Business Analytics for Management Decision
Prof. Rudra P Pradhan
Vinod Gupta School of Management
Indian Institute of Technology, Kharagpur

Lecture – 10
Exploring Data and Analytics on Spreadsheets (Contd.)

Hello everybody this is Rudra Pradhan here, and welcome you all to business analytics for management decision. We are still in second unit and today's discussion is on exploring data and analytics on spreadsheet. In fact in the, in this unit we have already discussed couple of things in the previous four lectures. So, the idea is that you have to understand the data structure then the kind of data entry data processing and that is how we have already discussed in details about the huge of you know excel that is excel spreadsheet.

So, now, once you have a data to solve some of the problems you know business related problems by using business analytics tools. So, it is the mandatory requirement that you know you have to have data first, understand the data get insights and then you think about some kind of advanced analytics tool to solve the business problem. But now having the data you may not to get directly some insights and you cannot it is not actually good to use without knowing anything about the data for a particular variable or for a set of variables.

So, of course, the problem can be actually summarized in the form of models and the model can be summarized again with respect to some of the variables and then the analysis will be done on the basis of available data for these variables. This is the basic kind of framework altogether. Now once you have actually data for all these variables and for a particular problem. So, first and foremost requirement is to understand the data corresponding to each variables right and that is how we have already discussed so many things and I frequently or constantly instructing you that you should know more about excel spreadsheet because without knowing excel spreadsheet it is very difficult to use some kind of statistical software to solve some of the business problems or to analyze the business problem. Because excel spreadsheet will give you some kind of exposure or it will give you some kind of structure through which the techniques can be used very

easily and then you can predict the particular you know problems as per the e problem requirement.

So, we have already discussed various ways to understand the data why view options, plotting, that is nothing but called as data visualization. So, we have discussed. So, many actually graphical visualization process starting with you know bar diagram line diagram pie diagram and so on.

So, these are all you know where you can you know visualize; that means, actually to get some kind of insights from the data you have a plenty of options. So, just you know same data you can actually plot in a different options or quantify in a different ways then by default you will get some kind of insights. Now, on the idea is that you know, the same data for same variables, but still we are trying to understand or trying to get insights in various angles by using some of the quantitative tools and graphical tools the only idea is that. So, that our insights or our inference from the data should not be kind of I mean some should not something wrong.

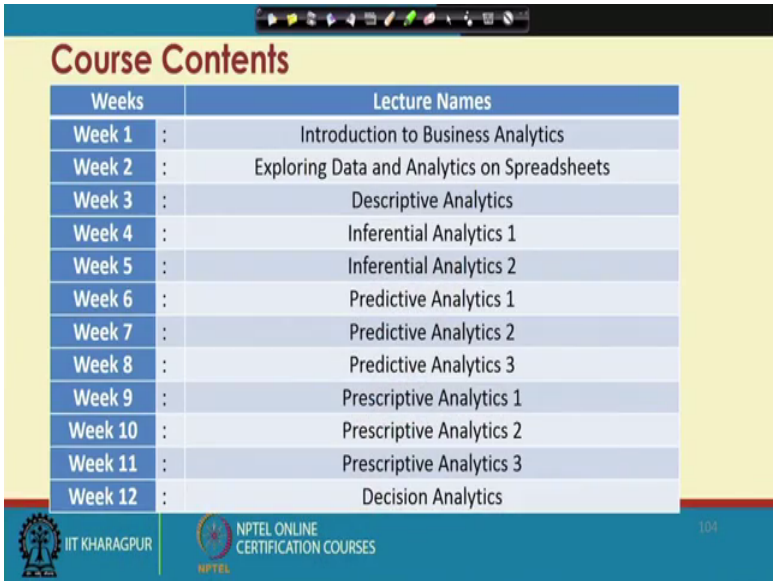
So, if you are getting something else then you know other options to check all these things will give you some kind of consistency in statistical world we call as you know kind of robustness kind of things because same things we are you know same requirement you are you know checking through various angles right. So, if actually your you know inference or insights are you know perfectly then by default by checking any angles in quantitative angles or qualitative angles the results will be uniquely same. So, it will not you know deviate or it will not change you know significantly.

So, it is always better you know same problem or whether to analyze the data or whether to understand the data. So, with you know at least you know more than one kind of tools or a structure through which you have to understand the situation or to analyze the situation. And in these particular lectures we have actually given you enough kind of an exposures by using you know excel spreadsheet. In fact, you know we have a many more kind of advanced softwares through which same inspection or same insights can be obtained for instance.

In the last lectures we have discussed you know excel spreadsheet through the graphic kind of in visualization, but I will sight to another example called as origin package. If you in the same data you take it to the origin software then you will get similar kind of

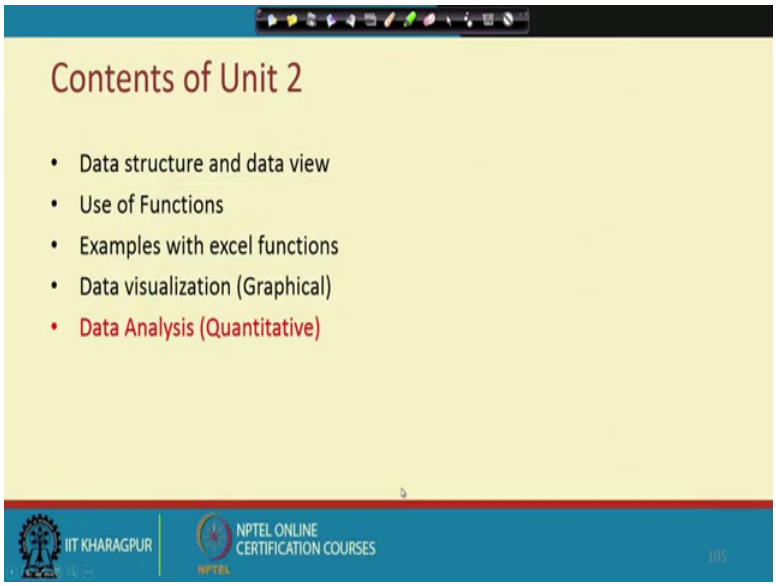
insights graphically also. That means, technically what I will like to say in order to understand the particular situation in order to understand the data or the kind of analysis. So, it is better to get better insights by various ways or by different kind of process right. So, here some of the things we are supposed to discuss more about this understanding and more of this particular requirement.

(Refer Slide Time: 06:02)



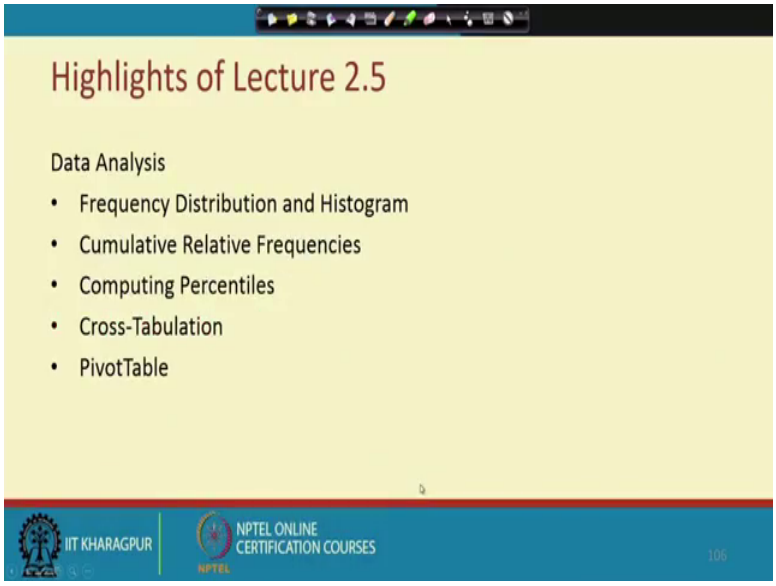
And corresponding to the last you know four lectures which we have actually highlighted the data structure, data view, use of excel functions then you know various examples with excel functions then graphical visualization.

(Refer Slide Time: 06:08)



Now, in this particular lectures I will give you some kind of quantitative approach you know or some kind of advanced kind of structure through which the we can justify that data is the insights of the data is more accurate. And then in the same times I like to highlight that you know excel spreadsheet has you know very handy and very useful for this kind of you know problems or particularly business analytics a kind of in a requirement right.

(Refer Slide Time: 06:56)

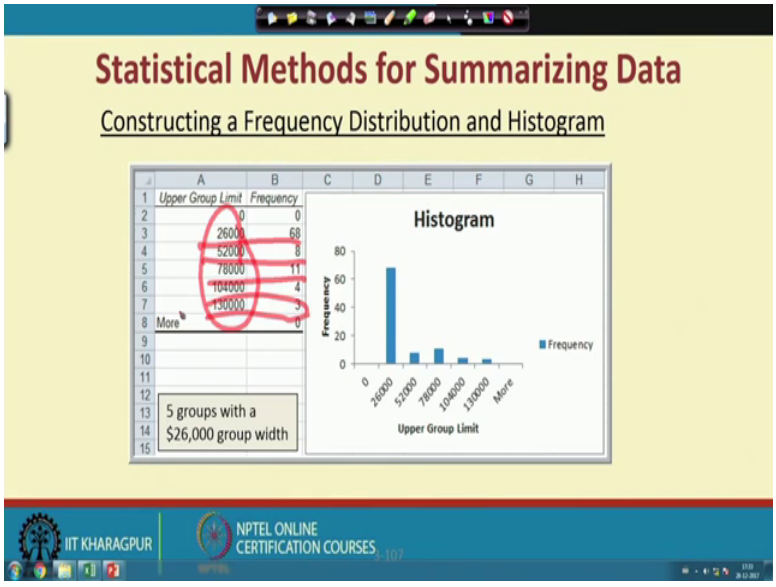


So, in this you know lectures we like to highlight the frequency distribution and histogram and cumulative relative frequency, a computings you know some of the things like you know percentiles, percentage, then cross tabulations, pivot tables. So, these are all you know kind of different options excel has of course, you know we have already discussed various options and various kind of features.

So, these are you know advanced features again excel has where you know the data can be visualized properly, gets some kind of better insights and then sometimes some analysis can be done through excel and that analysis can help you lot to get some kind of insights for the business problem and then accordingly you can take some kind of management decisions.

So, let me give you some kind of exposure here, what are the problems you know that can be investigated through these you know tools excel spreadsheets and then you know the kind of requirement and the kind of objectives right.

(Refer Slide Time: 08:08)



So, the standard structure is first structure is histogram. This is a kind of graphical kind of structure through which you have to know the details about the particular data set and then the kind of in a problem requirement. In fact, you will get more detail about this particular you know tools or graphs in the next you know unit that is in descriptive analytics, but still you know here the idea is actually to just justify that you know excel

spread sheet has an advantage to do something as per the business analytics requirement or need right.

So, here this is a standard kind of structures in this particular you know spreadsheet we have actually some information that is called as you know 5 groups with the 26,000 you know with and then corresponding to each particular you know entry there is you know frequency. That means, actually what we are you know supposed to do here this is actually some kind of process data or some way you can say that you know it is a structured data let me give you some kind of an exposure about the data and the kind of frequency.

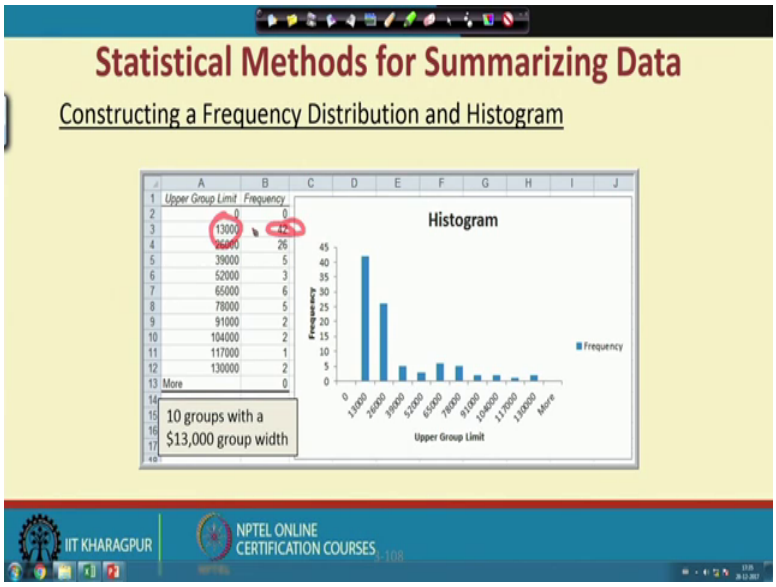
So, initially when you will collate the data it will be like you know here and there completely unstructured now you should just you know enter the data randomly in excel sheet from you know beginning to end then with the help of a particular command you arrange the data from you know highest to lowest or lowest to highest then you check actually for every every data point how many repetition is there. So, number of repetition by default we will take you to the kind of frequency. For instance if 26 in this particular spreadsheet, so 26,000 is repeated in 68 times; that means, the data is the 26,000, but you know the numbers available is equal as you know 68.

Similarly the data 52,000 you know repeated 8 times. So, as a result the frequency will be 8. So, likewise you know we have numbers. And then the corresponding frequency then you know excel spreadsheet has a kind of beauty that you know through the help of histogram it will give you some kind of exposures right. So, in fact, you know this table itself this table itself this table itself will give you some kind of structure so; that means, 26,00068 is the highest frequency, then this is 8 this is you know second highest and then this is you know something lower frequency.

So, this is actually clearly visible, but you know, now, the same things I am a plotting in a kind of structure called as a histogram that is readily available in excel spreadsheet. So, then it will give you some kind of better kind of visualizations. Now corresponding to 0 there is no entry, now, 26,000 so it is actually 68. So, this is the highest kind of entry highest frequency. So, 50,000 two layers then you know a little bit high then again low this is lower. So, this clearly you know visualize right so; that means, this is actually you know very a handy to understand the particular you know structure. That means,

histogram has a kind of structure through which you know it can visualize the data more accurate you know and then it will give you some kind of in a better insights compared to you know some of the other excel functions right.

(Refer Slide Time: 11:52)

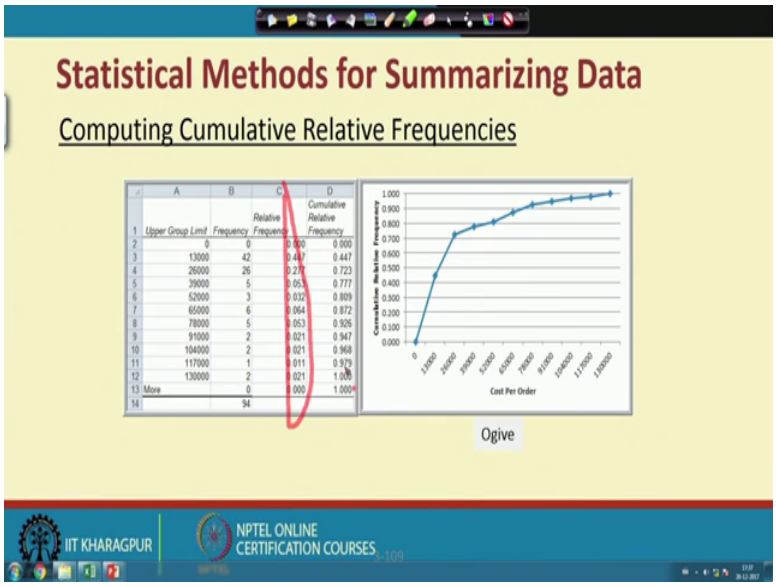


So, now, I will give you another kind of an exposure here. So, here the entry of data and there are you know compared to previous one, here there are you know 10 observations and with a you know with help you know 13,000 you know intervals.

Now, here you see here is again 0 there is no frequency so; that means, the there is non 0 in this particular you know series. So, it will be starting with the 13,000 then you know ending with you know 130000 right. So, then you know you have a different frequency. For instance again here 13,000, 13,000 is a kind of an item which is which is actually repeated 42 times; that means, by default frequency is 42 here. And followed by 26,000 and the number is repeated you know 26 times. So, by default frequency will be reported as you know 26. So, that means, again I am highlighting that you know excel spreadsheet has you know beauty that you know the unstructured data putting in a kind of in a structured format then this particular structure by default will give you some kind of in a better understanding, better insights and again with the help of some kind of graphical kind of an excel feature you can get more you know better visualization, better insights and by default it gives you know clear picture altogether.

So, that that is what actually you know what I can submit that you know excel spreadsheet has a kind of has an unique kind of structure through which you can understand the data and put in a kind of in a structured format right.

(Refer Slide Time: 13:39)



So, now coming to another kind of an angle to understand the data is called as you know cumulative frequencies right. So, same thing here same that I am you know using. So, the initially you start with you unstructured data then put in a structured data with you know frequency you know connections. So, 13,000 the repetition is a 42 times as a result the frequency is 42, 26,000 the repetition is a 26. So, as a result frequency is a 26 39000 repetition is a 5 as a result frequency is 5 like that you have to continue up to you know 130000 right. And then the next column compared to the previous one. So, here is the kind of relative frequency. So, this is a question of you know relative frequency.

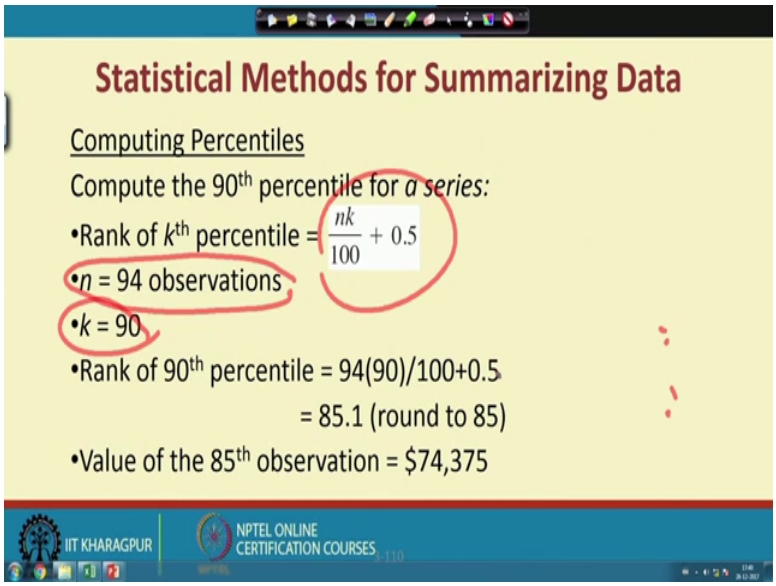
So, what we have done actually. So, relative frequency is nothing, but actually individual frequency by you know total right. So, then you will get some kind of relative frequency and then this relative frequency when you add up over the you know entry then by default the last entry will be actually 1. So, this gives you no clear cut you know signal or it is a kind of a particular format or structure format and that brings you know some kind of no confidence this is nothing, but actually the concept is you well connected with you know probability. So, how many structural and some of the probability will be equal

to 1 and then the free cumulative frequency will give you such kind of cross check that your entries are properly recorded and your structure is properly arranged.

So, we will get we will go, we will get to know details about the probability in the later stage, but you know this is a kind of structure which can you know give you better understanding, better exposure so far as you know particular data structure is concerned.

So, now I will you know means same things in transferring into a kind of in a graphical option again. So, instead of you know plotting actually you know numbers with the frequency. So, now, I am plotting numbers with these you know cumulative frequency. So, this is going in a kind of particular you know trend right. So, this gives actually some kind of structuring actually. So, you know you will get better insights you know how the cost per order actually over the kind of structure you know increasing or kind of in a decreasing. So, it will give you some kind of clarity, better clarity you can say and that can be actually easily you can do through excel spreadsheet. Initially you know when you have a firsthand data or raw data. So, that will not that will not give you better visualization and with the help of excel. So, you will get you know better visualization in fact, right. So, this is what actually the frequency, the cumulative frequency is all about.

(Refer Slide Time: 16:45)



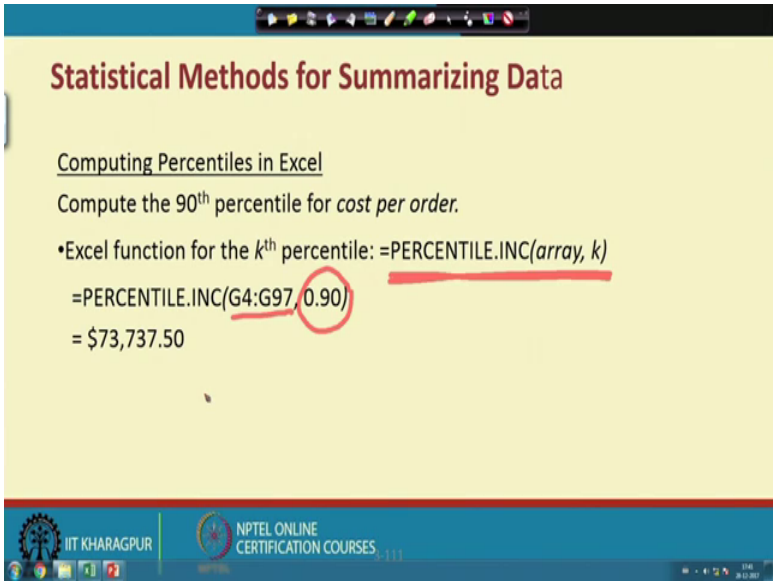
So, now another kind of excel function is you know percentiles. So, now, see whether it is a kind of histogram or is a kind of percentile or any kind of formula. So, excel has a beauty actually some operations are already inbuilt. So, just you have to give the

command automatically it will give you the kind of requirement or if not there you just you know write the formula and connect with the particular entry then by default it will help you to give some kind of value as per your you know particular requirement.

That means, so far as you know structuring or the requirement is concerned excel is a very handy spreadsheet which can help you lot to do or process the data as per the particular you know requirement. In the percent, in the percentile, this is a simple you know standard you know formula and when you have a spreadsheet then you know you can use this particular formula and calculate or else you know you go to the spreadsheet and use the command then by default it, it will give you the kind of value.

So, in this particular formula, the requirement is you to just signify what is the number of observations and how much percentile you know what is the percentile you want actually. For instance 90th percentiles, 94 and here the observation actually 90 divide by 100 and that too plus you know 0.5. So, this will give you the kind of percentile figures. So, I will give you better explanations when you will go to the excel spreadsheet right. So, let me first highlight it then I will take you to the excel spreadsheet right.

(Refer Slide Time: 18:27)

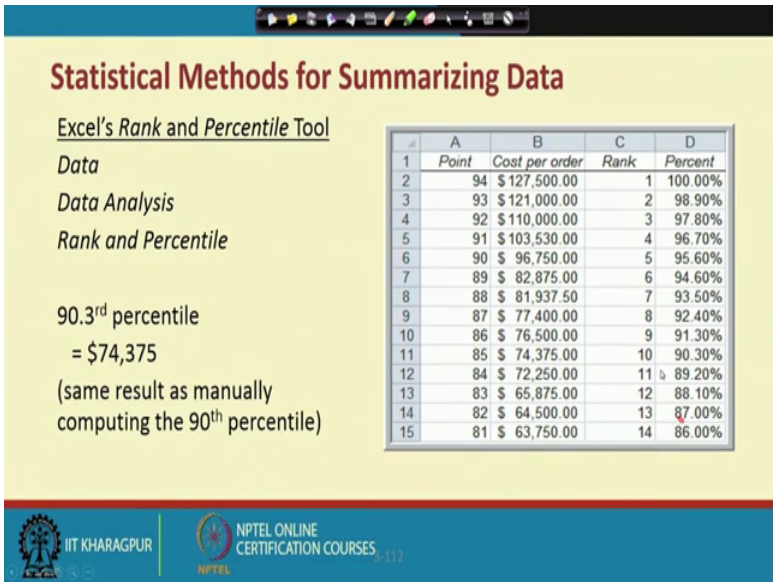


So, this is how the excel spreadsheet. So, same you know you have the data raw data then put you know equal to sign that is the command which you are you know asking a excel spreadsheet to you know give the particular figures that is what we called as you know one kind of know structuring or one kind of requirement. So, here actually this is

what the, this is what the range altogether so; that means, a we have actually entry G4 to G97 and we our requirement is 90 percentile then you just you know go to the particular you know column put entry then you ask for this particular you know requirement you enter the data then once you know enter, then by default it will give you some kind of figures right.

So, this is how the beauty of the kind of excel spreadsheet heads right. So, what I can say that you know it has advanced kind of structure through which, you can recognize the data and you can understand the data in a proper structure. The same things which you know excel spread sheet is usually like this.

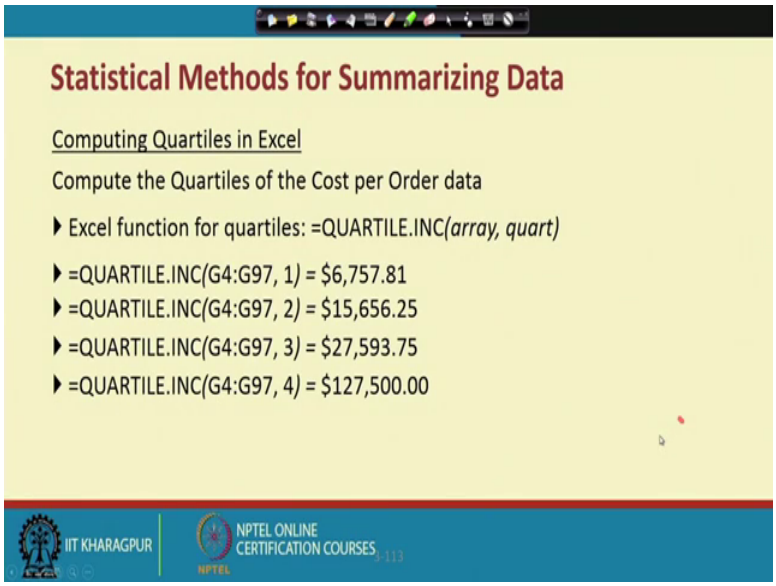
(Refer Slide Time: 19:29)



	A	B	C	D
1	Point	Cost per order	Rank	Percent
2	94	$ 127,500.00	1	100.00%
3	93	$ 121,000.00	2	98.90%
4	92	$ 110,000.00	3	97.80%
5	91	$ 103,530.00	4	96.70%
6	90	$ 96,750.00	5	95.60%
7	89	$ 82,875.00	6	94.60%
8	88	$ 81,937.50	7	93.50%
9	87	$ 77,400.00	8	92.40%
10	86	$ 76,500.00	9	91.30%
11	85	$ 74,375.00	10	90.30%
12	84	$ 72,250.00	11	89.20%
13	83	$ 65,875.00	12	88.10%
14	82	$ 64,500.00	13	87.00%
15	81	$ 63,750.00	14	86.00%

So, you have the reporting, just with the help of a particular in excel command then you can report actually see here. So, how the actual data is all about and then how it will be transferred into some kind of in a percentile right. So, this is how the; again I like to highlight that you know excel spread sheet has a beauty all these things.

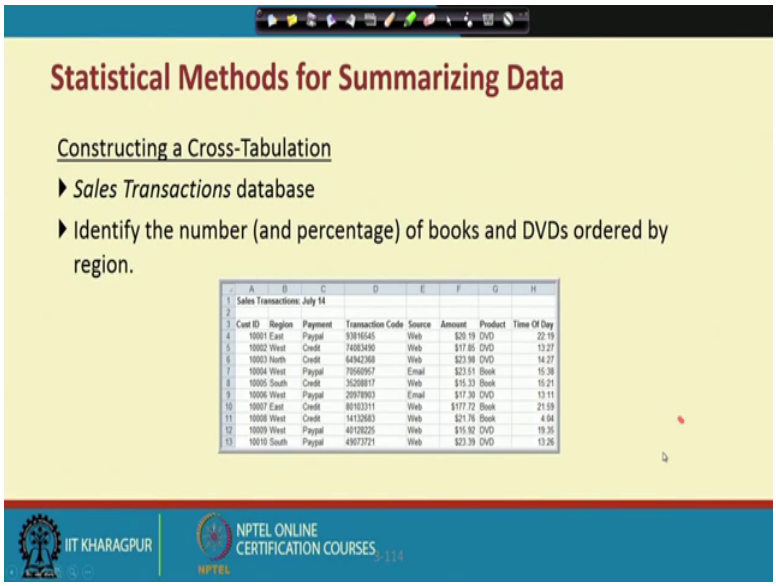
(Refer Slide Time: 19:55)



So, similarly you can find out you know quartiles; so in a kind of spreadsheet, so need actually quarter informations like first quartile, second quartile, third quartile, forth quartile and so on. So, you can get to know much better or all details in next unit where we like to d discuss a component called as a descriptive analytics, right.

So, in the mean times you should understand the use of you know excel spreadsheet and you should know the operations. Then corresponding to the operation and an excel you know kind of a requirement you are in a position to know more details and operate the data as per the particular requirement, right.

(Refer Slide Time: 20:44)



So, this is actually a real excel spreadsheet. And then you live you will find here different kind of entries right. And of course, this is a kind of structured entry and the there is a beautiful kind of in a specification. But you know having lots of you know data with lots of features lots of attributes its sometimes very difficult to analyze, but excel spreadsheet has lots of you know different functions and by using some of these functions you can better understand the particular problem. For instance one such example is called as you know cross tabulations.

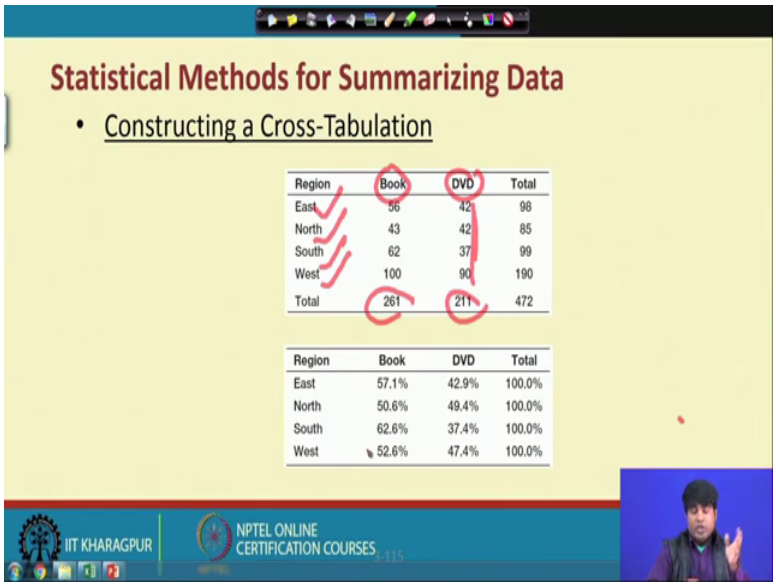
So, cross tabulations with this particularly complex tables now it will be a give you see a simple kind of in a structure where you can easily understand and you can get better insights and then you can analyze as per the particular unit requirement. But the thing is that I will give you two kind of situation. See when we are actually solving a business problem then we need actually data, without data the analytics are totally useless. So, you have a data and then with the help of a particular command or tools or a kind of structure you have to analyze and then get some kind of insight and with the help of these insights you can you know solve some of the business problem.

But, by the way when you are dealing with a problem and data so that may not be a in the first hand will be readily available as for your requirement. So, a particular organization or a particular sector they may record the data as per their requirement. But when you like to use this data so you have to transfer or you have to structure the data for

your requirement. So, that is the classic here, this is the classic examples. So, these data are you know recorded as per the particular requirement of a particular organization.

Now you need actually to understand the data and you need some kind of insights as per your program requirement and then you have to analyze accordingly. And in this case excel spreadsheet has a beauty to help you lot to get these insights as per the data availability and the kind of problem requirement, right.

(Refer Slide Time: 23:14)

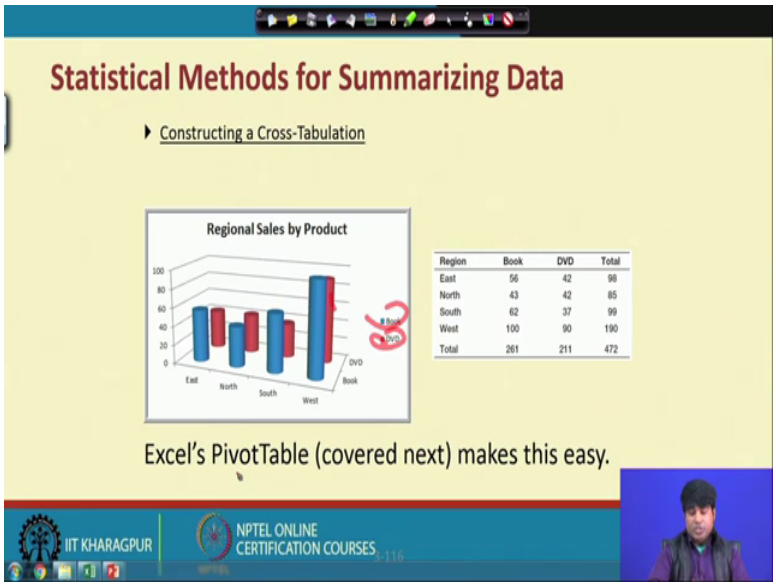


So, now the same in the same structure whatever available there. Now, if I will summarize in a particular structure through spreadsheet then it will give you better insights and it will give you very simple structure to understand the reality. You see whatever data is there with many attributes. So, now, I will be streamlined and little bit simplify and then I will indicate that this will give you some kind of in a better insight.

From this particular you know big data, so I am actually clustering here that you know region wise classification and then book sellings and DVD sellings. So, how book selling is you know differing from one region to another region and then like this actually. So, this is actually kind of book selling and this is a DVD selling and from one a region to another region in India. And then in percentage how it is actually changing. Sometimes you may not actually recognize properly in terms of absolute figure, but having percentage relative transformation you can you know better understand and can have a better comparison, right.

So, again this will give you the total of book sellings and then region wise classification, similarly DVD has a total kind of sell and this is region wise classification. Then here in this table you like to know which region is having high percentage of book selling and which region having you know high DVD selling. So, means against what I like to say that excel has a beauty to analyze these particular problems and simplify the particular structure and give you better insights as per your problem requirement, right.

(Refer Slide Time: 25:03)



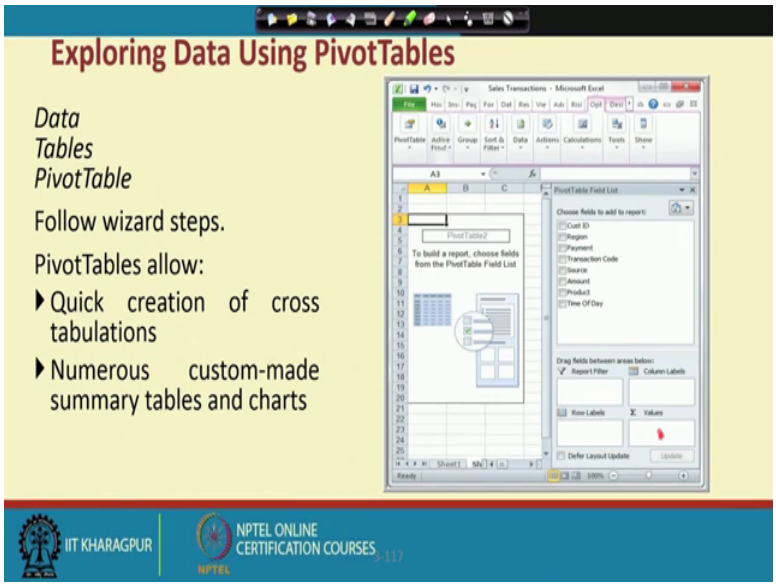
So, this is another kind of looks that was quantitative kind of structuring through which you have to get insights a as per you are problem requirement and problem analysis. And here is this is another kind of view graphical view through which actually you have to understand the particular structure and then get the particular insights is the same things, but you we are putting in a different structure. So, that you can understand very easily. Everything we are putting in a particular you know structures you know the red signal, the red signal is nothing, but called as you know DVD structure and the blue signal is nothing, but in a book structure. So, it will give you that you know.

So, from this particular you know plotting here actually you have to go in depth and look into the particular picture, but in the graphic you will find in a western zone it is actually highest you know selling both in book as well as you know DVD. It gives you know clear signal here where you know in the northern side and the eastern side, northern side it is a lower and compared to you know other regions. That means, again what I say like

to say that you know excel has a beauty to you know give some kind of better exposure and better insight to understand the problem, to understand the data and to get some kind of insights as for the problem requirement.

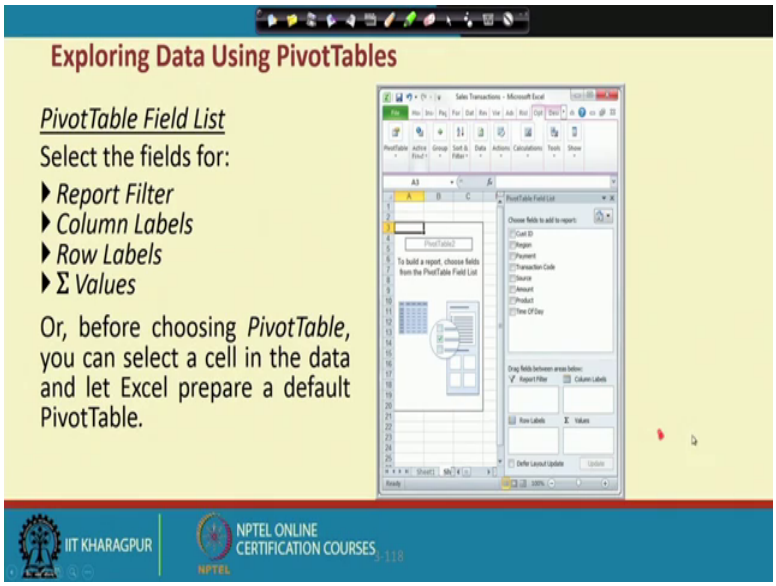
So, now I will give you another kind of you know look.

(Refer Slide Time: 26:23)



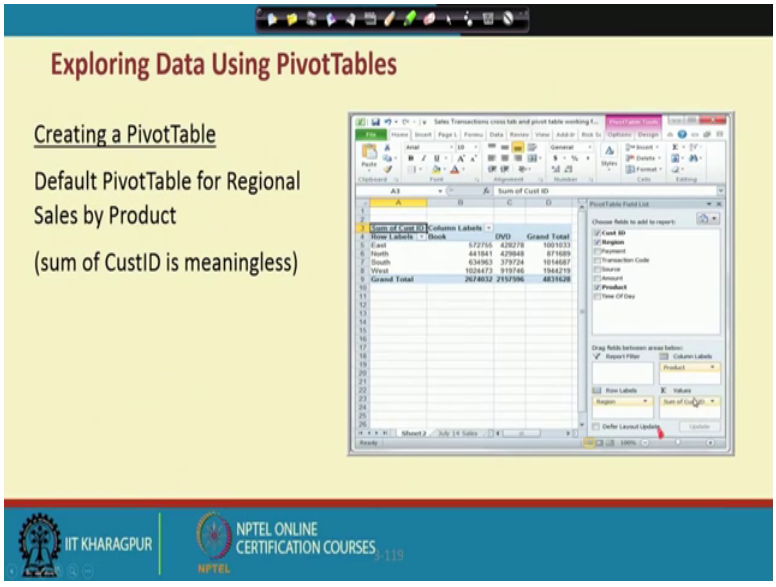
So, there is another you know tool works called as you know pivot tables. So, the pivot tables will help you again some kind of better insights from the complex you know kind no problem and excel has a beauty to you know bring the problem in a kind of means transfer the particular problem and bring the analysis as per the particular structure through which you can understand and get some kind of in a better insights right.

(Refer Slide Time: 26:50)



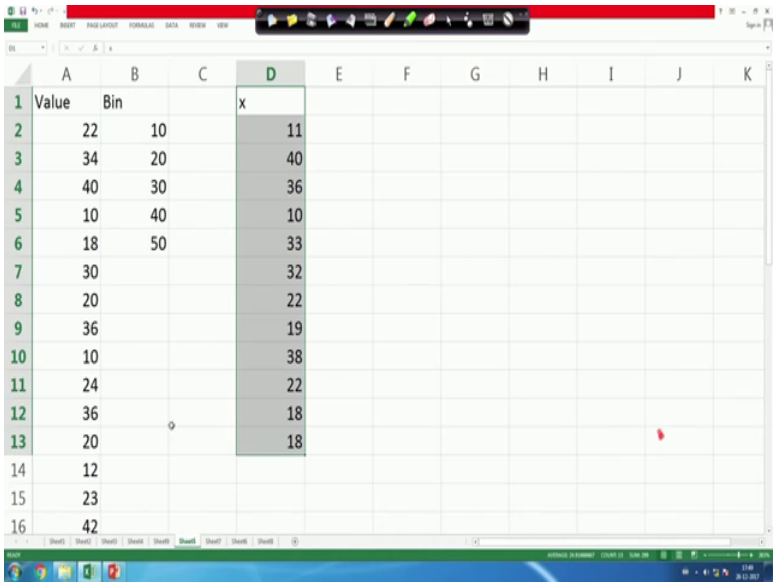
So, these are the things I am actually you know like to connect with the pivot tables. So, you have actually 4 different you know structure in the pivot tables then corresponding to particular problems you have to analyze as per the problem requirement. Let me take you to this particular you know structures.

(Refer Slide Time: 27:06)



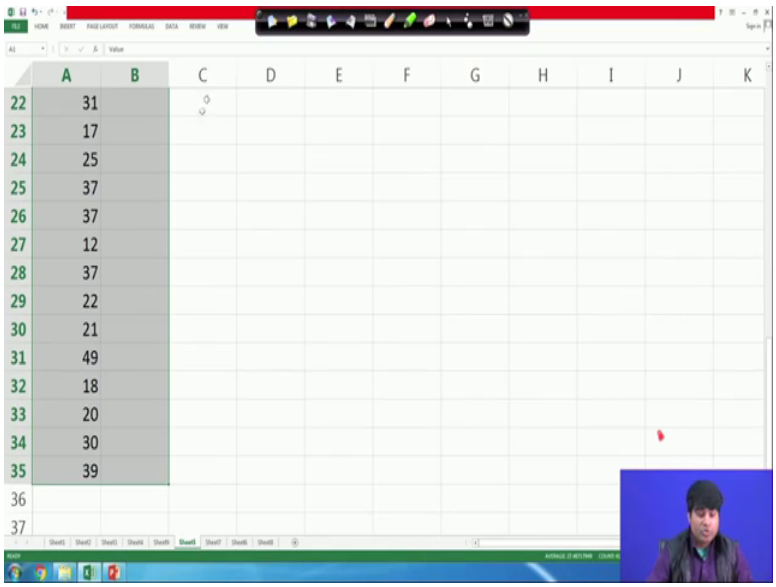
For that you know for pivot tables, what I will do. So, let me take you to problem in the excel sheet then I will connect with this you know problems right so.

(Refer Slide Time: 27:20)



So, first actually I like to show you know; how you know a histogram can be connected with these particular you know excel sheet. So, let us say this is a you know random data and we have actually generated here from 10 to 50 and then we like to actually show you how the diagram you know histo diagrams can you know visualize properly.

(Refer Slide Time: 27:45)

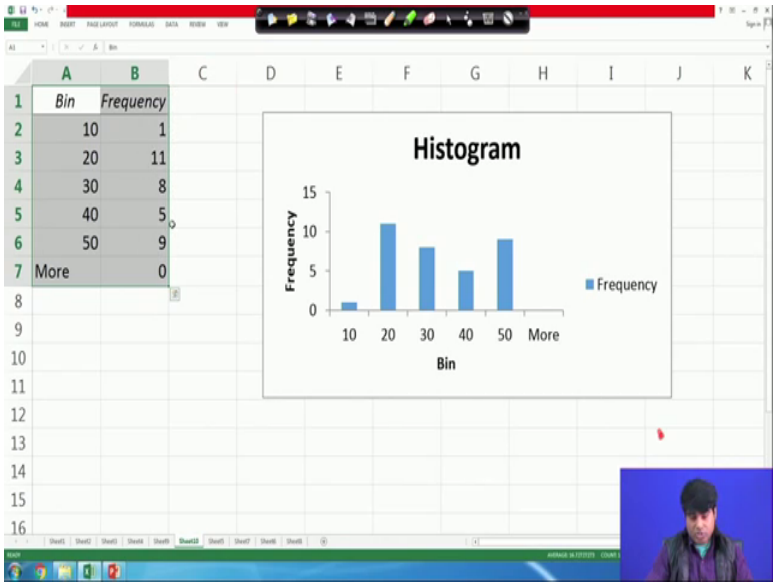


So, go to this you know data then you know go to the data analysis package and by default you will find here is a package called as in a histogram. Just you put ok, then you give the input range here is the input range is here actually A 2 to then whatever available

here is up to say you know 35 and then bin range is already have been given that is actually the range available in this particular series. So, I put in the wrong place.

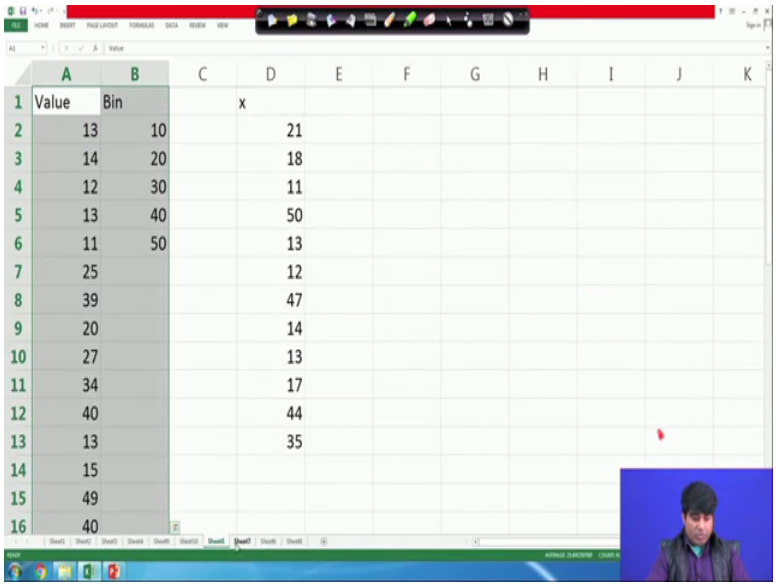
Let me give you another same histogram. So, first you put this you know entry A 2 to like you know a 35. Be careful you know if the you know entry or the command is not correct then you will not get actually the as per you know the requirement then you put the entry like this then you looks the chart output here, there are many different forms it is there. So, you just put in a chart output then it will you some kind of an expression this is what the histogram can be obtained data here right. I will little bit enhance so that you can visualize.

(Refer Slide Time: 28:46)



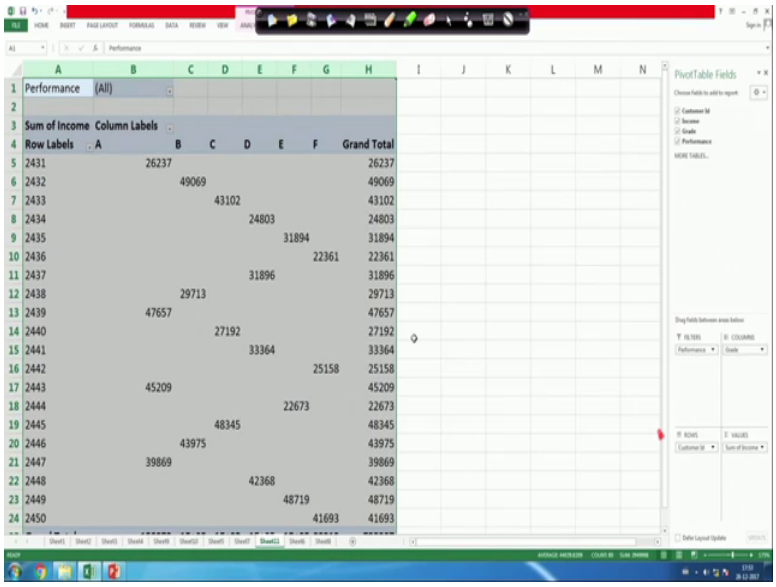
This time you know data it will be transfer into this frequency distribution and then with the help of you know series and the frequency distribution the historic histogram can be plotted. So that means, what I would like to say that you know excel has a beauty actually to you know use the data and with the help of some functions it will give you some kind of insights to understand the problem and to predict the problem as per the problem requirement. Then I will take you to the another kind of you know structure here.

(Refer Slide Time: 29:24)



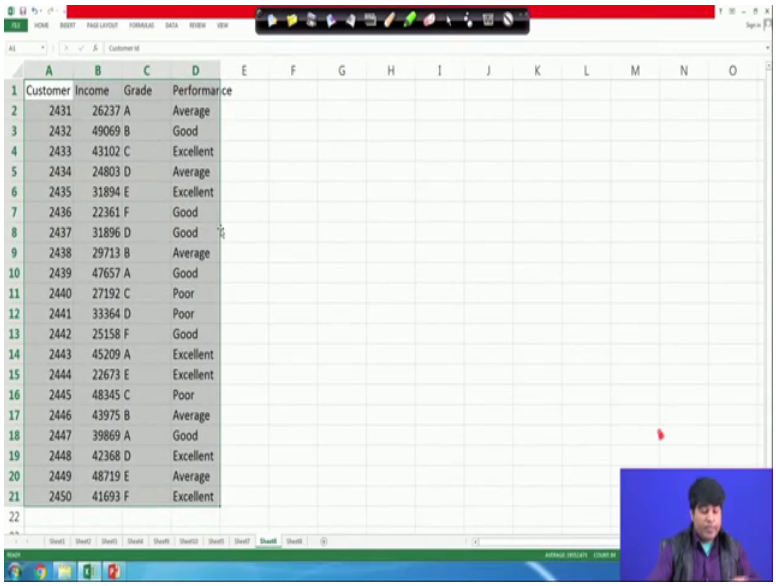
So, let us say the pivot table concept.

(Refer Slide Time: 29:26)



Here, let me let me give you a little bit enhance here. So, in this you know pivot tables case. So, I am taking these examples you know we have a customer id and against customer id we have income, grade then you know performance right.

(Refer Slide Time: 29:28)

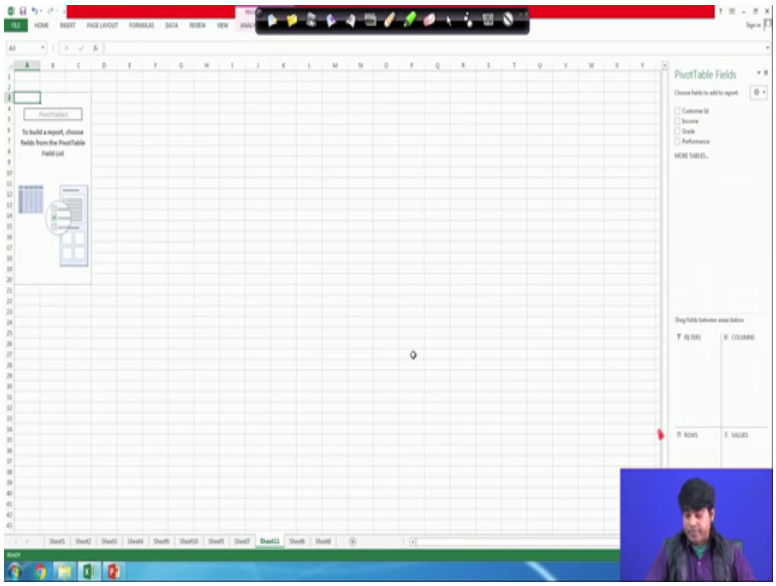


	A	B	C	D
1	Customer	Income	Grade	Performance
2	2431	26237	A	Average
3	2432	49069	B	Good
4	2433	43102	C	Excellent
5	2434	24803	D	Average
6	2435	31894	E	Excellent
7	2436	22361	F	Good
8	2437	31896	D	Good
9	2438	29713	B	Average
10	2439	47657	A	Good
11	2440	27192	C	Poor
12	2441	33364	D	Poor
13	2442	25158	F	Good
14	2443	45209	A	Excellent
15	2444	22673	E	Excellent
16	2445	48345	C	Poor
17	2446	43975	B	Average
18	2447	39869	A	Good
19	2448	42368	D	Excellent
20	2449	48719	E	Average
21	2450	41693	F	Excellent
22				

So, we like to analyze it equally because there are you know three attributes are there corresponding to customers. So, we like to grade them and now looking this data it is like you know it is a very complex to understand reality, but excel has a lots of you know functions. So, some of the function can help you lot to understand this particular problem and then it will help you lot to get some kind of insight as per the problem requirement.

So, now I like to solve this problem through you know pivot tables then you go to the excel sheet again then put the insert option then in the insert option extreme corners you will find a item called as you know pivot tables. We have already highlighted this then you just click then pivot table by default will be appeared here. So, once you will get the pivot table you just put you know options.

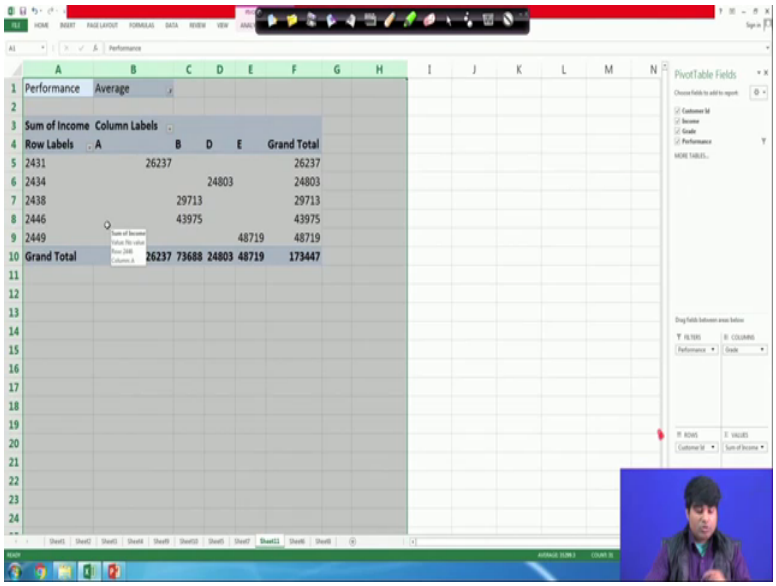
(Refer Slide Time: 30:41)



Now, here in the pivot tables you refined these are you know following 4 option corresponding to this particular you know problem. So, now, I am putting here, I am putting here actually some of the options, now I am putting income to the value options then I am putting here the customer id, customer id, a customer id to the raw operations and I am putting red into column, then I am putting performance into filter. So that means, you have to just give an indications corresponding to this data and now by default you will find the layout here by default you will find the layout here, I am just you know enlarging it. So, you see here the kind of requirement.

So, now check here, see here the first hand information and now the second the first hand the information is like this and the second hand in hand information after using the pivot tables. So, it will be appearing like this. So, this is actually giving you some kind of classifications with respect to different attributes right. So, your performance is actually corresponding to this performance you can change this particular option here.

(Refer Slide Time: 32:19)



We have actually different kind of a kind of divisions like you know average excellent good and performance then you just click one particular options then it will give you again it will give you some kind of requirement again you change this average to kind of an excellent then it will give you some kind of next exposure. So, now, responding to grade and your kind of income and id, you will get some kind of better kind of information from the original data then get some kind of you know insights and that insights will help you lot to solve some of the business problems.

Again the performance can be obtained again you change to you know goods then again it will change like this. That means, actually see, we are actually, what is actually required here. So, your first hand entry in the excel should be you know perfect. So, if you enter the data perfectly and in your kind of structured format and all entries are correct. So, you have to be very careful and cross check it. So, once your entry is correctly you know correct and correctly structured and then from that particular entry you can understand easily or through some kind of graphs or something like you know quantity tools you can get some kind of better insights better understanding then if still it is not enough feasible or possible. So, you can use some kind of excel you know tools like you know the sighted example which I have you know highlighted here like you know percentile quartile histograms then in a pivot tables cross tabulation. So, all these you know tools are there you know and that will help you lot to analyze the problem in a as per the particular requirement and to understand the particular problem in a more

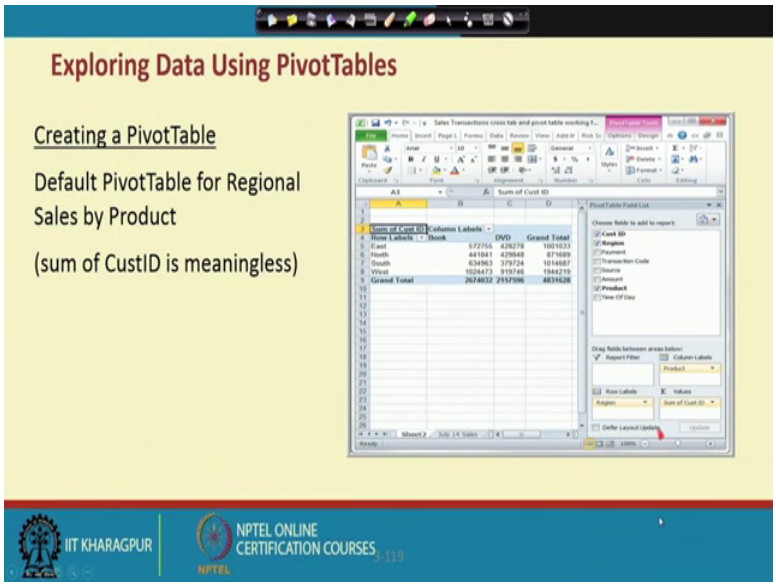
attractive way. Because ultimately the requirement is that only entering the data corresponding to the variables or corresponding to a particular attribute is not enough actually.

So, when you need actually a solution for a problem and that we through some kind of an analytics a business analytics tools you must be very careful how is your entry. So, your entries should be a very much what I can say perfectly correct and it should be in a kind of proper structure because if any kind of mistake or any kind of in a uneven situation will there. Then you may not be in a position to you know analyze these problems because it is well connected with whatever items or package we like to you know use for further investigation to get insights it is well connected with your you know initial entry or that is what we called as a first hand entry or raw data. If you are fast hand entry raw data is not actually correct or proper then the entire insights will be getting affected.

So, you must be very careful how you have to you know enter the data. So, be ensure that it will be correctly specified. Then you should know the entire function of you know excels and excel spreadsheet and there you know functionality, various features, various operation, because this will give you enough kind of exposure or opportunity to understand it data to get insights and then to get some kind of better inference as per your problem requirement and then you may be in a position to solve some of the business problems and then you will be in a position to take some kind of no management decision.

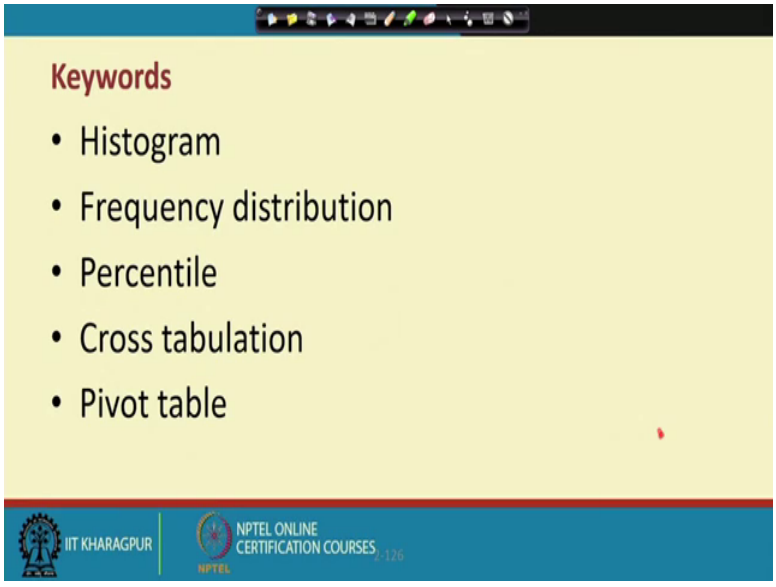
So, the final thing is that you know what I like to say is that you know.

(Refer Slide Time: 36:03)



So, your problem is nothing, but actually the understanding of you know data, the data visualizations and to get insights as per your you know problem requirement. So, in this particular unit we have discussed all these details and then I have already highlighted all these details.

(Refer Slide Time: 36:29)



So, in the mean times what I what I like to say that you know in these particular lectures we have discussed some of the important features of you know excels or some kind of functionality like histogram, frequency distribution, percentile, quartile cross tabulations,

pivot tables. So, likewise we have also discussed several you know functional or operations in the excel sheet and the only submission is that you know if you are very handy in excels and all these operation excel operations then by default some of the analytics can be easily you know process or then you can solve sub problems as per your you know problem requirement. And then again it will help you to take some kind of better management decision.

So, so the final submission is that you know be careful say about you know the data understanding data entry, data visualizations, then you know excel operations any kind of spread sheet operation. So, that you know the idea is that you know since it is the game of you know data and the data processing. So, you should be very much you know careful how to know, how to enter how to process all these data. So, once you are you know familiar with all these processing through excel spreadsheet then you know for advanced kind of requirement and advanced kind of an analysis, it will be help you lot to quickly proceed and quickly get some kind of inference. Because some of the complex problem when you have a more number of variables, more data then you know and if you have no idea how to enter, how to process, how to operate then it will have a some kind of negative kind of environment where you are not in a position to analyze the particular problem in a quick time and as per your requirement. So, you must be very careful how to know all these things and how to operate all these things.

With this we will stop here.

Thank you very much. Have a nice time.