Bike Sales Project

In this project, I will be using a dataset I discovered on the internet (Kaggle) to garner basic insights on the sales of bikes for a company.

I first downloaded and loaded the dataset on excel. The dataset consisted of the following columns:

- ID
- Marital Status
- Gender
- Income
- Children
- Education
- Occupation
- Home Owner
- Number of Cars
- Commute Distance
- Region
- Age
- Purchased Bike

As good practice, I will duplicate the dataset so I am not working with the raw dataset. This way if any errors are encountered or I lose track, I can refer back to the original dataset. The next thing I proceeded to do on the Working sheet, with the copied dataset, was to remove any duplicate data. This was done using the 'Remove Duplicate' function in excel. After carrying this out, I successfully removed 26 duplicates from the working set.

The next step would be to work on the Marital Status column. On the original dataset, "Married' and 'Single' was represented with 'M' and 'S' respectively. This, however, can be an issue due to readability for someone who may end up using our dashboard. So, to fix this, I used the find and replace function of excel to find any instance of 'M' on the column and replace that with 'Married'. I did the same thing with any instance of 'S' on the column and replaced that with 'Single'.

This issue also applies with the Gender Column where 'M' represents Male and 'F' represents Female. So, we use the Find and Replace function to replace the letters with the relevant Gender.

Looking at the income column, we can see that it needs to be set in the relevant format, so I switched the data format to currency.

Looking at the age column, we can see that a large variety of ages are held in the dataset, so making visuals may be more difficult. So, what I did was create a new column called ' Age Brackets'.

On this column we would write a query that consists of nested if statements as shown below:

```
=IF(M2>54,"Old",IF(M2>=31,"Middle Aged",IF(M2<31,"Adolescent","Invalid")))
```

This query/formula ensures that those who are less than 31 are adolescent, older than 31 are Middle aged and above 54 are old. This way we have grouped the ages into 3 bins that will allow us to make some useful visuals later.

After this, I created a new worksheet in which I would create PivotTables. These PivotTables will be used to help develop the visuals for the final dashboard.

Using the Pivot Table Function, I referenced the table on the working sheet that was created prior. With this, the columns of the table become the Pivot Table field which will allow us to construct views of the data in a tabular format, to visualise certain insights.

The first insight/visual of the data I wanted to create was the Average income of people who purchased a bike and people who didn't buy a bike. This data should also be split by gender for further analysis.

To do this, I set the value to be the Average aggregate of the income of all individuals. I set the rows to be equivalent to each gender. Finally, I set the column to be whether they had purchased a bike or not. As a result, we have the following table:

Average of Income	Column Labels		
Row Labels	No	Yes	Grand Total
Female	$53,440.00	$55,774.06	$54,580.78
Male	$56,208.18	$60,123.97	$58,062.62
Grand Total	$54,874.76	$57,962.58	$56,360.00

With this Pivot Table, we can create the first visual as a Bar chart and a simple table underneath that holds the data values for better readability:
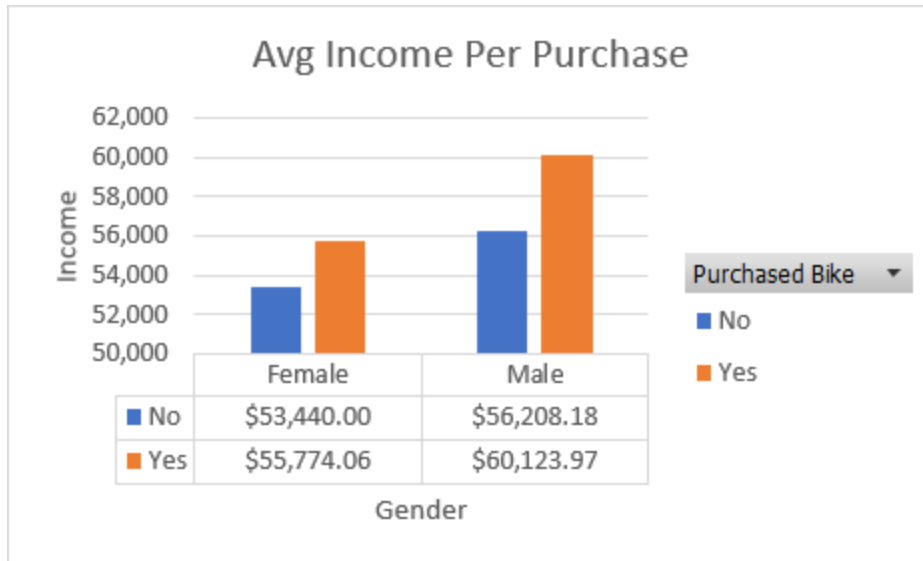
The next pivot table will be created from the same working sheet table; however, we will be doing an analysis on the customer commute and whether they had purchased a bike or not. To do this, the Commute distance ranges were set to be the rows the values was set to be the Count of the Purchased Bikes. Finally, the columns where set to whether they had purchased a bike or did not.

With this, the Pivot table has been created:

| Count of Purchased Bike	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

With this Pivot Table, I created the next visual which was a line graph, with the commute distance as the X- Axis, and the number of purchased bikes being the Y-axis. 2 lines are created: one for those who have purchased a bike, and then next for those who did not.

Customer Commute

Using the last pivot table, I wanted to gain an insight into the purchasing behaviours of people in the varying age brackets. To do this, I set the Pivot table to consist of the age brackets as the rows, the number of purchased bikes as the values, and whether a bike had been purchased or not as the columns. This produced the following Pivot Table:

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

With this Pivot Table, we can create our final visual:

Customer Age Brackets
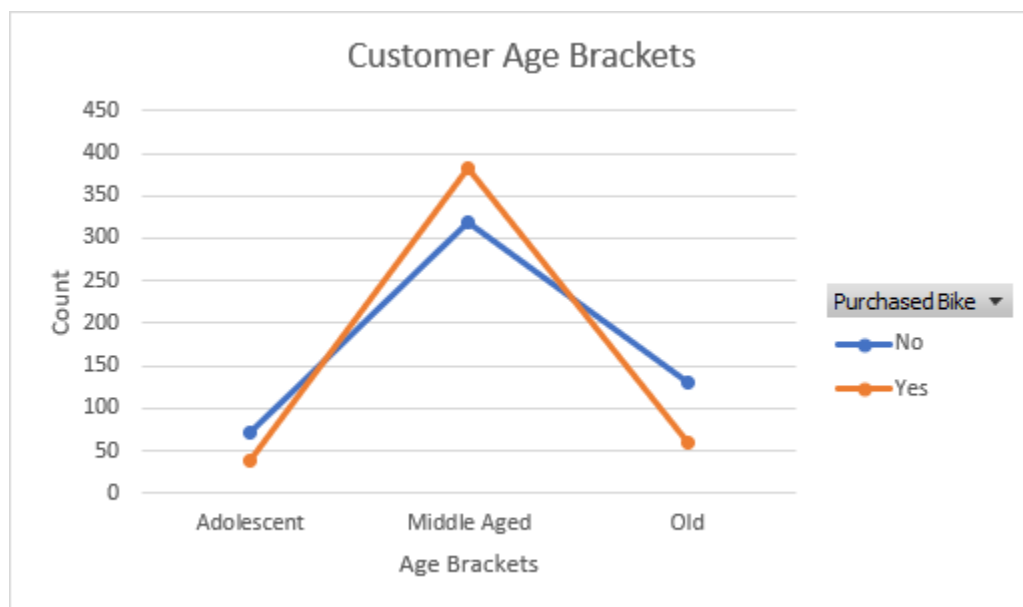
Dashboard

To complete this project, I created a new worksheet that will be used as a dashboard for others to interact with the data and attain insights.

To do this, I filled the cells to be white and pasted the visuals into the sheet in a neat format, with a clear title of the dashboard.

To allow the users of the dashboard to reach more specific information, to gain useful and niche insights, I decided to create 3 filters based of the columns of the dataset.

The 3 filters will be of the following:

- Marital Status (Married or Single)
- Region (Europe, North America or Pacific)
- Education (Bachelors, Graduate Degree, High School, Partial College, Partial High School)

To create filters, I used the Slicer function on a chart and selected one of the 3 columns for filtering and then established connections with the other 2 pivot tables that were used for the other visuals.

Here is the final dashboard with slicers implemented:
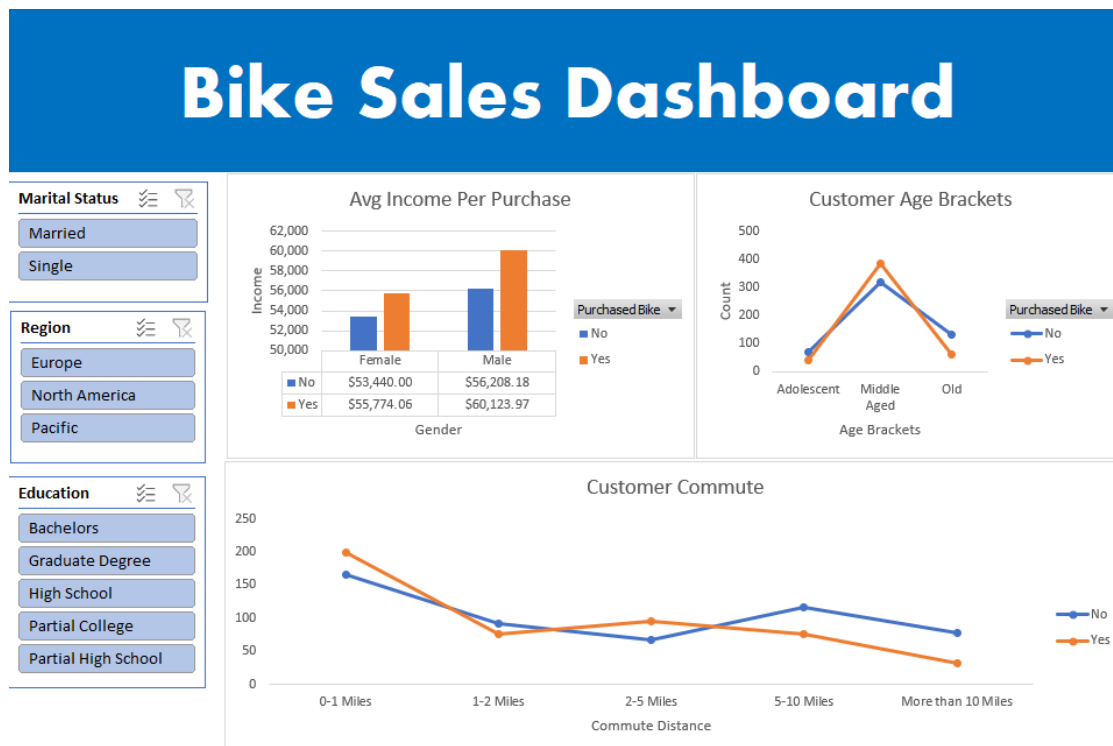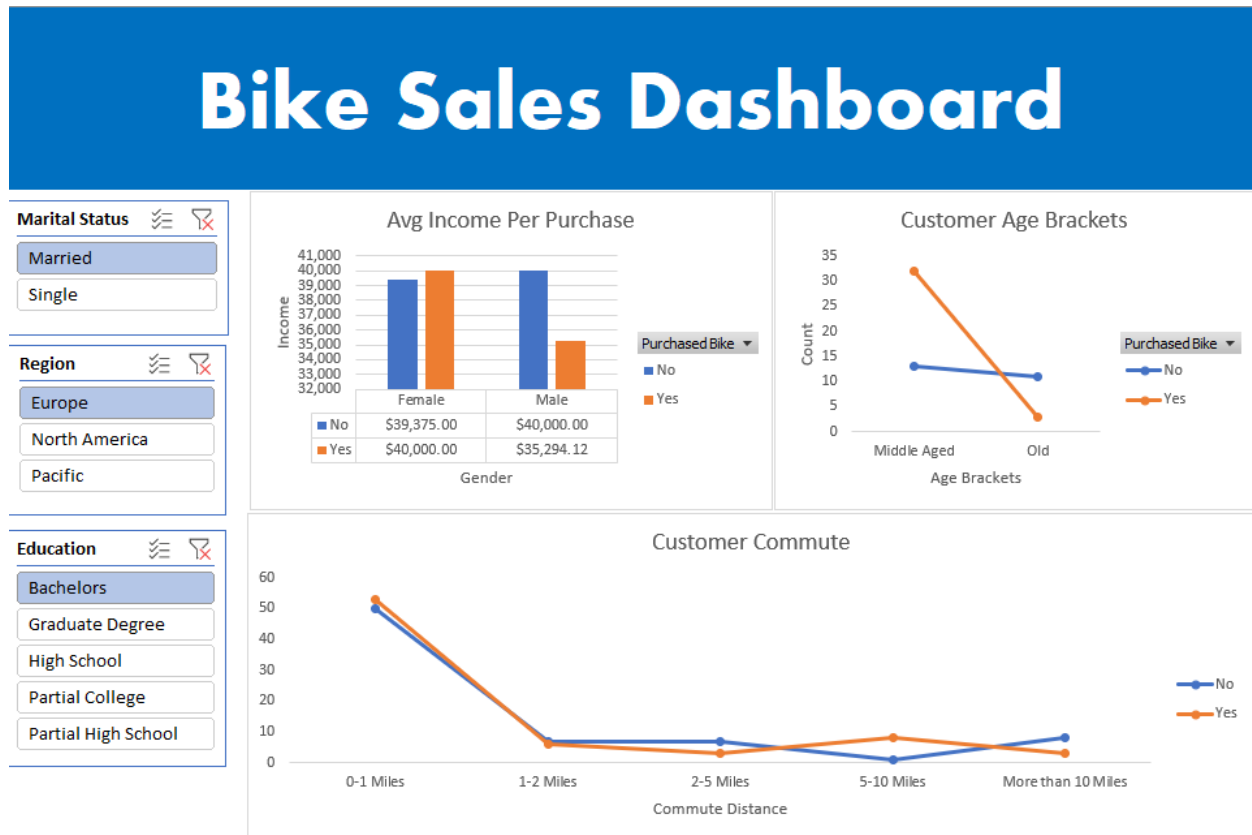
Insights

For this set of insights, I will set the slicers to select data from individuals who are:

- Married
- In Europe
- Completed their Bachelors

As a result, the dashboard alters to display the following:

Customer Commute

The data reveals a clear relationship between commute distance and bike purchases. Most individuals who purchased a bike had a daily commute distance of 2 miles or less, which aligns with the practicality of using a bike for shorter journeys. It is likely that these individuals find it inefficient to spend money on petrol, transport fares, or deal with traffic for such short commutes, making a bike an attractive alternative.

Beyond a 2-mile commute distance, the number of responses declines significantly. However, there remains an opportunity to target individuals with longer commutes. For example, promoting electric bikes or durable hybrid models could be an effective way to appeal to those travelling 2–5 miles. These bikes reduce effort and costs for customers, making them better suited for medium-distance commutes and potentially increasing sales from this group.

Additionally, the data highlights potential barriers to purchasing bikes for those commuting over 10 miles. Investigating these barriers through surveys or focus groups could provide valuable insights into why certain segments are less likely to buy, enabling the store to develop strategies to overcome these challenges.

Marketing strategies should focus on local populations and towns, as a large proportion of respondents travel within short distances. By highlighting the benefits of biking for convenience, cost savings, and sustainability, the client can strengthen engagement with these groups while expanding appeal to medium-distance commuters.

Average Income Per Purchase
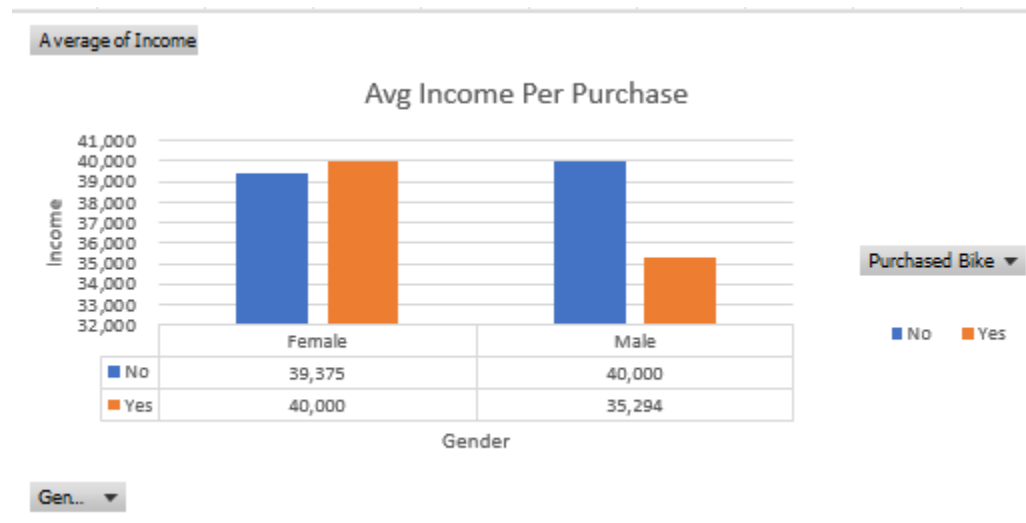
The data indicates that females are more likely to purchase bikes than males, as suggested by the income patterns. This insight could be valuable to the client, as it highlights an opportunity to optimise stock and marketing strategies by focusing on this demographic. Specifically, increasing the variety of bikes targeted towards females could help boost sales.

The higher average income of females who purchase bikes ($40,000) suggests potential for introducing premium products or high-end brands that appeal to this group. Offering bikes with advanced features or aligning marketing messages with their preferences and aspirations could prove particularly effective.

For males, the data shows that the average income of purchasers ($35,294.12) is lower than that of non-purchasers ($40,000). This suggests that affordability and value are likely key factors in male purchasing decisions. It is possible that the store's current range of bikes targeted towards men may focus on affordability but lack diversity or appeal for higher-income male customers. Additionally, this trend aligns with common male purchasing behaviour, where practicality, value for money, and discounts tend to play an influential role.

To increase sales among males, the store could:

1. Introduce premium or higher-quality bike options to attract higher-income male customers.

2. Conduct an analysis of the types of bikes currently purchased by males with lower incomes to identify trends, ensuring these products remain available to maintain revenue from this group.

Another key observation is that the average income of non-purchasers ($39,583.33) is relatively high compared to purchasers ($37,714.29). This indicates untapped sales potential among higher-income individuals who have not yet made purchases. To address this, the store could conduct a survey or sentiment analysis to identify potential barriers to purchase, such as product range, pricing concerns, or competition. Addressing these barriers could help capture a larger share of this market.

Customer Age Brackets and Purchasing
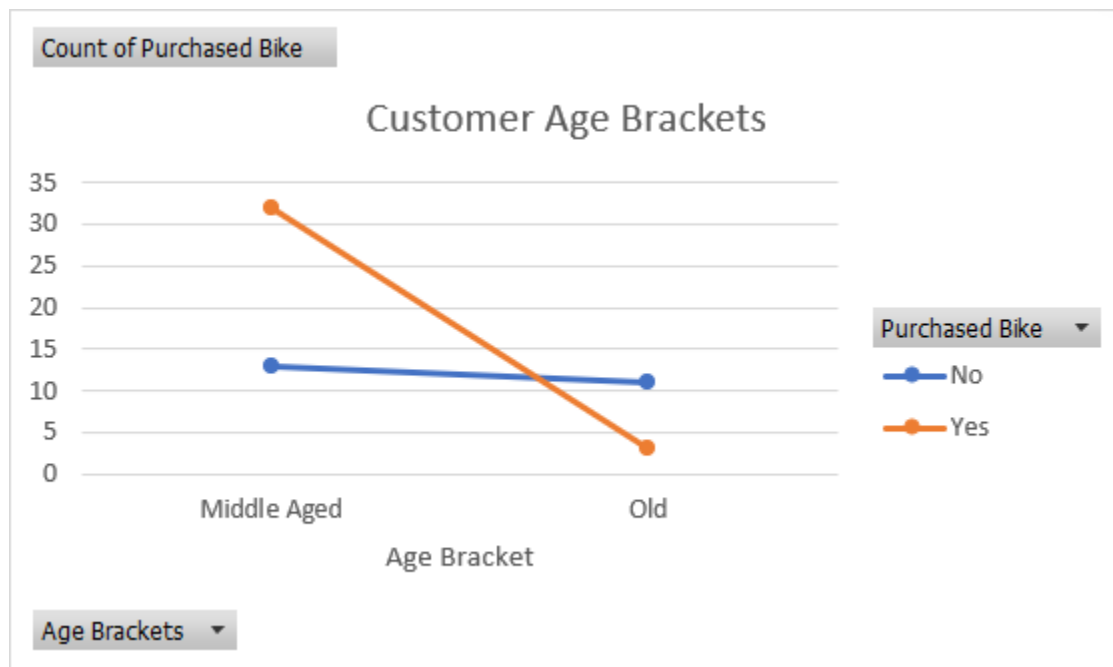
The analysis indicates that middle-aged individuals (31-60 years) represent the majority of bike purchases, accounting for 32 out of 45 purchases in this group. This suggests that individuals within this age bracket are more likely to purchase bikes, potentially due to factors such as financial stability, commuting needs, or leisure activities. In contrast, older individuals (over 54 years) show significantly lower engagement, with only 3 purchases out of 14.

The data for younger individuals (under 31 years) is absent, indicating a potential gap in the dataset or in the targeting strategy for this age group. This absence suggests an opportunity to explore the purchasing behaviour of younger, educated consumers. However, further data collection would be necessary to confirm this trend and to understand the preferences and potential barriers for this demographic.

To capitalise on this opportunity, the client should consider expanding data collection to include younger consumers, particularly those who are still in education or early employment stages and assess their willingness to purchase bikes. Additionally, focusing on the middle-aged segment will likely continue to drive sales, given their dominant share of the market.