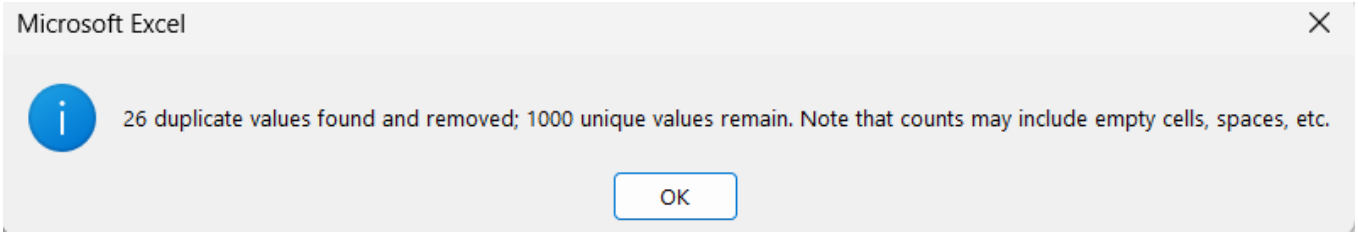
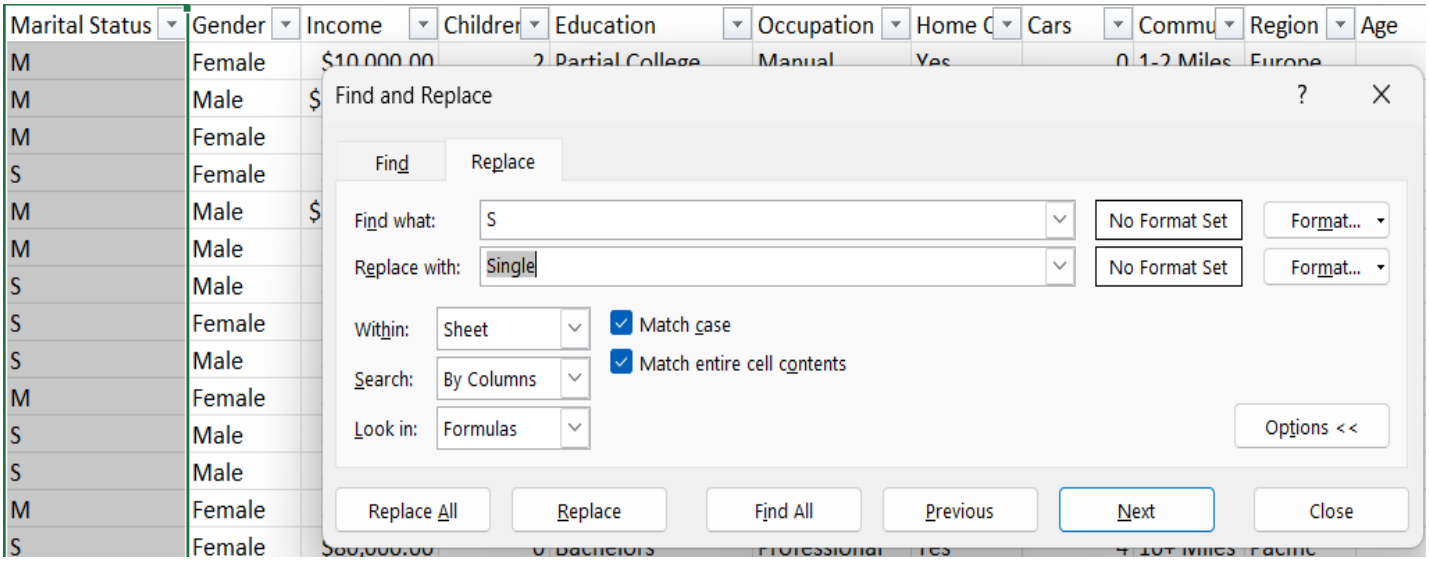
1 Contents

2 Data Cleaning and Preparation

First of all, let's remove the duplicates from the "ID" column, since they should be unique.



In the "Marital Status" and "Gender" columns, both of them have the letter "M." To avoid confusion, we will write the full forms in both columns.



To prevent any changes to the column names, we should enable the "Match case" and "Match entire cell contents" options.

In order to enhance readability and visualization, we need to eliminate specific zeros within the "Income" columns.

D
Income
$10,000.000
$120,000.000
$30,000.000
$90,000.000
$170,000.000
$40,000.000
$60,000.000
$10,000.000
$30,000.000
$30,000.000
$40,000.000

D
Income
$10,000
$120,000
$30,000
$90,000
$170,000
$40,000
$60,000
$10,000
$30,000
$30,000
$40,000
$20,000
$40.000

To facilitate interpretation from visualization due to the presence of numerous age values, a new column called "Age Brackets" has been created.

L	M
Age	Age Brackets
50	Middle Age
40	Middle Age
54	Middle Age
36	Middle Age
55	Old
35	Middle Age
45	Middle Age
38	Middle Age
59	Old
47	Middle Age
35	Middle Age
55	Old
36	Middle Age
35	Middle Age
35	Middle Age
56	Old
34	Middle Age
63	Old
29	Young

The "Age Brackets" column has been generated using the following function to categorize ages:

```
=IF(L2>54,"Old",IF(L2>=31,"Middle Age",IF(L2<31,"Young","Invalid")))
```

The code snippet is an Excel formula using the IF function to categorize ages into different brackets based on the value in cell L2. Let's break it down:

- IF(L2>54, "Old", ...): This is the first condition. If the value in cell L2 is greater than 54, it will be labeled as "Old".
- IF(L2>=31, "Middle Age", ...): If the first condition is not met, this condition checks if the value in cell L2 is greater than or equal to 31. If it is, the age will be labeled as "Middle Age".
- IF(L2<31, "Young", "Invalid"): If none of the previous conditions are met, this condition is evaluated. If the value in cell L2 is less than 31, it will be labeled as "Young". If it doesn't satisfy this condition, it is labeled as "Invalid".

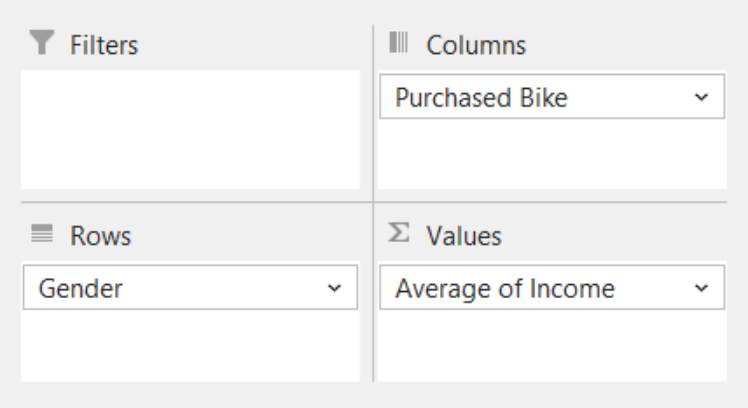
To summarize:

- Ages greater than 54 are categorized as "Old".
- Ages between 31 and 54 (inclusive) are categorized as "Middle Age".
- Ages less than 31 are categorized as "Young".
- Any other values or errors will be labeled as "Invalid".

Please note that this formula assumes the age value is in cell L2 and it has been applied to other cells accordingly.

3 Creating Pivot Tables and Data Analysis

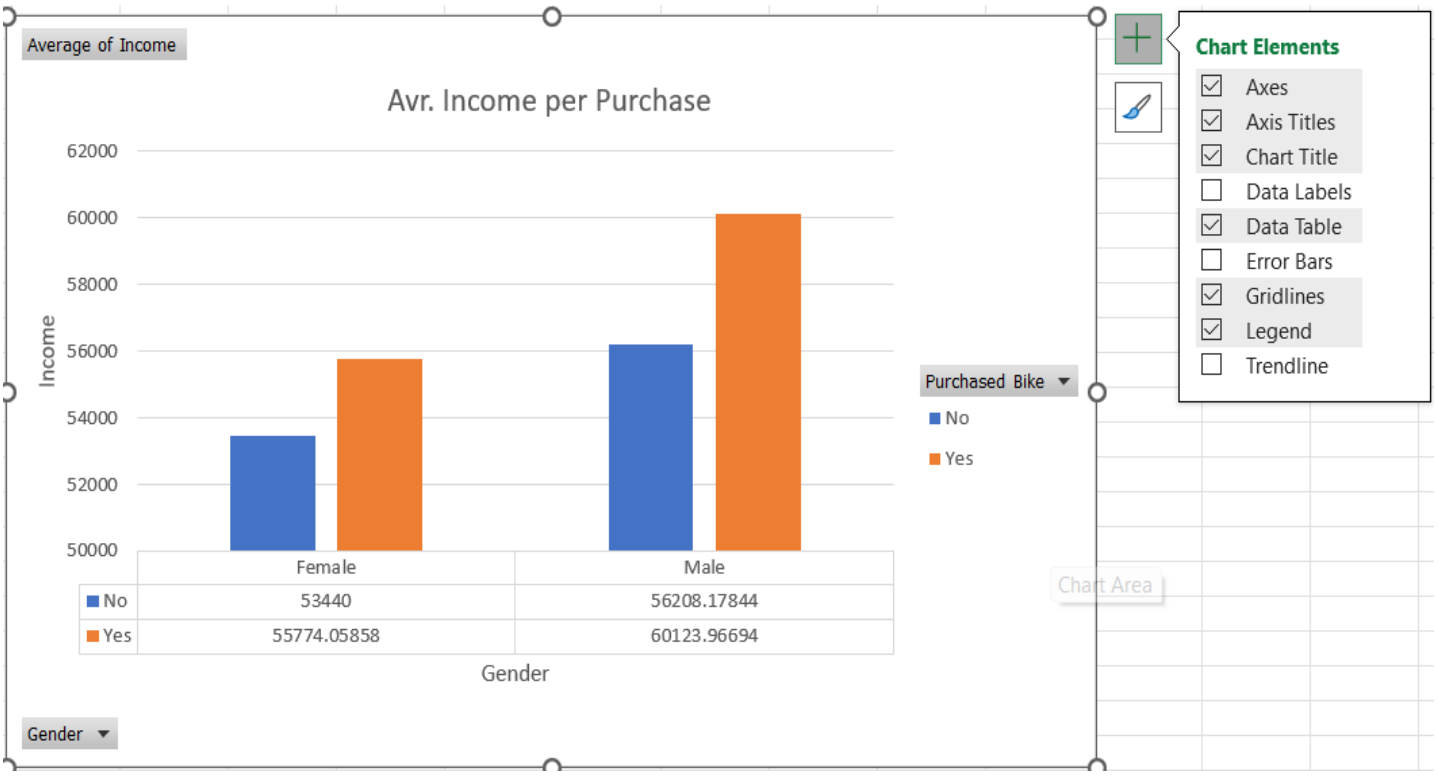
3.1 Average Annual Salary by Bike Purchase and Gender



In the Pivot table, I placed the "Gender" column in the rows section, the "Average of Income" column in the values section, and the "Purchased Bike" column in the columns section.

Average of Income	Column Labels		
Row Labels	**No**	**Yes**	**Grand Total**
Female	53440	55774.05858	54580.7771
Male	56208.17844	60123.96694	58062.62231
Grand Total	**54874.75915**	**57962.57796**	**56360**

The provided Pivot table displays the average yearly salary categorized by gender and whether individuals purchased bicycles. The data reveals that men earn $4,500 more on average than women. Among women, those who purchased bicycles earn roughly $2,300 more compared to those who did not. Similarly, among men, individuals who purchased bicycles earn approximately $4,000 more than those who did not. Now let's proceed to visualize this information on a graph.

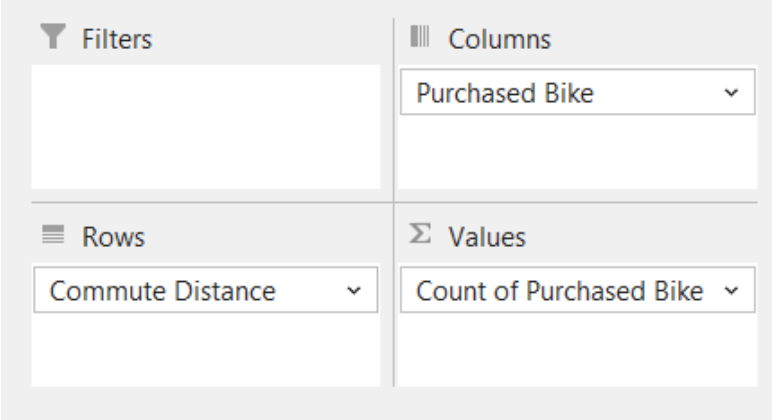


Here, we selected the recommended graph type, which is the clustered column. Additionally, we added a data table. However, the commas in the data table are visually distracting, so now we will remove these excessive decimal places.

Average of Income	Column Labels		
Row Labels	No	Yes	Grand Total
Female	53,440	55,774	54,581
Male	56,208	60,124	58,063
Grand Total	**54,875**	**57,963**	**56,360**

In the Pivot table, we selected the numeric values from which we want to remove commas, and then we chose "Number" from the "Home", decrease the decimals and clicked to the thousands seperator.

3.2 Bike Purchase by Commute Distance



In the Pivot table, the "Commute Distance" column was placed in the rows section, the "Count of Purchased Bike" column in the values section, and the "Purchased Bike" column in the columns section.

Count of Purchased Bike	Column Labels		
Row Labels	No	Yes	Grand Total
0-1 Miles	166	200	366
10+ Miles	78	33	111
1-2 Miles	92	77	169
2-5 Miles	67	95	162
5-10 Miles	116	76	192
Grand Total	**519**	**481**	**1000**

In this table, I need to edit the "10+ Miles" section so that the numerical values can be properly sorted.

Count of Purchased	Column Labels		
Row Labels	No	Yes	Grand Total
0-1 Miles	166	200	366
1-2 Miles	92	77	169
2-5 Miles	67	95	162
5-10 Miles	116	76	192
More than 10 Miles	78	33	111
Grand Total	**519**	**481**	**1000**

We changed "10+ Miles" to "More than 10 Miles," and as seen in the Pivot table, it is now sorted properly from smallest to largest.

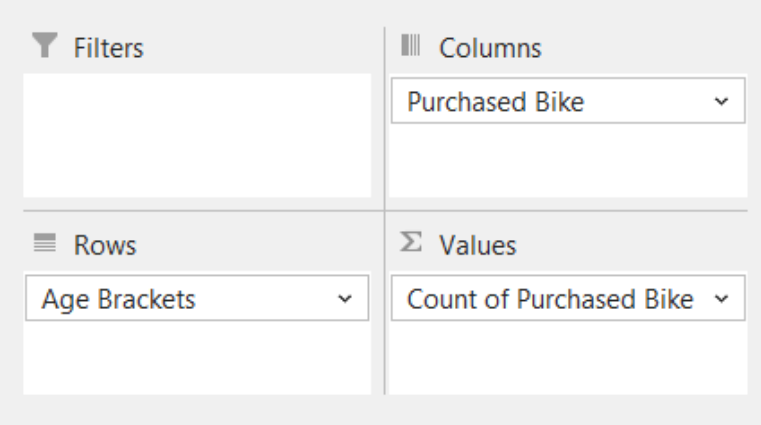
In these Pivot table and graph, we observe that individuals whose commute distance falls between 0 and 1 mile are more likely to prefer purchasing a bicycle compared to other distance intervals. More than half of the individuals within this distance range have chosen to purchase a bicycle. On the other hand, over half of the individuals whose commute distance ranges from 1 to 2 miles have chosen not to purchase a bicycle. In contrast, more than half of the individuals with a commute distance between 2 and 5 miles have preferred purchasing a bicycle. For individuals whose commute distance ranges from 5 to 10 miles, slightly over half have chosen not to purchase a bicycle. It is evident that the majority of individuals with a commute distance exceeding 10 miles prefer not to purchase a bicycle.

Count of Purchased Bike	Column Labels		
Row Labels	**No**	**Yes**	**Grand Total**
0-1 Miles	45.36%	54.64%	100.00%
1-2 Miles	54.44%	45.56%	100.00%
2-5 Miles	41.36%	58.64%	100.00%
5-10 Miles	60.42%	39.58%	100.00%
More than 10 Miles	70.27%	29.73%	100.00%
Grand Total	**51.90%**	**48.10%**	**100.00%**

To address the imbalance in data counts within the commute distance groups in this pivot table, I rearranged the rows using the "% of row totals" format as you can see above.

By using this approach, we can obtain a more reliable graph. Additionally, as observed from the graph, we can see that apart from the 2-5 mile range, as **commute distance** increases, the trend of **not purchasing** bicycles also increases.

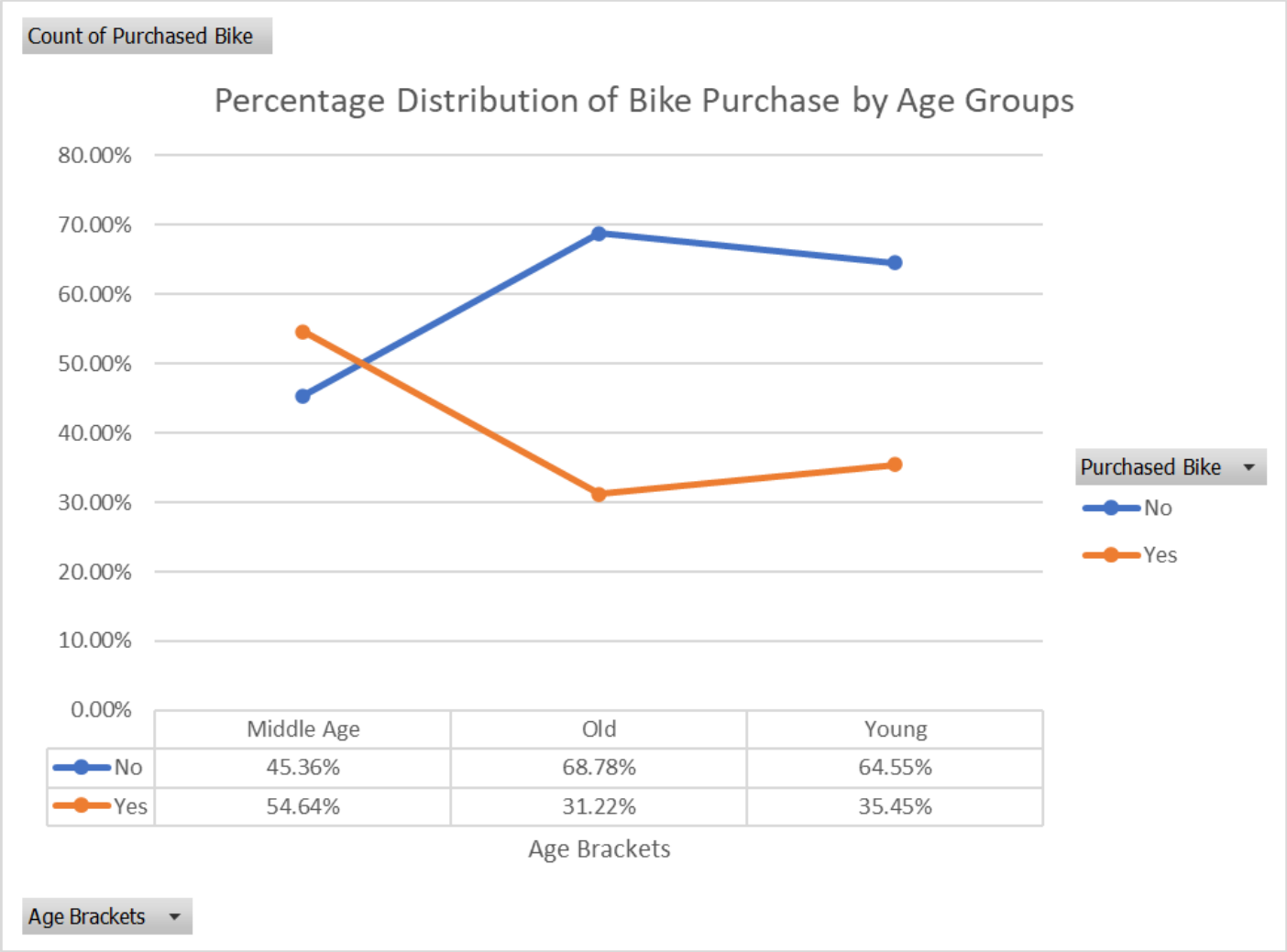
3.3 Bike Purchase by Age Groups



I have performed the necessary steps to create the Pivot table as described above, similar to the previous instructions.

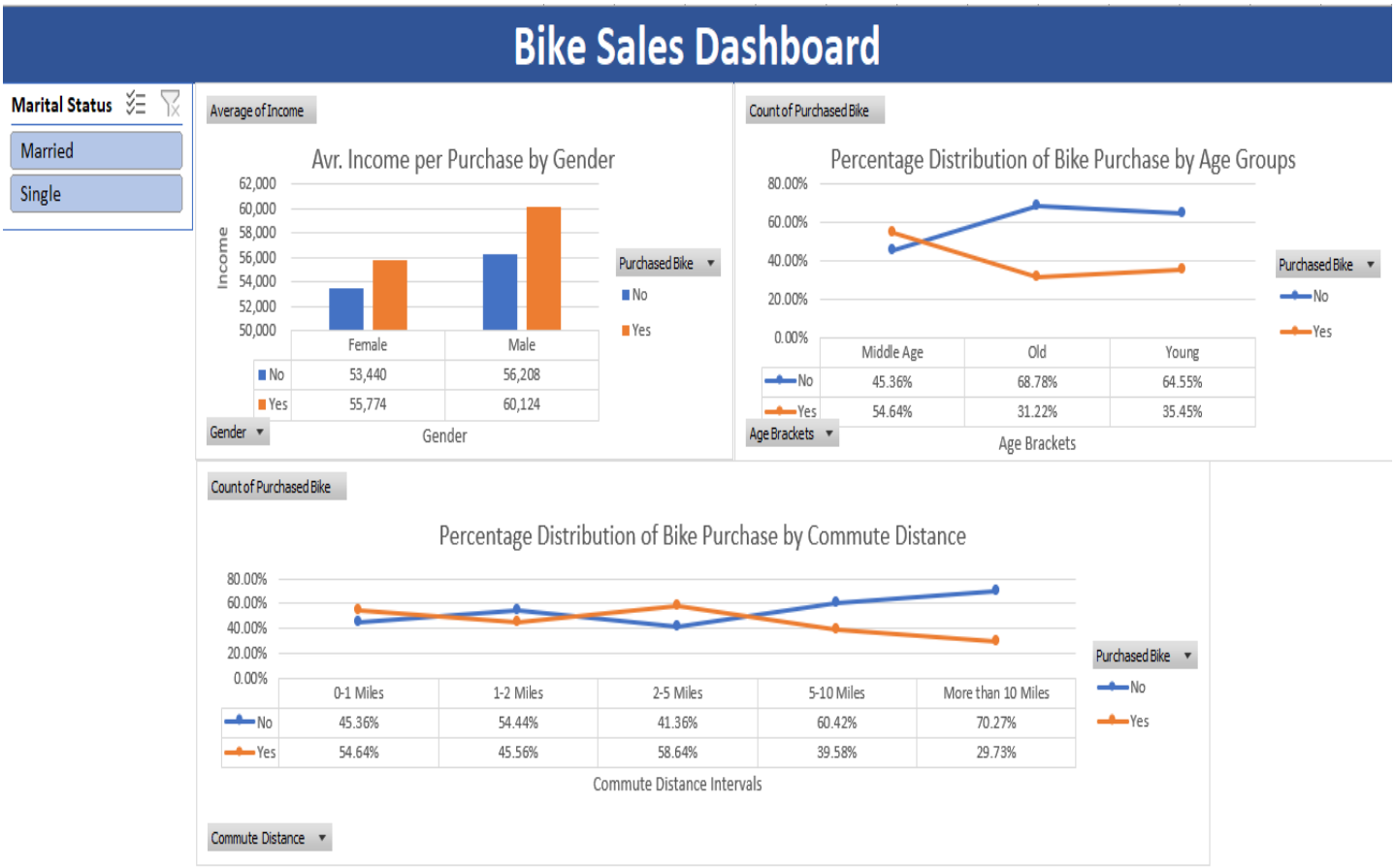
Count of Purchased Bike	Column Labels		
Row Labels	No	Yes	Grand Total
Middle Age	318	383	701
Old	130	59	189
Young	71	39	110
Grand Total	**519**	**481**	**1000**

In this Pivot table, we observe that there is a significantly higher number of individuals in the middle-aged group compared to the younger and older groups. Similarly to before, let's convert the rows into percentages to better observe the purchasing bike preference of each age group and visualize it in a graph.



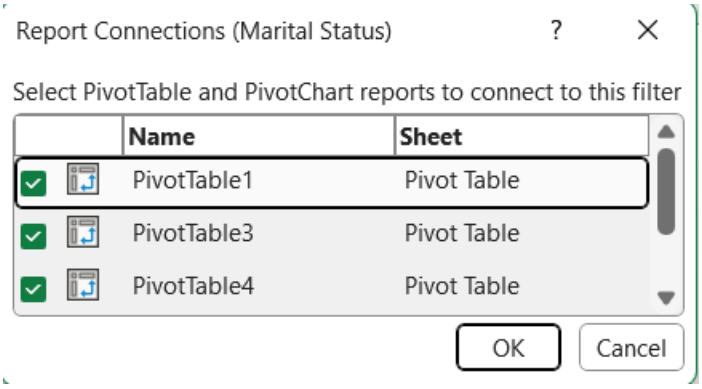
Based on the data, we can see that the majority of young and elderly individuals do not prefer purchasing bicycles. In the middle-aged group, however, there is a slightly higher number of individuals who choose to purchase bicycles. However, it should be noted that the validity of this analysis is debatable due to the relatively small number of individuals in the young and elderly groups.

4 Dashboard Creation

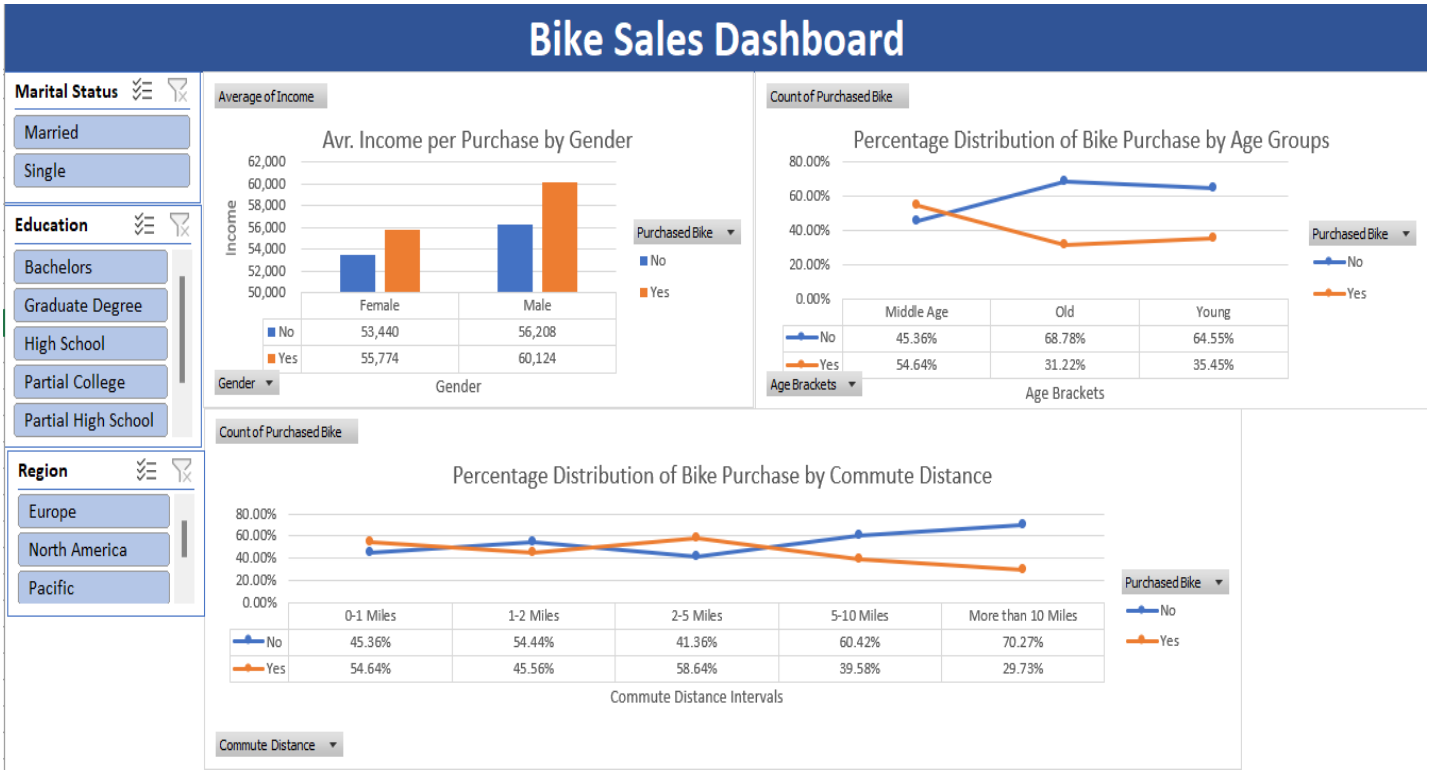


I have placed the necessary graphs in the dashboard, as shown in the figure, and added a slicer indicating whether the individual is married.

I selected "Report Connections" in the slicer section and chose all the Pivot tables so that the slicer in the figure interacts with all the graphs.



In addition, I added Education and Region slicers as you can see image below, ensuring they are also connected to all the Pivot tables for interactive functionality.



This dashboard provides the individual with controlled, efficient, and interactive data analysis capabilities. The filtering tools, namely slicers, allow for quick and interactive results according to the user's preferences.