oronto Business College

Data Visualization

Assignment (6)

Case study - Tableau charts

|  |  |
| --- | --- |
| **Course** | 2021F-T2 AISC2001 - Data Visualization 01 |
| **Instructor** | Gamal Ali |
| **Students** |  |
|  | Charitha Priya Dongari – 500191110 |
|  | Venkata Satya Narayana Reddy Katterapalli – 500190693 |
|  | Rumana Banu Iliyas Ahmed – 500186725 |
|  | Rohan Bhatt – 500187633 |
|  | Shivam Jolly – 500196452 |

**Table of Contents**

[LOADING DATA IN EXCEL INTO TABLEAU 1](#_Toc89372267)

[Most Sold Subcategory in Each Segment 3](#_Toc89372268)

[Top 5 Most Profitable Customer Name, City, and their Profit 6](#_Toc89372269)

[Top 5 Most Discounted Subcategory 12](#_Toc89372270)

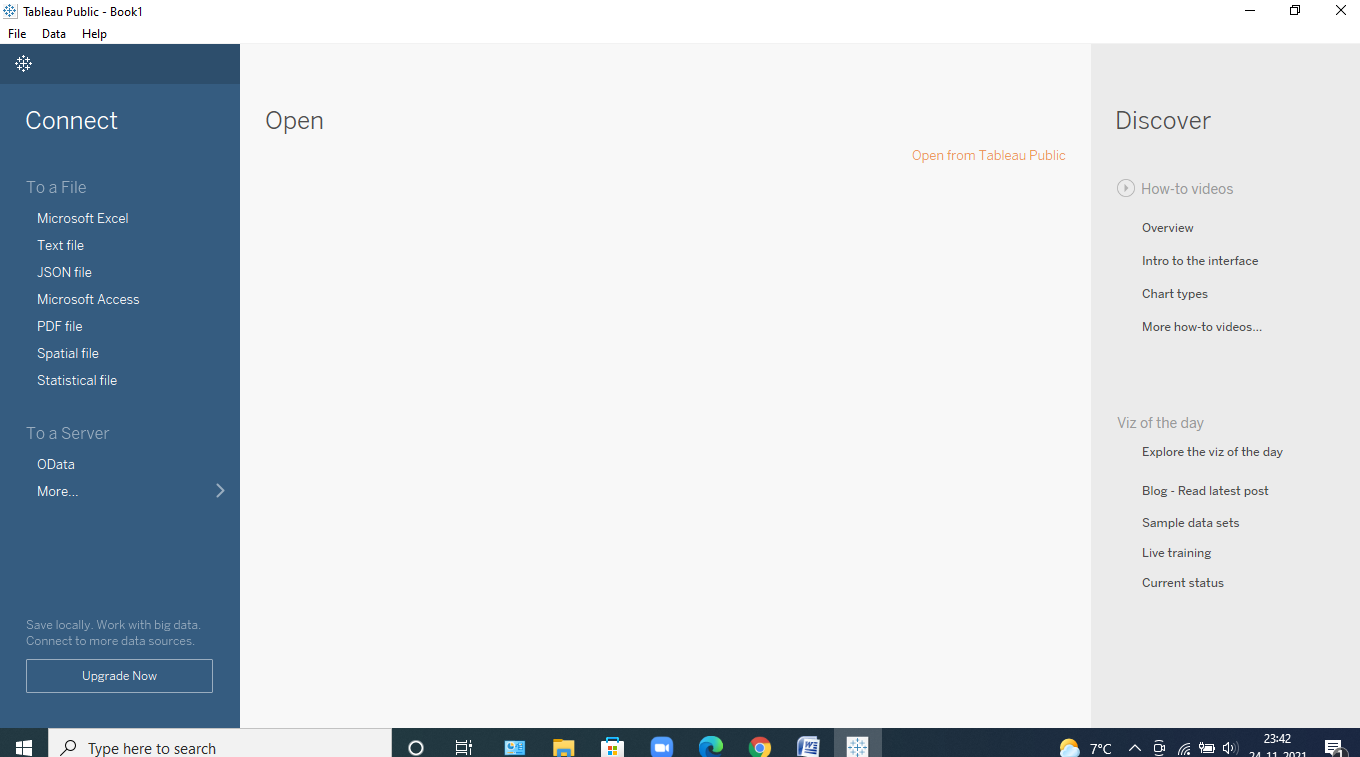
[Deciding which Category Performs Well (sale, profit, return, and quantity) 17](#_Toc89372271)

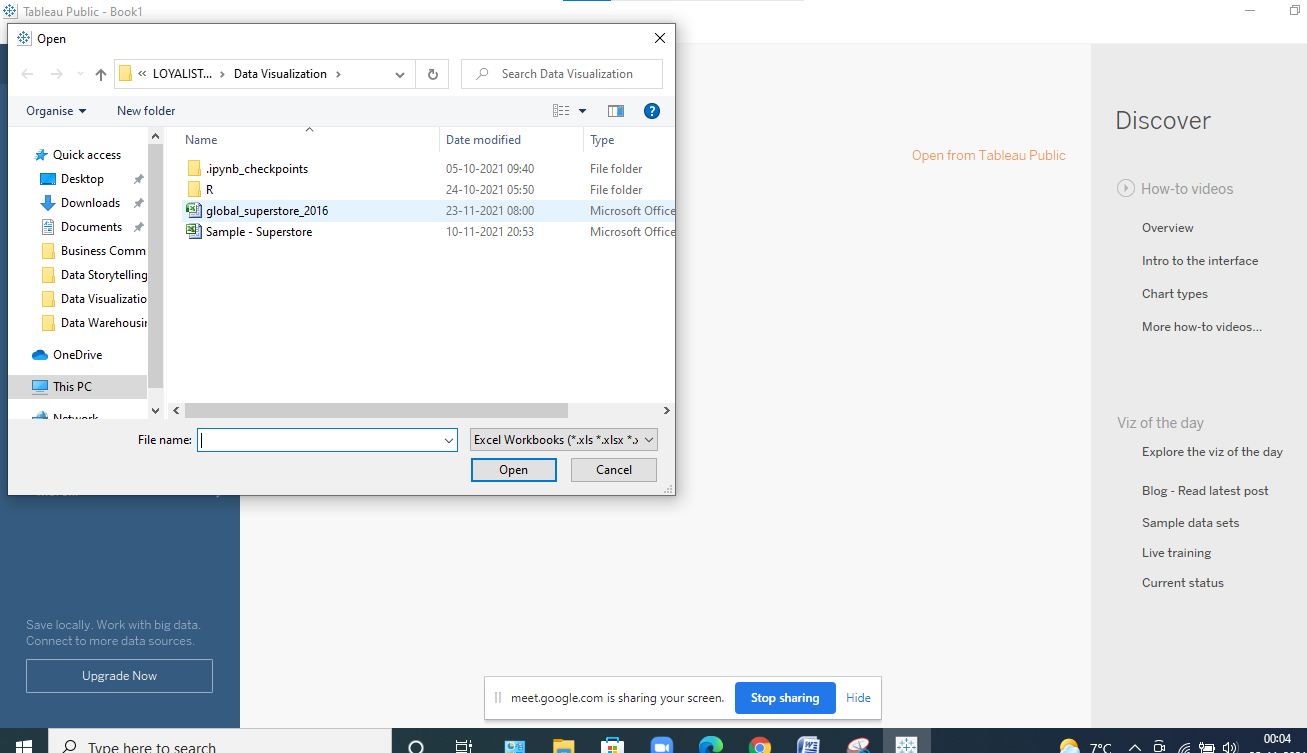
[Correlation Between Quantity and Sales using Scatter Plot 18](#_Toc89372272)

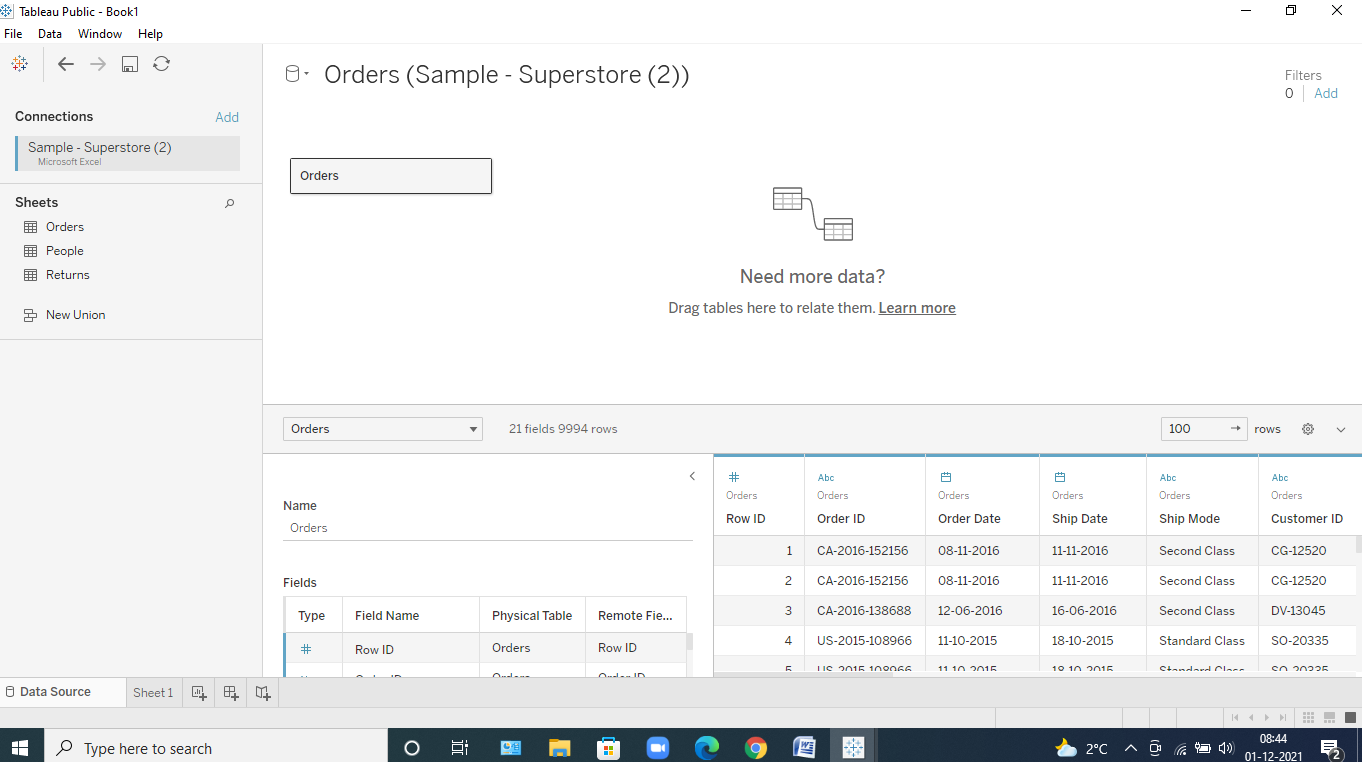
[Create a Dashboard 20](#_Toc89372273)

# 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.



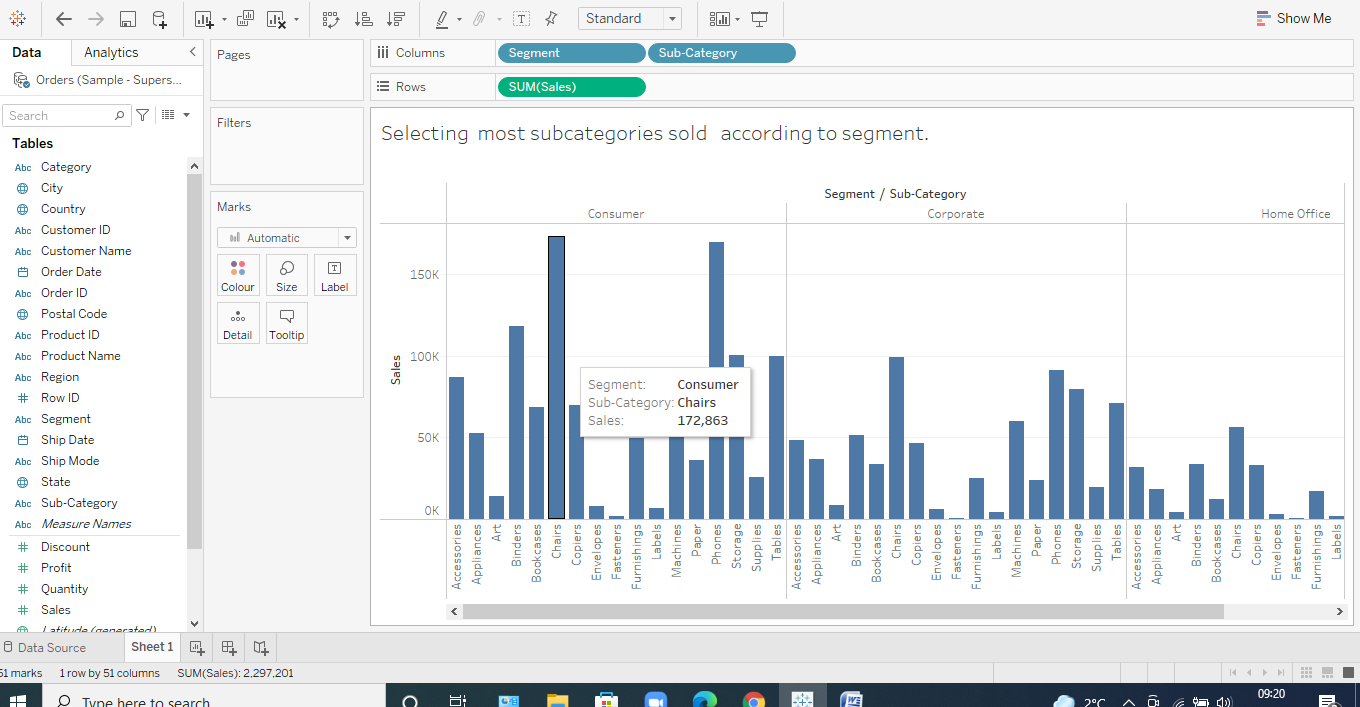
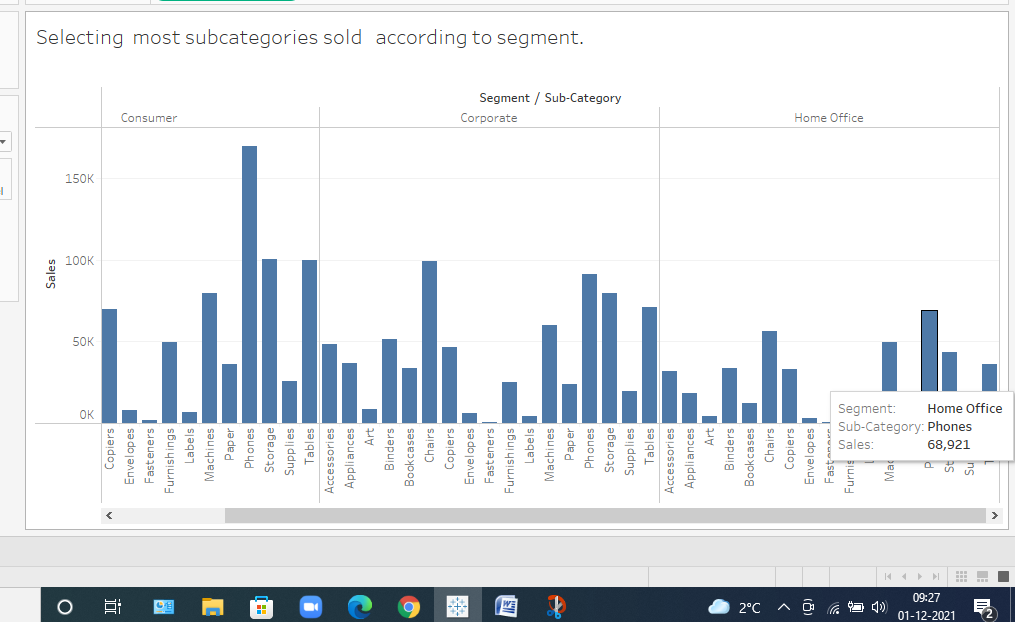
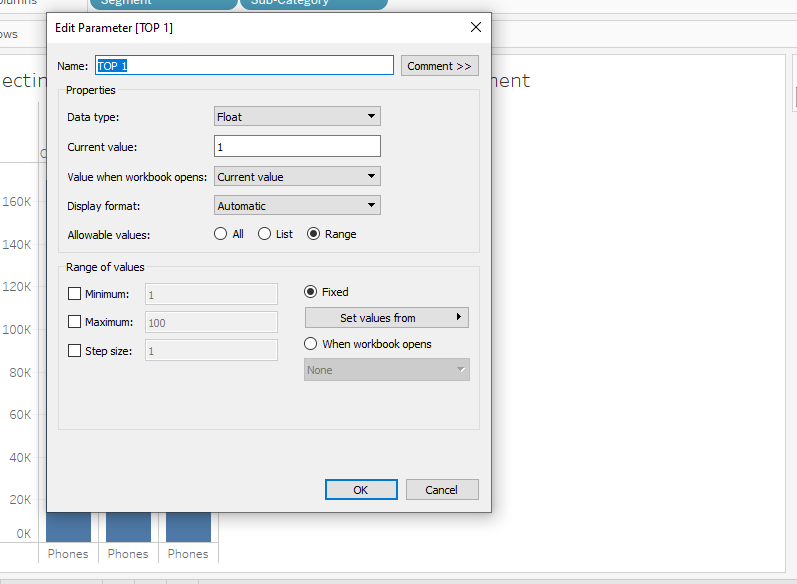
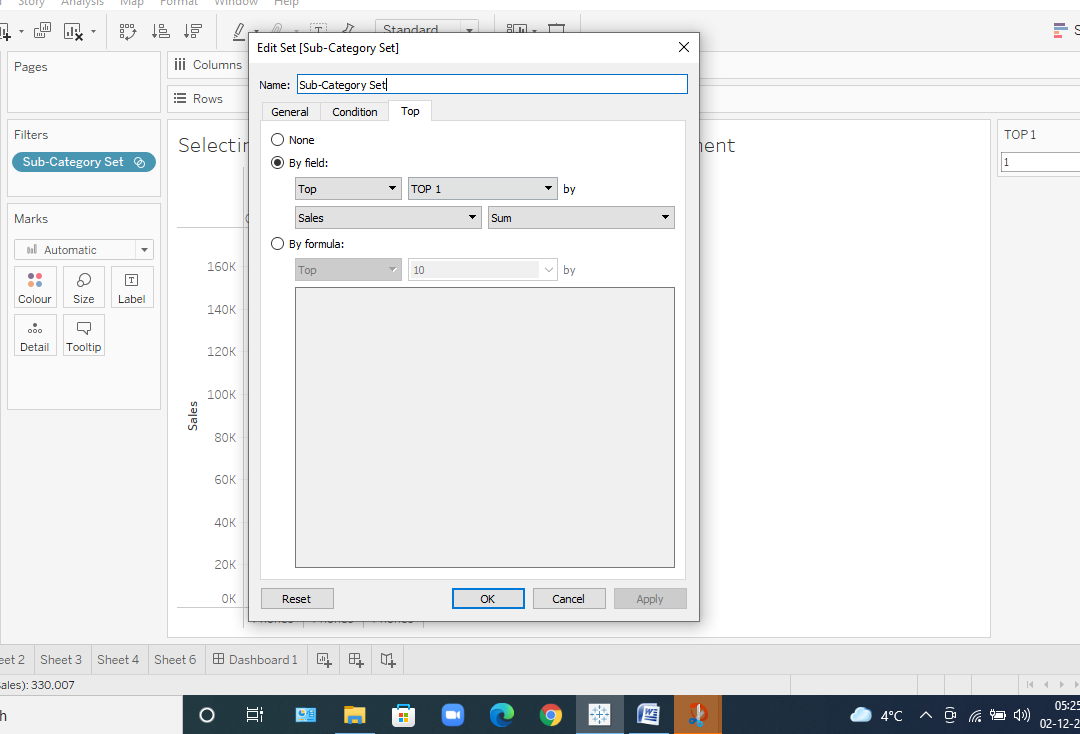
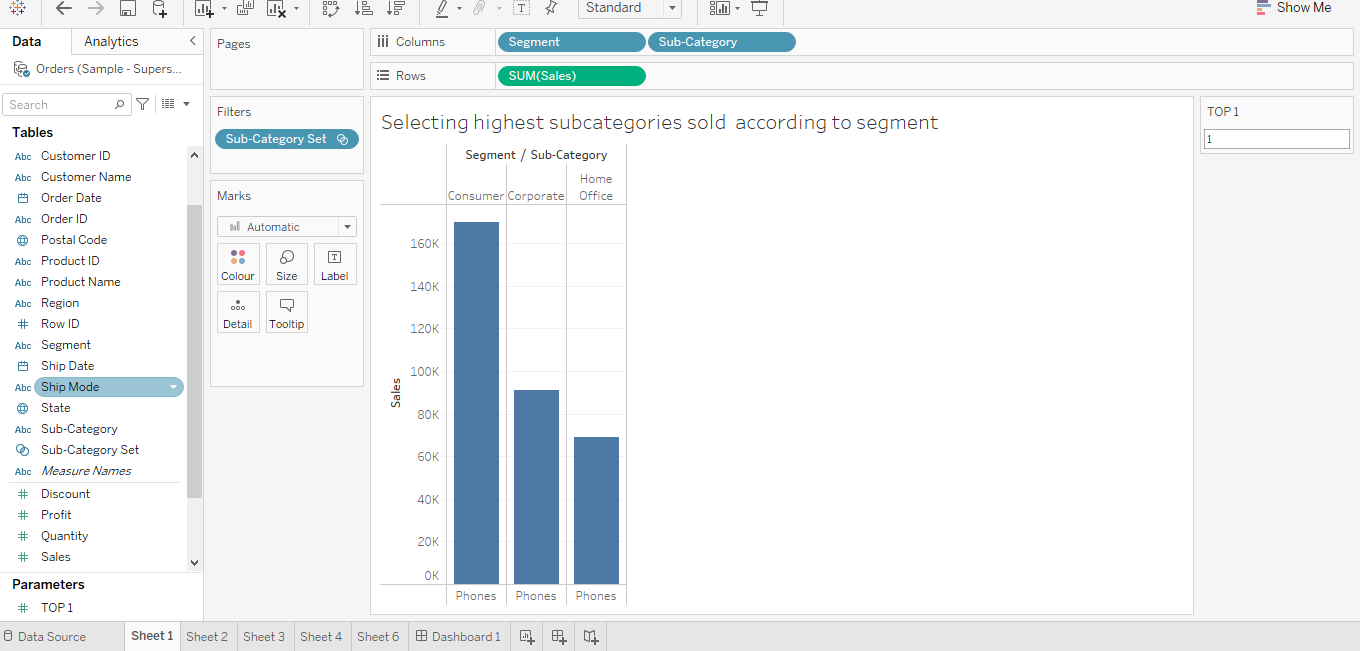


* Once the data gets loaded, we get to see how many tables are inside it like in this case we can clearly see that we have been provided with nearly about 3 tables and then we have to see whether they need to be joined or not like in this case there is no need to join tables and simply use only one dataset 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.   
  

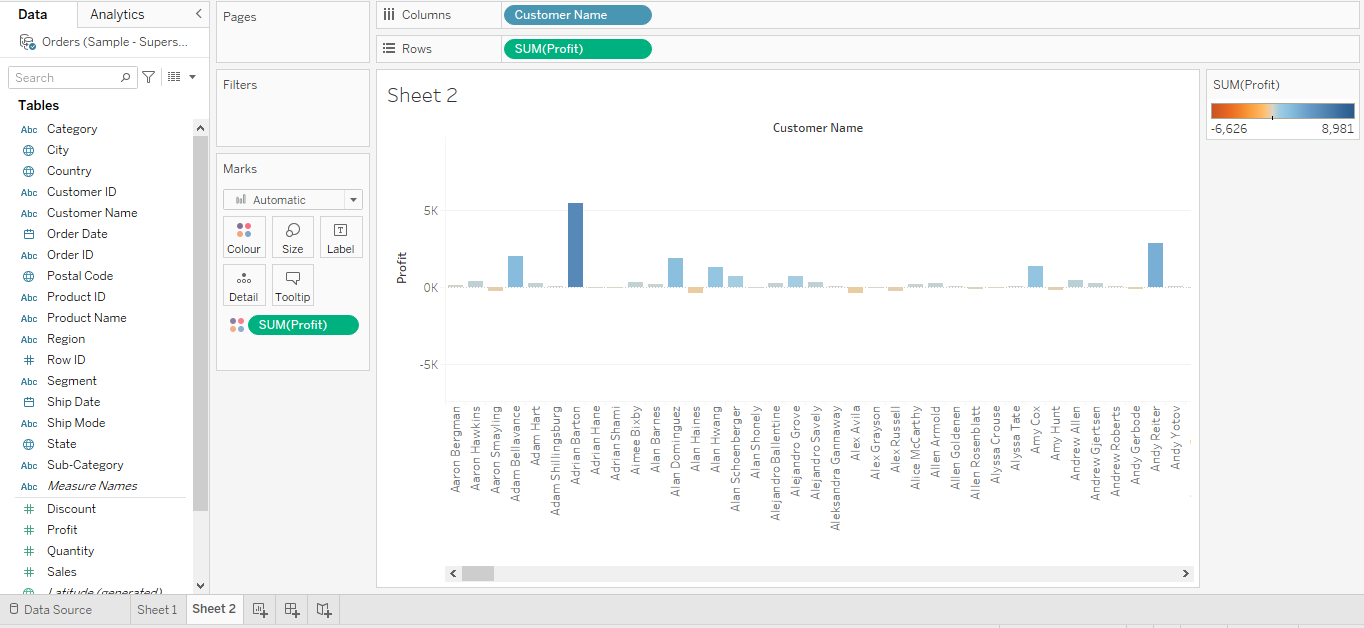
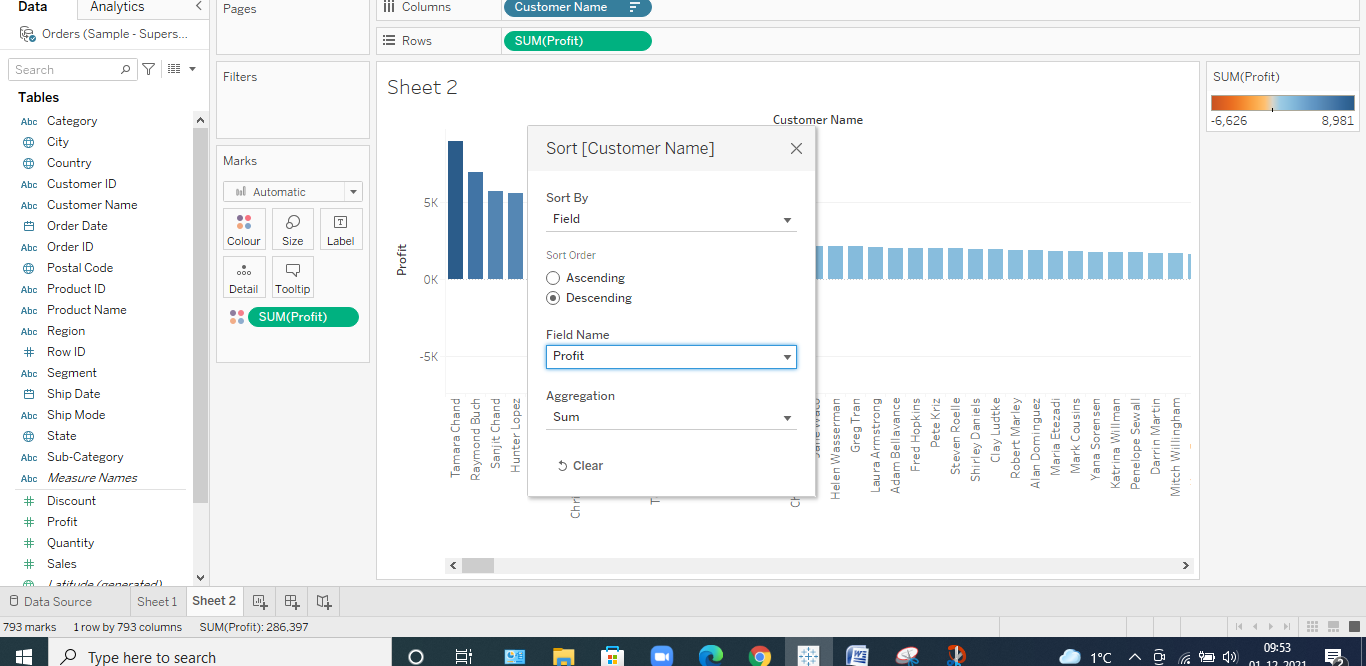
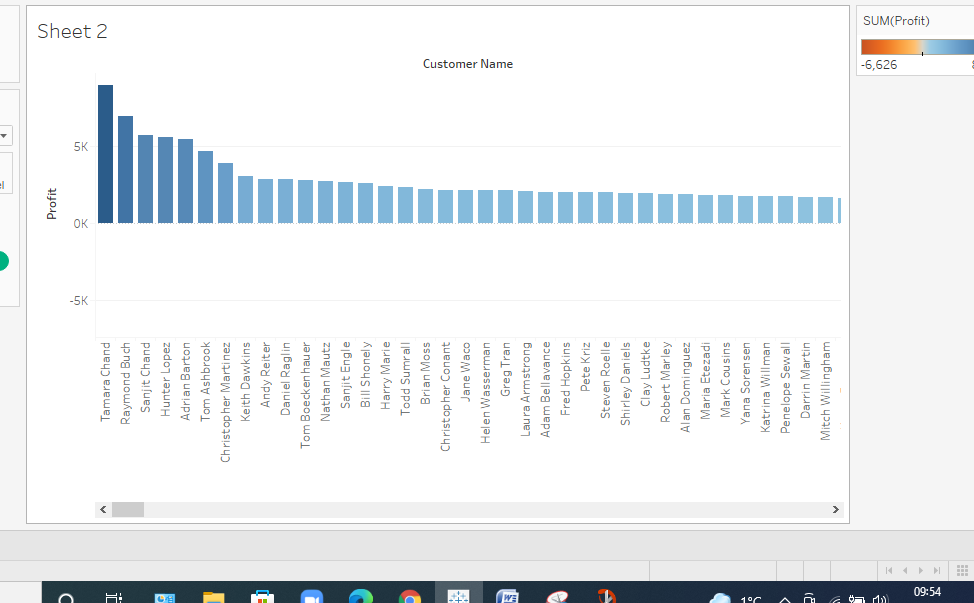
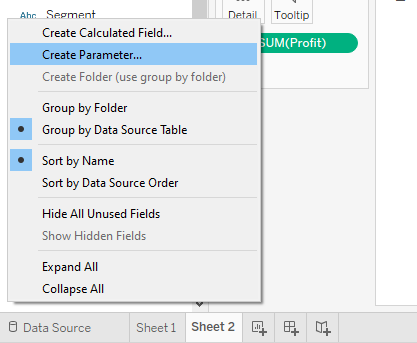
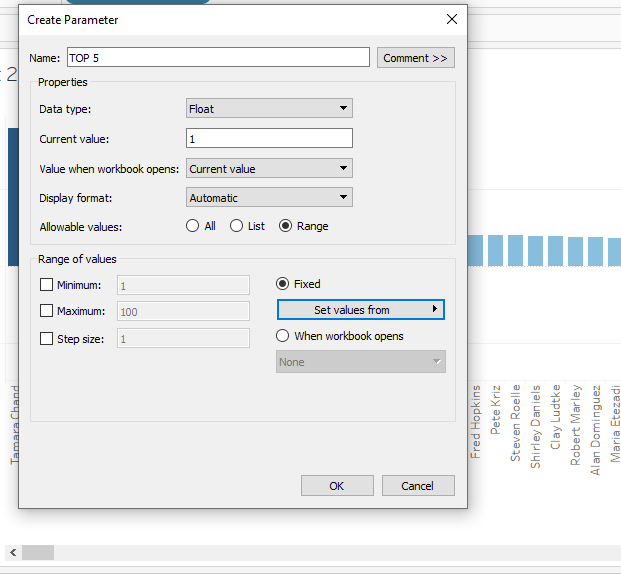
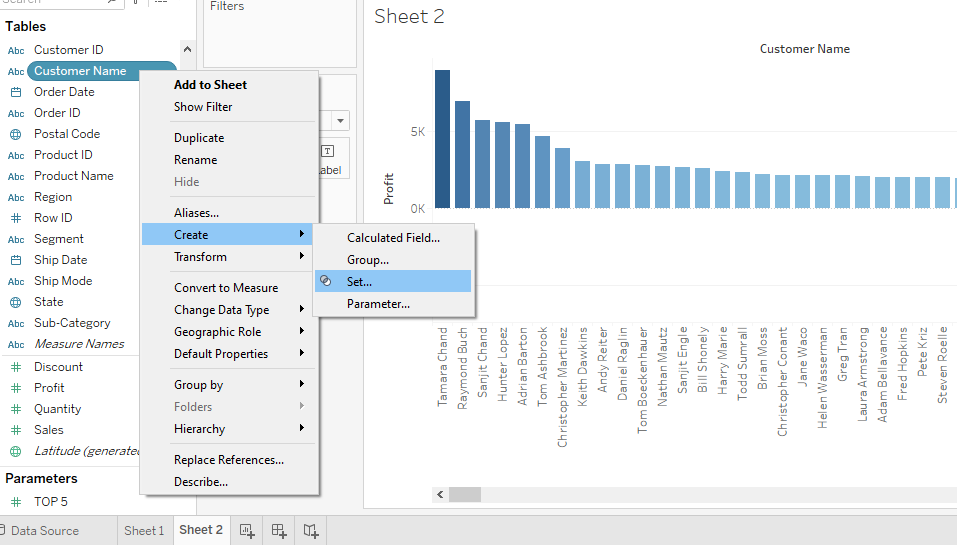
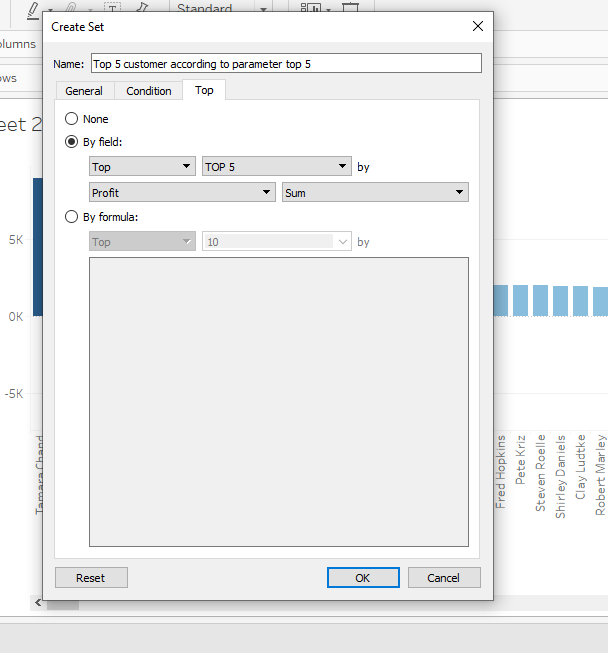
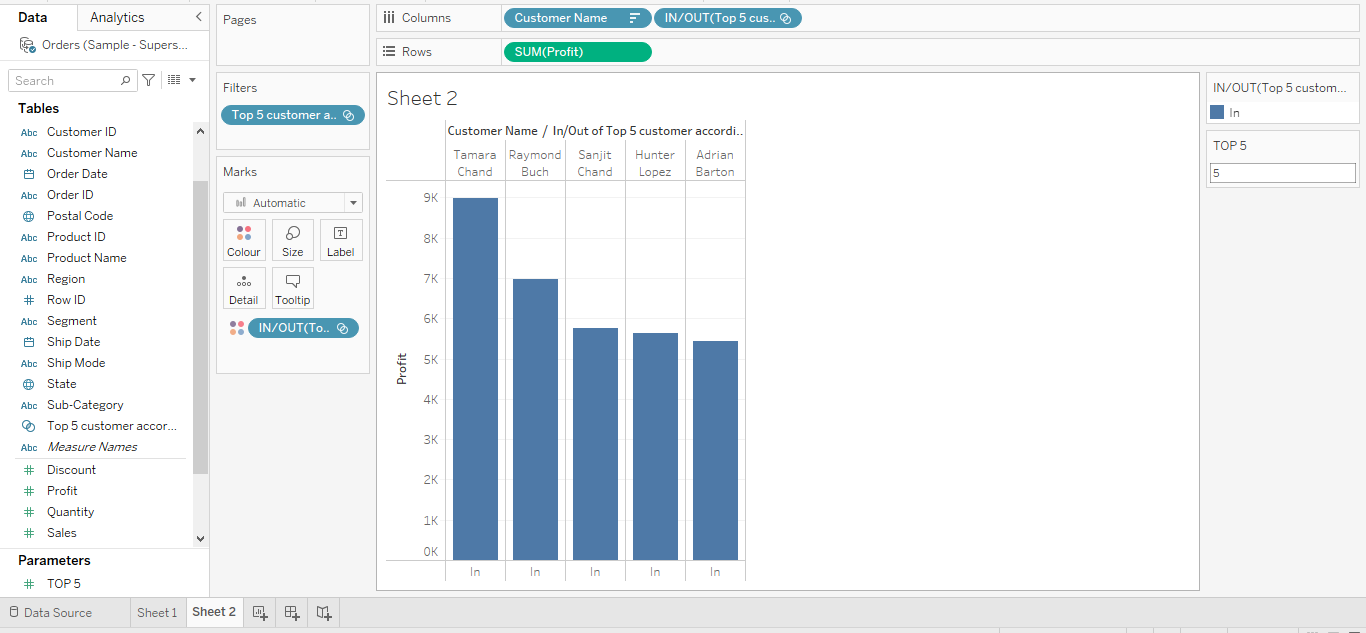
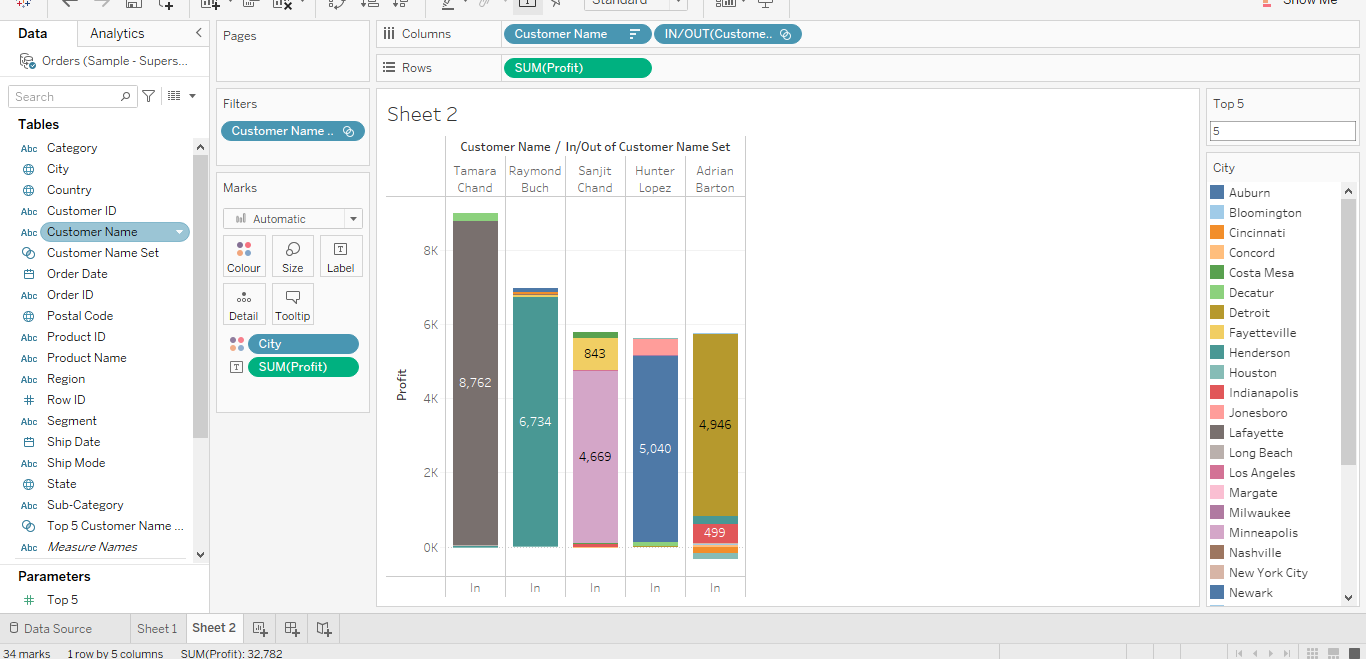
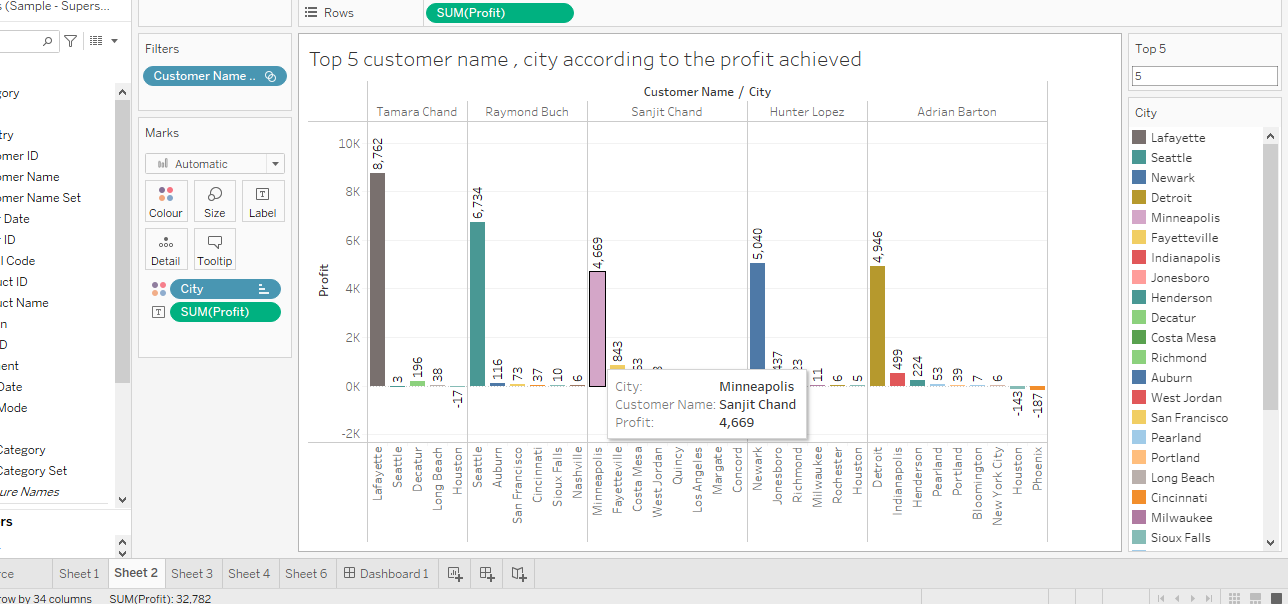
Now that we've uploaded the data to the software, let's get right into the Sheets, where we can see the analysis, we're going to perform.

* 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.

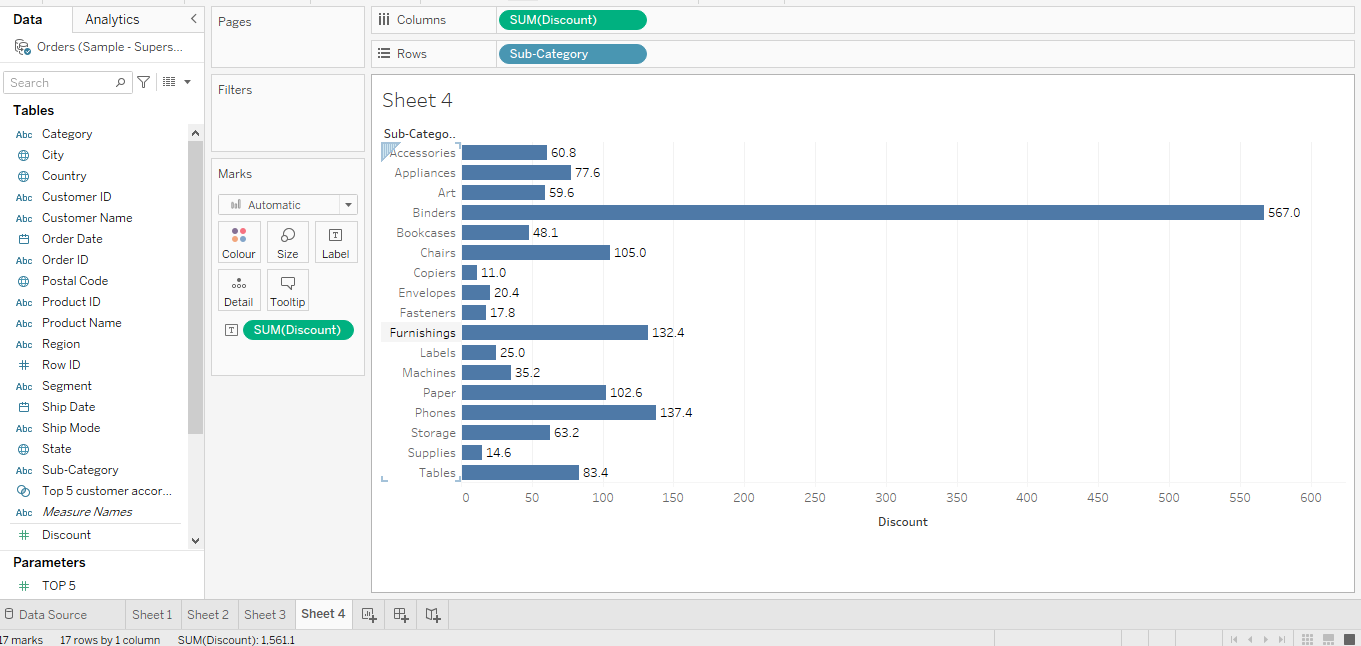
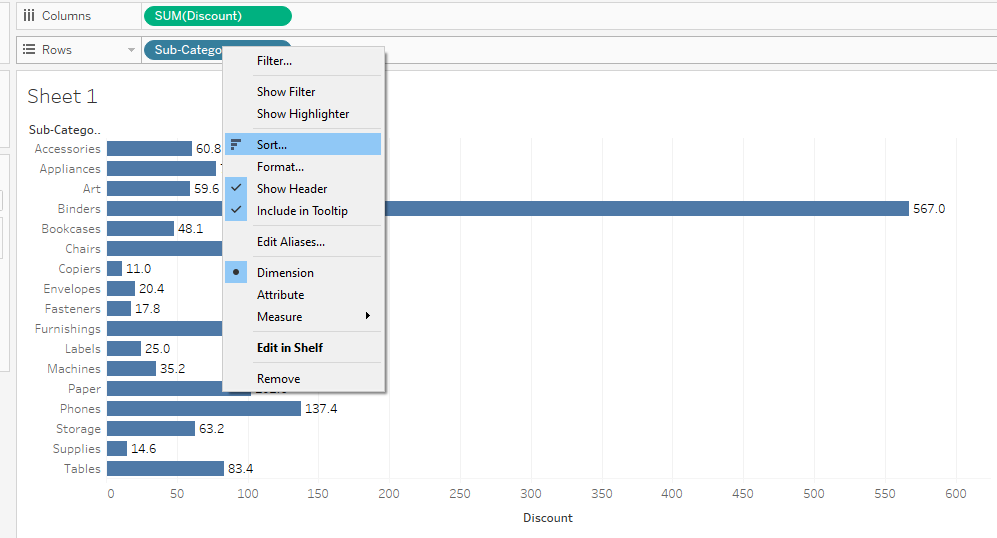
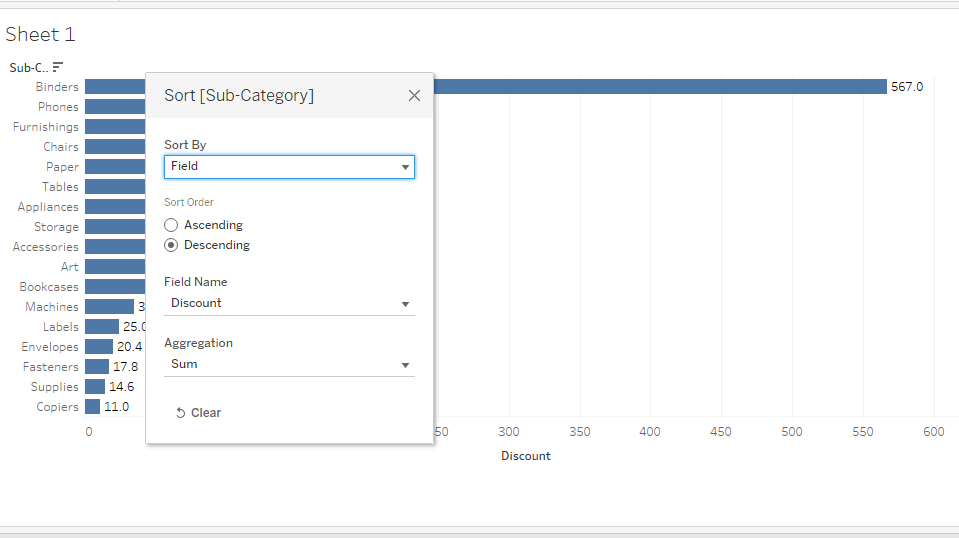
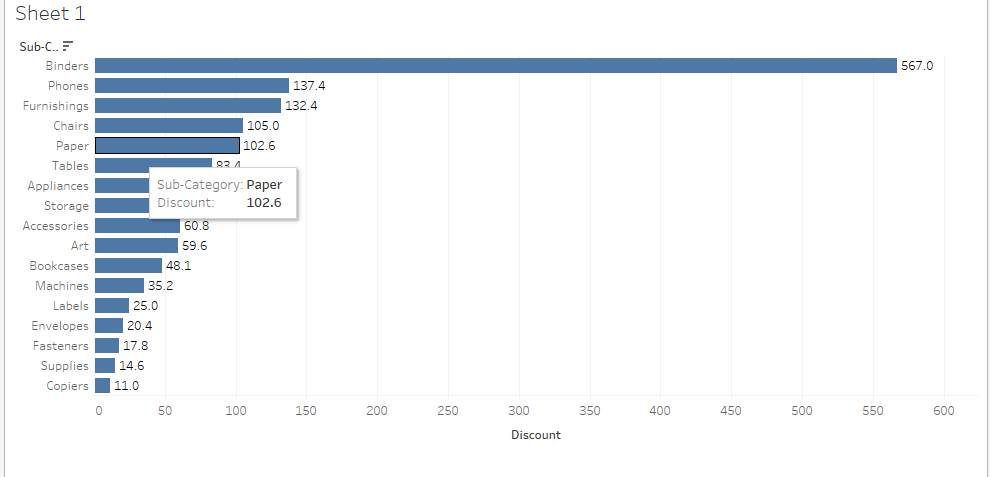
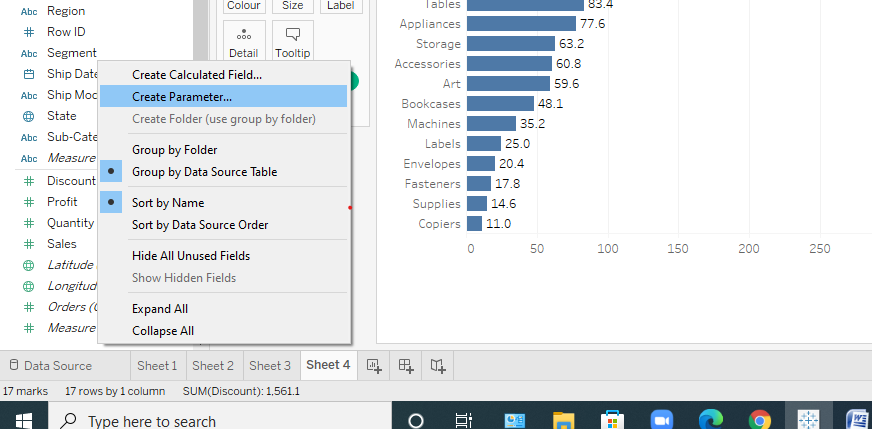
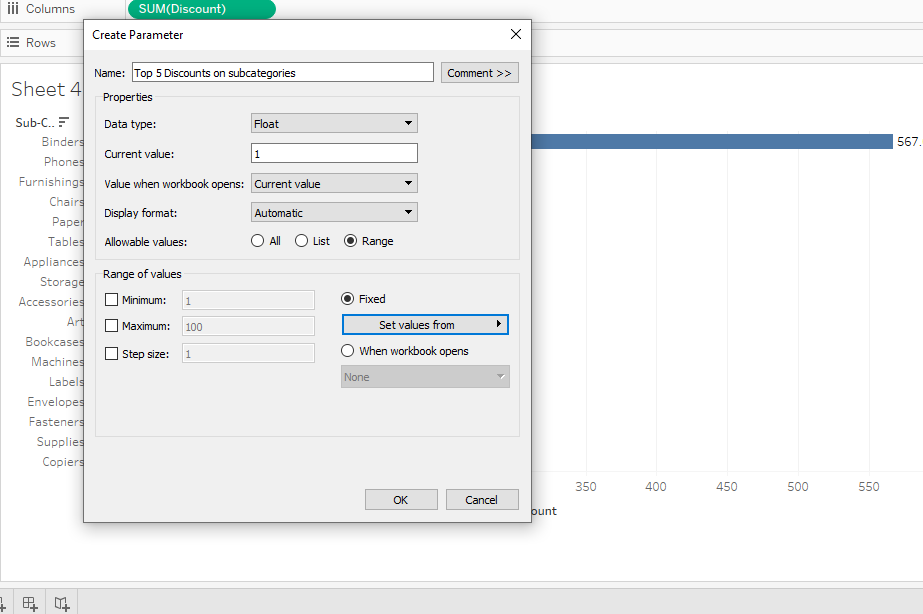
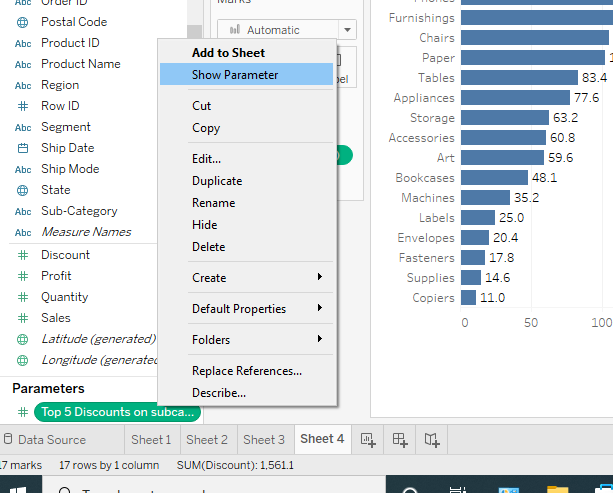
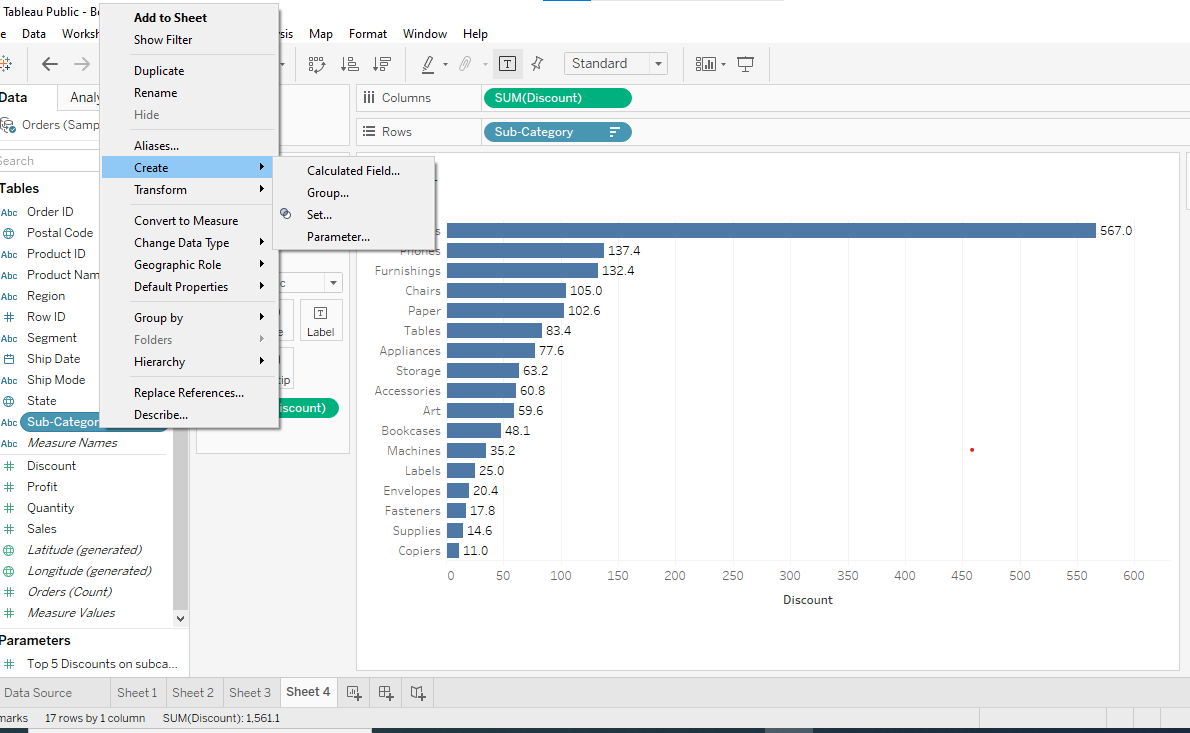
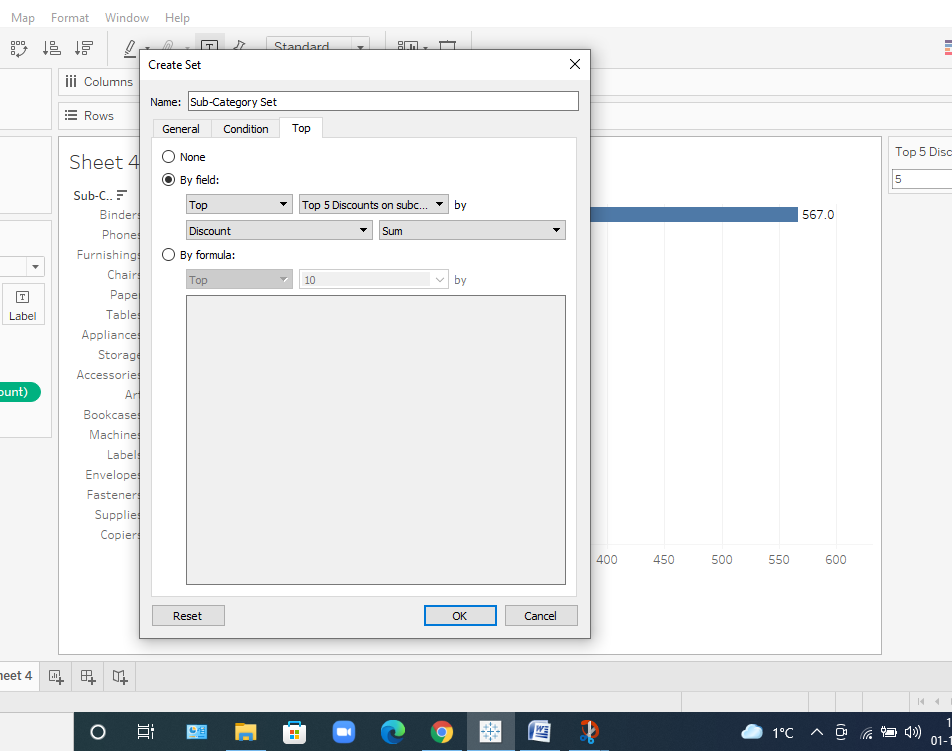
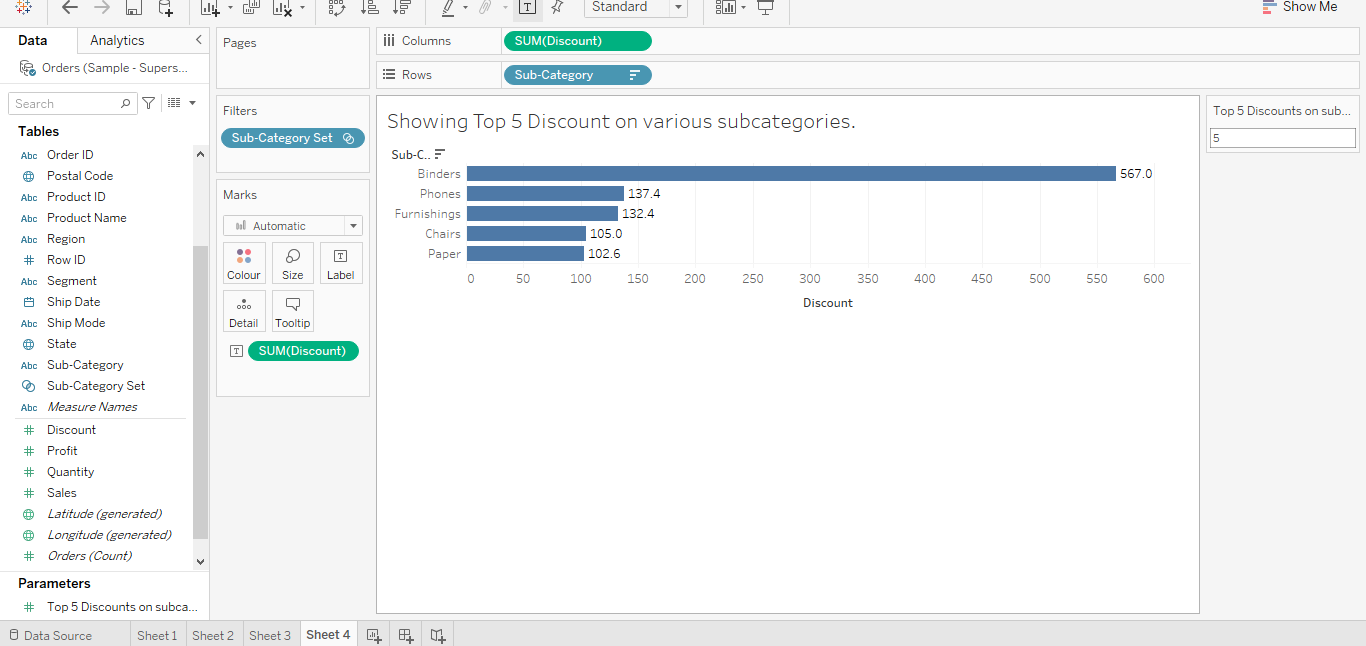
# Most Sold Subcategory in Each Segment

* Double on segment, sales and subcategories, which got placed inside the rows and columns 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.
* Let’s 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.   
    
  
* From this dataset we found that there are 3 segments corporate, home office and consumers.
* Sales in the phone’s subcategory were extremely strong, with 160,000 and 99141 units sold in the consumer and corporate segments, respectively, while 68921 units were sold in the home-office segment.
* The phone is the single subcategory with the most sales in each segment, followed by the chair.

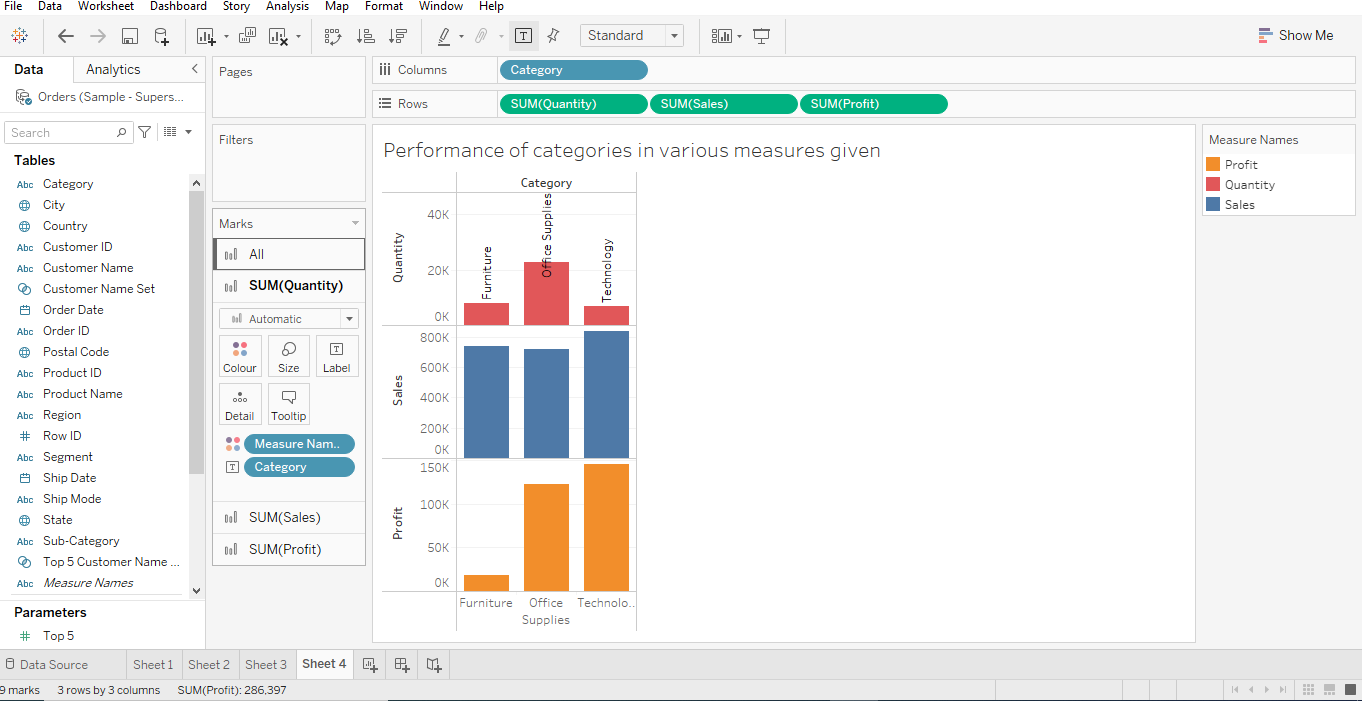
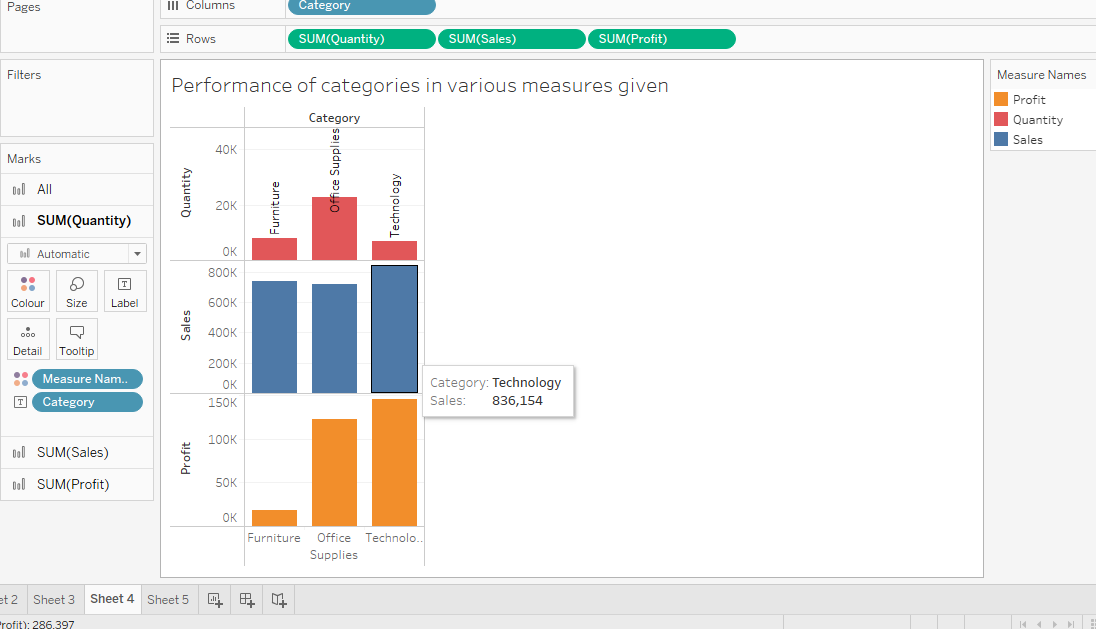
# Top 5 Most Profitable Customer Name, City, and their Profit

* So, lets show the top 5 profitable customer name with steps and their screenshots to achieve it.
* 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’s name select sort, change to filter profit and descending order and we get this,  
    
    
    
    
  
* Let’s create parameter for this for helping us to take out only 5 top customer names as below,  
    
  
* 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 let’s 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 as below,  
    
  
* According to our findings, Tamara Chan is the most profitable consumer, with sales totaling roughly $9,500, which includes all profits made in all of the cities where he sold his goods.
* The maximum profit range that any customer in any city might possibly achieve is up to $9,000.
* Over profit calculated marks both loss and profit earned, such as in the example of Adrian Barton, where we can plainly see that he has made a loss in the city of Houston.

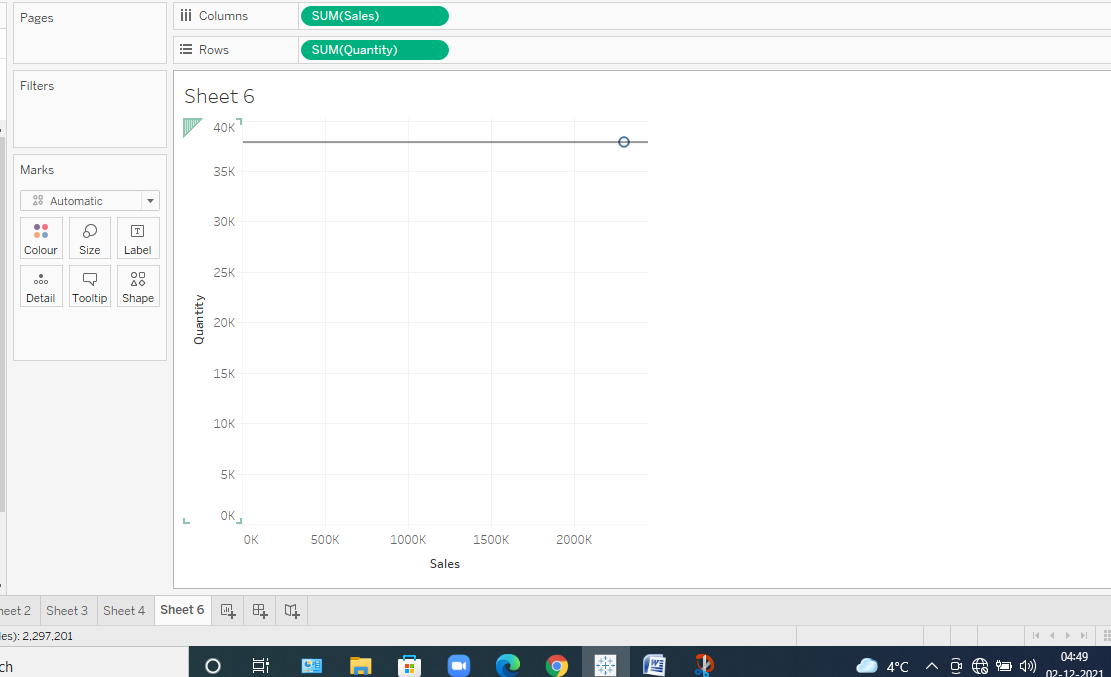
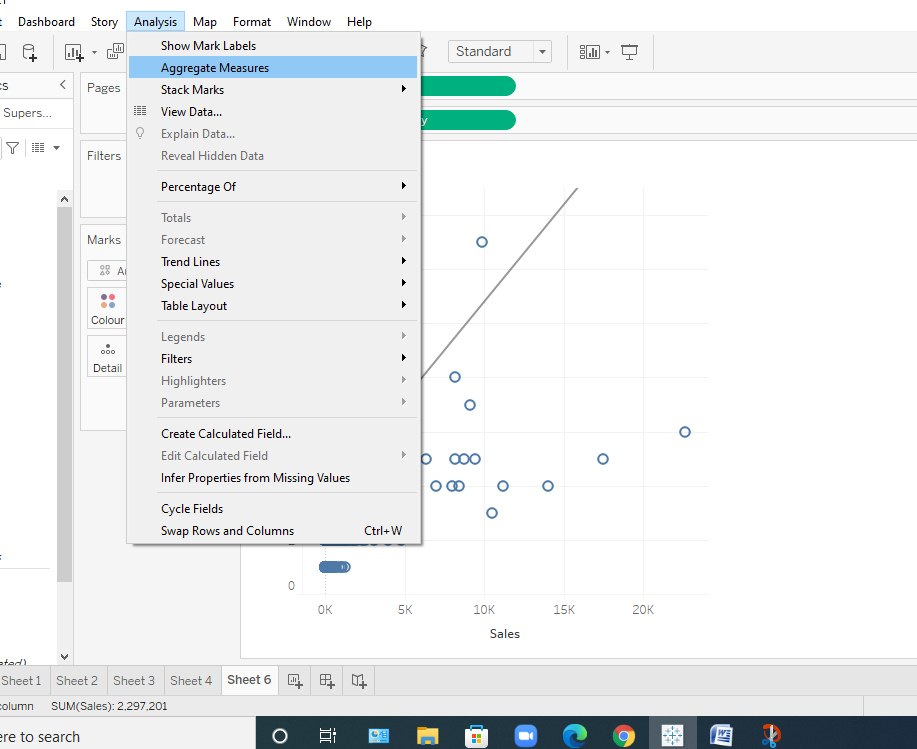
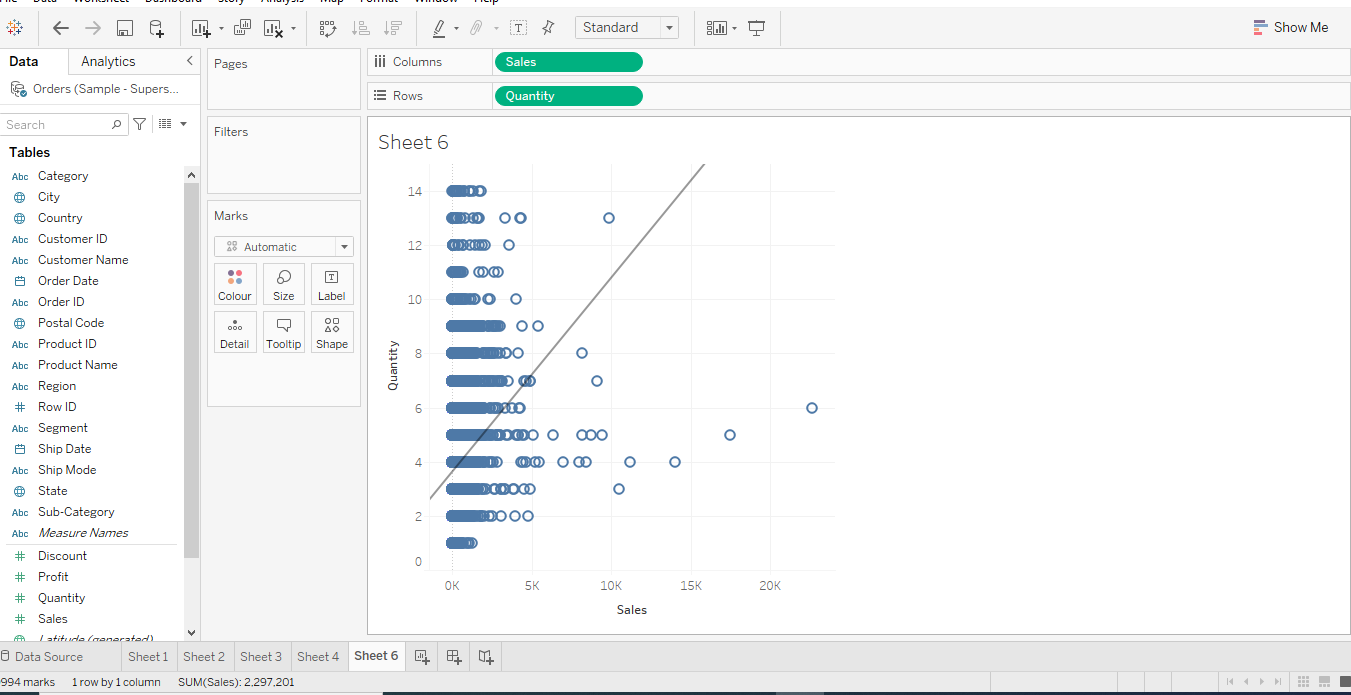
# Top 5 Most Discounted Subcategory

* 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 can see the total discount awarded on each subcategory.   
  
* 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 aligned according to discount shown below,  
    
    
    
    
    
  
* Let’s 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 can know 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 let’s 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, u will see a pop-up and then 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.  
    
  
* From the above graph achieved we can clearly see that most discounted subcategory is binder at around 567K dollars.
* whereas in the top 5 set, paper is the least to around 102.6 k.
* Another insight 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.

# Deciding which Category Performs Well (sale, profit, return, and quantity)

* It is fairly easy to perform 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.   
    
    
    
  
* 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 are the least.
* In case of technology, Sales and profit is very high nearly 836000 and 145000 respectively but this doesn’t justify those 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.

# Correlation Between Quantity and Sales using Scatter Plot

* 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 let’s see the correlation,  
    
  
* We can clearly see that the trend line is showing that yes there is some correlation between the quantity and sales.
* We can infer from it that as quantity is rising by 2 K the sales are 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.

# Create a Dashboard

* 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.  
    
  