren) für ein Ziel. Je weiter rechts sich der Einflussfaktor befindet, desto stärker ist er.

IBMCognos Analytics verwendet hoch entwickelte Algorithmen, mit deren Hilfe sehr gut interpretierbare Insights gewonnen werden können, die auf komplexen Modellierungen basieren. Sie müssen nicht wissen, welche statistischen Tests mit Ihren Daten ausgeführt werden müssen. Cognos Analytics wählt die richtigen Tests für die Daten aus.

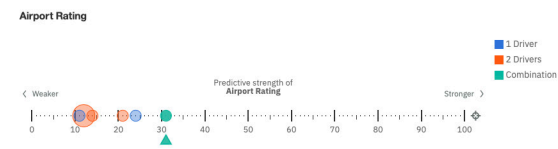
Wichtige Einflussfaktoren für sowohl fortlaufende als auch kategoriale Ziele sind in der Analyse der Einflussfaktoren in Dashboards und Explorationen verfügbar.

Weitere Informationen finden Sie in den Informationen zu *statistischen Tests* im *IBM Cognos Analytics Dashboards und Storys - Benutzerhandbuch*.

Die Visualisierung dieser Einflussfaktoranalyse zeigt beispielsweise, dass die Kombination der Gesamtzufriedenheit, der Bewertung von Beschilderungen, der Bewertung der Sicherheit und der Bewertung der Gestaltung die stärksten Einflussfaktoren des Ziels 'Flughafenbewertung' sind.

Wenn Sie zentrale Einflussfaktoren bearbeiten oder hinzufügen möchten, klicken Sie auf  im Zielfeld.

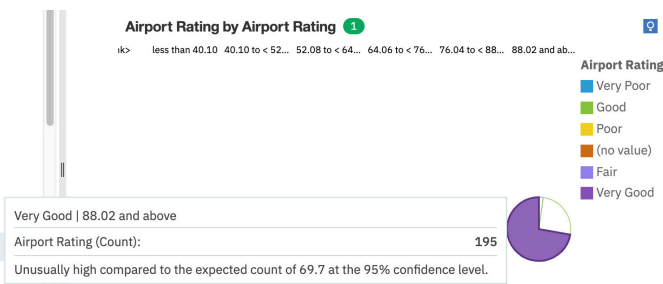
Um die Leistung zu verbessern, basiert die Analyse aufgrund der Anzahl der Zeilen in der Datenquelle auf einer repräsentativen Stichprobe der gesamten Daten.



Wenn Sie den Cursor über einen Datenpunkt bewegen, zeigt die Visualisierung der Einflussfaktoranalyse die Einflussfaktoren der Flughafenbewertung insgesamt.



Wenn Sie auf einen Datenpunkt in der Verzeichnisstruktur klicken, werden andere empfohlene Visualisierungen angezeigt.



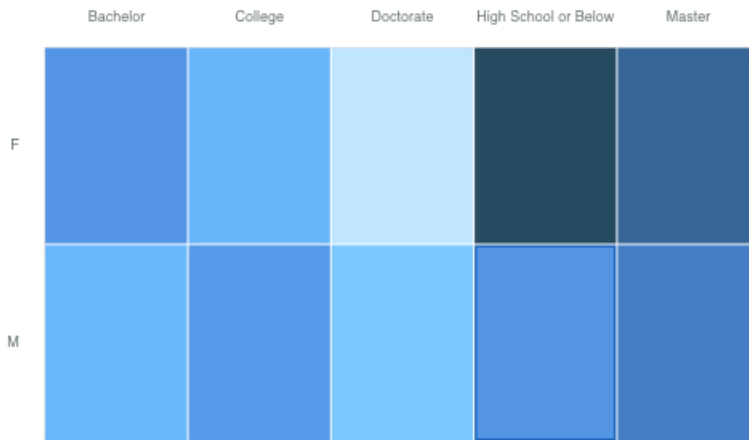
Anmerkung: Filter werden für Visualisierungen von Analysen von Einflussfaktoren nicht unterstützt.

Heat-Map

Verwenden Sie eine Heat-Map-Visualisierung zur Darstellung der Beziehung zwischen Spalten in einer matrixartigen Ansicht.

In einer Heat-Map-Visualisierung werden Farben und Farbintensitäten zur Darstellung der Beziehung zwischen zwei Spalten verwendet.

In dieser Heat-Map-Visualisierung wird der durchschnittliche Kundenertragswert nach Geschlecht und Bildungsabschluss dargestellt.



Erstellen Sie die Heat-Map-Visualisierung, indem Sie die folgenden Datenelemente aus der Anzeige **Quel-**

len  ziehen:

- Ziehen Sie **Geschlecht** in das Feld **Zeilen**.
- Ziehen Sie **Bildungsabschluss** in das Feld **Spalten**.
- Ziehen Sie **Kundenertragswert** in das Feld **Heat-Map**.

Beispiele

Beispiele für Visualisierungen finden Sie im Beispielbericht **Analyse des Kundenlebensdauerwerts**. Die Beispiele finden Sie hier: **Teaminhalt > Beispiele > Berichte > Analyse des Kundenlebensdauerwerts**.

Wenn eines der Beispielobjekte fehlt, wenden Sie sich an Ihren Administrator.

Hierarchie

Verwenden Sie eine Hierarchie, wenn Sie die Daten in Zeilen und Spalten angeordnet sehen möchten.

In dieser Hierarchie werden beispielsweise Produkttypen angezeigt.

Binoculars

Climbing Accessories

Cooking Gear

Eyewear

First Aid

Golf Accessories

Insect Repellents

Irons

Knives

Lanterns

Navigation

Packs

Putters

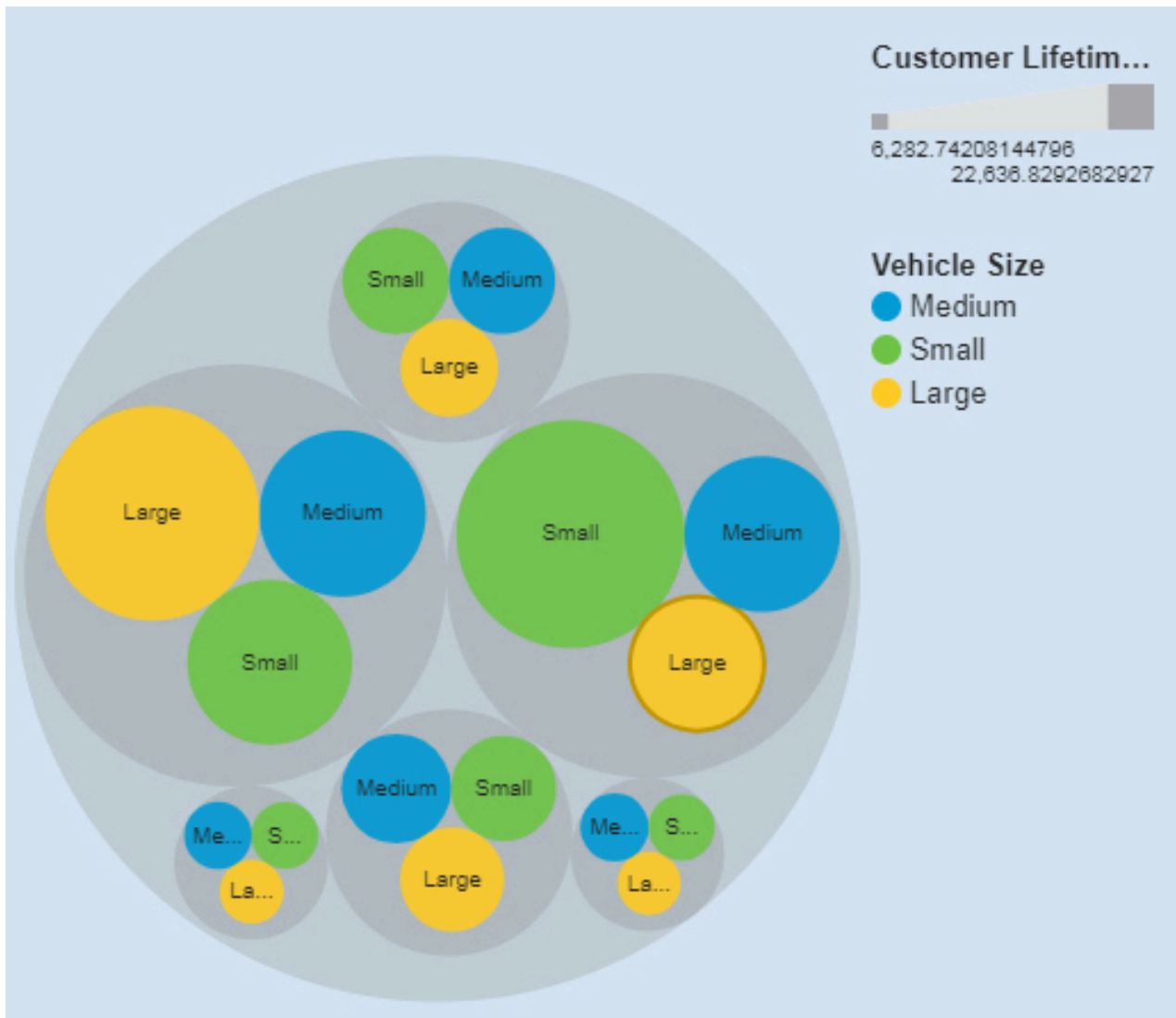
Blasendiagramm mit Hierarchien

Verwenden Sie eine Visualisierung eines Blasendiagramms mit Hierarchien, wenn Sie die Beziehungen zwischen Spalten, die Werte enthalten, wie beispielsweise einen Nettoverlust, darstellen möchten. Sie ist ähnlich der Blasendiagrammvisualisierung, aber die Blasen sind dicht gedrängt und nicht über ein Raster verteilt dargestellt. Die Blasen sind verschachtelt, um die Hierarchie darzustellen. Eine Visualisierung eines Blasendiagramms mit Hierarchien zeigt eine große Menge an Daten auf engem Raum.

Die Größe jeder Blase stellt die quantitative Dimension jedes Datenpunkts dar. Es werden zahlreiche Ebenen innerhalb einer Hierarchie und Beziehungen zwischen Gruppen dargestellt, basierend auf zugeordneten Attributen. Blasengröße und -farbe werden verwendet, um vergleichende Informationen zu Kategorien zu vermitteln.

Die Blasen haben verschiedene Größen und Farben.

Diese Visualisierung des Typs 'hierarchisches kompaktes Blasendiagramm' zeigt den Kundenertragswert nach Fahrzeugklasse pro Fahrzeuggröße an. Jede Blase steht für einen anderen Fahrzeugklasse in einer der drei Fahrzeuggrößen. Die Größe einer Blase wird durch den Kundenertragswert der betreffenden Fahrzeugklasse bestimmt. Die Farben der Blasen stehen für die Fahrzeuggröße.



Die Visualisierungen des Typs 'hierarchisches kompaktes Blasendiagramm' wurde erstellt, indem die folgenden Datenelemente aus der Anzeige 'Quellen' gezogen werden:

- Ziehen Sie **Fahrzeugklasse** und **Fahrzeuggröße** in das Feld **Blasen**.
- Ziehen Sie **Kundenertragswert** in das Feld **Größe**.
- Ziehen Sie **Fahrzeuggröße** in das Feld **Farbe**.

Beispiele

Im Beispielbericht **Kundenertragswertanalyse** wird ein Beispiel für eine Visualisierung des Typs 'Wortwolke' angezeigt. Sie finden das Beispiel hier: **Teaminhalt > Beispiele > Berichte > Kundenertragswertanalyse**.

Wenn eines der Beispielobjekte fehlt, wenden Sie sich an Ihren Administrator.

KPI

Mithilfe einer KPI-Visualisierung können Sie einen KPI (Key Performance Indicator) anzeigen, der zwei verwandte Kennzahlen enthält, zum Beispiel Umsatz und geplanter Umsatz. Optional können Sie eine Wortgrafik und eine aussagekräftige Form in Ihren KPI-Visualisierungen anzeigen.

In einer KPI-Visualisierung wird ein Basiswert mit einem Zielwert verglichen und es wird die Varianz zwischen den beiden Kennzahlen angezeigt.

Im vorliegenden Beispiel wird der tatsächliche Umsatz in der KPI-Visualisierung in grün mit einem Aufwärtspfeil angezeigt, um anzugeben, dass der Umsatz über dem Ziel liegt. In diesem Fall ist der Zielwert der geplante Umsatz. In einer Wortgrafik wird die Form der Varianz im Lauf der Zeit in derselben Farbe wie der Basiswert angezeigt.

TARGET_LINE compared to Revenue for
MonthsAsMember

\$228,762,440.00 ↑

Revenue

\$176,454,488.14 (+29.64%)

Planned Revenue



Erstellen Sie eine ähnliche KPI-Visualisierung, indem Sie Kennzahlen von Ihrer eigenen Datenquelle in die Felder in einer leeren KPI-Visualisierung ziehen:

1. Wählen Sie in der Symbolleiste **Neue Karte** aus.
2. Wählen Sie die leere Karte **Einzel** aus.
3. Klicken Sie auf **Typ auswählen** und wählen Sie den Visualisierungstyp **KPI** aus. Fügen Sie dann jedem Feld ein Datenelement hinzu.
4. Ziehen Sie eine Kennzahl in das Feld **Basiswert**. Dieser Wert ist das tatsächliche Ziel.
5. Ziehen Sie eine Kennzahl in das Feld **Zielwert**.
6. Ziehen Sie eine weitere Kennzahl in das Feld **Zeit**. Durch diesen Wert wird eine Wortgrafik für Ihre KPI-Visualisierung erstellt. Sie können mehrere Kennzahlen, z. B. Jahre und Monate, zum Feld **Zeit** hinzufügen.

Verwenden Sie die Eigenschaften, um eine KPI-Visualisierung anzupassen. Die Eigenschaften sind standardmäßig beispielsweise so festgelegt, dass die bedingte Farbe Grün angezeigt wird, wenn ein Ziel erreicht wird und die bedingte Farbe Rot, wenn das Ziel nicht erreicht wird. Wenn Sie das tatsächliche Ziel in einer anderen Farbe anzeigen möchten, erweitern Sie unter Eigenschaften die Regel und wählen Sie dann eine andere **Textfarbe** aus.

Führen Sie die folgenden Schritte aus, um eine Regel für eine bedingte Farbe zu bearbeiten, und wählen Sie eine angepasste Farbe aus:

1. Wählen Sie die KPI-Visualisierung in Ihrer Exploration aus.
2. Klicken Sie auf die Registerkarte **Eigenschaften**.
3. Erweitern Sie unter **Regeln** die Regel, die Sie bearbeiten möchten.
4. Wählen Sie unter **Textfarbe** eine Farbe aus.

Nachfolgend werden die KPI-Eigenschaften unter **Regelstil** beschrieben:

- **Textfarbe**

Legen Sie die Farbe für den Wert, die Wortgrafik und die Indikatorform fest.

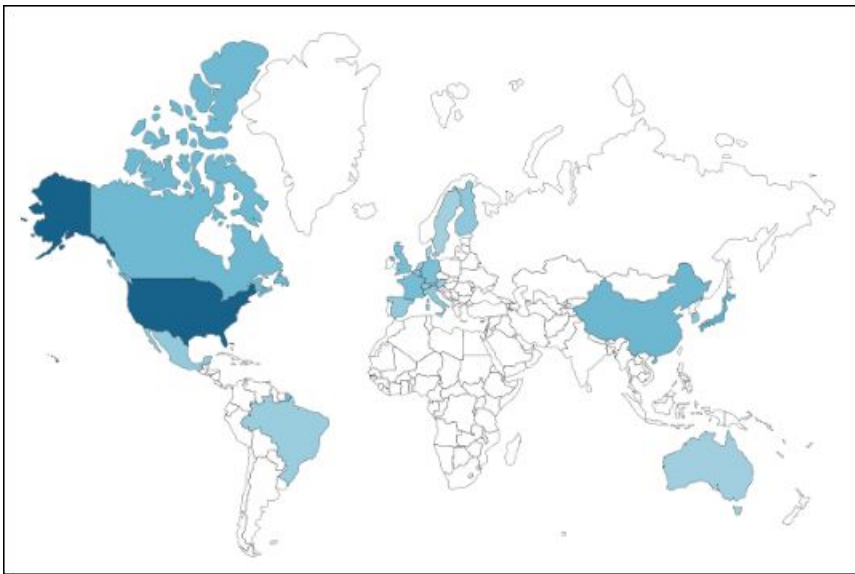
- **Indikator**

Wählen Sie eine Form aus, die in der KPI-Visualisierung angezeigt werden soll, wenn die Regel eingehalten wird. Das Anzeigen eines Abwärtspfeils empfiehlt sich beispielsweise, wenn der Basiswert unter einen bestimmten Schwellenwert im Vergleich zum Zielwert fällt.

Vorgängerlandkarte

Verwenden Sie eine Vorgängerlandkarte, wenn Sie Muster innerhalb Ihrer Daten unter geografischen Gesichtspunkten darstellen möchten. Sie können eine Vorgängerlandkarte verwenden, wenn Sie nicht mit dem Internet verbunden sind.

In dieser Vorgängerlandkartenvisualisierung werden beispielsweise die Einnahmen nach dem Land des Einzelhändlers dargestellt, wobei die dunklere Farbe auf höhere Einnahmen hinweist.



Weitere Informationen finden Sie in https://www.ibm.com/support/knowledgecenter/SSEP7J_11.1.0/com.ibm.swg.ba.cognos.ug_ca_legacymaps.doc/ug_ca_legacymaps.pdf.

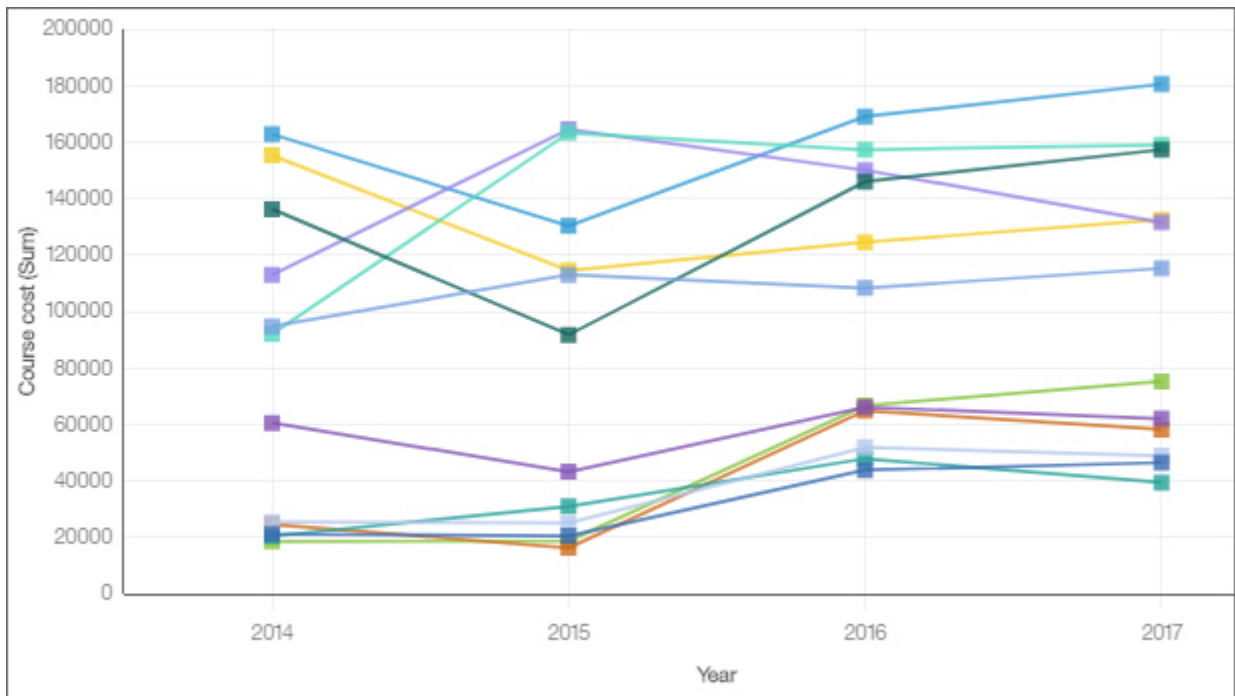
Liniendiagramm

Verwenden Sie eine Liniendiagrammvisualisierung zur Darstellung von Trends über bestimmte Zeiträume hinweg.

Eine Liniendiagrammvisualisierung kann Trends und Zyklen vergleichen, Beziehungen zwischen Variablen erkennbar machen oder darstellen, wie sich eine einzelne Variable über einen Zeitraum hinweg entwickelt hat.

Verwenden Sie für eine effektive Liniendiagrammvisualisierung eine Zeitspalte auf der X-Achse, wie zum Beispiel Jahre, Quartale, Monate oder Tage. Wenn auf der X-Achse etwas anderes dargestellt wird, beispielsweise Kanada, die Niederlande, Großbritannien und die USA, verwenden Sie eine Balken- oder Säulendiagrammvisualisierung.

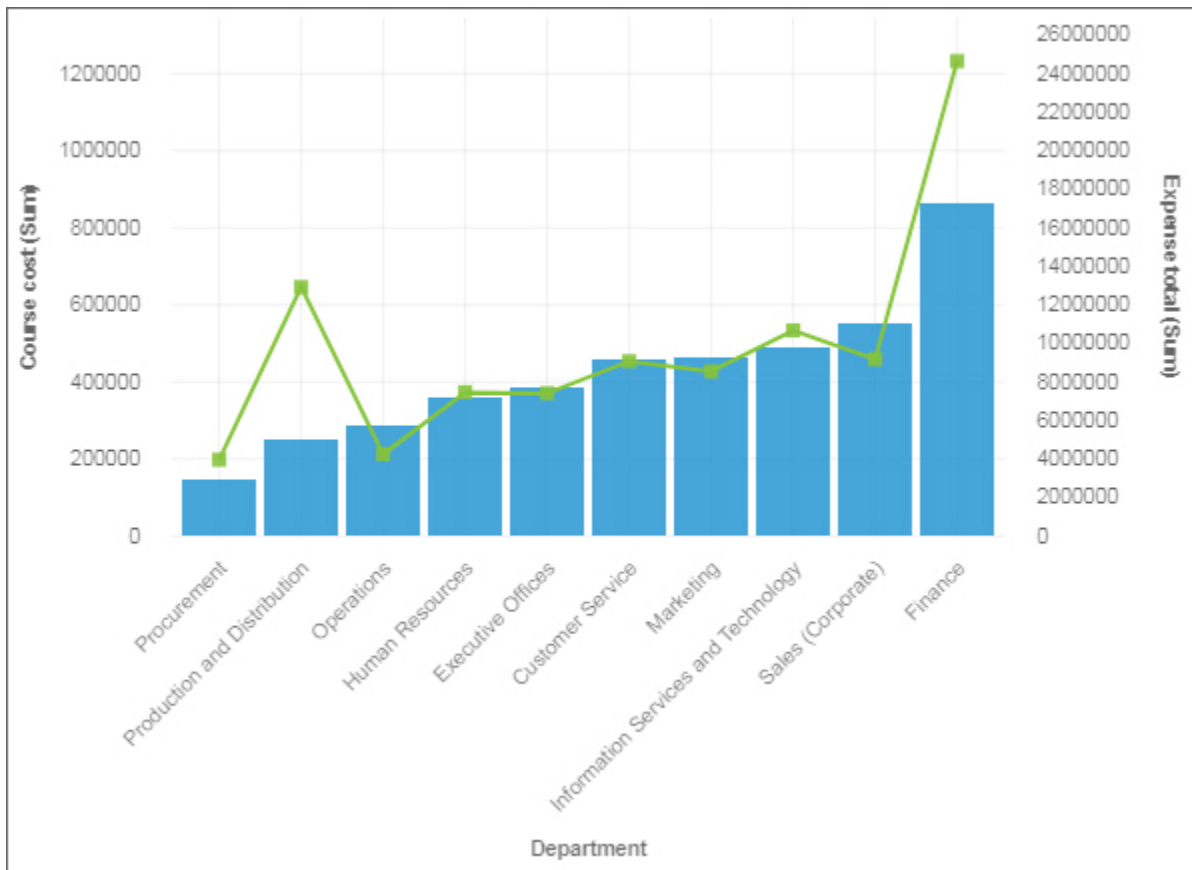
Diese Liniendiagrammvisualisierung stellt zum Beispiel den Trend für die Kursgebühren pro Abteilung über mehrere Jahre hinweg dar.



Linien- und Säulendiagramm

Verwenden Sie eine Linien- und Säulendiagrammvisualisierung, um die Beziehungen zwischen mehreren Datenserien hervorzuheben, indem Sie Balken und Linien in einer gemeinsamen Visualisierung kombinieren.

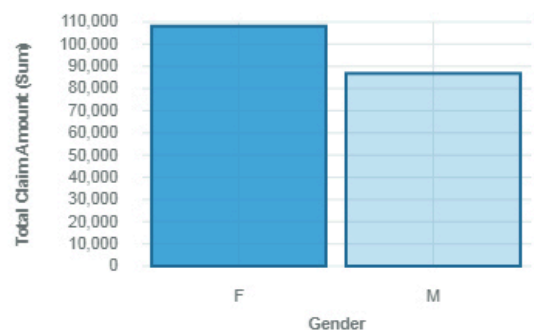
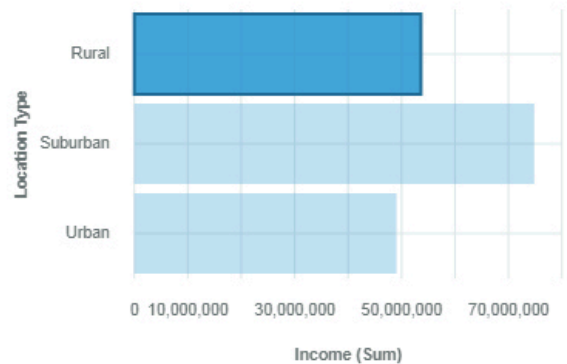
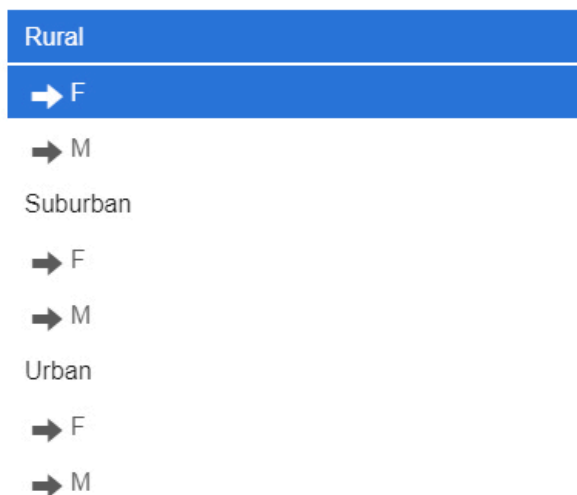
Diese Linien- und Säulendiagrammvisualisierung stellt beispielsweise die Beziehung zwischen Kursgebühren und den Gesamtausgaben pro Abteilung dar.



Liste

Verwenden Sie eine Listenvisualisierung zum Erstellen einer Übersicht der Daten in hierarchischer Form.

Eine weitere Verwendung der Listenvisualisierung ist das Erstellen von Filterwidgets. Im nächsten Beispiel wird gezeigt, wie Sie die Listenvisualisierung als Filterwidget verwenden können.



Karte

Verwenden Sie eine Landkarte, wenn Sie in Ihren Daten Muster unter geografischen Gesichtspunkten erkennen möchten.

Ihr Datenasset muss geografische Daten enthalten, wie zum Beispiel Länder, Bundesländer, Provinzen oder Kontinente.

Anmerkung: Wenn Sie in den Einstellungen des verwendeten Eingabehilfesystems festlegen, dass keine Animationen angezeigt werden sollen, werden in den Karten keine Animationen angezeigt.

Die Karten in Cognos Analytics unterstützen die folgenden Kontinente:

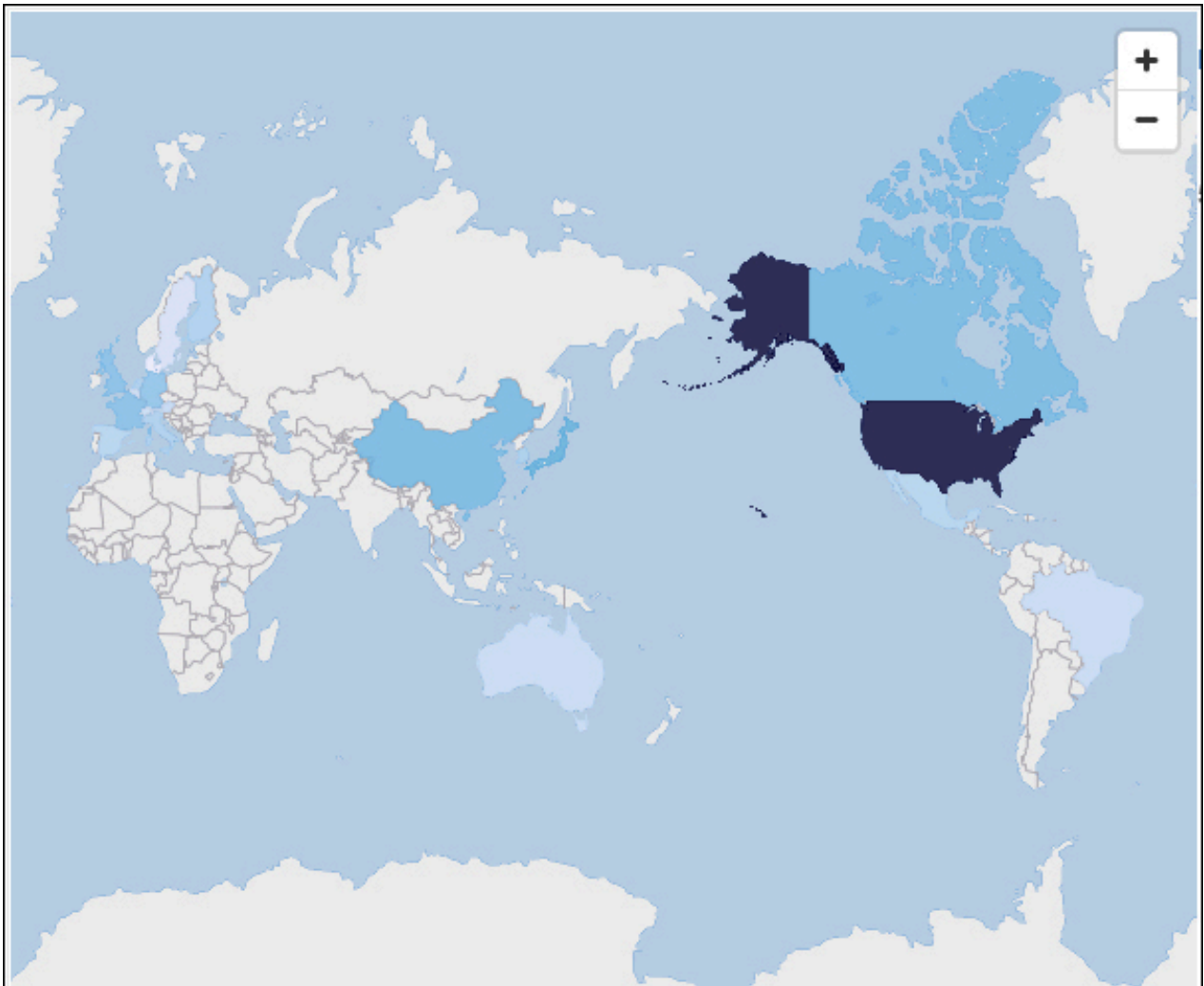
- Nordamerika
- Südamerika
- Afrika
- Asien
- Europa
- Antarktis
- Ozeanien

Um zu ermitteln, ob eine Spalte in einer Karte verwendet werden kann, analysiert Cognos Analytics eine Stichprobe von 2000 Werten in der Standortspalte und sucht dabei nach wiedererkennbaren geografischen Bezeichnungen. Wenn mindestens 80% als Landkartenwerte erkannt werden, erstellt Cognos Analytics eine Landkarte.

Sie haben beispielsweise vier Länder in Ihrer Positionsspalte: Brasilien, China, Indien und Russland. Die falsche Schreibweise von 'Indien' bedeutet, dass nur 75 % der Werte wiedererkennbare geografischen

Bezeichnungen sind und keine Karte als Ausgangspunkt angezeigt wird. Sind jedoch fünf Ländernamen vorhanden, von denen einer einen Schreibfehler enthält, kann eine Karte angezeigt werden.

In dieser Kartenvisualisierung werden beispielsweise die Einnahmen nach dem Land des Einzelhändlers dargestellt, wobei die dunklere Farbe auf höhere Einnahmen hinweist.



Marimekko-Diagramm

Eine Marimekko-Visualisierung ähnelt einer Visualisierung eines gestapelten Säulendiagramms. Sie stellt Daten durch unterschiedliche Höhen dar und bietet eine zusätzliche Datendimension durch unterschiedliche Säulenbreiten. Die Breite der Säulen basiert auf dem Wert, der dem Breitenfeld zugeordnet ist. Die individuelle Segmenthöhe entspricht dem prozentualen Anteil am betreffenden Spaltengesamtwert.

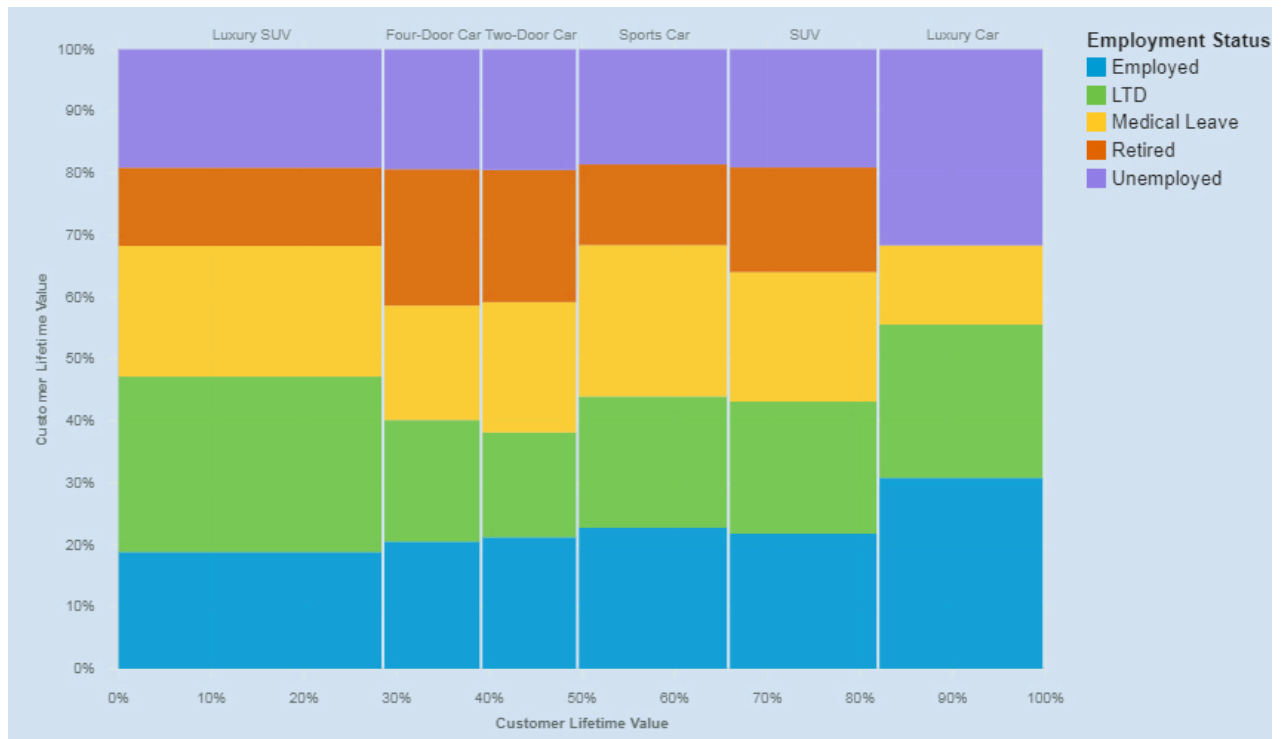
Sie können schnell größere Segmente erkennen, beispielsweise eine bestimmte Branche, die einen großen Anteil in einer Region hat. Sie können auch Leerräume erkennen, beispielsweise eine unterrepräsentierte Branche in einer bestimmten Region.

Die Marimekko-Visualisierung ist nützlich bei Teil-Ganzes-Vergleichen, bei denen eine zusätzliche Kennzahl/Variable dargestellt werden muss.

Die Marimekko-Visualisierung ermöglicht die Darstellung von Daten, die in Verbindung mit zwei Dimensionen gleichzeitig dargestellt werden sollen. Marktsegmente werden beispielsweise häufig auf der X-Achse dargestellt, wobei die Breite jeder Spalte dem finanziellen Wert eines Segments entspricht. Marimekko-Visualisierungen verwenden Sie beispielsweise in Fällen, in denen Sie pro Produktlinie den Anteil an den Einnahmen darstellen möchten. Oder das Bruttonationaleinkommen pro Land.

Die Marimekko-Visualisierung kann Gesamt- oder Teilwerte anzeigen. Wenn Sie gestapelte Prozentsätze statt der Anzahl verwenden möchten, verwenden Sie die Option **Als gestapeltes Prozentsatzdiagramm anzeigen**.

Im folgenden Beispiel werden bei aktivierter Option **Als gestapeltes Prozentsatzdiagramm anzeigen** der Anteil des Kundenertragswerts und des Beschäftigungsstatus für Produktlinien in unterschiedlichen Fahrzeugklassen dargestellt.



Die Marimekko-Visualisierung wurde erstellt, indem die folgenden Datenelemente aus der Anzeige 'Quellen' gezogen werden:

- Ziehen Sie den Typ **Fahrzeugklasse** in das Feld **Balken**.
- Ziehen Sie **Kundenertragswert** in das Feld **Länge**.
- Ziehen Sie **Beschäftigungsstatus** in das Feld **Farbe**.

Beispiele

Im Beispielbericht **Kundenertragswertanalyse** wird ein Beispiel für eine Visualisierung des Typs 'Wortwolke' angezeigt. Sie finden das Beispiel hier: **Teaminhalt > Beispiele > Berichte > Kundenertragswertanalyse**.

Wenn eines der Beispielobjekte fehlt, wenden Sie sich an Ihren Administrator.

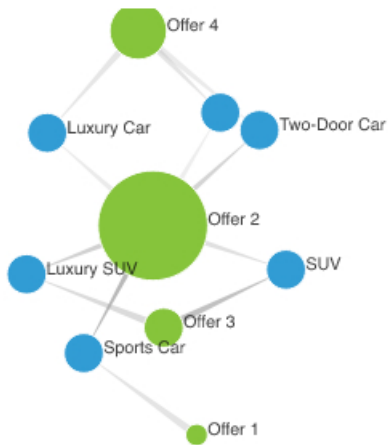
Netzdiagramm

Verwenden Sie eine Netzdiagrammvisualisierung, wenn Sie die Verbindungen zwischen Spalten in Ihrem Datenasset darstellen möchten. Eine Netzdiagrammvisualisierung ist eine gute Möglichkeit zur Darstellung von Verbindungen, Netzen und Schnittpunkten.

In Netzdiagrammvisualisierungen wird eine Reihe von Knoten, dargestellt durch Symbole, und Verbindungen, dargestellt durch Pfade, angezeigt, um die Beziehungen zwischen Entitäten oder Elementen darzustellen.

Verwenden Sie die Felder **Von** und **Bis** zum Definieren der Beziehung, die Sie untersuchen möchten.

Eine Netzdiagrammvisualisierung zeigt die Angebotsannahme nach Fahrzeugklasse an.



Erstellen Sie die Netzdiagrammvisualisierung, indem Sie die folgenden Datenelemente aus dem Abschnitt

Angebote in die Anzeige **Quellen**  ziehen:

- Ziehen Sie **Angebot** in das Feld **Von**.
- Ziehen Sie **Fahrzeugklasse** in das Feld **An**.
- Ziehen Sie **Akzeptiert** in das Feld **Linienstärke**.

Legen Sie dann die Größen- und Knoteneigenschaften fest.

1. Klicken Sie auf die Visualisierung und dann auf \approx . Legen Sie die folgenden Optionen in der Anzeige **Eigenschaften** fest:
 - **Größe - Breite:** 500 px, **Höhe:** 300 px
 - **Mindestgröße der Knoten:** 20
 - **Maximalgröße der Knoten:** 100
2. Klicken Sie auf \approx , um die Anzeige **Eigenschaften** zu schließen.

Beispiele

Beispiele für Visualisierungen finden Sie im Beispielbericht **Analyse des Kundenlebensdauerwerts**. Die Beispiele finden Sie hier: **Teaminhalt** > **Beispiele** > **Berichte** > **Analyse des Kundenlebensdauerwerts**.

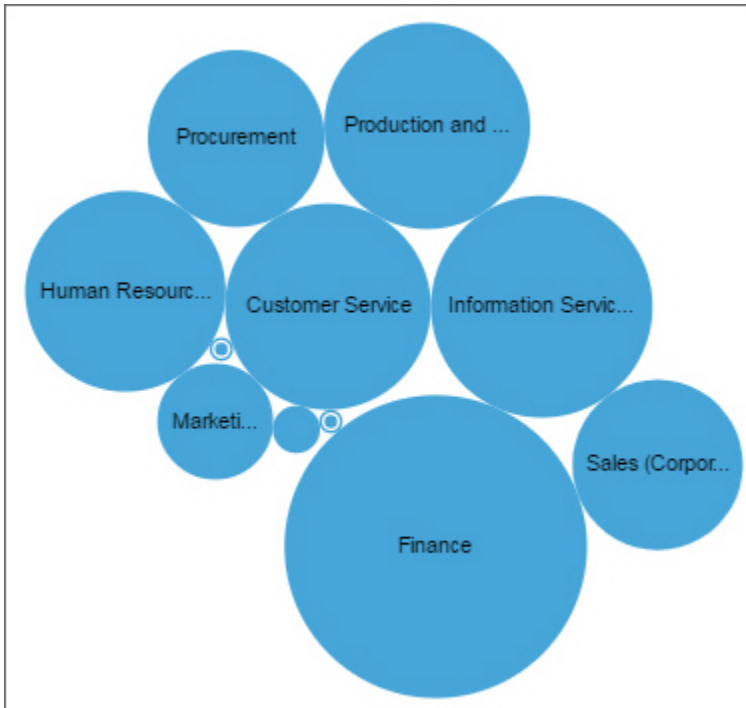
Wenn eines der Beispielobjekte fehlt, wenden Sie sich an Ihren Administrator.

Kompaktes Blasendiagramm

Verwenden Sie die Visualisierung eines kompakten Blasendiagramms, wenn Sie die Beziehungen zwischen Spalten darstellen möchten, die numerische Werte enthalten, beispielsweise Einnahmen. Sie ist ähnlich der Blasendiagrammvisualisierung, aber die Blasen sind dicht gedrängt und nicht über ein Raster verteilt dargestellt. Eine Visualisierung eines kompakten Blasendiagramms zeigt eine große Menge an Daten auf engem Raum.

Die Blasen haben verschiedene Größen und Farben.

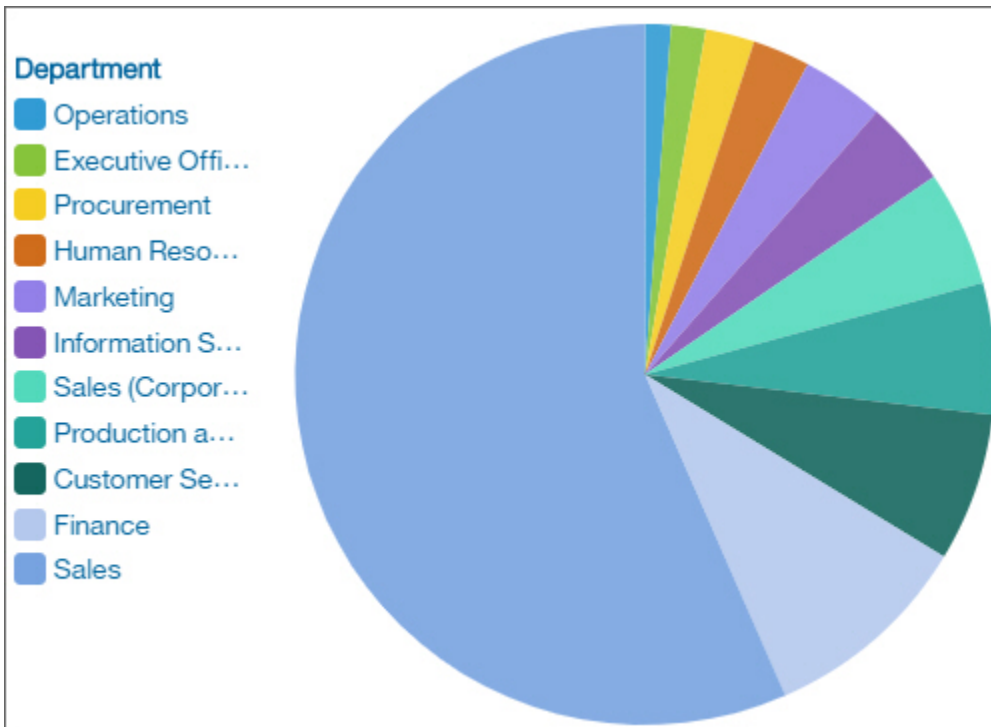
Diese Visualisierung eines kompakten Blasendiagramms stellt zum Beispiel die externen Rekrutierungen für jede Abteilung dar. Jede Blase steht für eine Abteilung. Die Größe einer Blase wird durch die Anzahl der externen Rekrutierungen der jeweiligen Abteilung bestimmt.



Kreisdiagramm

Verwenden Sie eine Kreisdiagrammvisualisierung zum Hervorheben von Anteilen. Die einzelnen Kreisabschnitte stellen die relative Beziehung der einzelnen Teile zum Ganzen dar.

Diese Kreisdiagrammvisualisierung zeigt zum Beispiel die Anzahl der Schulungstage pro Abteilung.



Punktogramm

Verwenden Sie eine Punktvisualisierung zur Darstellung von Trends über bestimmte Zeiträume hinweg.

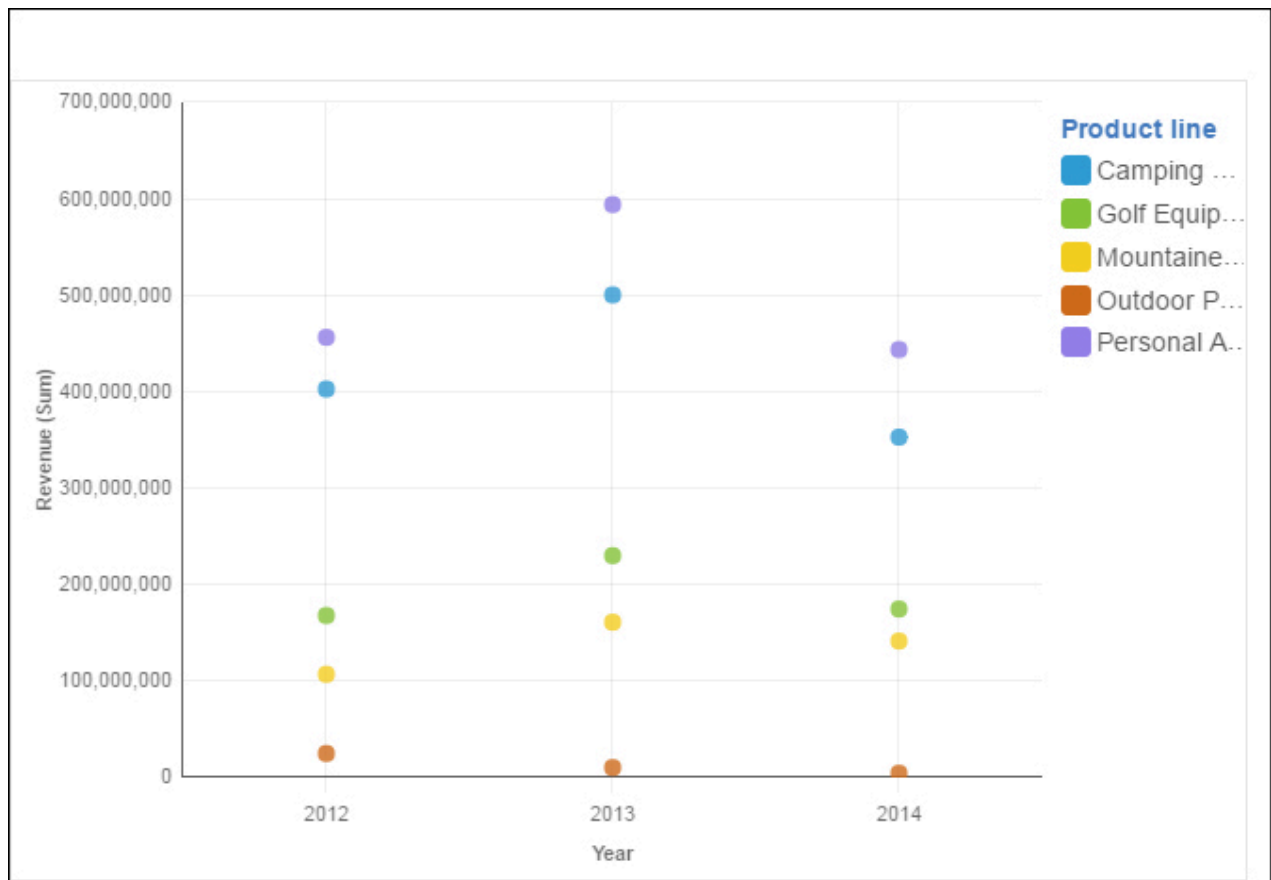
Eine Punktvisualisierung kann Trends und Zyklen vergleichen, Beziehungen zwischen Variablen erkennbar machen oder darstellen, wie sich eine einzelne Variable über einen Zeitraum hinweg entwickelt hat.

Eine Punktvisualisierung entspricht einem Liniendiagramm ohne die Verbindungslinien.

Für eine effektive Liniendiagrammvisualisierung sollten auf der X-Achse Zeitwerte dargestellt werden, wie zum Beispiel Jahre, Quartale, Monate oder Tage. Wenn auf der X-Achse etwas anderes dargestellt wird, beispielsweise Kanada, die Niederlande, Großbritannien und die USA, verwenden Sie eine Balkendiagrammvisualisierung.

Datenwerte werden vertikal dargestellt.

Diese Liniendiagrammvisualisierung stellt zum Beispiel die Einnahmen pro Quartal nach Bestellmethodentyp dar. Bestellungen über das Web haben in diesem Zeitraum dramatisch zugenommen.

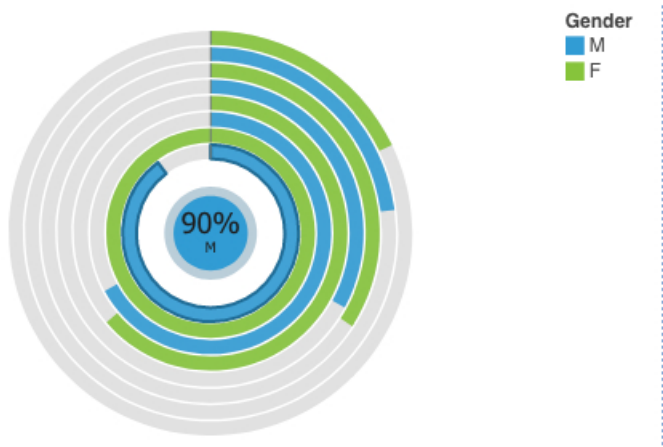


Radialdiagramm


In einer Visualisierung des Typs 'Radialdiagramm' wird jeder Balken in einem Kreis mit längeren Balken angezeigt, die größere Werte darstellen. Bewegen Sie den Cursor über einen Balken, um Details zu dem Balken anzuzeigen, beispielsweise den exakten Wert, der durch den Balken dargestellt wird. Jeder Balken beginnt an der 12-Uhr-Position und verläuft bei positiven Werten im Uhrzeigersinn und bei negativen Werten gegen den Uhrzeigersinn.

Bei Radialdiagrammvisualisierungen, auch bekannt als Skalendiagramme oder Tachometerdiagramme, werden Informationen wie Werte auf einer Messuhr dargestellt. Die Visualisierung des Typs 'Radialdiagramm' ist nur mit einer Kategorie gültig.

Diese Visualisierung zeigt Verlängerungen nach Angebotstyp und Geschlecht an.



Erstellen Sie die Visualisierung des Typs 'Radialdiagramm', indem Sie die folgenden Datenelemente aus

dem Abschnitt **Kundenanalyse** in die Anzeige **Quellen**  ziehen:

- Ziehen Sie **Angebotstyp erneuern** in das Feld **Balken**.
- Ziehen Sie **Anzahl der Policen** in das Feld **Länge**.
- Ziehen Sie **Geschlecht** in das Feld **Farbe**.

Im nächsten Schritt legen Sie die Sortierungseigenschaften für **Angebotstyp erneuern** und **Geschlecht** fest.

1. Klicken Sie auf die Visualisierung, und klicken Sie dann in der Anzeige **Daten** auf das Datenelement **<Angebotstyp erneuern>**.
2. Klicken Sie auf \approx .
3. Wählen Sie in der Anzeige **Eigenschaften** für die **Sortierreihenfolge** die Option **Aufsteigend** aus.
4. Klicken Sie in der Anzeige **Daten** auf das Datenelement **<Geschlecht>**.
5. Wählen Sie in der Anzeige **Eigenschaften** für die **Sortierreihenfolge** die Option **Absteigend** aus.
6. Klicken Sie auf \approx , um die Anzeige **Eigenschaften** zu schließen.

Beispiele

Beispiele für Visualisierungen finden Sie im Beispielbericht **Analyse des Kundenlebensdauerwerts**. Die Beispiele finden Sie hier: **Teaminhalt > Beispiele > Berichte > Analyse des Kundenlebensdauerwerts**.

Wenn eines der Beispielobjekte fehlt, wenden Sie sich an Ihren Administrator.

Streudiagramm

In Streudiagrammvisualisierungen werden Datenpunkte verwendet, um zwei Kennzahlen an beliebigen Stellen entlang einer Skala darzustellen, nicht nur an regelmäßigen Teilstrichen.

Streudiagrammvisualisierungen sind nützlich bei der Untersuchung von Korrelationen zwischen verschiedenen Datensets.

Im folgenden Beispiel wird die Korrelation zwischen Einnahmen und Bruttogewinn für jeden Produkttyp dargestellt.



Spiraldiagramm

Eine Spiraldiagrammvisualisierung zeigt Ihnen die wichtigsten Einflussfaktoren (oder Prädiktoren) für ein angegebenes Ziel. Je näher sich ein Einflussfaktor zur Mitte hin befindet, desto stärker ist er.


IBMCognos Analytics verwendet hoch entwickelte Algorithmen, mit deren Hilfe sehr gut interpretierbare Insights gewonnen werden können, die auf komplexen Modellierungen basieren. Sie müssen nicht wissen, welche statistischen Tests mit Ihren Daten ausgeführt werden müssen. Cognos Analytics wählt die richtigen Tests für die Daten aus.


Wichtige Einflussfaktoren für sowohl fortlaufende als auch kategoriale Ziele sind in der Spiraldiagrammvisualisierung in Dashboards und Explorationen verfügbar.

Weitere Informationen finden Sie in der Dokumentation zu *statistischen Tests* in der Veröffentlichung *IBM Cognos Analytics Dashboards und Storys - Benutzerhandbuch*.

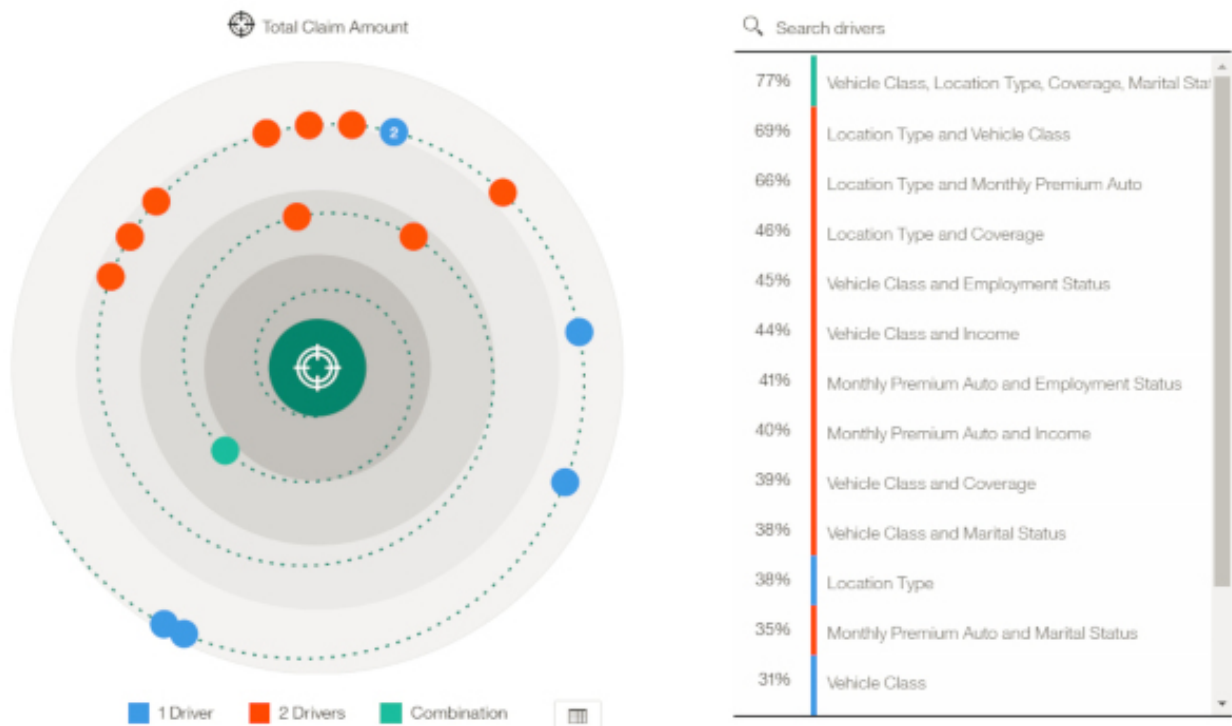
Diese Spiraldiagrammvisualisierung zeigt beispielsweise, dass die Kombination von Fahrzeugklasse, Standorttyp, Abdeckung, Familienstand und Beschäftigungsstatus die stärksten Einflussfaktoren des Ziels (Gesamtbetrag der Schadensmeldungen) sind.

Sie können Einflussfaktoren von der Analyse ausschließen. Klicken Sie mit der rechten Maustaste auf

einen Einflussfaktor und klicken Sie auf das Symbol **Einflussfaktoren bearbeiten** . Wählen Sie die Einflussfaktoren aus, die in die Analyse einbezogen werden sollen.

Wenn Sie zentrale Einflussfaktoren bearbeiten oder hinzufügen möchten, klicken Sie auf  im Zielfeld.

Um die Leistung zu verbessern, basiert die Analyse aufgrund der Anzahl der Zeilen in der Datenquelle auf einer repräsentativen Stichprobe der gesamten Daten.

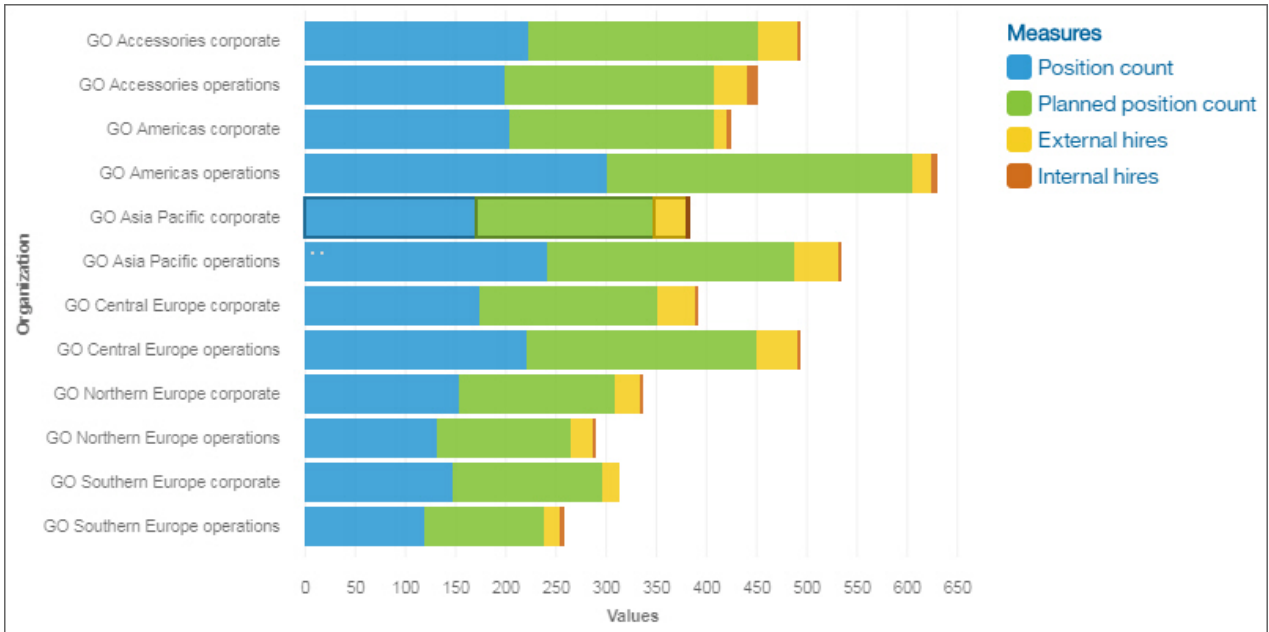


Anmerkung: Filter werden für Spiraldiagrammvisualisierungen nicht unterstützt.

Gestapeltes Balkendiagramm

Verwenden Sie eine Visualisierung eines gestapelten Balkendiagramms zum Vergleichen der proportionalen Beiträge der einzelnen Elemente zum Gesamtwert, wie zum Beispiel der Umsatz für Produkte und der Umsatz für Produkte pro Monat.

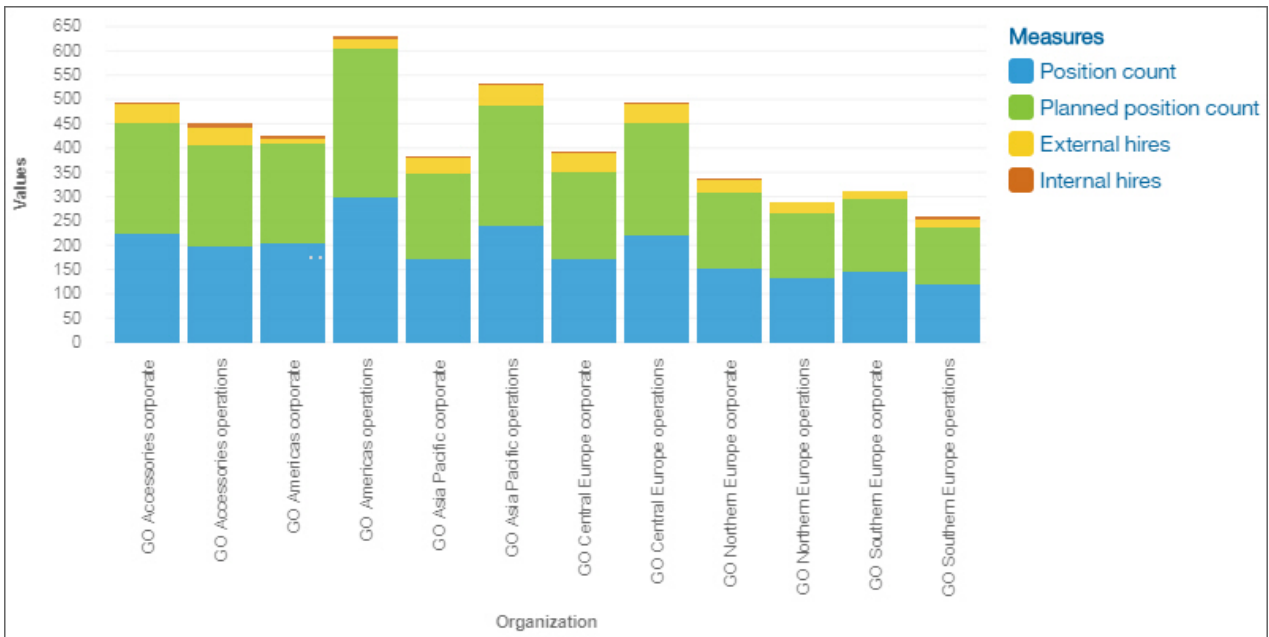
Der Visualisierung eines gestapelten Balkendiagramms können Veränderungen über einen bestimmten Zeitraum hinweg entnommen werden oder sie kann zum Vergleichen der proportionalen Beiträge der einzelnen Elemente zum Gesamtwert verwendet werden. Falls es so viele Balken gibt, dass die Beschriftung nicht mehr lesbar ist, filtern Sie die Daten, um sich auf ein Subset der Daten zu konzentrieren, oder verwenden Sie eine Treemap.



Gestapeltes Säulendiagramm

Verwenden Sie eine Visualisierung eines gestapelten Säulendiagramms zum Vergleichen der proportionalen Beiträge der einzelnen Elemente zum Gesamtwert, wie zum Beispiel der Umsatz für Produkte und der Umsatz für Produkte pro Monat.

Der Visualisierung eines gestapelten Säulendiagramms können Veränderungen über einen bestimmten Zeitraum hinweg entnommen werden oder sie kann zum Vergleichen der proportionalen Beiträge der einzelnen Elemente zum Gesamtwert verwendet werden. Falls es so viele Balken gibt, dass die Beschriftung nicht mehr lesbar ist, filtern Sie die Daten, um sich auf ein Subset der Daten zu konzentrieren, oder verwenden Sie eine Treemap.



Auswertung

Verwenden Sie eine Auswertungsvisualisierung, wenn Sie den Gesamtwert einer Kennzahl oder die Anzahl für eine kategoriale Spalte sehen möchten.

Diese Auswertungsvisualisierung zeigt beispielsweise den Gesamtbetrag der Einnahmen aller Produkttypen.



Diese Auswertungsvisualisierung zeigt beispielsweise die Anzahl der Abteilungen in Ihrem Unternehmen.



Sunburst-Diagramm

Eine Sunburst-Diagrammvisualisierung wird verwendet, um darzustellen, wie die zugrunde liegenden Daten ein gewähltes Ziel prognostizieren, und hebt wesentliche Insights hervor.

Weitere Informationen zur Sunburst-Diagrammvisualisierung finden Sie in [„Entscheidungsstrukturvisualisierung untersuchen“](#) auf Seite 20.

Tabelle

Verwenden Sie eine Tabelle zum Anzeigen detaillierter Informationen aus Ihrer Datenbank, wie zum Beispiel Produktlisten und Kundenverzeichnisse. In einer Tabelle werden Daten in Zeilen und Spalten angezeigt. Jede Spalte enthält alle Werte für ein Datenelement in der Datenbank oder eine Berechnung, die auf den Datenelementen in der Datenbank basiert.

In dieser Tabelle werden beispielsweise die Kursgebühren für jede Abteilung dargestellt.

Department	Course cost
Customer Service	459,250
Executive Offices	384,000
Finance	863,750
Human Resources	361,250
Information Services and Technology	491,750
Marketing	465,750
Operations	288,000
Procurement	146,250
Production and Distribution	249,450
Sales	4,056,000
Sales (Corporate)	549,550
Summary	8,315,000

Einer Tabelle weitere Spalten hinzufügen

Sie können sich auf Punkte konzentrieren, die für Sie von Interesse sind, indem Sie der Visualisierung weitere Daten hinzufügen.

1. Ziehen Sie eine weitere Spalte zu dem Feld, in dem Sie zusätzliche Daten wünschen.
2. Legen Sie die Spalte neben der vorhandenen Spalte ab.

Ab Cognos Analytics Version 11.1.4 können Sie Daten aus der Anzeige **Ausgewählte Quellen** ziehen und Sie können Daten in eine Zeile/Spalte einfügen oder die Daten auf vorhandenen Daten ablegen, um diese zu ersetzen.

Treemap

Verwenden Sie eine Treemap-Visualisierung zum Ermitteln von Mustern und Ausnahmen in einem großen, komplexen Datenasset.

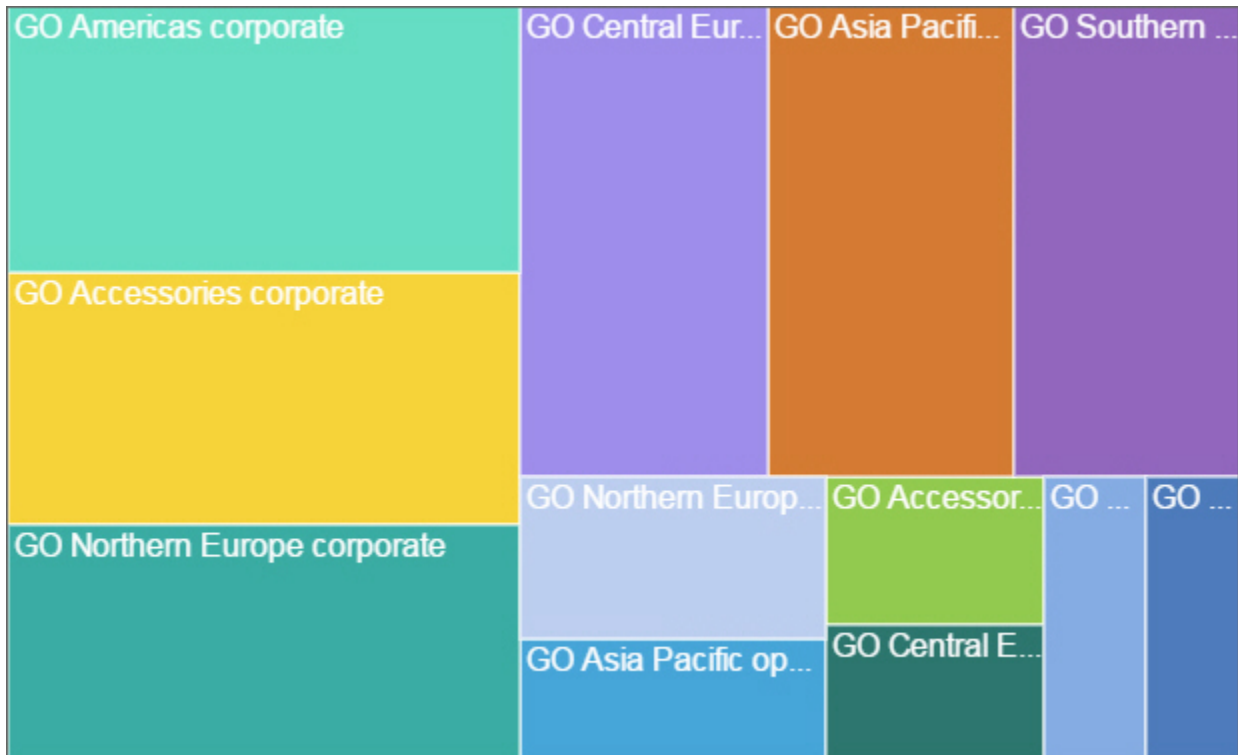
In einer Treemap werden Beziehungen zwischen einer großen Zahl von Komponenten mithilfe der Größe und der farblichen Codierung einer Gruppe von verschachtelten Rechtecken dargestellt.

In einer Treemap, in der die Farbgebung nach Kategorien erfolgt, wird die Kategorie der Ebene 1 durch die Farbe angegeben. Die Größe der Rechtecke stellt die Werte dar. In einer Treemap, in der die Farbgebung nach Werten erfolgt, stellt die Größe der Rechtecke einen der Werte dar, während die Farbe eine zweite

Gruppe von Werten darstellt. Verwenden Sie keine Daten, die negative Zahlen enthalten. Eine Treemap ignoriert negative Zahlen.

Viele Datenassets weisen eine hierarchische Struktur auf. Sie haben beispielsweise Daten zur Gewinnspanne einzelner Artikel in einem Lebensmittelgeschäft. Innerhalb der allgemeinen Kategorie 'Früchte' gibt es die Kategorie 'Zitrusfrüchte'. Es werden verschiedene Zitrusfrüchte aufgelistet, beispielsweise Grapefruit, Orange und Zitrone. In einer Treemap können Sie darstellen, wie die einzelnen Zitrusfrüchte im Vergleich untereinander und im Vergleich mit anderen Lebensmitteln abschneiden.

Diese Treemap-Visualisierung zeigt zum Beispiel die Kursgebühren pro Organisation.



Klicken Sie bei gedrückter Steuertaste auf ein ausgewähltes Rechteck, um das Rechteck abzuwählen.

Wasserfalldiagramm

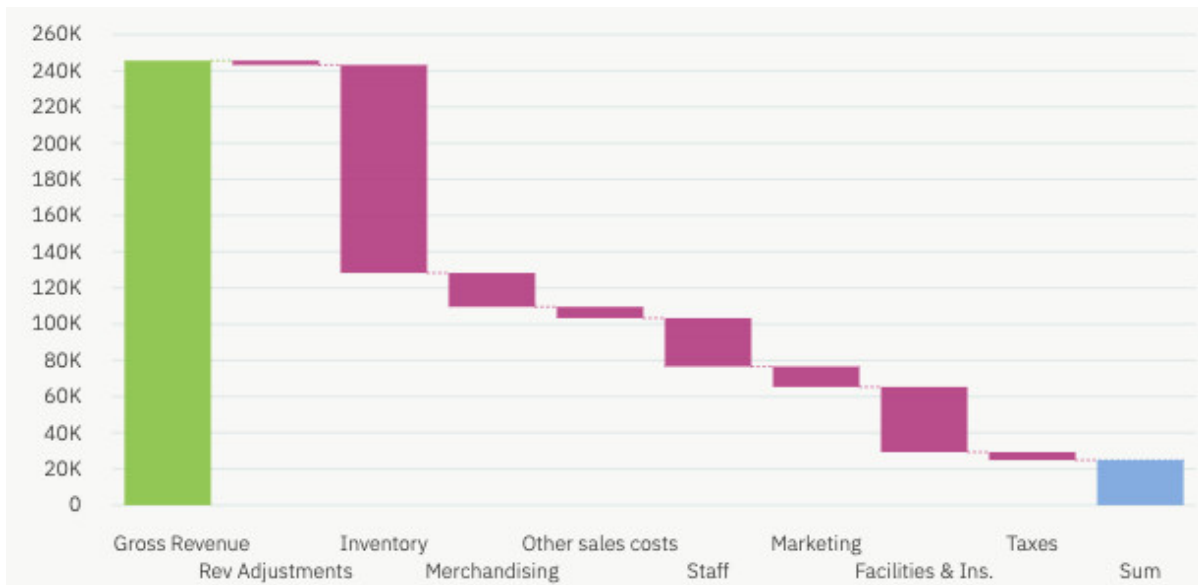
Verwenden Sie eine Wasserfallvisualisierung, um den kumulativen Effekt zu verstehen, der eine Reihe positiver und negativer Werte auf einen Anfangswert hat. Die Balken in einer Wasserfallvisualisierung sind keine Summen.

Eine Wasserfallvisualisierung zeigt, wie ein Anfangswert erhöht und durch eine Reihe von Zwischenwerten verringert wird, was zu einem letzten kumulativen Wert führt, der in der rechten Spalte angezeigt wird. Die Zwischenwerte können entweder zeitbasiert oder categoriesbasiert sein.

Einige Beispiele für Wasserfallvisualisierungen sind wie folgt:

- Anzeigen des Nettoeinkommens, nachdem Sie die Erhöhungen und Abnahmen von Einnahmen und Kosten für ein Unternehmen über ein Viertel hinzugefügt haben.
- Kumulative Verkäufe für Produkte über ein Jahr hinweg mit einem jährlichen Gesamtvolumen.

Diese Wasserfall-Visualisierung zeigt den Policenhalter Delta im Monat an.



Wasserfallvisualisierung erstellen

1. Erstellen Sie eine neue Exploration. Weitere Informationen finden Sie in [„Neue Exploration über das Menü Neu starten“](#) auf Seite 1.
2. Öffnen Sie das Beispieldatenmodul über **Quelle auswählen > Teaminhalt > Beispiele > Daten > Kundenanalyse**.
3. Klicken Sie auf **Visualisierungen**, und klicken Sie anschließend auf **Wasserfalldiagramm**, um die Wasserfallvisualisierung zur Exploration hinzuzufügen.
4. Klicken Sie auf **Quellen**



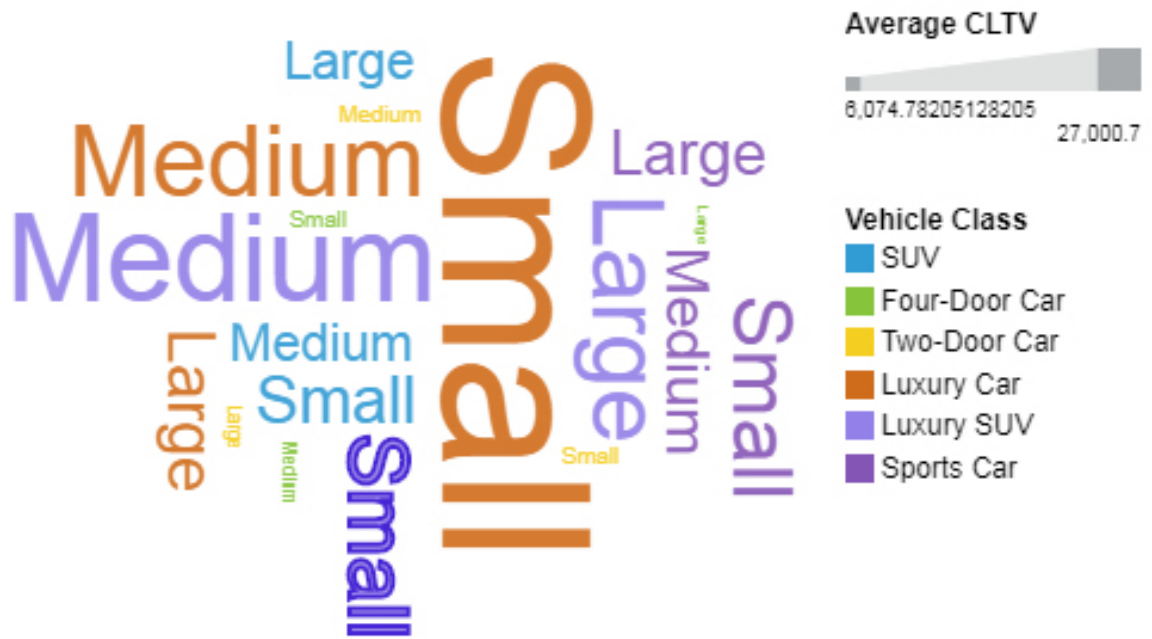
5. Ziehen Sie die folgenden Datenelemente aus dem Bereich für Versicherungsnehmer:
 - Ziehen Sie **Monat** auf die **X-Achse**.
 - Ziehen Sie **Delta** auf die **Y-Achse**.

Wortwolke

Verwenden Sie eine Visualisierung des Typs 'Wortwolke', wenn Sie eine textbasierte Visualisierung einer Spalte anzeigen möchten. Die Textgröße stellt die Größenordnung dar. Der Name selbst stellt die verschiedenen Mitglieder der Spalte dar.

Tipp: Das Datenasset sollte mindestens 15 Spalten und mindestens 100 Zeilen enthalten, damit eine effektive Wortwolke erstellt werden kann.

Diese Visualisierung des Typs 'Wortwolke' stellt zum Beispiel den Kundenertragswert (Customer Lifetime Value, CLV bzw. CLTV) anhand der Fahrzeuggröße und -klasse dar.



Die Wortwolke wurde erstellt, indem die folgenden Datenelemente aus der Quellenanzeige gezogen wurden:

- Typ **Fahrzeuggröße** in das Feld **Wörter** ziehen.
- **Durchschnittlicher CLTV** in das Feld **Größe** ziehen.
- **Fahrzeugklasse** in das Feld **Farbe** ziehen.


Beispiele

Im Beispielbericht **Kundenertragswertanalyse** wird ein Beispiel für eine Visualisierung des Typs 'Wortwolke' angezeigt. Sie finden das Beispiel hier: **Teaminhalt** > **Beispiele** > **Berichte** > **Kundenertragswertanalyse**.


Wenn eines der Beispielobjekte fehlt, wenden Sie sich an Ihren Administrator.

Insights in Visualisierungen

IBMCognos Analytics bietet analytische Insights, die Sie dabei unterstützen, wichtige Beziehungen und aussagekräftige Unterschiede zu erkennen und auszuwerten, basierend auf den Daten, die in der Visualisierung dargestellt werden.

Insights sind verfügbar, wenn Sie auf das Symbol **Insights**  bei auswählbaren Visualisierungen klicken. Wenn Sie Insights aktivieren, wird die Auswertung im Feld **Insights** angezeigt und zugehörige Visualisierungselemente werden hervorgehoben und Details werden in der entsprechenden QuickInfo-Nachricht bereitgestellt. Sie können jede verfügbare Insight separat steuern.

Vorgehensweise

1. Klicken Sie in einer Visualisierung, die Insights unterstützt, auf das Symbol **Insights** .
2. Abhängig von der Visualisierung werden die folgenden Insights angezeigt:
 - **Durchschnitt** Gibt den Durchschnitt des angezeigten Zielwerts an.
 - **Prognosestärke** Zeigt die Prognosestärke der Beziehung zwischen Ziel- und Erläuterungsfeldern an.


- **Anpassungslinie** Zeigt an, ob es eine lineare oder quadratische Beziehung zwischen Ziel- und Erläuterungsfeldern gibt.
- **Aussagekräftige Unterschiede** Zeigt Werte an, die wesentlich höher oder niedriger sind als der Durchschnitt oder Trend.
- **Größte Häufigkeit** Zeigt Werte an, die am häufigsten gemeldet werden.

Korrelierte Insights auswählen

Basierend auf Ihrer Visualisierung werden Ihnen auf Statistiken basierende, korrelierte Insights angezeigt.


Informationen zu diesem Vorgang

Wenn korrelierte Insights verfügbar sind, die mit der Hauptvisualisierung in Verbindung stehen, wird

auf der X- oder der Y-Achse ein grünes Symbol mit einer Zahl  angezeigt. Die Zahl steht für die verfügbaren korrelierten Insights.

Führen Sie die folgenden Aktionen aus, um auf korrelierte Visualisierungen zuzugreifen:


Vorgehensweise

1. Klicken Sie in der Visualisierung auf das grüne Symbol .
2. Klicken Sie auf eines der auf Statistiken basierenden Insights, die im Menü dargestellt werden.
Eine neue Karte wird erstellt.

Empfohlene Visualisierungen auswählen

Empfohlene Visualisierungen sind Piktogramme, die Visualisierungen anzeigen, die möglicherweise für Ihre Daten geeignet sind.

Vorgehensweise

1. Wählen Sie im Fenster **Karten** die Karte aus, die für die Visualisierung steht, die Sie öffnen möchten.
2. Klicken Sie in der Symbolleiste auf **Visualisierungstyp auswählen** .
Die empfohlenen Visualisierungen werden angezeigt.
3. Klicken Sie auf das Piktogramm für die empfohlene Visualisierung, mit der Sie arbeiten möchten.

Zugehörige Visualisierungen auswählen

Wenn eine Visualisierung in Ihrer Exploration im Fokus steht, empfiehlt das System einige zugehörige Visualisierungen, die Sie nicht ausdrücklich angefordert haben. Aufgrund der Datenanalyse könnten diese zugehörigen Visualisierungen für Sie von Interesse sein.

Informationen zu diesem Vorgang

Zugehörige Visualisierungen ersetzen eines der Datenelemente in der Visualisierung oder fügen ein anderes Datenelement hinzu, um eine neue Visualisierung zu erstellen. Zugehörige Visualisierungen verwenden eine Kombination aus gelernten Benutzerinteraktionen, Statistiken und Interessantheit, um nützliche nächste Schritte zu empfehlen.

Führen Sie die folgenden Aktionen aus, um auf zugehörige Visualisierungen zuzugreifen:

Vorgehensweise

1. Wählen Sie im Fenster **Karten** die Karte aus, die für die Visualisierung steht, die Sie öffnen möchten.

2. Klicken Sie in der Symbolleiste auf **Zugehörig**



Kapitel 4. Forecasting

Prognose

Sie können die Prognose in IBM Cognos Analytics verwenden, um Trend, Saisonalität und Zeitabhängigkeit in Daten zu erkennen und zu modellieren.

Sie können in IBM Cognos Analytics Prognosen erstellen, indem Sie automatisierte Tools verwenden, mit denen zeitabhängige Daten modelliert werden können. Automatisierte Modellauswahl und -optimierung machen die Prognose einfach verwendbar, selbst wenn Sie mit der Zeitreihenmodellierung noch nicht vertraut sind.

Prognosen und ihre Konfidenzgrenzen werden in Visualisierungen als Weiterführung von historischen Daten angezeigt. Sie können auch die statistischen Details für generierte Modelle anzeigen, wenn Sie den technischen Hintergrund sehen möchten.

Für die Angabe von Zeitreihen in Prognosen ist oft eine Datenbearbeitung erforderlich. Cognos Analytics unterstützt eine breite Palette von Zeitreihen, ohne dass eine Änderung erforderlich ist, von Standardtypen für Datum und Uhrzeit bis hin zu verschachtelten periodischen und zyklischen Zeitfeldern. Wenn Daten als eine Zeitreihe erkannt werden, wird die Datenaufbereitung automatisiert. Geeignete Trends und saisonale Zeiträume werden erkannt und Modelle aus einer Gruppe von neun unterschiedlichen Modelltypen ausgewählt.


Die Prognose ist in Liniendiagramm-, Balkendiagramm- und Säulendiagrammvisualisierungen möglich. Die Prognose ermöglicht die Analyse von Hunderten von Zeitreihen pro Visualisierung. Prognosen und Konfidenzgrenzen werden für jede Zeitreihe einzeln berechnet und in der Visualisierung als Erweiterungen der aktuellen Daten angezeigt. Sie können jede Zeitreihe einzeln überprüfen und die Prognose und Ergebnisse an Ihre eigenen Daten und Anforderungen anpassen.

Wenn Sie mit Prognosemodellen vertraut sind, können Sie den ausgewählten Modelltyp, die geschätzten Modellparameter, die Standardgenauigkeitskennzahlen und die Informationen zur Verarbeitungsauswertung anzeigen.

Prognosefunktionen

Die Prognose bietet Zeitreihendatenmodellierung und auf Daten basierende Prognosen in Visualisierungen.

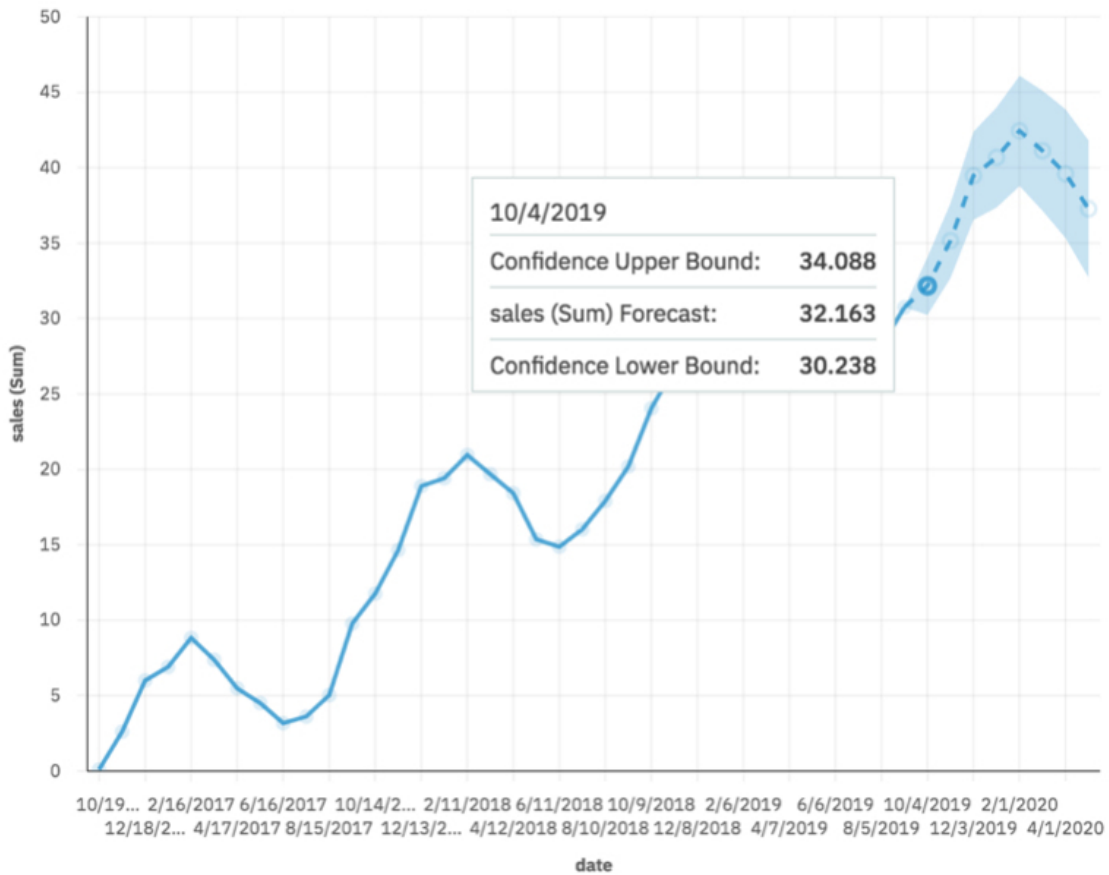
Zur Anwendung der Prognose muss die Visualisierung entweder eine Liniendiagramm-, Balkendiagramm- oder Säulendiagrammvisualisierung sein. Die Daten müssen für Prognosen unterstützt und die Prognose-

funktion muss aktiviert sein. Ist die Prognosefunktion aktiviert, so ist ein Dialogfeld **Prognose**  oben rechts in einer Visualisierung verfügbar, wo Sie die Einstellungen für Modell und Prognose sowie die Konfidenzgrenzen bearbeiten können. Geeignete Zeitreihenmodelle für die Visualisierung werden geschätzt und Prognosen werden in der Visualisierung angezeigt. Außerdem werden auch die Zeitreihenmodellspezifikation und die Auswertung der Datenverarbeitung in der Datenablage angezeigt.

Das folgende Beispiel zeigt Prognosewerte und Konfidenzgrenzen in einer Visualisierung.

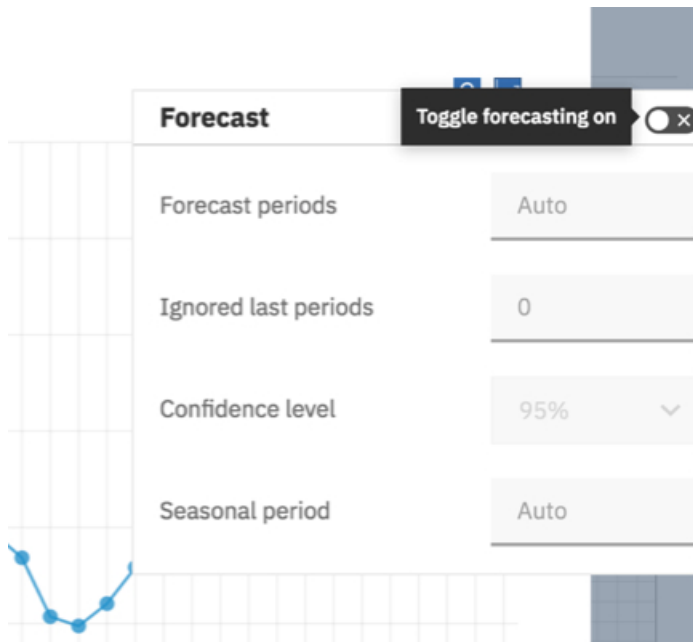


Forecast



Prognoseoptionen

Sie können Ihre Prognosen ändern, indem Sie im Dialogfeld **Prognose** einige Optionen für Zeiträume und Konfidenzniveaus festlegen.



Ein Zeitraum ist das kleinste Zeitintervall zwischen benachbarten Punkten in den Daten.

Die folgenden Optionen sind verfügbar.

Prognosezeiträume

Die Anzahl der Schritte, die prognostiziert werden sollen.

Der Standardwert ist **Automatisch**, d. h. 20 % der Länge der historischen Daten. Alle fehlenden Werte am Ende einer bestimmten Reihe werden ebenfalls prognostiziert, sie werden jedoch nicht auf die angegebene Anzahl von Prognosezeiträumen angerechnet.

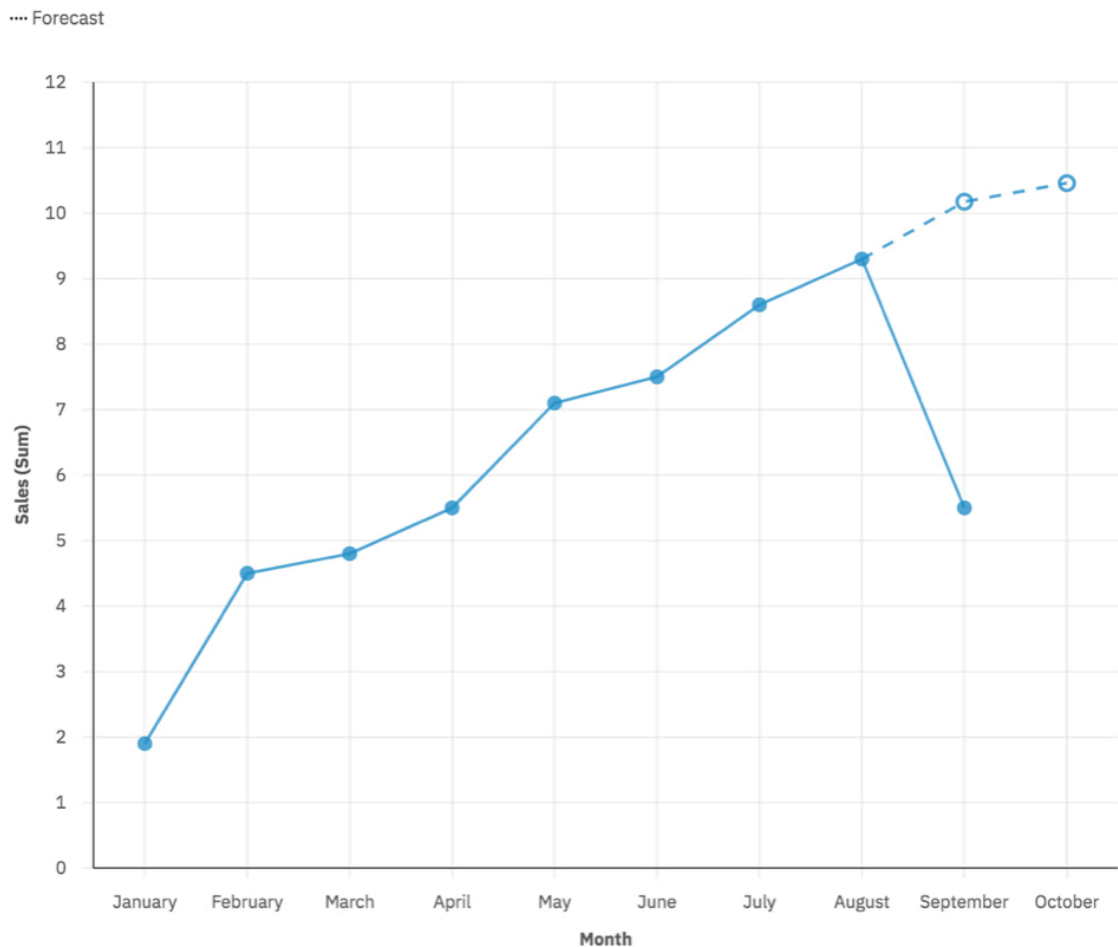
Ignorierte letzte Zeiträume

Ignoriert eine bestimmte Anzahl von Datenpunkten am Ende einer Zeitreihe beim Erstellen des Modells und bei der Berechnung der Prognosen. Alle fehlenden Werte am Ende eines nicht ignorierten Teils einer Reihe werden ebenfalls prognostiziert. Ignorierte letzte Zeiträume müssen als nicht negative ganze Zahl angegeben werden, z. B.: 0, 1, 2, 3.

Der Standardwert ist 0. Wenn keine fehlenden Werte vorhanden sind, werden alle historischen Daten zur Modellgenerierung verwendet und der erste Prognosepunkt liegt nach dem letzten Datenpunkt der historischen Daten. Bis zu 100 Datenpunkte können ignoriert werden.

Das Ignorieren des letzten Datenzeiträume kann nützlich sein, wenn die Daten unvollständig sind. Sie erstellen beispielsweise gerade eine Prognose über einen halben Monat. Schließen Sie diesen Monat aus der Prognose aus, indem Sie **Ignorierte letzte Zeiträume** auf 1 setzen.

In der folgenden Visualisierung ist eine Prognose dargestellt, bei der die Ergebnisse von September durch Festlegen des Werts 1 für die Option **Ignorierte letzte Zeiträume** ignoriert werden.



Konfidenzniveau

Die Gewissheit, mit der der wahre Wert innerhalb des angegebenen Bereichs erwartet wird. Sie können das entsprechende Konfidenzintervall in einer QuickInfo anzeigen, indem Sie den Cursor über einen beliebigen Prognosewert bewegen. Das Konfidenzintervall wird in Form der oberen und unteren Grenze angezeigt.

Sie können drei unterschiedliche Konfidenzniveaus auswählen: 90 %, 95 % und 99 %. Der Standardwert ist 95 %. Die untere und obere Grenze definieren den Bereich, in dem Sie zu 95 % sicher sein können, dass der wahre Wert innerhalb dieses Bereichs liegt.

Saisonaler Zeitraum

Die Saisonalität, mit der das Modell erstellt werden soll. Saisonalität bedeutet, dass die Zeitreihe eine vorhersehbare zyklische Variante hat. Dies kann z. B. ein Urlaubszeitraum in jedem Jahr sein.

Der Standardwert ist **Auto**. Mit **Automatisch** wird die Saisonalität automatisch erkannt, indem mehrere Modelle mit unterschiedlichen saisonalen Zeiträumen erstellt werden und die beste Option ausgewählt wird.

Sie können eine Saisonalität angeben, indem Sie eine nicht negative Ganzzahl wie z. B. 0, 1, 2, 3 als saisonalen Zeitraum angeben.

Zur Angabe eines nicht saisonalen Modells legen Sie **Saisonaler Zeitraum** auf 0 oder 1 fest. Ein Modell mit benutzerdefinierter Saisonalität wird nur dann angezeigt, wenn das saisonale Modell genauer ist als alle nicht saisonalen Modelle.

Insights

Wenn sowohl Insights als auch Prognosen aktiviert und verfügbar sind, wird neben dem Dialogfenster "Prognose" das Dialogfenster "Insights" angezeigt. Jedes Feature stellt eine unabhängige Gruppe von

Analyseergebnissen bereit. Weitere Informationen finden Sie in „Insights in Visualisierungen“ auf Seite 70.

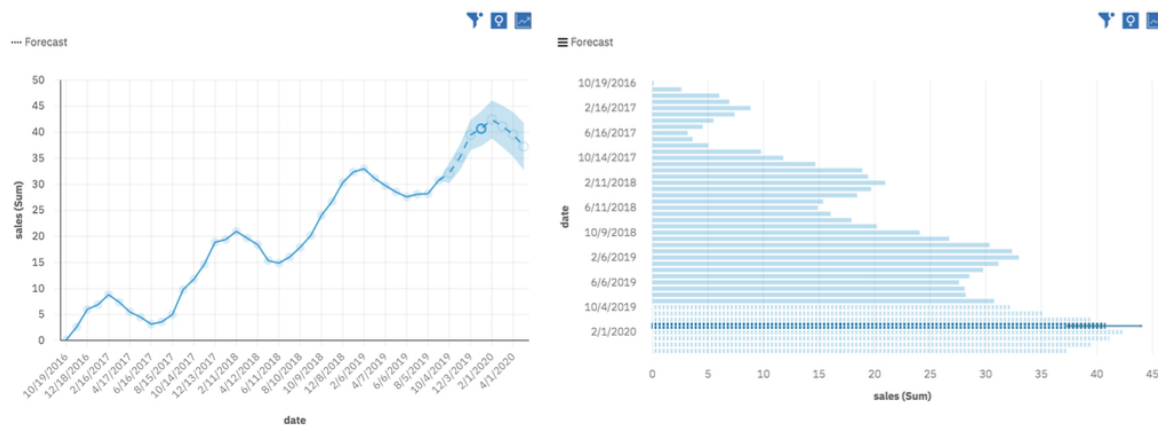
Visualisierungstypen, die die Prognose unterstützen

Die Prognose wird in Liniendiagramm-, Balkendiagramm- und Säulendiagrammvisualisierungen unterstützt.

In der folgenden Tabelle werden die Prognoseanzeigefunktionen für die einzelnen Visualisierungen verglichen.

Prognosefunktionen	Liniendiagramm	Balkendiagramm	Säulendiagramm
Prognosepunkte	Offener Kreis	Gestreifter Balken	Gestreifte Säule
Anzeige des Konfidenzintervalls	Schattierte Region	Durchgezogene Linie	Durchgezogene Linie
Konfidenzintervall aktivieren	Auf beliebigen Punkt klicken	Auf einen Prognosebalken klicken	Auf eine Prognosesäule klicken
Anzahl angezeigter Konfidenzintervalle	Alle	1	1

Die folgende Abbildung zeigt eine Prognose für Liniendiagramm- und Balkendiagrammvisualisierungen mit aktivierten Balkenintervallen.



Insights und Prognosen

Insights in Visualisierungen bieten analytische Insights, die Sie dabei unterstützen, wichtige Beziehungen und aussagekräftige Unterschiede zu erkennen und auszuwerten, basierend auf den Daten, die in der Visualisierung dargestellt werden. Insights können in unterstützten Visualisierungen neben Prognosen verwendet werden. Insights bilden eine separate Gruppe von Analyseergebnissen. Die Ergebnisse haben nur historischen Wert. Weitere Informationen finden Sie in „Insights in Visualisierungen“ auf Seite 70.

Prognosedaten

Für die Prognose geeignete Daten haben Kennzahlenwerte, die Zeitpunkten mit regelmäßigen Abständen entsprechen. Zeit und Kennzahlen in Visualisierungen geben Sie an, indem Sie Zeitfelder und Kennzahlenfelder in die Visualisierungsfelder ziehen. Optional können Sie auch Gruppenfelder angeben, die die Kennzahlenwerte nach Kategorien aufteilen.

Die folgende Tabelle enthält eine Zusammenfassung der Feldtypen und entsprechenden Visualisierungsfelder, die für die Prognose unterstützt werden:

	Zeitfelder (erforderlich)	Kennzahlenfelder (erforderlich)	Gruppenfelder (optional)
Feld 'Liniendiagramm'	X-Achse	Y-Achse	Farbe
Feld 'Balkendiagramm'	Balken	Länge	Farbe
Feld 'Säulendiagramm'	Balken	Länge	Farbe

Andere Visualisierungen oder Visualisierungsfelder werden nicht unterstützt, mit Ausnahme des Feldes **Lokale Filter**.

Zeitfelder in Prognosedaten

Ein Zeitfeld wird durch ein Zeitsymbol vor der Feldbeschriftung im Teilfenster **Daten** angegeben.

Sie können die Eigenschaften eines Zeitfelds mit den folgenden Eigenschaften anpassen: **Datentyp** oder **Repräsentiert Zeit**.

Datentyp

Ein Feld wird als Zeitfeld erkannt, wenn es einen der folgenden Datentypen aufweist: Datum, Uhrzeit oder Zeitmarke. Der Datentyp wird von der Datenquelle übernommen und kann nicht geändert werden.

Die Datentypen Datum, Uhrzeit und Zeitmarke unterstützen das gesamte Spektrum der Datums- und Uhrzeitformate, die von den Basis- und erweiterten Formaten nach ISO 8601 abgedeckt werden. In der folgenden Tabelle sind die unterstützten Datentypen zusammen mit einem Beispiel für das Format und jeweiligem Datenbeispiel aufgeführt.

Datentyp	Formatbeispiel	Datenbeispiel
Datum	jjjj-mm-tt	2019-07-01
Zeit	hh:mm:ss	12:34:56
Zeitmarke	jjjj-mm-tt'T'hh:mm:ss	2019-07-01T12:34:56

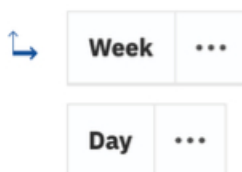
Repräsentiert Zeit

Ein Feld wird als Zeitfeld erkannt, wenn die Dateneigenschaft **Repräsentiert** auf **Zeit** festgelegt ist. Text- und Ganzzahlfelder, die Zeitdaten enthalten, werden ebenfalls als Zeitfelder erkannt. Zeitfelder werden beim Datenimport oder bei der Datenanreicherung automatisch definiert. Mögliche Definitionen sind Datum, Jahr, Quartal, Saison, Monat, Woche, Tag, Stunde, Minute oder Sekunde.

Wenn Zeitfelder nicht automatisch erkannt werden, können Sie sie als Zeitfelder angeben. Stellen Sie sicher, dass die Feldwerte eines der unterstützten Formate aufweisen. Andernfalls erhalten Sie möglicherweise den Fehler **Nicht unterstütztes Datumsformat**.

Verschachtelte Zeitfelder

Sie können mehrere Zeitfelder in dasselbe Visualisierungsfeld ziehen, um ein verschachteltes Zeitfeld anzugeben. Zum Beispiel kann ein Feld, das die Woche darstellt, zusammen mit einem Feld, das den Tag darstellt, in das Feld gezogen werden, um eine Prognose nach Wochentagen zu erstellen.



Verschachtelte Felder im Feld müssen zeitlich hierarchisch angeordnet sein. Zum Beispiel muss die Woche über dem Tag liegen.

Verschachtelte Felder können Ebenen in der Zeithierarchie nicht überspringen, wenn dies zu Mehrdeutigkeiten führen würde. In der folgenden Tabelle sind zulässige Hierarchien beschrieben.

Zeitfeld	Zulässige untere Felder
Jahr	Quartal, Monat, Woche, Tag
Quartal	Monat
Monat	Tag
Woche	Tag
Tag (von Jahr, Monat oder Woche)	Stunde, Uhrzeit
Stunde	Minute
Minute	Sekunde

Wenn Jahr nicht in der Zeithierarchie enthalten ist, wird das System standardmäßig auf das aktuelle Jahr gesetzt. Dies kann zu Problemen aufgrund der Unterschiede zwischen den Schaltjahren und den anderen Jahren führen. Erwägen Sie, in solchen Fällen das Jahr anzugeben.

Chronologische Datenreihenfolge

Angegebene Zeitfelder definieren eine chronologische Reihenfolge für die Zeitpunkte in der Visualisierung. Sie werden verwendet, um die Punkte in der Visualisierung in chronologischer Reihenfolge zu sortieren, wenn die Prognose aktiviert ist. Die chronologische Reihenfolge umfasst die historischen Punkte, zusammen mit den neuen prognostizierten Punkten. Alle anderen Sortierkriterien, die für die Visualisierung angegeben werden, werden bei aktivierter Prognose ignoriert.

Ungültige Zeitbeschriftungen werden an den Beginn der Sequenz verschoben und aus der Erstellung des Modells und der Berechnung der Prognose ausgeschlossen.

Zeitintervallerkennung

Eine Zeitintervallerkennung ist möglich, wenn die Daten chronologisch sortiert sind. Das Zeitintervall ist die Größe des kleinsten Intervalls zwischen zwei benachbarten Zeitpunkten, wie zum Beispiel "2 Wochen". Wenn variierende Zeitintervalle erkannt werden, müssen sie alle ganzzahlige Vielfache des kleinsten Intervalls sein. Andernfalls werden die Daten als unregelmäßig betrachtet und es kann keine Prognose erstellt werden. Fehlende Zeitpunkte, die sich durch mehrfache Intervalle ergeben, werden für das erkannte Intervall angegeben. Entsprechende Kennzahlenwerte werden auf 'fehlend' gesetzt. Übersteigt die Anzahl fehlender Werte 33 % der Serienlänge, wird der Fehler Zu viele fehlende Werte gemeldet.

Kennzahlenfelder

Ein oder mehrere Felder eines beliebigen Typs können als Kennzahlenfelder für die Prognoseanalyse angegeben werden, indem sie einem entsprechenden Visualisierungsfeld hinzugefügt werden. Jedes Kennzahlenfeld wird separat analysiert. Es können auch mehrere Zeitreihen angegeben werden, indem ein Feld zum Feld **Farbe** hinzugefügt wird, wobei die Kennzahlenwerte anhand der Kategorien des angegebenen Felds aufgeteilt werden.

Alle Kennzahlenfeldwerte, die demselben Zeitpunkt entsprechen, werden anhand einer der folgenden Auswertungsstufen ausgewertet: **Summe**, **Minimum**, **Maximum**, **Durchschnitt**, **Anzahl** und **Anzahl eindeutiger Elemente**. Das Feld muss numerisch sein, um die Auswertung **Summe**, **Minimum**, **Maximum** oder **Average** zu unterstützen. Alle zulässigen Datentypen und Auswertungsstufen werden für die Prognose unterstützt. Folgende Aspekte gilt es jedoch zu beachten:

- Eine geringe Anzahl unterschiedlicher Kennzahlenwerte kann zu unerwarteten oder wenig informativen Vorhersagen führen. Das kann z. B. der Fall sein, wenn **Anzahl eindeutiger Elemente** verwendet wird.
- Null Kennzahlenwerte können die Ergebnisse unangemessen beeinflussen, besonders dann, wenn sie fehlende Kennzahlen repräsentieren.

Fehlende Werte interpolieren

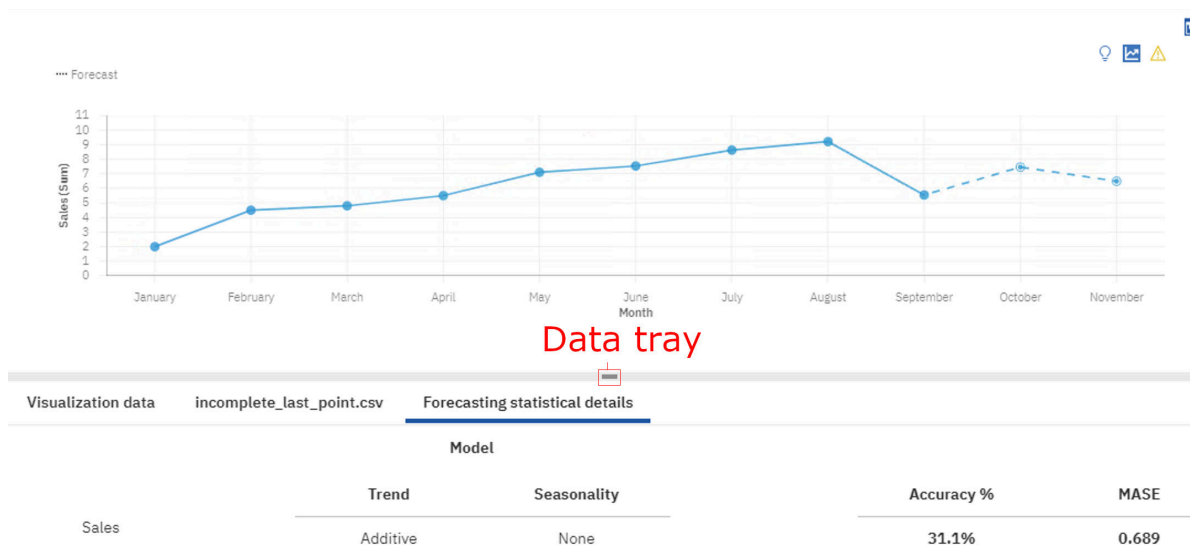
Fehlende Werte werden mit dem Algorithmus **Lineare Interpolation** berechnet und angegeben. Die Berechnung basiert auf den nächsten Nachbarn in einer chronologisch geordneten Zeitreihe mit erkanntem Zeitintervall. Der neue Wert ist $(\text{vorheriger Wert} + \text{nächster Wert}) / 2$. Beispiel: Bei den Werten [3, 6, fehlend, 12] ist der interpolierende Wert, der den fehlenden Wert ersetzen soll, $(6 + 12) / 2$ oder 9. Der Interpolationsalgorithmus kann außerdem aufeinanderfolgende fehlende Werte verarbeiten.

Datenpunkte mit fehlenden Werten an den ersten oder letzten historischen Zeitpunkten werden vor dem Erstellen des Modells aus der Reihe ausgeschlossen. Fehlende Werte an den letzten historischen Zeitpunkten werden ebenfalls prognostiziert.

Prognose - Statistikdetails

Ein Prognoselauf generiert Prognosen und statistische Details zur Prognose. Die statistischen Details zur Prognose befinden sich in der Datenablage unten in den einzelnen Visualisierungen. Für jede Zeitreihe in der Visualisierung gibt es eine einzelne Zeile mit statistischen Details. Prognosedetails werden generiert, solange die Zeitpunkte gleichmäßig verteilt sind.

Die Prognoseinformationen enthalten einen **Status** der Prognose für die angegebene Zeitreihe. Wenn der Status **Erfolg** lautet, werden in den anderen Feldern Details zu dem Modell und zu den für die Prognose verwendeten Daten bereitgestellt. Lautet der Status **Fehler**, werden in einigen der anderen Felder, einschließlich **Hinweise**, Details zur Ursache des Fehlers bereitgestellt. Auswertungen aller Fehler werden stets in den Visualisierungswarnungen bereitgestellt.



In den Informationen zu **Modell** wird für **Trend** und **Saisonalität** der Typ angegeben, der bei erfolgreicher Ausführung für die Schätzung der Zeitreihendaten ausgewählt wird. In der folgenden Tabelle sind die verschiedenen verfügbaren Typen aufgelistet.

Trendkomponente	Saisonale Komponente		
	N	A	M
	KEINE	ADDITIV	MULTIPLIKATIV

Trendkomponente	Saisonale Komponente		
N KEINE	(N, N)	(N, A)	(N, M)
A ADDITIV	(A, N)	(A, A)	(A, M)
Ad ADDITIV_GEDÄMPFT	(Ad, N)	(Ad, A)	(Ad, M)

Genauigkeitskennzahlen

Die Modellgenauigkeitskennzahlen 'Mittlerer absoluter Fehler' (Mean Absolute Error, MAE), 'Mittlerer absoluter skaliertes Fehler' (Mean Absolute Scaled Error, MASE), 'Genauigkeit Prozent' (Accuracy %), 'Fehler-Effektivwert' (Root Mean Squared Error, RMSE), 'Mittlerer absoluter Fehler Prozent' (Mean Absolute Percent Error, MAPE) sowie 'Akaike Information Criterion' (AIC) basieren auf den Zeitreihendaten, mit denen das Modell generiert wird. Alle Genauigkeitskennzahlen basieren auf den historischen Daten. Genauigkeitskennzahlen können auch als Indikator für die Prognosegenauigkeit verwendet werden; sie werden aber nicht als zukünftige Werte übernommen.

Mittlerer absoluter Fehler (Mean Absolute Error, MAE)

Berechnet als die durchschnittliche absolute Differenz zwischen den vom Modell angepassten Werten (stichprobeninterne 'Einen-Schritt-voraus-Prognose') und den beobachteten historischen Daten.

Mittlerer absoluter skaliertes Fehler (Mean Absolute Scaled Error, MASE)

Die Fehlerkennzahl, die für die Modellgenauigkeit verwendet wird. Dies ist der MAE, geteilt durch den MAE des nativen Modells. Das naive Modell ist ein Modell, das den Wert zum Zeitpunkt t als den vorherigen historischen Wert vorhersagt. Die Skalierung nach diesen Fehler bedeutet, dass Sie evaluieren können, wie leistungsfähig das Modell gegenüber dem naiven Modell ist. Ist der MASE größer als 1, ist das Modell weniger leistungsfähig als das naive Modell. Je niedriger der MASE, desto leistungsfähiger ist das Modell gegenüber dem naiven Modell.

Genauigkeit Prozent (Accuracy %)

Der primäre Indikator für die Modellgenauigkeit basierend auf den Prädiktoren. Die Genauigkeit wird als Verkleinerungsprozentsatz des mittleren absoluten Fehlers im Verhältnis zum naiven Modell angegeben. Sie wird berechnet, indem MASE von 1 subtrahiert und als Prozentsatz ausgedrückt wird. Wenn MASE größer-gleich 1 ist, wird die Genauigkeit auf 0 % gesetzt, da sich das Modell gegenüber dem naiven Modell nicht verbessert. Eine höhere Genauigkeit gibt einen geringeren Modellfehler im Verhältnis zum naiven Modell an.

Mittlere Fehlerquadratsumme (Mean Squared Error, MSE)

Die Differenz in der Summe der Quadrate zwischen den vom Modell angepassten Werten und den beobachteten Werten, geteilt durch die Anzahl der historischen Punkte minus die Anzahl der Parameter im Modell. Die Anzahl der Parameter im Modell wird von der Anzahl der historischen Punkte subtrahiert, um Konsistenz zu einer unverzerrten Modellvarianzschätzung herzustellen.

Fehler-Effektivwert (Root Mean Squared Error, RMSE)

Die Quadratwurzel der MSE. Befindet sich in der gleichen Skalierung wie die beobachteten Datenwerte.

Mittlerer absoluter Fehler Prozent (Mean Absolute Percent Error, MAPE)

Die durchschnittliche absolute prozentuale Abweichung zwischen den durch das Modell angepassten Werten und den beobachteten Datenwerten.

Akaike Information Criterion (AIC)

Eine Kennzahl zur Modellauswahl. Das AIC ahndet Modelle mit vielen Parametern und versucht auf diese Weise, das beste Modell bei Präferenz von einfacheren Modellen auszuwählen. Das AIC ist die Summe aus dem Logarithmus des nicht bereinigten MSE, multipliziert mit der Anzahl der historischen

Punkte und der Anzahl der Modellparameter und anfänglichen Glättungsstatus, die mit 2 multipliziert sind.

Parameter

Erkannter saisonaler Zeitraum und Schätzungen für andere Parameter, die im ausgewählten Modell mit exponentiellem Glätten verwendet werden, sind verfügbar.

Saisonaler Zeitraum

Die Anzahl der Zeitschritte in einem saisonalen Zeitraum, verwendet im Modell mit exponentiellem Glätten.

Alpha

Der Glättungsfaktor für Stufenstatus im Modell mit exponentiellem Glätten. Niedrige Werte für Alpha erhöhen die Glättung, d. h. bei geringem Alpha werden mehr historische Daten berücksichtigt. Hohe Werte für Alpha reduzieren die Menge der Glättung, was bedeutet, dass mehr Gewichtung auf die neueren Beobachtungen gelegt wird. Wenn Alpha 1 ist, wird die ganze Gewichtung auf die aktuelle Beobachtung gelegt.

Beta

Der Glättungsfaktor für Trendstatus im Modell mit exponentiellem Glätten. Dieser Parameter verhält sich ähnlich wie Alpha, gilt jedoch für den Trend und nicht den Stufenstatus.

Gamma

Der Glättungsfaktor für Saisonalitätsstatus im Modell mit exponentiellem Glätten. Spielt eine ähnliche Rolle wie Alpha, jedoch für die saisonale Komponente des Modells.

Phi

Der Dämpfungskoeffizient im Modell mit exponentiellem Glätten. Lange Prognosen können zu unrealistischen Ergebnissen führen. Es ist sinnvoll, mit einem dämpfenden Faktor den Trend im Zeitverlauf zu dämpfen und konservativere Prognosen zu erzeugen.

Diagnosen

Zu den Informationen gehören: Fehlende Anzahl, Länge der Serie, Ignorierte Zeiträume, Trendstärke, Saisonalitätsstärke und Datums-/Zeitintervall.

Fehlende Anzahl

Gibt die Anzahl der Datenzeilen an, die entweder fehlende Werte oder fehlende Zeitpunkte aufweisen und zwischen dem ersten und dem letzten gültigen Serienwert positioniert sind. Ungültige Zeitpunkte sowie Punkte mit fehlenden Werten am ersten oder letzten historischen Zeitpunkt sind nicht enthalten.

Länge der Serie

Gibt die Anzahl der Datenpunkte an, die für die Zeitreihenmodellierung verwendet werden. Es sind nur die Punkte zwischen dem ersten und dem letzten gültigen Serienwert (Zeitreihenwert) enthalten.

Ignorierte Zeiträume

Eine Ganzzahl, m , die die letzten m Datenpunkte der Reihe ignoriert, wenn das Modell mit exponentiellem Glätten erstellt wird und die Prognosen berechnet werden. Alle fehlenden Werte am Ende eines nicht ignorierten Teils einer Reihe werden ebenfalls prognostiziert. Der Standardwert für diesen Parameter ist 0, was bedeutet, dass alle historischen Daten bei der Modellgenerierung verwendet werden, wenn keine fehlenden Werte vorhanden sind. Maximal 100 Datenpunkte können ignoriert werden. 'Ignorierte Zeiträume' schließt die Datenpunkte bei der Erstellung eines Modells aus, sodass eine Prognose aufgrund von Faktoren wie Mindestdatenlänge-Anforderungen und fehlender Werteproportion, die 33 % überschreitet, fehlschlagen könnte.

Trendstärke

Vergleich des ursprünglichen Modells (M) mit dem gleichen Modell mit entfernter Trendkomponente. Die Trendstärke von M ist der Unterschied in der Genauigkeit zwischen Modell M und Modell M mit entfernter Trendkomponente.

Saisonalitätsstärke

Vergleich des ursprünglichen Modells (M) mit dem gleichen Modell mit entfernter saisonaler Komponente. Die Saisonalitätsstärke von M ist die Differenz in der Genauigkeit zwischen Modell M und Modell M mit entfernter saisonaler Komponente.

Datums-/Zeitintervall

Das Datums-/Zeitintervall stellt das erkannte Zeitintervall der chronologisch sortierten Daten dar. Das Zeitintervall wird als kleinste Differenz zwischen benachbarten Punkten in den Daten identifiziert, wenn sie in chronologischer Reihenfolge sortiert sind.

Prognosemodelle

Modelle mit exponentiellem Glätten sind eine vielfach eingesetzte Klasse von Zeitreihenmodellen.

Modelle mit exponentiellem Glätten sind auf einen einzelnen Satz von Werten anwendbar, die nur in gleichen Zeitinkrementen aufgezeichnet werden. Sie unterstützen jedoch Dateneigenschaften, die in Geschäftsanwendungen häufig zu finden sind, wie Trend, Saisonalität und Zeitabhängigkeit. Alle angegebenen Modellfeatures werden auf Basis der verfügbaren beobachteten Daten geschätzt. Ein geschätztes Modell kann anschließend zur Prognose zukünftiger Werte und für die Bereitstellung von oberen und unteren Konfidenzgrenzen für die Prognosewerte verwendet werden.

Jeder Modelltyp eignet sich für die Modellierung einer anderen Kombination von Eigenschaften, die in den Daten gefunden werden. Der Modelltyp mit der besten Übereinstimmung mit den beobachteten Daten wird zum Modellieren der beobachteten Daten ausgewählt und für die Prognose zukünftiger Werte verwendet.

Modellschätzalgorithmen

Modelle werden durch die Glättungsgleichungen angegeben, die die Modellparameter und die anfänglichen Glättungsstatus enthalten. Die Modellparameter werden mit Werten geschätzt, die den Modellfehler minimieren.

Glättungsgleichungen

Modelle mit exponentiellem Glätten leiten ihren Namen aus den Glättungsgleichungen ab, die das Modell angeben. Sie stellen Formeln für die Berechnung von Glättungsstatus für jeden beobachteten Punkt bereit, indem sie den aktuellen beobachteten Wert und die vorherigen Glättungsstatus heranziehen. Glättungsgleichungen liefern gewichtete Durchschnittswerte des aktuellen Werts und der vorherigen Status in der Zeitreihe. Die Gewichtung für den aktuellen Wert oder aktuellen Status wird durch einen Modellparameter zwischen 0 und 1 angegeben, während die Gewichtungen für vorherige Werte exponentiell abnehmen.

Gleichungen für die Stufenglättung

Alle Modelltypen berechnen einen Stufenstatus für jeden Zeitreihenpunkt unter Verwendung der entsprechenden Stufenglättungsgleichung. Die Stufenstatus für das Modell ohne Trend- und saisonale Komponenten werden als gewichteter Durchschnitt des Zeitreihenwerts am aktuellen Punkt und des Stufenstatus am vorherigen Punkt berechnet. Die Gewichtung, die dem aktuellen Wert zugeordnet ist, ist ein Parameter **alpha**, dessen Wert zwischen 0 und 1 liegen muss. Bei anderen Modellen werden auch die vorherigen Trend- und saisonalen Status in die Stufenglättungsgleichung einbezogen.

Gleichungen für die Trendglättung

Modelltypen mit additivem oder gedämpft-additivem Trend berechnen einen Trendstatus für jeden Zeitreihenpunkt unter Verwendung der entsprechenden Trendglättungsgleichung. Der Trendstatus für den aktuellen Punkt basiert auf der Differenz der Stufenstatus am aktuellen und am vorherigen Punkt sowie auf dem Trendstatus am vorherigen Punkt. Die Gewichtung, die der Differenz der Stufenstatus beim aktuellen und beim vorherigen Punkt zugeordnet ist, ist der Parameter **beta**, dessen Wert zwischen 0 und 1 liegen muss. Bei gedämpften Trendglättungsgleichungen wird ein weiterer Parameter (**phi**) hinzugefügt. **Phi** multipliziert den Trendstatus-Beitrag vom vorangehenden Punkt; sein Wert muss ebenfalls zwischen

0 und 1 liegen. Zweck dieses Parameters ist die Schätzung des Grades an Trenddämpfung von einem Punkt zum nächsten.

Gleichungen für die Saisonalitätsglättung

Modelltypen, die additive oder multiplikative Saisonalität unterstützen, berechnen für jeden Zeitreihenpunkt einen Saisonalitätsstatus. Die Saisonalitätsstatus werden anhand von Glättungsgleichungen für die Saisonalität berechnet. Der Saisonalitätsstatus für den aktuellen Punkt beinhaltet bei additiver Saisonalität die Differenz zwischen dem Zeitreihenwert und dem aktuellen Stufenstatus bzw. bei multiplikativer Saisonalität das Verhältnis zwischen diesen beiden Werten. Die Gewichtung, die dieser Bedingung zugeordnet ist, ist der Parameter **gamma**, dessen Wert zwischen 0 und 1 liegen muss. Der restliche Beitrag stammt aus dem entsprechenden Saisonalitätsstatus im vorherigen saisonalen Zeitraum. Beachten Sie, dass der saisonale Zeitraum eine feste Länge hat, und während sich der Saisonalitätsstatus für jeden Punkt ändern kann, werden in den Glättungsgleichungen für die Saisonalität nur übereinstimmende saisonale Indizes aus unterschiedlichen Zeiträumen zusammen berücksichtigt.

Anfängliche Glättungsstatus

Werte müssen für Stufen-, Trend- und Saisonalitätsstatus für die Punkte angegeben werden, die der Zeitreihe vorangehen. Die Werte sind für die Glättungsgleichungen erforderlich. Um die verschiedenen Status am ersten Punkt der Zeitreihe zu berechnen, sind Statuswerte an den entsprechenden vorherigen Punkten erforderlich.

Modellparameter

Jede Glättungsgleichung verwendet entsprechende Modellparameter:

alpha

Steuert die Stufenstatus.

beta

Steuert die Trendstatus.

gamma

Steuert saisonale Indizes über saisonale Zeiträume hinweg.

phi

Ein zusätzlicher Parameter zur Angabe des gedämpften Trends.

Alle vier Parameter haben Werte zwischen 0 und 1. Höhere Werte für **alpha**, **beta**, and **gamma** bedeuten, dass neuere Beobachtungen höher gewichtet wurden; niedrigere Werte stehen für die höhere Gewichtung von älteren Beobachtungen. Ein höherer Wert von **phi** entspricht einem höheren Grad der Dämpfung des Prognosetrends.

Modellschätzung

Modellparameter in den Glättungsgleichungen werden basierend auf den Zeitreihendaten geschätzt. Parameter können nicht direkt anhand einer Formel geschätzt werden. Sie werden durch einen iterativen Prozess geschätzt, der nach Parameterwerten sucht, die den Modellfehler minimieren. Der Modellfehler wird als Mittlerer absoluter skaliertes Fehler (Mean Absolute Scaled Error, MASE) berechnet. Die Iterationen werden gestoppt, wenn keine weitere Verkleinerung des Modellfehlers erreicht werden kann. Die entsprechenden Parameterwerte zusammen mit den anfänglichen Glättungsstatus geben das geschätzte Modell vollständig an. Sie werden verwendet, um die Modellstatus für alle anderen Datenpunkte zu berechnen und die Modellprognosen unter Verwendung einer entsprechenden Prognosegleichung zu generieren.

Prognosealgorithmen

Bei der Prognose werden zahlreiche Algorithmen verwendet.

Einen Schritt voraus

Jedes Modell unterstützt Einen-Schritt-voraus-Prognosen auf Basis der entsprechenden Prognosegleichung. Einen-Schritt-voraus-Prognosen werden zur Berechnung von Modellfehlern während des Modellschätzprozesses benötigt.

Einen-Schritt-voraus-Prognosen werden sequenziell für jeden Datenpunkt anhand der berechneten Stufen- und Trendstatus für den aktuellen Punkt und der saisonalen Status für den letzten saisonalen Zeitraum berechnet.

Der Prognosefehler wird berechnet, indem der Prognosewert am vorherigen Punkt vom beobachteten Wert am aktuellen Punkt subtrahiert wird. Der Gesamtmodellfehler, anhand dessen das Modell geschätzt wird, wird als Durchschnittswert für absolute Prognosefehler berechnet. Kleinere Fehler entsprechen einer besseren Modellanpassung. Die in **Prognose - Statistikdetails** angezeigten Genauigkeitskennzahlen bieten mehrere Modellauswertungen der Fehler der Einen-Schritt-voraus-Prognose.

k Schritte voraus

k-Schritte-voraus-Prognosen werden verwendet, um Vorhersagen für eine beliebige Anzahl zukünftiger Werte nach den beobachteten Zeitreihendaten zu erstellen. Sie basieren auf denselben Vorhersagegleichungen wie die 'Einen-Schritt-voraus'-Prognosen für das angegebene Modell.

Die Anzahl der Prognosewerte, die generiert werden, beträgt standardmäßig 20 % der Länge der historischen Datenserie. Im Dialogfeld **Prognose** können Sie eine genaue Anzahl von Werten angeben, die prognostiziert werden sollen. Alle fehlenden Werte am Ende einer bestimmten Reihe werden ebenfalls prognostiziert, sie werden jedoch nicht auf die angegebene Anzahl von Prognosezeiträumen angerechnet.

Konfidenzgrenzen

Die Konfidenzgrenzen geben den Grad der Unsicherheit an, der den einzelnen Prognosewerten zugeordnet ist. Typischerweise weiten sich diese Grenzen aus, je weiter sie in der Zukunft liegen, da entfernte Prognosen an Zuverlässigkeit abnehmen. Konfidenzgrenzen liefern relevante Einblicke in das zukünftige Verhalten der beobachteten Zeitreihen.

Die Berechnung von Konfidenzgrenzen basiert auf der Gesamtvarianz der Prognosefehler, die anhand der beobachteten Daten geschätzt werden, und auf einem Faktor, der vom angegebenen Modell und von der Anzahl der Schritte ab dem letzten beobachteten Punkt abhängt.

Automatische Modellauswahl in Prognosen

Es werden mehrere Modelltypen verwendet, um Kandidatenmodelle für jede Zeitreihe in einer Prognose zu erstellen. Normalerweise werden alle neun verfügbaren Modelltypen verwendet, es sei denn, eine saisonale Komponente fehlt. Es sind nur drei Modelltypen verfügbar, die die Saisonalität in den Daten nicht berücksichtigen.

Der Standardwert **Automatisch** für die Option "Saisonaler Zeitraum" erkennt den saisonalen Zeitraum durch einen Vergleich mehrerer Modelle mit einem jeweils anderen saisonalen Zeitraum.

Anhand eines Modellfehlers und der Anzahl der Modellparameter werden mehrere Modelle verglichen. Wenn zum Beispiel die Modellfehler für zwei Modelle gleich sind, wird das Modell mit weniger Parametern bevorzugt. Das letztere Modell liefert eine stärker verdichtete Darstellung der beobachteten Daten und generiert auch eher eine zuverlässige Prognose.

Kapitel 5. Principles of advanced data analytics

Richtlinien für die erweiterte Datenanalyse

IBMCognos Analytics ist ein Business-Intelligence-Tool zur Verwaltung und Analyse von Daten. Es enthält eine Reihe unterschiedlicher Self-Service-Funktionen, die es den Benutzern ermöglichen, Daten aufzubereiten, zu untersuchen und zwecks gemeinsamer Nutzung mit anderen zu teilen. Im Rahmen dieses Produktangebots umfasst Cognos Analytics verschiedene Verfahren für Prognosen, Beschreibungen und Explorationen, die auch als numerische Intelligenz bekannt sind. Cognos Analytics verwendet zahlreiche statistische Tests für die Datenanalyse.

Es ist wichtig, die Definitionen dieser Tests in Cognos Analytics zu kennen.

Als Teil des Workflows werden numerische Algorithmen verwendet, um den Benutzern Features zur Verfügung zu stellen, die Informationen zu den numerischen Eigenschaften und Beziehungen in den Benutzerdaten vermitteln.

Geschäftsorientiert

Im Gegensatz zu konventioneller Statistiksoftware, deren Zielgruppe erfahrene Datenanalysten sind, richten sich die Algorithmen von Cognos Analytics an Benutzer, die sich mit Datenanalysen zwar auskennen, aber keine Experten auf diesem Gebiet sind. Dies bedeutet, dass bei der Berücksichtigung von Kompromissen in Cognos Analytics mehr Wert auf Zweckmäßigkeit als auf Komplexität gelegt wird.

Vertrauenswürdig

Geschäftsdaten sind in Wirklichkeit viel komplizierter als das, was in Lehrbuchbeispielen in Statistikkursen verwendet wird oder in Beispielen bei der Websuche gefunden wird. Cognos Analytics verwendet stabile Algorithmen, die in der Lage sind, ein breites Spektrum an ungewöhnlichen Daten zu verarbeiten. Bei Cognos Analytics hat man sich aus gutem Grund für stabile Algorithmen entschieden. Zwar lassen sich mit labileren Algorithmen geringfügig bessere Ergebnisse erzielen als mit stabilen Algorithmen, doch als Voraussetzung hierfür muss der Benutzer sicherstellen, dass die labilen Algorithmen auch anwendbar sind und Daten korrekt konvertiert werden, damit die Ergebnisse aussagekräftig sind. Eine geringfügig niedrigere Genauigkeit wird durch die Zuverlässigkeit aufgewogen, die robuste Algorithmen bieten, da sie auch dann keine ungültigen Ergebnisse liefern, wenn die Daten nicht in der Form vorliegen, wie sie sollten.

Intelligent

Bei nahezu allen Algorithmen müssen bestimmte Entscheidungen getroffen werden, die mit den Algorithmen selbst zu tun haben. Dies betrifft beispielsweise Aspekte wie das Konfidenzniveau, die zu untersuchenden Kombinationen von Feldern und die Datenkonvertierung. Einzelheiten zu diesen Entscheidungen finden Sie in den Beschreibungen.

Geeignete Werte werden von Cognos Analytics automatisch ausgewählt. Diese Auswahl beruht auf einer Untersuchung der Eigenschaften der betreffenden Daten. Als Benutzer ist es unter Umständen nicht immer möglich, alle getroffenen Entscheidungen nachzuvollziehen.

Auswertung

In Cognos Analytics sind die numerischen Algorithmen und Prozeduren so konzipiert, dass vertrauenswürdige Ergebnisse automatisch erzeugt werden. Die bestmöglichen Vorhersagen, Klassifizierungen oder Analysen werden erzielt, wenn ein professioneller Statistiker die Daten mithilfe von IBM SPSS Statistics oder IBM SPSS Modeler analysiert. Das Ziel von Cognos Analytics besteht darin, Ihnen qualitative Insights zu liefern, die Sie dabei unterstützen, Ihre Daten und deren Beziehungen untereinander besser zu verstehen, und dies soll automatisch für eine Vielzahl von Datentypen erfolgen. Cognos Analytics ist darauf

ausgerichtet, ähnliche Ergebnisse zu liefern wie ein professioneller Statistiker, ohne dem Geschäftsbutzer im Wege zu stehen.

Datenaufbereitung

Bei der Datenaufbereitung handelt es sich um einen Schritt im Rahmen der Vorabanalyse, der von den meisten Datenanalysealgorithmen verwendet wird, um sicherzustellen, dass die Daten für die analytische Verwendung geeignet sind.

Übersicht

Die Datenaufbereitung ist in IBM Cognos Analytics von wesentlicher Bedeutung. Nur aufbereitete Daten werden in die Analyse für zentrale Einflussfaktoren, Entscheidungsstrukturen und Beziehungen eingegeben, die in den Visualisierungen erweiterter Analysen angezeigt werden: Spiraldiagramm, Einflussfaktoranalyse, Entscheidungsstruktur, Sunburst-Diagramm und Explore-Beziehungen. Daten werden nicht automatisch für andere Visualisierungen und ihre entsprechenden Insights aufbereitet.

Algorithmen

Alle angewandten Algorithmen basieren jeweils immer nur auf den Werten eines einzelnen Felds. Fehlende Werte werden für jedes Feld entfernt oder verarbeitet, alle numerischen Prädiktorfelder für Einflussfaktoren werden in Klassen (Bins) eingeteilt. Alle kategorialen Felder werden im Hinblick auf eine große Anzahl von Kategorien entsprechend angepasst, und Ausreißer werden im Zielfeld bearbeitet. Wenngleich die Datenaufbereitung immer einen Einfluss auf die Analyseergebnisse hat, werden derzeit keine entsprechenden Auswertungen zur Datenaufbereitung an die Benutzer berichtet.

Details

Die Datenaufbereitung sowie die nachfolgenden zentralen Einflussfaktoren, Entscheidungsstrukturen und Beziehungen basieren auf einer Datenstichprobe mit ungefähr 10.000 Zeilen, wenn die ursprünglichen Daten umfangreicher sind. Auf hochgeladene Daten und alle verbundenen Datenquellen, die Zufallsstichproben unterstützen, wird die Zufallsstichprobe nach Bernoulli angewandt (gleiche Wahrscheinlichkeit ohne Ersatzzufallsstichprobe). Andernfalls wird eine systematische Stichprobe verwendet.

Datenaufbereitung für numerische Felder

Ein Feld wird immer dann als numerisch behandelt, wenn es numerische Informationen enthält und seine Verwendungseigenschaft auf 'Kennzahl' gesetzt ist.

Übersicht

Da numerische Daten unterschiedlich verteilt sein können, wandelt IBM Cognos Analytics nicht zielbezogene numerische Felder in ordinale Klassen (Bins) um, wodurch sich die Abhängigkeit der Analysealgorithmen vom Format der numerischen Daten reduziert.

Algorithmen

Der verwendete Basisalgorithmus entspricht der Klasseneinteilung (Binning) nach Häufigkeit. Numerische Daten werden in eine festgelegte Anzahl von Bins aufgeteilt; hierbei wird versucht, in jeden Bin dieselbe Anzahl an Datenzeilen zu stellen. Fehlende Werte werden in einen eigenen Bin gestellt. In Cognos Analytics wird versucht, das Wissen über fehlende Werte in Prädiktorfeldern entsprechend zu verwenden, um ein besseres Modell zu erstellen. Beispiel: Wenn ein Datenfeld darstellt, wann ein Element getestet wurde, verwendet Cognos Analytics fehlende Werte (die möglicherweise darstellen, dass ein Element niemals getestet wurde), um bei der Vorhersage der Werte anderer Felder zu helfen.

Details

Für numerische Felder gelten bestimmte Feldausschlusskriterien. Ein numerisches Feld wird aus der weiteren Analyse ausgeschlossen, wenn es nur einen einzigen Wert enthält, einschließlich des fehlenden Werts. Andernfalls wird das numerische Feld einem Bin zugeteilt, wobei die Standardanzahl der Bins 5 beträgt. Wenn ein Feld maximal 10 eindeutige numerische Werte enthält, wird kein Binning versucht, und jeder eindeutige Wert erhält eine eigene Kategorie. Wenn in mehr als 40 % der Zeilen Null vorkommt, wird dafür immer eine eigene Kategorie eingerichtet. Fehlende Werte werden in einen eigenen Bin gestellt und haben keine Auswirkungen auf die Binning-Prozedur.

Datenvorbereitung für kategoriale Felder

Ein Feld wird immer dann als kategorial behandelt, wenn seine Verwendungseigenschaft auf 'Attribut' oder 'Identifizierung' gesetzt ist.

Übersicht

Die Hauptinformation, die aus kategorialen Feldern extrahiert wird, ist die beobachtete Häufigkeit für jeden eindeutigen Kategorienwert. Auf kategoriale Felder werden entsprechend geeignete Analysemethoden angewandt. Die Genauigkeit und Leistung dieser Methoden kann jedoch bei einer steigenden Anzahl unterschiedlicher Kategorien negativ beeinflusst werden. Der Hauptschritt bei der Datenaufbereitung besteht darin, die Zusammenführung von Kategorien einzuleiten, wenn deren Anzahl ansteigt.

Algorithmen

Der grundlegende Algorithmus, der verwendet wird, besteht in der Zusammenführung von Kategorien. Kategorien werden nach ihrer Häufigkeit in absteigender Reihenfolge sortiert. Übersteigt die Anzahl der Kategorien einen zuvor festgelegten Standardwert, werden die überzähligen Kategorien zu einer einzelnen Kategorie zusammengeführt. Fehlende Werte werden als einzelne separate Kategorie behandelt. Mit anderen Worten: IBM Cognos Analytics verwendet fehlende Werte auf ähnliche Weise wie für numerische Felder. Kategoriale Felder werden als nominal behandelt. Bei den Kategorien untereinander wird keine intrinsische Reihenfolge angenommen.

Details

Für kategoriale Felder gelten bestimmte Feldausschlusskriterien. Ein kategoriales Feld wird aus der weiteren Analyse ausgeschlossen, wenn es nur einen einzigen Wert aufweist oder wenn die Anzahl der eindeutigen, nicht zusammengeführten Kategorien größer ist als 50 % der Anzahl gültiger Datenzeilen.

Andernfalls wird das kategoriale Feld zusammengeführt; die Standardanzahl nicht zusammengeführter Kategorien beträgt 49. Die verbleibenden Kategorien werden zu einer einzelnen zusätzlichen Kategorie zusammengeführt. Alle Kategorien mit weniger als drei Zeilen werden ebenfalls zusammengeführt. Ein kategoriales Feld wird ebenfalls ausgeschlossen, wenn der Prozentsatz von gültigen Datenzeilen, die der zusammengeführten Kategorie entsprechen, den Wert von 25 % überschreitet.

Fehlende Werte werden als separate Kategorie behandelt und im Zusammenführungsschritt als solche berücksichtigt.

Datenaufbereitung für Zielfelder

Die Angabe eines Zielfelds ist für Visualisierungen des Typs 'Zentrale Einflussfaktoren' und 'Entscheidungsstruktur' erforderlich.

Übersicht

Geben Sie immer das Zielfeld und mindestens ein zusätzliches Feld an. Modelle werden durch die Verwendung von bereitgestellten Zielwerten trainiert und auf dieser Grundlage dazu verwendet, Prognosebeziehungen zu erkennen und schließlich Zielwerte anhand der Werte in den Eingabefeldern vorherzusagen. Die Datenaufbereitung für das Zielfeld unterscheidet sich von der Datenaufbereitung für den Rest der

Felder. Fehlende Werte im Ziel werden nicht für die Erstellung von Modellen verwendet; die übrigen Informationen bleiben jedoch erhalten und werden bisweilen korrigiert, um unverzerrte Modelle zu erhalten.

Algorithmen

Der Hauptschritt bei der Datenaufbereitung im Zusammenhang mit Zielfeldern besteht darin, alle Datenzeilen mit fehlendem Zielwert zu entfernen. Dies geschieht vor allen anderen Datenvorbereitungsschritten. Durch diesen Schritt wird zwar sichergestellt, dass nur zuverlässige Informationen für die Erstellung von Modellen verwendet werden. Allerdings kann die Anzahl der entfernten Zeilen erheblich sein. In diesem Fall ist es möglich, dass das resultierende Modell nur einen eingeschränkten Geltungsbereich hat. Numerische Zielfelder werden nicht in Klassen eingeteilt, das heißt, es wird kein Binning durchgeführt. Dennoch werden extreme Ausreißer verarbeitet, um die zu einem späteren Zeitpunkt erstellten Modelle nicht zu beeinträchtigen. Kategoriale Zielfelder werden überwiegend wie andere kategoriale Felder behandelt. Der einzige Unterschied besteht darin, dass für die kategorialen Ziele die fehlenden Werte entfernt worden sind.

Details

Extreme Ausreißer werden anhand von Unter- und Obergrenzen ermittelt. Die Obergrenze wird ermittelt, indem zunächst ein oberes Perzentil so verwendet wird, dass lediglich noch 2,5 % der Zielwerte einen größeren Wert aufweisen. Die Differenz zwischen dem oberen Perzentil und dem Median wird mit 2,5 multipliziert und zum Median addiert, um die Obergrenze zu erhalten. Die Untergrenze wird durch ähnliche Berechnungsschritte ermittelt. Die Zielwerte, die jenseits der berechneten Grenzen gefunden werden, werden in allen nachfolgenden Analysen durch den entsprechenden Grenzwert ersetzt.

Unidirektionale zentrale Einflussfaktoren

Bei den unidirektionalen Einflussfaktoren handelt es sich um ein modellbasiertes Explorationstool.

Übersicht

Anhand eines Zielfelds verwendet das Tool ein statistisches Modell zur Analyse aller anderen verfügbaren Datenfelder und schätzt deren Stärke bei der Vorhersage der Zielwerte. Solche Datenfelder werden als Zielprädiktoren oder Einflussfaktoren bezeichnet. Jedes potenziell relevante Datenfeld wird analysiert, wobei jedoch nur die Einflussfaktoren mit den höchsten Werten hinsichtlich ihrer Prognosestärke angezeigt werden. So erhalten Sie Insights im Hinblick auf die verfügbaren Einflussfaktoren und ihre Rangordnung gemäß ihrer Prognosestärke für das in den Daten angegebene Ziel. Die Ergebnisse der Analyse unidirektionaler Einflussfaktoren sind sowohl in Visualisierungen vom Typ 'Einflussfaktoranalyse' als auch in Visualisierungen vom Typ 'Spiraldiagramm' verfügbar. Visuelle Drilldown-Operationen für jeden einzelnen Einflussfaktor sind nur für die Visualisierung von Einflussfaktoranalysen in Explore aktiviert.

Algorithmen

Die Analyse jedes einzelnen unidirektionalen Einflussfaktors basiert auf einem statistischen Modell, das das Ziel und einen einzelnen kategorialen Prädiktor umfasst. Das Modell wird nach dem Datenvorbereitungsschritt für das Zielfeld und alle potenziellen Prädiktorfelder angewendet. Beispiel: Alle numerischen Prädiktorfelder werden während des Datenaufbereitungsschritts in Klassen, sogenannte Bins, eingeteilt und in der Analyse als kategoriale Felder behandelt. Die unidirektionale VA wird für numerische Ziele angewendet und der Chi-Quadrat-Test der Unabhängigkeit wird für kategoriale Ziele mit der Chi-Quadrat-Anpassung für spärliche Daten angewendet.

Für jedes Feld in der Liste potenzieller Einflussfaktoren wird ein Hypothesentest durchgeführt, um zu ermitteln, ob das Feld signifikante Auswirkungen auf das Ziel hat. Nur diejenigen Felder, die den Test bestehen und eine ausreichend hohe Prognosestärke aufweisen, werden als mögliche unidirektionale zentrale Einflussfaktoren ausgewählt.

Details

In einigen Fällen wird die Anzahl der potenziellen Einflussfaktoren mithilfe einer vorläufigen Analyse auf der Grundlage von intelligenten Funktionen reduziert. Hierbei besteht das Ziel darin, irrelevante oder redundante Felder zu entfernen. Die Liste der verwendeten Einflussfaktoren ist in der Benutzerschnittstelle verfügbar und Sie können alle anfänglich ausgeschlossenen Einflussfaktoren wieder zur Analyse hinzufügen. Die resultierenden 20 größten Einflussfaktoren mit einer Prognosestärke von mehr als 10 % können angezeigt werden.

Zur Verbesserung der Leistung und Verarbeitungsgeschwindigkeit werden einige Einschränkungen im Hinblick auf die Datengröße umgesetzt. Wenn die Daten mehr als 250 Felder enthalten, werden die Felder mit der geringsten Relevanz bereits vor der Einflussfaktoranalyse ausgeschlossen. Wie vorstehend beschrieben, haben Sie die Möglichkeit, die ausgeschlossenen Felder über die Benutzerschnittstelle wieder zur Analyse hinzuzufügen. Wenn die angegebenen Daten mehr als 10.000 Zeilen enthalten, kann es sein, dass zwecks Einflussfaktoranalyse eine Stichprobe von ungefähr 10.000 Zeilen erstellt wird. In solchen Fällen wird eine Warnung angezeigt: *Um die Leistung zu verbessern, basiert die Analyse aufgrund der Anzahl der Zeilen in der Datenquelle auf einer repräsentativen Stichprobe aller Daten.* Es wird davon ausgegangen, dass die Ergebnisse nahe an die Ergebnisse herankommen, die erzielt würden, wenn sämtliche Zeilen in den ursprünglichen Daten verwendet würden.

Bidirektionale zentrale Einflussfaktoren

Bei bidirektionalen Einflussfaktoren kommt es darauf an, Paare von kategorialen Prädiktoren gleichzeitig zu modellieren und nach Rangordnung einzustufen.

Übersicht

Anhand eines Zielfelds verwendet IBM Cognos Analytics ein statistisches Modell zur Analyse eines Paares anderer Datenfelder und schätzt dessen Stärke bei der Vorhersage der Zielwerte. Die Suche in unterschiedlichen Prädiktorpaaren ist in der Regel nicht erschöpfend und es kann auch sein, dass einige hochrangige Paare aus den Endergebnissen herausgefiltert werden. Das Ziel besteht darin, eine Übersicht und Vielfalt von Prädiktorpaaren bereitzustellen, die die Prognosestärke einzelner Vorhersagemodelle verbessern, die als unidirektionale Einflussfaktoren angezeigt werden. Daher werden die von unidirektionalen Einflussfaktoren gewonnenen Insights erweitert und Benutzer erhalten relevante Informationen zu den Feldpaaren in den Daten. Die Ergebnisse der Analyse sowohl von unidirektionalen als auch von bidirektionalen Einflussfaktoren sind in den Einflussfaktoranalyse- und Spiraldiagrammen verfügbar. Sie können separat angezeigt werden, indem Sie eine entsprechende Diagrammanzeigeoption auswählen. Jeder angezeigte unidirektionale oder bidirektionale Einflussfaktor kann direkt über die Visualisierung einer Einflussfaktoranalyse in Explore in eine neue Visualisierung erweitert werden.

Algorithmen

Die Analyse jedes einzelnen bidirektionalen Einflussfaktors basiert auf einem statistischen Modell, das das Ziel und ein Paar kategorialer Prädiktoren umfasst. Das Modell wird im Anschluss an die Datenaufbereitung und die Erstellung aller unidirektionalen Einflussfaktoren angewendet. Der erste Prädiktor im Paar wird aus den obersten 50 unidirektionalen Einflussfaktoren ausgewählt und der zweite wird aus den obersten 25 unidirektionalen Einflussfaktoren ausgewählt. Mit dieser Suchstrategie wird sichergestellt, dass die meisten der höchstrangigen Prädiktorpaare für das Modellieren berücksichtigt werden. Die bidirektionale VA (Varianzanalyse) wird für numerische Ziele angewendet und der Chi-Quadrat-Test der Unabhängigkeit wird für kategoriale Ziele mit der Chi-Quadrat-Anpassung für spärliche Daten angewendet.

Für jedes berücksichtigte Feldpaar wird ein Hypothesentest durchgeführt, um zu ermitteln, ob das Paar signifikante Auswirkungen auf das Ziel hat. Nur diejenigen Paare, die den Test bestehen und eine ausreichend hohe Prognosestärke aufweisen, werden als mögliche bidirektionale Einflussfaktoren ausgewählt.

Details

Die Einschränkungen, die bei der Auswahl von Datenfeldern und Datenzeilen für unidirektionale Einflussfaktoren gelten, finden auch bei den bidirektionalen Einflussfaktoren Anwendung. Dies ist auch zu erwarten, da die potenziellen Prädiktorfelder für bidirektionale Einflussfaktoren aus den obersten unidirektionalen Einflussfaktoren auf Basis ihrer jeweiligen Prognosestärke ausgewählt werden. Nicht erforderlich für den Einschluss unidirektionaler Einflussfaktoren in ein bidirektionales Modell sind hingegen deren Modellsignifikanz und die minimale Prognosestärke. Der resultierende bidirektionale Einflussfaktor muss eine Prognosestärke von über 10 % aufweisen und eine relative Verbesserung von über 10 % im Vergleich zur jeweiligen Prognosestärke der einzelnen enthaltenen unidirektionalen Einflussfaktoren mit sich bringen. Die relative Verbesserung wird als Prozentsatz der Differenz zwischen 100 % und der Prognosestärke des verschachtelten unidirektionalen Einflussfaktors berechnet. Die resultierenden bidirektionalen Einflussfaktoren, die diese Kriterien erfüllen, werden nach ihrer Prognosestärke in eine Rangordnung gebracht und die obersten 20 werden zum Anzeigen verfügbar gemacht.

Entscheidungsstruktur

Bei Entscheidungsstrukturen handelt es sich um komplexere Modelle als unidirektionale und bidirektionale Einflussfaktoren. Als Kombinationsmodelle erweitern Sie die Modellreihe. Der Hauptunterschied besteht darin, dass Entscheidungsstrukturen die Erkennung von Interaktionen unter mehreren Prädiktoren unterstützen und daher aussagekräftigere Informationen liefern als die Einflussfaktoren.

Übersicht

Anhand des Zielfelds durchsucht der Algorithmus sämtliche anderen Datenfelder und fügt diese dem Modell entsprechend hinzu, um dessen Prognosestärke bei der Vorhersage der Zielwerte zu verbessern. Die Durchsuchung unterschiedlicher Prädiktoren ist iterativ. Nachdem die Suche einen Prädiktor hinzugefügt hat, wird der Suchvorgang fortgesetzt, um den nächsten Prädiktor hinzuzufügen, der das Modell am meisten verbessert. Ziel ist es, die bestmögliche Menge von Prädiktoren zu finden und diese Prädiktoren optimal zu kombinieren, sodass ein optimales Modell berechnet wird. Die aus Entscheidungsstrukturen gewonnenen Insights werden in Form von Entscheidungsregeln dargestellt, wobei die Kombination aus Prädiktoren und zugehörigen Werten eine einzelne Prognose für den Zielwert liefert. Die Entscheidungsregeln werden nach Stärke sortiert, sodass Sie ohne großen Aufwand die Regeln finden können, die am relevantesten und interessantesten sind. Die von der Entscheidungsstruktur generierten Entscheidungsregeln schließen sich gegenseitig aus. Darüber hinaus liefern die Entscheidungsregeln einen vollständigen Regelsatz, sodass für jede Kombination der Prädiktorwerte in den Daten eine entsprechende Regel vorhanden ist. Des Weiteren wird die Gesamtprognosestärke der Entscheidungsstruktur angegeben, die die relative Verbesserung im Vergleich zum Basismodell darstellt. Die Ergebnisse werden mithilfe von drei unterschiedlichen Visualisierungen zur Verfügung gestellt: Sunburst, Verzeichnisstruktur und Entscheidungsregeln. Jede dieser Visualisierungen hat bestimmte Vorteile hinsichtlich der Art und Weise, wie die Entscheidungsstruktur und der Inhalt der zugehörigen Entscheidungsregeln angezeigt werden. In der Visualisierung der Einflussfaktoranalyse ist die Gesamtprognosestärke der Entscheidungsstruktur ebenfalls verfügbar.

Algorithmen

Das Entscheidungsstrukturmodell wird im Anschluss an die Datenaufbereitung und die Erstellung aller unidirektionalen Einflussfaktoren berechnet. Der erste Prädiktor der Entscheidungsstruktur wird als höchster unidirektionaler Einflussfaktor ausgewählt. Die Kategorien des Prädiktors werden zusammengeführt, wenn die negative Auswirkung auf die Prognosestärke unterhalb eines bestimmten Schwellenwerts liegt. Der nächste Schritt besteht in der Ermittlung des besten Prädiktors zur Aufteilung aller Strukturknoten, die jeweils aus den zusammengeführten Kategorien bestehen. Der Prozess wird so lange fortgesetzt, bis eine Stoppregel auf einen Strukturknoten Anwendung findet. Folgende Optionen sind für einen Stopp möglich: Sämtliche Kategorien für jeden als Kandidaten infrage kommenden Prädiktor werden zu einem einzelnen Knoten zusammengeführt, oder die Anzahl der Knoten überschreitet den hierfür maximal zulässigen Wert. Kategorien, die weniger Zeilen als die geltende Mindestanzahl enthalten, werden immer mit einer anderen Kategorie zusammengeführt. Dies bedeutet, dass keiner der Knoten in der Verzeichnis-

struktur weniger als die Mindestanzahl von Zeilen enthalten kann. Dieselbe Prozedur wird für fortlaufende und kategoriale Ziele verwendet; lediglich die Funktion zur Ermittlung der Unreinheit ist unterschiedlich.

Details

Funktionen zur Ermittlung der Unreinheit

Die Werte der Funktionen zur Ermittlung der Unreinheit werden als Hauptkriterium für das Aufteilen und Zusammenführen potenzieller Strukturknoten herangezogen. Der Gesamtwert der Unreinheitsfunktion für fortlaufende Verzeichnisstrukturen ist die Summe der Quadrate pro Knoten, während für die kategorialen Ziele das Gini-Unreinheitsmaß verwendet wird. Der Gesamtwert der Gini-Unreinheit ist die Summe der Quadrate der Zähleranteile in allen Zielkategorien pro Knoten, die von eins subtrahiert wird, wobei die Ergebnisse mit der Anzahl der Zeilen multipliziert werden. Eine Verbesserung des Unreinheitsfunktionswerts ist ein Informationsgewinn.

Beim Aufteilen der einzelnen Knoten sucht IBM Cognos Analytics nach einem Prädiktorfeld mit einem größtmöglichem Informationsgewinn. Dieser wird berechnet, indem die Gesamtunreinheit über alle potenziellen untergeordneten Knoten hinweg ermittelt und von der Unreinheit des übergeordneten Knotens subtrahiert wird. Bevor Cognos Analytics den Prädiktor auswählt, versucht Cognos Analytics, einige der potenziellen untergeordneten Knoten zusammenzuführen, die ursprünglich jeweils den einzelnen Prädiktorkategorien entsprechen. Dabei wird der Informationsverlust berechnet, indem die Unreinheit der nicht zusammengeführten Knoten von der Unreinheit der zusammengeführten Knoten subtrahiert wird. Solange der Informationsverlust unterhalb eines bestimmten Schwellenwerts liegt, werden die Knoten zusammengeführt. Dieser Prozess hilft bei der Erstellung relativ kleiner Verzeichnisstrukturen, die einfach zu visualisieren und zu begreifen sind, während die Gesamtstärke der Struktur erhalten bleibt.

Stoppregeln

Kandidatenknoten werden immer zusammengeführt, wenn sie auf weniger als 25 Zeilen basieren. Wenn alle Kategorien eines Prädiktors zusammengeführt werden, kann dieser nicht zum Aufteilen eines bestimmten Knotens verwendet werden. Wenn der betreffende Knoten von keinem der Prädiktoren aufgeteilt werden kann, wird der Prozess für diesen Knoten gestoppt. Wenn keiner der Knoten aufgeteilt werden kann oder die Anzahl der generierten Knoten größer als 36 ist, dann wird der Gesamtprozess zum Generieren der Verzeichnisstruktur gestoppt.

Wichtigkeit von Variablen

Die Wichtigkeit von Variablen entspricht der relativen Reduzierung des Verzeichnisstrukturfehlers, die sich ergibt, wenn der entsprechende Prädiktor in die Struktur eingeschlossen wird. Sie wird berechnet, indem der Fehler der ursprünglichen Verzeichnisstruktur mit dem Fehler einer eingeschränkten Verzeichnisstruktur verglichen wird, die anhand der verbleibenden Prädiktoren in der ursprünglichen Struktur generiert wird. Der Fehler der ursprünglichen Verzeichnisstruktur wird vom Fehler der eingeschränkten Verzeichnisstruktur subtrahiert und das Ergebnis wird durch den Fehler der eingeschränkten Struktur dividiert. Variablen mit einer Wichtigkeit von null oder einer negativen Wichtigkeit werden aus der Verzeichnisstruktur entfernt. Der Verzeichnisstrukturfehler wird als Summe der Quadrate für fortlaufende Ziele und als Klassifikationsfehler für kategoriale Ziele berechnet.

Prognosestärke

Die Berechnung der Prognosestärke für Verzeichnisstrukturen mit fortlaufenden Zielen erfolgt ähnlich wie bei zentralen Einflussfaktoren. Es wird der Inhalt von Endknoten berücksichtigt. Der Varianzbeitrag für jeden einzelnen Endknoten wird addiert und durch die Gesamtvarianz für die Daten dividiert. Dies ist der relative Fehler für die Verzeichnisstruktur. Er wird von Eins subtrahiert, um die Prognosestärke zu ermitteln, die mit der R-Quadrat-Kennzahl kompatibel ist, die für zentrale Einflussfaktoren verwendet wird.

Bei kategorialen Zielen berechnet Cognos Analytics die Klassifikationsgenauigkeit auf Grundlage des Klassifikationsfehlers, der aus allen Endknoten addiert wird. Die relative Verbesserung der Klassifikationsgenauigkeit im Vergleich zum Basismodell, die auch als korrigiertes R im Quadrat bezeichnet wird, wird als Prognosestärke der Verzeichnisstruktur gemeldet. Der Wert wird berechnet, indem der Verzeichnisstrukturfehler vom Basismodellfehler subtrahiert und das Ergebnis durch den Basis-

modellfehler dividiert wird. Beispiel: Die Klassifikationsgenauigkeit des Modells kann bis zu 95 % betragen; wird jedoch für 90 % der Zeilen in den Daten die Mehrheitsklasse angezeigt, dann wird die Prognosestärke der Verzeichnisstruktur nur als 50 % gemeldet. Dies ist eine Parallele zum Fall mit dem fortlaufenden Ziel, in dem das Basismodell durch den Gesamtmittelwert dargestellt wird. Die Prognosestärke, die durch R im Quadrat gemessen wird, basiert auf der relativen Verbesserung der Verzeichnisstruktur hinsichtlich der Reduzierung der Gesamtvarianz.

Cognos Analytics zeigt nur diejenigen Verzeichnisstrukturen an, die eine Prognosestärke von mehr als 10 % aufweisen. Verzeichnisstrukturen für fortlaufende Ziele werden in Visualisierungen vom Typ 'Einflussfaktoranalyse' oder 'Spiraldiagramm' angezeigt, wenn ihre Prognosestärke größer ist als die Prognosestärke des stärksten zentralen Einflussfaktors. Andernfalls werden sie in diesen Diagrammen nicht angezeigt, da die zentralen Einflussfaktoren bereits alle relevanten Insights bereitstellen.

Die Prognosestärke für eine Entscheidungsstruktur wird anhand derselben Daten berechnet, die auch zum Generieren der Entscheidungsstruktur verwendet werden. Dies führt bekanntermaßen zu einer gewissen Verzerrung und zur Angabe optimistischer Schätzungen hinsichtlich der Leistung der Entscheidungsstruktur bei ähnlichen Daten aus derselben Datenquelle. In Cognos Analytics ist die Diskrepanz reduziert worden, indem der Algorithmus so optimiert wurde, dass eine Überanpassung der Trainingsdaten minimiert wird.

Insights in Visualisierungen

Insights in Visualisierungen liefern analytische Insights, die Benutzern dabei helfen können, wichtige Beziehungen und aussagekräftige Unterschiede auf Grundlage der Daten zu erkennen und zu validieren, die in der Visualisierung dargestellt werden.

Übersicht

Insights werden vom Feld **Insights** gesteuert und dort ausgewertet. Dieses Feld ist in jeder infrage kommenden Visualisierung verfügbar. Wenn Sie Insights aktivieren, wird die Auswertung im Feld **Insights** angezeigt, und verwandte Visualisierungselemente werden hervorgehoben. Details werden in der entsprechenden QuickInfo-Nachricht bereitgestellt. Sie können jede verfügbare Insight separat steuern.

Algorithmen

Der Typ der Insights hängt von den durch die Visualisierung angezeigten Daten ab. Folgende Insights-Typen sind verfügbar: **Durchschnitt**, **Prognosestärke**, **Aussagekräftige Unterschiede**, **Anpassungslinie** und **Größte Häufigkeit**. 'Durchschnitt' stellt den Mittelwert der angezeigten Auswertungen bereit und 'Größte Häufigkeit' die Kategorie, die am häufigsten in den Daten vorkommt. Die übrigen Insights hängen von erweiterten Analysen und statistischen Tests ab. Ziel ist es, zuverlässige Informationen bereitzustellen, die Sie für eine erweiterte Beschreibung der angezeigten Daten verwenden können sowie zur Erkennung von Beziehungen, die in der von diesen Daten dargestellten Grundgesamtheit an Datenwerten zu finden sein sollten.

Details

Die Analyse von Insights basiert stets auf denselben Datenzeilen, die auch zur Erstellung der in der Visualisierung angezeigten Auswertungen verwendet werden. Dies bedeutet, dass für die Insights die vollständigen Daten verwendet werden, sofern kein Filter auf die ursprünglichen Daten angewendet wird.

Für einige statistische Tests und Analysen, die in Insights verwendet werden, sind nicht nur die in der Visualisierung angezeigten Datenauswertungen erforderlich, sondern darüber hinaus noch einige weitere Auswertungen. Beispiel: Für den Test auf aussagekräftige Unterschiede über mehrere Kategorien eines Erläuterungsfelds hinweg sind zusätzlich zu den angezeigten Daten auch Zähler und Varianzen für jede Kategorie erforderlich. Diese zusätzlichen Auswertungen werden zusammen mit den für die Visualisierung erforderlichen Auswertungen aus einer Datenbank abgerufen. Sämtliche Auswertungen werden von den Insights verarbeitet, wobei jedoch nur die erforderlichen Auswertungen in der Visualisierung verfügbar sind. Die Analyse von Insights basiert stets auf denselben Datenzeilen, die auch zur Erstellung der in der Visualisierung angezeigten Auswertungen verwendet werden.

Einschränkungen

Wenn Insights in einer Visualisierung nicht unmittelbar verfügbar sind, kann unter Umständen einer der folgenden Gründe vorliegen:

- Der Visualisierungstyp selbst unterstützt keine Insights.
- Die Daten in der Visualisierung wurden möglicherweise abgeschnitten.
- Die Kombination aus Auswertungsstufe, Feldtyp und Feldrolle eines ausgewählten Felds entspricht nicht den Anforderungen der verfügbaren Insights.

Unterstützte Visualisierungstypen für Insights

Die folgenden Visualisierungstypen unterstützen Insights:

- Flächendiagramm
- Balkendiagramm
- Blasendiagramm
- Säulendiagramm
- Heat-Map
- Blasendiagramm mit Hierarchien
- Liniendiagramm
- Linien- und Säulendiagramm
- Karte
- Kompaktes Blasendiagramm
- Kreisdiagramm
- Punktdiagramm
- Radialdiagramm
- Streudiagramm
- Gestapeltes Balkendiagramm
- Gestapeltes Säulendiagramm
- Treemap
- Wortwolke

Small Multiples-Erweiterungen werden für einige Insights unterstützt, z. B. 'Größte Häufigkeit' und 'Aussagekräftige Unterschiede'.

Auswertungsstufen

Folgende Auswertungsstufen werden unterstützt: **Anzahl**, **Durchschnitt**, **Summe**, **Minimum** und **Maximum**. Alle anderen Werte wie beispielsweise **Anzahl eindeutiger Elemente** können unter Umständen dazu führen, dass keine Insights vorgeschlagen werden. Einige Algorithmen unterstützen nur bestimmte Auswertungsstufen. Wenn Sie die Standardauswertungsstufe in einen der unterstützten Werte ändern, lassen sich die Insights dadurch möglicherweise aktivieren.

Feldtypen

Feldtypen können intern als 'fortlaufend' oder 'kategorial' festgelegt sein. Dies hängt von den Werten des ausgewählten Felds ab.

Feldtyp	Beschreibung
Kategorial	Eine Variable kann einen Wert aus einer begrenzten und in der Regel festgelegten Anzahl von möglichen Werten annehmen. Eine kategoriale Variable ordnet jede Person oder andere Beobachtungseinheit einer bestimmten Gruppe oder nominalen Kategorie zu. Diese Zuordnungen ba-

Feldtyp	Beschreibung
	sieren auf gewissen qualitativen Eigenschaften. Ein Beispiel hierfür ist das Land, in dem eine Person lebt.
Fortlaufend	Eine Variable, die zum Beschreiben numerischer Werte verwendet wird, wie beispielsweise eines Bereichs von 0 - 100 oder 0,75 - 1,25. Ein fortlaufender Wert kann eine ganze Zahl, eine reelle Zahl oder ein Datum und eine Uhrzeit sein.

Feldrollen

IBMCognos Analytics ordnet jedem Feld in einer unterstützten Visualisierung eine Rolle zu. Je nach dem Bereich der Visualisierung kann eine Feldrolle als eine der folgenden Rollen festgelegt werden.

Feldrolle	Beschreibung
Antwort	Eine Variable, die vorhergesagt wird. Sie kann auch als Ziel oder abhängige Variable bezeichnet werden. Sie befindet sich im Allgemeinen auf der Y-Achse.
Erläuterung	Eine Variable, die dabei hilft, Änderungen in der Antwort zu erklären und auch als Prädiktor oder unabhängige Variable bezeichnet wird. Sie befindet sich im Allgemeinen auf der X-Achse.
Gruppe	Eine Variable, die als Erläuterung oder als optionaler gruppierter Faktor behandelt wird, der dabei hilft, die Anzahl der im Algorithmus erstellten Modelle zu bestimmen. Dieser Wert kann zum Beispiel dem Farbfeld einer Säulendiagrammvisualisierung entsprechen.
Gewichtung	Eine Variable, die die optionalen Regressionsgewichtungen definiert, die zur Berechnung des Regressionsmodells verwendet werden. Dieser Wert kann zum Beispiel dem Größelfeld einer Blasendiagrammvisualisierung entsprechen.
Wiederholen	Eine Variable, die Small Multiples erstellt, wobei die Visualisierung einmal für jeden einzelnen Wert der Variablen wiederholt wird. Dieser Wert kann zum Beispiel dem Feld 'Wiederholen (Zeile)' einer Kreisdiagrammvisualisierung entsprechen.
Punkte	Eine Variable, die die Gestaltung der Daten und Datenpunkte, die zur Berechnung des Regressionsmodells verwendet werden, definiert. Dieser Wert kann zum Beispiel dem Punktefeld einer Streudiagrammvisualisierung entsprechen.

Hierzu ein allgemeines Beispiel: In einer Visualisierung vom Typ 'Balkendiagramm' mit den folgenden Feldern können die Rollenzuordnungen in der Visualisierung wie folgt definiert sein:

- Balken (Y-Achse), Erläuterung
- Länge (X-Achse), Antwort
- Farbe, Gruppe

Insights in Visualisierungen für Zähler

Insights für Zähler sind immer dann verfügbar, wenn für jede Kategorie eines einzelnen kategorialen Felds eine Anzahl angezeigt wird.

Sie sind auch verfügbar, wenn für jede Kombination von Kategorien eines Paares kategorialer Felder in der Visualisierung eine Anzahl angezeigt wird. In letzterem Fall kann es sich bei dem Paar um zwei Erläuterungsfelder handeln, wie in den Zeilen und Spalten einer Heat-Map, oder um ein Erläuterungs- und ein Wiederholungsfeld, wie bei den Balken und Wiederholungen (Spalte) eines Balkendiagramms.

Insights für Zähler kombinierter Kategorien aus drei kategorialen Feldern werden unterstützt für ein Erläuterungsfeld und die Kombination von zwei Wiederholungsfeldern, wie bei den Segmenten, Wiederholungen (Spalte) und Wiederholungen (Zeile) einer Kreisdiagrammvisualisierung.

Übersicht

Verwenden Sie diesen Visualisierungstyp, wenn Sie die Anzahl der Elemente in verschiedenen Kategorien oder Kombinationen von Kategorien vergleichen wollen.

Algorithmen

IBMCognos Analytics gibt die durchschnittliche Anzahl für alle Kategorien des angegebenen Antwortfelds zurück und wendet statistische Tests an, um Kategorien zu ermitteln, in denen die Zähler statistisch gesehen am meisten vom Durchschnitt abweichen.

Visualisierungen mit zwei oder drei kategorialen Feldern und Zählern für jede Kombination von Kategorien werden anders behandelt. Cognos Analytics vergleicht nicht nur die Zähler der Kategorien, sondern ermittelt auch alle Beziehungen zwischen den kategorialen Feldern. Cognos Analytics behandelt ein Feld als Antwortfeld und die anderen als Erläuterungsfelder.

Cognos Analytics meldet die häufigste Kategorie in Visualisierungen mit einem kategorialen Erläuterungsfeld, einem oder zwei kategorialen Wiederholungsfeldern und einem Zählerantwortfeld.

Details

Einzelnes kategoriales Feld

Der erste Test, der angewendet wird, ist der Chi-Quadrat-Test gleicher Häufigkeiten, um festzustellen, ob Zähler verfügbar sind, die signifikant vom Durchschnitt abweichen. Ist das Testergebnis signifikant, wendet Cognos Analytics den Chi-Quadrat-Beeinflusstest für jede Kategorie separat an. Cognos Analytics berechnet die Effektgröße für Kategorien, bei denen das Ergebnis des Beeinflusstests statistisch signifikant ist, und gibt die Kategorien mit der höchsten Effektgröße unter 'Aussagekräftige Unterschiede' an.

Einschränkungen

In der folgenden Tabelle werden die Bedingungen beschrieben, von denen abhängt, ob Insights (d. h. aussagekräftigere Informationen) für diesen Algorithmus vorgeschlagen werden.

Antwort	Erläuterung	Gruppe	Gewichtung	Punkte	Insight
Genau 1 Auswertungsstufe = Anzahl	Nicht zutreffend	Nicht zutreffend	Nicht zutreffend	Nicht zutreffend	Durchschnitt Aussagekräftige Unterschiede

Zwei kategoriale Felder

Cognos Analytics behandelt ein kategoriales Feld als Antwortfeld und das andere als Erläuterungsfeld. Das ursprüngliche Zählerfeld wird als Eingabe für die Algorithmen verwendet.

Der Chi-Quadrat-Test der Unabhängigkeit mit der Anpassung für spärliche Daten wird verwendet, um festzustellen, ob eine Beziehung zwischen dem Antwortfeld und dem Erläuterungsfeld besteht. Ist das Testergebnis signifikant, berechnet Cognos Analytics die Prognosestärke für dieses Modell als korrigiertes R im Quadrat mit gefilterten Kategorien mit niedrigen Häufigkeiten. Liegt die Prognosestärke über 10 %, wird die Prognosestärke gemeldet und die Beziehung als zuverlässig deklariert.

Wenn das obige Testergebnis signifikant ist, werden alle Kombinationen von Erläuterungs- und Reaktionskategorien weitergehend analysiert, indem der Chi-Quadrat-Beeinflusstest für jede Kombination angewendet wird. Kombinationen von Erläuterungs- und Antwortkategorien, bei denen der Beeinflusstest signifikant ist, gelten als einflussreich. Für jede einflussreiche Kombination von

Kategorien wird die Effektgröße berechnet und die Kombinationen mit der höchsten Effektgröße werden unter 'Aussagekräftige Unterschiede' aufgeführt.

Wenn die beiden kategorialen Felder die Rollen 'Erläuterung' und 'Wiederholung' haben, wird der Algorithmus für die größte Häufigkeit angewendet. Die Zähler werden für jede einzelne Kategorie des Erläuterungsfelds summiert. Die größte Summe wird zusammen mit der Anzahl der Kategorien, die diese Summe haben, gemeldet. Beachten Sie, dass das Wiederholungsfeld nicht von diesem Algorithmus verwendet, sondern nur ausgelöst wird, wenn der Algorithmus angewendet wird.

Einschränkungen

In der folgenden Tabelle werden die Bedingungen beschrieben, von denen abhängt, ob Insights (d. h. aussagekräftigere Informationen) für diesen Algorithmus vorgeschlagen werden.

Antwort	Erläuterung	Gruppe	Gewichtung	Punkte	Wiederholen	Insight
Genau 1 Auswertungsstufe = Anzahl	Genau 2 Kategorial	Nicht zutreffend	Nicht zutreffend	Nicht zutreffend	Nicht zutreffend	Prognosestärke Aussagekräftige Unterschiede
Genau 1 Auswertungsstufe = Anzahl	Genau 1 Kategorial	Nicht zutreffend	Nicht zutreffend	Nicht zutreffend	Genau 1 Kategorial	Prognosestärke Aussagekräftige Unterschiede Größte Häufigkeit

Drei kategoriale Felder

Diese Algorithmen werden nur angewendet, wenn es ein Erläuterungsfeld und zwei Wiederholungsfelder gibt. Die Kombination der beiden Wiederholungsfelder wird so behandelt, als handelte es sich um ein einziges kategoriales Feld, wobei die Kategorien die Paare der Kategorien aus den beiden Wiederholungsfeldern sind.

Die Prognosestärke wird genau wie im Fall von zwei kategorialen Feldern berechnet, wobei die paarweise verbundenen Wiederholungsfelder als Prädiktor des Erläuterungsfelds verwendet werden. Der Chi-Quadrat-Test der Unabhängigkeit mit der Anpassung für spärliche Daten wird zum Testen der Signifikanz der Beziehung verwendet und das korrigierte R im Quadrat mit gefilterten Kategorien mit niedrigen Häufigkeiten wird zum Feststellen der Prognosestärke verwendet.

Die aussagekräftigen Unterschiede werden genau wie im Fall von zwei kategorialen Feldern berechnet, wobei Kombinationen aus dem Erläuterungsfeld und den paarweise verbundenen Wiederholungsfeldern identifiziert werden, für die der Zähler ungewöhnlich ist. Der Chi-Quadrat-Beeinflussungstest wird zum Testen der Signifikanz für jede Kombination verwendet und die Kombinationen mit der größten Effektgröße werden gemeldet.

Der Algorithmus für die größte Häufigkeit wird genau wie im Fall von zwei kategorialen Feldern angewendet, wobei die Zähler für jede einzelne Kategorie des Erläuterungsfelds summiert werden. Die größte Summe wird zusammen mit der Anzahl der Kategorien, die diese Summe haben, gemeldet. Beachten Sie, dass die Wiederholungsfelder nicht von diesem Algorithmus verwendet werden, sondern nur ausgelöst werden, wenn der Algorithmus angewendet wird.

Antwort	Erläuterung	Gruppe	Gewichtung	Punkte	Wiederholen	Insight
Genau 1 Auswertungsstufe = Anzahl	Genau 1 Kategorial	Nicht zutreffend	Nicht zutreffend	Nicht zutreffend	Genau 2 Kategorial	Prognosestärke Aussagekräftige Unterschiede Größte Häufigkeit

Unterschiede zwischen Cognos Analytics Version 11.1 R2 und R3

Bei Visualisierungen mit zwei kategorialen Feldern, wenn das Antwortfeld eine Kategorie enthielt, die fehlende Daten darstellte (die Kategorie "(kein Wert)"):

- In Cognos Analytics Version 11.1 R2 wurden Datenwerte für die Kategorie fehlender Daten bei der Berechnung des korrigierten R im Quadrat für die Prognosestärke übergangen. In Cognos Analytics Version 11.1 R3 werden die Werte einbezogen. Dies kann sich auf die Prognosestärke auswirken, die für Heat-Maps gemeldet wird, bei denen das Heat-Map-Datenfeld zähleraggregierte Daten enthält und es sich bei den Zeilen und Spalten um zwei kategoriale Felder handelt.
- In Cognos Analytics Version 11.1 R2 haben die aussagekräftigen Unterschiede keine ungewöhnlichen Zellen in Verbindung mit der Kategorie fehlender Daten gemeldet. In Cognos Analytics Version 11.1 R3 werden die Zellen gemeldet. Dies kann sich auf die aussagekräftigen Unterschiede auswirken, die für Heat-Maps angezeigt werden, bei denen das Heat-Map-Datenfeld zähleraggregierte Daten enthält und es sich bei den Zeilen und Spalten um zwei kategoriale Felder handelt.

Insights in Visualisierungen für Auswertungen anhand mindestens eines Erläuterungsfelds

Insights für Auswertungen sind verfügbar, wenn die Auswertungsstufe für ein kontinuierliches Antwortfeld 'Durchschnitt', 'Summe', 'Minimum' oder 'Maximum' ist. Insights werden in jeder Kategorie eines einzelnen kategorialen Erläuterungsfelds oder in jeder Kombination von Kategorien eines Paares kategorialer Erläuterungsfelder in der Visualisierung berechnet und angezeigt.

Übersicht

Verwenden Sie diesen Visualisierungstyp, wenn Sie Werte eines Antwortfelds in verschiedenen Kategorien oder Kombinationen von Kategorien vergleichen wollen.

Algorithmen

Wenn die Auswertungsstufe 'Durchschnitt' lautet, ermittelt IBM Cognos Analytics jegliche Beziehung zwischen dem Antwortfeld und Erläuterungsfeldern und berechnet die Prognosestärke des entsprechenden Modells. Wenn die Unterschiede bei den Durchschnittswerten in den Erläuterungskategorien statistisch signifikant sind, gibt Cognos Analytics die Erläuterungskategorien bzw. Kombinationen von Kategorien mit den größten Abweichungen unter 'Aussagekräftige Unterschiede' an.

Wenn die Auswertungsstufe für Antworten für ein numerisches Antwortfeld 'Summe' lautet, berechnet Cognos Analytics die durchschnittliche Summe für die Erläuterungskategorien oder Kombinationen von Kategorien. Wenn die Unterschiede bei den Summen in den Kategorien statistisch signifikant sind, gibt Cognos Analytics die Erläuterungskategorien bzw. Kombinationen von Kategorien mit den größten Abweichungen unter 'Aussagekräftige Unterschiede' an.

Bei alle Diagrammen, die dies betrifft, zeigt die Insight 'Durchschnitt' den mittleren ausgewerteten Antwortwert für alle Erläuterungskategorien an. Wenn die Auswertungsstufe für die Antwort 'Durchschnitt' lautet, wird unter Verwendung des angezeigten Wertes und der Anzahl für jede Erläuterungskategorie der gewichtete Mittelwert berechnet.

Details

Durchschnitt anhand eines einzelnen Erläuterungsfelds

Wenn die Auswertungsstufe für das Antwortfeld 'Durchschnitt' lautet und ein einzelnes kategoriales Erläuterungsfeld verfügbar ist, wendet Cognos Analytics die unidirektionale Varianzanalyse (VA) an. Anhand der F-Statistik überprüft Cognos Analytics, ob die Durchschnittswerte über alle Erläuterungskategorien hinweg gleich sind. Gibt es signifikante Unterschiede, berechnet Cognos Analytics das korrigierte R im Quadrat als Prognosestärke der Beziehung zwischen dem Antwortfeld und dem Erläuterungsfeld. Liegt die Prognosestärke über 10 %, wird dem Benutzer eine zuverlässige Beziehung mit der zugehörigen Prognosestärke gemeldet.

Ist der Unterschied unter den Durchschnittswerten signifikant, führt Cognos Analytics einen Beeinflussungs-t-Test durch, um die Kategorien zu ermitteln, die am stärksten vom Gesamtmittelwert abweichen. Dies umfasst die Berechnung des Standardfehlers für jeden Kategoriedurchschnitt sowie den Vergleich des Durchschnitts mit dem Gesamtmittelwert anhand der t-Teststatistik. Bei Kategorien mit signifikanten Unterschieden berechnet Cognos Analytics auch die entsprechende Effektgröße und meldet die Kategorien mit der höchsten Effektgröße unter 'Aussagekräftige Unterschiede'.

Einschränkungen

In der folgenden Tabelle werden die Bedingungen beschrieben, von denen abhängt, ob Insights (d. h. aussagekräftigere Informationen) für diesen Algorithmus vorgeschlagen werden.

Antwort	Erläuterung	Gruppe	Gewichtung	Punkte	Insight
Genau 1 Auswertungsstufe = Durchschnitt Fortlaufend	Genau 1 Kategorial	Nicht zutreffend	Optional Beliebig	Nicht zutreffend	Durchschnitt Prognosestärke Aussagekräftige Unterschiede

Durchschnitt anhand von zwei Erläuterungsfeldern

Im Fall von Diagrammen, bei denen die Auswertungsstufe für das Antwortfeld 'Durchschnitt' lautet und zwei kategoriale Erläuterungsfelder verfügbar sind, wendet Cognos Analytics die bidirektionale Varianzanalyse (VA) an. Anhand der F-Statistik überprüft Cognos Analytics, ob die Durchschnittswerte über alle Kombinationen von Erläuterungskategorien hinweg gleich sind. Gibt es signifikante Unterschiede, berechnet Cognos Analytics das korrigierte R im Quadrat als Prognosestärke der Beziehung zwischen dem Antwortfeld und den beiden Erläuterungsfeldern. Cognos Analytics berechnet das korrigierte R im Quadrat auch für unidirektionale Modelle, die jeweils ein einzelnes Erläuterungsfeld enthalten. Wenn die Prognosestärke eines bidirektionalen Modells größer als 10 % ist und die zugehörige relative Verbesserung der Prognosestärke im Vergleich zu entsprechenden unidirektionalen Modellen mehr als 10 % beträgt, zeigt Cognos Analytics die Prognosestärke des bidirektionalen Modells an und meldet eine zuverlässige Beziehung zwischen dem Antwortfeld und den beiden Erläuterungsfeldern. Andernfalls gilt Folgendes: Wenn die maximale Prognosestärke von unidirektionalen Modellen mehr als 10 % beträgt, meldet Cognos Analytics eine zuverlässige Beziehung zwischen dem Antwortfeld und dem entsprechenden einzelnen Erläuterungsfeld zusammen mit der zugehörigen Prognosestärke. Liegt die maximale Prognosestärke von unidirektionalen Modellen nicht über 10 %, meldet Cognos Analytics keine Beziehung zwischen der Antwort und den Erläuterungsfeldern.

Ist der Unterschied unter den Durchschnittswerten über alle Kategoriekombinationen hinweg signifikant, führt Cognos Analytics auch einen Beeinflussungs-t-Test durch, um die Kategoriekombinationen zu ermitteln, die am stärksten vom Gesamtmittelwert abweichen. Dieser Test ähnelt dem Test, der für ein einzelnes Erläuterungsfeld verwendet wird. Der Hauptunterschied besteht darin, dass Cognos Analytics nicht die Kategorien eines einzelnen Erläuterungsfelds berücksichtigt, sondern Kombinationen von Kategorien aus den beiden Erläuterungsfeldern. Die Kategoriekombinationen mit der höchsten Effektgröße werden unter 'Aussagekräftige Unterschiede' aufgelistet.

Einschränkungen

In der folgenden Tabelle werden die Bedingungen beschrieben, von denen abhängt, ob Insights (d. h. aussagekräftigere Informationen) für diesen Algorithmus vorgeschlagen werden.

Antwort	Erläuterung	Gruppe	Gewichtung	Punkte	Insight
Genau 1 Auswertungsstufe = Durchschnitt Fortlaufend	Genau 2 Kategorial	Optional (als Erläuterung behandeln) Kategorial	Optional	Nicht zutreffend	Durchschnitt Prognosestärke Aussagekräftige Unterschiede

Summe anhand von einem oder zwei Erläuterungsfeldern

Im Fall von Diagrammen, bei denen die Auswertungsstufe für das Antwortfeld 'Summe' lautet und ein oder zwei kategoriale Erläuterungsfelder verfügbar sind, wendet Cognos Analytics den Summenvergleichstest an. Mithilfe dieses Tests wird ermittelt, ob es Summen gibt, die vom durchschnittlichen Summenwert über alle Erläuterungskategorien oder Kombinationen von Kategorien hinweg abweichen. Ist das Ergebnis dieses Tests signifikant, führt Cognos Analytics anschließend den Summenbeeinflussungstest durch, mit dem die Summe für jede einzelne Kategorie oder Kombination von Kategorien mit der Durchschnittssumme verglichen wird. Für jedes signifikante Testergebnis berechnet Cognos Analytics auch die entsprechende Effektgröße. Die Kategorien oder Kombinationen von Kategorien mit der höchsten Effektgröße werden unter 'Aussagekräftige Unterschiede' aufgelistet.

Einschränkungen

In der folgenden Tabelle werden die Bedingungen beschrieben, von denen abhängt, ob Insights (d. h. aussagekräftigere Informationen) für diesen Algorithmus vorgeschlagen werden.

Antwort	Erläuterung	Gruppe	Gewichtung	Punkte	Insight
Genau 1 Auswertungsstufe = Summe Fortlaufend	1 oder 2 Kategorial	Optional (als Erläuterung behandeln) Kategorial	Nicht zutreffend	Nicht zutreffend	Durchschnitt Aussagekräftige Unterschiede

Minimum oder Maximum anhand von einem oder zwei Erläuterungsfeldern

Für die Auswertungsstufen 'Minimum' und 'Maximum' ist nur die Insight 'Durchschnitt' verfügbar. Sie wird als der Durchschnittswert des Antwortminimums oder -maximums über alle Erläuterungskategorien oder Kombinationen von Kategorien hinweg berechnet.

Einschränkungen

In der folgenden Tabelle werden die Bedingungen beschrieben, von denen abhängt, ob Insights (d. h. aussagekräftigere Informationen) für diesen Algorithmus vorgeschlagen werden.

Antwort	Erläuterung	Gruppe	Gewichtung	Punkte	Insight
Genau 1 Auswertungsstufe = Minimum oder Maximum Fortlaufend	1 oder 2 Kategorial	Nicht zutreffend	Nicht zutreffend	Nicht zutreffend	Durchschnitt

Insights in Visualisierungen für zwei fortlaufende (stetige) Felder

Insights für zwei fortlaufende (stetige) Felder sind verfügbar, wenn eine Visualisierung zwei fortlaufende Felder und ein optionales kategoriales Gruppenfeld oder Punktfeld einbezieht.

Übersicht

Verwenden Sie Visualisierungen wie etwa ein Streudiagramm für die zwei fortlaufenden Felder, möglicherweise mit Ausschnitten entsprechend den Kategorien des Gruppenfeldes. Das Hauptziel besteht darin, jedwede Beziehung zwischen dem fortlaufenden Feld zu erkennen und das kategoriale Gruppenfeld ebenfalls einzuschließen. Die Ergebnisse enthalten die Prognosestärke der erkannten Beziehungen, eine Beschreibung der Beziehungen anhand von Anpassungslinien und sämtliche Punkte mit großen Abweichungen von den Anpassungslinien als 'Aussagekräftige Unterschiede'.

Algorithmen

IBMCognos Analytics berechnet mehrere Regressionsmodelle, die eines der fortlaufenden Felder als Antwort und die anderen fortlaufenden Felder als Erläuterungsfeld einbeziehen. Das optionale kategoriale Gruppenfeld wird als Modellfaktor verwendet. Zusätzlich zu additiven Modellbeiträgen, die dem Erläuterungsfeld entsprechen, berücksichtigt Cognos Analytics auch das Quadrat des Erläuterungsfelds sowie alle Wechselwirkungsterme, die einen Faktor enthalten. Aus einer Reihe möglicher Modelle wird ein Regressionsmodell ausgewählt, das eine optimale Anpassung für die Daten bietet. Die entsprechende Anpassungslinie wird vom linearen oder quadratischen Modell abgeleitet. In dem Fall, in dem ein optionales kategoriales Gruppenfeld angegeben wurde, kann es für jede Kategorie des Faktors eine andere Linie oder quadratische Kurve erzeugen. Derzeit wird ein Faktor mit bis zu drei Kategorien berücksichtigt, um die Visualisierung nicht zu überfrachten.

Jeder Punkt in einer Visualisierung stellt eine Anzahl von Zeilen in den Daten dar und wird durch das Feld **Punkte** definiert. Eine entsprechende Zeilenanzahl, die auf dem Antwortfeld basiert, definiert die Häufigkeitsgewichtung, die für die Erstellung der Regressionsmodelle verwendet wird. Regressionsgewichtungen werden von Cognos Analytics bei der Berechnung der Regressionsmodelle unabhängig von der Häufigkeitsgewichtung verwendet.

Details

Zwei fortlaufende (stetige) Felder

Wenn Cognos Analytics eine mehrfache lineare Regression für zwei fortlaufende (stetige) Felder anwendet, wird ein Feld als Antwort und das andere als Erläuterungsfeld im Modell ausgewählt. Cognos Analytics berücksichtigt sowohl lineare als auch quadratische Modellbegriffe. Wenn das quadratische Modell gemäß F-Test signifikant ist und die entsprechende relative Verbesserung der Prognosestärke im Vergleich zum linearen Modell mehr als 10 % beträgt, meldet Cognos Analytics die Prognosestärke des quadratischen Modells und zeigt die quadratische Kurve auf Grundlage des berechneten Modells an. Diese Kurve zeigt vorhergesagte Werte der Antwort basierend auf den entsprechenden Werten des Erläuterungsfelds an. Andernfalls wird das lineare Prädiktormodell berücksichtigt. Wenn es signifikant ist und seine Prognosestärke mehr als 10 % beträgt, meldet Cognos Analytics seine Prognosestärke und zeigt eine Linie an, die die vorhergesagten Werte des Antwortfelds auf Grundlage der entsprechenden Erläuterungswerte darstellt. Wenn das lineare Modell nicht infrage kommt, wird der Mittelwert als Anpassungslinie angegeben, und es wird keine Beziehung zwischen den beiden fortlaufenden (stetigen) Feldern gemeldet.

Wenn eine lineare oder quadratische Beziehung erkannt wird, überprüft Cognos Analytics auch die Unterschiede zwischen den vorhergesagten und den beobachteten Werten des Antwortfelds. Diese Unterschiede werden als 'Residuen' bezeichnet und Cognos Analytics führt den Test für studentisiertes Residuum durch, um Ausreißer zu erkennen. Punkte mit einer großen Abweichung von den erkannten Beziehungen werden im entsprechenden Diagramm unter 'Aussagekräftige Unterschiede' angezeigt.

Einschränkungen

In der folgenden Tabelle werden die Bedingungen beschrieben, von denen abhängt, ob Insights (d. h. aussagekräftigere Informationen) für diesen Algorithmus vorgeschlagen werden.

Antwort	Erläuterung	Gruppe	Gewichtung	Punkte	Insight
Genau 1	Genau 1	Nicht zu-treffend	Optional	Optional	Prognosestärke
Auswertungsstufe = Beliebig	Fortlaufend		Fortlaufend	Beliebig	Anpassungslinie
Fortlaufend					Aussagekräftige Unterschiede

Kategoriales Gruppenfeld

Wenn ein kategoriales Gruppenfeld zusätzlich zu zwei fortlaufenden (stetigen) Feldern angegeben wird, wird es als Faktor in der mehrfachen linearen Regression verwendet, wo eines der beiden

fortlaufenden (stetigen) Felder als Antwortfeld und das andere als Erläuterungsfeld ausgewählt wird. Cognos Analytics betrachtet lineare und quadratische Modellterme für fortlaufende Erläuterungen kombiniert mit Beiträgen aus dem Faktor. Wenn das quadratische Modell oder das lineare Modell, das den Faktor enthält, gemäß F-Test signifikant ist und die relative Verbesserung der Prognosestärke mehr als 10 % über der des linearen Modells liegt, das nur fortlaufende Erläuterungen enthält, generiert Cognos Analytics vier zusätzliche Modelle. Diese Modelle schließen alle möglichen Interaktionen fortlaufender Erläuterungen und Faktoren ein. Ein Modell mit ebenfalls signifikantem maximalem korrigiertem R im Quadrat wird als Endmodell ausgewählt. Dieses Modell wird verwendet, um eine Anpassungslinie für jede Kategorie des kategorialen Prädiktors zu erstellen. Andernfalls wird das lineare Modell mit fortlaufender Erläuterung auf Signifikanz getestet und gemeldet, wenn seine Prognosestärke größer als 10 % ist. Wenn das lineare Modell nicht infrage kommt, wird keine zuverlässige Beziehung zwischen den Feldern erstellt, und der Gesamtmittelwert wird als Anpassungslinie angegeben.

Wenn eine zuverlässige Beziehung erkannt wird, führt Cognos Analytics auch eine Überprüfung auf Unterschiede zwischen den vorhergesagten und den beobachteten Werten des Antwortfelds durch. Cognos Analytics führt den Test für studentisiertes Residuum durch, um Ausreißer zu erkennen und diese im entsprechenden Diagramm unter 'Aussagekräftige Unterschiede' anzuzeigen.

Einschränkungen

In der folgenden Tabelle werden die Bedingungen beschrieben, von denen abhängt, ob Insights (d. h. aussagekräftigere Informationen) für diesen Algorithmus vorgeschlagen werden.

Antwort	Erläuterung	Gruppe	Gewichtung	Punkte	Insight
Genau 1	Genau 1	Genau 1	Optional	Optional	Prognosestärke
Auswertungsstufe = Beliebig	Fortlaufend	Kategorial	Beliebig	Beliebig	Anpassungslinie
Fortlaufend					Aussagekräftige Unterschiede

Feld für Regressionsgewichtungen

Es kann ein optionales fortlaufendes (stetiges) Feld verwendet werden, um Regressionsgewichtungen für das Modell anzugeben. Die Regressionsgewichtung für einen verfügbaren Wert entspricht dem Einfluss der Beobachtung auf die berechneten Modellparameter.

Details zur natürlichen Sprache

Bei 'Details zur natürlichen Sprache' handelt es sich um eine Textfunktion, die angezeigte Visualisierungen um zusätzliche Auswertungen erweitert. Die Details liefern Insights, die aus einer entsprechenden, für Sie relevanten Datenanalyse abgerufen werden.

Übersicht

Diese Funktion steht für Visualisierungen zur Verfügung, die in einer Exploration erstellt werden, und die textbasierten Details werden in der zugehörigen Detailanzeige aufgeführt. Anhand dieser Informationen können Sie Informationen, die für die angezeigten Daten von größter Relevanz sind, im Format einer natürlichen Sprache abrufen. Darüber hinaus werden weitere Auswertungen und Details angezeigt, die in der Visualisierung nicht verfügbar sind.

Algorithmen

Grundlegende Details bieten einfache Auswertungen zu den Daten, die in der angezeigten Visualisierung nicht ohne Weiteres zu sehen sind. Wengleich diese Informationen von Ihnen durch Angabe anderer zugehöriger Visualisierungen abgerufen werden können, werden solche Explorationsschritte unnötig, da zugehörige Auswertungen als textbasierte Details verfügbar gemacht werden.

Details, die auf den Insights in Visualisierungen basieren, stellen Textbeschreibungen der Details bereit, die über das Dialogfeld für Insights in der angezeigten Visualisierung oder den zugehörigen Visualisierungen abgerufen werden können. Diese Beschreibungen verdeutlichen die angezeigten Insights und ermöglichen darüber hinaus das Hinzufügen weiterer Insights, die in der angezeigten Visualisierung nicht verfügbar sind.

Details

Die Details basieren auf den Feldern, die in der entsprechenden Visualisierung angezeigt werden. Die zugehörige Analyse kann auf zusätzliche Auswertungen zurückgreifen, enthält jedoch keine Felder, die nicht in der Visualisierung angegeben sind. Auswertungen und Details werden mithilfe von Vorlagen in übersetzbaren Text konvertiert, anstatt eine vollständige Funktion für die Generierung natürlicher Sprachen zu verwenden. Dies führt zu gewissen sprachlichen Redundanzen im ausgegebenen Text, wodurch die Menge und Qualität der angezeigten Informationen jedoch nicht beeinträchtigt werden.

Grundlegende Details zur natürlichen Sprache

Anhand von grundlegenden Details zur natürlichen Sprache werden zusätzliche Auswertungen zu den Daten bereitgestellt, die in einer Visualisierung angezeigt werden, oder zusätzliche Details zu den bereits verfügbaren Informationen hervorgehoben.

Übersicht

Diese Informationen runden die angezeigten Daten ab und sind überdies auch für die angezeigten Auswertungen relevant, die Sie berücksichtigen. Wird die für das Antwortfeld angegebene Auswertungsstufe geändert, dann ändern sich auch die grundlegenden Details in Teilen oder im Ganzen, da sich durch eine Änderung der Auswertungsstufe der Schwerpunkt der Visualisierung verschiebt. Weitere Informationen hierzu werden im Kontext der Haupttask bereitgestellt.

Algorithmen

Die berechneten grundlegenden Details hängen von der Auswertungsstufe ab, die für das Antwortfeld in der entsprechenden Visualisierung angegeben ist. Die folgenden Auswertungsstufen für Antworten sind möglich: 'Anzahl eindeutiger Elemente' und 'Anzahl' für alle Felder sowie 'Summe', 'Durchschnitt', 'Minimum' und 'Maximum' für numerische Felder. Während für die meisten Auswertungsstufen für Antworten zählerbasierte Details verwendet werden, werden bei den Auswertungsstufen 'Summe', 'Durchschnitt', 'Minimum' und 'Maximum' für Antworten zusätzliche passende Auswertungen bereitgestellt. Für die Auswertungsstufen 'Anzahl' und 'Anzahl eindeutiger Elemente' für Antworten werden dieselben zählerbasierten Auswertungen generiert.

Details

Gesamtzahl

Die Gesamtzahl wird für die Antwort und alle Erläuterungsfelder in der Visualisierung angezeigt. Diese Anzahl enthält keine fehlenden Werte der Antwort und wird immer berechnet, sofern die Auswertungsstufe für Antworten nicht 'Summe' oder 'Durchschnitt' lautet. Alle in der Visualisierung angegebenen kategorialen Felder werden zwecks Bereitstellung von Details als Erläuterungsfelder behandelt.

Anzahl für Erläuterungsfelder

Die Anzahl wird auch für jede Kategorie der angezeigten Erläuterungsfelder berechnet, sofern die Auswertungsstufe für Antworten nicht 'Summe' lautet. Der Algorithmus wählt die obersten Kategorien, die entsprechenden Zähler und deren Prozentsätze im Verhältnis zur Gesamtzahl für das Erläuterungsfeld aus und berichtet diese. Diese Prozedur wird auf jedes einzelne Erläuterungsfeld im Diagramm angewendet.

Summe

Wenn die angezeigte Auswertungsstufe für ein numerisches Antwortfeld 'Summe' lautet, wertet IBM Cognos Analytics die entsprechende Gesamtsumme für das Antwortfeld aus. Wenn die Visualisierung mehrere kategoriale Erläuterungsfelder enthält, wird die Summe jeweils für jede Kategorie und jedes Erläuterungsfeld berechnet. Der Algorithmus wählt die obersten Kategorien, die entsprechenden Summen und deren Prozentsätze im Verhältnis zur Gesamtsumme für jedes Erläuterungsfeld aus und berichtet diese.

Durchschnitt

Wenn die angezeigte Auswertungsstufe für ein numerisches Antwortfeld 'Durchschnitt' lautet, wertet Cognos Analytics den entsprechenden Gesamtdurchschnitt für das Antwortfeld aus.

Bereich

Bereichsauswertungen werden durch das Minimum und Maximum angegeben, wenn für ein numerisches Antwortfeld die Auswertungsstufe 'Summe' oder 'Durchschnitt' für die Kombinationen von Kategorien der angegebenen Erläuterungsfelder berechnet wird. Ist nur ein einzelnes Erläuterungsfeld angegeben, werden auch die Kategorien angezeigt, in denen das Minimum und Maximum vorkommen.

Minimum und Maximum

Wenn die angezeigte Auswertungsstufe für ein numerisches Antwortfeld 'Minimum' oder 'Maximum' lautet, zeigt Cognos Analytics den entsprechenden Minimal- bzw. Maximalwert in allen Kategorien eines kategorialen Erläuterungsfelds an. Werden mehrere kategoriale Erläuterungsfelder angegeben, wird das Minimum bzw. das Maximum für alle möglichen Kombinationen von Kategorien berechnet.

Auf Insights basierende grundlegende Details zur natürlichen Sprache

Details, die auf Insights basieren, stellen Textbeschreibungen zu den Insights bereit, die über das Dialogfeld 'Insights' in der Visualisierung abgerufen werden können.

Übersicht

Details, die auf Insights basieren, stellen auch Insights aus zugehörigen Visualisierungen bereit, die im Kontext des aktuellen Diagramms informativ und leicht verständlich sind. Dies ermöglicht die Bereitstellung von umfassenderen Details, die mit der aktuellen Visualisierung in Zusammenhang stehen.

Algorithmen

Für Details, die direkt auf Insights basieren, wird das gesamte Spektrum an Berechnungen und statistischen Tests verwendet, die von den Insights unterstützt werden. Details enthalten auch die abgerufenen Ergebnisse für zugehörige Visualisierungen und kompilieren diese zusammen in einer aussagekräftigen Nachricht. Darüber hinaus bieten diese Details auch einige zusätzliche Analysen auf Basis der Insights-Ausgabe, indem weitere Details generiert werden, die für eine Ausgabe in Textform geeignet sind.

Details

Einzelnes Erläuterungsfeld

Anhand eines Antwortfelds und eines kategorialen Erläuterungsfelds werden für Diagramminformationen Insights verwendet, um Beziehungen zwischen dem Antwortfeld und dem Erläuterungsfeld zu erkennen. In Abhängigkeit davon, ob die Auswertungsstufe für Antworten 'Durchschnitt' oder 'Anzahl' lautet, wird eine entsprechende Analyse angewendet und es wird die Prognosestärke angegeben, wenn eine relevante Beziehung erkannt wird.

Wenn das Erläuterungsfeld numerisch ist und Insights eine Anpassungslinie generiert, dann gibt IBM Cognos Analytics eine positive oder negative Neigung für die lineare Anpassung zurück und zeigt an, ob eine quadratische Beziehung erkannt worden ist. Wenn die Beziehung quadratisch ist, gibt Cognos Analytics auch den äußersten Punkt (Extremwert) zurück. Cognos Analytics berechnet den Minimal- bzw. Maximalwert der Antwort sowie den Erläuterungswert, in dem der Extremwert der Antwort auftritt.

Zwei Erläuterungsfelder

Wenn zwei Erläuterungsfelder verfügbar sind, ermittelt Cognos Analytics die Beziehung zwischen der Antwort und beiden Erläuterungsfeldern sowie die separaten Beziehungen zwischen der Antwort und den einzelnen Erläuterungsfeldern. Wenn die Prognosestärke für die Beziehung mit beiden Erläuterungsfeldern eine relative Verbesserung von mehr als 10 % im Vergleich zu jeder separaten Beziehung bietet, wird von Cognos Analytics berichtet, dass die Antwort von beiden Erläuterungsfeldern zusammen beeinflusst wird. Andernfalls wird die Antwort von einem einzelnen Erläuterungsfeld beeinflusst oder von beiden Erläuterungsfeldern jeweils getrennt, aber nicht zusammen.

Aussagekräftige Unterschiede

In den Details werden alle aussagekräftigen Unterschiede angegeben, die von den Insights für die verwendeten Daten erkannt werden, wenn die Auswertungsstufe für Antworten 'Anzahl', 'Durchschnitt' oder 'Summe' lautet. Diese aussagekräftigen Unterschiede werden für die Kategorien jedes einzelnen Erläuterungsfelds separat angegeben sowie für die Kombinationen von Kategorien für beide Erläuterungsfelder. Darüber hinaus wird angezeigt, ob Werte ungewöhnlich hoch oder niedrig sind.

Entscheidungsstrukturen

Details enthalten die Prognosestärke für eine Entscheidungsstruktur sowie eine Liste der Prädiktorfelder, die zum Aufteilen der Strukturknoten verwendet werden. Prädiktorfelder mit der höchsten Wichtigkeit von Variablen werden zusammen mit der entsprechenden Faktorverbesserung der Variablenwichtigkeit im Vergleich zu anderen Feldern in der Entscheidungsstruktur gemeldet, sofern der Faktor größer als 2 ist. Ähnliche Details können für Felder mit der niedrigsten Variablenwichtigkeit angezeigt werden.

Details zur natürlichen Sprache für Zeitreihen

Details für Zeitreihen stellen Text-Insights bereit, die auf der Analyse der Zeitreihendaten und der entsprechenden Prognosemodelle basieren.

Übersicht

IBMCognos Analytics berichtet Details für Zeitreihen für eine Visualisierung, die in einer Exploration immer dann erstellt wird, wenn die Visualisierungsdaten eine einzelne Zeitreihe enthalten und ein Prognosemodell berechnet wird. Wenn die Daten geeignet sind, werden Insights für Zeitreihen generiert, auch wenn das Dialogfeld **Prognose** nicht in der Visualisierung enthalten ist. Wenn das Dialogfeld **Prognose** vorhanden ist, erzeugt es bei der Aktivierung dasselbe Standardmodell wie die Insights für Zeitreihen. Zeitreihenpunkte werden automatisch in chronologischer Reihenfolge zum Zweck der Insight-Erkennung sortiert, aber im Gegensatz zur Prognosefunktion werden die in der Visualisierung angezeigten Zeitpunkte nicht sortiert.

Algorithmen

Die Details für Zeitreihen basieren auf einem Modell mit exponentiellem Glätten für die beobachteten Zeitreihendaten. Beobachtete Zeitreihenwerte und berechnete Modellkomponenten werden verwendet, um Insights für die Zeitreihe zu erstellen: ungewöhnliche Werte, saisonale Effekte und Insights zu Trends. Jeder Insight-Typ hängt von einer bestimmten Kombination von Daten und entsprechenden Modellen mit exponentiellem Glätten ab.

Details

Ungewöhnliche Werte

Ein Modell mit exponentiellem Glätten stellt einen Prognosewert für jeden beobachteten Zeitpunkt bereit. Ein Prognosewert zu einem Zeitpunkt ist die Ein-Schritt-voraus-Prognose zum vorherigen Zeitpunkt. Für jeden Prognosewert wird ein Konfidenzintervall berechnet, das die entsprechende vom Modell abhängige Prognosewertvarianz verwendet. Ein beobachteter Zeitreihenwert, der außerhalb des Konfidenzintervalls für den entsprechenden, auf dem Modell basierenden Prognosewert gefunden wird, wird als *ungewöhnlicher Wert* betrachtet.

Ungewöhnliche Werte werden basierend auf dem ausgewählten Modell mit exponentiellem Glätten für die Zeitreihe erkannt. Das Konfidenzniveau, das für die Berechnung der Prognosekonfidenzintervalle verwendet wird, liegt bei 99,74 %. Bis zu fünf ungewöhnliche Werte werden durch Auflisten der entsprechenden Zeitpunkte berichtet. Cognos Analytics listet die Punkte nicht in chronologischer Reihenfolge, sondern entsprechend dem Abstand vom Konfidenzintervall (in absteigender Reihenfolge) auf. Weitere ungewöhnliche Punkte werden zuerst aufgelistet. Ungewöhnliche Werte werden - falls möglich - als ungewöhnlich hoch oder ungewöhnlich niedrig angegeben.

Ein ungewöhnlicher Wert, der zum letzten Zeitpunkt erkannt wird, wird separat gemeldet. Dies kann bedeuten, dass die Daten unvollständig sind. Der ausgewertete Wert für den letzten Monat könnte die täglichen Daten beispielsweise nur für einen halben Monat wiedergeben.

Saisonale Effekte

Die Insight *Saisonale Effekte* zeigt die saisonale Länge für eine Zeitreihe, die vom Modell angegeben wird. Die saisonale Länge entspricht einer festen Dauer eines saisonalen Musters, das in der Zeitreihe festgelegt wurde. Eine durchschnittliche Temperaturschwankung über 12 Monate legt zum Beispiel ein Jahresmuster fest. Diese Insight liefert auch die Stärke der saisonalen Effekte und berichtet die Zeiträume mit den größten und kleinsten saisonalen Werten.

Die saisonale Länge wird aus dem ausgewählten Modell abgerufen. Sie wird aus dem saisonalen Zeitraum und dem Datum oder dem Zeitintervall abgeleitet, die in den Prognosestatistikdetails gemeldet werden. Ein saisonales Modell wird nur ausgewählt, wenn es sich besser eignet als alle nicht saisonalen Modelle. Der saisonale Zeitraum für das ausgewählte saisonale Modell wird durch den Vergleich von Modellen mit mehreren möglichen saisonalen Zeiträumen ermittelt.

Saisonale Effekte werden abhängig vom Wert der berechneten Stärke als schwach, moderat oder stark gemeldet. Die Stärke der saisonalen Effekte wird berechnet als Verringerung des Modellfehlers des saisonalen Modells im Vergleich zu einem entsprechenden nicht saisonalen Modell und dividiert durch den nicht saisonalen Modellfehler. Dies unterscheidet sich von der saisonalen Stärke, die in den Prognosestatistikdetails gemeldet wird, wo der Unterschied in der Genauigkeit zwischen den beiden Modellen angegeben wird.

Die größten und kleinsten saisonalen Werte werden auf der Basis der zugrunde liegenden Durchschnittswerte für die saisonalen Modellkomponenten über alle saisonalen Muster in der Zeitreihe berechnet. Die entsprechenden Zeiträume werden gemeldet, wenn die Durchschnittswerte über die Mehrheit der saisonalen Muster durchgängig die größten oder die kleinsten Werte sind.

Trend

Die Insight *Trend* meldet eine insgesamt positive oder negative Richtung der Zeitreihenwerte (falls vorhanden). Sie liefert außerdem die Stärke des Trends.

Die Komponenten 'Stufe' und 'Trend' werden beide aus dem entsprechenden Modell mit exponentiellem Glätten extrahiert. Wenn das Modell keine Komponente 'Trend' aufweist, wird nur die Komponente 'Stufe' verwendet. Dadurch wird eine Trendkurve für die Zeitreihendaten definiert. Für die Trendkurve werden dann der Rangkorrelationskoeffizient nach Kendall und der entsprechende statistische Test berechnet. Sie erkennen eine insgesamt positive oder negative Richtung der Zeitreihenwerte. Unterschiedliche Tau-Wertebereiche definieren den gemeldeten Stärkegrad für den Trend: schwach, moderat oder stark.

Weitere Informationen zu Modellen mit exponentiellem Glätten finden Sie unter [„Prognosemodelle“](#) auf Seite 59.

Beziehungen

Visualisierungen von Beziehungen in einer Exploration werden erstmals angezeigt, wenn Sie Daten für die Exploration angeben.

Übersicht

IBMCognos Analytics bietet eine schnelle Übersicht über die Beziehungen zwischen Paaren von Feldern, die sich auf ein einzelnes für Sie interessantes Gebiet konzentrieren. Die Visualisierung umfasst mehrere Registerkarten. Jede Registerkarte wird für ein anderes Feld verwendet, das für Sie von Interesse ist.

Diese Informationen sind sehr nützlich, wenn Sie sich einen Überblick über die Vielzahl relevanter Beziehungen verschaffen wollen, die in den Daten verfügbar sind und gegebenenfalls weiter untersucht werden sollen.

Algorithmen

Während das erste Feld von Interesse anhand einer semantischen Datenanalyse ermittelt wird, haben Sie die Möglichkeit, ein anderes Feld anzugeben, das für Sie von Interesse ist. Jede Registerkarte enthält ein Netzdiagramm mit Feldern als Knoten und Links zwischen Paaren von Knoten zur Darstellung der relativen Stärke der entsprechenden Beziehung zwischen den Knoten. Wenngleich die Links von dem Feld, das von besonderem Interesse ist, im Diagramm vorherrschen, werden auch andere zugehörige Feldpaare mit starken Beziehungen angezeigt. Mithilfe eines Schiebereglers können Sie das Diagramm so anpassen, dass eine größere oder kleinere Anzahl an Knoten im Netz angezeigt wird.

Details

Daten für die Analyse

Für Beziehungen werden nicht ausgewertete Daten verwendet, um die Stärke der Beziehungen unter allen berücksichtigten Feldpaaren zu berechnen. Um die Beziehungsstärke als Kennzahl zu standardisieren und sie für sämtliche Feldpaare vergleichbar zu machen, werden zunächst alle numerischen Felder in Bins (Klassen) eingeteilt. Alle Felder in den Daten werden als kategoriale Felder behandelt. Das angewendete Binning (Klasseneinteilung) entspricht der Einteilung nach Häufigkeit. Hierbei werden fünf Bins generiert. Weitere Informationen hierzu finden Sie im Abschnitt zur Datenaufbereitung für numerische Felder.

Beziehungsstärke

Zunächst werden die Daten für jedes einzelne Paar kategorialer Felder tabuliert, und zwar für alle Kombinationen von Feldkategorien, die in den Daten gefunden werden. Auf Grundlage der tabulierten Daten wendet IBM Cognos Analytics den Chi-Quadrat-Test der Unabhängigkeit an, um zu beurteilen, ob die Felder unabhängig sind. Ist die Abweichung bei der Unabhängigkeit signifikant, berechnet Cognos Analytics die Effektgröße auf Grundlage der Chi-Quadrat-Statistik. Hierbei handelt es sich um Cramérs V, ein häufig verwendetes Maß für den Zusammenhang zwischen zwei kategorialen Feldern. Die Werte für diese Kennzahl liegen im Bereich von 0 bis 1; Cognos Analytics gibt den Wert der Beziehungsstärke an, der als Prozentsatz ausgedrückt wird. Beziehungen mit einer Stärke von weniger als 10 % werden nicht aufgeführt, da sie als zu schwach angesehen werden, um einen praktischen Nutzen zu haben.

Leistungseinschränkungen

Eine Berechnung der Beziehungsstärke zwischen sämtlichen Feldpaaren im Datensatz ist bei großen Datenmengen nicht tragbar. In Cognos Analytics ist die Anzahl der verarbeiteten Felder auf 100 begrenzt, um eine schnelle Antwort bereitstellen zu können. Diese Felder werden jedoch von einem anderen Prozess ausgewählt und ein möglicher Verlust relevanter Beziehungen ist minimal. Wenn die Daten mehr als 10.000 Zeilen enthalten, ruft Cognos Analytics aus Leistungsgründen eine Zufallsstichprobe dieser Größe (also mit 10.000 Zeilen) ab. Mit dieser Datengröße wird sichergestellt, dass der Verlust an Genauigkeit bei der Schätzung der Beziehungsstärke minimal ist.

Unterschiede zwischen Cognos Analytics und Watson Analytics in Bezug auf erweiterte Datenanalyse

In IBM Cognos Analytics wurden die Methoden von Watson Analytics für die erweiterte Analyse angepasst und verbessert. Infolgedessen lassen sich eine Reihe von Unterschieden feststellen.

Datenaufbereitung

Die Datenaufbereitung ist sowohl in Cognos Analytics als auch in IBM Watson Analytics von entscheidender Bedeutung, da ausschließlich aufbereitete Daten in Analysen für zentrale Einflussfaktoren und

Entscheidungsstrukturen aufgenommen werden. Die Datenaufbereitung unterscheidet sich zwischen den beiden Packages hinsichtlich mehrerer wichtiger Aspekte.

Unterschied	Cognos Analytics	Watson Analytics
Fehlende Werte	<p>Fehlende Werte im Zielfeld werden in beiden Produkten auf gleiche Art und Weise verarbeitet: Die Zeilen mit fehlendem Wert im Zielfeld werden aus allen nachfolgenden Analysen ausgeschlossen. Fehlende Werte für Prädiktorfelder werden hingegen unterschiedlich verarbeitet.</p> <p>Es gibt eine separate Kategorie für kategoriale Prädiktoren.</p> <p>Numerische Prädiktoren werden immer in Klassen (Bins) eingeteilt und fehlende Werte werden als separater Bin behandelt.</p>	<p>Fehlende Werte werden auf Grundlage der Beziehung zwischen dem Ziel und einem Prädiktor imputiert.</p> <p>Prädiktorfelder mit mehr als 25 % fehlenden Werten werden in Watson Analytics ausgeschlossen.</p>
Zusammenführung von Kategorien	<p>Cognos Analytics ordnet Kategorien nach der Häufigkeit ihres Auftretens in den Daten und gruppiert alle Kategorien ab der 50. Kategorie zu einer einzelnen Kategorie zusammen. Alle Kategorien mit einer Häufigkeit von weniger als 3 werden ebenfalls zusammengeführt. Eine Anforderung besteht darin, dass zusammengeführte Kategorien weniger als 25 % der Datenzeilen darstellen müssen. Andernfalls wird das Feld nicht für die Analyse verwendet. Auch Zielkategorien werden zusammengeführt.</p>	<p>Die Zusammenführung von Kategorien für Prädiktoren hängt von den Zielwerten ab. Es werden nur solche Kategorien zusammengeführt, die Zielfeldwerte mit ausreichender Ähnlichkeit aufweisen.</p>
Verarbeitung von Ausreißern	<p>Für Ausreißer in fortlaufenden Zielfeldern werden extreme Grenzwerte festgelegt und die Werte jenseits der Grenzen werden zu Analyse Zwecken durch die entsprechenden Grenzwerte ersetzt. Dadurch wird die Analyse sowohl für zentrale Einflussfaktoren als auch für Entscheidungsstrukturen verbessert, da sich auf diese Weise Beziehungen mit dem Ziel leichter erkennen lassen. Ausreißer, die bei der Datenaufbereitung für Prognosezwecke verarbeitet werden, werden in der Visualisierung nicht zurückgegeben.</p>	<p>Eine automatisierte Verarbeitung von Ausreißern ist nicht verfügbar.</p>
Stichprobenentnahme	<p>Cognos Analytics analysiert Datenstichproben, die jeweils etwa 10.000 Zeilen enthalten, wenn die ursprünglichen Daten diese Größe überschreiten. Wenn die entsprechende Analyse anhand einer Datenstichprobe durchgeführt wird, wird stets eine Warnung angezeigt. Al-</p>	<p>Watson Analytics verwendet zum Generieren von zentralen Einflussfaktoren und Entscheidungsstrukturen den vollständigen Zeilensatz in den Daten. Dies kann zu einer langsamen Verarbeitung führen.</p>

Unterschied	Cognos Analytics	Watson Analytics
	lerdings liegt die Prognosestärke generierter zentraler Einflussfaktoren und Entscheidungsstrukturen erwartungsgemäß nahe an der Prognosestärke der Modelle, die auf den entsprechenden vollständigen Daten basieren.	

Zentrale Einflussfaktoren

Unterschied	Cognos Analytics	Watson Analytics
Zentrale Einflussfaktoren	<p>Cognos Analytics bietet zentrale Einflussfaktoren sowohl für fortlaufende als auch für kategoriale Ziele in den Visualisierungen des Typs 'Spiraldiagramm' und 'Einflussfaktorenanalyse'.</p> <p>Während sich die Schritte zur Datenaufbereitung sowohl auf unidirektionale als auch auf bidirektionale zentrale Einflussfaktoren auswirken, unterscheidet sich die Cognos Analytics -Analyse für bidirektionale zentrale Einflussfaktoren von Watson Analytics im Hinblick auf die Kriterien zur Auswahl der Einflussfaktoren, die den Benutzern angezeigt werden sollen.</p>	Zentrale Einflussfaktoren sind sowohl für fortlaufende als auch für kategoriale Ziele in Spiraldiagrammvisualisierungen verfügbar.
Statistische Tests	Für kategoriale Ziele verwendet Cognos Analytics den Chi-Quadrat-Test der Unabhängigkeit, um festzustellen, ob zwei kategoriale Felder unabhängig sind.	Watson Analytics verwendet den Likelihood-Quotiententest für kategoriale Ziele.
Auswahlkriterium	Cognos Analytics setzt voraus, dass die Prognosestärke eines bidirektionalen Einflussfaktors eine relative Verbesserung von mehr als 10 % im Vergleich zur Prognosestärke des entsprechenden unidirektionalen Einflussfaktors bietet. Die relative Verbesserung der Prognosestärke wird anhand der prozentualen Reduzierung des Fehlers des bidirektionalen zentralen Einflussfaktors im Verhältnis zum minimalen Fehler für die beiden entsprechenden unidirektionalen zentralen Einflussfaktoren gemessen. Dieses Kriterium stellt sicher, dass nur die relevanten bidirektionalen Einflussfaktoren ausgewählt werden.	Watson Analytics verwendet die bidirektionale Interaktion als Auswahlkriterium für bidirektionale zentrale Einflussfaktoren. Um ausgewählt zu werden, muss ein bidirektionaler zentraler Einflussfaktor eine Wechselwirkungseffektgröße aufweisen, die größer ist als ein bestimmter Schwellenwert.
Prognosestärke	Die Prognosestärke für Spiraldiagramme und Einflussfaktorenanalyse mit fortlaufendem Ziel wird als korrigiertes R im Quadrat berechnet.	<p>Die Prognosestärke für fortlaufende Ziele wird als R im Quadrat berechnet.</p> <p>Die Prognosestärke für kategoriale Ziele wird als Klassifikationsgenauigkeit berechnet.</p>

Unterschied	Cognos Analytics	Watson Analytics
	<p>Die Prognosestärke für das kategoriale Ziel wird als korrigiertes R im Quadrat berechnet. Sie wird ermittelt, indem die Verbesserung der Klassifikationsgenauigkeit des Modells im gesamten Konstantenmodell berechnet und durch den Konstantenmodell-Klassifikationsfehler dividiert wird. Das Konstantenmodell prognostiziert immer den Zielmodus und die zugehörige Klassifikationsgenauigkeit wird durch die Modushäufigkeit geschätzt. Eine zuverlässige Prognosebeziehung wird berichtet, wenn die Prognosestärke des Modells größer als ein Standardschwellenwert von 10 % ist.</p> <p>Dies führt dazu, dass die Prognosestärke für eine Spiraldiagrammvisualisierung mit kategorialem Ziel in Watson Analytics und Cognos Analytics unterschiedlich ausfallen kann, und zwar selbst dann, wenn Ziel und Eingaben vergleichbar sind.</p>	

Entscheidungsstruktur

Unterschied	Cognos Analytics	Watson Analytics
Entscheidungsstruktur	<p>Zusätzlich zu den Visualisierungen des Typs 'Verzeichnisstruktur' und 'Entscheidungsregeln' stellt Cognos Analytics auch die neue Sunburst-Visualisierung bereit. Die Unterschiede beim Generieren einer Entscheidungsstruktur in Cognos Analytics oder Watson Analytics hängen vom fortlaufenden bzw. kategorialen Ziel ab. Die Prognosestärke für Entscheidungsstrukturen wird in Visualisierungen des Typs 'Spiraldiagramm' und 'Einflussfaktoranalyse' in Cognos Analytics angezeigt, jedoch nur dann, wenn sie größer ist als die Prognosestärke des größten zentralen Einflussfaktors.</p>	<p>Visualisierungen des Typs 'Entscheidungsstruktur' und 'Entscheidungsregeln' sind verfügbar. Die Prognosestärke der Struktur ist in der Spiraldiagrammvisualisierung immer verfügbar.</p>
Fortlaufendes Ziel	<p>Entscheidungsstrukturen für fortlaufende (stetige) Ziele werden sowohl in Watson Analytics als auch in Cognos Analytics auf Grundlage der Summe der Quadrate generiert.</p> <p>Cognos Analytics verwendet den Informationsgewinn, der in Watson Analytics im Zusammenhang mit der Effektgröße</p>	<p>Watson Analytics verwendet den CHAID-Algorithmus und die entsprechenden F-Tests beim Zusammenführen und Aufteilen von Strukturknoten. Dies ermöglicht mehrfache Aufteilungen für die Entscheidungsstrukturknoten.</p>

Unterschied	Cognos Analytics	Watson Analytics
	für F-Tests steht, direkt bei der Zusammenführung und Aufteilung von Strukturknoten. Dies ermöglicht mehrfache Aufteilungen für die Entscheidungsstrukturknoten.	
Stoppregeln	<p>In Cognos Analytics ist die Anzahl der Verzeichnisstrukturknoten auf 36 begrenzt.</p> <p>Die minimale Anzahl der Zeilen, die den Strukturknoten entsprechen, ist 25.</p> <p>Felder ohne Beitrag zur Prognosestärke werden aus den Verzeichnisstrukturen von Cognos Analytics entfernt. Cognos Analytics generiert kleinere Verzeichnisstrukturen, die für die Visualisierung und Insights besser geeignet sind. Gleichzeitig weisen die generierten Verzeichnisstrukturen eine vergleichbare und mitunter sogar größere Genauigkeit auf als in Watson Analytics.</p>	<p>In Watson Analytics wird die Größe von Verzeichnisstrukturen begrenzt, indem die Anzahl der Verzeichnisstrukturebenen auf 5 limitiert wird.</p> <p>Die minimale Anzahl von Zeilen, die den Strukturknoten entsprechen, ist 50.</p> <p>Strukturen in Watson Analytics sind tendenziell nur dann genauer, wenn sie eine wesentlich größere Anzahl von Knoten enthalten.</p>
Wichtigkeit von Variablen	Die Wichtigkeit von Variablen wird als relative Verbesserung der Prognosestärke berechnet, die erreicht wird, wenn die betreffende Variable der Entscheidungsstruktur hinzugefügt wird. Prädiktoren mit einer Variablenwichtigkeit an einem Schwellenwert oder unterhalb eines Schwellenwerts werden aus der Entscheidungsstruktur ausgeschlossen. Für die meisten Verzeichnisstrukturen wird der Schwellenwert auf null gesetzt. Bei Verzeichnisstrukturen, die auf einer geringeren Anzahl an Datenzeilen basieren und eine geringere Prognosestärke aufweisen, ist der Schwellenwert geringfügig höher, um mögliche Überanpassungseffekte zu reduzieren.	Die Wichtigkeit von Variablen wird auf Grundlage der Sensitivität der Ergebnisse berechnet, wenn die Variable unterschiedliche Werte annimmt.
Prognosestärke	<p>Die Prognosestärke für Entscheidungsstrukturen mit fortlaufendem Ziel wird als R im Quadrat berechnet.</p> <p>Die Prognosestärke für das kategoriale Ziel wird als korrigiertes R im Quadrat berechnet.</p> <p>Beim korrigierten R im Quadrat wird zunächst der Klassifikationsfehler der Verzeichnisstruktur vom grundlegenden Klassifikationsfehler subtrahiert, der sich dadurch ergibt, dass stets die häufigste Zielkategorie ausgewählt wird. Anschließend wird das Ergebnis durch den</p>	<p>Die Prognosestärke für Entscheidungsstrukturen mit fortlaufendem Ziel wird als R im Quadrat berechnet.</p> <p>Die Prognosestärke für kategoriale Ziele wird als Klassifikationsgenauigkeit berechnet.</p>

Unterschied	Cognos Analytics	Watson Analytics
	<p>grundlegenden Klassifikationsfehler dividiert. Dieser Wert stellt die relative Verbesserung dar, die durch die Entscheidungsstruktur erreicht wird. Dies führt dazu, dass die Prognosestärke für Entscheidungsstrukturen mit kategorialen Ziel in Watson Analytics und Cognos Analytics unterschiedlich ausfallen kann, und zwar selbst bei vergleichbaren Entscheidungsstrukturen.</p>	

Steuerelemente

Unterschied	Cognos Analytics	Watson Analytics
Steuerelemente	<p>Ein Feld wird als fortlaufend angegeben, indem die Dateneigenschaft 'Verwendung' auf 'Kennzahl' gesetzt werden. Ein Feld wird als kategorial (nominal) angegeben, indem 'Verwendung' auf 'Identifizierung' oder 'Attribut' gesetzt wird.</p> <p>Es ist möglich, Prädiktoren direkt in die entsprechenden Visualisierungen einzuschließen oder direkt aus diesen auszuschließen. Hierfür wird das Dialogfeld 'Einflussfaktoren bearbeiten' verwendet.</p>	<p>Felder können als fortlaufende, nominale oder ordinale Felder angegeben werden, indem unter der Option für Präzisierung (Refine) die entsprechenden Dateneigenschaften festgelegt werden.</p> <p>Mithilfe einer weiteren Dateneigenschaft für Präzisierung können Prädiktoren in die Modelle eingeschlossen oder aus diesen ausgeschlossen werden.</p>

Kapitel 6. Assistent

Assistent


IBM Cognos Analytics verfügt über einen integrierten Assistenten, der textbasierte Eingaben unterstützt, mit deren Hilfe Sie sich schnell Insights für Ihre Daten verschaffen und Ihre Analysen vereinfachen können. In nur wenigen Schritten können Sie auf Schlüsseldatenquellen zugreifen, Visualisierungen erstellen und sie in Ihren Erstellungsbereich 'Exploration' oder 'Dashboard' ziehen. Die Texteingabe wird nur in englischer Sprache unterstützt.

Sie können zwei Schnittstellen verwenden:

1. Vollständige Assistentenanzeige
2. Kompakte Assistentenanzeige


Vollständige Assistentenanzeige

Die vollständige Assistentenanzeige besteht aus drei Bereichen: Symbolleiste, Ausgabe und Eingabefeld. Sie können Aktionen in der Symbolleiste ausführen, wie z. B. das Zurücksetzen Ihres Protokolls. Die Ausgabe wird in einem Blätterbereich angezeigt, der alle früheren Antworten enthält. Im Eingabefeld, das standardmäßig mit **Frage stellen** beschriftet ist, können Sie textbasierte Dialogeingaben vornehmen.

Öffnen Sie die vollständige Anzeige durch Klicken auf das Symbol **Assistent** . Wenn Sie die kompakte Anzeige verwenden (nur in **Exploration**), können Sie zur vollständigen Anzeige wechseln, indem Sie auf **In vollständiger Ansicht öffnen** klicken.

Kompakte Assistentenanzeige

Die kompakte Anzeige ist nur in der Schnittstelle **Exploration** verfügbar. Sie können die kompakte Anzeige öffnen, indem Sie im Kompaktanzeigebereich auf **Frage stellen** klicken.

In der kompakten Anzeige werden nur die letzte Eingabe und Antwort angezeigt. Klicken Sie zum Anzeigen des vollständigen Protokolls auf den Link **In vollständiger Ansicht öffnen** in der kompakten Anzeige oder auf das Symbol **Assistent** .


Im Feld **Frage stellen** können Sie mithilfe der Taste 'Aufwärtspfeil' oder 'Abwärtspfeil' eine Liste der empfohlenen Befehle anzeigen. Bei der Texteingabe bietet Ihnen das Type-ahead-Feature Vorschläge zu möglichen Fragen an. Dies ist eine wunderbare und zeitsparende Möglichkeit, um passende Fragen zu stellen.

Wenn Sie die letzte Frage wiederholen möchten, klicken Sie auf die Antwortblase oder verwenden Sie die Tastenkombination (Umschalttaste + Aufwärtspfeil) und drücken Sie anschließend die Eingabetaste. Durch Drücken der Umschalttaste + Aufwärts-/Abwärtspfeil können Sie durch Ihre vorherigen Fragen blättern.

Visualisierungen in der Anzeige **Assistent** sind komprimiert und schließen möglicherweise einige Informationen aus. Die vollständigen Visualisierungsdetails können unter **Erkunden** oder **Dashboard** angezeigt werden. Sie können Visualisierungen vom **Assistenten** zu einem Erstellungsbereich unter **Exploration** oder **Dashboard** ziehen.

The screenshot displays the IBM Cognos Analytics interface. On the left, a sidebar shows a list of fields: 'Total number of rows' (5397), 'Street Name', 'Hundred Block', 'Local Area', 'Base Type', and 'Minute'. A button labeled 'show Hour and Local Area' is visible. Below this, a visualization titled 'Hour and Local Area' is shown, which is a bubble chart. A legend for 'Local Area' is provided, listing: Marpole (blue), Grandview-Woodland (purple), Kensington-Cedar Cottage (green), Mount Pleasant (pink), Arbutus Ridge (yellow), and Hastings-Sunrise (orange). On the right, a table is displayed with the following data:

Hour	Local Area
0	Grandview-Woodland
0	Kensington-Cedar Cottage
0	Mount Pleasant
1	Arbutus Ridge
1	Hastings-Sunrise

Sie können den Inhalt Ihres Dialogprotokolls löschen und das Protokoll auf den Anfangsstatus zurücksetzen, indem Sie auf das Symbol **Zurücksetzen**  in der Symbolleiste klicken. Die Eingabe von reset oder clear erzielen Sie dasselbe Ergebnis. Das Symbol **Zurücksetzen** wird nur in der vollständigen Anzeige unterstützt und ist in der kompakten Anzeige nicht verfügbar.

Wichtig: Ihr Protokoll wird auch beim Aktualisieren Ihrer Cognos Analytics-Browserregisterkarte zurückgesetzt.

Funktionen des Assistenten

Der Assistent ist ein leistungsfähiges Feature, das es Ihnen ermöglicht, Fragen in natürlicher Sprache zu stellen. Im vorliegenden Abschnitt werden einige Befehle beschrieben, die mit dem Assistenten verwendet werden können.

Stellen Sie Fragen in natürlicher Sprache. Beispiel: show Profit ist ein Synonym für list Profit, what is Profit?, tell me about Profit und Profit.

Hier finden Sie einige häufig verwendete Fragen zur Einführung:

help

Übersichten und allgemeine Nutzungsinformationen werden angezeigt.

show data

Alle unterstützten Datenquellen, auf die Sie in **Eigener Inhalt** und **Teaminhalt** zugreifen können, werden aufgelistet. Zu den unterstützten Datenquellen gehören:

- Hochgeladene Dateien (.csv, .tsv, .xls, .xlsx und .zip)
- Datenmodule
- Dataset
- Framework Manager-Packages

Anmerkung: Weitere Informationen finden Sie im Abschnitt *Packages aufbereiten* im Dokument *Verwaltung*.

Wenn die Anzahl der Ergebnisse die Anzahl der angezeigten Elemente überschreitet, klicken Sie auf **Mehr**, um bis zu 100 Datenquellen anzuzeigen. Blättern Sie zum Ende der Liste und klicken Sie auf **Weniger**, um die Liste zu reduzieren.

show source <quellenname>

Eine Liste von für die angegebene Datenquelle relevanten Feldern und Details wird angezeigt. In der Symbolleiste wird der jeweilige <quellenname> angezeigt. Cognos Analytics verwendet standardmäßig die aktive Quelle in Ihrer Dashboard- oder Explorationsanzeige. In den Ergebnissen können Sie auf Feldnamen klicken, um weitere Informationen zu diesen jeweiligen Feldern zu erhalten. Das Klicken auf Feldnamen entspricht der Frage `show column <spaltenname>`.

Wenn die Anzahl der Ergebnisse die Anzahl der angezeigten Elemente überschreitet, klicken Sie auf **Mehr**, um bis zu 100 Datenquellen anzuzeigen. Blättern Sie zum Ende der Liste und klicken Sie auf **Weniger**, um die Liste zu reduzieren.

show column <quellenname>

Für die angegebene Spalte werden Informationen und zugehörige Felder angezeigt. Das Klicken auf die zugehörigen Felder entspricht der Frage `show chart <spalte1> and <spalte2>`. Wenn die angegebene Spalte Einflussfaktoren aufweisen soll, können Sie `what influences <spaltenname>` eingeben, damit die Liste der Einflussfaktoren angezeigt wird.

what influences column <spaltenname>

Eine Liste von Feldern, die die Ergebnisse in der angegebenen Spalte beeinflussen, wird angezeigt.

show chart <spalte1> und <spalte2>

Zeigt Visualisierungen an, die die Beziehung zwischen <spalte1> und <spalte2> darstellen. Blättern Sie in den Visualisierungen, indem Sie auf den Linkspfeil bzw. den Rechtspfeil klicken. Jede Visualisierung enthält ein Informationssymbol in der rechten oberen Ecke. Bewegen Sie den Mauszeiger über das Symbol, um Beschreibungen zu den zugrunde liegenden Daten anzuzeigen. Optional können Sie weitere Spalten eingeben, eine zu großen Anzahl an Spalten kann jedoch zu weniger effektiven Visualisierungen führen.

Wenn Sie auf `Show related visualizations` klicken, werden Visualisierungen auf der Basis von Einflussfeldern und zugehörigen Feldern zurückgegeben.

Durch Klicken auf `Create dashboard from the charts` wird ein neues Dashboard basierend auf den zuletzt generierten Diagrammen erstellt. Die Eingabe von `Create related dashboard` erzeugt dasselbe Dashboard als Ergebnis. Wenn die Diagramme oberste (top) oder unterste (bottom) Aggregationen enthalten, werden diese Modifikatoren auf das erstellte Dashboard angewendet.

Durch die Anwendung von Aggregationen und Filtern können bestimmte Daten hervorgehoben und aussagekräftigere Visualisierungen erstellt werden. Zu den gängigen Aggregationen gehören 'total', 'average', 'count', 'maximum/minimum', 'top/bottom', 'best/worst' usw. Im Folgenden sehen Sie einige Aggregationsbeispiele:

show top <num> <spalte1> by <spalte2>

Zeigt die oberen Werte von <spalte1> auf Basis des Kontexts von <spalte2> an. Beispiel: `show top 5 Sales by Region`. Ist <num> nicht angegeben, wird der Standardwert 10 verwendet.

<spalte1> ist eine aggregierte oder nicht aggregierte Kennzahl, bei <spalte2> handelt es sich um eine kategoriale Spalte.

show average <spaltenname>

Zeigt den Durchschnitt für alle Werte an, die in <spaltenname> gefunden wurden.

how many <spaltenname>

Wenn <spaltenname> eine Kategorie ist, wird die Anzahl der eindeutigen Elemente zurückgegeben. Wenn <spaltenname> eine Kennzahl ist, wird die Gesamtzahl zurückgegeben.

show maximum <spaltenname>

Zeigt den höchsten Werte an, der in <spaltenname> gefunden wurden.

show minimum <spaltenname>

Zeigt den niedrigsten Werte an, der in <spaltenname> gefunden wurden.

show total <spaltenname>


Zeigt die Gesamtzahl für alle Werte an, die in <spaltenname> gefunden wurden.

Sie können Filter für Zeichenfolgen für geografische Begriffe (z. B. Country oder State) oder für zeitliche Begriffe (z. B. Month oder Year) hinzufügen. Die Filterung nach Datum wird derzeit nicht unterstützt. Aggregationen und Filter können kombiniert werden, um differenziertere Ergebnisse zu erzielen. Im Folgenden finden Sie einige Beispiele, die auf Beispieldaten basieren:

- show Education by Income where Income is less than 1000
- show Education by Income where Income > 100K

Verwenden Sie optional K (für 'Tausend') oder M (für Millionen).

- show Revenue in 2017 and 2018
- show Income by Month for New York City
- what are the top 5 States by average Inventory, excluding California

Gefilterte Visualisierungen enthalten ein Filtersymbol () , das sich in der rechten oberen Ecke des Diagramms befindet. Bewegen Sie den Mauszeiger über das Symbol, um die angewandte Filterung anzuzeigen.

create dashboard

Generiert ein neues Dashboard auf der Basis der derzeit ausgewählten Datenquelle. Sie können die Visualisierungen, Registerkarten, Reihenfolge usw. ändern und das neue Dashboard speichern. Standardmäßig enthält das generierte Dashboard erweiterte Analyse- und Prognosedigramme.

Die automatische Generierung von Dashboards für umfassendere Datenquellen kann zu Leistungsproblemen führen. Um dies zu vermeiden, können Sie `create simple dashboard` eingeben, um ein Basisdashboard zu generieren. Anschließend können Sie das Dashboard ändern, indem Sie Diagramme durch komplexere Visualisierungen wie Analysen von Einflussfaktoren oder Spiraldiagrammen ersetzen.

