Use SPSS Statistics direct marketing analysis to gain insight

Analyze customer history using RFM

David Gillman

May 12, 2014
(First published October 23, 2012)

Introduction

Breaking customers into groups is a natural tendency. Companies want to know who are their best customers, who are their worst customers, who has potential, who is new, and so on. Marketing and sales departments do this regularly and often. Their goal is to expend limited effort to achieve maximum return (sales, in this case).

Classifying and grouping customers may be a natural function of human nature and business operations, but doing it well is a subject of study, discussion, and practice. One type of segmentation modeling built into wizards in IBM SPSS Statistics is recency, frequency, and monetary value (RFM) segmentation. RFM is a proven and widely used method for dividing customers into groupings based on their behaviors. A quick scan of the customer list when grouped by the RFM score shows you who your best customers are and who your bad or dead customers are.

RFM modeling is not the only way of segmenting customers, and it isn't necessarily the best way. It is, however, a good method of segmenting customers that anyone can easily understand and put to use quickly.

Knowing how your customers break into groups is useful. You can use that information to predict customer behavior in the near future. Even more useful is monitoring how individual customers' RFM scores change over time. Using that knowledge, you can change business processes to
maximize a customer's life cycle. And you can glean all of this information from an easy-to-use wizard in SPSS Statistics.

**RFM models**

If you are not in marketing, you might not have heard of RFM segmentation. No worries, it's easy to grasp:

- **Recency**  
  *Recency* refers to how long ago a customer placed their last order. This metric is used, because in many situations, it has been shown that customers who last ordered a long time ago are far less likely to order from you again compared to customers who ordered more recently.

- **Frequency**  
  *Frequency* refers to how many times a customer ordered from you over his or her lifetime. This metric is used, because someone who has ordered from you once is far less likely to order again compared to someone who has ordered from you many times. Frequency is sometimes tweaked a bit. After review and examination in your operations, you may come up with a slightly different definition of *frequency*. For example, you might use the number of orders per year rather than orders over an entire lifetime. Another variant is to use orders only over a certain value in the frequency calculation (negating small orders and the effect that some customers might have by placing many tiny orders, which drives up processing, delivery, and receivables costs).

- **Monetary value**  
  *Monetary value* refers to the worth of the customer. Most RFM analyses use either gross revenue or net profit over the lifetime of the customer. Which you use depends on the opinion of influential people in the company. You can define monetary value other ways as well. Using net profit per order may change the outcome. Seeing the difference in how a customer is ranked between the different monetary value metrics can be insightful.

Having defined the *R*, *F*, and *M*, let's look at the model.

Think of each category (*R*, *F*, or *M*) as an ordered list of customers based on the value of the metric. Divide that ordered list into equal parts—typically, three or five, but any number will work. For example, customers who order most frequently will all receive a 5 out of 5; customers who ordered only one time will get a 1 out of 5.

Use the same ranking system for the other metrics. Each customer then has a three-number score, such as 114, 352, or 445. In the default SPSS Statistics case, the lower each number, the better. Though simple in outcome, many industries use RFM models for quick but powerful segmentation. RFM modeling comes originally from the direct marketing industry (think catalogs by mail). The modern equivalent to mail order catalogs is e-commerce. Companies use RFM modeling to send targeted offers to get customers to come back to the site and maintain name recognition by email.

Another variant industry using RFM is business-to-business distribution. Here, a business can use knowledge about the customer to determine price lists—more discounts to more active
and valuable customers. One could also use recency to quickly see when good customers stop ordering, and then prepare an offer to get them to come back.

**Historical analysis using RFM**

A single RFM model is a snapshot in time. Comparing several models over time is a way to model the customer lifecycle.

Seeing how customers move from different RFM classes over their lifespan gives marketing and salespeople a lot of insight into customer behavior. Often, several tracks are visible. Knowing how different types of customers progress through the RFM model over time provides a foundation for altering business processes, making marketing offers or moving direct sales resources to the point of greatest impact.

For example, you may notice that new customers in one industry enter at an RFM score of 513 (meaning they are recent, not frequent, and have a medium value if scored out of 5). Their next move could be an improved RFM score of 534.

Next, you could see a split. Some customers could go down, while others go up in their RFM scores. Determining the difference between such customers could result in designing better offers, incentives, or service programs that get more customers onto the good track.

Keeping a data table where you records each customer's RFM score every time you run the models is the easiest way to do this.

**Other segmentation modeling techniques**

Before I dive into how to use SPSS Statistics wizards to create an RFM model, let me describe some of the other ways to segment customers in this tool. As you would expect, there are many ways of grouping customers, and SPSS Statistics supports many of the statistical processes used to accomplish the task. Clicking the **Analyze** menu, you can see several general categories of statistical analysis, including one labeled **Classify** (see Figure 1).
The **Classify** submenu shows the main algorithms available. These more advanced options will be useful to you when you move to creating custom segmentation models of your customers. However, their effective use does require a moderate level of statistical knowledge and, in reality, will be a learning process as you adapt them to fit the needs of your organization and the data you possess.

**RFM analysis in SPSS Statistics**

Let’s get to work. Before you start in SPSS Statistics, you need to gather your data, which you extract from your transactional systems. The type of data and the low complexity of the query you need might surprise you. Using some fairly basic queries that return the count of transactions, the sum of the amount, and the maximum value for the date, gather data that represents the:

- Customer number or other unique identifier;
- Last order date for each customer;
- Number of transactions that customer has had; and
- Total revenue for the customer.

As mentioned, you can use other definitions for the number of transactions and the total revenue for each customer. But the above list is a good starting point.

When the data is put together, it might look something like **Figure 2**. In this example, the data is in a spreadsheet, but you can have it in other formats. Just make sure that SPSS Statistics can read that file type.
With the data file put together, you are ready to start the analysis:

1. Start SPSS Statistics, then make a connection to the data file. You see the familiar Data Editor window filled with your customer file, as shown in Figure 3.

2. Click Direct Marketing > Choose Technique. The Direct Marketing window appears (see Figure 4).
Figure 4. The Direct Marketing window

3. Double-click **Help identify my best contacts (RFM Analysis)**.
4. In the **RFM Analysis: Data Format** window (see **Figure 5**), select **Customer data**, and then click **Continue**.

**Figure 5. Data organization choices**
The multi-tabbed **RFM Analysis from Customer Data** window appears in which you specify all the parameters for the RFM modeling process.

5. Click the **Variables** tab, shown in [Figure 6](#).
This tab has four data elements that you must define for the RFM modeling process to work. You must tell SPSS Statistics which variable in the incoming data (think columns in the spreadsheet) translate to the last transaction date, the number of transactions, and the amount.

**Figure 6. Defining data elements for RFM modeling**

6. After you map the data variables to the modeling input variables, include an identifier so that the model can give a score to each customer. For this example, specify the **Customer ID** field from the spreadsheet (see [Figure 7](#)).
7. Click the **Binning** tab, and then select the number of bins you want from the **Recency**, **Frequency**, and **Monetary** lists. **Binning** refers to how many bins, or divisions, you want for each metric. The default for each metric is **5**, which is a common number to use in real life. For simplicity, I adjusted my examples to work with **3** (see **Figure 8**).

**Figure 8. Selecting the number of divisions on the Binning tab**
Determining the number of bins you need

Determining how many divisions you should use can be tricky. Too many (nine or more), and you will create so many scores as to make the model difficult for people to understand. Too few, and the model will not be meaningful.

Common practice is to use an odd number, although I have never seen a satisfying explanation as to why. Regardless, using an odd number creates a middle category against which to compare the ends. That explanation is far from definitive, but I typically follow the custom of using an odd number, anyway.

8. In the **Binning Method** area, select **Nested** or **Independent**, as appropriate. The option you choose alters where people are placed for the frequency and monetary value scores. One method is not necessarily better than the other. Making a flow chart of the difference and discussing the procedure with your business users and decision-makers is the best way to decide. When you do make a decision on which method to use, stick with it for subsequent modeling so the comparisons over time will be valid.

9. Click the **Save** tab.

10. Choose where to write the model output (see **Figure 9**). For this example, use the default `output.sav`. I usually select **Write a new data file** in the **Location** area, and then click **Browse** to name a new file. The only format for this file is the native SPSS Statistics .sav format.

**Figure 9. Saving the output**

11. Click the **Output** tab, as shown in **Figure 10**. This tab controls the output that is displayed in SPSS Statistics Viewer. Selections and changes on this tab do not affect the output file you indicated on the **Save** tab.
Figure 10. The Output tab

12. Click **OK** to run the RFM model. The output data will look like **Figure 11** in the Data Editor after you run the modeling procedure.

Figure 11. The output.sav file in the SPSS Statistics Data Editor window
After the modeling process is complete, SPSS Statistics Viewer displays windows that look like Figure 12, Figure 13, and Figure 14. You must access the output data file separately using the Data Editor window.
Figure 12. Screen in SPSS Statistics Viewer that results from the RFM modeling process (1 of 3)

These scatterplots show the relationships between the three variables used to calculate recency, frequency, and monetary scores.

It's common to see noticeable linear groupings of points on the frequency scale, since frequency often represents a relatively small range of discrete values. For example, if the total number of transactions doesn't exceed 15, then there are only 15 possible frequency values (unless you count fractional transactions), whereas there could be hundreds of possible recency values and thousands of monetary values.

The interpretation of the recency axis depends on the type of recency measure: date or time interval. For dates, points closer to the origin represent dates further in the past. For time intervals, points closer to the origin represent more recent values.
Figure 13. Screen in SPSS Statistics Viewer that results from the RFM modeling process (2 of 3)

The table of bin counts displays the bin distribution for the selected binning method. Each cell represents the number of customers that will be assigned to the RFM model. Although you typically want a fairly even distribution, with all (or most) cells containing a similar number of customers, a certain amount of variance should be expected when using the default binning method that assigns tied values to the same bin. Extreme fluctuations in cell counts and/or many cells with a count of 0 may indicate that you should try another binning method (fewer bins and/or random assignment of ties) or reconsider the suitability of RFM analysis.
Use the charts and graphs in the viewer window to communicate how the model is presenting data to your analysts and business decision-makers. These windows also include basic statistics about the histograms. The histograms show the relative distribution of values for the three fields used to calculate recency, frequency, and monetary scores. It is not unusual for these histograms to indicate somewhat skewed distributions rather than a normal or symmetrical distribution.

The horizontal axis of each histogram is always ordered from low values on the left to high values on the right. With recency, however, the interpretation of the chart depends on the type of recency measure: date or time interval. For dates, the bars on the left represent values further in the past (a less recent date has a lower value than a more recent date). For time intervals, the bars on the left represent more recent values (the smaller the time interval, the more recent the transaction).

Figure 14. Screen in SPSS Statistics Viewer that results from the RFM modeling process (3 of 3)
the mean value of each input variable metric, with standard deviations. Consider making your own graphs and tables that you tailor to your audience, as well.

**Note:** You can save the output.sav file to other formats, and then integrate it into queries and databases to be able to give RFM scores for customers in different applications.

**Repeat frequently**

A single RFM model is a snapshot of your customers' past behavior from today's perspective. Running the model over time and using the results to show how customers move between categories provides a depth that a single result cannot give.

The easiest way to do so is to create a simple data file that stores the RFM scores for each customer by date. Using equally simple queries, you can pull the time series of RFM scores for individual customers and groups of customers over time. To make analysis more accurate, run the RFM model regularly and at equally spaced time intervals. In this way, you will have created a foundation for customer life cycle analysis.

You can use this data to see many things about how your customers' order behavior changes over time. One of the best ways is to combine your analysis with a demographic segmentation to see how different groups move through the RFM scores over time. One insight you may gain is how to identify patterns that indicate when a customer is likely to stop ordering (some people call this churn). Targeting those customers with incentives or extra attention may change their upcoming actions and retain them longer.

**Conclusion**

Using the RFM modeling capabilities within SPSS Statistics is a quick way to get others on board for more analysis. You can use RFM modeling to gain deeper insight into your customers' behavior, whether it is in retail, e-commerce, distribution, or other commercial industries. Even charities can apply this model to improve interactivity with donors.

RFM analysis is, relatively speaking, an easy modeling process to understand. Business users can see the value quickly. Use it to leverage a deeper use of analytics in your organization. It is a great starting point for finding more and interesting ways to bring data mining and predictive analytics into your company.
Related topics

- Learn more about RFM from the blog, *Statistical Concepts and Analytics Explained*.
- Learn more about *SPSS Statistics*.
- Evaluate IBM products in the way that suits you best.

© Copyright IBM Corporation 2012, 2014
Trademarks
(www.ibm.com/developerworks/ibm/trademarks/)