Toronto Business College

Data Visualization

Assignment (6)

Case study - Tableau charts

Hand Out Date: 25/11/2021

Hand In Date:02/12/2021

Assignment Type: Workgroup(Maximum 4 Students Per Group)

Due Marks: 15%

Please put the Group’s students’ names and IDs clearly on top of each work document.

**REQUIREMENTS**

Note: **sample\_superstore**file (XLS format) is attached with this assignment.

Using Tableau and the Superstore sample data, answer the following questions:

* Display the most sold subcategory in each segment?
* Display top 5 most profitable customer name, city, and their profit?
* Display top 5 most discounted subcategory
* Which category performs well (sale, profit, return, and quantity)?
* Is there a correlation between quantity and sales? Answer the question by displaying a scatter plot
* Create a dashboard and download it as pdf for submission

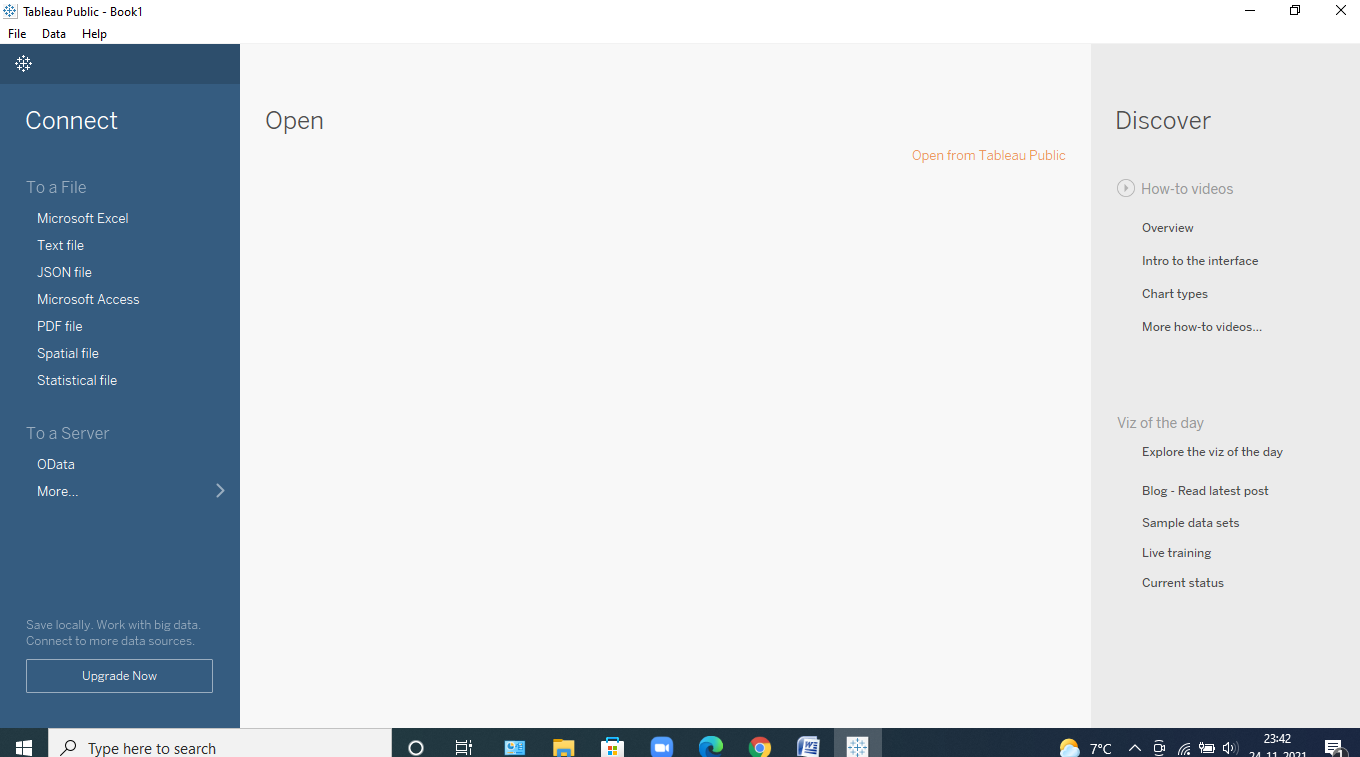
**Deliverable: -**

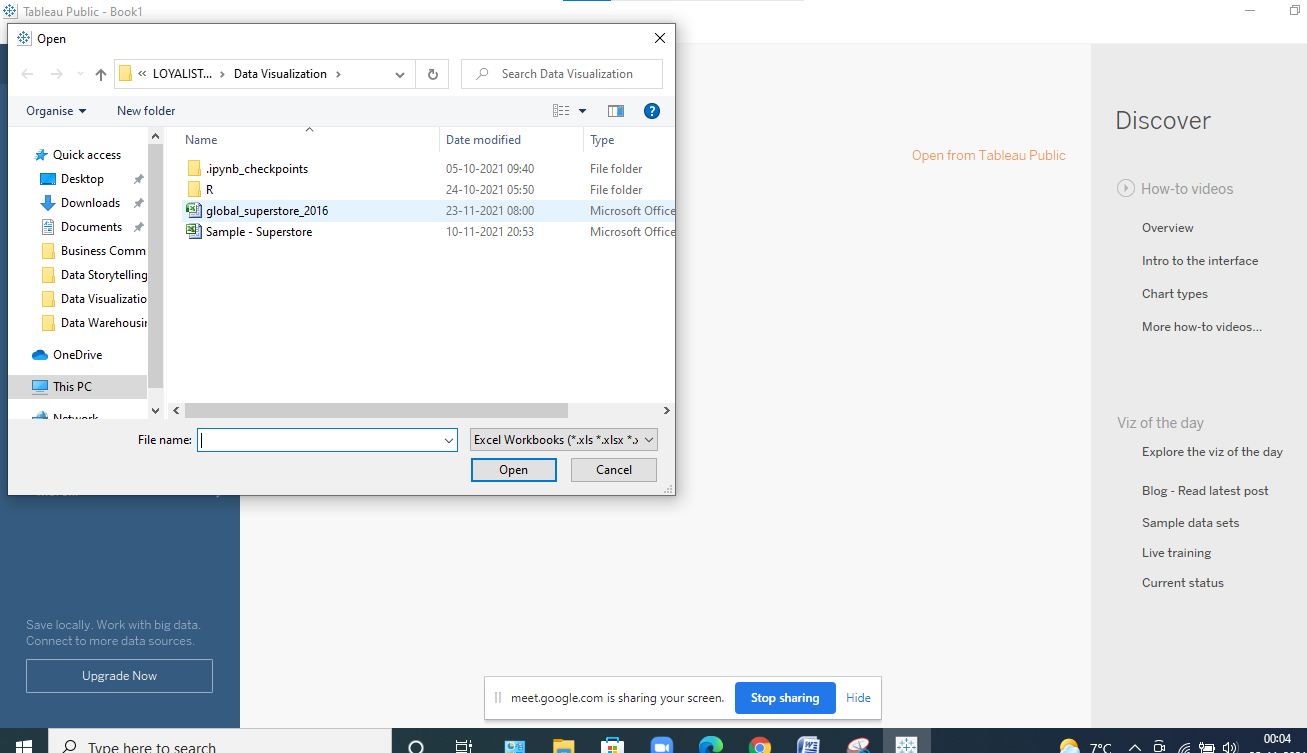
Word document that shows: -

1. –Describe the steps you followed to generate the charts
2. - Screen shots of the generated graph/graphs.

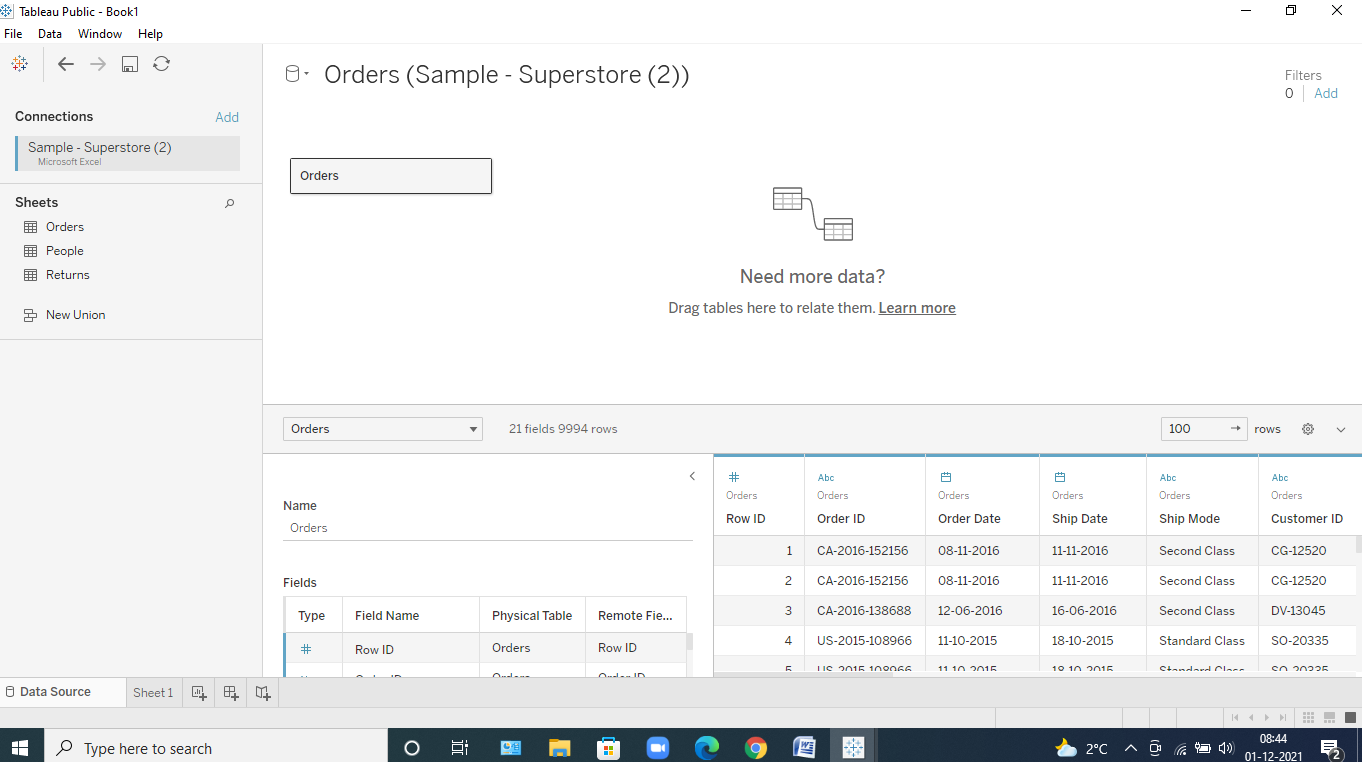
# **Loading Data in Excel into Tableau : -**

* Let's get started by downloading the dataset that has been provided to us. We've been given an excel file on sample superstore.
* We clicked on the excel sheet and the data got loaded.
* We clicked on Microsoft excel and then we got a pop up and we selected that very excel file.





* Once the data gets loaded we get to see how many tables are inside it like in this case we can clearly see that we have been provided with nearly about 3 tables and then we have too see whether they need to be joined or not like in this case there is no need to join tables and simply use only one dataset that is the Order table and drag and drop it in the space.
* Now, we can clearly see the data in the dataset, and the rows and columns as in the figure mentioned below. We can even think about it to see like which is categorical, numerical, Null and many more.



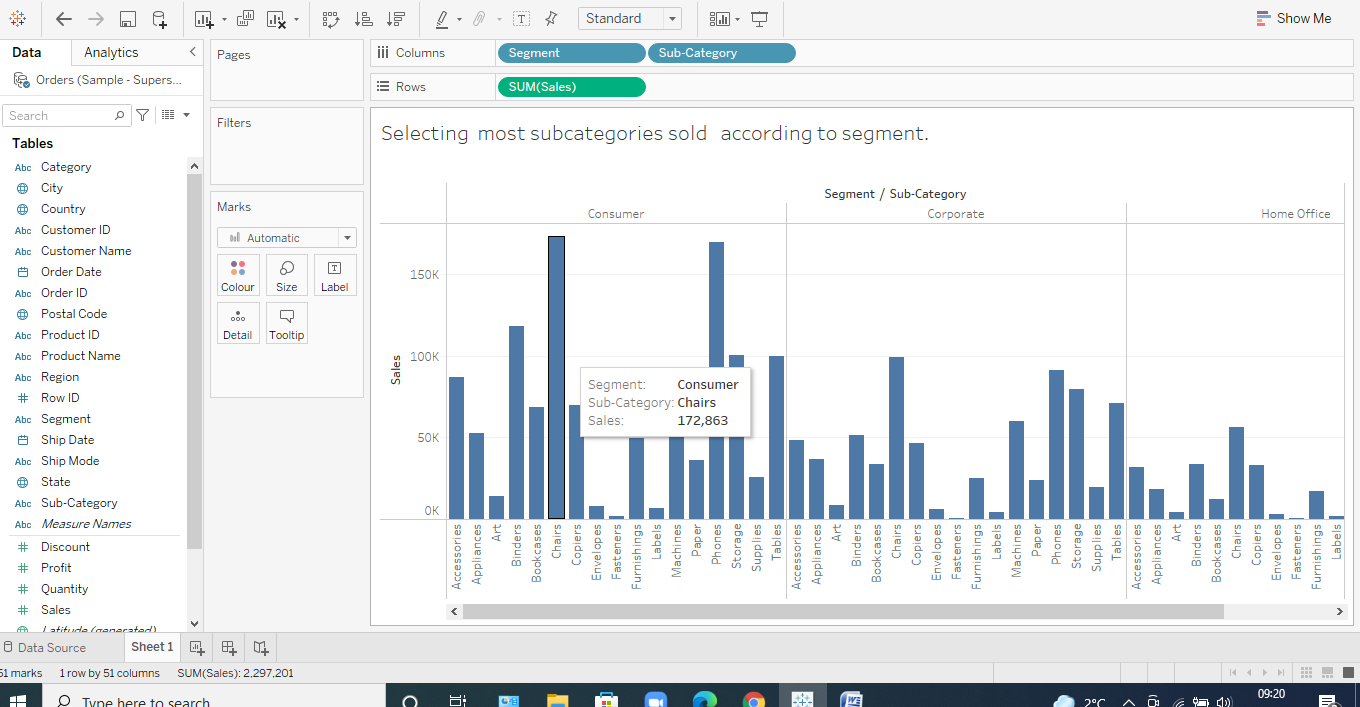
Now that we've uploaded the data to the software, let's get right into the Sheets , where we can see the analysis we're going to perform.

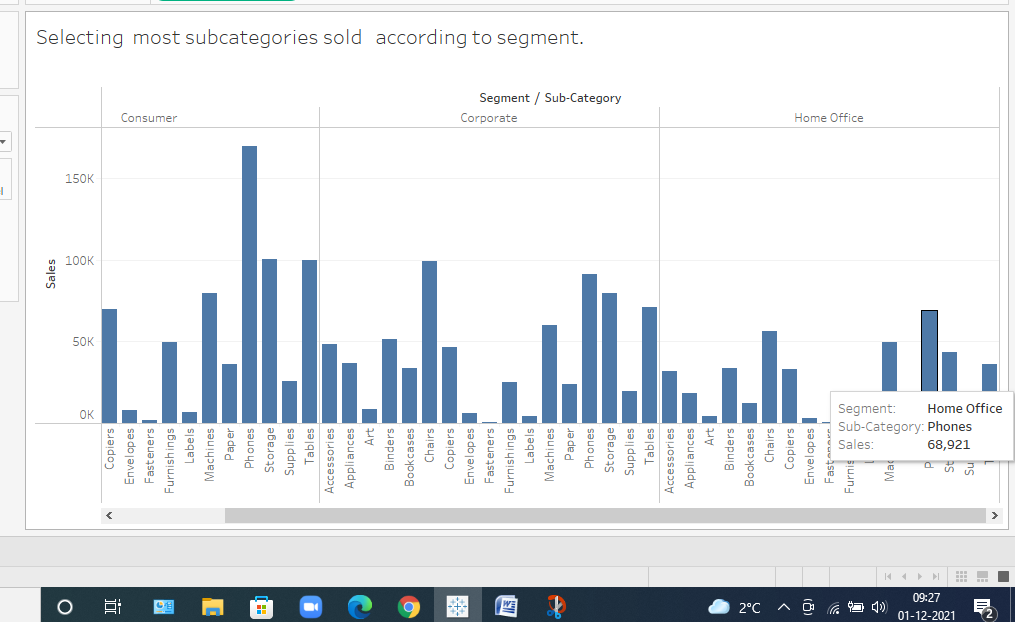
* As and when we go inside it, we can clearly see the Dimensions and the Measures in it
* Measure depicts the calculative values and Dimensions are usually the categorical values.

Now lets start explaining each and every question step wise, so that we get to know how to work on it : -

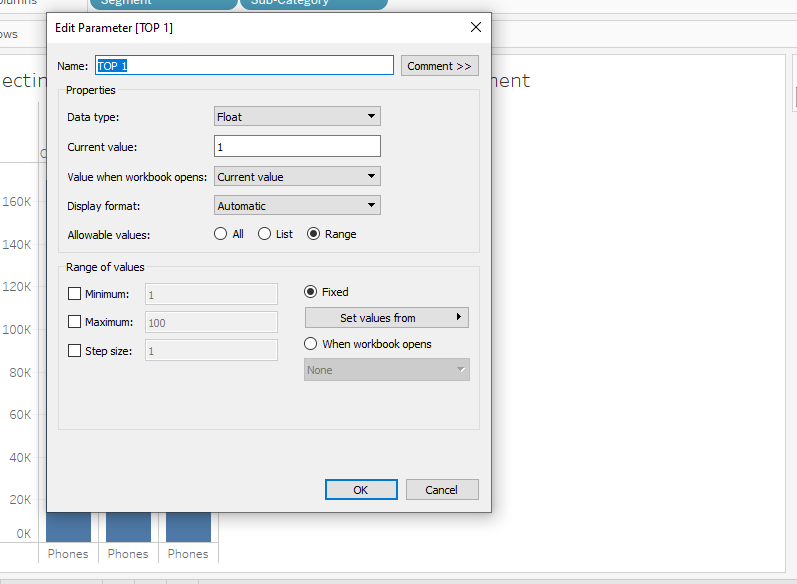
**Answer 1 :-**

* According to the question, we need to select the total sales achieved in each subcategory under every segment.
* And for that we double clicked on segment, sales and subcategories, which got placed inside the rows and coloumns automatically.

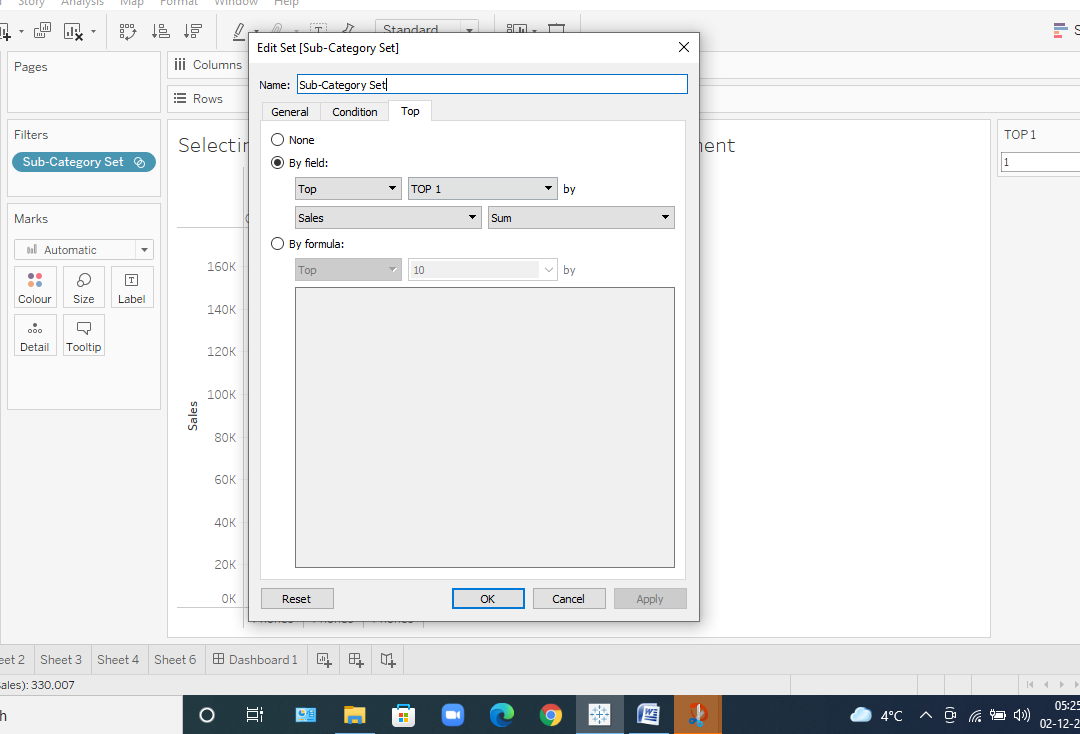




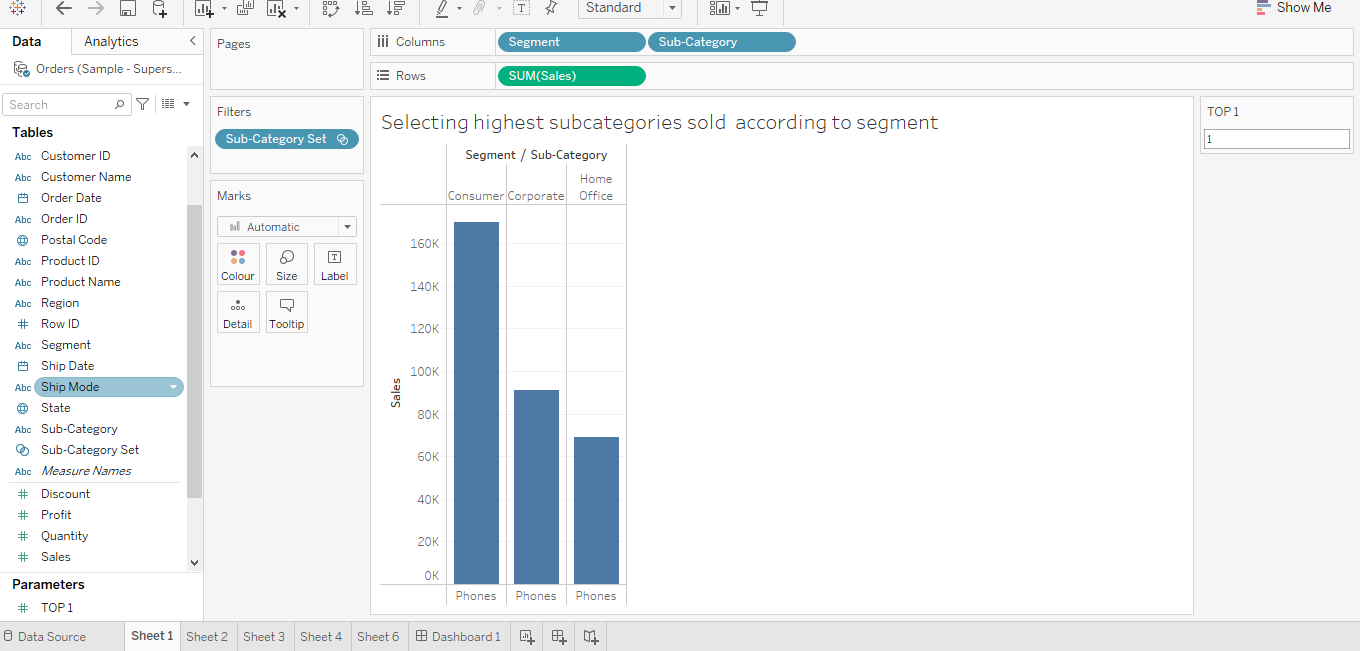
* Now that we can see that we need to find only highest of all the subcategories so for that we need to create a parameter
* Lets name it to be top 1
* For it to be created randomly right click on the set of measures and then name it



* After creating it go to subcategory and then go to create and then go to set
* To create a set of number on which we can apply this parameter on to it.
* Then apply the parameter and sum of it



* After creating it we can easily apply this new attribute to the filter and colors and get this.



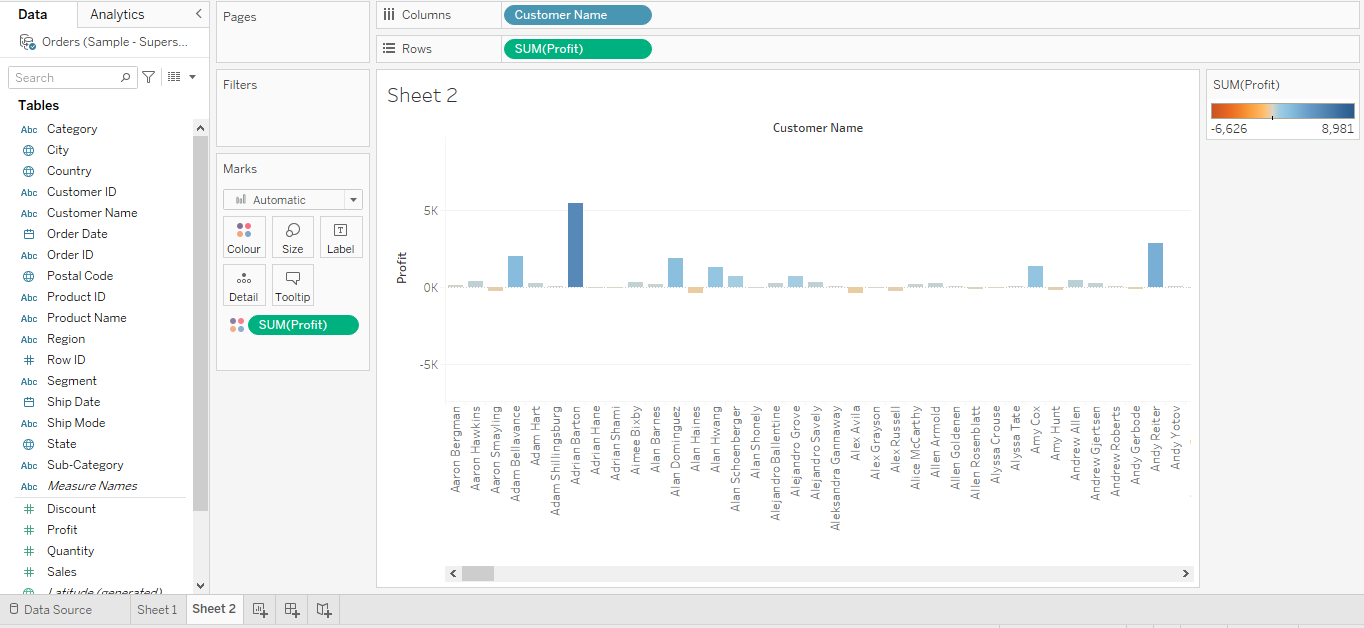
**Insights we Found out : -**

* **From this dataset we found that there are 3 segments corporate , homeoffice and consumers.**
* **The sales achieved under phones subcategory has been very high around 160,000 and 99141 in consumer, corporate segment respectively, whereas in case of Home – office phones took the lead by 68921.**
* **Phone is the only subcategory insight each segment with maximum number of sales after it comes the chair.**

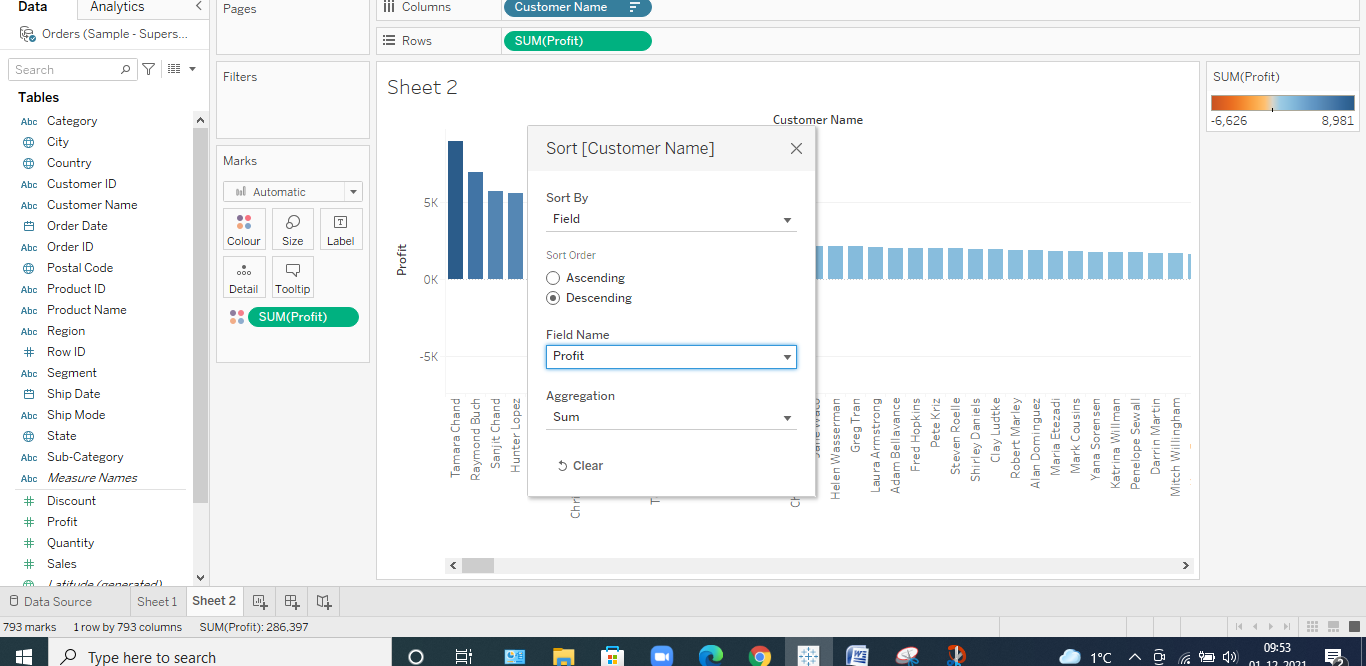
**Answer 2 : -**

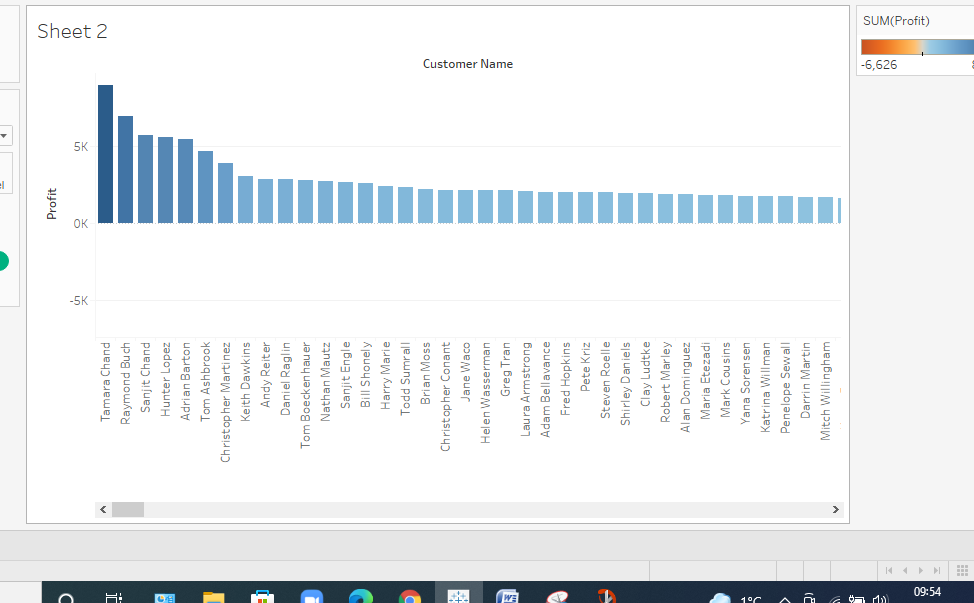
Now, we want to go ahead with top 5 most profitable customer name, city and there on their total profit : -

* So lets show the top 5 profitable customer name with steps and their screenshots to achieve it : -
* So first select profit and customer name by double clicking on it.
* Put profit under color and label under marks. We got this.

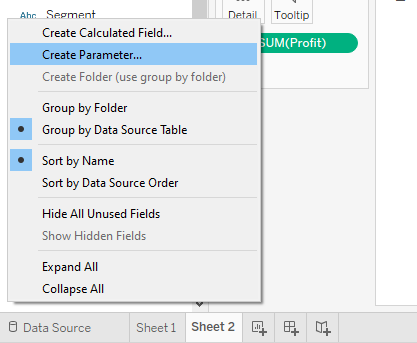


* We cannot find anything so lets right click the customer name select sort , change to filter profit and descending order and we get this : -

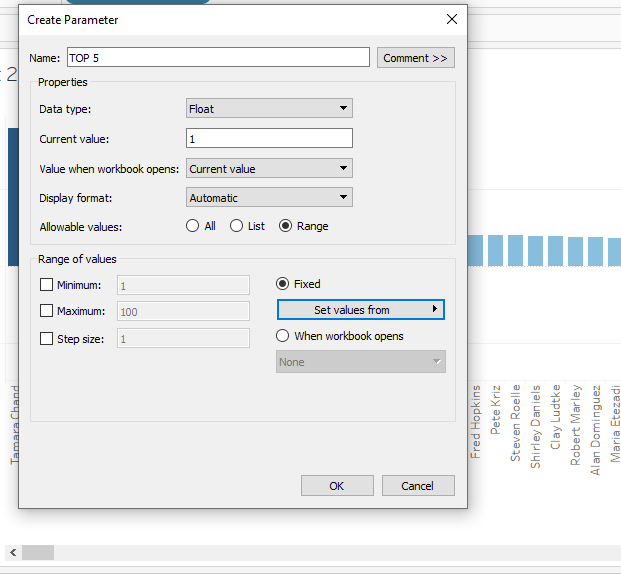




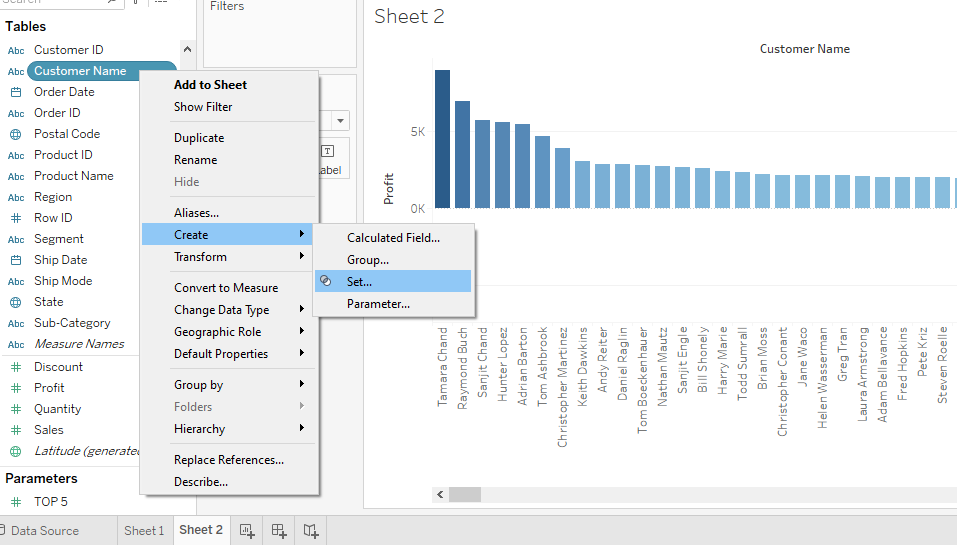
* Lets create parameter for this for helping us to take out only 5 top customer names, lets see : -



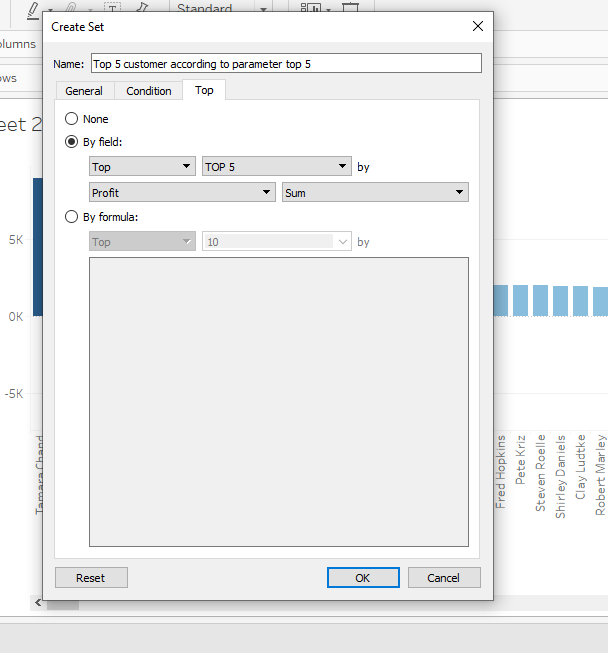
* Setting the values of parameter according to our choice we get this : -
* As we need to see the top 5 from current value and then click on show parameter and apply the values to it.



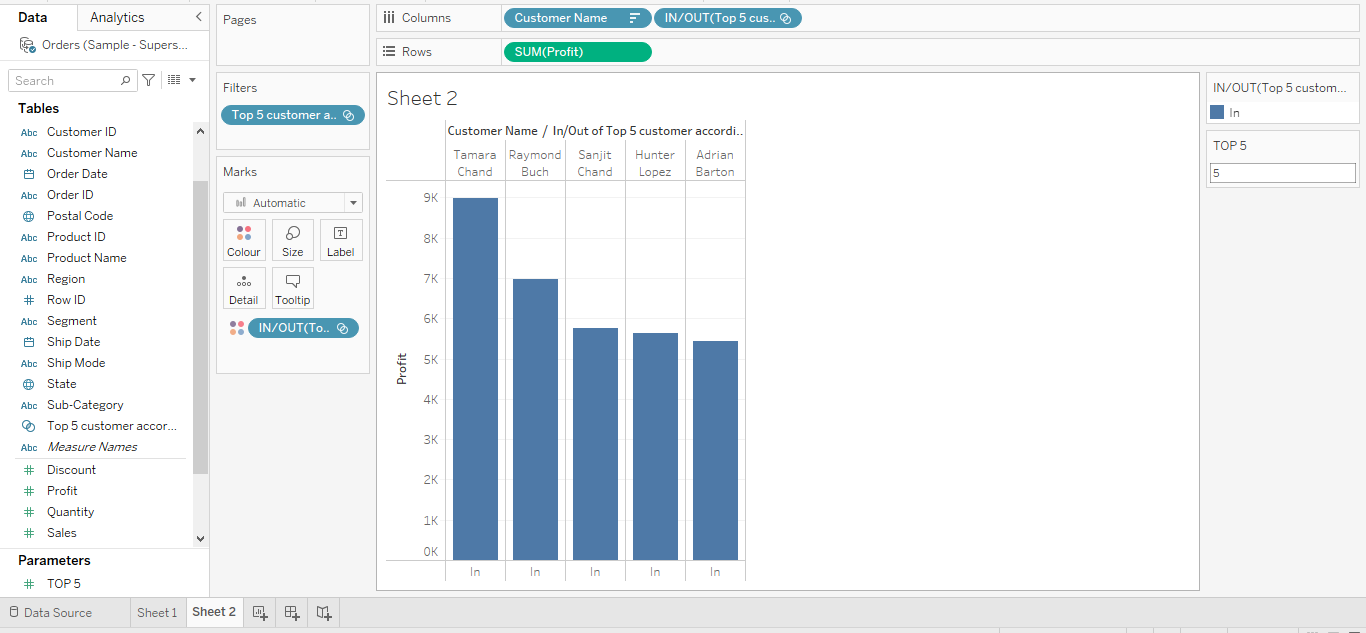
* To make it function we need to click on the attribute where we need to apply it : -
* Customer name and then create a set of 5 given name to it and then apply the parameter to select the so called top 5 customers .



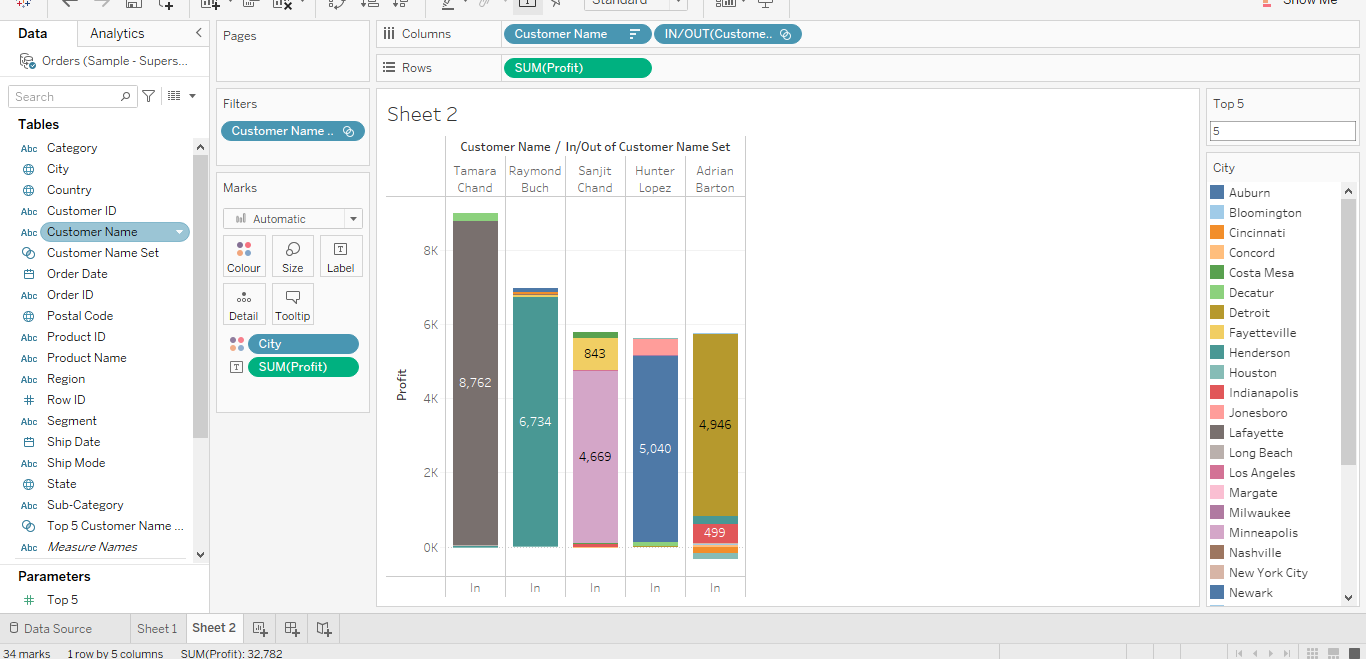
* And then apply the parameter to select the so called top 5 customers .



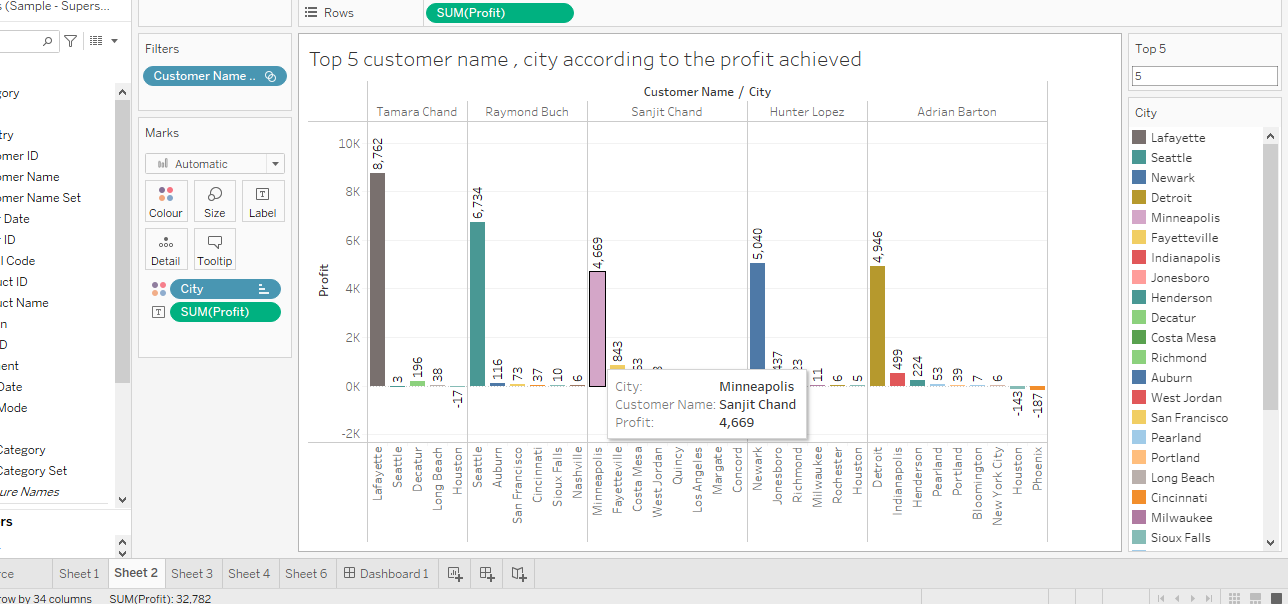
* And after selecting it just apply it to filter and we get our answer : -



* Now we can see clearly all the top 5 customers in the graph but lets check out their task according to the city and then see at which city they achieved the most profit.
* So, just select city and place it at colors on marks.



Another way to show the above graph is like this : -



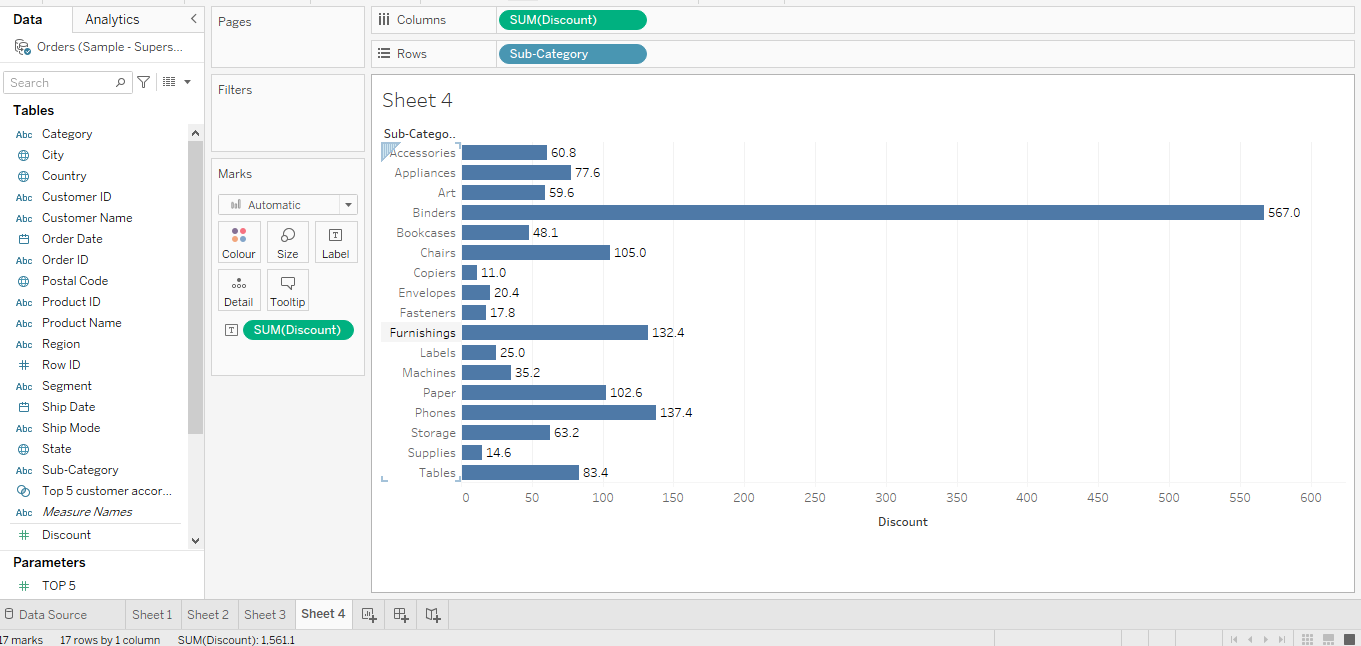
**Insights we get : -**

* **The insights we get is that the Tamara chand is the most profitable customer after the sales achieved by him at around 9 k dollars in total that means all the profit at all the cities he sold his stuff.**
* **Maximum profit range is till 9K dollars that could ever be achieved by any customer in any particular city.**
* **Over profit calculated marks on the fact of both loss as well as profit achieved like in the case of Adrian barton we can clearly see that , he has achieved a loss in city Houston.**

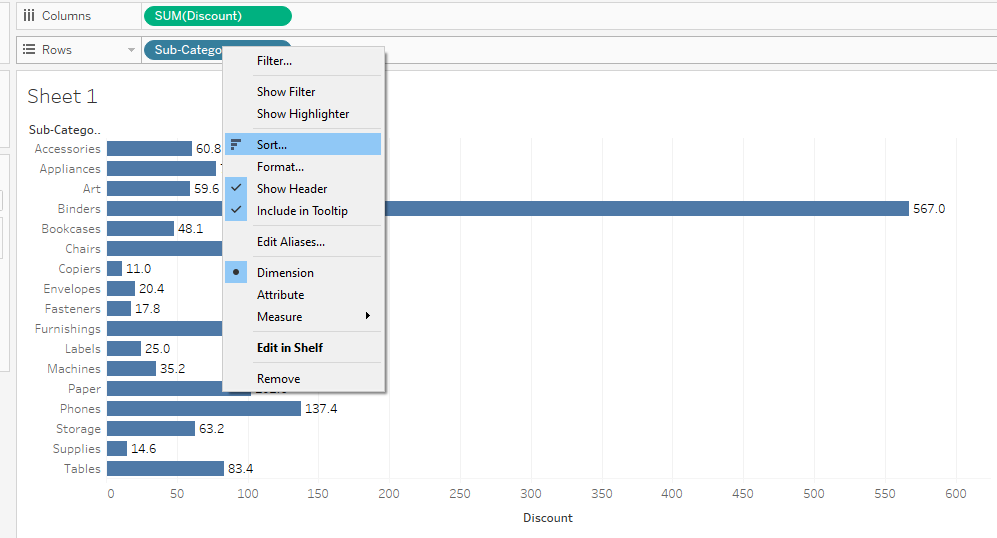
**ANSWER 3 : -**

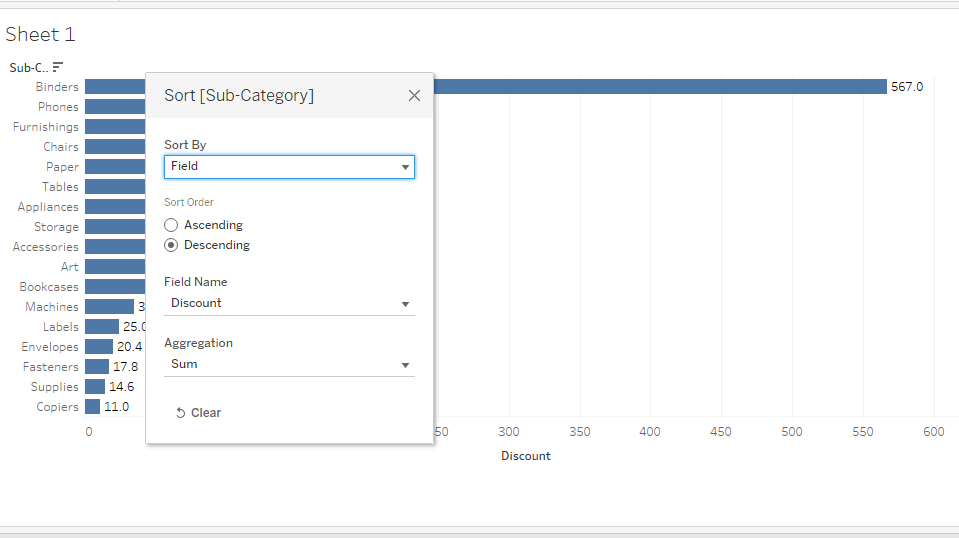
Now, we want to go ahead with top 5 most discounted subcategory with screenshots taken for each step: -

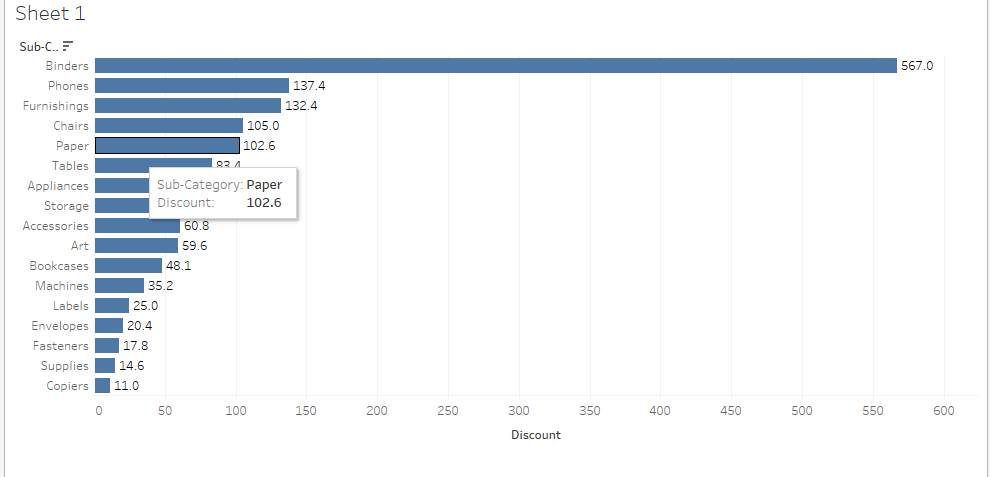
* Mainly, to achieve this first let double click the attribute which we want that is the subcategory
* Now, select the Discounted measure by same double clicking on it and it would then automatically take its place in columns.
* Putting sum of discount on the label Mark we we can see the total discount awarded on each subcategories.



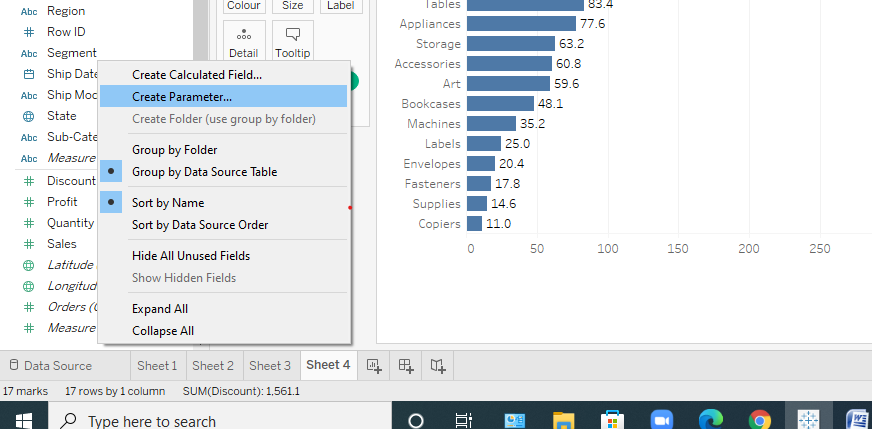
* But this doesn’t solve our problem to find out the top 5 discounted subcategories
* So lets right click on the subcategory attribute to make it more sorted and allinged according to discount shown Below: -

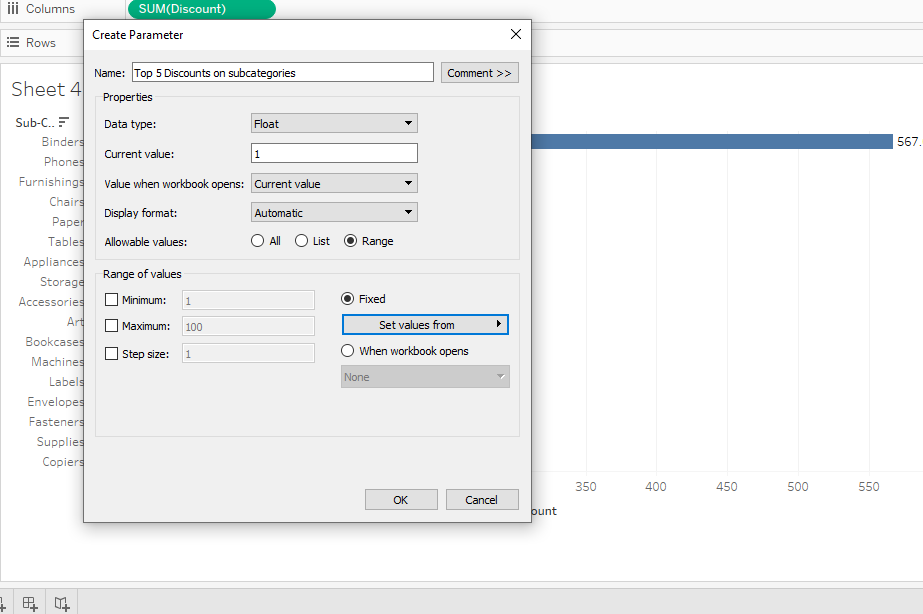




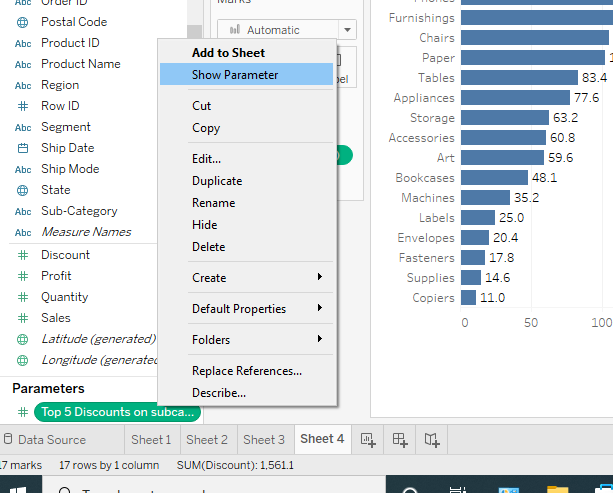


* Lets create a parameter that would act as a function to find the top 5 discount by explaining all the steps of it : -
* Right click on the side and click parameters
* Then give the name of parameter as Top 5 discount on subcategories so that we canknow which Top 5 to use.
* Set the range to be around 1 to 100 as we have taken here but we can take any range.

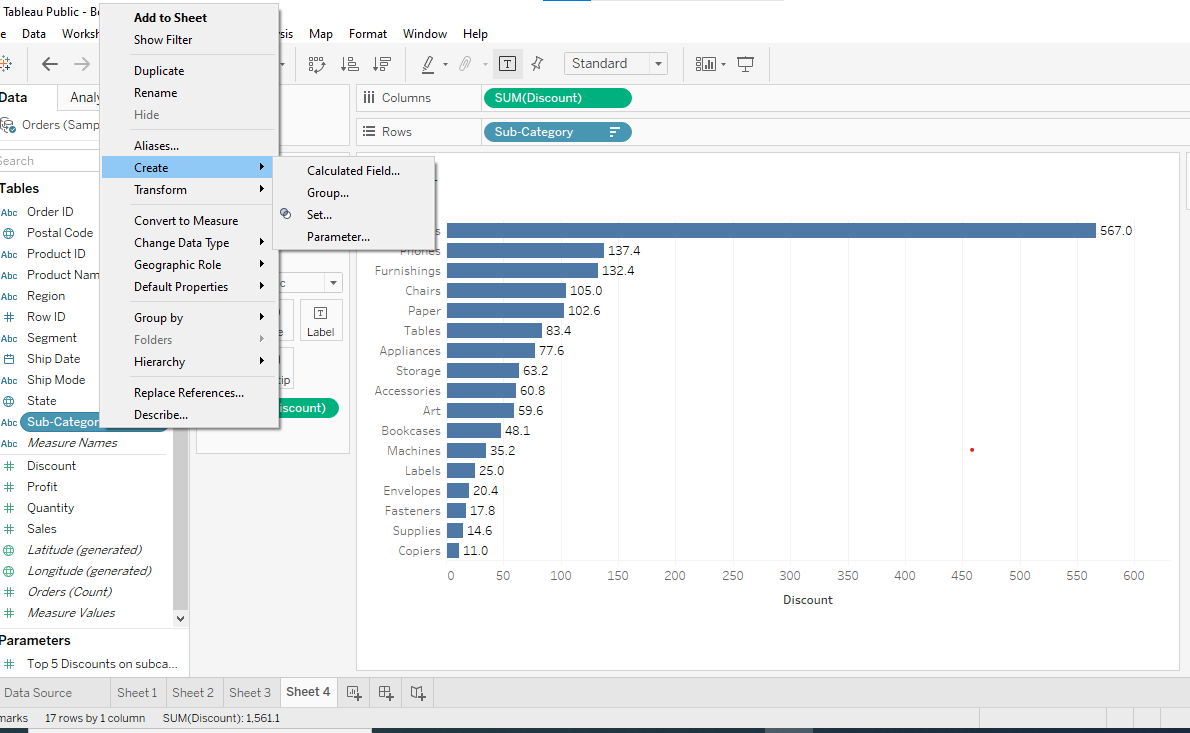




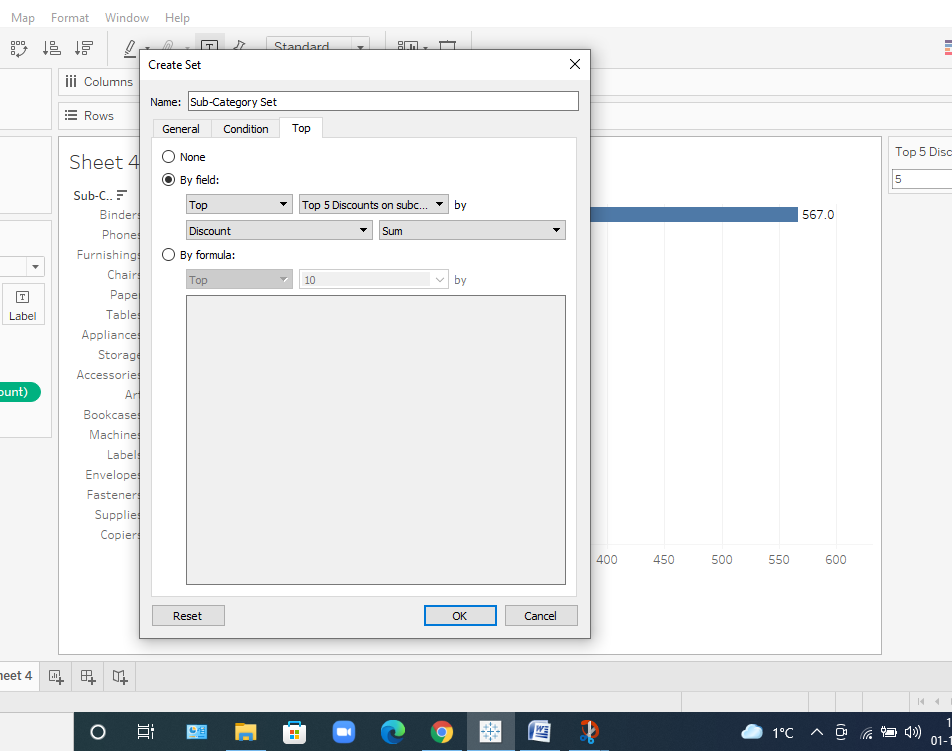
* Now lets show the parameter for our understanding by clicking show on the parameter created by us.



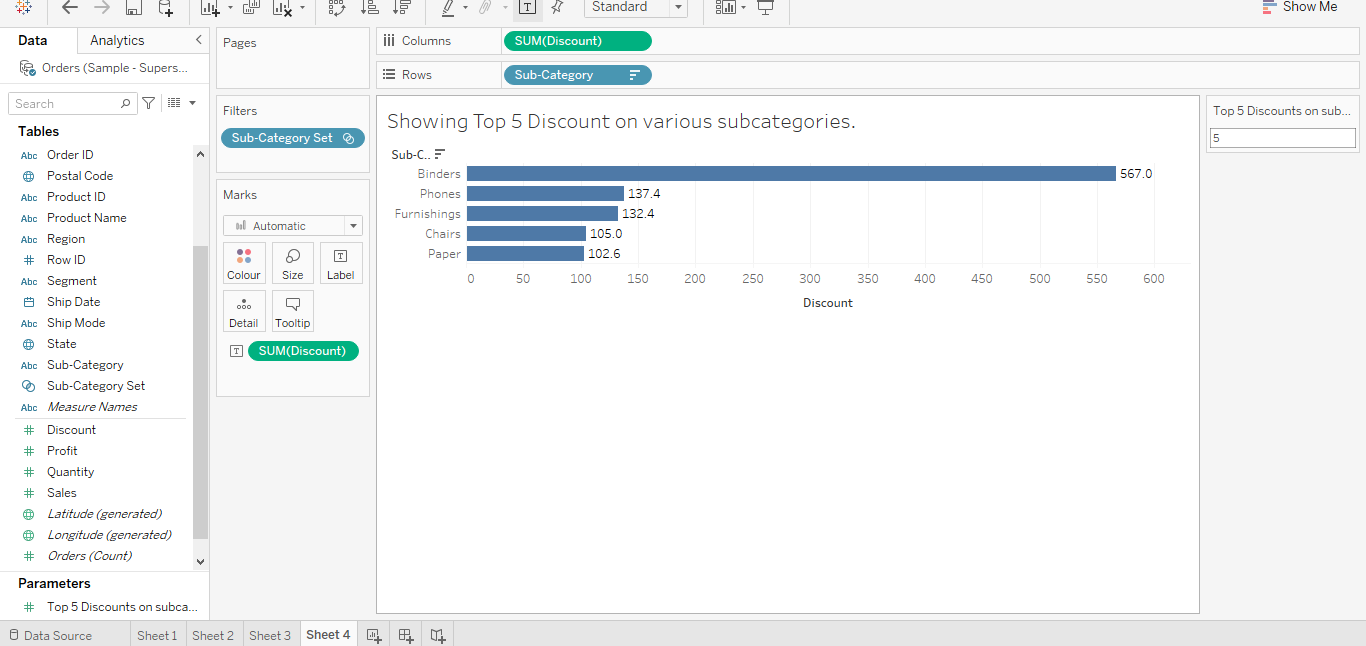
* But it cannot work upon the data so to make it happen we will need to create the set by
* Right clicking on the attribute for which we want the set to come and then clicking on the create
* After clicking on create , just click on Set like this : -



* Then clicking on the set button , a pop up will arise go to top attribute of it and then
* Make the changes by top 5 discount parameter created by us and apply it and we will get the final result as this.



* After achieving this we need to apply the new attribute where we have applied our parameter to the filter column and we will get the answer.



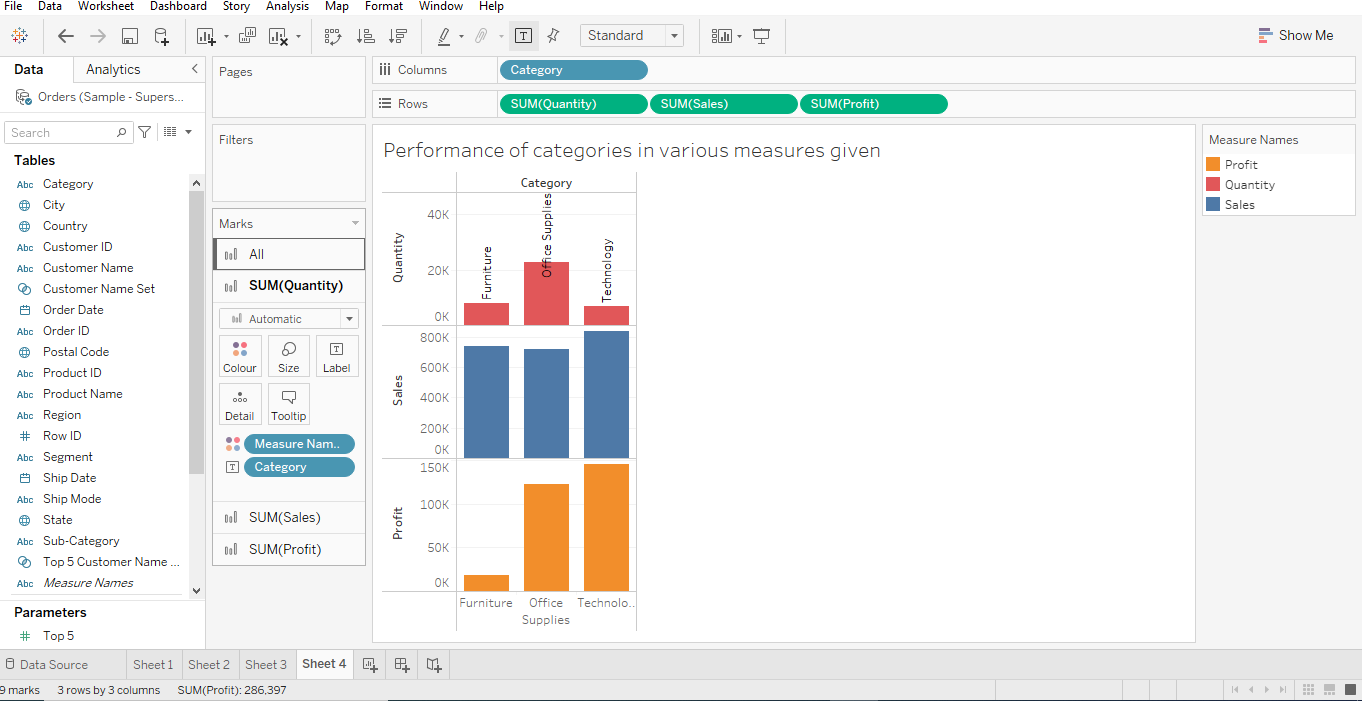
**Insights we get : -**

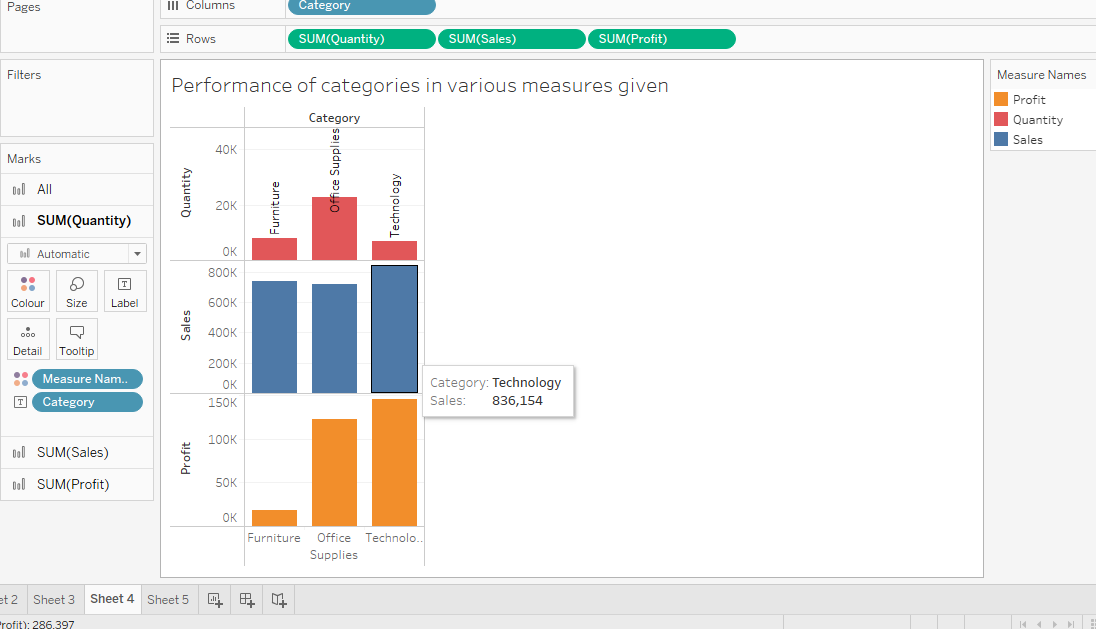
* **From the above graph achieved we can clearly see that most discounted subcategory is binder at around 567 k dollars .**
* **whereas in the top 5 set, paper is the least to around 102.6 k.**
* **Another insights we can see is that there is a huge gap of discount awarded between the second highest and first highest subcategory nearly around 430 k dollars.**

**ANSWER 4 : -**

Now, we want to go ahead to do analyse on categories with different measures: -

* It is fairly easy to perform the this analysis as we just have to double click on the category attribute that would give us the 3 categories :- furniture , technology and office supplies.
* After which we just have to double click on these measures like profit, quantity and sales and by default it would show us the sum of it and hence it is done.





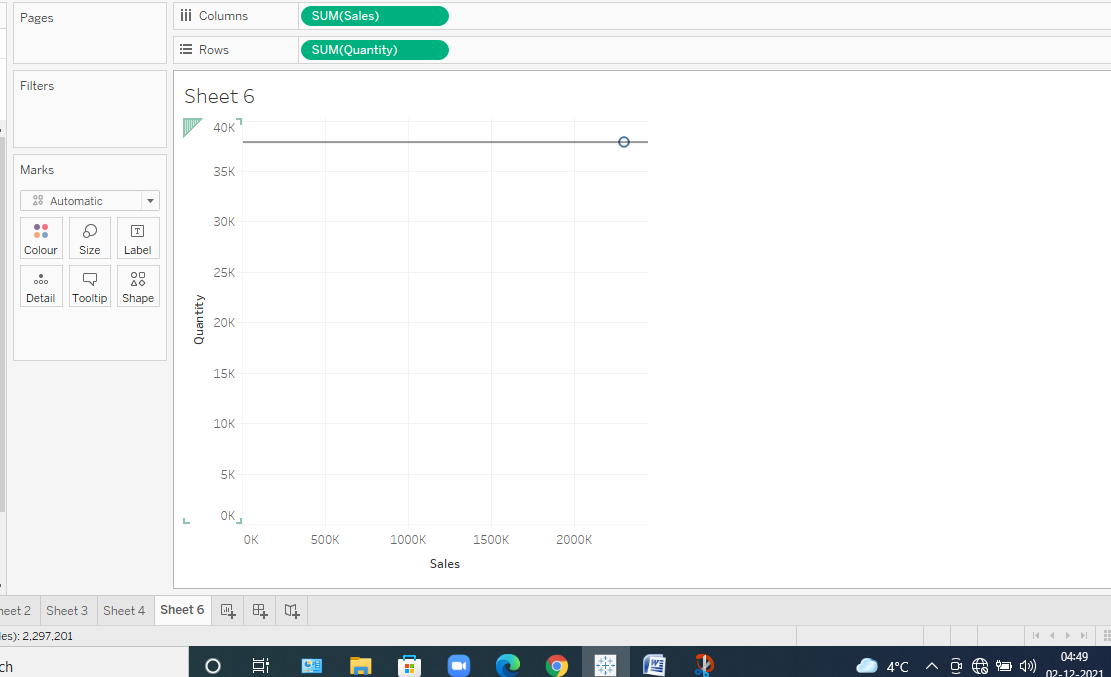
**Insights we get : -**

* **Quantities for office supply has been the maximum with it being 22, 900.**
* **We can see that there is no such relation between quantity and sum of sales as more quantity of items were produced in case of office supply but their overall sales is the least.**
* **In case of technology, Sales and profit is very high nearly 836000 and 145000 respectively but this doesn’t justify that sales is directly proportional to sum of profit earned as in cases of furniture there is good amount of sales but still the profit earned is very low near about 18000.**

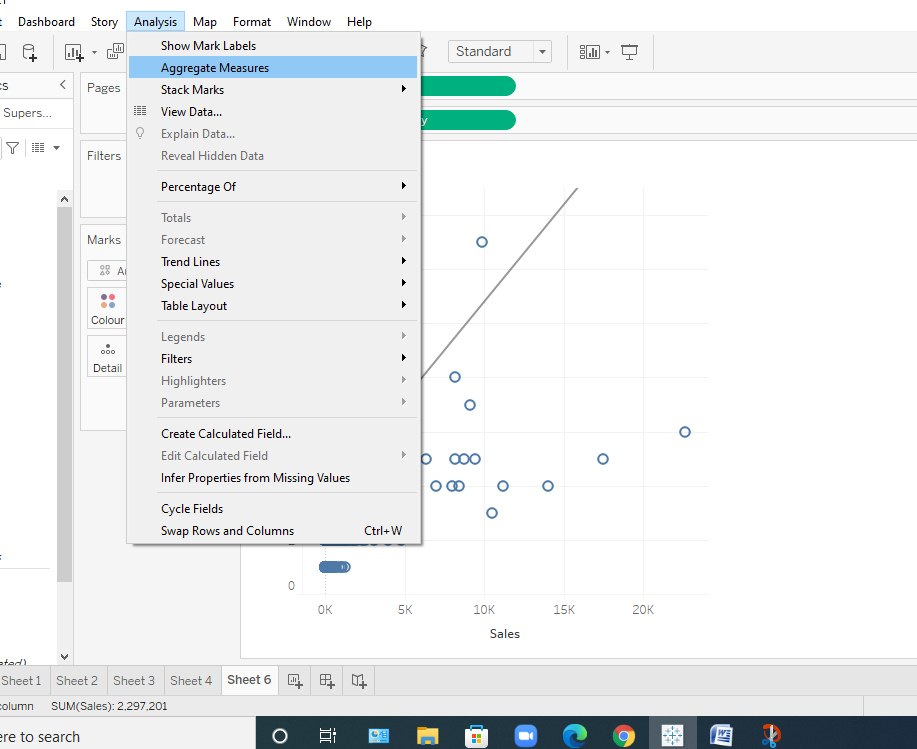
**ANSWER 5 : -**

Now, we want to go ahead and try to find out that whether there is any correlation between the quantity and the sales produced

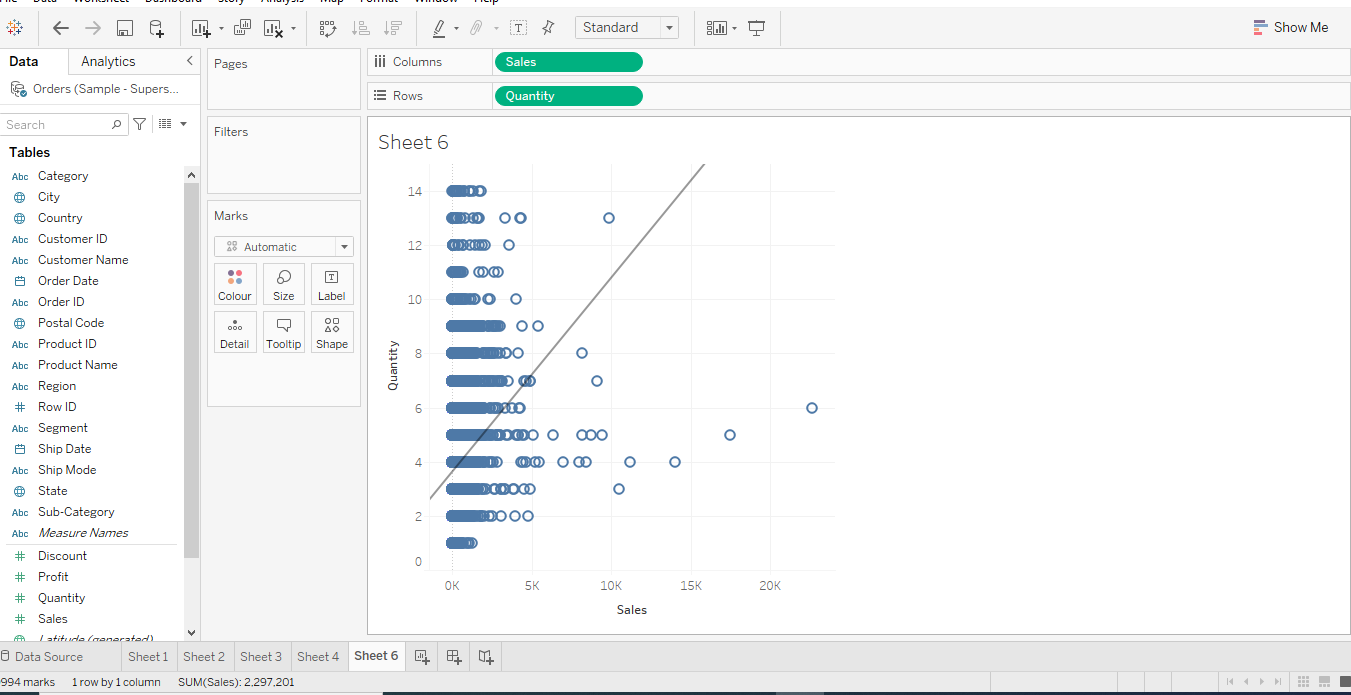
* It is fairly easy to perform this analysis as we just have to double click on the quantity and the sales
* The problem that would arise would be around this corner that it is showing sum of it and showing just a dot in analysis graph.



* This is the dot and it is not showing any correlation
* So we first get rid of aggregated values and then look for it



* Now lets see the correlation



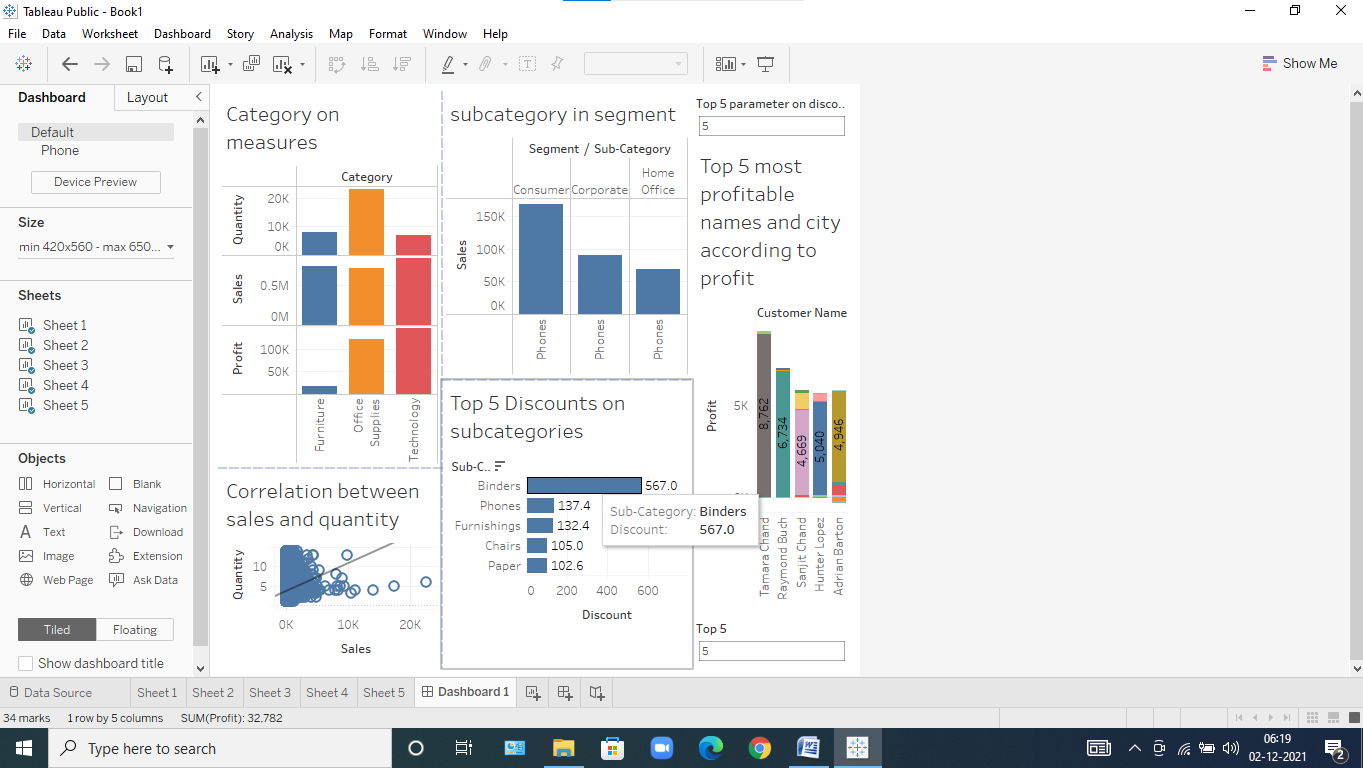
**Insights which we get : -**

* **We can clearly see that the trend line is showing that yess there is some correlation between the quantity and sales.**
* **We can infer from it that as quantity is rising by 2 K the sales is rising by 5 k but after that when the quantity is rising and reaching to 10k the sales remain constant with it.**
* **According to me, the correlation will be around 0.5 at a max as it is neither negative nor positive.**

**ANSWER 6: -**

By this time we have created so many charts on different sheets now is the time to collaborate them in one single sheet and try to understand them all in one go and try to find out different insights from it.

* Mainly to achieve this we need to go ahead with and place all the sheets inside it and for that to take place we just have to drop the sheets in the dashboard and adjust the features of each sheet to represent it in desirable manner.



We have explained about the correlation of almost all of the sheets in earlier stages.

THANK YOU