Course Name:Business Intelligence and Analytics
Professor Name:Prof. Saji.K.Mathew
Department Name:Department of Management Studies
Institute Name: Indian Institute of Technology Madras
Week:02
Lecture:05

CASE: Bizocity scoring at AT&T | BI&A

Here is a case related to AT& T and a specific subsidiary of AT&T called AT&T Long Distance. So, the case discusses a problem faced by AT&T Long Distance at some point in time and how a team of engineers from Bell Labs came and proposed a solution for that problem. That is what you see in this case. The case is titled Bizocity score at AT&T. I would read this case for you to follow the nuances of or the threats of the problem and also the how the problem was addressed by the consultants from Bell Labs. So, this is how the case goes.

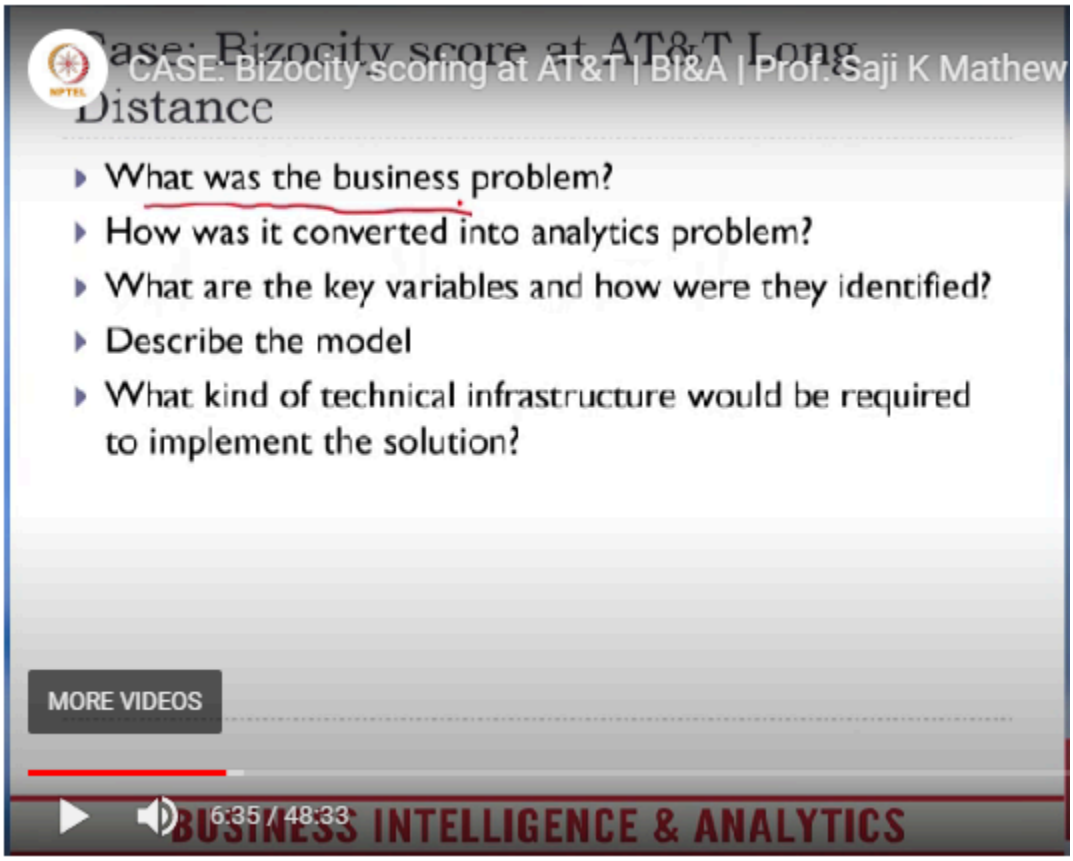
One very interesting application of data mining to the task of finding patterns corresponding to predefined customer segments is the problem that AT&T Long Distance uses to decide whether a phone is likely to be used for business purpose. AT&T views anyone in the United States who has a phone and is not already a customer, as a potential customer. For marketing purposes, they have long mentioned maintained a list of phone numbers called the universe list. This is a comp, this is as complete as possible.

A list of US phone numbers for both AT&T and non AT&T customers flagged as either business or residence. The original method of obtaining non AT&T customers was to buy directories from local phone companies and search for numbers that were not on AT&T customer list. This was both costly and unreliable and likely to become more so, as the companies supplying the directories competed more and more directly with AT&T. The original way of determining whether a number was a home or business was to call and ask. AT&T Long Distance had separate services for business and home customers.

Corina Cortes and Daryl Pregibom, researchers at Bell Labs, then a part of AT&T, came up with a better way. AT&T, like other phone companies, collects call detail data on every call that traverses its network. They are legally mandated to keep this information for a certain period. Many of these calls are either made or received by non customers. The telephone numbers of non customers appear in the call detail data when they dial AT&T 800 numbers and when they receive calls from AT&T customers.

They record, these records can be analyzed and scored for likelihood to be business based on a statistical model of business like behavior, derived from data generated by non business. This code which AT&T calls bizocity is to determine which services should be marketed to the prospects. Every telephone number is scored every day. AT&T switches process several hundred millions calls each day, representing about 65 million distinct phone numbers. Over the course of a month, they see over 300 million distinct phone numbers.

Each of those numbers is given a small profile that includes the number of days since the number was last seen, the average daily minutes of use, the average time between appearances of the number on the network and the bizocity score. The bizocity score is generated by a regression model that takes into account the length of calls made and received by the number, the time of day that calling peaks and the proportion of calls that the number makes to unknown business. Each day's new data adjusts the score. In practice, the score is a weighted average over time with the most recent data counting the most. Bizocity can be combined with other information in order to address particular business segments.



One segment of particular interest is the past, in the past is home business. These are often not recognized as businesses, even by the local phone company that issued the number. A phone number with high bizocity that is at a residential address or one that

has been flagged as residential by the local phone company is a good candidate for services aimed at people who work at home. So, this case highlights a problem as I said and also a solution and also certain additional insight they got when they in the process of addressing that problem, which they did not expect to get, but they got in the process of analyzing the data, as is discussed towards the end of this case. So, let me now try analyze this case using certain specific questions that can be used to guide us in analyzing the problem.

So, the first case question is, what was the business problem? What was the business problem? So, please keep in mind that when you go to a business organization or a company or a non-profit organization and when they describe a problem to you, they describe the problem in business terms, not in analytics or data mining terms. They will not talk about any algorithms, they may not be even aware of. So, you are walking into the premise of a company as a consultant or an analyst or a data scientist as it is, as the community is called today. And your job is to listen to them and understand the problem and the way they describe the problem should be understood clearly. So, if I ask you the question, what was the business problem in the case of AT&T Long Distance? It clearly says this, let us take it as question number 1.

It says they had a problem in acquiring new customers, the effort. So, it seems to be related to the marketing department. The marketing department is trying to acquire new customers or convert competitors' customers into their fold. And in marketing, we can call this activity as customer acquisition. Customer acquisition, customer acquisition is a specific head of budget of marketing department in an organization.

And this budget is to be invested such that they get the best returns. So, therefore, you must target, let me use that term again, target those who are likely to become customers. Why is that important? This has a budget. It is budgeted. You cannot target all the people of the whole world.

You do not have the money for that. A customer acquisition or marketing has certain budget for this purpose. And therefore, the number of people, targeting potential customers as the case says, potential customers in marketing will also be known as prospects. Who are the prospects? And this activity can also be termed as prospecting. If you define the business problem using business language, it is a customer acquisition problem, where in prospecting or in reaching out to potential customers, they are facing certain problem.

And that problem is how to reach out to potential customers. If from the first description, which is about, they say we have customer acquisition problem. So, if you

have to make the problem more specific, convert that into a question like how to reach out to potential customers. So, then as someone who is trying to understand the business problem clearly, you need to ask more questions, so that the sponsor of the project or the organization which is facing the problem, the representative of that organization is able to think about the problem and articulate it more clearly.

You see some articulation already in the case description, but it is through more detailed interactions, you get more finer details of the problem. So, let us try understand that. So, what is the problem? So, we first ask, what is the 'as is'? Oftentimes, in understanding problem, before you go into a solution, 'as is' is the current status. What is the 'as is' process or what is the 'as is' status? Then they would say, in order to reach out to the customers, they say we use competitors telephone directories. They buy telephone directory from competitors or that could be from the market.

So, you can imagine a directory consists of phone numbers and addresses. So, you get phone numbers and some addresses. But what is wrong? The phone numbers, there will be, you know, if you buy all the directories, possibly you will get the phone numbers of all potential customers. But the problem is that these numbers are too many. And the company has a specific requirement in terms of targeting potential customers.

They need to know, as is clearly mentioned in the case, they need to know who are business customers. So, AT&T Long Distance has separate products for business, business customers and residential customers. Unfortunately, if you look through a directory, which was used at the time of writing this case, of course, many years ago, it was not clear whether a number belong to a business or a residence. It is not written there, it is not clear whether it is a business number or a non-busines's number. So, since these two categories are important to add a customer to AT&T Long Distance's customer base or to target them before adding them, of course, they it is important for the company to know whether the number belongs to a business or a residence.

This is another part of the problem. They use directories, directories do not tell them whether the number is business or residence. And in order to solve this problem, as is what do they do? So what do they do? They employ someone to call each number, call each number, number means number of the, of the competitor. And then what do they ask? You imagine having a large database, you got access to a large database of potential customers of your business. And now you want to reach out to them with certain promotions or whatever.

So the effort is what do you call a telemarketing campaign. And each of us today get a lot of telemarketing calls, you know, from insurance companies, from healthcare

companies and so on. As soon as they start talking about what do we do? As soon as we know that it is a telemarketing call, either we try to block it, or we cut the call and keep it back, is not it? That is what, we do not have time to listen to unsolicited calls. And keep in mind a call like this, which AT&T Long Distance is making is an unsolicited call, which typically someone in business does not want to attend to. And therefore, it is not mentioned in this case, but what you can reasonably imagine is that they have a problem of response or more specifically, the response rate.

Response rate is, if you call 100 people like this, how many people respond? What is the question that you are asking in this case? You ask a respondent of someone who has non AT&T phone number and you call, hey, can you tell me, are you a business customer or a residential customer? And you can imagine what response you will give. So, why do you, what do you have to do with it? Why should I respond to you? So, the response rate is so low, and could be so low. So we call it say 1 percent, 2 percent, 3 percent, 1 to 5 percent, very low. If you call 100 people, maybe 1 to 5 people may respond if you are lucky. And therefore, AT&T Long Distance found that this, as is process of identifying potential customers is a very inefficient process.

It is very inefficient, in the sense that it takes a lot of time. And you do not get useful response or enough response to do your targeting effectively. But at the same time, it is a marketing effort. And as we already saw, marketing wants to add new customers to its business. And it wants to spend its budget on the right target groups.

It needs to identify the right target group to reach out to them with the products of the company. That is a very important requirement for the company. But the company is facing a real problem. And as we are trying to understand it here, the problem is twofold. One problem is related to data, the source of data, I will create the data source is the problem.

The data source is unreliable because it is published by the competitors. It does not have the information that you want, because you want to know whether a customer is business or residence. So the data source is one issue that the company is facing. What is the other issue that they are facing? When you detail out the problem, you find that you are not able to get the specific information of business or the customer segments, whether a given number is business or residence. So, you can call it a targeting problem.

Targeting problem in terms of specifically classifying customer segments or customers into potential business customers and potential or residence customers. So, in this particular case, the company is more interested as the case sort of alludes to, is trying to target business customers. And with that objective it is not , it is not efficient in the

current process to create that prospect list. And prospect list means what they want is list of business, sorry, list of potential business customers. Here potential meaning people who are likely to, likely to respond or people, one is likely to be business customers, likely to be business customers.

And if they respond, they should be good prospects, in the sense they should be valuable. There is no point in reaching out to customers who do not use the phone at all or they very rarely use the phone etc. Such prospects are not potential prospects or good potential, good customers to reach out to or good target. So, they are looking at it from two perspective, who are likely to be business customers and also business customers who are valuable. There is a point, there has to be some point in reaching out to them.

They should create value when they are added to the customer base. See how this problem which is described as a case in one page can be parched out into details when you look at the case closely and here what we have done is, we have used our understanding. For example, we understand what is telephone, we generally understand what is a business customer or a residence customer etc. So, some domain knowledge is available with everyone who is participating in the course.

It is not an unknown domain. So, therefore, we have some ease in going ahead with this case. So, let us continue to work with AT&T Long Ddistance's problem and what we have just seen is detailing of the problem in terms of the business side of the problem. So, this is purely a business understanding that we are trying to develop first and then we look at the solution side. So, what we have seen is the 'as is' part of the case and now we are going to look at what is the solution. So, we say in projects or in problem solving, the problem state is not desirable, but the state after the solution is more desirable than the current state.

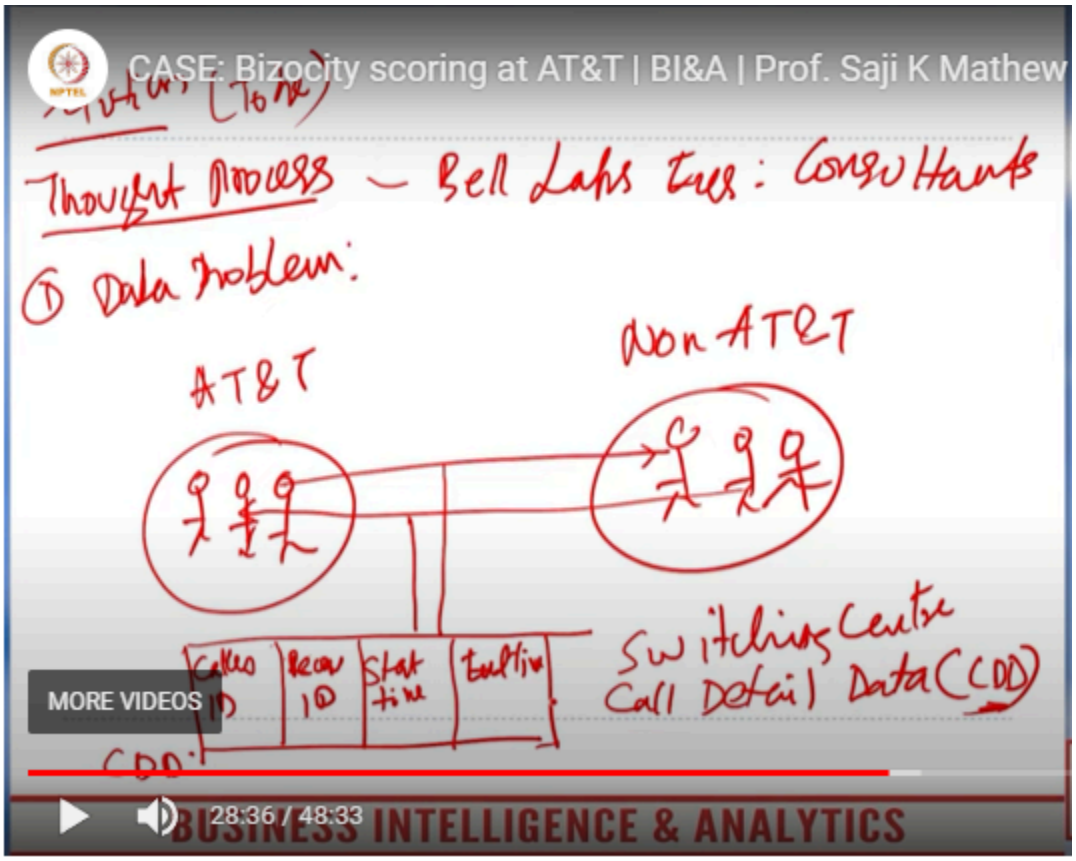
So, 'as is' not desirable. So, there has to be a, to be what should be a more appropriate or more desirable state where the company is able to reach out to potential customers effectively and those whom they are reaching out are really potential or valuable customers. These are the things to keep in mind when you actually look at the solution. Now, look at the approach or I would say thought process. The thought process followed by the Bell Labs engineers.

Bell Labs has come into picture. They are like the consultants here. What is Bell Labs? You know that Bell Labs is a group company of AT&T. Till the mid 80s, AT&T was one company. The sole telecommunications company in the United States was AT&T, American Telephone and Telegraph Company which was started in the early part of the industrial revolution, maybe early 1900s. And Alexander Graham Bell was one of the

founders of the AT&T.

So, in that sense, he was an inventor and also an entrepreneur. So, AT&T continued to be the sole telecom company of America. And then there was an antitrust judgment against AT&T because it became a monopoly. So which is not in favor of customers or citizens. So therefore, the company broke down into subunits and Bell Labs, which is the R&D of AT&T continued to be a group company, we can say.

I do not know about the exact registration of these entities in USA, but Bell Labs is the R&D of the AT&T. So they bring consultants from their R&D organization or their consulting organization, in terms of the IT industry. So they come and they actually look at the problem the way we actually try to do and try to give a solution. Look at how they address the problem. They give the solution is twofold, you can see one, they are trying to address the data problem.



How are they addressing the data problem? They say why are you going to buy the directories from the market? Instead of that, why do not you look at your own database? Your problem is data source problem, but you have not carefully looked at your own data. Have you looked at your database? So that is a very important point or observation that they make. What are they actually trying to suggest? There is AT&T, AT&T companies and telecom companies. So, they have customers, AT&T customers, and then there is the

world of non-AT&T.

And there are customers here as well. What AT&T is trying to do is to get customers of non-AT&T into their fold. So, that is their ultimate objective, but they do not know about them and they do not know how to target them etc. We saw the problem. So, what the consultant say is that AT&T customers do make call to non-AT&T customers and non-AT&T customers also make calls to AT&T customers. In both cases, the phone numbers of the customers of the competitors do get registered in the database of AT&T.

How does that happen? So, then we have to look at the infrastructure of telecommunications where whenever a call happens between two persons A and B. So, therefore, there is certain mandatory data that is captured and stored in switching centers. So a switching center for telecom is like the data center of a typical company. So, this is the data center and in data center, they store a record of each call, known as call detail data, call detail data or CDD. CDD is an important record that every telecom company is mandated by law in all countries, I would say, because this is very important in crime and investigation.

So, if you go to a cybercrime unit, or any police investigation, you would say they contact the telecom companies for CDD, they make a specific request. The competent authority in police department would make a request to a telecom company to share the CDD. CDD is a record of a call. And what does CDD consists of? CDD or call detail data consists of, CDD consists of caller ID, receiver ID, then start time, end time. If there is a start time, there has to be an end time.And that is it. That is it, just four field.

Do they have to store the call duration, the voice, etc? No, voice is not stored, voice is not supposed to be stored, or what actual content of communication happened between two persons is not supposed to be stored in any database unless, you know, some law enforcement authority ask a telecom company to do it. Otherwise, what they call is, what they store a CDD is caller ID, receiver ID, start time, end time. This looks like a very trivial set of data. But it is a very rich data when this data is aggregated, when this data is aggregated. So, this is basically telling who is calling whom, at what time of the day, for how long.

Look at it closely, who is calling whom at what time of the day, for how long is what CDD is all about. But what is the suggestion here, instead of going to directory, look at the CDD. Because when AT&T caller makes a non-AT&T call or vice versa, the non-AT&T phone number, either as caller ID or receiver ID is entering the database, the switching center of AT&T. Therefore, you have the phone numbers of the competitors already in your database. That is one part of your requirement, which is to get access to the numbers.

And one could argue that suppose a non-AT&T customer makes a call to a non-AT&T customer like this, the number may not get registered here. But AT&T is the largest telecom provider in the country. And also when you look at the calls over a period of say, five years or so, then you know, it is quite likely that every non-AT&T customer's number has entered the switching center of AT&T. And therefore, it is a fairly exhaustive database, which carries the phone numbers of most of the non-AT&T customers.

Number one, they are trying to address the data problem. But AT&T's marketing department is not satisfied with that. They also want to know whether a given non-AT&T is a business or residence. That information is not stored in the database directly. So, what are they trying to do? What are they trying to do? How are they addressing the second problem? So, in order to solve the second problem, they are applying the analytics.

The analytics solution is now coming in. They do not know whether a number belongs to a business or residence. That is the problem definition. And now, the analysts are thinking, are there some indicators in the database? Database means only four fields like caller ID, receiver ID, start time and end time. So two date stamps, essentially two date stamps. So how do you get indicators whether a given number is a business customer or not? You can see that effort to derive behavioral, what we today call behavioral indicators.

Based on customers behavior in terms of their phone calls, are there some indicators which can be used to model a business like behavior? So that is the sort of problem translated into the analytics domain. It is not directly given whether a number is business or residence, but from the behavior, from the calling behavior, are there indicators of business like behavior? And can that be understood? So, what are they trying to do? They want their outcome variable is the probability of being a business customer. So, business customer behavior which is called bizocity score or a probability score, we can call it between 0 and 1. 0 being non-business, 1 being business, and anything between 0 and 1 is a probability.

So, what is that? This can be a function of certain indicators. And they are trying to derive those indicators from the call detail data. For, they use a few indicators. So, they use time of the call, duration of the call, call to non-business. Three indicators from the database of the call detail data, what time the call is made in, essentially meaning if a call is made during day hours, business generally functions during day hours, then it is likely to be a business call and duration of the call.

So, you have to understand the context of the case. The case was written say about 20 years ago when the cell phones as we use it today was not available, nor the internet based calls, the VoIP calls were not common those days. So the call was very expensive. And only business or rich people could actually make long duration calls. So a long duration call here is taken as an indicator of business like behavior.

Day hour call is made is taken as an indicator of business behavior. And then call to non-business, meaning business to business calls will be more than business to customer calls. So, or a call to a business means it is a business call. So, it is likely to be a business number. So, these are sort of assumptions they make. And these are sort of indicators, reasonable indicators they assume or they hypothesize to be correct.

Of course, this has to be tested. And how do they arrive at these indicators? You can see the twofold thinking. One is these are likely indicators of business like behavior. And they must have discussed with the AT&T Long Distance service providers, their managers or their own experience, you can think of how they arrived at this. One aspect is consultancy or domain knowledge, etc. The other aspect is each of this indicator, the data is available in their database.

They are not dependent for this data on any other source. So availability of data and useful indicators of the outcome they want to model. So, these are two principles they follow in identifying variables. We can call them as variables. The variables are identified, three potential variables. And then they can actually derive these indicators of customer behavior from the database, aggregate it over a period of time.

And then use appropriate modeling technique. They say they score each number. So, you can imagine there are numbers x x x x x, y y y y y, z z z. Numbers like this, and each number gets a score.

So, some number gets 0.1, this may be 0.8, this may be 0.6, 0.4 here, 0.3 here and so on. So, this is bizocity score. Each number is scored using a scoring model or an appropriate model. When we will discuss classification in near future in one of the sessions, we will explore what are the potential techniques that can be used. It could be a logit, it could be any other classification technique like the decision tree, which we are going to discuss in one of the sessions.

So, those algorithms could be employed to build a model like this based on the variables and their data. And then test whether this bizocity score is predicting business like behavior accurately or not. So, the prediction performance of the model, of course, need to be tested. So, that is a purely analytics or data analysis exercise.

That is the, that is the modeling part where they have to validate the model. And the plan is, the proposed model is this, they give a score to each customer. And then you can also see that based on the score, now you can sort the phone numbers. The yyy will come first, then will come zzz, then will come this two dots, and then will come this, let me call it three dots, or three dots, or then comes xxx, which we, this is a sorted list, where you have the highest score on the top and the lowest score on the bottom. Why are we doing like this? Because the business wants to know whom they should target.

And the business will also have limited budget. They cannot, they may not be wanting to target all these numbers, which are millions of numbers potentially. But they want to reach out to those who are very likely to be business customers, or these are the ones they actually want to target. So in effect, these consultants are solving a business problem. Let me say that again, analytics is about business, business analytics is about business problem solving.

So therefore, analytics is not just data analysis. Analytics is not about algorithm performance. That is involved in analytics, but analytics is basically to solve a business problem, think through the business problem and give a solution to the decision maker, which can be used by the decision maker. So here in this case, you can sitting at the chair of a decision maker or a marketing head, whom, where should I spend my budget on, who are the targets of this campaign from a sorted list like this, based on the likelihood of being a business customer, they can decide whom to target and how many to target. And that was a problem they had. And they had a data problem, which is already solved. They also had a difficulty in identifying who is a business customer, who is a residential customer.

That problem is solved through a proper classification, classification method of problem solving called the scoring model. The modeling aspect, we will be discussing later, but here our purpose is to understand what is the thought process that goes into the solving of this business problem. Now, one aspect of the problem was to understand who is likely;there are two aspects of the problem, identifying likely business customer. So the solution is bizocity score. So a potential customer or a prospect should be a likely business customer and should also be valuable.

As I said previously, if a number is, you know, is dead or, you know, is not using phone number for a long time and is not make, even if the customer is active, but very rarely use the phone, etc it is not worth perceiving that particular prospect. So the other aspect, which you can see in the case description, they also give a profile, they create in addition to bizocity score, they profile each number. How do they profile? There is a, there is a

phone number and then they look at the recency of the call. Has this number made any calls recently? Recency, they also look at how frequently this number appears in the network.

So they look at frequency and they also look at the case size duration of the calls. How long calls are made? And then they also add the bizocity score. So each number is profiled using four characteristics or four attributes; recency, frequency, duration of the call, bizocity score. And that is a very brilliant solution that the analysts are giving to a business. An analyst here you can see is someone who understands business and someone who understands algorithms or mining techniques, I would say. Someone who understands technology, someone who understand business domain and someone who understand algorithms and statistics or the sciences, which are the three pillars of analytics, as we saw in the previous session.

So, how is this solution brilliant in identifying valuable customers? Recency, frequency and monetary value, RFM are they are called. These three variables are widely used in marketing as indicators of customer, customer loyalty and customer value. How good a customer is, how loyal a customer is, etc. These three indicators are used together.

And it is the loyal customers who makes profit for the company. And therefore, it is important to acquire and invest on them, a simple principle in marketing. So they have brilliantly, although they are engineers, they have brilliantly brought together these three variables, duration as a proxy for monetary value. You may not agree with it in today's context, because duration of the call may not be directly indicating economic value.

But let us understand in the context of the case. And bizocity score indicating business like behavior. So, a manager can now look at the overall profile and filter potential targets, not just based on bizocity score alone, but also based on loyalty. Loyalty and value, customer value being the m variable or the duration variable. So, you can look at a problem from business as well as analytics.

And you can see how brilliant is the solution that is suggested. And this case is discussed in this detail, in order to give an understanding of the role of analytics in business and management. It is to solve business problems. And therefore, engaging closely with business and solving their problem is the purpose. It is not to demonstrate how good you are in algorithms, or how, what are the algorithms you know, and therefore, since you are good in certain algorithms, just plug that in wherever possible. That does not mean analytics. Analytics is first of all, understanding business problems and seeking the knowledge that is required to solve the business problems, so that

business value increases, where analytics contributes to business in increasing its business value.

You know, I think I gave indicators of that in the previous session, where the profits of the business increases. So from the, so that is one aspect of this discussion. We have discussed how analytics involves translation of business problem into analytics problem. So that is point number one in terms of learning from this case. And the second aspect of learning is to understand if AT&T has to implement this solution real time, what kind of investments or what kind of resources they will require.

Because this is like a proposal, but it is for the company to implement it real time because we, we, we noted in the case description that they have gone for real time implementation, meaning the model gets retrained. So that is something that we learn, meaning that the, once the data gets updated, they train the model again, and make the model more sort of accurate. So, that shows that they have implemented the solution real time. And if so, what is the infrastructure that they need to implement the solution? That gives us a sense of the business intelligence, architecture and infrastructure requirements. From an analyst's point of view, from an analytical expert, that aspect of technology is also very important.