

Project on excel: Using Excel formula, charts, pivot tables, creating dashboards

The very first step is to clean the data by checking if there is any duplicate values by using the tools below.

The screenshot shows the Microsoft Excel interface with the 'Remove Duplicates' dialog box open. The dialog box has a title bar 'Remove Duplicates' and a close button. Below the title bar, it says 'To delete duplicate values, select one or more columns that contain duplicates.' There are two buttons: 'Select All' and 'Unselect All'. A checkbox 'My data has headers' is checked. Below this, there is a list of columns: ID, Marital Status, Gender, Income, and Children. All these columns are selected with checkboxes. At the bottom of the dialog box are 'OK' and 'Cancel' buttons. The background spreadsheet shows columns A through M with data for bike buyers. The first row is a header row with columns: ID, Marital Status, Gender, Income, Children, Education, Occupation, Home Owner, Cars, Commute Distance, Region, Age, and Purchased Bike. The data rows follow, with values for each column.

After selecting OK button, we got the results as 26 duplicate values as shown in the below screenshot

The screenshot shows the Microsoft Excel interface after the 'Remove Duplicates' operation. A message box is displayed in the center of the screen, stating: '26 duplicate values found and removed; 1000 unique values remain. Note that counts may include empty cells, spaces, etc.' The message box has an 'OK' button. The background spreadsheet shows the same data as before, but with some rows highlighted in green. The first row is a header row with columns: ID, Marital Status, Gender, Income, Children, Education, Occupation, Home Owner, Cars, Commute Distance, Region, Age, and Purchased Bike. The data rows follow, with values for each column.

While looking at the marital status and the gender, it looks very similar and some don't know the actual meaning of "M" or "S". It is better to use the full abbreviation to make the data look readable and understandable. We will use the Find and Replace function and replace all the necessary data as shown in the figure below.

Excel screenshot showing a data table with columns: ID, Marital Status, Gender, Income, Children, Education, Occupation, Home Owner, Cars, Commute Distance, Region, Age, Purchased Bike. The 'Marital Status' column is highlighted in green. A 'Find and Replace' dialog box is open, showing 'Find what: M' and 'Replace with: Married'. A 'Microsoft Excel' status bar indicates 'All done. We made 539 replacements.' The status bar at the bottom shows 'Count: 1001'.

Similarly performing the same function for “Single”

Excel screenshot showing the same data table. The 'Marital Status' column is now 'Single'. The 'Find and Replace' dialog box is open, showing 'Find what: S' and 'Replace with: Single'. A 'Microsoft Excel' status bar indicates 'All done. We made 463 replacements.' The status bar at the bottom shows 'Count: 1001'.

Similarly performing the same function gender F and M

Excel screenshot showing the same data table. The 'Gender' column is highlighted in green. The 'Find and Replace' dialog box is open, showing 'Find what: M' and 'Replace with: Male'. A 'Microsoft Excel' status bar indicates 'All done. We made 511 replacements.' The status bar at the bottom shows 'Count: 1001'.

Next column is the salary column, if we need to do some calculation, we might need mess with data, hence converting this column to no "\$" and converting from text to currency using the format tool as shown below.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	ID	Marital Status	Gender	Income	Children	Education	Occupation	Home Owner	Cars	Commute Distance	Region	Age	Purchased Bike
2	12496	Married	Female	\$40,000	1	Bachelors	Skilled Manual	Yes	0	0-1 Miles	Europe	42	No
3	24107	Married	Male	\$30,000	3	Partial College	Clerical	Yes	1	0-1 Miles	Europe	43	No
4	14177	Married	Male	\$80,000	5	Partial College	Professional	No	2	2-5 Miles	Europe	60	No
5	24381	Single	Male	\$70,000	0	Bachelors	Professional	Yes	1	5-10 Miles	Pacific	41	Yes
6	25597	Single	Male	\$30,000	0	Bachelors	Clerical	No	0	0-1 Miles	Europe	36	Yes
7	13507	Married	Female	\$10,000	2	Partial College	Manual	Yes	0	1-2 Miles	Europe	50	No
8	27974	Single	Male	\$160,000	2	High School	Management	Yes	4	0-1 Miles	Pacific	33	Yes
9	19364	Married	Male	\$40,000	1	Bachelors	Skilled Manual	Yes	0	0-1 Miles	Europe	43	Yes
10	22155	Married	Male	\$20,000	2	Partial High School	Clerical	Yes	2	5-10 Miles	Pacific	58	No
11	19280	Married	Male	\$120,000	2	Partial College	Manual	Yes	1	0-1 Miles	Europe	40	Yes
12	22173	Married	Female	\$30,000	3	High School	Skilled Manual	No	2	1-2 Miles	Pacific	54	Yes
13	12697	Single	Female	\$90,000	0	Bachelors	Professional	No	4	10+ Miles	Pacific	36	No
14	11434	Married	Male	\$170,000	5	Partial College	Professional	Yes	0	0-1 Miles	Europe	55	No
15	25323	Married	Male	\$40,000	2	Partial College	Clerical	Yes	1	1-2 Miles	Europe	35	Yes
16	23542	Single	Male	\$60,000	1	Partial College	Skilled Manual	No	1	0-1 Miles	Pacific	45	Yes
17	20870	Single	Female	\$10,000	2	High School	Manual	Yes	1	0-1 Miles	Europe	38	Yes
18	23316	Single	Male	\$30,000	3	Partial College	Clerical	No	2	1-2 Miles	Pacific	59	Yes
19	12610	Married	Female	\$30,000	1	Bachelors	Clerical	Yes	0	0-1 Miles	Europe	47	No
20	27183	Single	Male	\$40,000	2	Partial College	Clerical	Yes	1	1-2 Miles	Europe	35	Yes
21	25940	Single	Male	\$20,000	2	Partial High School	Clerical	Yes	2	5-10 Miles	Pacific	55	Yes
22	25598	Married	Female	\$40,000	0	Graduate Degree	Clerical	Yes	0	0-1 Miles	Europe	36	Yes
23	21564	Single	Female	\$80,000	0	Bachelors	Professional	Yes	4	10+ Miles	Pacific	35	No
24	19193	Single	Male	\$40,000	2	Partial College	Clerical	Yes	0	1-2 Miles	Europe	35	Yes
25	26412	Married	Female	\$80,000	5	High School	Management	No	3	5-10 Miles	Europe	56	No
26	27184	Single	Male	\$40,000	2	Partial College	Clerical	No	1	0-1 Miles	Europe	34	No
27	12590	Single	Male	\$30,000	1	Bachelors	Clerical	Yes	0	0-1 Miles	Europe	63	No
28	17844	Single	Male	\$30,000	0	Partial College	Clerical	No	1	0-1 Miles	Europe	38	Yes

Checking all the columns if there is any spelling mistakes, extra spaces and other errors.

If we observe the Age column it looks very difficult to go through each record and to do the visualization hence, we try to group then by giving them an age range. I am going to do this using the Nested IF statements which can help us categorize the age range.

Lets see the below formula **=IF(L2>54,"Old",IF(L2>=31,"Middle Age",IF(L2<31,"Young","Invalid")))**

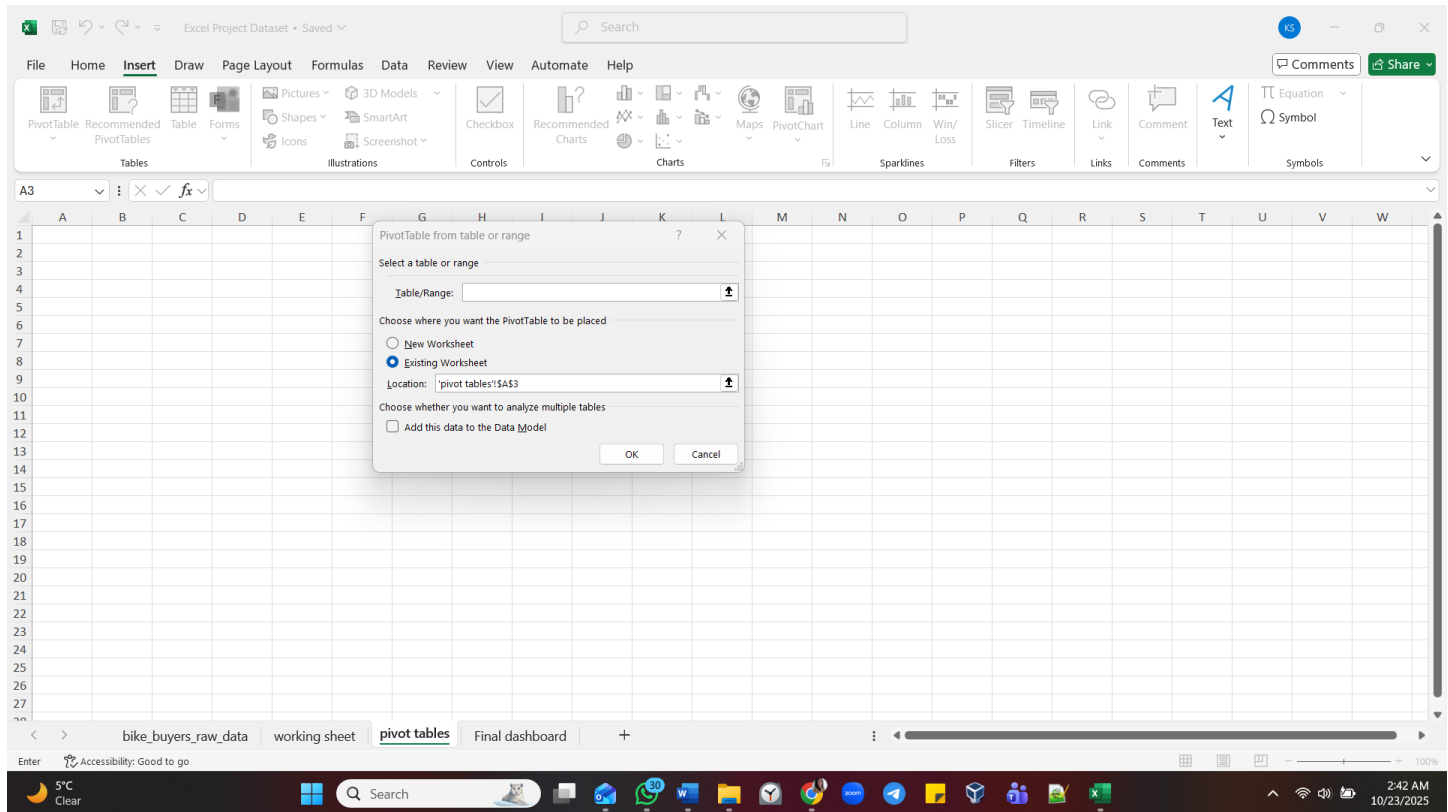
<div> <div> <div>Get Data</div> <div> <div>From Text/CSV</div> <div>From Web</div> <div>From Table/Range</div> <div>From Picture</div> <div>Recent Sources</div> <div>Existing Connections</div> </div> </div> <div> <div>Queries & Transform Data</div> <div>Refresh</div> <div>Properties</div> <div>Workbook Links</div> </div> <div> <div>Queries & Connections</div> <div>Stocks</div> <div>Currencies</div> </div> <div> <div>Sort</div> <div>Filter</div> <div>Advanced</div> </div> <div> <div>Text to Columns</div> <div>Data Tools</div> </div> <div> <div>Forecast</div> <div>What-If Analysis</div> <div>Forecast Sheet</div> </div> <div> <div>Gr</div> <div>Un</div> <div>Su</div> </div> </div>													
<div> <div>M2</div> <div> <div>fx</div> <div>=IF(L2>54,"Old",IF(L2>=31,"Middle Age",IF(L2<31,"Young","Invalid")))</div> </div> </div>													
	B	C	D	E	F	G	H	I	J	K	L	M	
1	Marital Status	Gender	Income	Children	Education	Occupation	Home Owner	Cars	Commute Distance	Region	Age	Age-Range	
2	96	Married	Female	\$40,000	1	Bachelors	Skilled Manual	Yes	0	0-1 Miles	Europe	42	Middle Age
3	07	Married	Male	\$30,000	3	Partial College	Clerical	Yes	1	0-1 Miles	Europe	43	Middle Age
4	77	Married	Male	\$80,000	5	Partial College	Professional	No	2	2-5 Miles	Europe	60	Old
5	81	Single	Male	\$70,000	0	Bachelors	Professional	Yes	1	5-10 Miles	Pacific	41	Middle Age
6	97	Single	Male	\$30,000	0	Bachelors	Clerical	No	0	0-1 Miles	Europe	36	Middle Age
7	07	Married	Female	\$10,000	2	Partial College	Manual	Yes	0	1-2 Miles	Europe	50	Middle Age
8	74	Single	Male	\$160,000	2	High School	Management	Yes	4	0-1 Miles	Pacific	33	Middle Age
9	64	Married	Male	\$40,000	1	Bachelors	Skilled Manual	Yes	0	0-1 Miles	Europe	43	Middle Age
10	55	Married	Male	\$20,000	2	Partial High School	Clerical	Yes	2	5-10 Miles	Pacific	58	Old
11	80	Married	Male	\$120,000	2	Partial College	Manual	Yes	1	0-1 Miles	Europe	40	Middle Age
12	73	Married	Female	\$30,000	3	High School	Skilled Manual	No	2	1-2 Miles	Pacific	54	Middle Age
13	97	Single	Female	\$90,000	0	Bachelors	Professional	No	4	10+ Miles	Pacific	36	Middle Age
14	34	Married	Male	\$170,000	5	Partial College	Professional	Yes	0	0-1 Miles	Europe	55	Old
15	23	Married	Male	\$40,000	2	Partial College	Clerical	Yes	1	1-2 Miles	Europe	35	Middle Age
16	42	Single	Male	\$60,000	1	Partial College	Skilled Manual	No	1	0-1 Miles	Pacific	45	Middle Age
17	70	Single	Female	\$10,000	2	High School	Manual	Yes	1	0-1 Miles	Europe	38	Middle Age
18	16	Single	Male	\$30,000	3	Partial College	Clerical	No	2	1-2 Miles	Pacific	59	Old
19	10	Married	Female	\$30,000	1	Bachelors	Clerical	Yes	0	0-1 Miles	Europe	47	Middle Age
20	83	Single	Male	\$40,000	2	Partial College	Clerical	Yes	1	1-2 Miles	Europe	35	Middle Age
21	40	Single	Male	\$20,000	2	Partial High School	Clerical	Yes	2	5-10 Miles	Pacific	55	Old
22	98	Married	Female	\$40,000	0	Graduate Degree	Clerical	Yes	0	0-1 Miles	Europe	36	Middle Age
23	64	Single	Female	\$80,000	0	Bachelors	Professional	Yes	4	10+ Miles	Pacific	35	Middle Age
24	93	Single	Male	\$40,000	2	Partial College	Clerical	Yes	0	1-2 Miles	Europe	35	Middle Age
25	12	Married	Female	\$80,000	5	High School	Management	No	3	5-10 Miles	Europe	56	Old
26	84	Single	Male	\$40,000	2	Partial College	Clerical	No	1	0-1 Miles	Europe	34	Middle Age
27	90	Single	Male	\$30,000	1	Bachelors	Clerical	Yes	0	0-1 Miles	Europe	63	Old
28	41	Single	Male	\$30,000	0	Partial College	Clerical	No	1	0-1 Miles	Europe	38	Yes

The above formula is a nested if statement which helped me categorize the age group young, middle age and Old. This looks more usable than the individual age group.

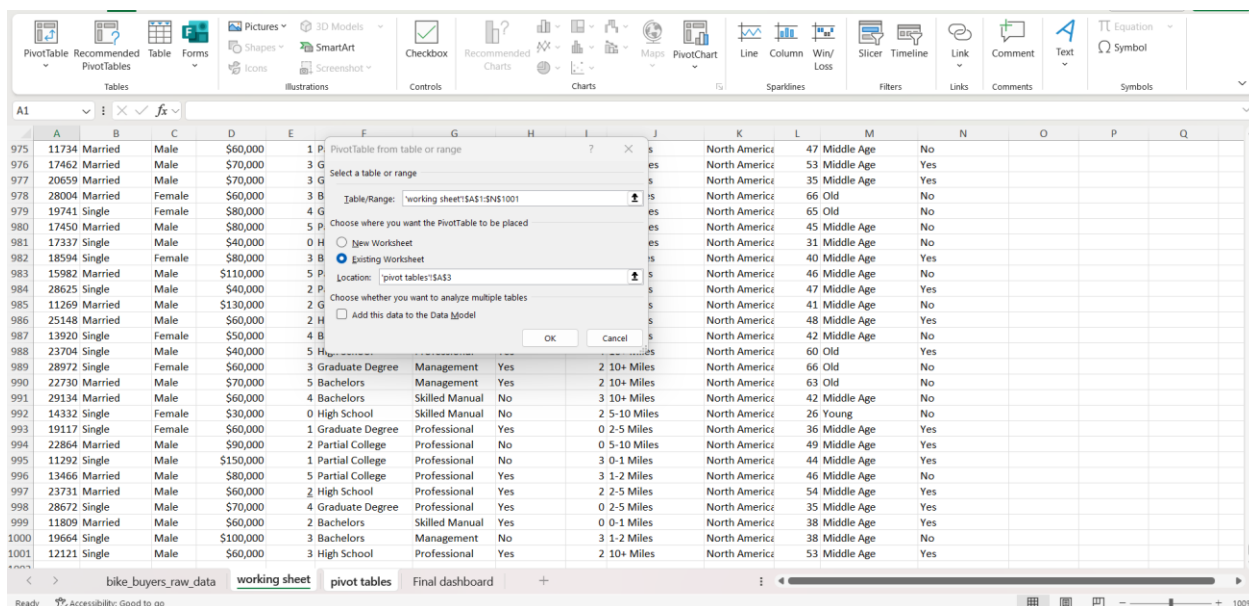
Data looks good and I am going ahead to build the pivot Tables which can help us to build our dashboards.

Now building the pivot table using the steps below:

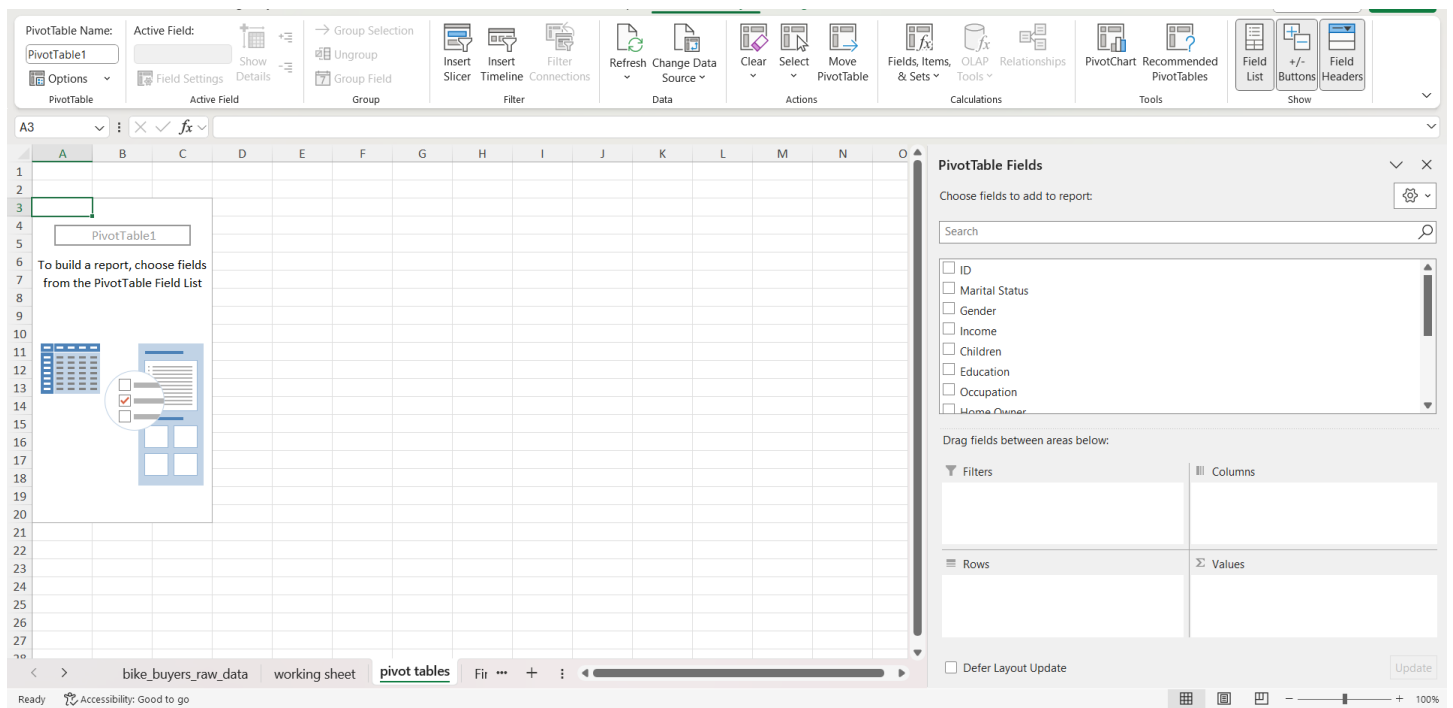
1] clicked on the Pivot Table as shown and a dialogue box will appear as asking us what range and what must be copied and create the table



2] Now copy the cleaned data from my working sheet which I created earlier

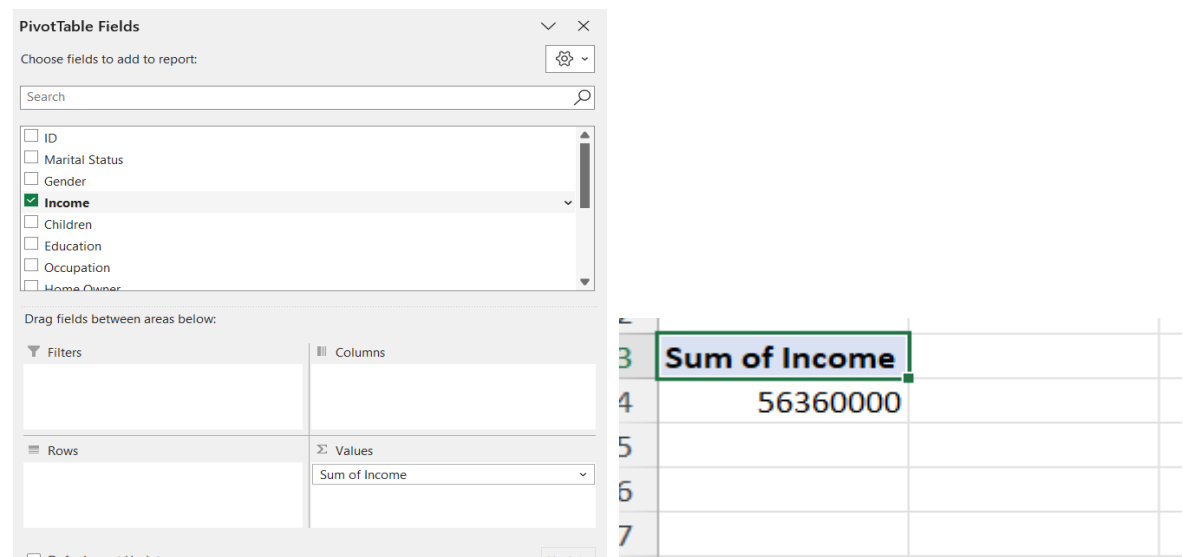


Once I clicked on Ok button it creates the pivot table as shown below, we have all the necessary fields and the drag and drop option to perform our functions.

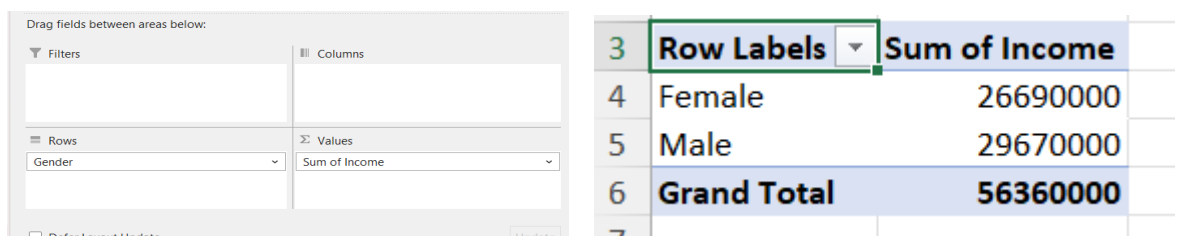


Case 1: Create a dashboard to find the average income of somebody who either bought or did not buy a bike.

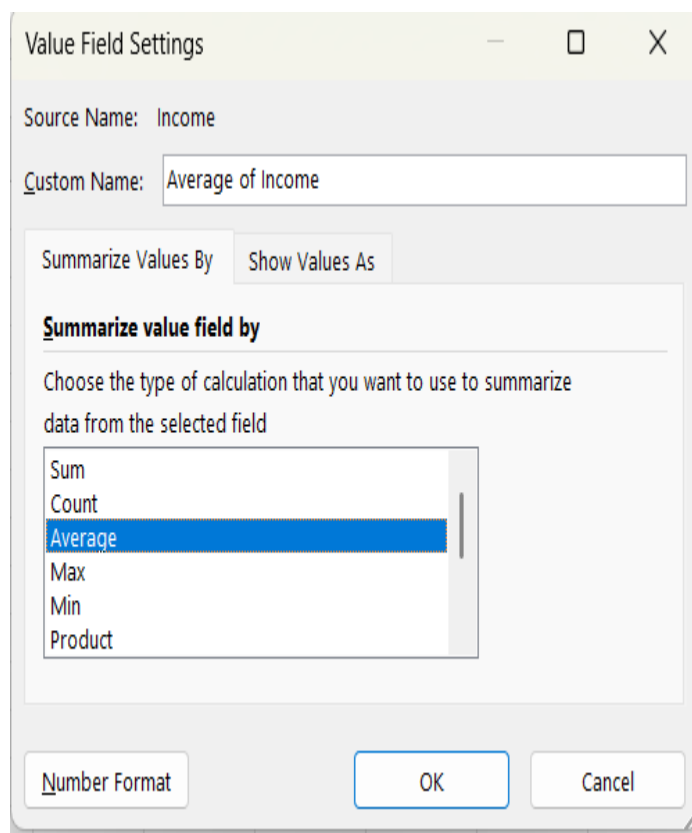
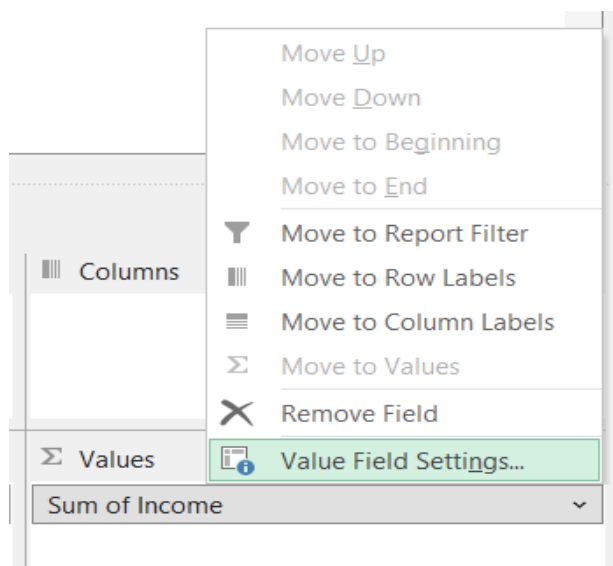
Step 1: to select the Income field and place it in the values tab as shown below



Step2: Break down the income based on the gender, we choose the gender to the rows tabs as shown in the below

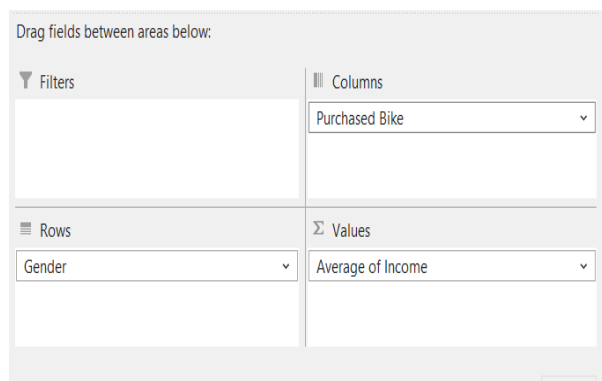


Since we need the avg of the income, we will change the calculation by doing the following



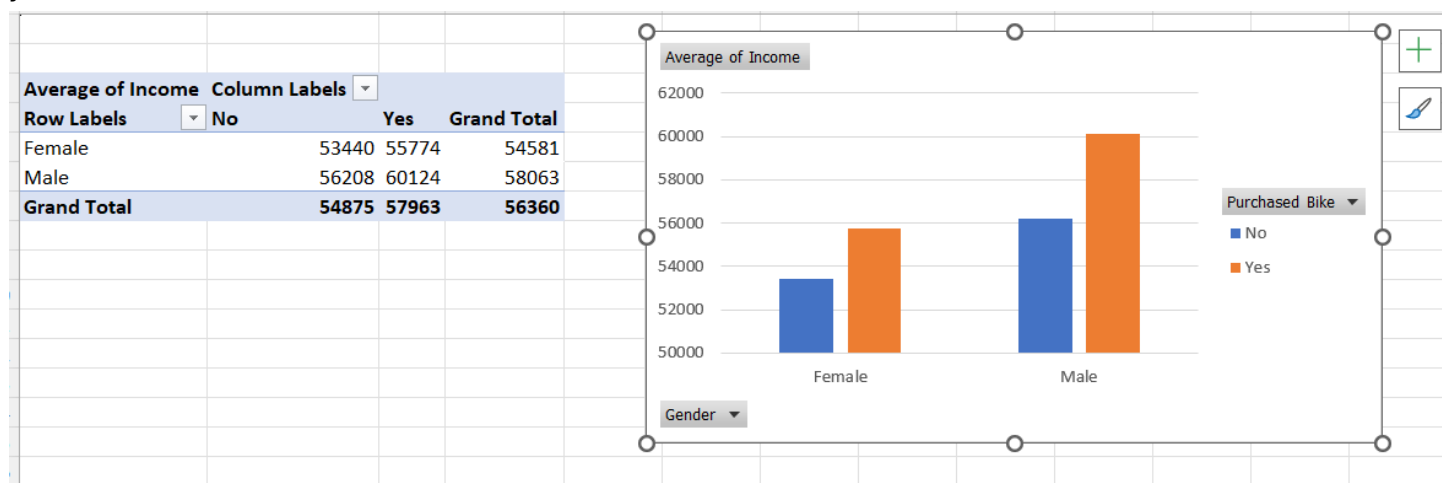
2		
3	Row Labels	Average of Income
4	Female	54581
5	Male	58063
6	Grand Total	56360
7		

Step3: we need to check if they purchased a bike or not so we need to add those values in the column tab as shown in the below data.

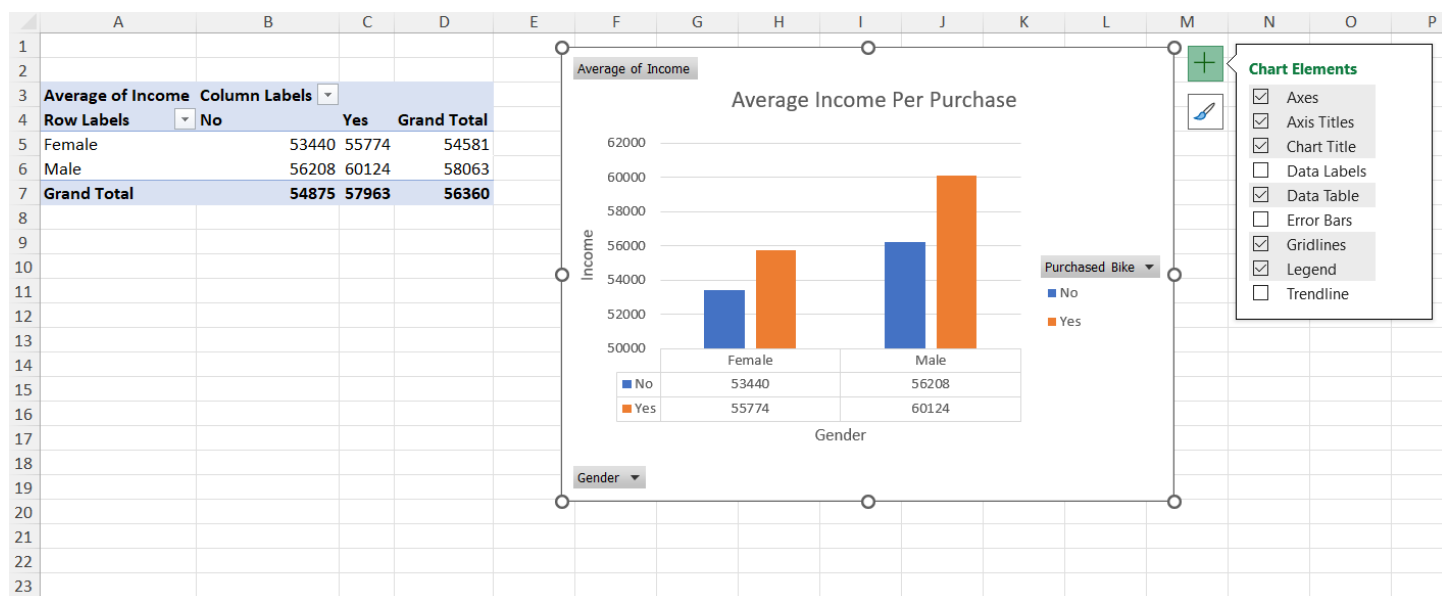


2				
3	Average of Income	Column Labels		
4	Row Labels	No	Yes	Grand Total
5	Female	53440	55774	54581
6	Male	56208	60124	58063
7	Grand Total	54875	57963	56360
8				

Step4: Performing the visualization using the charts as shown below, choose the best chart describes your data. I have chosen this as shown below:

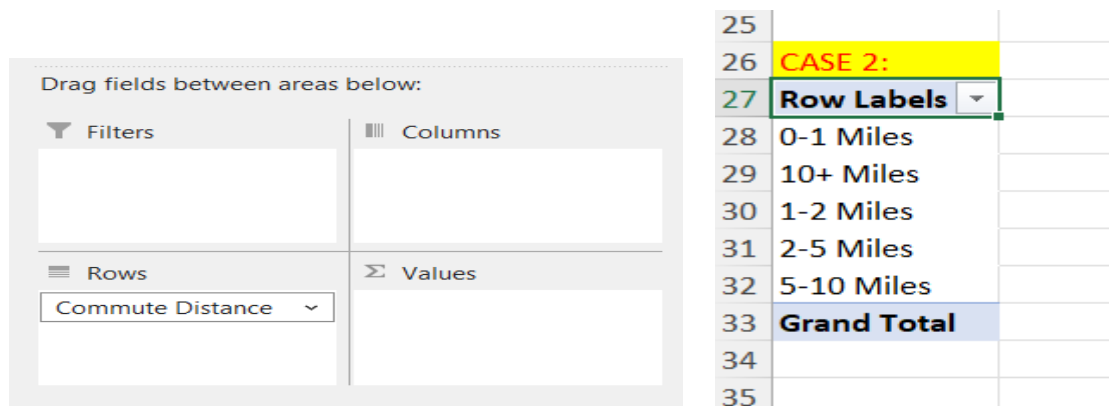


Now if you want to add any additional details to the chart, we can use the filter option that is available and give the chart title, Axis details and other features. I have added chart name, axis names and added the data table as shown below



Case 2: Lets build a dashboard using the commute distance and see how many miles do they live away from work. This can help us visualize who is buying the bike.

step1: Drag the commute distance field to the rows column as shown below



Step 2: Checking if they have bought a bike by moving the purchased bike field to columns and the value tab as shown below

Drag fields between areas below:

Filters	Columns
	Purchased Bike

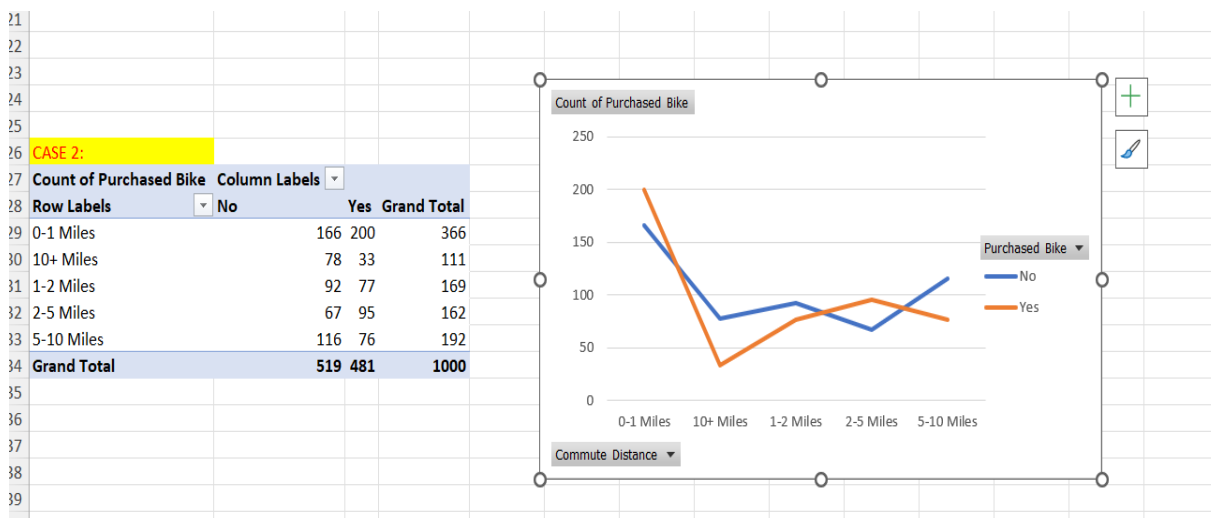
Rows	Values
Commute Distance	Count of Purchased ...

☐ Defer Layout Update

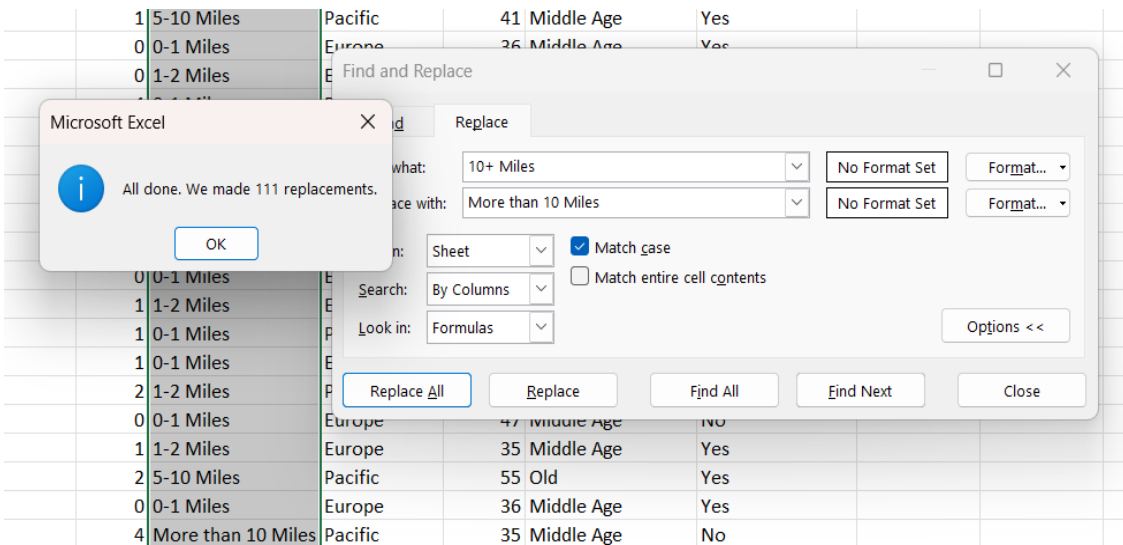
26	CASE 2:			
27	Count of Purchased Bike	Column Labels		
28	Row Labels	No	Yes	Grand Total
29	0-1 Miles	166	200	366
30	10+ Miles	78	33	111
31	1-2 Miles	92	77	169
32	2-5 Miles	67	95	162
33	5-10 Miles	116	76	192
34	Grand Total	519	481	1000
35				

We also have the count of the purchased a bike or not as shown above.

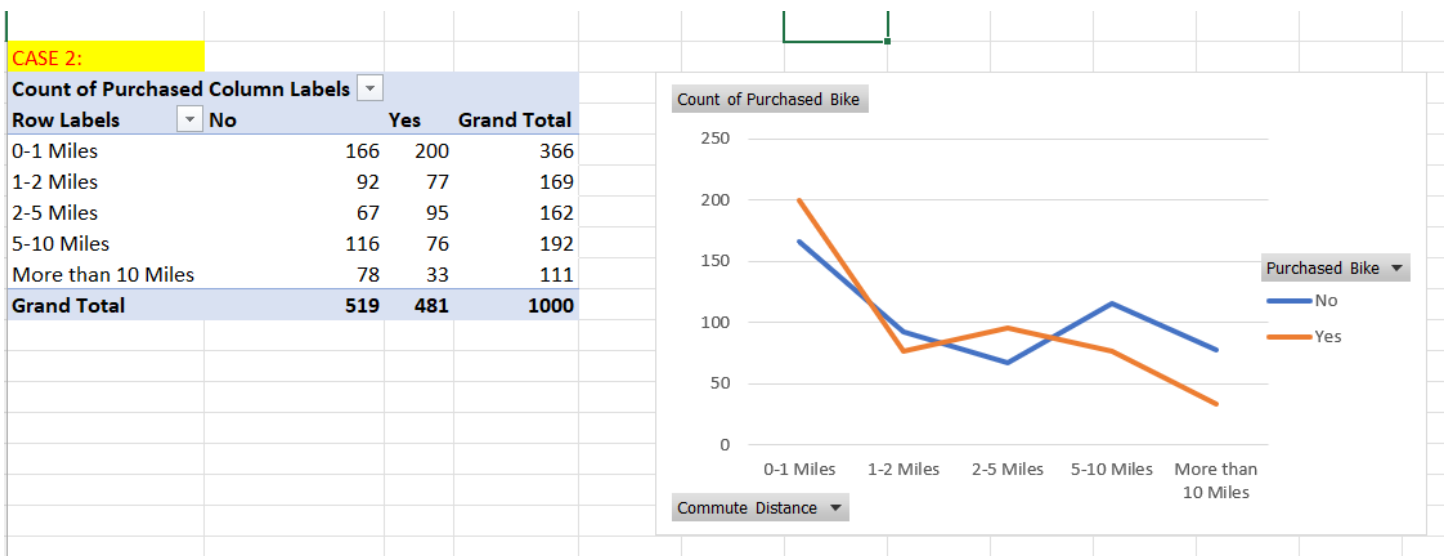
I found an issue while I was observing the row Lales, after 0-1 miles then comes the 10+miles which is not the correct way, and we need to change that because while we are creating the Charts it will cause the issue. Let's just show how it looks for a better understanding. If we see the below line chart we can see that the C- axis has 0-1 miles an dthen 10+miles followed by 1-2 miles. This is not a good approach



Let's try to fix this issue by finding and replace tool and change the 10+miles such that it comes to end of the X-axis. As shown below



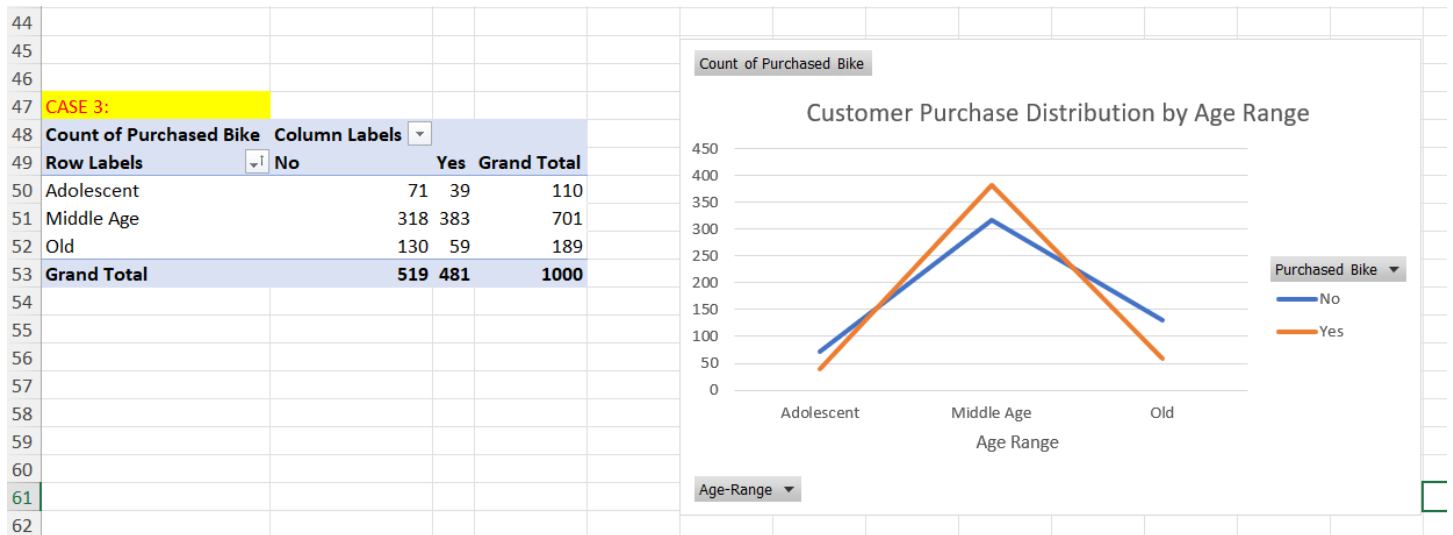
Now, let's try to refresh the charts and how they look. This looks a clean line chart now



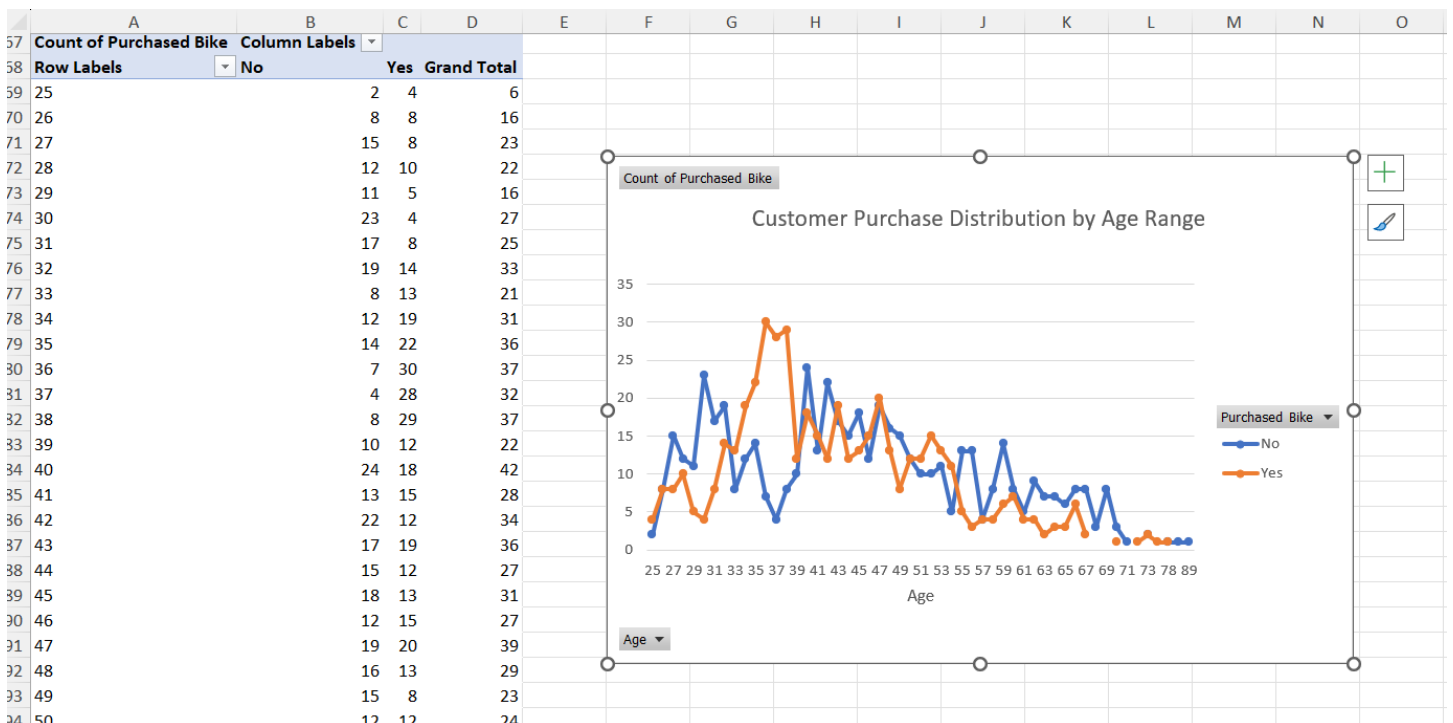
Added some filters like chart titles and other required fields and the final chart and the pivot table are as below



CASE 3: Checking on the age range if they bought the bike or not. Making the age range is convenient to analyze the data for the future dashboard making.



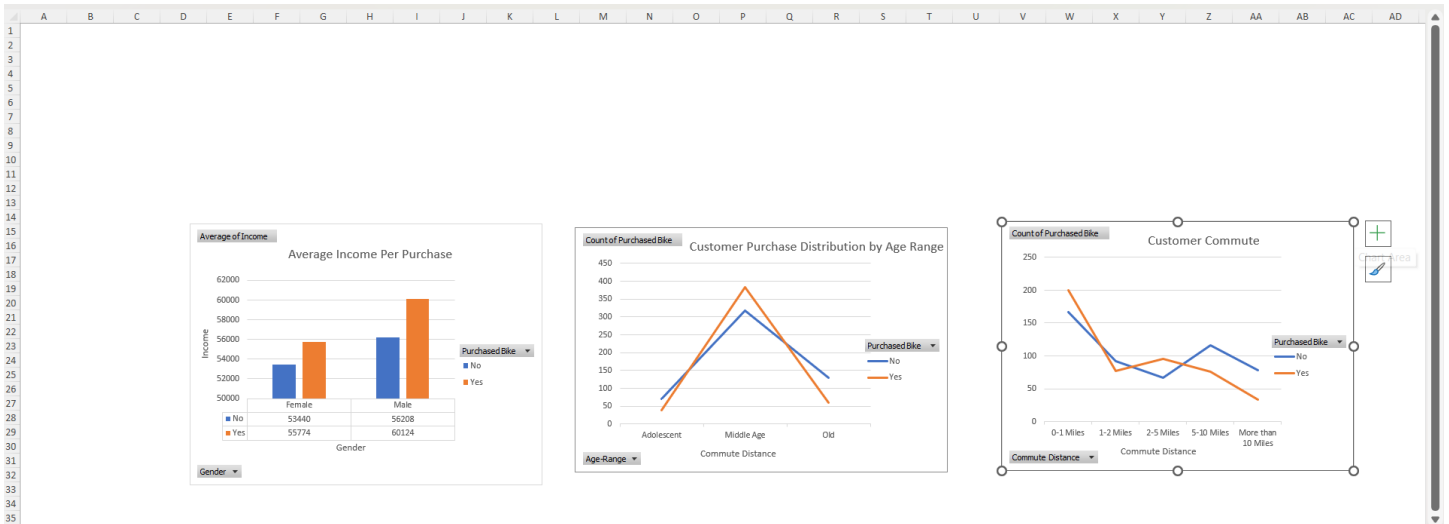
Suppose the age group division was not divided then the data would look something like the below chart and it is very difficult for anyone to understand the purchase details vs their ages.



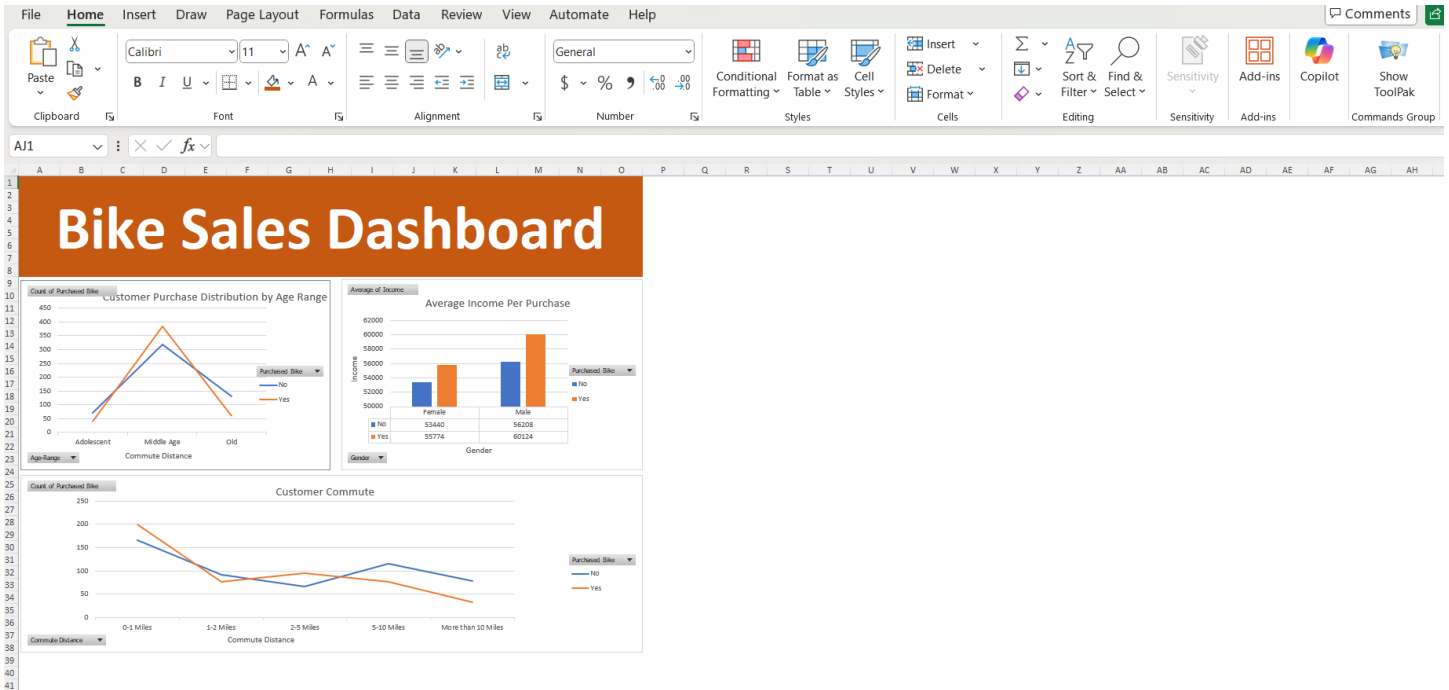
In this case, we would just go ahead with the case 3 First chart to build our dashboard in a simple.

Building of the dashboard.

Step1: To build the dashboard you need to get rid of the grid lines of the excel to make it look visually beautiful.



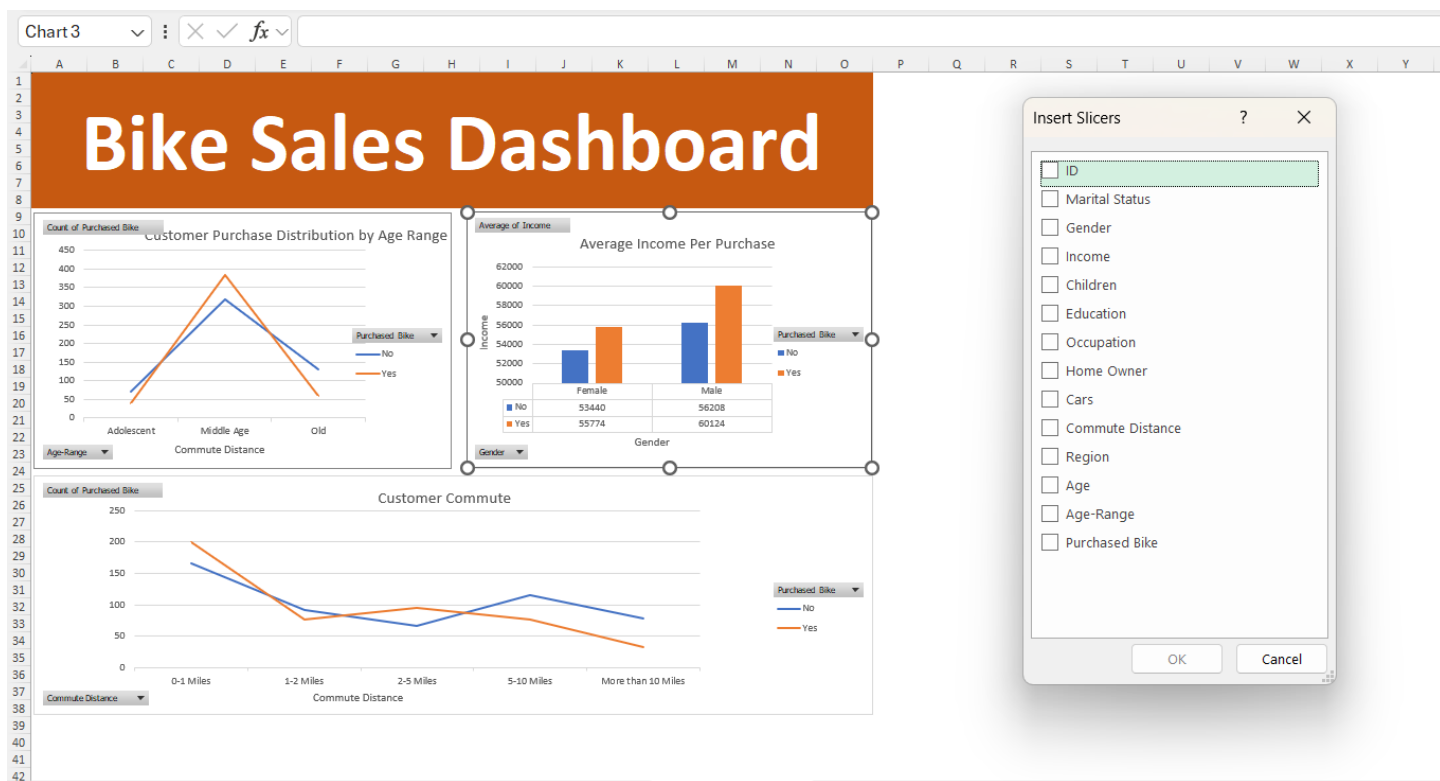
Step2: Let's try to create a Header for this dashboard. This is a simple visualization dashboard which is created.



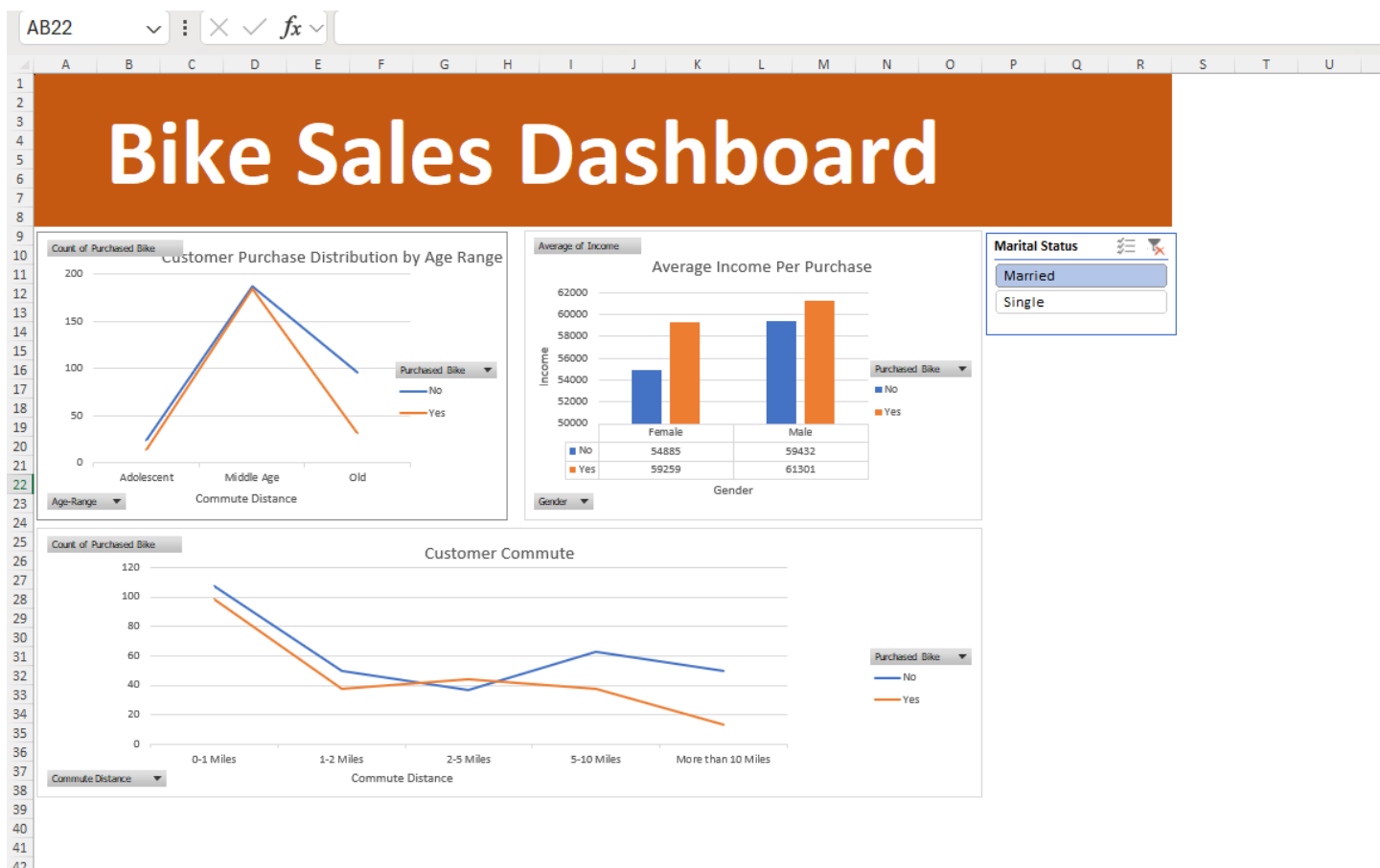
To make it more dynamic, we will add some filters to the dashboard. If anyone wants to analysis people who are married and bought the bike or are still single. We will build a filter such that it is easy to analyze

Adding a slicer by clicking on one of the charts in which you want to add the slicers. The following steps are followed.

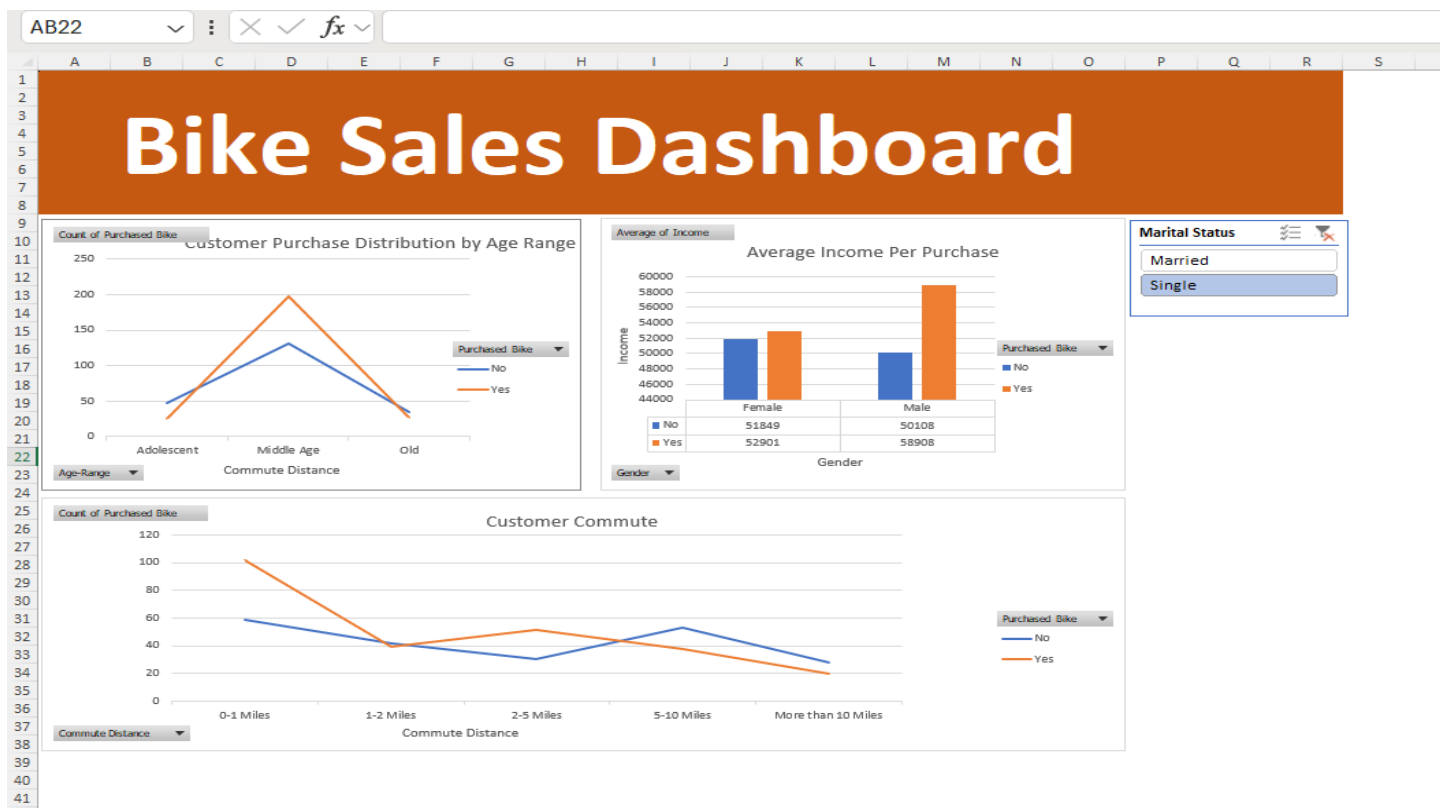
1] select the chart and click on the PivotChart Analyzer



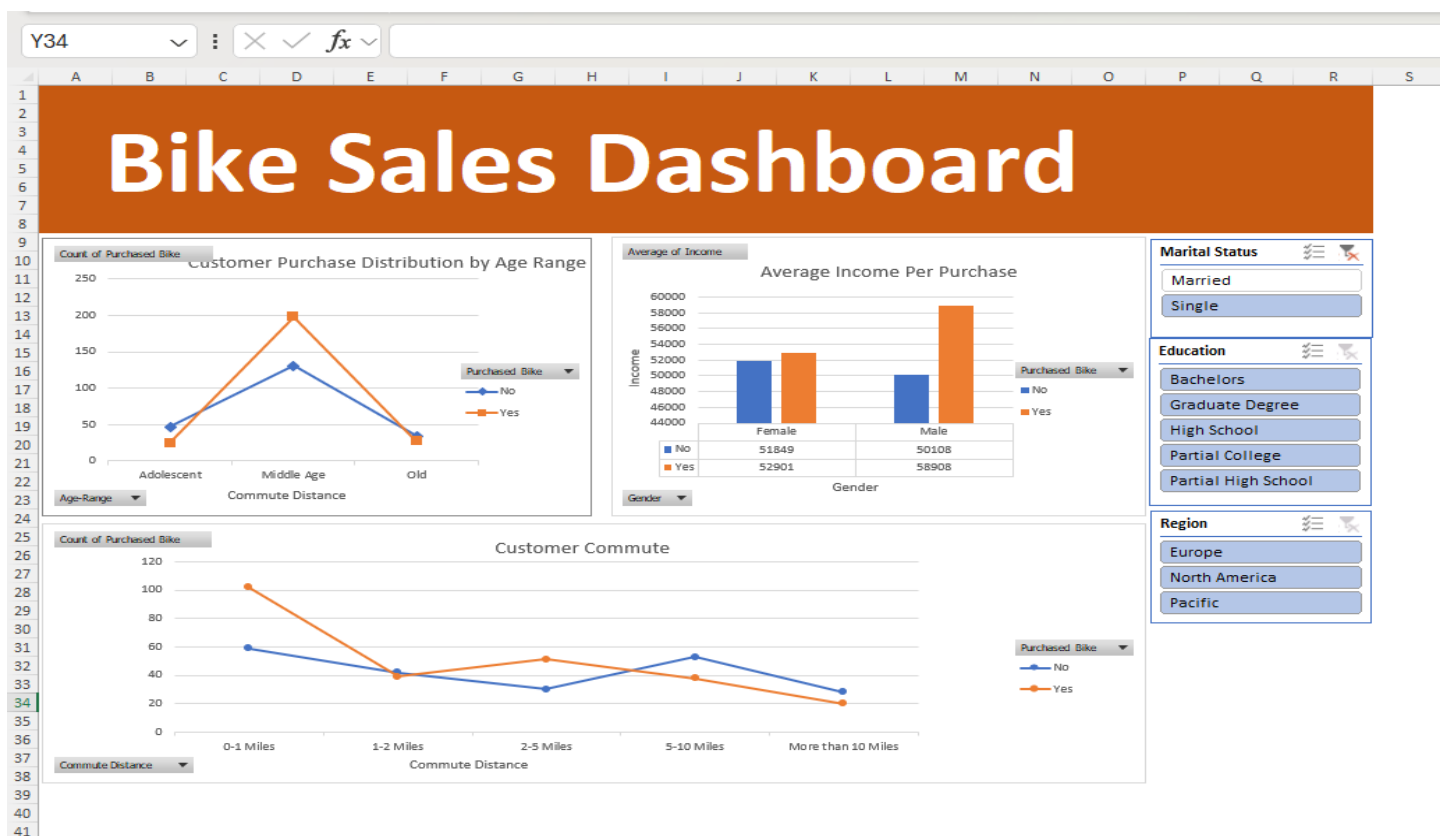
Select the slicers, here we are going to select Marital status. I selected the Marital Status as Married, hence we got different results.



Let's try marital status for single. We get the below results



Added some more slicers such as education and the region. This is the final Dashboard which I created.



In the screenshot above I have selected Marital status as Single and Education as Bachelors. The output defines how many people purchased a bike have completed Bachelors and are single.