Predictive Analytics - Regression and Classification

Prof. Sourish Das
Department of Mathematics
Chennai Mathematical Institute

Lecture - 59
Hands on with R: Dynamic Pricing with Cheese Data

Hello all, in this video I am going to use show how can we use regression analysis in dynamic pricing. What is dynamic pricing? Sometimes you will see from if you want to buy air ticket from say Chennai to Bombay it might cost you maybe 7000 rupees in one way, in another day it might cost you 9000 rupees and another day it might cost you 6000 rupees. Why airline charge different price for different days or different time slot for different, though it is travelling the it is in the same route.

The air travel the price of these cases depends on the availability of the seat and the demand for the seat. If the demand for the seat is high then airline charges higher fees or higher fare for the same ticket. If the sub demand for the seat is low then airline charges a bare minimum base price for the air fare.

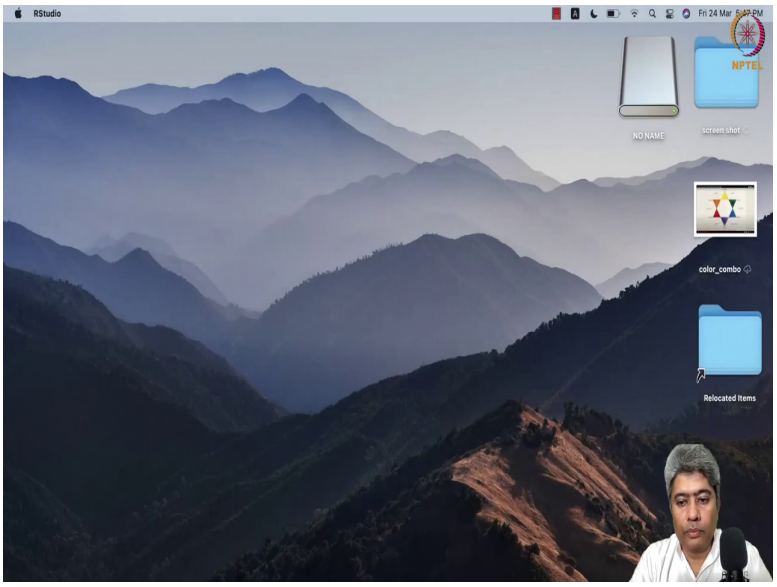
So, this kind of discriminatory pricing often called in industry dynamic pricing based on the demand and supply of the product or the service that are available. So, in this video I am going to use cheese data set from bayesm package to demonstrate how can you use regression analysis for dynamic pricing.

(Refer Slide Time: 02:15)

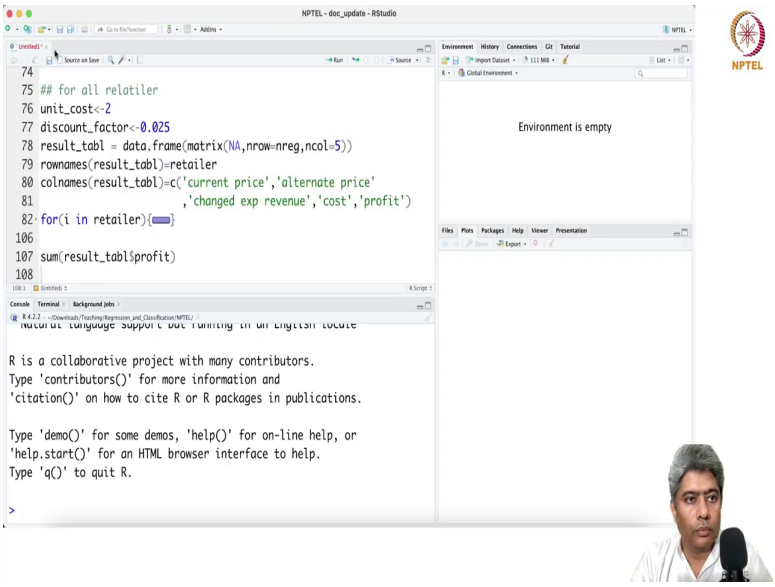
So, first I will open my R.

(Refer Slide Time: 02:19)

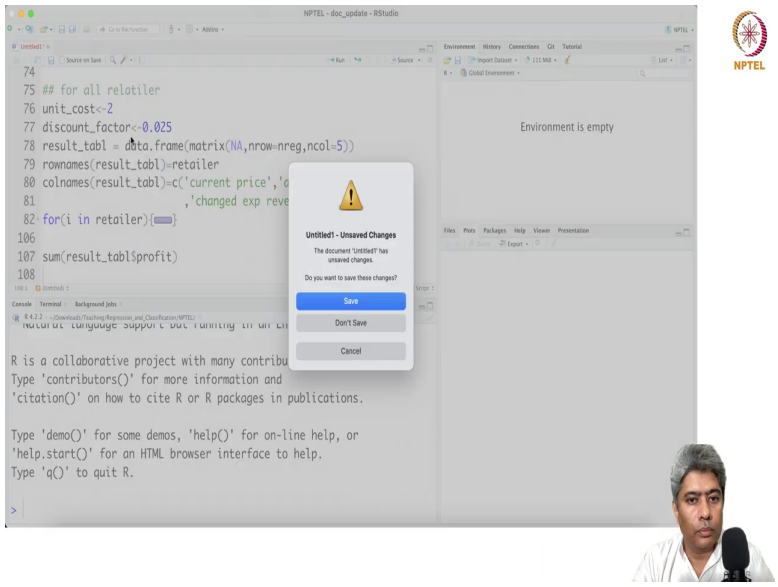
(Refer Slide Time: 02:21)



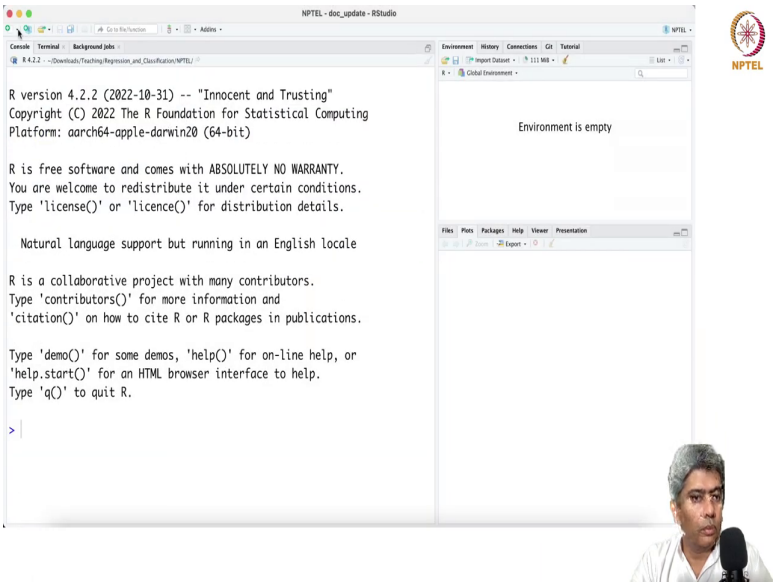
(Refer Slide Time: 02:25)



(Refer Slide Time: 02:23)

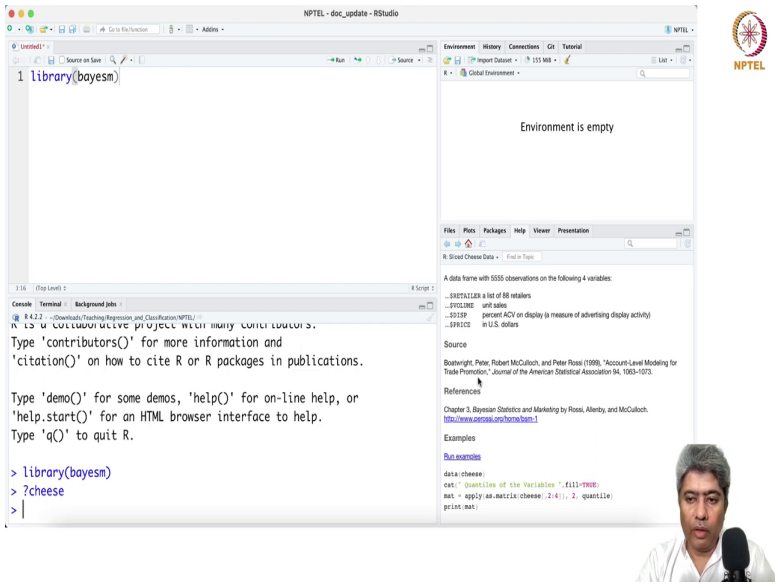


(Refer Slide Time: 02:27)



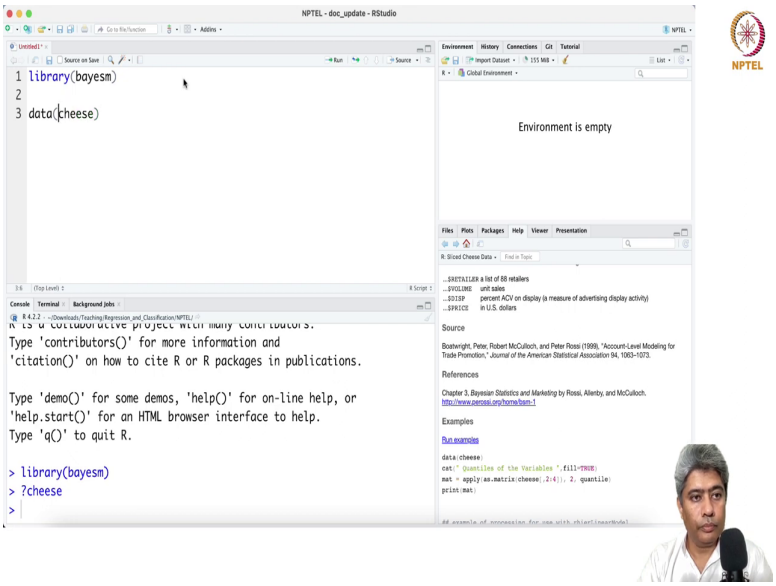
Ok.

(Refer Slide Time: 02:31)



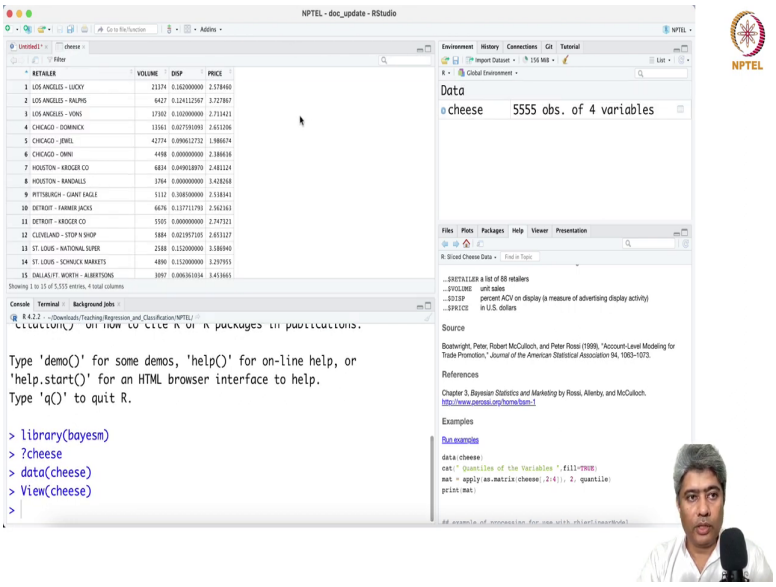
So, first I am going to call library bayesm. In the If you do not have this library just, please install this library and in this bayesm package there is a called data set available called sliced cheese data set ok. The reference is also available in McCulloch and you know Peter McCulloch and Peter Rossis Robert McCulloch and Peter Rossis book on Account-Level Modeling for Trade Promotion.

(Refer Slide Time: 03:17)



And in this paper the this data set first appeared. So, first I am going to call this data set data cheese.

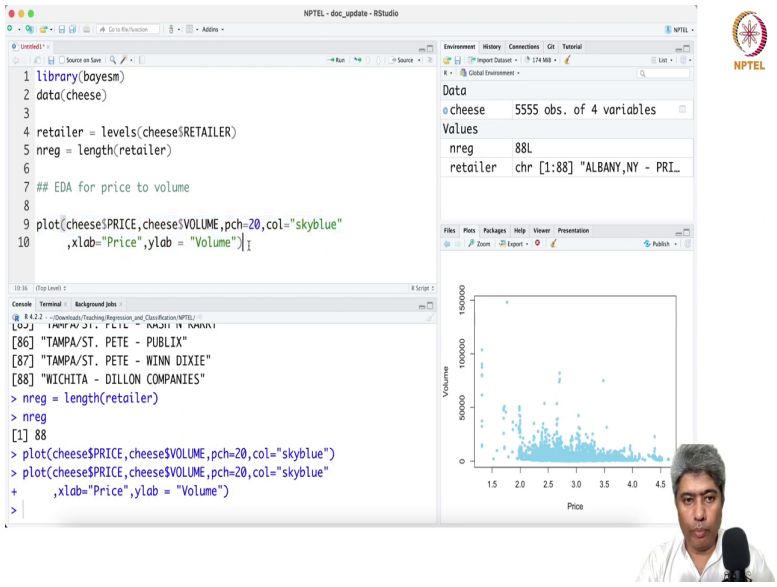
(Refer Slide Time: 03:43)



So, here you have the data set. So, you have four columns retailer name of the retailer there are about 88 it says 88 retailers LOS ANGELES-LUCKY, LOS ANGELES- RALPHS, LOS ANGELES-VONS. So, Lucky, Ralph, Vons, Dominick these are the retailer chain and this is the city name.

So, city name and retailer chain they are different and; obviously, and then along with that the VOLUME, VOLUME or the unit sales the number of volume sales that happened in that week. This typically called display activity that has happened some marketing activity they measured in the scale of 0 to 1 and the PRICE per unit price in US dollar is given.

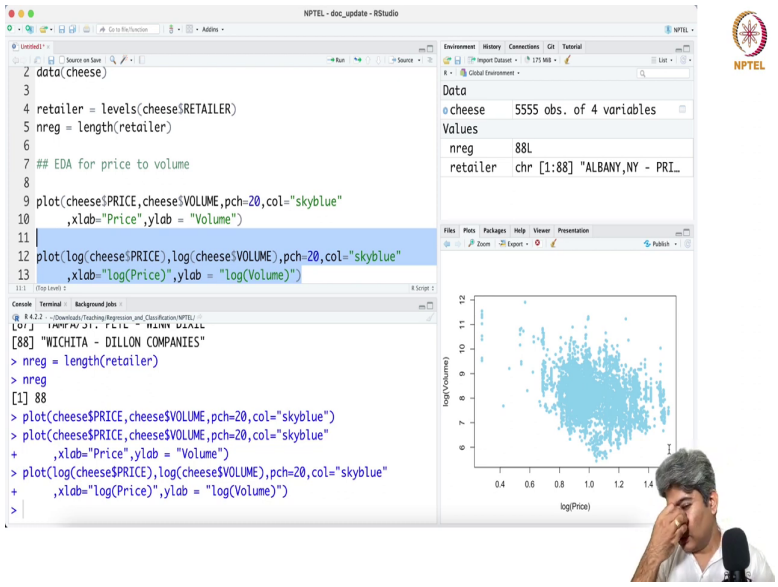
(Refer Slide Time: 04:40)



So, so, the first thing I am going to do I am going to take the retailer equal to check what are the levels that are available in cheese data set. So, you can see there are 88 retailer different retailer are available. So, I am going to say there reg. So, these are the number of leng retailer is 88, ok. Now, I am going to do some Exploratory Data Analysis EDA for price to volume data to volume.

So, the first thing is plot cheese dollar PRICE and cheese dollar VOLUME for each price what are the cells that you see. And say ch equal to say 20 and color equal to “skyblue”, ok. So, it looks like this how that is how the price and so, xlab equal to “Price” and ylab equal to “Volume” or unit cell number of unit cells that has happened ok. Let me just run it.

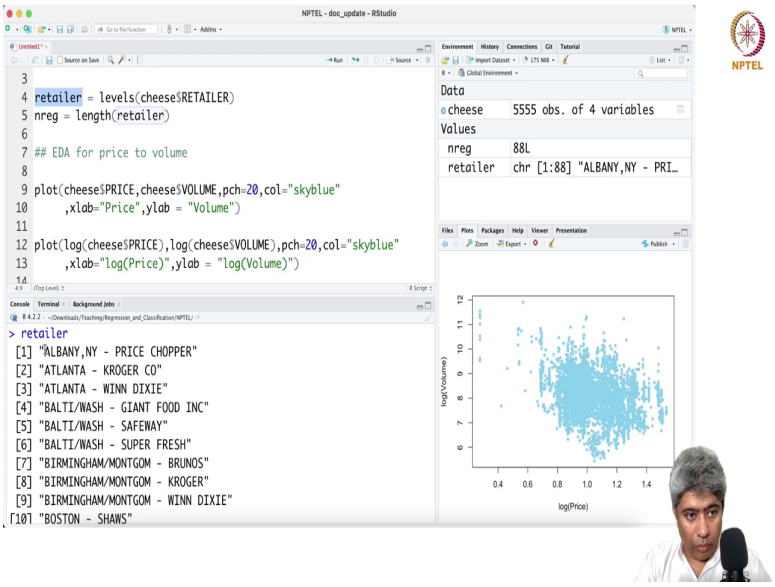
(Refer Slide Time: 07:08)



And now, as I told you in the in a previous video that, if it is any economic variable one should use log transformation; so, I am going to use log transformation over PRICE and VOLUME and. So, it will be “log Price” versus “log Volume”. So, looks like log Price versus log Volume has a sort of a you know expected slope, which is sort of fitting one line kind of thing you see.

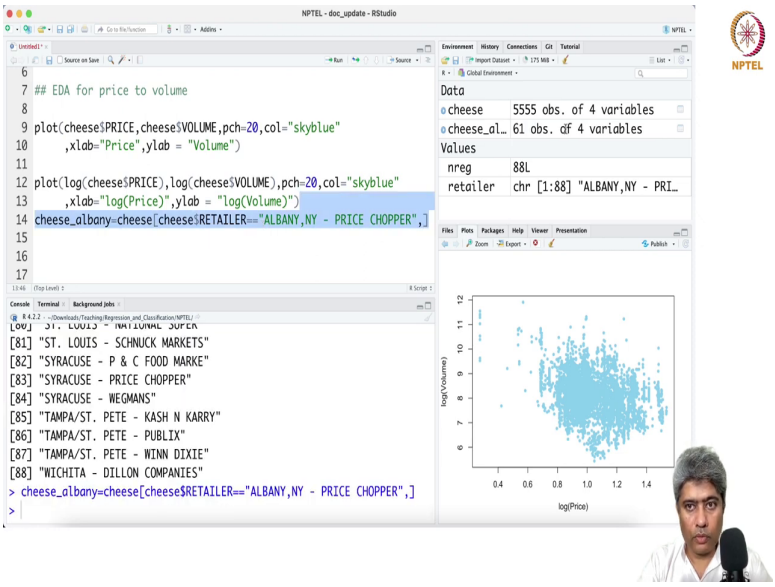
And now so; that means, most likely Log Price to Log Volume modeling Log Price and Log Volume will be much more you know effective than you know trying to do something different. So, next what I am going to do is say if you say retailer.

(Refer Slide Time: 08:26)



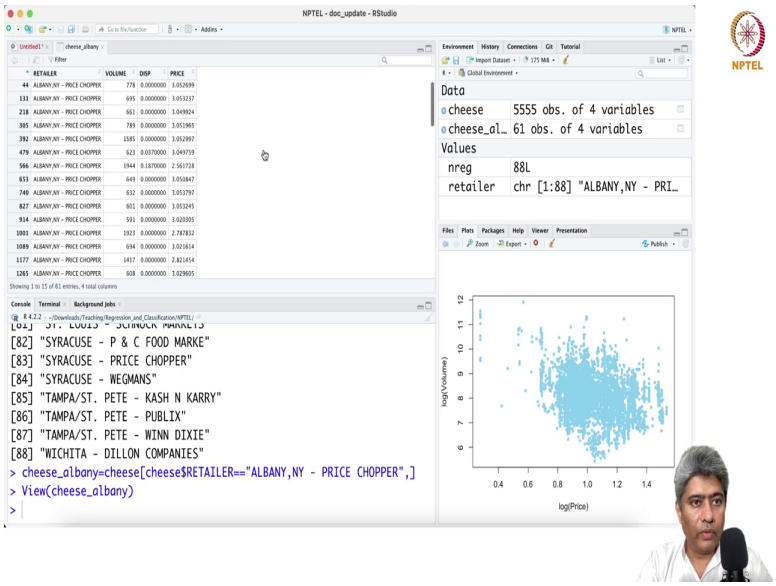
So, ALBANY PRICE CHOPPER and so, Albany is a small place and I we can have a look like you know maybe Los Angeles is a big place “LOS ANGELES – LUCKY” and they are also bit different places. So, let me what I am going to do is something like this.

(Refer Slide Time: 08:54)



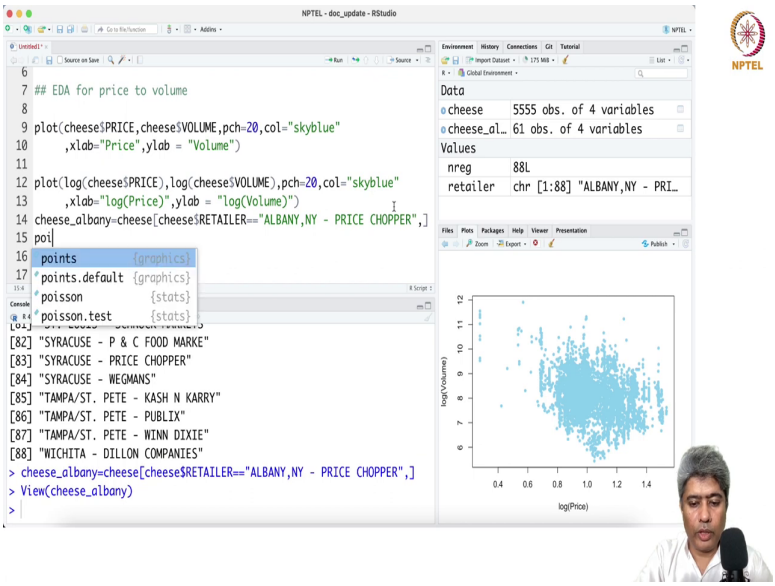
So, first I am going to make a subplot or you know subset of the data. So, cheese, but its only for the albany ok. So, the cheese dataset dollar RETAILER equal to. So, let me just copy this and comma if I just do that.

(Refer Slide Time: 09:42)



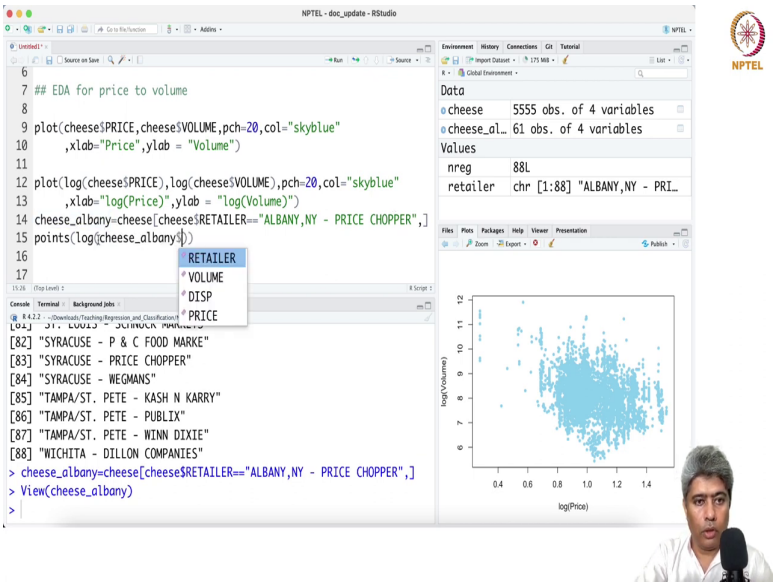
So, these are the values of all the 61 observations all belongs to ALBANY PRICE CHOPPER, ok.

(Refer Slide Time: 09:53)



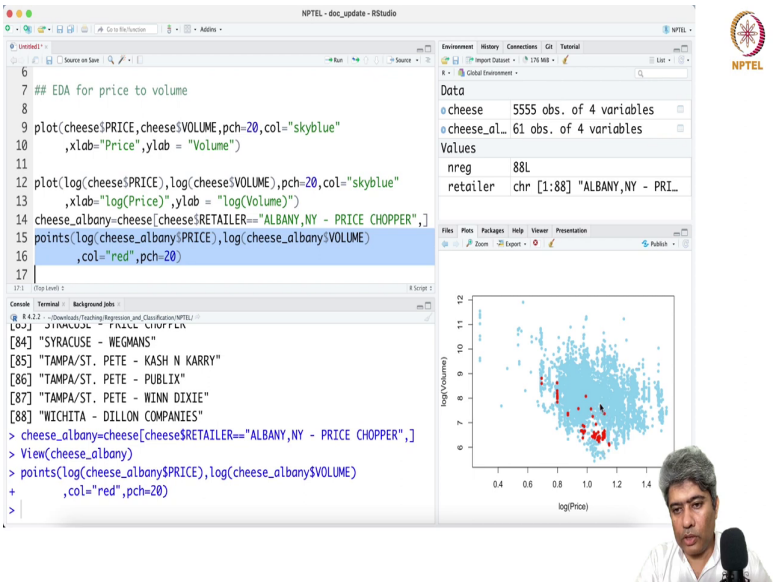
And now what I am going to do.

(Refer Slide Time: 10:07)



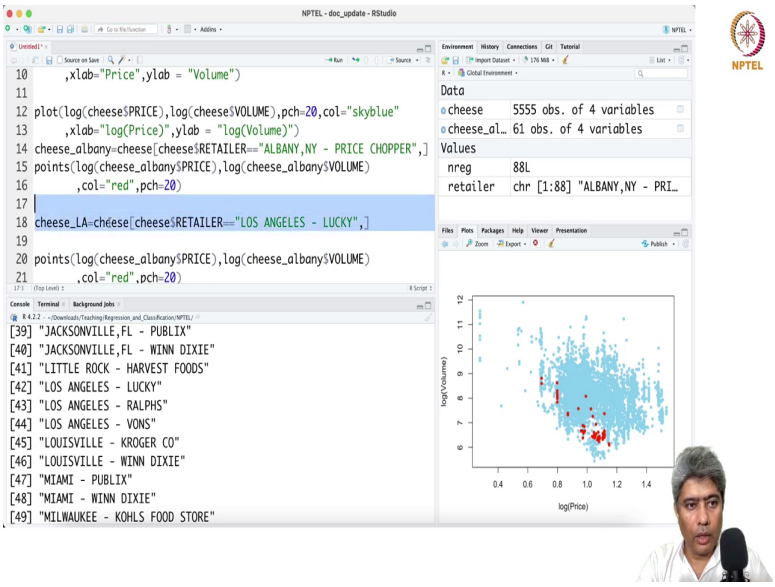
I am going to point plot the points of only log cheese albany dollar PRICE comma log cheese albany dollar VOLUME and color equal to “red”.

(Refer Slide Time: 10:27)



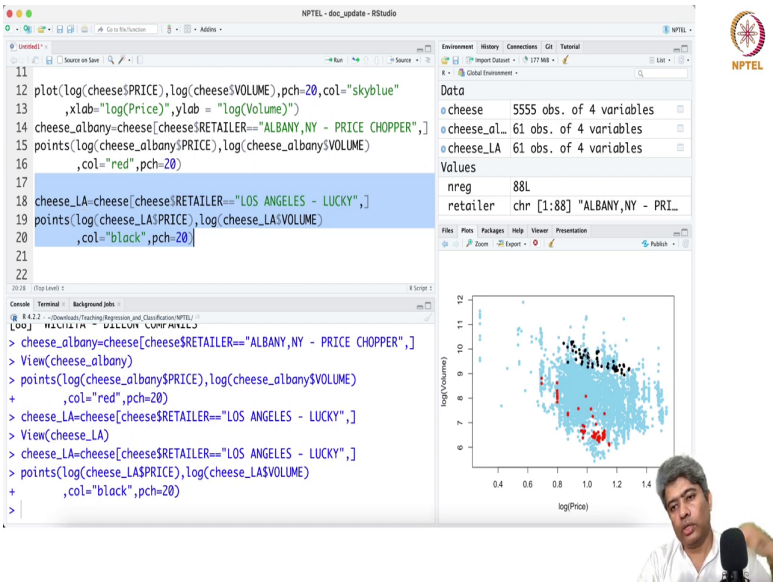
And pch equal to 20. So, you can see the price is these are the points, which all belongs to price chopper. So, chopper Albany price chopper in this range. So, you can give a different for different price at different price level the number of sales behaves like this, ok.

(Refer Slide Time: 10:58)



Now, I am going to curve compare it with “LOS ANGELES – LUCKY” let me see same thing, but instead of I am going to do with “LOS ANGELES - LUCKY” ok cheese LA.

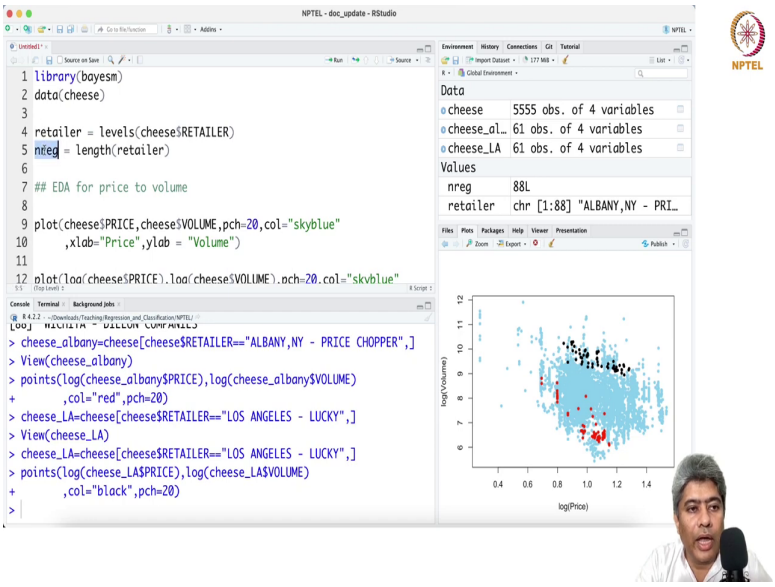
(Refer Slide Time: 11:31)



Now, if you see it also has 61 observations VOLUME and these are the PRICE and VOLUMES. Now, if I just do instead of cheese LA, LA and instead of “red” I will use “black” right. So, all these are just different cluster altogether. So, now, if I fit a model here based on using all the models then the line will pass through somewhere here.

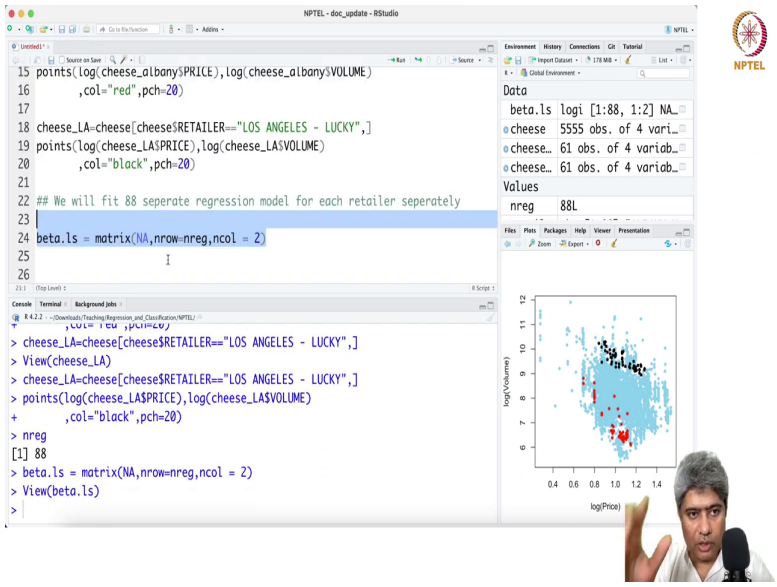
And if I use this line right because if I use all these points to fit a model that line will pass through the middle of this and that line will neither fit price chopper in Albany nor it will fit Los Angeles Lucky. So, what we want? We want a particular for a for each retailer we want to fit a model for each model or we want to fit a separate model. So, that is the idea. So, what we will do so; that means, there are how many retailer?

(Refer Slide Time: 12:54)



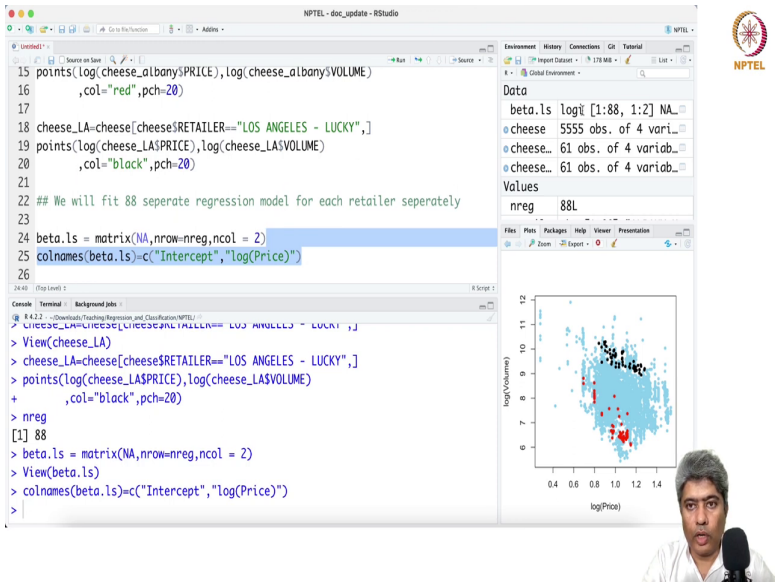
88 retailer, we want to fit 88 regression model.

(Refer Slide Time: 13:00)



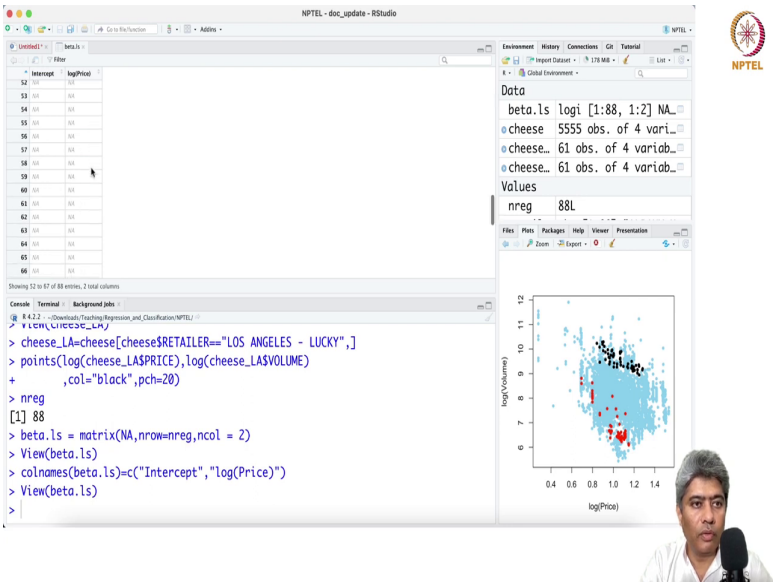
So, let us fit 88 regression model, we will fit we will fit 88 separate regression model for each retailer separately, correct. Now, that is my goal that is my goal. So, first what I am going to do beta dot ls equal to matrix NA, nrow equal to nreg and ncol equal to 2. So, this is my beta I am just creating a place holder this is just a place holder.

(Refer Slide Time: 14:04)



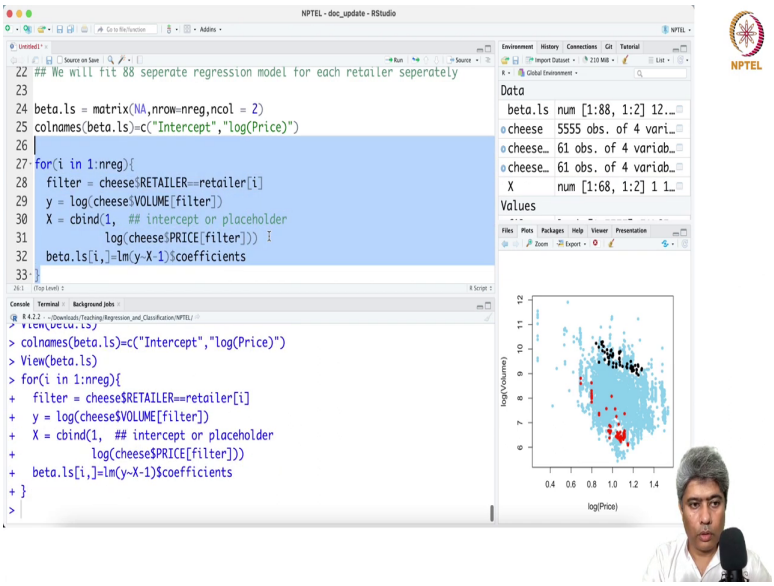
And colnames will be c first is “Intercept”, ok and the second is “log Price” correct.

(Refer Slide Time: 14:23)



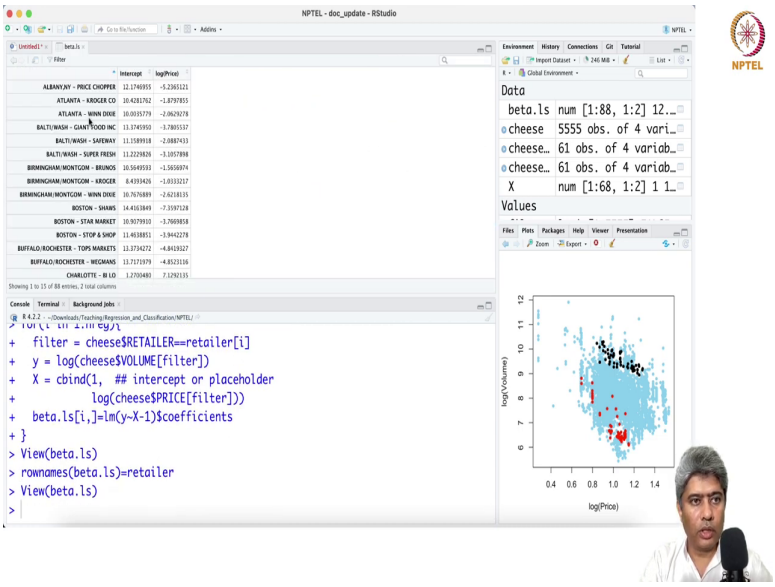
So, what I am going to do these are the two things. So, intercept and the coefficient of log Price these two will be stored will calculate and store for all 88 retailers separately. So, I am going to fit 88 separate model for this. This is Los Angeles Lucky I will have a separate line this is Albany you know price chopper I will have a separate line. So, we will fit a separate line for each of them. So, what we are going to do now is for i in 1 is to nreg.

(Refer Slide Time: 15:24)



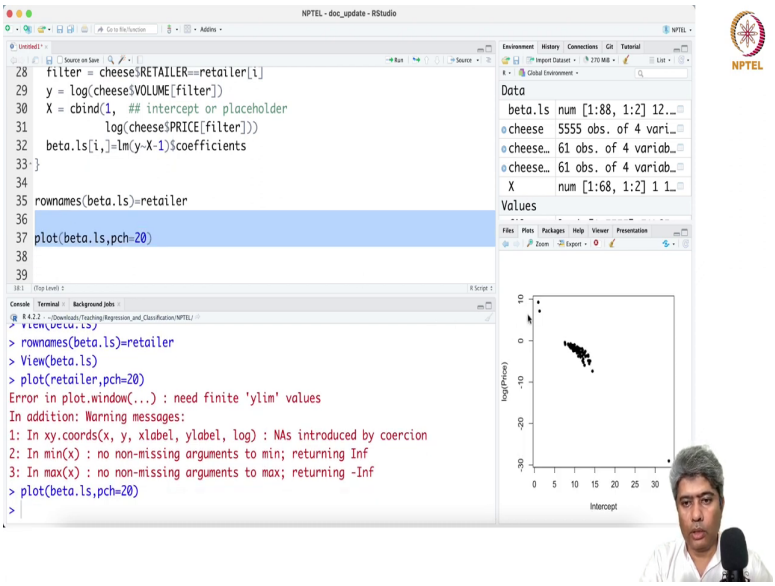
Filter equal to cheese RETAILER retailer i and then what I am going to do y equal to log of cheese dollar VOLUME filter. And X equal to cbind 1 comma this is essentially intercept or placeholder. And then this is log cheese dollar PRICE filter, correct, and then beta dot ls i comma lm y tilde X minus 1 dollar coefficients.

(Refer Slide Time: 17:19)



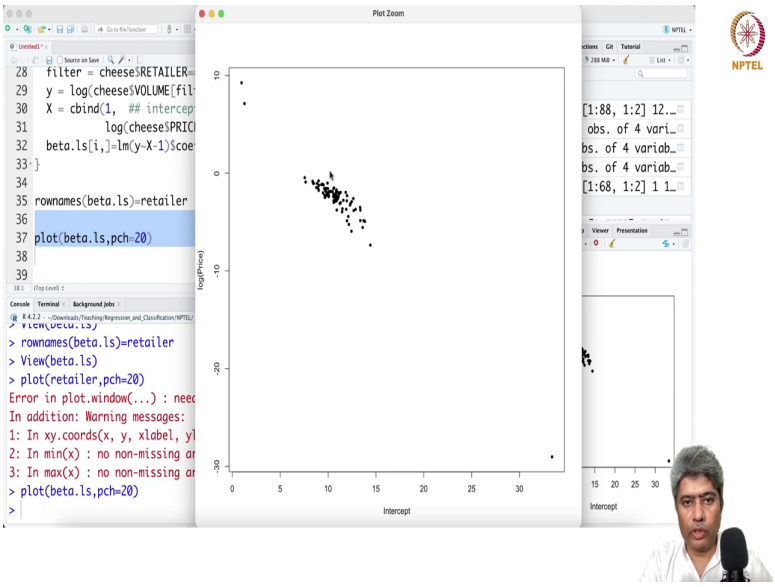
Now, if you just say beta dot ls these are the coefficient values of intercept and log price coefficient of log price. So, rownames equal to beta dot ls equals to you can put retailer let me put retailer. So, these are the names that we have for each guys you have the retailer names.

(Refer Slide Time: 17:55)



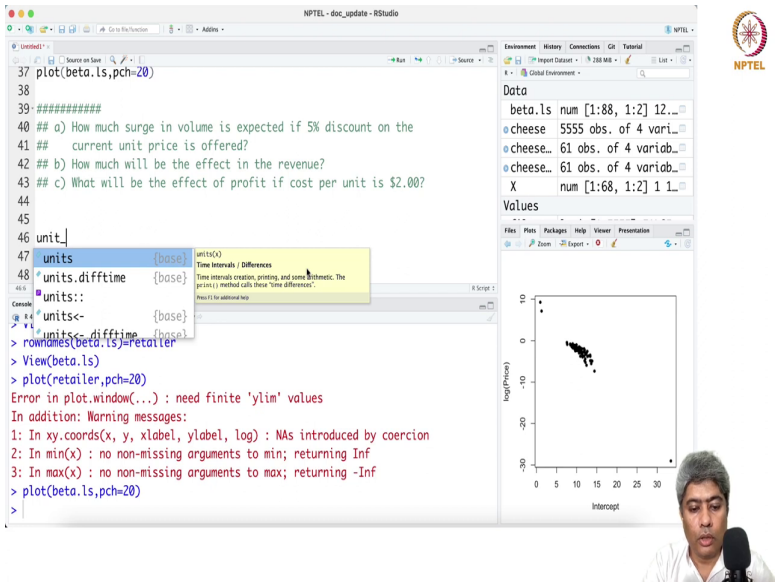
Now, if I just put plot beta dot ls pch equal to 20 sorry this is beta dot ls, yeah.

(Refer Slide Time: 18:16)



So, if you just so, these are the typical coefficients there are two outlier and there is one outlier here. There are 3 two different outliers. So, we have to do a little bit of outlier analysis later.

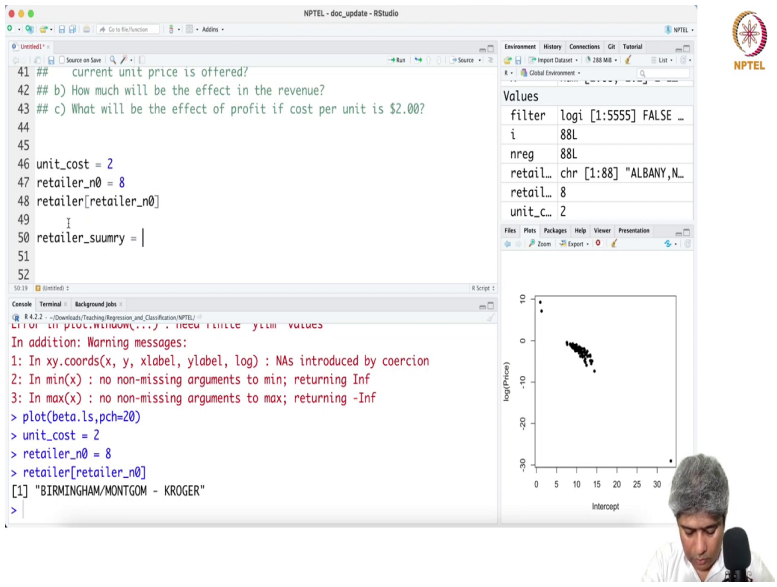
(Refer Slide Time: 18:34)



So, now what I am going to do? I am going to try to answer following questions, ok. How much; first question how much surge in volume surge in volume is expected if 5 percent discount on the current unit price on the current unit price is offered? Ok, so, this is the if I give 5 percent discount on the current unit price then how much surge I would in terms of volume sales I can expect.

Then how much will be the effect in the revenue how much will be the effect in the revenue? And what will be the effect of cost profit if cost per unit is dollar 2? So, these are the three questions we are going to discuss, ok. So, the first thing is unit cost is 2, right, unit cost is 2.

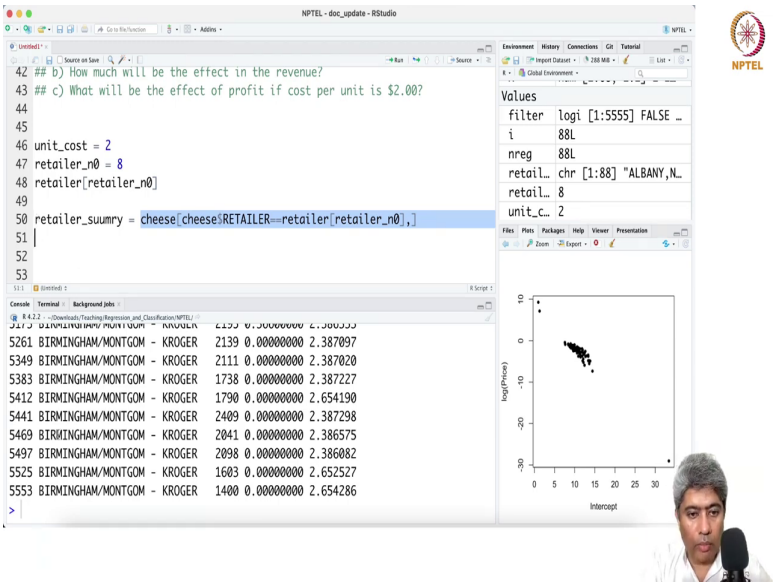
(Refer Slide Time: 20:59)



And let us take a retailer number is some number maybe 8, ok. Now, retailer number 8 who is this guy, ok? Birmingham Kroger, Kroger is a big chain and in Birmingham Montgom Montgomery district or county Kroger retailer this is the retailer they are talking about.

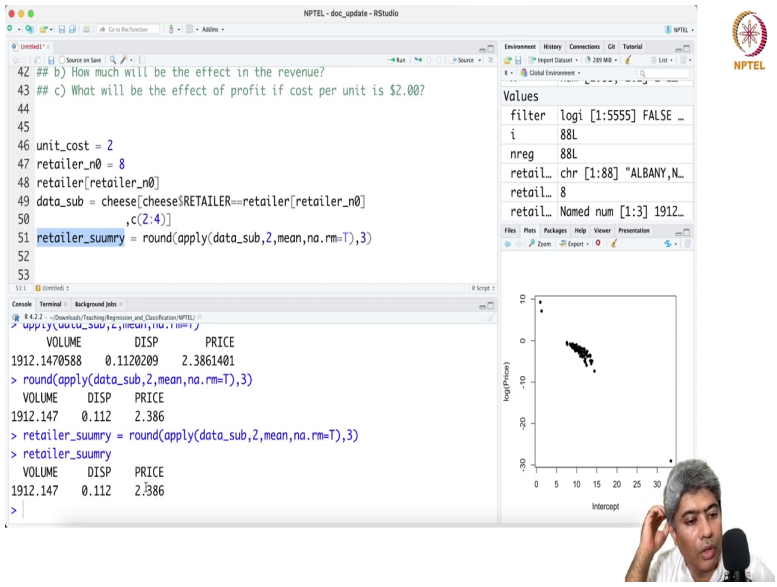
So, let us have a look into the retailer summary typical what is the summary statistics of this particular retailer ok. So, the first thing I am going to do is I am going to take this guys whoever this guy is retailer with retailer number, right.

(Refer Slide Time: 22:27)



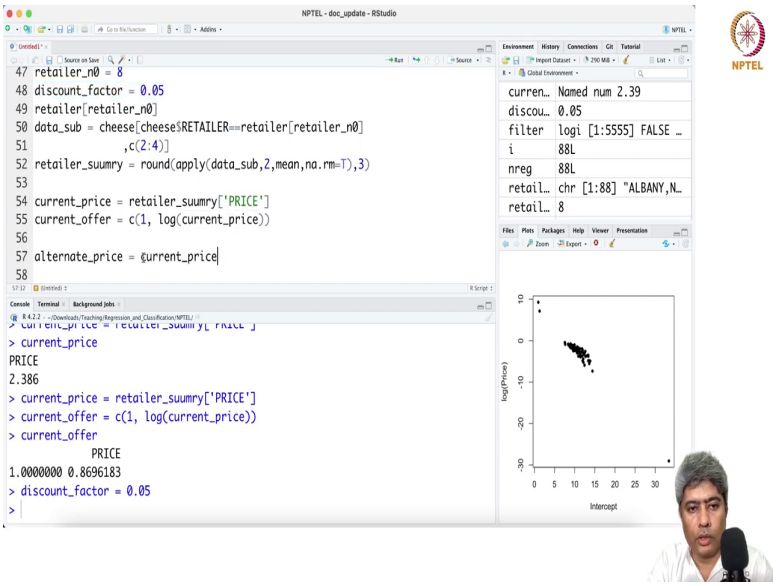
And in that particularly I am interested in 2 and 4 the volume and price. Particularly I am interested in 2 is to 4, volume, display and price; why not I apply? So, this is the so, first let me just take a data sub if I just take data sub data set. So, this is the data set.

(Refer Slide Time: 23:25)



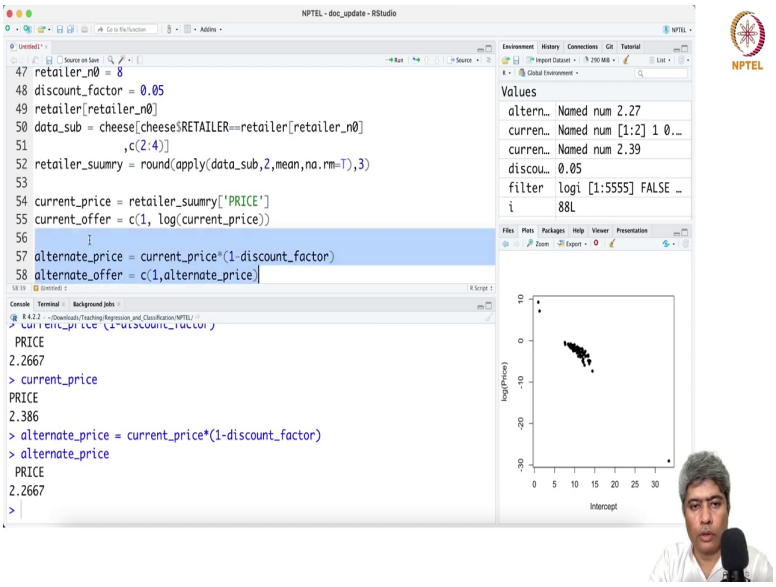
And data set comma data sub comma 2 comma mean na dot rm equal to true, ok. And then if I just round up to 3 decimal places. So, that gives me these values. So, that means, this particular retailer is selling at 2.386 on an average the price that they are selling and volume is they selling about 1900; 1900 1912 many units per week.

(Refer Slide Time: 24:31)



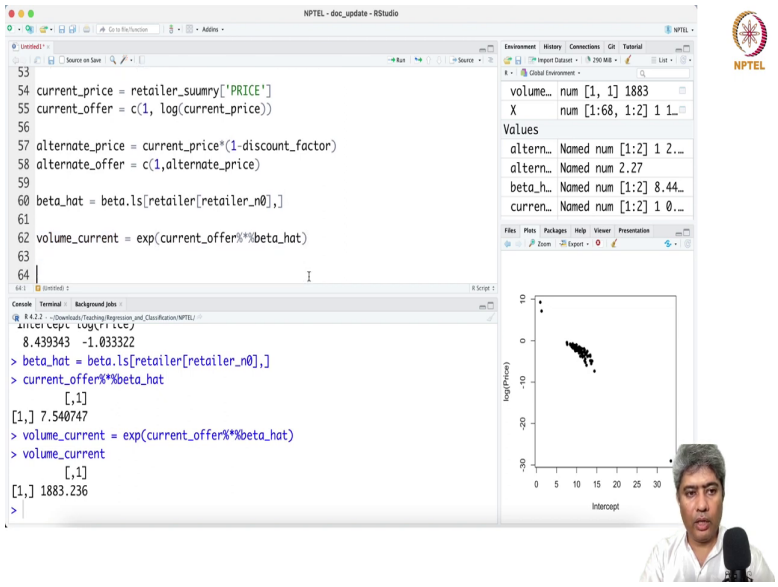
So, let me take the current price is this price the average price as the current price. So, this is suppose current price, ok. If this is the current price and then I will call say current offer equal to 1 comma log equal to log of current price. So, this is my current offer. Similarly, we have to offer make a alternate price, which is say discount we have to define a discount factor discount factor. So, maybe 0.05 5 percent discount I am giving.

(Refer Slide Time: 26:02)



So, whatever the current price times 1 minus whatever the discount factor. So, to so, the current price is 2.386 after giving say 5 percent discount alternate price is 2.267. Now, so, my alternate offer alternate offer is 1 comma alternate price, ok.

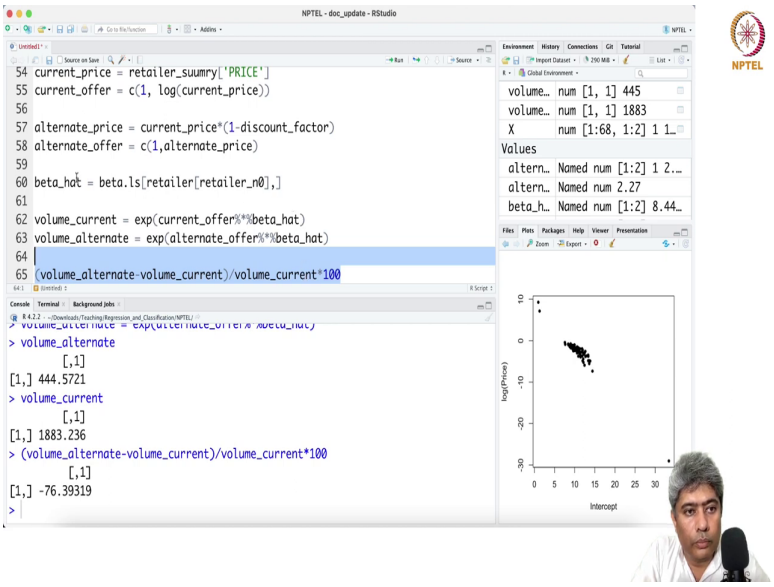
(Refer Slide Time: 26:53)



So, the first thing now I have to do is beta hat equal to beta dash beta dot ls retailer equal to retailer number. So, if I just take so, this is my coefficient for this particular retailer, ok. Now, volume if for this current offer and alternate offer what would be the volume expected sales of the volume, that is, clearly current offer beta hat percentage star percentage beta hat.

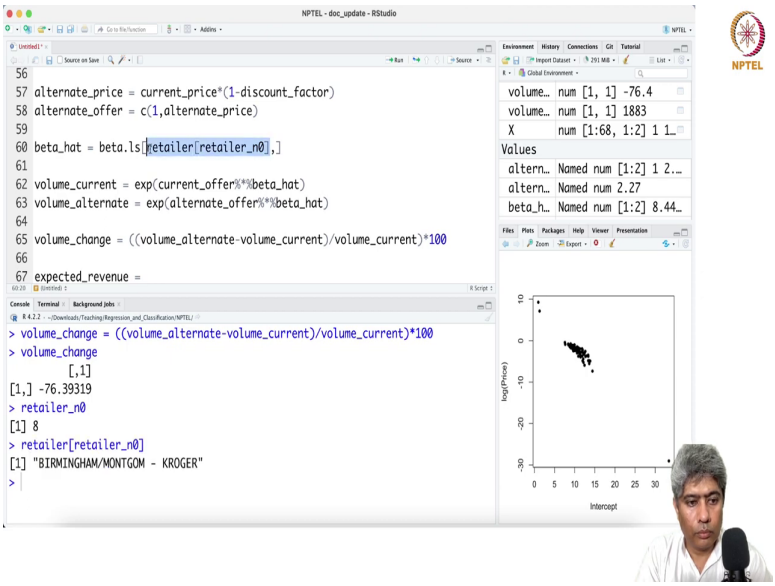
And then you have to make expected because remember that we are fitting log volume versus log price. So, volume current sales, ok. So, this is 8 you are expected to make a sale of 1883.

(Refer Slide Time: 28:25)



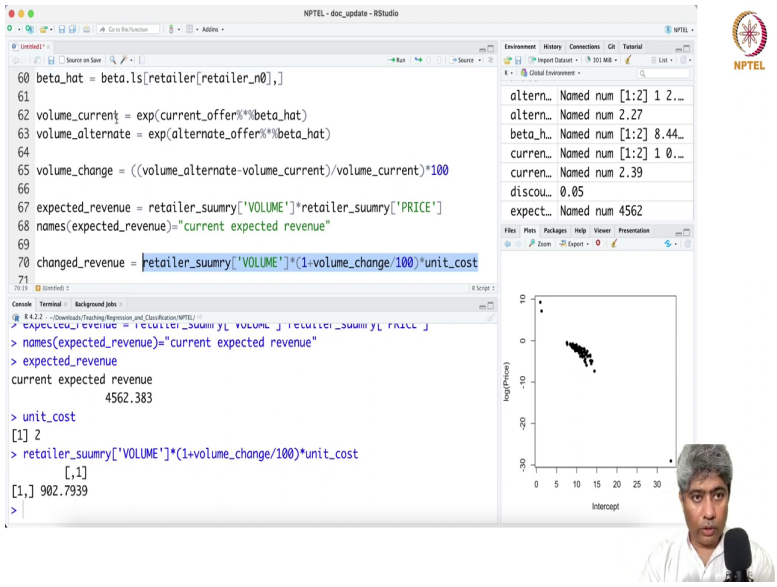
Now, similarly I can also make prediction for the alternate offer so, alternate offer, ok. So, this is my current offer this is my alternate offer, ok. So, alternate offer minus current offer divided by volume of current offer times 100, ok.

(Refer Slide Time: 29:48)



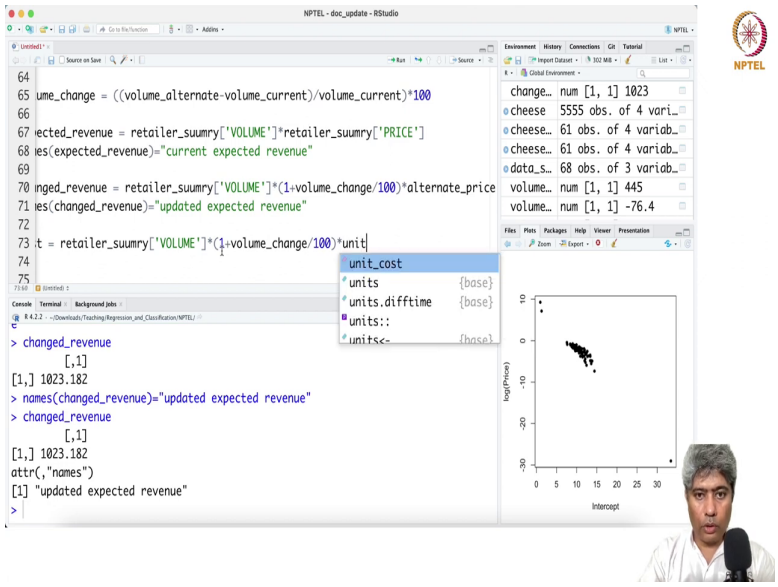
Let me just do one thing. Percentage star percentage I think I am making 7.75, right. And 2.67 this seems very different alternate minus volume current, ok. So, let me just take this volume change. Maybe there will be a serious drop in the volume. I think that we can take it as at face value. And we can compute the expected revenue expected revenue, yeah, ok.

(Refer Slide Time: 31:26)



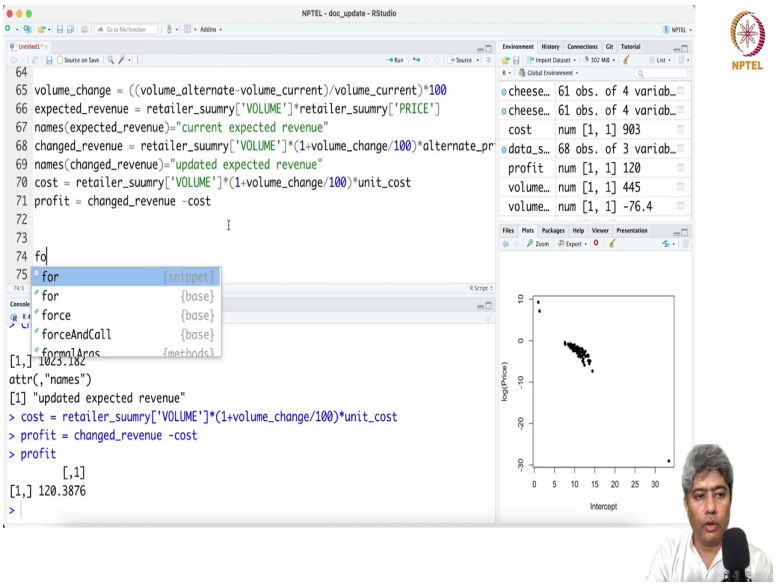
Expected revenue is whatever be the retailer summary and that is the “VOLUME” V o l u m e Times retailer summary of the PRICE, right. So, this is the volume that you expect, ok. So, we can call it current expected revenue, ok, ok. Now, what we will do is we will check what would be the changed revenue. The changed revenue will be first we have to take the this times 1 plus whatever the volume change divided by 100 times unit cost right. That will be the sorry that will be not unit cost that will be alternate price.

(Refer Slide Time: 33:33)



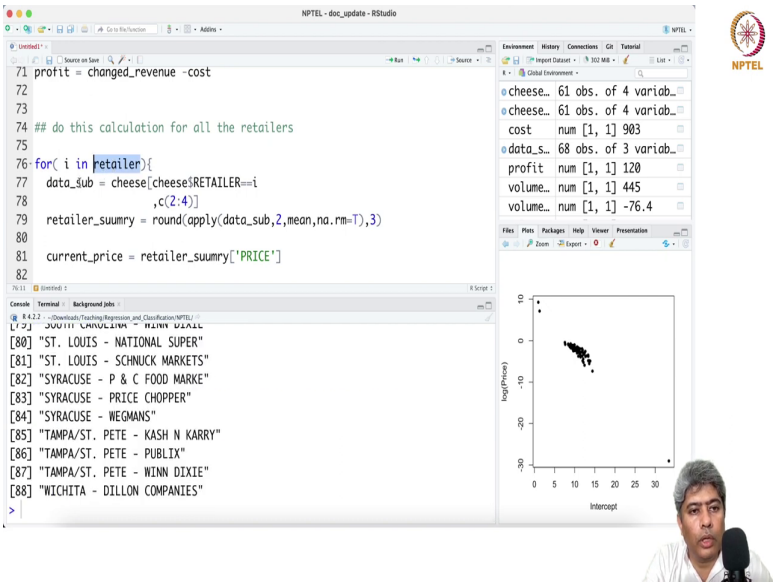
That is the X that will be the so, the expected revenue is 4562 and the change revenue is sorry 1000 something 1023 right. So, this is sort of a names of changed revenue is “updated expected revenue”, ok. Now, total cost will be cost will be whatever the “VOLUME” times this changed percentage times unit cost. So, this will be the total cost.

(Refer Slide Time: 35:05)



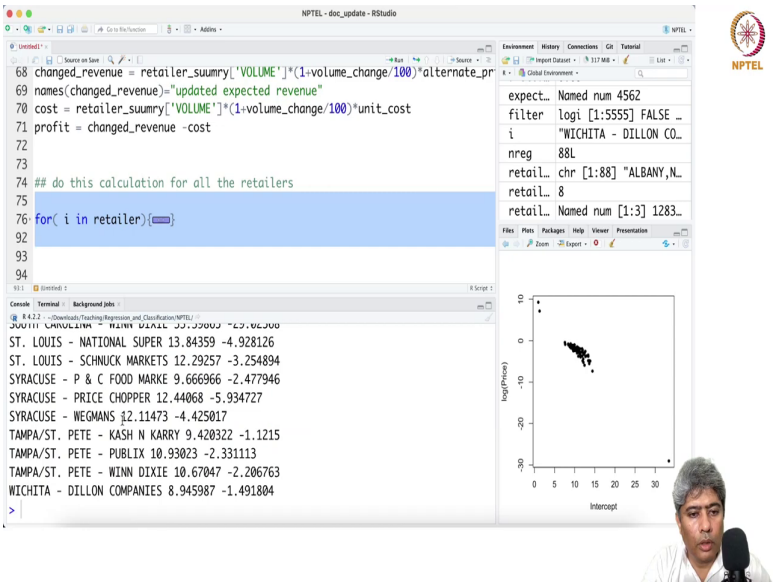
And changed revenue is this. So, this will be the total cost and profit would be profit would be changed revenue minus cost. So, the total profit will be 120 dollar, ok; from this one particular guy. So, let me now what we will do we will put it into sort of a we will do have to do this calculation for all the retailer.

(Refer Slide Time: 35:52)



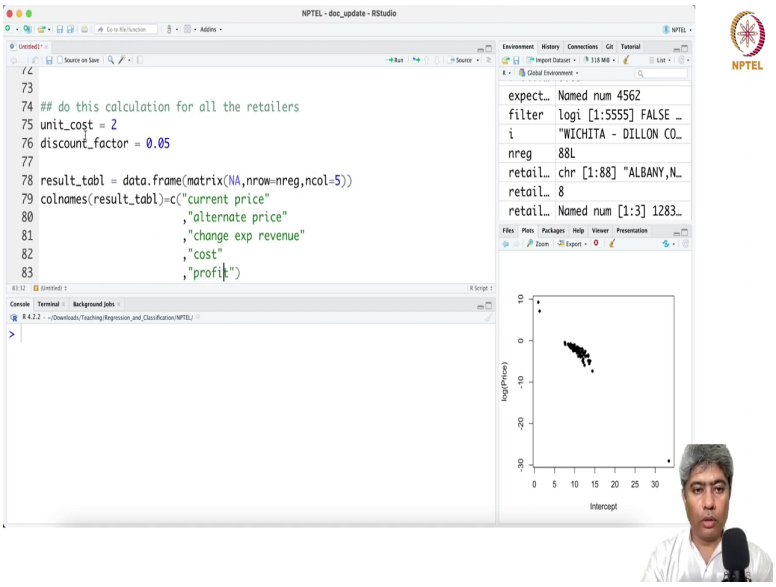
For do this calculation for all the retailer. For i in retailer ok. So, the first thing I have to do is perhaps data sub retailer sub i right, ok.

(Refer Slide Time: 37:06)



So, retailer summary and let me just run this part. Let me see if this is coming properly, alright i comma, yeah. So, it is alright that is happening. That is fine. So, now what we have to do, we have to just take the current price; retailer summary is done, right. And then we have to just take current price. I so, let me just take if it is doing well, yeah ok, fine.

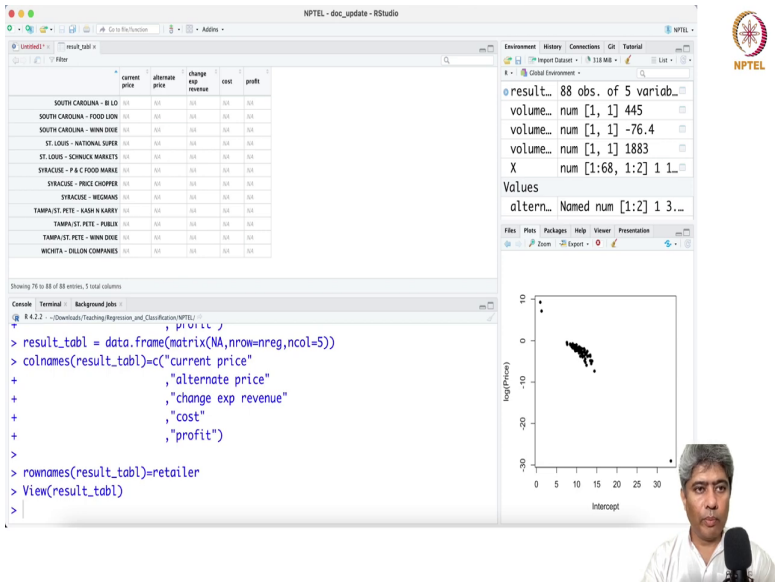
(Refer Slide Time: 38:30)



Now, what I am going to do is I am going to calculate the current volume and the expected volume. This, these things will be exactly same actually. No change. So, this will be, there will be no change. But so, now, what we need, we need essentially, we I need a discount factor. So, discount factor is here actually, you know.

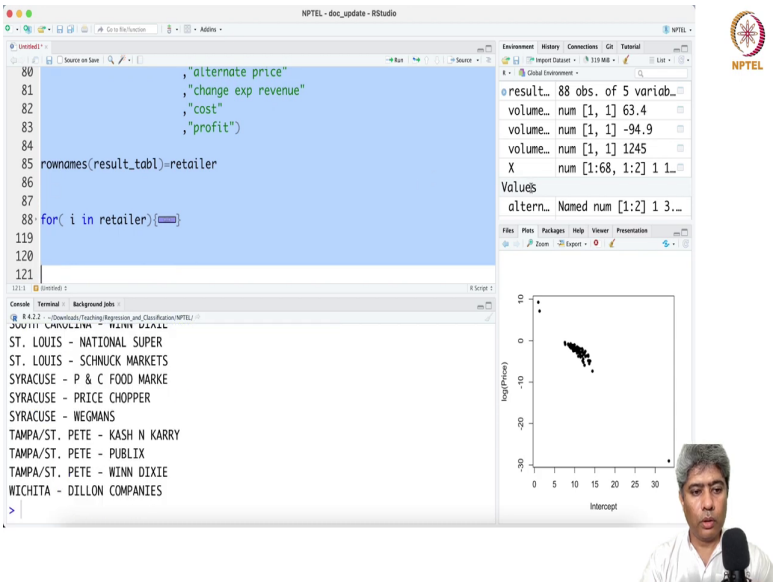
So, let me just take these things, two things. So, I do not need this ha. Unit cost is 2, retail. And then essentially, I need result table, ok. Result table data dot frame matrix NA nrow equal to nreg and ncol equal to 5, ok. And colnames equal to result table equal to c. First is “current price e”. Second is “alternate price” price. Third is changed in “change in expected revenue”. How much change we are expecting total “cost” and final “profit”, ok, alright.

(Refer Slide Time: 40:47)



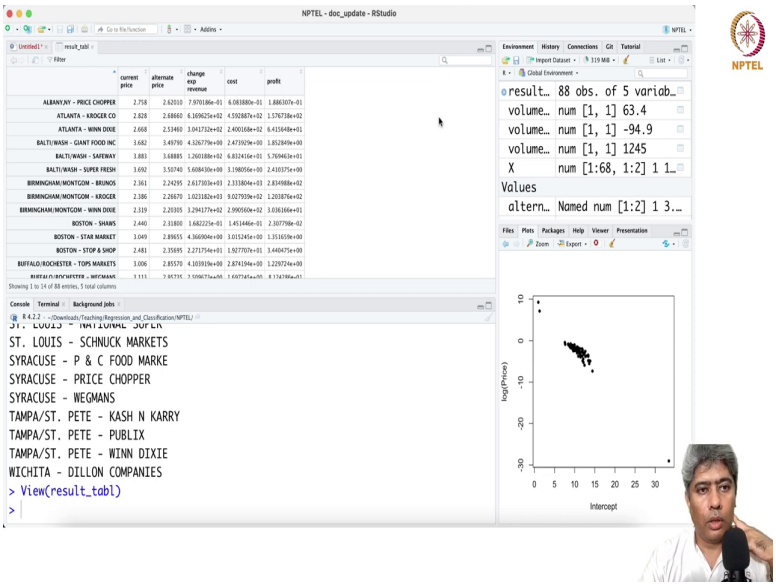
So, now if I just run it, and result table equals to, I have to do one more thing. We have to do that is col names. I have to do row names also, rownames. Result table equal to retailer, right, ok. So, if I just now define it, you can see all these names are there. I have just created a place holder for that.

(Refer Slide Time: 41:35)



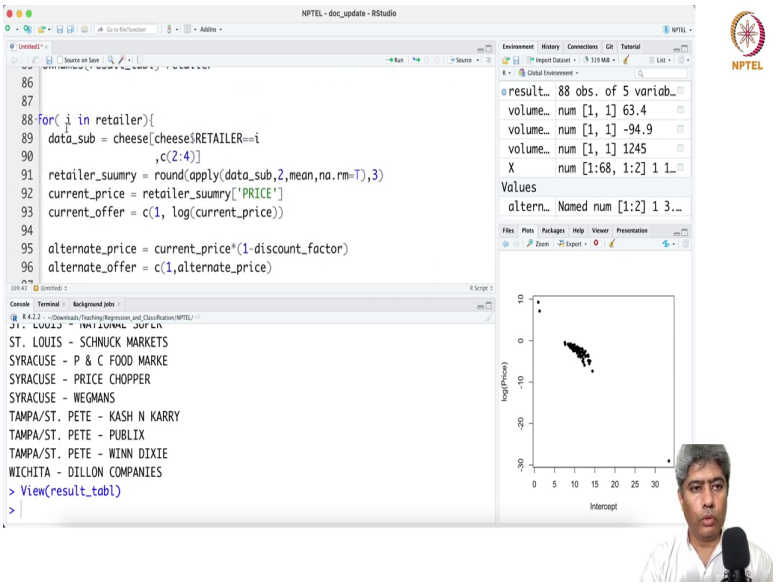
Now, result table equals to i comma c. So, what I have to now fit is current price. What is current price? This is the current price. Then alternate price. Then current price, alternate price, change in expected revenue; so, changed revenue, then cost because of the change offer and what is the final profit ok. So, let me see alright. Let me just run it, ok.

(Refer Slide Time: 42:38)



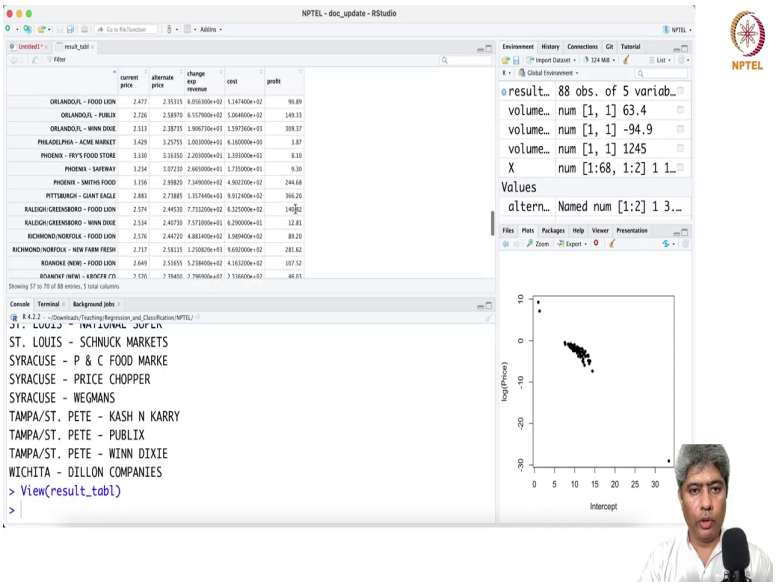
Now, here is the things, ok.

(Refer Slide Time: 42:53)



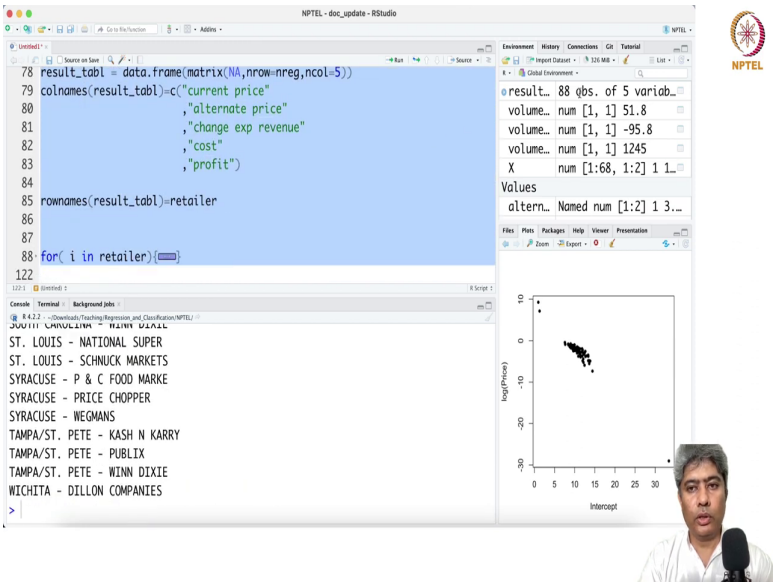
Do one thing boss what I am going to do is round it off to cost I am going to make it round it off to 2 and profit also comma 2 and changed revenue equal to change round, round it off to 2, ok.

(Refer Slide Time: 43:42)



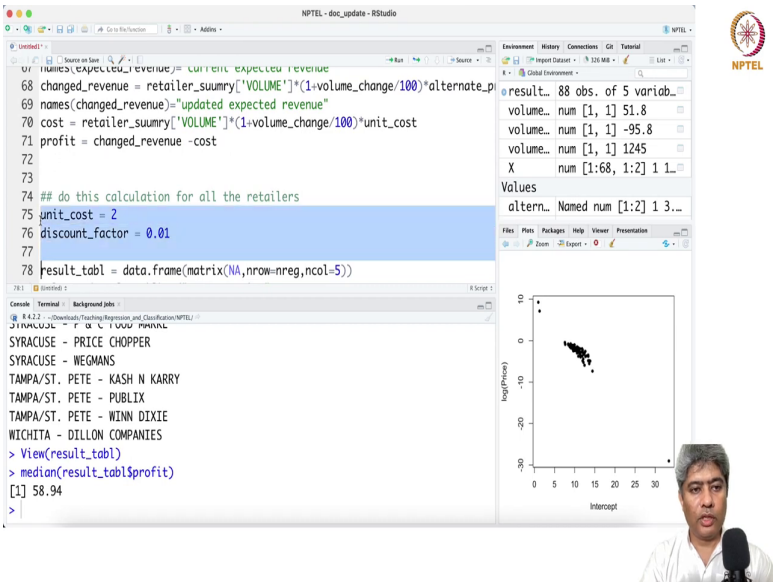
So, let me just run it once more. Looks like fine ok. Here is the for CHARLOTTE - BI LO. There is a huge profit that are coming up. BI LO is also giving us quite a bit of huge profit.

(Refer Slide Time: 44:00)



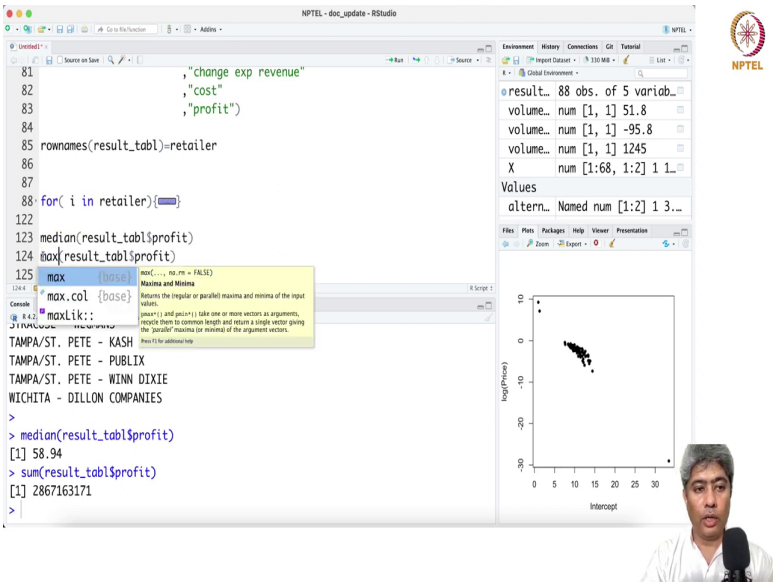
So, there is something to do with the Bi Lo Charlotte and South Carolina. Why it is, I do not know, maybe there is must there could be some issue with the coefficient. Remember that maybe these are the these one of the outlier coefficient that is getting reflected. But if you now, instead of 0.05, if I just give 1 discount, well, let us see what happened.

(Refer Slide Time: 44:38)



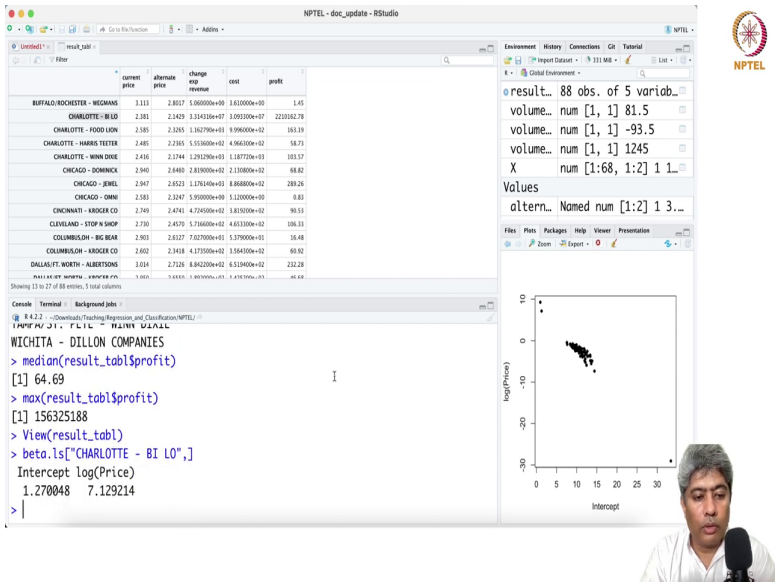
You see ok. So, what I am going to do is result table is dollar profit kind of median profit. What is my median profit from a 58 dollar? Ok so, that makes sense perhaps if I just give 1 percent discount, my median profit is 58 dollar.

(Refer Slide Time: 45:12)



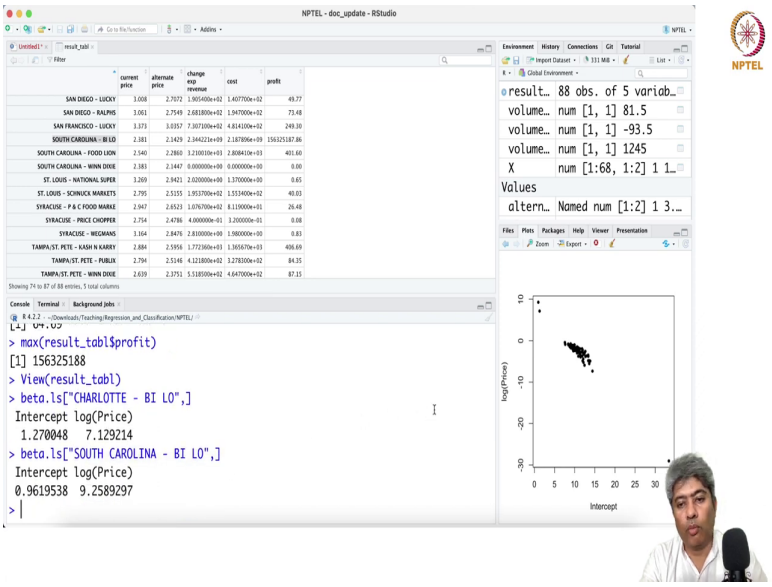
And total profit of if I give this to everybody, everybody, then this is my huge profit or maybe I am just getting I can give, you know, max profit that can I get is this is a big number. This is definitely not right, something wrong. And then if I give say 10 percent profit, 10 percent discount, then what happens is median profit is 64 and maximum profit.

(Refer Slide Time: 45:52)



So, I think I am pretty much. So, this particular guy, whoever this is, this is Bi Lo Charlotte Bi Lo. Let me just take copy and beta dot ls, yeah.

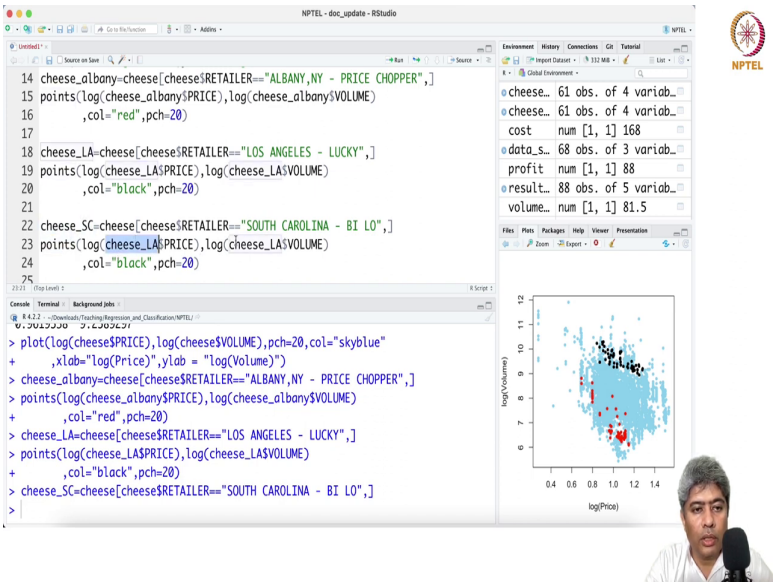
(Refer Slide Time: 46:17)



So, 1.27 is somewhere here, no, yeah. And 7 point something on the log price. So, this is definitely one these two guys. Here is, I am pretty sure these two guys here is giving us the problem. So, if I just put it there, yeah, 0.96 and 9.5. So, my log price, the coefficient is 9.25.

So, most of my coefficient are negative, you see, most of the coefficient, but this guy is giving me very large positive coefficient. That means and intercept is 0.96, which is near 1, whereas, most of the intercept is near 10, at least above 5. So, this tells us this two estimates for South Carolina Bi Lo may be. What I will do, I will do a plot.

(Refer Slide Time: 47:32)



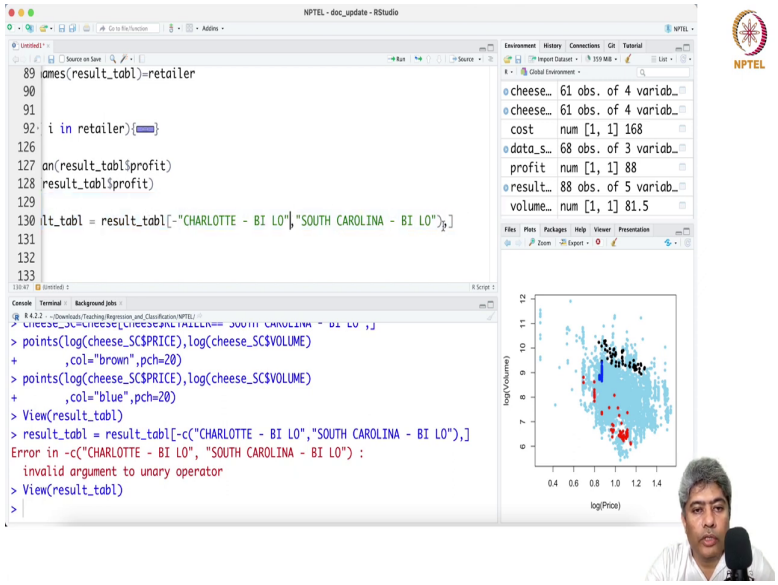
So, in this, let me just draw a plot. So, here is the plot. Here is the Chinese cheese data set Albany Chopper. This is our LA Los Angels. This is SC Bi Lo South Carolina. And this is "SOUTH CAROLINA - BI LO", right. And then if I just plot that guy, cheese SC versus cheese SC, and instead of "black", maybe "brown" ok.

Let us see where this guy is. You see, this is the problem. All points are here and that is why. So, there is no diversity in the points. All points are plotted here. That is the issue. So, all I can just say "blue". Let me see if I can show you guys, yeah.

See, all points are in one line. And as a result, probably the line is very sharp. So, these kind of weird point, set of points, if you get, you cannot do much effectively, you cannot do much you have, yeah, one possibility is you can try some Bayesian; hierarchical Bayesian models, but that is bit out of the question here. But we will see if we can fix it. But I hope you

understood what is happening. So, if we even if we decide to, ok, let us drop these two guy, ok. We have this, if we drop these two guy for the time being from the.

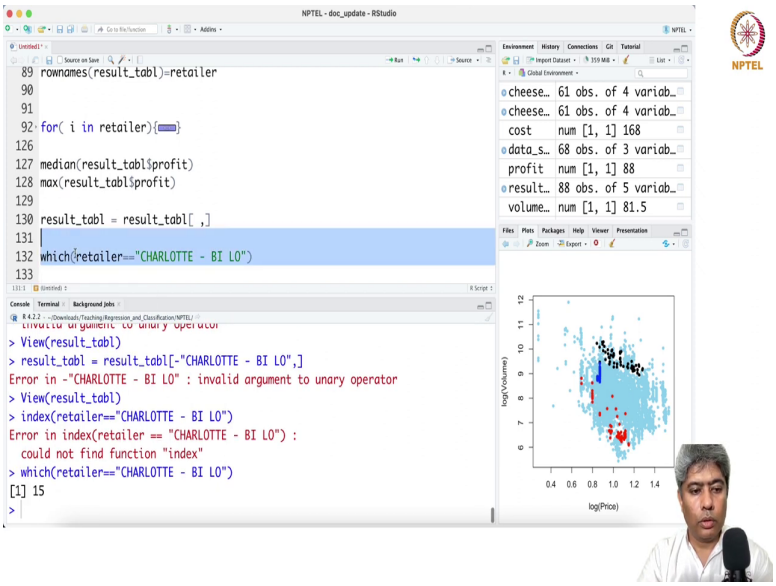
(Refer Slide Time: 49:44)



So, so, what are those two names? Let us see. One is CHARLOTTE will BI LO. And the other guy is other person, other one where we got a terrible estimate is South Carolina Bi Lo. So, this, two retailer I think have some issues. So, what we are going to do? We are going to result table, minus these two person result table.

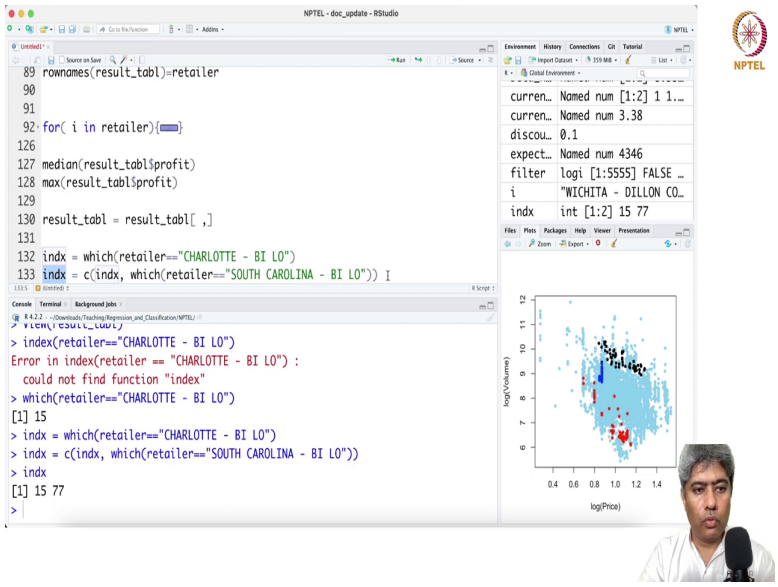
Now, if I just run this, I am not sure if we can drop it in this way, alright. So, it is not looks like it is not mutable ok. If I just drop this guy like this, invalid line argument, ok. What I can do is. So, I just, why not I do this? Just for the time being, I can come up with a much better index, i n d x of (Refer Time: 51:58) tab retailer.

(Refer Slide Time: 51:51)



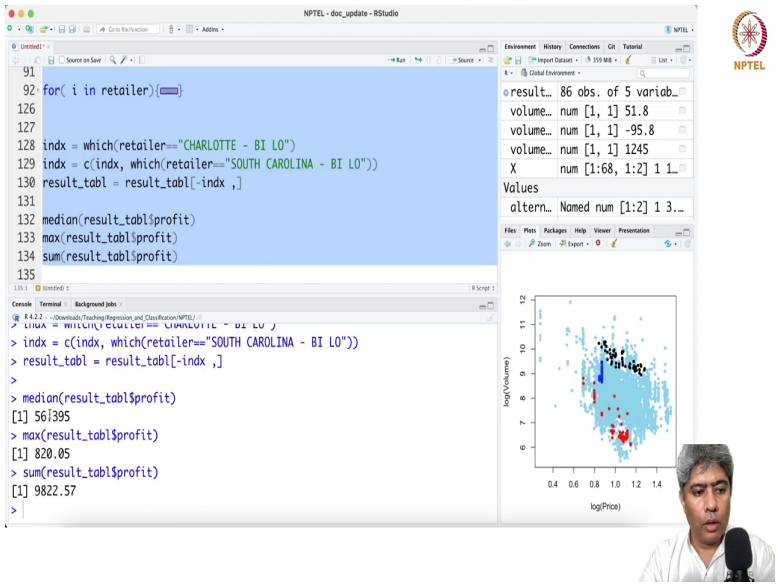
Equals to, CHARLOTTE will, “CHARLOTTE - BI LO”, which 15.

(Refer Slide Time: 52:32)



And so, x and then equal to c index comma “SOUTH CAROLINA - BI LO” so, these are the two cases and then result table minus index if I just say that. And now if you see I have dropped these two guys kind of.

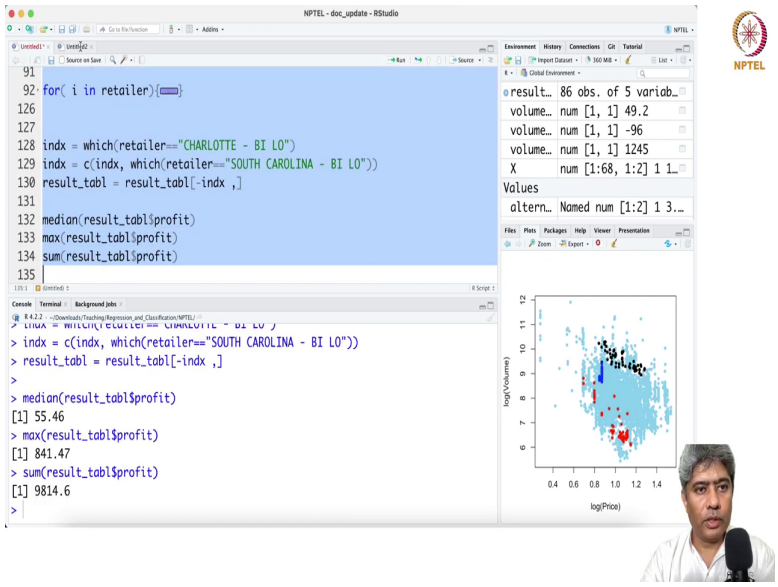
(Refer Slide Time: 53:27)



Now, what is happening is if I now median profit and this is the max profit and sum if the result table and dollar profit is this, right. Now, if I just go and say 0.01 percent if I give 1 percent discount then median profit is 56. Let me just start with no discount. So, I will not give any discount.

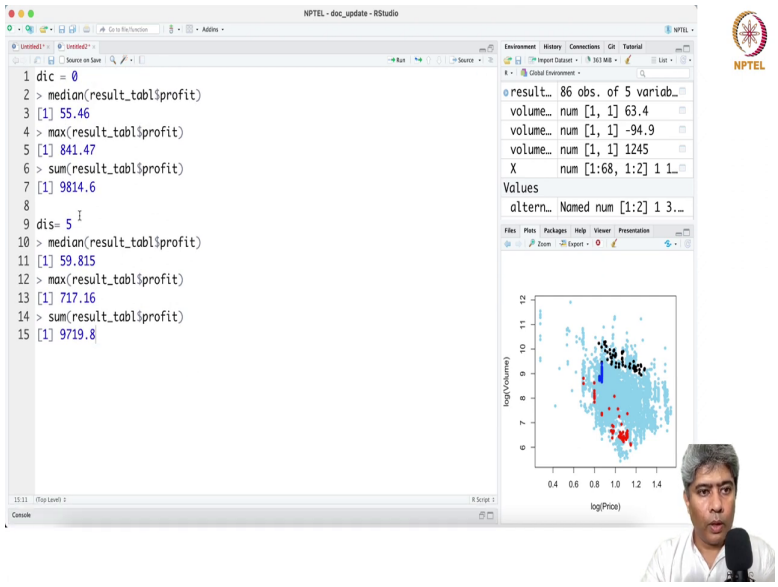
Current price, my kind of median profit is 55; max I can earn 841 from one particular retailer. And total profit can be expected to 9,014 dollar. Now, if I give 5 percent discount then it is median profit sort of increased 59.717. Actually, let me just do one thing. Let me not print this line otherwise it is creating lot of (Refer Time: 55:05).

(Refer Slide Time: 55:11)



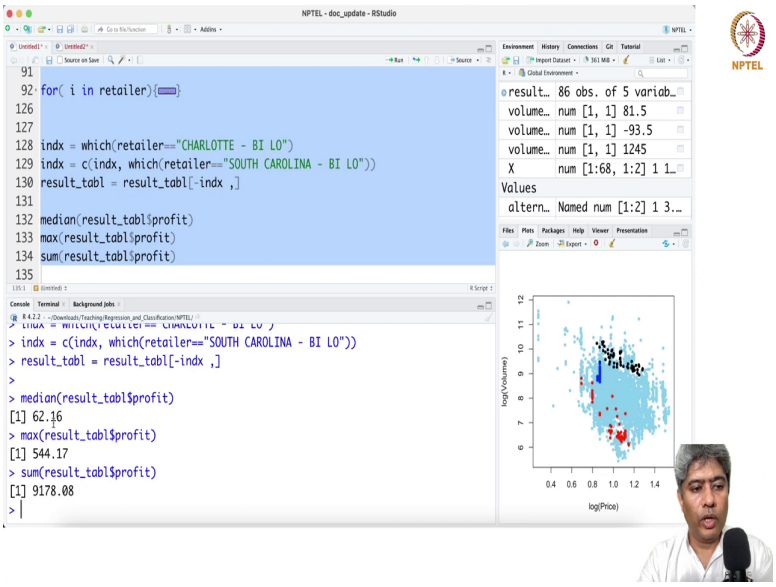
So, so, let me just create another thing, ok. If I do not give discount then that is what happen so, discount 0, discount.

(Refer Slide Time: 55:24)



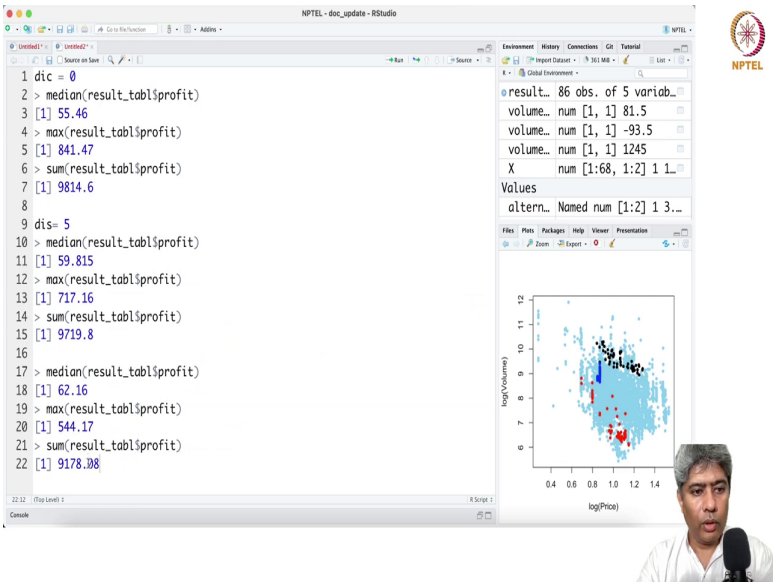
Discount equal to 0 I get median profit of these guys. If I give 5 percent discount if I give 5 percent discount then I get this. Discount is 5 percent. Now, my total profit reduce a bit.

(Refer Slide Time: 56:07)



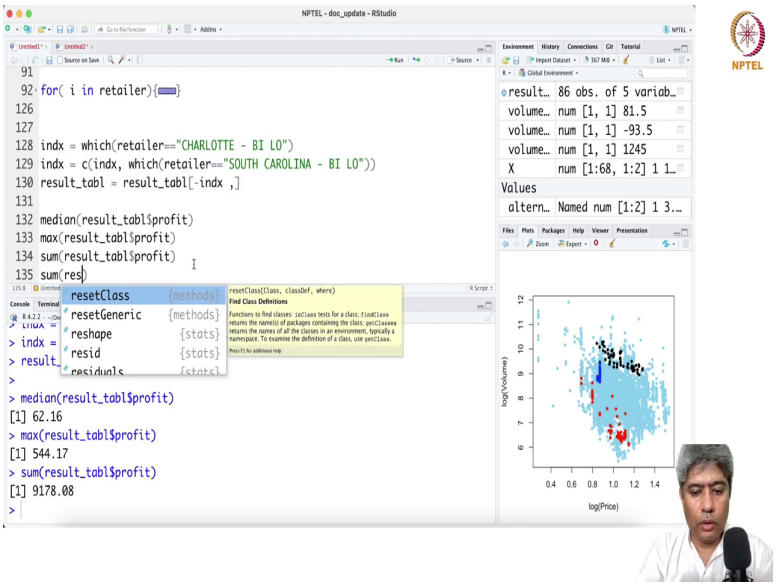
And then if I give 10 percent discount my median profit increases probably 62.

(Refer Slide Time: 56:28)



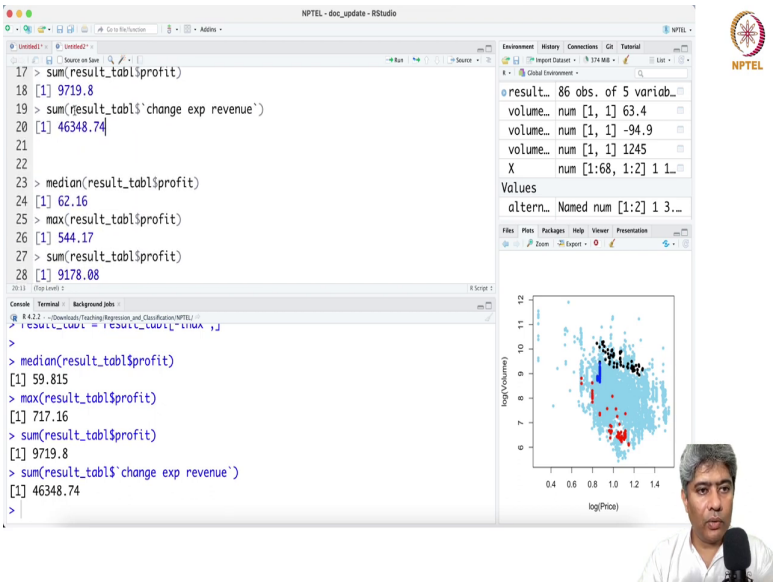
But total profit drops. So, total profit drops, max profit drops, but overall drops.

(Refer Slide Time: 56:49)



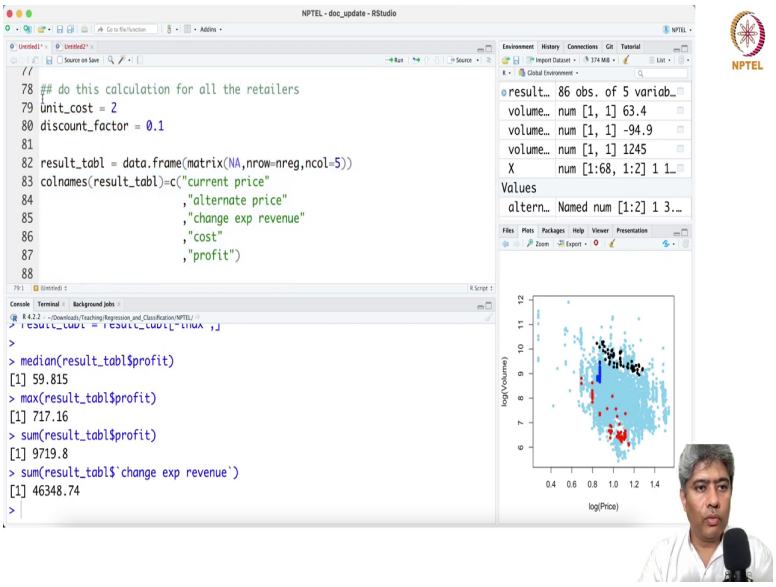
And then if I just say sum of result table dollar expected revenue "change of expected revenue". So, at 0 discount what happens is my total revenue is 39,000 dollar when I am giving 5 percent discount, when I am giving 5 percent discount my total revenue is 46,000 dollar. My total profit, my total revenue goes up from 39,000 to 46000.

(Refer Slide Time: 57:49)



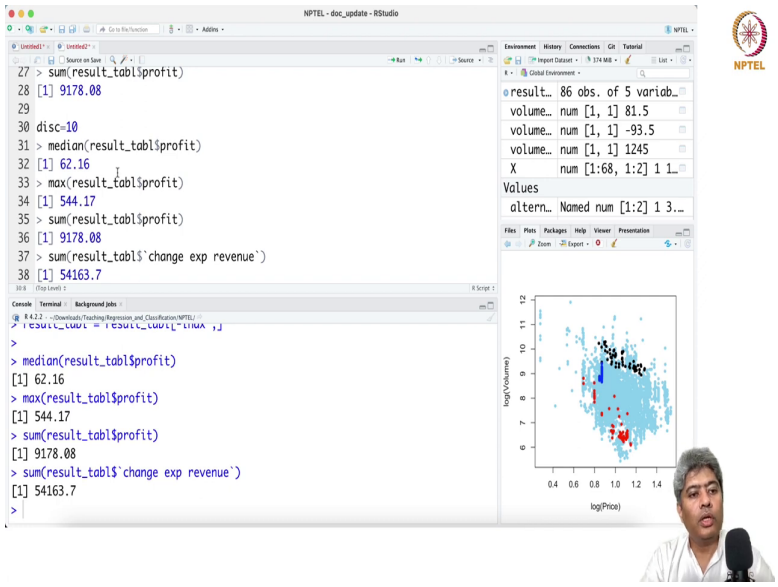
But total profit came down from 9,800 to 9,700, but median profit raised by 4 dollar, ok. Median is the 50 percent at a lower level overall profit goes up and then what happened at when I give 10 percent discount at 10 percent discount.

(Refer Slide Time: 58:16)



So, revenue I expect will go up further, but profit may not go up. So, the median profit again goes up, but total profit goes down, but the revenue goes up quite significantly; discount is 10 percent.

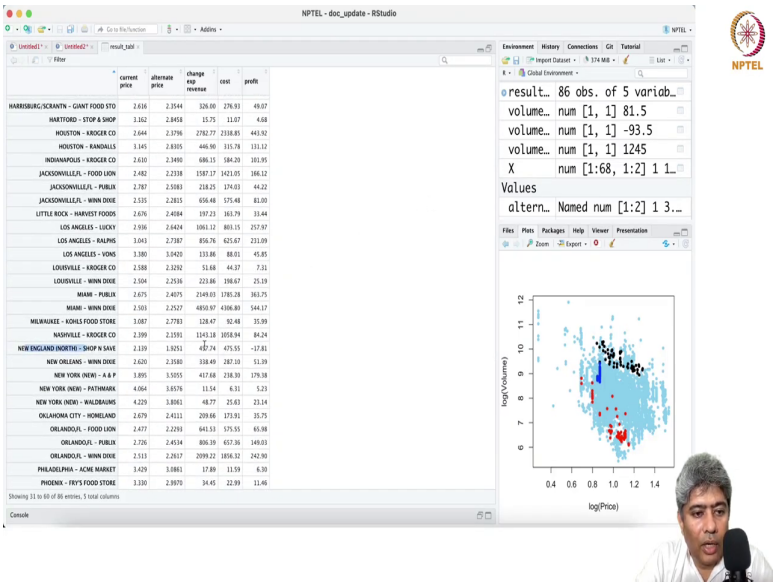
(Refer Slide Time: 58:36)



So, total profit going up to 46,000 to 54,000. So, total revenue stoods up, but total profit goes down marginally. So, sometimes the company wants to show the, you know, grab the market share and in order to, if they see that if my total profit is not taking hit, but my overall market share is going up, why not I give some kind of promotional activity by giving discount.

Now, based on the, this is like giving overall, but at the same time what people can do they can check where overall profit going up and where the profit is not going up. The where the profit is not going up they can like.

(Refer Slide Time: 59:31)



Here in this NEW ENGLAND NORTH SHOP N if the profit is negative then they should not give the discount here, ok. Or anywhere if the profit is less than certain value they will not get profit discount; otherwise, they will get a discount. So, this kind of discriminatory pricing practice sometimes helps companies to improve their profitability.

So, I will stop here. I hope you enjoyed that how simple predictive model and regression analysis can help companies and industry to implement dynamic pricing and improve their profitability, improve the total revenue and maintain the cost and pay. We will continue with such more data analysis in the next few videos.

Thank you very much. See you in the next video.