Project: Bike Sales Data Analysis

Overview

The purpose of this project is to analyze a data set of bike sales and provide insights into the factors that influence bike purchases. The data set contains information about the customers' demographics, income, education, commute distance, and bike purchase status.

The data set was obtained from: Click here

Introduction

The main objectives of this project are to:

- Clean and format the data set using Excel functions and tools.
- Create new variables based on the existing data, such as age categories and income distribution.
- Perform data aggregation using pivot tables to summarize the data by different dimensions.
- Create a dashboard using charts and filters to visualize the data and answer key questions.

The main questions that I want to answer are:

- How does income distribution vary by gender among bike purchasers and non-purchasers?
- How does commute distance affect bike purchase behaviour?
- How does age category influence bike purchase behaviour?

Phases of Work

The project consists of four phases:

Phase 1: Data Cleaning and Formatting

- Apply filters and conditional formatting to check the data quality and identify errors.
- Correct the income format by using the home ribbon in Excel.
- Divide the customers into three age categories (Adolescent, Middle-Aged, and Old) using nested IF conditions.

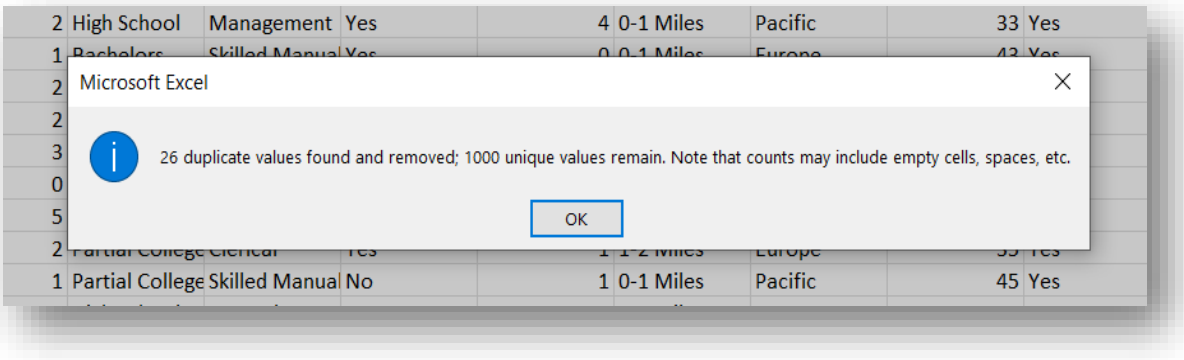
Data Cleaning and Formatting

In this section, I will explain how I cleaned and formatted the data set using Excel functions and tools. I will show the steps that I followed and the results that I obtained.

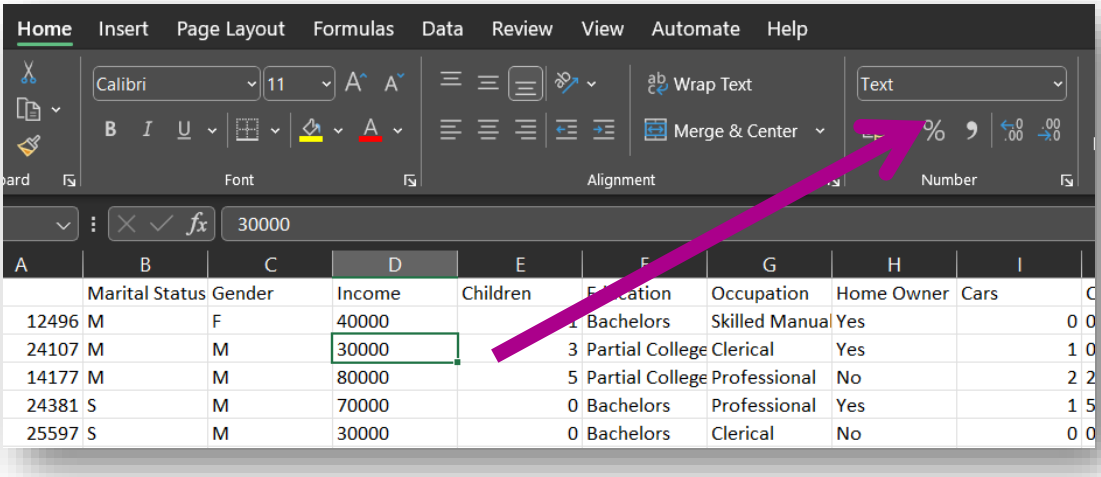
Checking the Data Quality

The first step of the data cleaning and formatting process was to check the data quality and identify errors. I applied filters and conditional formatting to the data set to examine the values and formats of each column. I found that:

- There were no missing values but 26 duplicates in the data set.



- The income column was not correctly formatted as a number, but as a text.



Correcting the Income Format

The second step of the data cleaning and formatting process was to correct the income format. I used the home ribbon in Excel to change the format of the income column from text to number. I also added a comma separator and a dollar sign to make the income values more readable.

Creating Age Categories

The third step of the data cleaning and formatting process was to create age categories based on the age of the customers. I used nested IF conditions to divide the customers into three age categories:

- Adolescent (age <= 30)
- Middle-Aged (30 < age < 54)
- Old (age >= 54).

I added a new column called Age Category to the data set and assigned the corresponding values based on the age column.

Here is a screenshot of the cleaned and formatted data set:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ID	Marital Stati	Gender	Income	Children	Education	Occupation	Home Owne	Cars	Commute D	Region	Age	Age Bracket	Purchased Bike
2	12496	Married	Female	$ 40,000.00	1	Bachelors	Skilled Manual	Yes	0	0-1 Miles	Europe	42	Middle Age	No
3	24107	Married	Male	$ 30,000.00	3	Partial College	Clerical	Yes	1	0-1 Miles	Europe	43	Middle Age	No
4	14177	Married	Male	$ 80,000.00	5	Partial College	Professional	No	2	2-5 Miles	Europe	60	Old	No
5	24381	Single	Male	$ 70,000.00	0	Bachelors	Professional	Yes	1	5-10 Miles	Pacific	41	Middle Age	Yes
6	25597	Single	Male	$ 30,000.00	0	Bachelors	Clerical	No	0	0-1 Miles	Europe	36	Middle Age	Yes
7	13507	Married	Female	$ 10,000.00	2	Partial College	Manual	Yes	0	1-2 Miles	Europe	50	Middle Age	No
8	27974	Single	Male	$ 1,60,000.00	2	High School	Management	Yes	4	0-1 Miles	Pacific	33	Middle Age	Yes
9	19364	Married	Male	$ 40,000.00	1	Bachelors	Skilled Manual	Yes	0	0-1 Miles	Europe	43	Middle Age	Yes
10	22155	Married	Male	$ 20,000.00	2	Partial High Sc	Clerical	Yes	2	5-10 Miles	Pacific	58	Old	No
11	19280	Married	Male	$ 1,20,000.00	2	Partial College	Manual	Yes	1	0-1 Miles	Europe	40	Middle Age	Yes
12	22173	Married	Female	$ 30,000.00	3	High School	Skilled Manual	No	2	1-2 Miles	Pacific	54	Middle Age	Yes
13	12697	Single	Female	$ 90,000.00	0	Bachelors	Professional	No	4	More than 10+	Pacific	36	Middle Age	No
14	11434	Married	Male	$ 1,70,000.00	5	Partial College	Professional	Yes	0	0-1 Miles	Europe	55	Old	No

I used the following code to create the age categories:

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

Phase 2: Data Aggregation using Pivot Tables

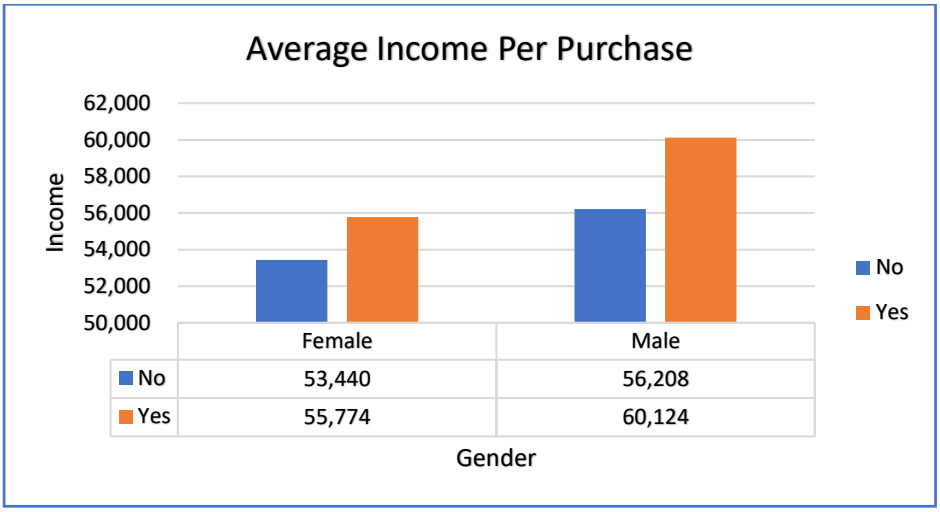
- Calculated and presented the count and percentage of bike purchases by various dimensions using pivot tables.
- Filtered and analyzed the pivot tables by applying slicers and timelines for different criteria.
- Designed and built pivot tables to measure bike purchases across multiple dimensions.

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**

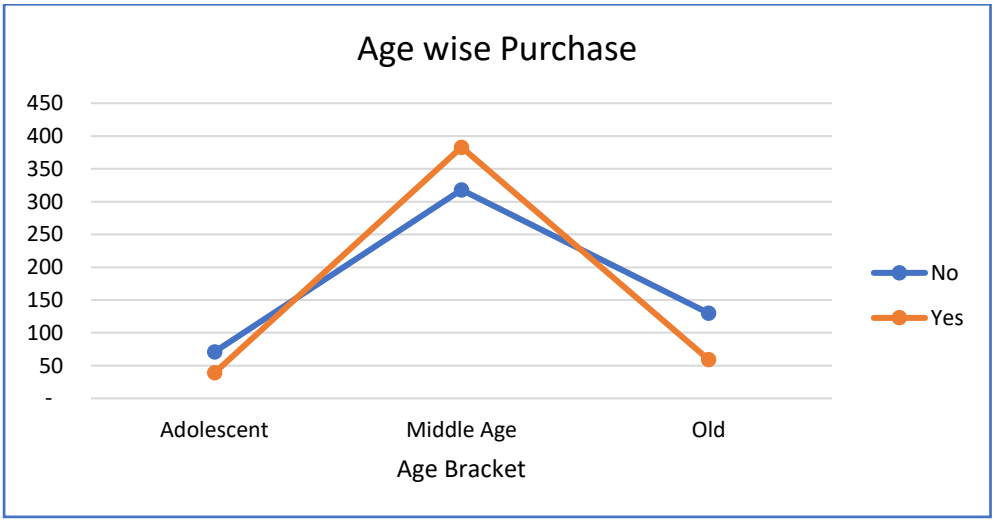
- Utilized slicers and timelines to refine and compare the pivot tables by various criteria.

Phase 3: Data Visualization using Charts.

- Displayed and communicated the data using charts based on the pivot tables.



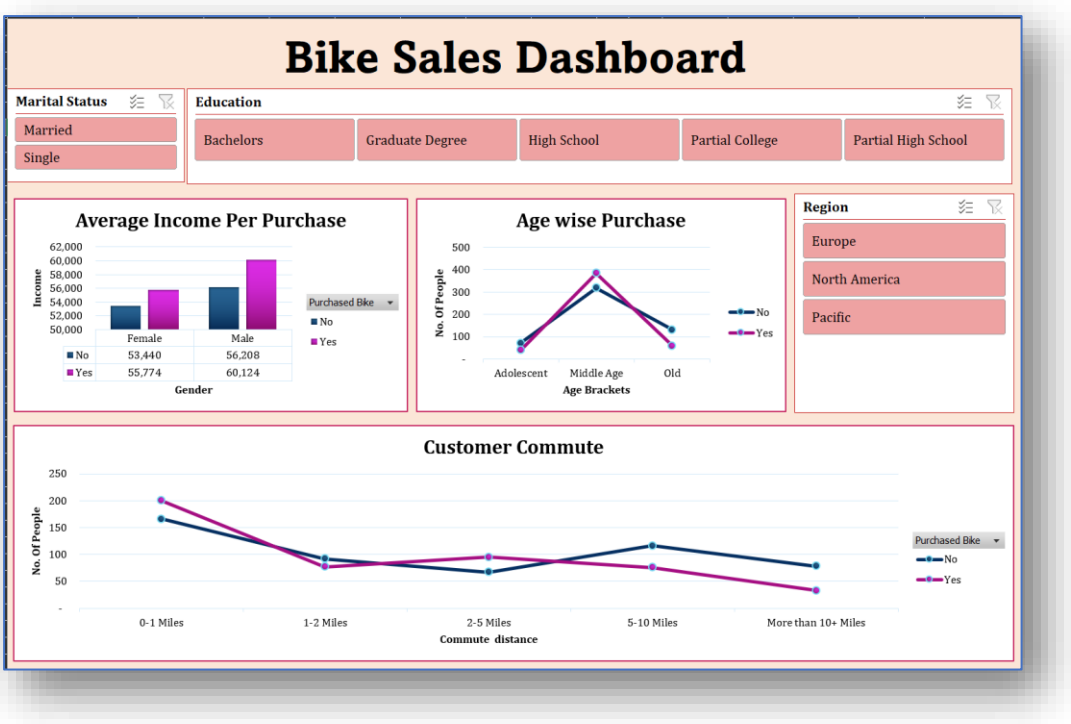
- Selected and applied suitable chart types, such as pie, column, bar, and line charts, to demonstrate the distribution, comparison, and trend of bike purchases.



- Enhanced and clarified the charts by adding titles, labels, legends, and data labels.

Phase 4: Dashboard Creation

- Arranged and integrated the charts and filters in a single worksheet to create a dashboard.
- Applied and maintained a consistent colour scheme and layout to enhance the dashboard's appearance and readability.



- Added and wrote a title and a brief introduction to the dashboard to clarify its purpose and scope.
- Enabled and facilitated the users to interact with the dashboard by applying filters and changing the views.

Conclusion

In this project, I have analyzed a data set of bike sales using Excel and obtained some interesting insights. I have found that:

- The income distribution among bike purchasers and non-purchasers is different by gender. Male bike purchasers tend to have higher incomes than female bike purchasers, while male non-purchasers tend to have lower incomes than female non-purchasers.
- The commute distance affects the bike purchase behaviour. Customers who have shorter commute distances are more likely to purchase bikes than customers who have longer commute distances.
- The age category influences the bike purchase behaviour. Middle-aged customers are more likely to purchase bikes than Adolescent and old customers.

These findings have some implications for the bike sales and marketing strategies. For example, the bike sellers can:

- Target customers who have higher income, shorter commute distances, and (31-53) age
- Offer discounts or incentives for customers who have lower income, longer commute distances, and older age.
- Segment the customers by gender, income distribution, age category, and other variables and tailor the bike products and services accordingly.

Limitations:

However, this project also has some limitations and challenges. For example, the data set is relatively small and may not represent the general population of bike customers. The data set also lacks some important variables, such as the

bike type, price, and quality, that may affect the bike purchase decision. The data analysis methods and tools are also limited by the Excel functions and features and may not capture the complex relationships and patterns in the data.