

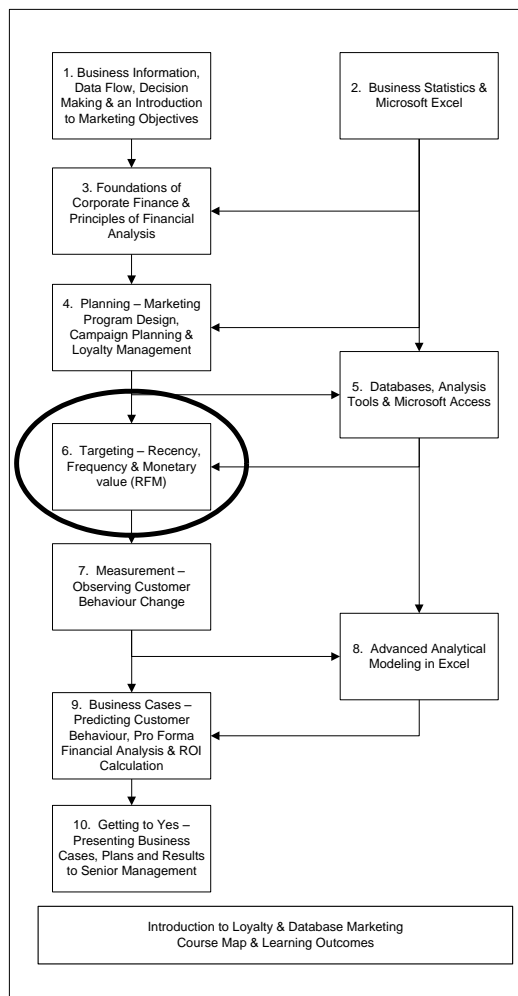
Lesson 6 – Targeting Customers using RFMv Analysis

Mike Sparling – Algonquin College

“I know that half of my advertising dollars are wasted ... I just don't know which half.”¹

Overview of Lesson 6

Welcome to lesson 6, the point where many students feel they first see “the rubber hit the road” so to speak. Starting with this unit we’re going to be getting our hands dirty working with data and marketing program design.



The best marketing message, sent to the wrong prospects, is a sure way to fail and few companies today remain committed to mass market advertising since the results achieved by targeted messages have consistently been proven to produce higher Return On Investment (ROI).

Targeting, as we defined in unit 4, is the science of selecting the appropriate prospects for receipt of a marketing message. The objective, clearly, is to maximize the return generated by each marketing dollar through the selection of prospects with a high propensity to adopt the behavior change found in our message.

As Mr. Wanamaker’s quote above demonstrates, the science of targeting is much sought, by the teams designing marketing programs as well as by the people writing the cheques to pay for the program.

In this lesson we’re going to focus on one of the most powerful targeting techniques used in marketing program design and execution – **RFMv analysis**.

The idea behind RFMv analysis, which stands for **Recency**, **Frequency** and **Monetary value**, is that an existing customer is easier to understand than a new prospect given the behavioral data available and they are easier to

¹ John Wanamaker, considered the father of both the department store and modern advertising. http://en.wikipedia.org/wiki/John_Wanamaker

target since we have an existing relationship.

Depending on the industry, the cost of acquiring a new customer versus expanding the relationship with an existing one can be significant. As a result techniques that derive incremental value from existing relationships, provided the cost of the program doesn't exceed the return generated, are techniques that are readily invested in by the organization.

Learning Objectives

Upon successful completion of this lesson, you will be able to:

- Define behavior-based variables and their role in creating target lists.
- Define and describe what Recency, Frequency and Monetary value measures are.
- Describe how RFMv analysis works, and what outcomes we expect each measurement value to represent.
- Using Microsoft Excel, and a pre-populated data set, perform RFMv analysis.
- Demonstrate the use of RFMv analysis to select a list of target customers that meet the selection criteria for a marketing program.

Process

1. Read the lecture notes in this document.
2. Read pp. 568 – 577 in Berry & Linoff [i].
3. Skim pp. 297 – 358 in Winston [ii].
4. Read the case studies provided, and answer the questions found in the Assignments section of this document.

The Goal of Analytical Marketing – Creating Lasting & Profitable Customer Relationships By Modifying Behaviour

Without customers, businesses wouldn't exist. Simple statement, yet it's probably the number 1, 2 and 3 objective on every executive's mind throughout their career. Marketing plays a key role in identifying, securing and nurturing a steady stream of new and existing customers upon which the firm builds its business. Analytical marketing is one of the ways that this takes place.

Recap the Previous Five Lessons

Recall that in Lesson 1 we defined the core objectives of any for-profit business as being the creation of shareholder value through the distribution of profit generated by improved revenue and decreased operating costs.

We established that there were three ways to increase revenue²:

1. Retention of an existing, profitable, customer relationship.

² While Enron, Worldcom and other companies did pioneer alternatives to these three strategies, they were subsequently found to have limited value in the long term.

2. Capture of more of a customer's spending in our current market segment or in an adjacent segment that we can serve effectively and efficiently.
3. Establishment of new customer relationships which are then subject to items 1 & 2.

Lesson 1 concluded with the definition of the core deliverables of marketing being:

1. The establishment of measurements that value the revenue contribution of a customer to the business.
2. The identification of opportunities to address one or more of the methods of increasing revenue (above) as a result of a competitive advantage held by the corporation or the exploitation of a weakness exhibited by a competitor with market share we covet.
3. The definition of hypothetical programs that would use assets of the corporation to execute against the opportunities identified in #2.
4. Execution of experiments that test the hypothesis, producing revenue expansion for the organization when successful.
5. Learning from the experiments what does or doesn't work.

In Lesson 3 we examined the techniques used to create and execute a marketing program. We identified a program as being composed of a number of campaigns that support the objectives of the program, and we briefly touched on techniques for measurement of program performance relative to expectation. We covered the details of creating a marketing message that is intended to be compelling enough to the targeted customer or prospect that they would modify behavior and exhibit the outcomes we anticipate (generally buying more or different products from our firm, remaining or expanding loyalty or abandoning our competitors in favour of our products and services). Finally we considered, at a high level, the financial implication of delivering our message to the targeted prospects.

The Key Marketing Question

This summary of our previous 5 lessons, when pulled together and put in context of this course, should raise one of many very important questions right now... **How do I find the right people to send my marketing message to? People who are predisposed to acting on the offer I'm making and who will fulfill their side of the agreement in response to my marketing offer?**

The answer to that question is found in the art and science of targeting and for this lesson specifically in behavior-based predictive analysis³ - the use of historic behaviour data to predict future behaviour in response to external stimuli which in this case is a marketing offer.

Understanding Historic Customer Behaviour through Data Analysis

Within the databases of a modern corporation is a wealth of data. Data on sales, proposals, cost-of-goods sold, price charged, payment terms offered, customer payment records and information on

³ The other primary method of targeting comes from an adjacent field of study generally thought of as Indirect Predictive Analysis, where various forms of 3rd party or indirect data is used to construct a profile and segmentation model believed to support the marketing program hypothesis. Sources of indirect data include demographics, psychographics, geographical data & survey results. These types of data are also often used in behavior-based predictive analysis to refine the segmentation so the fields are not that far apart.

competitors. When that data is grouped with supportive / related data and analyzed, it produces insight that can be acted upon. These insights are generally thought of as reflecting a **historic behaviour**.

If you can identify the events that surrounded the observed behaviour, be they geopolitical, financial or related to activities initiated by you, your firm, the customer or a competitor, the behaviour can be analyzed for cause and effect. If you can assemble a series of cause and effect events, either sequentially for a given customer or dimensionally for a set of customers with similar causes and either similar or diverse effects, you are conducting **behaviour-based analysis**.

If in analyzing a sequence of cause and effect events you spot what you judge to be a trend or pattern you believe can be used to achieve a corporate goal, such as a desire to increase revenue or to decrease the stockpile of wide screen LCD TV's presently in inventory, you have made a prediction of future behaviour based on analysis of historic behaviour – also known as **predictive behaviour-based analysis**.

Matching Behaviour Data with Circumstantial Data

Companies don't operate in a vacuum and as a result a pure analysis of historic patterns won't reliably predict the future, yet an absence of appreciation of historic patterns ensures any errors or missed opportunities from the past have a high probability of being repeated in the future. As a result you will most often use historic behaviour data, and findings from historic analysis, in combination with other input values in order to refine or focus the future prediction.

For example the hotel industry frequently uses predictive behaviour-based analysis within their *revenue management* systems to maximize profitability by optimizing hotel room rates. The goal is to identify the optimal price to charge for each room type while still producing maximum occupancy.

Setting the price to high usually leads to unoccupied rooms, as customers generally have choices in destination (more on that in a moment). Setting the price to low will ensure each room is filled but at a reduction in maximum profitability. Since a hotel room has a high fixed cost, occupied or not each room is "responsible" for nightly revenue contribution to pay the mortgage and the salaries of a core of the staff, and a relatively low variable cost for additional staff and resource consumption when the room is occupied, everything above the fixed cost threshold is largely profit.

Understanding the patterns of occupancy is critical to setting the optimal room rate. Knowing that in general hotel rooms located near business offices attract more occupants during the week rather than the weekend while a hotel near Disney World is probably busier on weekends and school holidays is a key finding. Noticing that two weeks out of 52 have inverse occupancy patterns for a reason you can't identify is equally important, since these periods probably warrant a reduced rate to try and raise occupancy. Understanding why these anomalies exist needs to be a key goal of the analysis cycle.

Possible questions to ask that might explain these anomalies could include:

- Was there a convention of relevance to the bulk of the nearby businesses that drew the usual occupants elsewhere?

- Was there an attraction in the area that drew more customers to the businesses and thus your rooms?
- Did a competitor send out an offer or conduct a campaign designed to draw customers from your business to theirs?
- Did you, for a brief period of time, have a particularly rude employee answering the phone and alienating prospective customers from making reservations?

Any of these or an unlimited number of alternatives exist to help explain historic performance. You usually won't find this type of data in the pure analysis of customer behaviour, so you have to go beyond just the basic data to understand the context and perspective that surround the data. Keep in mind the fact that understanding these factors is often the difference between making future predictions subjective or speculative.

With an understanding of the forces that modified historic behaviour, we can also see the role these inputs may provide to future prediction.

Modify Strategy Based on Behaviours Recognized

Keeping with the hotel example, with greater understanding of the environment the customer behaviour data was influenced by, and by gathering information about upcoming events, the hotel may find that:

- A conference being held at the hotel is clearly an opportunity to raise rates.
- A conference at a nearby competitor may offer opportunity to lower rates and steal customers from a competitor.
- Knowing the holiday patterns of different school districts helps the hotel near Disney World maximize rates against changing dates.
- Recognizing the geographical areas where the bulk of customers come from ensures focusing on their holiday periods.

Predicting the future by learning from the past is a core technique in any business. Understanding the environment that generated the visible historic behaviour and analyzing to understand the root cause is critical to ensuring future prediction is as accurate as possible. Finally matching up an understanding of past behaviour with insight into future events that may modify expected outcomes raises the probability of a successful prediction.

In the next few sections we'll look at one of the foundational techniques used to identify prospects for a campaign or offering based on historic behaviour analysis.

Past Behaviours Tend to Repeat

We are all creatures of habit. The things we've done recently, we will probably redo in response to similar stimuli. The things we do consistently, it can be reasonably expected, we will continue to do and if we've exhibited a tendency towards a behaviour we are highly probable to increase the frequency of that behaviour if we have reason to do so.

That we are creatures of habit is the basis upon which RFMv analysis is built. If we can identify the customers who have exhibited one or more behaviours in the past, we should attempt to incent the customers we presently have to respond since generally that will be cheaper, and hopefully more profitable as a result, than try to gain new customers and modifying their behaviour to meet our expected outcomes.

This is not a fool proof proposition and any good marketing campaign not exclusively focused on new customers will have, in addition to pursuing existing customers, initiatives and offers restricted to new customers since we will need to replace those existing customers who are either not sufficiently valuable to warrant extra attention or who attrite for any number of reasons.

What we need to identify is a technique for looking at our existing customers and finding the ones we want to work with to generate continued or expanded value. There are many techniques for doing customer analysis but one of the longest serving is RFMv analysis.

Recency, Frequency and Monetary value (RFMv) Analysis

RFMv analysis is a technique used to group or segment existing customers based on historic behaviour in the hopes that history can, with the right motivators, be caused to repeat or even improve upon its self.

The acronym is short for Recency, Frequency and Monetary value and each of these measures aligns to one or more of the three methods of increasing revenue for a business.

Recency (R) – When did the customer last place an order, visit our store or interact with us in a material way? A customer who recently had a favourable interaction with our firm is, we hope, predisposed to repeating that interaction and thus susceptible to an offer that would encourage future business. Similarly a customer who hasn't done business with us for sometime may be open to an offer of resumption that draws them back.

Frequency (F) – How many interactions, over a period of time, has the customer had with us? Assuming the interactions have been favourable for both parties, we would hope that we can sustain or increase the frequency of the interactions to our advantage. As with a customer who has not done business with us recently, frequency of interaction is a trigger you will want to pay attention to when it falls off over a period of time. This is where the frequency measure is often correlated to the recency one.

Monetary value (M or Mv) – Over a given period of time, or number of interactions, what is the value of the customers business either in terms of revenue or profitability. Grouped in with monetary analysis is often inventory and channel analysis to get a sense of customers whose purchases reflect higher margin activities for the business such as buying large volumes through automated channels or the purchase of inventory items that have higher margins, are slow moving in various periods or are ends or remnants of other jobs.

Each measure on its own has immense value to understanding the historic patterns of a customer or group of customers, but it's when you combine the measures together with each other and / or add in additional insights about the environment within which the behaviour is observed that the true power comes through.

We've all experienced RFMv in action. Some examples include:

- Catalog and direct-mail marketers were early adopters of RFMv techniques to determine which customers got which catalogs, how often and with what special incentives, coupons or savings. With the advent of high capacity colour digital presses, many companies now custom print each catalog, varying the items on pages, prices for items and even specialized promotional offers for each customer based on the findings of RFMv analysis.
- RFMv analysis forms the basis of every customer loyalty program in operation from frequent flyer or hotel guest programs to retail shopper reward cards.
- If you've ever been to a casino you've seen RFM analysis combined with life-time value analysis⁴. These are the principles upon which casinos issues complementary (aka comp) hotel rooms, meals, show tickets and everything else they offer "for free" to patrons of their establishments. Even the so called "free drinks" you can get in Las Vegas casinos are carefully distributed based on a real time size-up of your value to the casino based on RFMv analysis.

In the next section we'll examine in depth how RFMv scores are created.

Calculating an RFMv Score – The Basic Approach

The first step in calculating an RFMv score is to agree on what the measures represent to our business and how to segment the values within each measure. Doing so will lead to three primary, and possibly a number of secondary, scores with different meanings and interpretations.

One basic approach to calculating scores is to use quintiles, five segments or buckets, to break the data into. Five is a number used often because it means each bucket will contain 20% of the values of a given measure provided we use linear distribution⁵. It's always easier to set up and use a system with round numbers.

For example, if you had a list of customers who have bought books from a vendor, you could sort them based on date of last purchase and then place the top 20% in segment 5, the next 20% in segment 4 and so on to produce a Recency score for these customers. Because we went with a linear distribution, we'll have an even distribution across the five quintiles provided the number of customers is evenly divisible. If we have an odd customer or two they usually wind up in the last segment for simplicity.

⁴ Life-time value analysis looks at how much financial contribution we believe the exhibited behaviours of the customer, with allowance for improvement based on our campaigns, marketing programs and other interventions, will be worth to us over the projected life of the relationship the company expects to have with the customer.

⁵ Linear distribution means we evenly distribute the values. Other distributions may place artificial boundaries in place to observe different groupings and as such the distribution usually isn't even across the segments. Linear is the easiest approach.

You'd follow the same process for Frequency, using a value that represents the total number of purchases or interactions across a period of time (more on time periods and sources of data in a moment), and finally you'd use a value that represents the total value of each of those purchases or interactions over the same time period to create a Monetary value score. The end result is a basic RFMv score made up of three numbers that have broken our customers down into three sets of five groups each where each of the groups represent 20 % of the customers within that segment and for that measure.

What do those three numbers, which represent the customers ranking on all three measurement axes, mean when we look at them?

It's possible there are customers who are 5-5-5 across the board and it's possible that there's some who are 1-1-1. It's also possible no customers from the data set fit every possible combination of values since there are 125 unique combinations of scores. With a large enough sample set you'll eventually see a nearly normal distribution but it's not typical to process that much data⁶.

For now it's worth eyeballing the numbers for things that stand out. You'll find as you grow in confidence with your analysis skills that eyeballing the data is often where some of your best insights come from.

Creating a Summary Score

To get yet another perspective on relative ranking of customers, we should create a **summary score**.

To create a summary score, we add each of the R, the F and Mv scores together to get a number between 3 and 15. This summary score is the ranking across all three measurement axes of the "customer value" and while it's a valuable number, it's also deceptive in its simplicity as we'll see later in the lesson. For now recognize that you won't always have an even or full distribution of all possible values between 3 and 15 unless the data set you are working with is sufficiently large as to contain normal distribution.

What Do All These Scores Mean?

A customer who had a score of 5-5-5, and a summary score of 15, would be one who was top 20% of customers who had purchased something recently⁷, is someone who falls in the top 20% of number of purchases and is also one of our top 20% in spenders. If you see a 5-5-5 or a 15 summary score, they're probably a pretty important customer to your business, and you probably want to answer the phone when they call!

It is important to put a couple caveats on statements of customer worth based only on RFMv analysis. First the ranking is within the data we examined – the **sample set** – which may not represent all our customers. You'll often work with a subset of all data for cost and efficiency reasons. Second the data

⁶ Lesson 2 Basic Statistics & Microsoft Excel covered normal distribution and probability.

⁷ Keeping in mind this ranking is of customers within our data set, not necessarily from our total customer base.

for Frequency and Monetary value are **derived values**⁸ rather than **fact values** and as such they have both a derivational algorithm that may or may not impact the overall worth of the measure and they have a time boundary that is important to consider.

Completing the Analysis Model

You need to document, along with your findings, the assumptions, boundaries and derivational methods used to create non-fact values. If you don't document these values, much of your work can later be questioned or even found to be inaccurate. The conveyance of information and supporting reference material is such an important topic that we will cover it in depth in lessons 8, 9 and again in lesson 10. For now just be sure you document the values somewhere in your model, in a journal or on the sheet of paper you're doing the analysis on.

As you calculate out the scores, you'll also generate some additional data of importance namely the boundary values for each of the segments. Once you've filled a segment, say the first segment (number 5) of the Recency axes, the first value in the segment is the most recent transaction and it forms the **segment upper boundary** while the last value in the segment represents the **segment lower boundary**. These values let you size-up the segments and as you'll see later in this section you may adjust the boundaries and thus skew the linear distribution to gain greater insight from the segment contents. For now simply label the segments with the upper and lower boundaries.

Finally you need to consider where these score values, the assumptions, source data set boundaries, derivational algorithms and segment boundary values should be stored. The simple answer is "where it makes most sense given your tools and data".

When working with a database you'll probably add columns to an existing table, or create a new table, to hold the scores. If you are doing extensive analysis within the database, you may consider a more logical structure to hold the results so multiple experiments can be conducted against the base data while retaining the results for future analysis and reporting.⁹

In Microsoft Excel you'd add columns to each row of data¹⁰ and if you were working on a sheet of paper you could jot the values in the margin or footer of the page.

This represents the basic RFMv scoring techniques. Ready to give it a try?

⁸ Derived values are values that are calculated or inferred from other values, usually from simple facts. If you take all the purchases a customer made over a 12 month period and added those purchases up, you've derived from the base data a value for Monetary value. It's as accurate as the input data, but if the day before or the day after the dates that bound the 12 month period the customer made the largest purchase ever in company history, that fact will be missed in the derivation.

⁹ The meaning of and method to achieve this was covered in Lesson 5 – Databases, Analysis Tools & Microsoft Access.

¹⁰ The meaning of and method to achieve this was covered in Lesson 2 – Business Statistics & Microsoft Excel.

Widget Manufacturing - An Example

Included with this lesson is an Excel spreadsheet entitled “Lesson6Example1.XLS”. The data for the following example can be found within.

The Widget Manufacturing Company Inc. provided the data in the spreadsheet to you for analysis. It is a subset of data taken from their sales system looking at the period of 12 months prior to today. The database support team did a lot of work for us¹¹, including summing data and creating three columns of derived content.

- Using today’s date as the point of origin, they’ve used the last date of a purchase transaction and identified how many months ago it took place, with that value going into the **Months Since Last Purchase** column. They’ve culled the data set at customers whose last purchase was more than 12 months ago¹².
- Having identified customers who fall within the 12 month purchase horizon, they’ve counted the number of unique sales in that period for this customer and placed that value in the column labeled **Num. Purchases Last 12 Months**.
- Finally they’ve taken each of those purchase transactions and summed the dollar value of each to produce the value in the **Dollar Value Purchases Last 12 Months** column¹³.

A sample of the data set is shown below.

¹¹ The work done against the database was the topic of Lesson 5 – Databases, Analysis Tools & Microsoft Access

¹² It’s important to note that you can’t assume this to be true. It’s possible the data wasn’t limited to customers with purchase in the last 12 months and an assumption it was could produce invalid results. **Remember, when in doubt always confirm don’t assume.**

¹³ When working with real data, the value of purchases column usually shows the widest variance in interpretation and data source. Looking at total value of purchases is one thing, looking at net or gross margin generated is another. Each is an important measure and ensuring you know the source and meaning of the data in this column is important to your analysis.

	A	B	C	D
1	Customer ID	Months Since Last Purchase	Num. Purchases Last 12 Months	Dollar Value Purchases Last 12 Months
2	52029	7	3	\$ 72
3	93179	2	16	\$ 112
4	99162	1	11	\$ 66
5	89691	9	3	\$ 9
6	71701	12	16	\$ 496
7	26601	11	6	\$ 54
8	82051	0	14	\$ 266
9	12993	6	16	\$ 48
10	41407	9	15	\$ 525
11	83511	11	9	\$ 36
12	60508	7	6	\$ 192
13	71210	6	7	\$ 175
14	68130	10	15	\$ 375
15	66762	7	14	\$ 182
16	64402	9	11	\$ 341
17	12257	10	5	\$ 95

We've conveniently been given the raw data required for RFMv analysis, so let's get to work. The first thing to do is to add some columns at the end of the spreadsheet to house the RFMv values. We'll creatively label three columns R, F and Mv respectively.

	A	B	C	D	E	F	G	H
1	Customer ID	Months Since Last Purchase	Num. Purchases Last 12 Months	Dollar Value Purchases Last 12 Months		R	F	Mv
2	52029	7	3	\$ 72				
3	93179	2	16	\$ 112				
4	99162	1	11	\$ 66				
5	89691	9	3	\$ 9				
6	71701	12	16	\$ 496				

Next we need to create our segments and sort the data. We'll follow the recommendation to use quintile segmentation, so we'll respectively sort the data by **Months Since Last Purchase** to find the **R value**, then sort by **Num. Purchases Last 12 Months** to find the **F value** and lastly sort on the **Dollar Value Purchases Last 12 Months** column to find the **Mv value**.

A quick check of the data shows there are 40 customers in this sample set and so each quintile will contain 8 customers (40 customers / 5 segments = 8 customers / segment). Once you've sorted the data by the appropriate column, put a 5 in the respective score column / row for the first 8 customers, a 4 for the next 8 and so on.

After working through the sorts and assigning the scoring values, the first few rows of data would look like this.

	A	B	C	D	E	F	G	H
1	Customer ID	Months Since Last Purchase	Num. Purchases Last 12 Months	Dollar Value Purchases Last 12 Months		R	F	Mv
2	41407	9	15	\$ 525		4	5	5
3	71701	12	16	\$ 496		5	5	5
4	53959	9	15	\$ 435		3	5	5
5	68130	10	15	\$ 375		4	5	5
6	39209	3	12	\$ 360		1	4	5
7	64402	9	11	\$ 341		4	4	5
8	51355	7	10	\$ 320		2	4	5
9	89412	8	11	\$ 319		3	4	5

This is after the third sort, on the **Dollar Value Purchases Last 12 Months** column, so if your sort order is not on this column, you'll possibly have different data. The full set of values is in the example spreadsheet on tab "RFMv2". You should have the same values if you've sorted and segmented correctly.

Now lets go ahead and add a summary score column to the right of the raw Mv score. Use the SUM() feature in Excel as we covered in Lesson 2, then sort the data on the RFMv summary score column. The first few rows of data should look like this when you're done, and the tab "RFMv3" contains the full data set.

	A	B	C	D	E	F	G	H	I	J
1	Customer ID	Months Since Last Purchase	Num. Purchases Last 12 Months	Dollar Value Purchases Last 12 Months		R	F	Mv		RFMv Summary Score
2	71701	12	16	\$ 496		5	5	5		15
3	41407	9	15	\$ 525		4	5	5		14
4	68130	10	15	\$ 375		4	5	5		14
5	53959	9	15	\$ 435		3	5	5		13
6	64402	9	11	\$ 341		4	4	5		13
7	97457	11	11	\$ 308		5	4	4		13
8	77364	12	15	\$ 150		5	5	3		13
9	89412	8	11	\$ 319		3	4	5		12
10	35631	11	6	\$ 180		5	3	4		12
11	51355	7	10	\$ 320		2	4	5		11
12	66762	7	14	\$ 182		2	5	4		11
13	85368	10	7	\$ 168		4	3	4		11
14	74752	11	13	\$ 52		5	4	2		11
15	39209	3	12	\$ 360		1	4	5		10
16	26601	11	6	\$ 54		5	3	2		10

With this analysis, that should have taken no more than 15 or 20 minutes to produce, you've gained some insight into the customers who've done business with Widget Manufacturing Company Inc. and we can make some predictions about future behaviour based on this analysis.

We can see that we've got some pretty big scores in the first dozen or so rows. We've got one customer way down at the bottom with a summary score of 3. We've also got a reasonably distributed scoring across all 40 customers.

Making Business Decisions From Analysis and Insights

Lets take a look at what sorts of decisions Widget could make with this analysis.

If Widget Manufacturing Company Inc. was holding a golf tournament this weekend, and wanted to invite their best 5 customers to play, which 5 would you recommend they invite?

That's a bit of a trick question actually because although we've got the raw R, F & Mv values as well as a summary score, we don't immediately know what value or values combine to represent a "best customer" to Widget. One assumption would be to go with the summary score and invite customers 71701, 41407, 68130, 53959 and 64402, and that would be a safe assumption but Widget may elect to define "best customer" in different ways. They could define "best customer" as:

- The one who buys most often, so we'd sort only on the F value column.
- The one who spends the most money, sort on the Mv column.
- Or the one who generates the most profit, which we can't really calculate from the data provided. You could take a guess by assuming that fewer transactions, a lower F score, with highest Mv score may have been less costly to have serviced, but that's a complete guess.

This illustrates the fact that analysis is an imprecise science and you have to be careful not to jump to conclusion you can't support with the available data.

For bonus points, clean up your model so you could present it to the management team at Widget. Ensure the assumptions and the conditions under which your analysis findings were generated are understood by a reader and improve the general appearance of the model. One example of a cleaned up version is on tab "RFMv4". We'll cover model building and presentation skills in future lessons but for now remember that you want your work to be capable of standing on its own without your narrative to go with it.

Bonus Question: Using the RFMv scores just calculated, and the Microsoft Excel sorting and pivot table skills developed in Lesson 2, create a spreadsheet model that simplifies looking at the RFMv scores in different ways to answer questions about customers, starting with the different ways Widget could determine who to invite to golf.

Applying and Enhancing RFMv Scores

The basic approach can yield some strong insight about customers, and clearly a customer with a 15 summary score is probably pretty valuable to the firm. With this analysis technique alone you can produce a wealth of insight and make some very informed decisions about future marketing offers.

To take things even further, consider the following additional categories of enhancements, refinements and improvements on our analysis techniques:

- We can **change the number of segments for a given measure**. For example we may only have 90 days of operating data for our customers and splitting that time period into 5 segments may make less sense where 3 might make more. Narrowing the number of segments analyzed can produce meaningful insight that a broader segmentation may not.
- We can **force fit the data to a prescribed set of segments within one or more measures**. If we have 90 days of operating data, and we ran a newspaper advertisement for the first 30 days, placed signs outside our store the next 30 days and called all our customers with a personal appeal the last 30 days, then establishing the Recency segment as 0 – 30, 31 – 60 & 61 – 90 days will allow us to evaluate responses to our different message channels¹⁴.
- If you have a lot of data, or you want to gain insight into a particular direction, a technique that is often used in commercial software packages is to “overload” the segmentation by breaking each primary segment – something that gets a number in the 5-5-5 scheme – into a number of sub-segments by doing an RFMv or part there of analysis within a segment. For example once you’ve identified the top 20% of customers in a Recency quintile, take only that data segment and perform either a full RFMv analysis or just analyze for FMv or even for just a single measure. Doing this level of drill in analysis will show additional patterns and draw out specialized insights that a very large data sample may gloss over when viewed with only a single RFMv analysis pass.
- In addition to any of the techniques above, you may want to work within a subset of the primary segments and replace some data with other data that augments the analysis value. For example the data used to generate the primary value for total spending in the period may be gross sales data, but within certain or all segments you may want to sub-segment using profitability values to understand the absolute value of a customer to your business once you’ve identified them as being of interest. A casino may want to identify customers within a “big spender” category who truly are the core contributors of value to the bottom line, while a manufacturer may want to examine a number of perceived lower value segments to ensure there are no “diamonds in the rough” that could be made more valuable with the right attention or who warrant not being included in a customer list clean-up process.

Once you get working with RFMv analysis, and experimenting with how you segment, how many segments you create and most interesting to the outcomes you may produce, how you derive and populate your data sets, you’ll come to appreciate the analysis strength in what is really a fairly simple technique to master.

¹⁴ This is slightly misleading in that the segmentation proposed assumes immediate take up and cessation of buying based on a message channel. Clearly not how it would work in practice and in later lessons we’ll look at the difficulty faced in tracking customer behavior back to root cause when multiple channels for messages exist. For now accept the basic idea here as illustration rather than it being completely accurate.

As we'll see later in this course another strength of RFMv analysis is that you're using little in the way of statistical techniques and the input data, analysis activity and most importantly the outcomes and recommendations are readily understood by business people.

Like anything worth learning, mastering RFMv analysis requires practice. The assignments for this lesson will give you a number of opportunities to try the basic and advanced techniques we've just covered.

Assignments

Case: Industrial Hose Supply Group Ltd.

Read the case as distributed, and using the provided Microsoft Excel spreadsheet, build a model that answers the following questions.

1. Using the spreadsheet entitled "IndustrialHose.XLS" calculate the raw values for Recency, Frequency & Monetary value, then calculate an RFMv sum score. Who are the 20 best customers? The 20 worst customers?
2. Recall the definitions of Revenue and Gross Profit from lesson 1, using the spreadsheet tab entitled "CostPrice", the tab entitled "SalesJournal" and the assumptions listed below, calculate the Gross Profit of each of the top 20 and bottom 20 customers. Can you make a case for changing the score for any customers? (Hint - you will need to use skills from previous modules to complete this analysis including **step function accounting and analysis** from lesson 2 and **sort and sum techniques in Excel** from lesson 4)
 - i. Assume all customer sales are inbound calls (ie they call our sales desk, so the **cost of sales** for each customer transaction is the same)
 - ii. Assume that there is a flat cost to assemble and package each order regardless of the number of items in the order
 - iii. Assume there is a flat cost to locate, cut to length and restock **each product**
 - iv. Assume there is never a shortage of inventory and that every order was filled completely each time (nirvana for any business owner!)
3. Consider the following additional information - All hose products come on 25 foot spools and the minimum length of purchase is 5 feet. Using the top 20 customers based on their RFM sum score and the analysis from the last question are any of these customers less appealing to us?

The next set of assignments introduce the main case studies we will use the remainder of this course. There will be material in the cases that will be used for future lessons so don't become concerned by facts that don't fit your current modeling and analysis.

Case: Skinny Joes Coffee House

Read the case material as distributed. Using the provided data, and as a consultant to Kirby McInnis the owner, prepare an answer to the following questions:

1. Using the data gathered from the "11th cup free" promotion cards, who are the most frequent customers (using RFM analysis)?
2. In your opinion, is there any value in understanding the Recency aspect of RFM analysis given Kirby's business?
3. What additional information not present in the case would you need to conduct a monetary analysis (using RFM analysis)? Based on the information Kirby presents in the case, and your experience with coffee shops in general, how difficult would it be to get the information you feel is unavailable?
4. Can you think of an example recently where a national or international coffee chain has taken steps to gather this type of customer tracking data for analysis? (Hint - Take a trip to a Starbucks or Tim Hortons and look at the signage and devices located around the Point Of Sale)

Case: Digital Cable TV Company - A New Product Launch

Read the case material as distributed. Using the provided data, and in the role of Steve Smith VP of Consumer Marketing, prepare answers to the following questions:

1. Who are DCTV's best customers from an RFM perspective?

2. Are all three aspects of the RFM analysis relevant to the business model of DCTV?
3. If you change any of the "classic definitions" of R F or M in the analysis to measure something equivalent, would it change your outcomes? Do you have the data available to you to conduct this revised analysis?

References

- i. Michael Berry & Gordon Linoff (2004). *Data Mining Techniques*. Indianapolis: Wiley Publishing Inc.
- ii. Winston, W. (2007). Microsoft Office Excel 2007 *Data Analysis and Business Modeling*. Seattle: Microsoft Publishing Ltd.