Marketing Analytics
Professor Swagato Chatterjee
Vinod Gupta School of Management
Indian Institute of Technology, Kharagpur
Lecture 40
RFM and Market Basket Analysis

(Refer Slide Time: 00:28)

Hello, everybody welcome to marketing analytics course. This is Doctor Swagato Chatterjee from VGSOM IIT Kharagpur who is taking this course, we are in week 8 and in this week we will be discussing about Recency Frequency Monetary and Market Basket Analysis. So, we are actually continuing our discussion on retail analytics and that is for these two things are mostly used.

These are two different concepts, probably a little bit older concepts, but still heavily used in in today's world also at least in the when we do simple analytics we when we do not focus on a lot on predictive and etc. When we keep our try to keep our lives simple but still meaningful or fast in those kind of situations we generally use RFM and Market Basket Analysis.

So, initially on this particular week first we will discuss about RFM Recency Frequency Monetary Analysis and the later part of this week that means, in the later videos we will talk about the Market Basket Analysis. So, first the Recency Frequency Monetary Analysis. Now, this is a analysis technique, which is used generally to as a as a tool to segment customer.

So, before we discuss this I have in a previous video I have discussed that there are various types of segmentation that you can do for a customer that those can be like demographic,

geographic, attitudinal or psychographic and then behaviour. So, behavioural segmentation is majorly focused on person's behaviour, how they are expressing themselves, how in the behaviour that they are doing or other places.

So, now here in marketing analytics, what we will be doing is, in the context of retail mainly or in the context of many other service business, we will focus on these behavioural analytics and behavioural segmentation and we will actually focus on a specific part of behavioural segmentation where people are giving they are purchased behaviour as the variable and that is the variable that we are using for further analysis and segmentation of the customers.

Now, why that is needed? Why do we have to segment into segment our customers based on their purchase behaviour? Because purchase behaviours is the most I would say, a behaviour which consumers cannot specifically hide that is the most crude raw version of their behaviour that you can see.

You might not be able to see from the data that you have that whether the consumer has seen your advertisement or what is the preference pattern of the consumers, or which kind of movies you watch, which kind of the probably which kind of websites he visit and from where the traffic is coming, a little bit of vague idea you can get about the customer from the data that you generally collect.

That very concrete, very strong behaviour that is available with you is from the scanner data, scanner data means the data that you can collect based on the customer purchases, for example, in a retail store, when you check out, let us say I bought 4, 5 products and you are planning to checking out right now, so you went to the billing counter and the billing counter guy scans your products one by one. So, whatever data has been created from that kind of scanner is called scanner data.

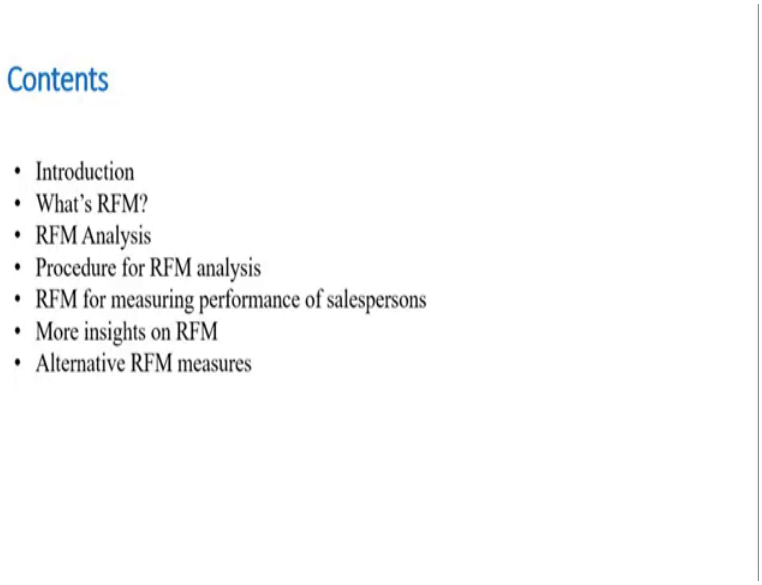
So, that is basically one individual's purchase data you do not have anything else about that individual. Probably while he registered for the card, the loyalty card, you might have a little bit of more idea his name, his email id’s, phone number, his address, his family members name family members number at max.

Some basic details you have. But that purchase data is actually the data that you can use further to create an idea or 360 degree view about the customer from the purchase point of view. And that is why that that behavioural the purchase behaviour because very important factor.

Now, next decision point that comes into this picture that what in this purchase behaviour will I focus on? So, if I if I have customers purchase history, what exactly will I focus on? Will I focus on how much groceries he bought? How much apparels he bought? How much FMCG products he bought? Or whether I will focus on how much is the total monetary value of their basket? Or will I focus on how many units, how many different types of skews he brought? What exactly will I focus on?

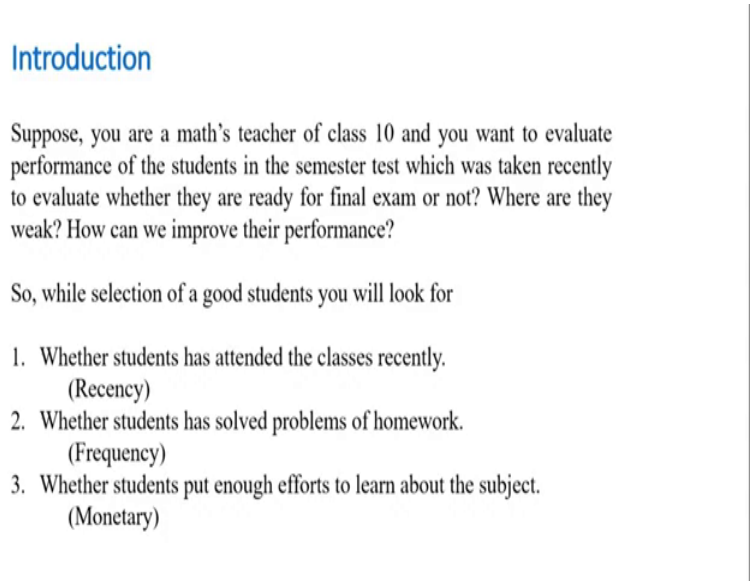
So, researchers went ahead and did a lot of research and found out that there are certain aspects of their purchase behaviour, lots of characteristics that can be created from the purchase behaviour, but there are certain aspects which better predict the profitability of the customer than other aspects. So, those are the aspects that primarily they found out and put them, these estimates are called Recency, R stands for Recency, F stands for Frequency and M stands for Monetary, this three things and I will explain what these things are.

(Refer Slide Time: 05:54)



So, my presentation will have an introduction, what is RFM? RFM analysis how it is done. Procedure and then RFM for measuring performance of salesperson, so we will actually show that for two sales person how you can use RFM to know what is the performance of that particular person. Then there are certain insights of RFM, we will talk about how? Why in the industry RFM is important and etc.? And at the last we will talk about the alternative RFM measures. So, these are some of the details that we will be covering in this current particular presentation.

(Refer Slide Time: 06:34)



Now, introduction. So, suppose you are a maths teacher of class 10 and you want to evaluate performance of the students in the semester test which was taken recently to evaluate whether they are ready for final exam or not? So, you want to know that whether your students are ready for the final exam or not, so you have taken a test. So, this is something that we all probably have given at various points of time.

Now, the question is, the next thing is, why are they weak? This is something that you also want to know that in which aspects they are weak and which aspects they are strong? And you also want to know, how you can improve their performance? So, these are the major goals that you have, so were they are weak, were they are strong, who is weak, who is strong, and how you can improve their performance.

So, keeping that as the goal you are trying to do something. So, what will you do? So, while selecting the good students, if you try to select the good students whom you will send it to some Olympiad exam let us say and the bad students whom you should focus on a mode for their educational purpose and etc. Some of the things that you come that comes to in your mind is probably if you are a professor or a or a teacher.

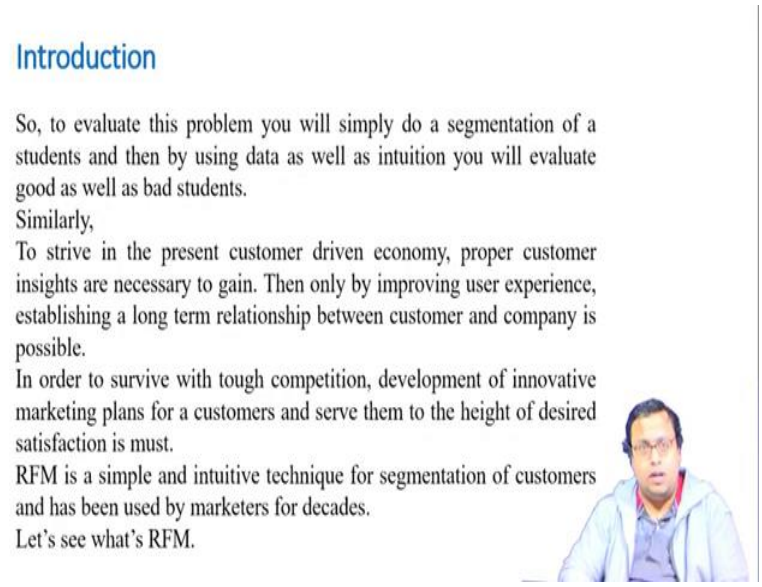
Then whether students has attended the classes recently. So, how recently they have attended your classes, if they very recently they have attended your classes then they are more more accustomed to you or probably they are more attracted to you, or there are you means so teaching and then they are they are probably more receptive of your whatever teaching ideas or whatever this thing will give, they will be more deceptive. So, that is something which is

recency, recency stands for the recent, it comes from recent, how recent it is, so students have attended the classes recently.

Then the second aspect that becomes important is, whether students have solve problems of homework. So, the homework that you have given, in what frequency not whether actually what frequency they have solved, how many of your class tests, how many of your quizzes, how many of your problems that you have given to them they have worked on that, so that stands for frequency.

And the last one is monetary, where students put enough efforts to learn about the subject, so whether they are spending enough amount of time, so that is also related to monetary. So, there are three aspects that is why, one is decency, one is frequency and one is monetary, which can be which can be told in the context of a student giving exam and etc and like this, but we will actually focus on extend this kind of an idea in a in a in a better form as we go ahead.

(Refer Slide Time: 09:39)



So, to evaluate this problem you will simply do a segmentation of students and then by using data as well as intuition you will evaluate good as well as bad students, so this is something that you are going to do. Now, this is the same method that a marketing manager will use for the customer.

Similarly, to strive in the present customer driven economy, so now the economy is driven by customers, proper customer insights are necessary to gain. So, you have to know your who

your customer is? What kind of things they like? There is you have to gain insights about the customer.

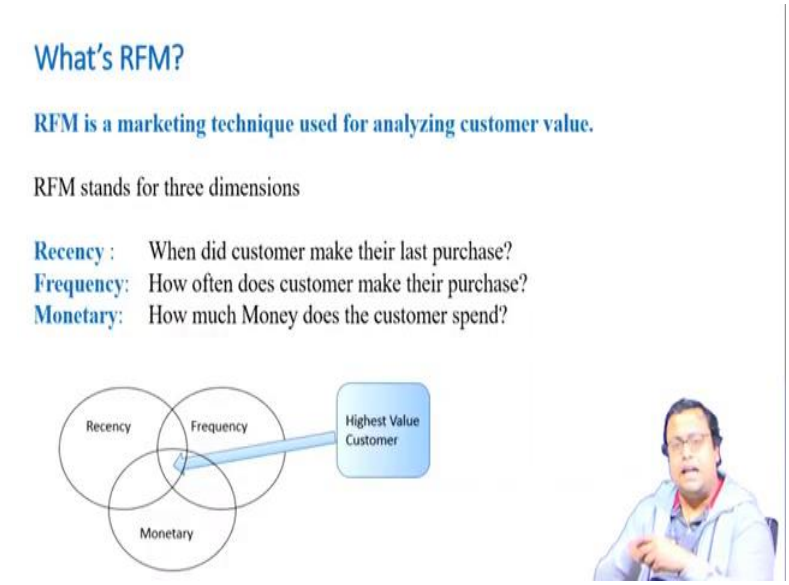
Then only by improving user experience establishing a long term relationship between customer and company is possible. So, the only way is to create a very good user experience, very good product experience and etc. So, in order to survive with tough competition, development of innovative marketing plans for a customer and solve them to the highest the height of desired satisfaction is must.

So, to do all of this thing RFM is required. So, that is the basic premise for RFM is defined. So, if to make your customers happy, you have to know your customers and all of these things you cannot do for all your customers, so you should focus on those customers who are more profitable as simple as that.

So, RFM is a simple and intuitive technique for segmentation of customers and has been used by marketing managers for decades. So, let us see what is RFM. So, this is a segmentation tactic because you cannot handle everybody you are actually creating a segmentation tactic while you are breaking the customers into multiple groups and you are saying that okay, I will only focus on the on the groups which are more I would say profit making.

And I will not focus on those groups which are less profit making and this is how you are making sure that the more profit making guys you are making them more happy, they are having long term relationship with you and by keeping that long term relationship with you, you are making more profit, so as simple as that, but basically RFM is used for segmentation.

(Refer Slide Time: 11:59)



What is RFM? RFM is a marketing technique used to analyze customer value and there are three things, RFM stands for as I told Recency, Frequency and Monetary. Recency stands for, when did customers make their last purchase. Frequency is, how often customers make their purchase? That means a customer who buys quite frequently even if small, small amounts but if he buys quite frequently then that is something that is very important?

Small example remember, so you get push notifications in Ola, Uber and etc. When will you get this push notifications? You remember, all the coupons and offers and push notifications comes to you when the two cases when it comes to you, let us say if you had a new customer, and you have just used now, for the next for further usage they will put, so that is a recency. So, you have recently used, they are putting some offers so that you further use.

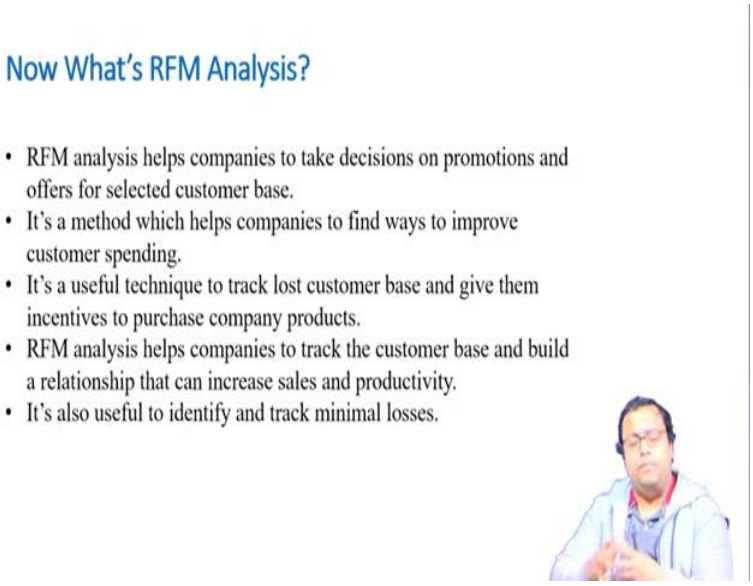
Now, let us see, when will they send offers irrespective of your recency that means even if you have bought wide lot a time back, let us say one year back you have used, still they will send you an offer, when? When your monetary component or frequency component were high. So, let us say in the last year, let us say in 2018 or 2019 early you are spending quite a lot of money on Ola, so you were you were actually commuting on Ola and then you stopped commuting.

Now, this guy in December 2019 or January 2020 or early 2020 this guy will actually identify that this guy, last year he was one year back he was using now he was not he was using, how will he identify? He will identify based on frequency and monetary, so how much money you

have spent on Ola, how much money does the customer spend on Ola in and how many how often does customer make the purchase, which is frequency.

So, when all these three things happen, that is the sweet spot where the highest value of customer. So, the more recent, more frequent and more monetary spending monetary expenditure is high that is the sweet spot that every company is trying to get, so that is something that is there.

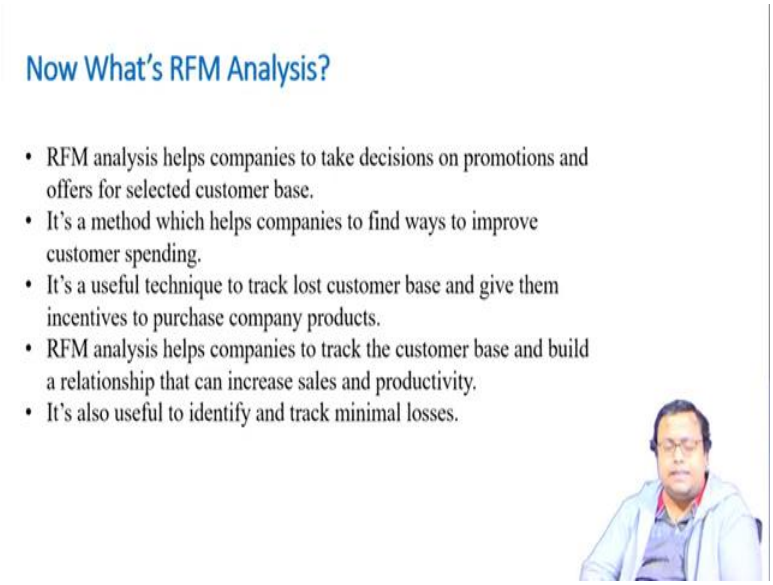
(Refer Slide Time: 14:26)



Now, what is RFM analysis? So, I understood this is RFM the meaning of recency, frequency and monetary, but what is the analysis? RFM analysis helps companies to take decisions on promotions and offers for selected customer base, so you do not send a email which is you sent targeted emails, you do not spend send emails to everybody, you do not do mass communication to everybody.

Because if you do mass communication, the cost is much much higher and the (())(14:58) conversion ratio is much smaller than when you do very targeted marketing. But targeted marketing is not so easy, you have to gain lots of data, so that you know whom to target and then you have to have analysis technique to find out some insights from the data, RFM analysis is one such insight generation technique.

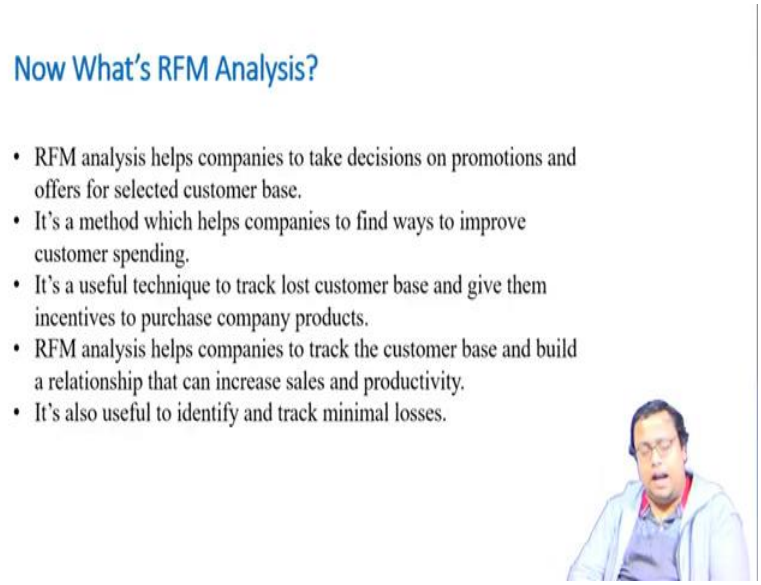
(Refer Slide Time: 15:22)



It is a method, which helps companies to find ways to improve customer spending. So, you know, so you have to increase your ROI of customer spending, and RFM will help you. It is a useful technique to track lost customer base, so whoever is lost, whoever is lower in RFM score, you can also track them and give them incentives to purchase company products. Just like I told that last year you were using Ola, last year many people were using Ola who have stopped using now.

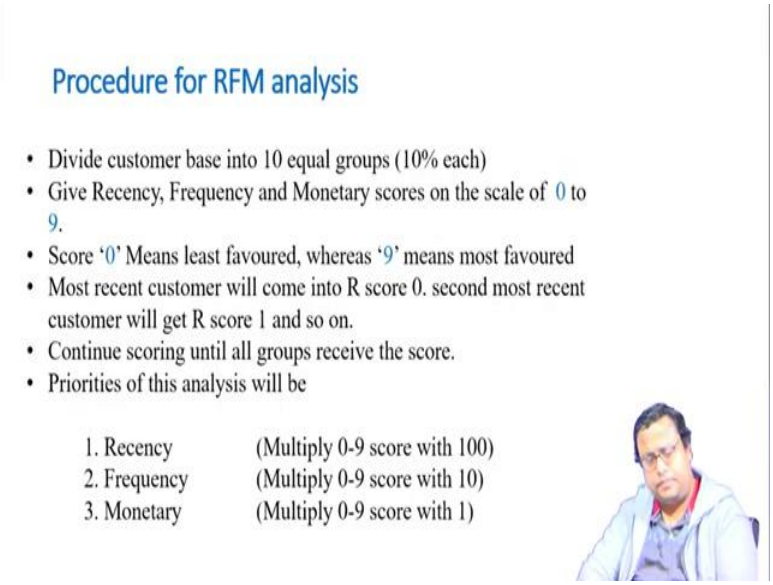
Now, let us say 1 million customers, or 1 million is a very big number, let us have 50,000 customers or 1 lakh customers were using Ola last year who have stopped this year. Now, they cannot send their offers to all these 1 lakh people, whom do will they send? Everybody has bad recency, so they will focus on frequency and monetary. So, at the end of the day RFM is used to not only create a good ROI of your advertisement expenditure, but it also helps to find out that whom to target out of those customers who have gone away.

(Refer Slide Time: 16:32)



RFM analysis helps companies to track the customer base and build a relationship that can increase sales and productivity. So, ultimately you have to create a relationship which will lead to sales and productivity. And it also useful to identify and track minimum losses. So, these are basic RFM needs.

(Refer Slide Time: 16:51)



What is the procedure? There are multiple procedure given by many people, the most common one is something that we will discuss here and we will run that also in our r code in the next video. But I will also give you an idea about the other ones other types of RFM analysis that is there in the market, but here we will use the easiest one.

So, it is saying that divide the customer base into 10 equals customer groups. So, if I break my customer base into 10, equal customer groups, basically I will get 10 deciles. Now, based on what based on what will I break? So, give recency, frequency and monetary scores on the scale of 0 to 9.

So, remember, so what I will do? First I will take the data and sort it based on let us say recency first thing. Then top 10 percent of recency will be getting a score of let us say 9, then 10 percent will get 8, then 10 percent will 7, 6, 5, 4 as I come down 9 to 0 each 10 percent will get that kind of a score.

So, I will just write down 9 then 8 then 7 I will create another column and I will write down the top 10 person guys if they you have 1000 people then top 100 guys are 9 then another 100 is 8 another 100 is 7, and so on. Now, with this new data set you again sought now sought based on frequency and do the same job 9, 8, 7, 6, 5 up to 0.

Again sort based on monetary, so three times sorting, last sorting based on monetary and again give 9, 8, 7, 6, 5 up to 0. So, score 0 means least favoured and whereas 9 means most favoured, so this is something that we have to understand. Now, there is nothing called recency, we do not measure recency, we measure basically the date of the purchase and current date, what is the distance?

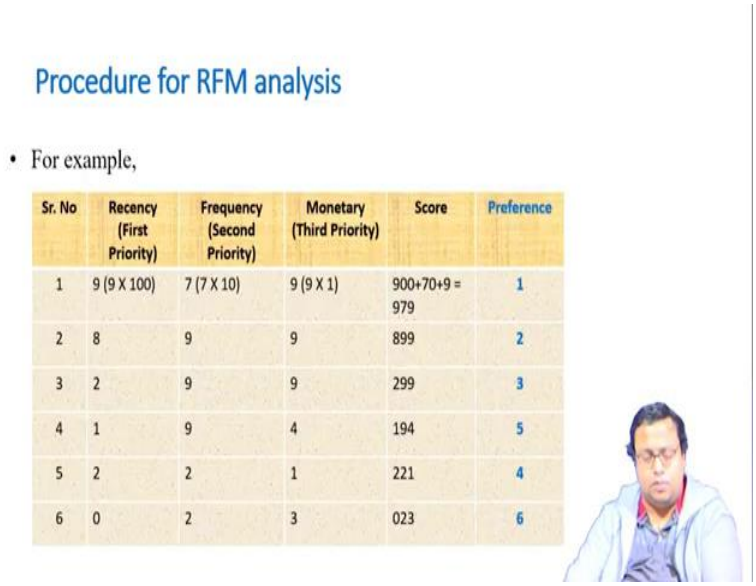
Now, the distance the smaller the better. So, here that distance variable when we create, we will create our increasing order sorting, on the other hand for frequency and monitory the higher the better, so I will create a decreasing order sorting. So, remember this is some basic new senses, that you have to keep in your mind that 0 means least favoured, what is least favoured?

High distance, high purchase time distance from today that is least favoured, low frequency is least favoured, low monetary is least favoured. Similarly, what is most preferred? Those at the stuff that you give the score of 9. So, most recent customer will come into R score 0, second most recent customer will get the R score 1 and so on, actually, most recent customer will get a 9, so this is wrong actually this one I have written wrong, this one will be 9 and this one will be will be will be 0 so 9, 8, 7, that is how.

And then continue scoring until all groups received the score and prioritize with the levels of. So, now this here is something. Now, I have given the score you got a recency score, you got a frequency score, you got a monetary score, so how will I get your ultimate score? There are

different ways. One is if by chance if you so that we will talk about that later, where how to put weights. But for example, now I am putting 100 weight to decency, 10 weight to frequency and 1 weight to monetary, why 100, 10, 1?, I will show you why.

(Refer Slide Time: 20:43)



So, if I give 100 to. So, let us say this is recency. So, recency these are the scores, the same there are 6 people who got some scores on recency, then some scores on frequency and some scores on monetary, not necessarily a guy who has high recency will also have high frequency, sometimes it might have not happened. For example, this guy his recency is 2 and 1, but his frequency is very high, monetary is also very high.

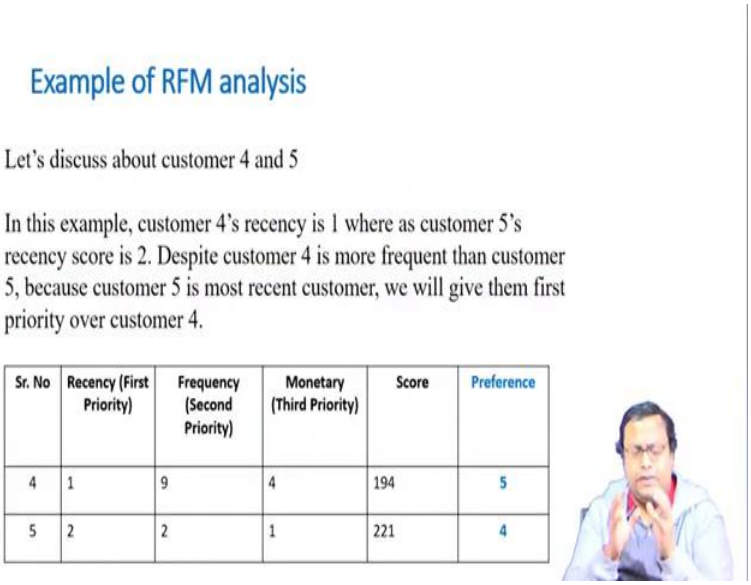
Now, what I do is? I multiply the first column, this column with 100, the second column with 10, third column with 1, so 900 plus 70 plus 9 gets 979, why? Because I have given it 100, 10 and 1, because now I know 979 means actually9 score for recency, 7 score for frequency and 9 score for monetary that particular information is stored here by giving 100 weightage and 10 weightage.

You might say that I am giving very high weightage to recency, actually this guy is not giving any weightage we will later see that how some weightages which are meaningful can be given, right now I am just writing 979. For this guy it is 899, so 899, 299, so 299, 194 so 194 and similarly the score is getting created.

Now, once this score has been created, I will create the preference based on the I will rank them. So, whoever has the highest score, he will get a rank of 1, whoever has the lowest score he will get a rank of whatever number of customers that I have.

So, here you will see that the most preferred guy is customer number 1 and then customer number 2, then customer number 3, then the fourth preference is customer number 5 and fifth preference is customer number 4 and sixth preference is customer number 6. So, this is how I am actually adorning these guys based on the preferences.

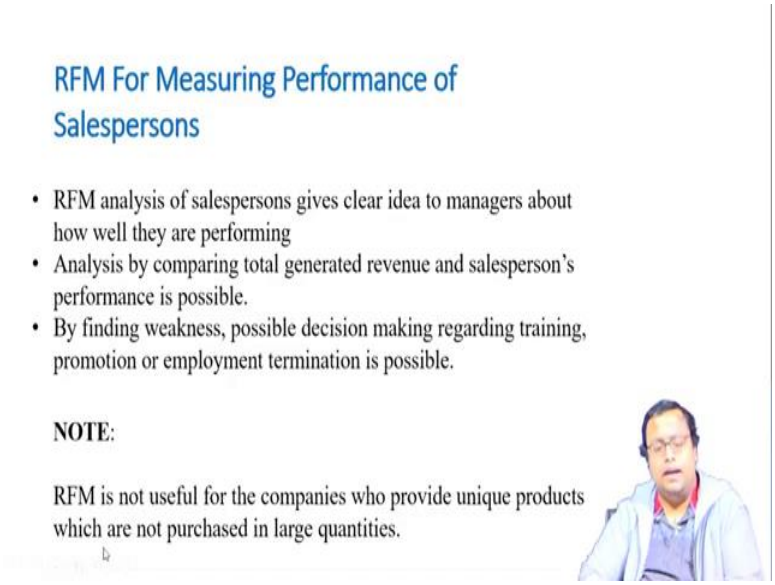
(Refer Slide Time: 22:55)



Now, if I focus on customer number 4 customer number 5, just checked 194 and 221. Now, the guy who has 194 has much higher frequency and monetary than customer number 5, 94 and 21, just check these two is much higher than these two, but still this guy will get lower reference and this guy will get higher preference because this guy is more recent according to the technique that I were I am using right now.

So, we are using right now, this 221 guy will get more preference than 194. So, here we are saying that recency is the top most story it has huge importance, but in real life, many people may want to argue with that and we will talk about that later as we come up.

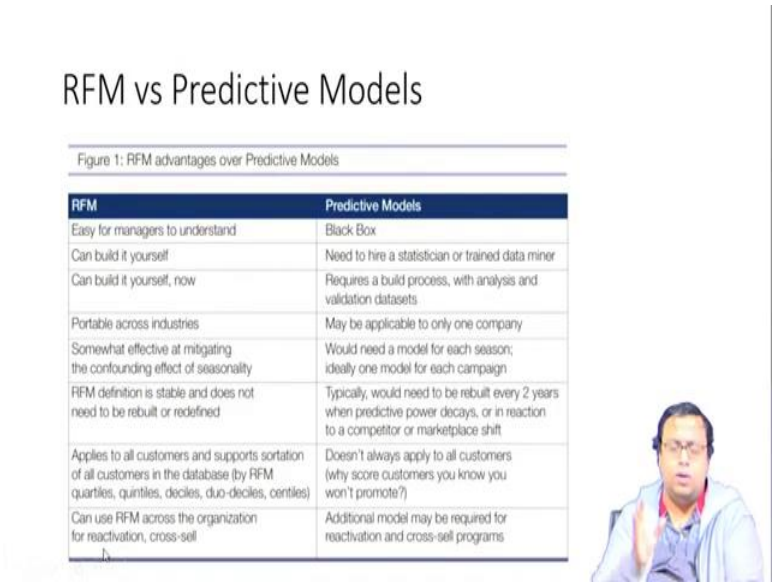
(Refer Slide Time: 23:45)



RFM for measuring of performance of salesperson, this is also another application of RFM. RFM analysis of salespersons gives clear idea to managers about how well they are performing. Analysis by comparing total generated revenue that is the monetary part and salesperson's performance is possible. By finding weakness, possible decision-making regarding training, promotion or employment termination is possible.

So, you can rank, like you rank the customers you can also rank the salesperson based on how recently he got an order, how many orders he has got and what is the total volume is of his order? So, RFM is not useful for companies who provide unique products which are not purchased in large quantities that is something is another important factor.

(Refer Slide Time: 24:38)



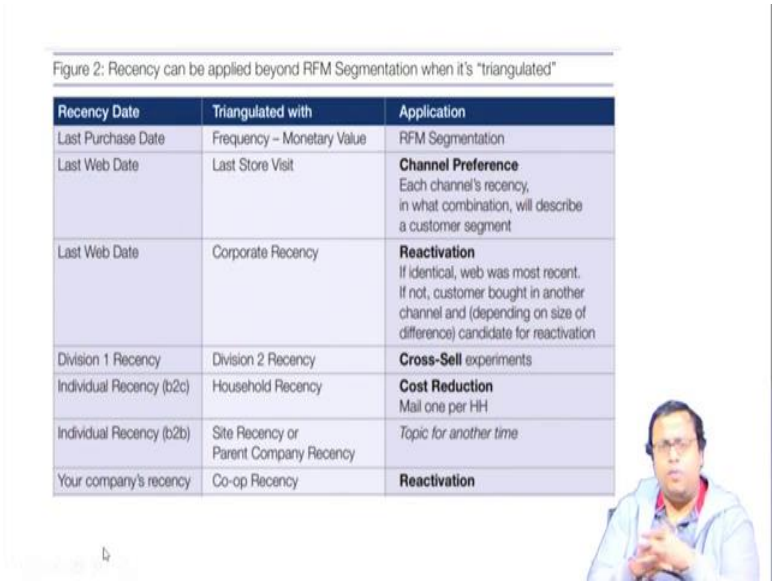
Now, if I compare this technique with some other predictive model, what are the various places where it is advantageous? So, RFM is easy for managers, where predictive models are black box basically. Can build it yourself, so you can build it yourself as a manager here you have to hire a statistician or a trained data miner. Can build it yourself obviously and then again requires a building process which analysis validation data set.

In RFM, you can have portable across industries, the same thing can be applied the same technique can be applied across industries, which cannot be done in case of a predictive model. So, one particular model that you develop is very contextual, you cannot use in another specific context. And then somewhat effective at mitigating the confounding effect of seasonality, here the seasonality as an issue.

RFM definition is stable and does not need to rebuild or redefine on the other hand here for predictive models typically would need to be rebuilt every 2 years. So, you have to (())(25:49) I would say rebuild your model you have to find out the parameter estimates once more, recalibrate the model basically. Applies to all the customers and support sortation of all customers in the database and this one does not always apply to all customers because the customer segments might be different.

And RFM can use you can use RFM across the organization for reactivation or cross selling on the other hand additional models would be required for reactive and cross-sell segment programs in case of predictive models. So, their sudden ease of use related implementation, related, generalizability related, advantages that RFM has over predictive models.

(Refer Slide Time: 26:36)



For we talk about recency, I have till now discuss the time, the time distances, now that recency variable has major I would say usage. For example, when last purchase date you combine with frequency and monetary you can do RFM analysis. On the other hand last web data web date, so last time you have somebody has seen in your web and if you can compact with it a last store visit of that particular person, then you know that each channels recency, in what combination, will describe a customer segments ,so how customers are challenged specific you can segment them.

And if you can find out the corporate recency means, when he has purchased? Then then if it is identical that web date and corporate recency is very close then he has purchased based on that web visit, but if that purchase has some difference between them or you did not see the purchase then probably he has searched in your in your platform and bought from somewhere else. So, that is a that is basically a lost sale and you have to reactivate those kind of customers.

So, division 1 recency and division 2 recency, if the if two products have different recency, then there is a chance of cross selling. The individual recency versus household recency, then mean mail one per household. So, let us say if individual recency and household recency are close, that means any purchase from one household and any purchases from one individual that recency is closed that means that is the person who is making a decision.

So, you focus on that person, you send one mail, rather than sending mails to everybody in the household. Similarly, individually with site recency that is not something that will discuss now. Your company's recency versus your competitors or co corporate recency, then that is also reactivation.

(Refer Slide Time: 28:37)

Non Buyer (no purchase recency)	Co-op Recency	Acquisition Better Recency on Co-op indicates better target
Non Buyer (no purchase recency)	List Recency ("selects")	**Acquisition** Better "selects" indicates better target (e.g., "hotline" names bought within last 90 days)
Recency	Products Purchased	**Product Recommendations** "People who bought x also bought y" recommender systems
Last Ship Date for Series	# of installments	**Tenure promotion**
Last Order Date	# of remaining items in series	Expected date when supply runs out
Last Subscription Date	# of issues in subscription	**Renewal Campaigns**
Last Response Date	NCOA move date	**Winback campaign**
Last Response Date	Job Change date	Bring us with you campaign
Last Purchase Date	Item Shipment Date(s)	Fulfillment issue?
Last Purchase Date	Pattern of Purchase Dates	Seasonality, Velocity, Segmentation (gift giver?)
Last Purchase Date	Time going by without repeat purchase	**Retention campaign**
Last Purchase Date	First Purchase Date	**Tenure** and Loyalty analysis

And then there are lots of such kind of things that recency can lead to, you can read them up and if you do not understand any one combination, for example, let us say recency and products purchased, people who bought x also bought y, so, this can this kind of recommendation system can be created. So, last time you bought x and then bought y, you can combine these two data set and created a competition engine. So, all of these kinds of combinations can create some kind of managerial insight which we can also focus on.

(Refer Slide Time: 29:11)

Figure 3: Response Summary by Recency

Response Summay by Recency

Last Purchase Recency	Customers	6 Month Season			% Resp.	Sales/ Piece
		Campaigns	Visits	Revenue		
0-3 Months	170862	1215314	49401	$21797541	4.10%	$17.94
4-6 Months	128238	1034895	21715	$8552928	2.10%	$8.26
7-12 Months	202443	1436914	21004	$8102208	1.50%	$5.64
13-18 Months	178411	912021	7217	$2636832	0.80%	$2.89
19-24 Months	154214	592846	3812	1259712	0.60%	$2.12
25-36 Months	294001	641028	3105	$1071216	0.50%	$1.67
37-48 Months	141888	226515	857	$239701	0.40%	$1.06
49-60 Months	46071	72184	215	$74715	0.30%	$1.04
Total	**1316128**	**6131717**	**107327**	**$43734853**	**1.80%**	**$7.13**

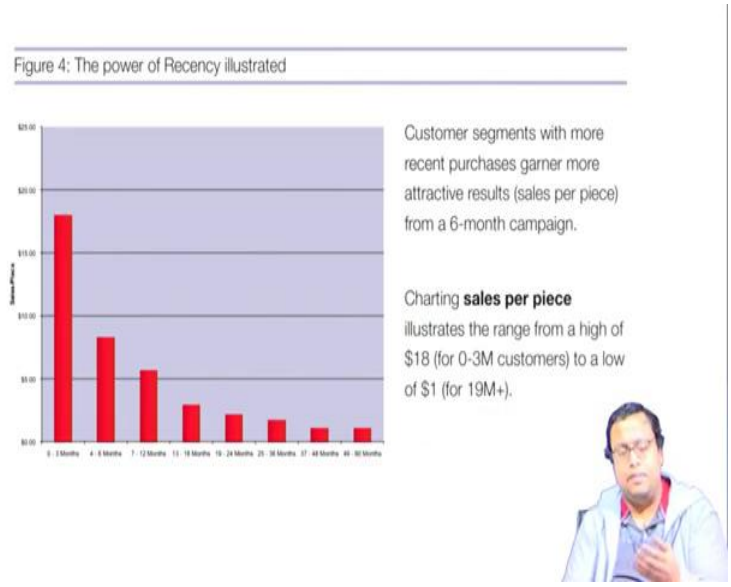
Best

Worst

It is been seen actually empirically, why recency frequency monetary came? It has been seen that people whose last purchase is very close, see if you see that their purchase revenue, their

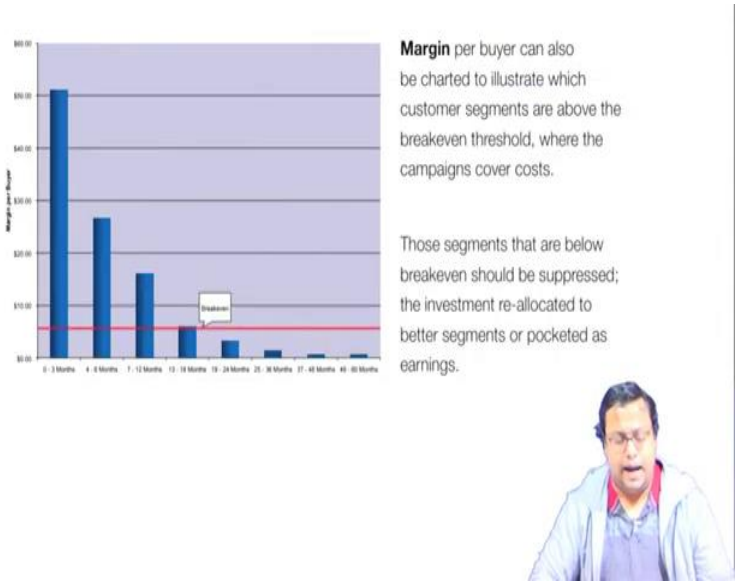
margin, etc. is much much higher when the purchase is recent verses the purchase is very far away. So, from best to worst if you come down that is basically from recent to past.

(Refer Slide Time: 29:38)



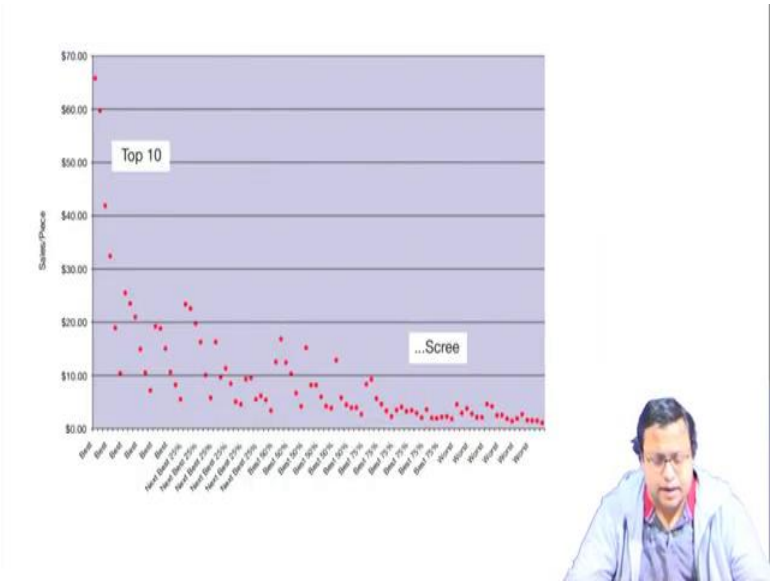
And that can be seen in case of the revenue generated. So, this is basically sales per piece that means, how much money you generate per piece that is also high when it is lower.

(Refer Slide Time: 29:50)



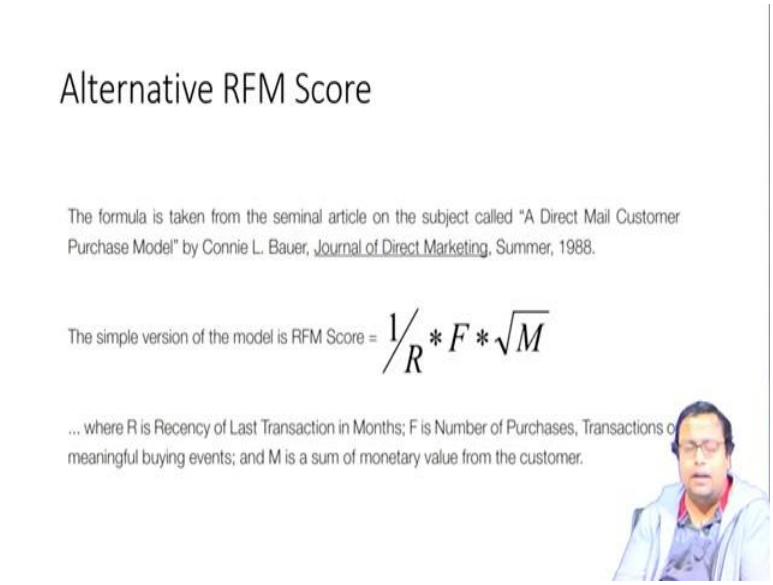
And the overall margin is also high when you buy, when your purchase distance is lower that means time distance is lower that means you have bought something very recently, then the margin is high, sales per piece is high, means the revenue per piece is also high and overall revenue is also high. So, generally, that is why which focus so much on recency.

(Refer Slide Time: 30:15)



And generally, when we create RFM, if you see the best RFM guys are here, their sales per piece is much, much higher than probably almost in an average probably 20-30 times higher than the worst guys. So, if you can find out that who are the best guys in case of RFM analysis, you are actually targeting the good customers and you can use that in the later period of time.

(Refer Slide Time: 30:42)

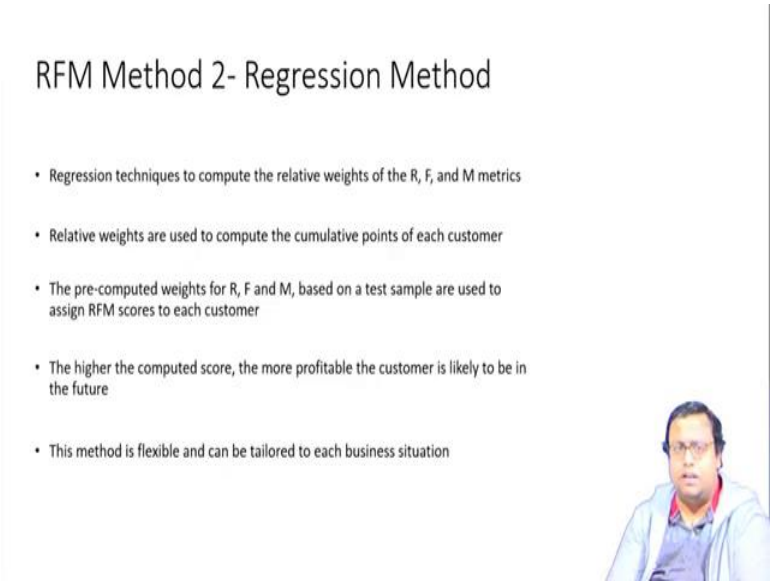


Now, I told that there are alternative models the same thing, but the calculation the ultimate calculation after finding out the R score, F score an M score, I have given 100, 10 and 1 weightage, somebody weight says how can you give that 100, 10 and 1 weightage you should give some weightages which are more meaningful.

So, one professor which is Connie L Bauer in Journal of direct marketing in 1988, so that is why I am saying this old concept, but it is still very used. So, he has written a paper on a direct mail customer purchase model and he told that RFM score should be 1 by R into F into root over of M, but now still, if you just do calculation, R has the highest effect on RFM score, delta R, delta RFM by delta R might be much higher than other ones, but still I am giving opposites.

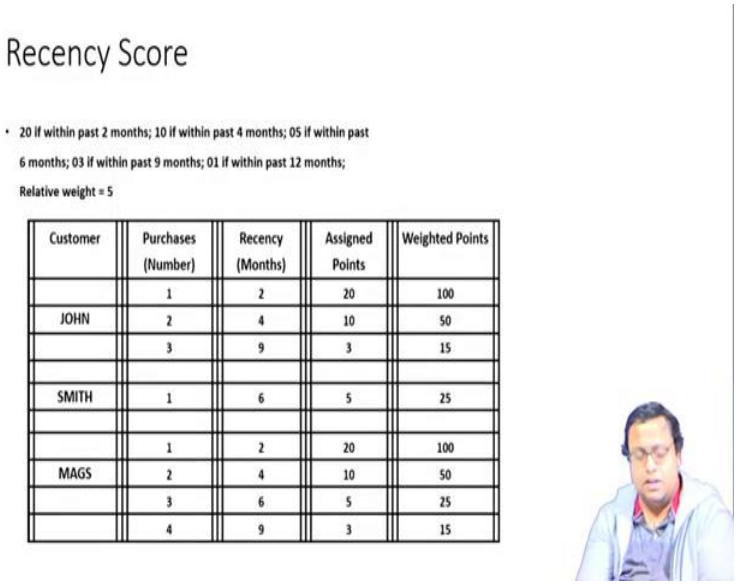
So, here R is recency that means, that the distance between, so 1 by R means the higher the distance the lower will be the RFM score, so R has been defined differently here. And then F is the numbers of purchase and M is the monetary value and R is in months, not in days, so that is something that we have to remember. So, 1 by R into F into root over of M and then you get the RFM score and again you just probably create a sought form in a decreasing order. So, higher the RFM score the better is the customer in terms of the profitability.

(Refer Slide Time: 32:18)



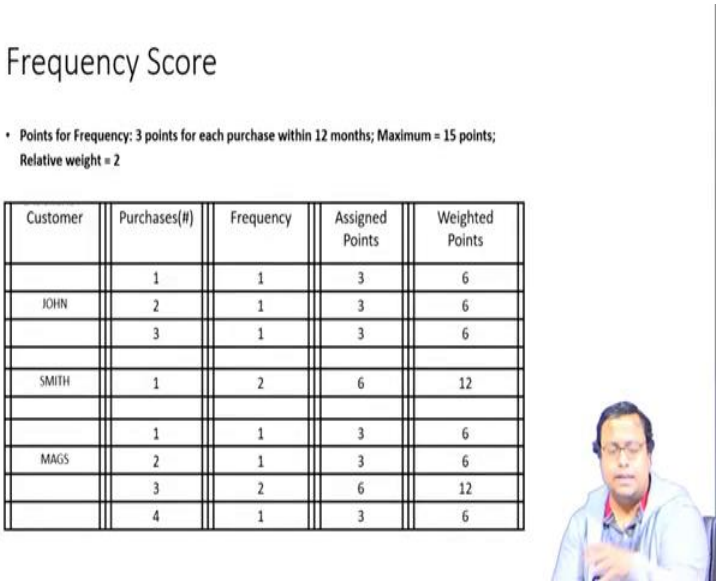
Alternative method, you can also use recreational techniques. So, you can use R, F, M and ultimate profitability as your y variable, do a regression and find out what is the weightage that is coming for recency or is a weightage that is coming for frequency or is the weightage that is a coming from monetary and use that weightage for the ultimate RFM score analysis.

(Refer Slide Time: 32:40)



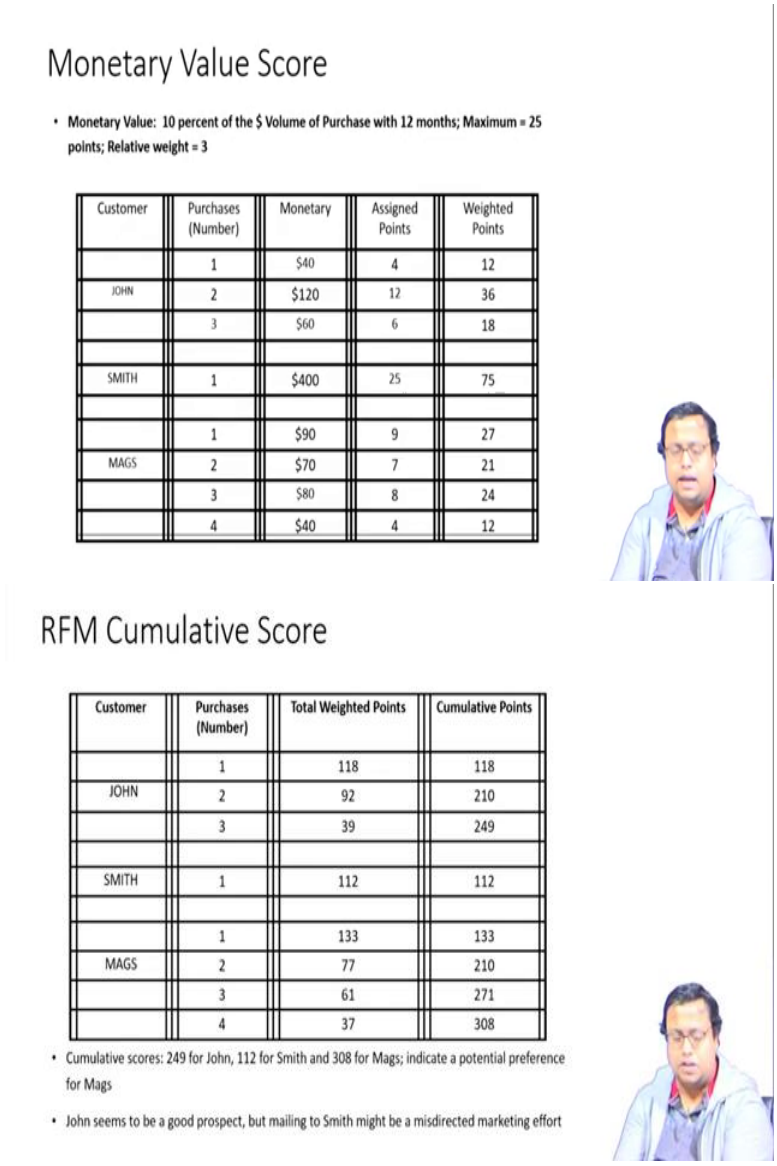
So, this is what here they have given. So, 20 is within past 2 months they have given points, 10 is within 4 months, and 5 if within 6 months, 3 points if within 9 months and 1 point if within 12 months, so that is how the scoring has been done. The relative weightage for recency score has been given 5, so that gives the weightage points. So, 1 you will see that recency months and corresponding points assigned has been written, as points assigned into 5 gives me the weightage points.

(Refer Slide Time: 33:12)



Similar thing they did for frequency, here the weight is 2 and the scores are, for 3 points for each purchase within 12 months and maximum up to 15 points, so that is how they have done the frequency.

(Refer Slide Time: 33:28)



And the monetary is, monetary value is 10 percent of the dollar volume of purchase within 12 months, maximum 25 they give a cut off. And the relative weightage is 3, so that is how they have calculated the weighted points for monetary also. So, ultimately when I find out the weighted point for recency frequency monetary, I am just add them up and then the cumulative point whoever is higher, so whoever is higher comes at the top most, I would say a preferred customer and whoever is lower comes as the least preferred customer.

And the cumulative points will actually talk about who is more preferred, who is less preferred. So, John seems to be a good prospect, but mailing to Smith might be a misdirect effort, because Smith has very low weighted, total weighted points also and cumulative points also. So, that is how we do RFM analysis.

(Refer Slide Time: 34:23)

And these are some of the reference from which I have used software contents and certain ideas as well. And thank you. In the next video, we will talk about how to do RFM analysis in R actually with a real data set. Thank you very much. See you in the next video.