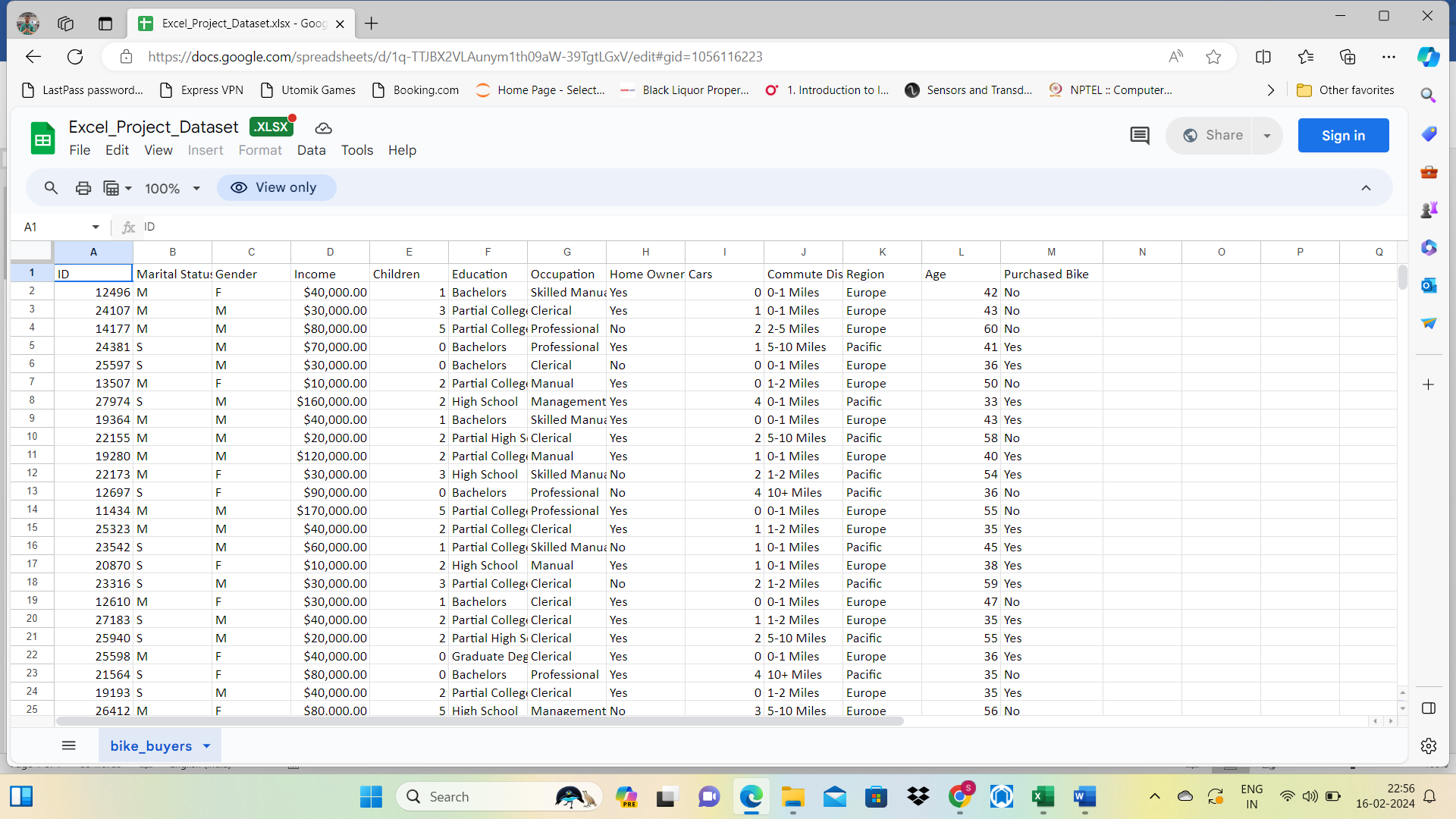
PROJECT: BIKE SALES DATA ANALYSIS

**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:



**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.

**Objectives**

The main objectives 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

Phase 2: Data Aggregation using Pivot Tables

Phase 3: Data Visualization using Charts.

Phase 4: Dashboard Creation

**1.1 Data Cleaning and Formatting**

1. Apply filters and conditional formatting to check the data quality and identify errors.

2. Correct the income format by using the home ribbon in Excel.

3. Divide the customers into three age categories (Adolescent, Middle-Aged, and Old) using nested IF

conditions.

4. Convert marital Status “M” To Married and “s” to “single”

**1.2 Checking the Data Quality**

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

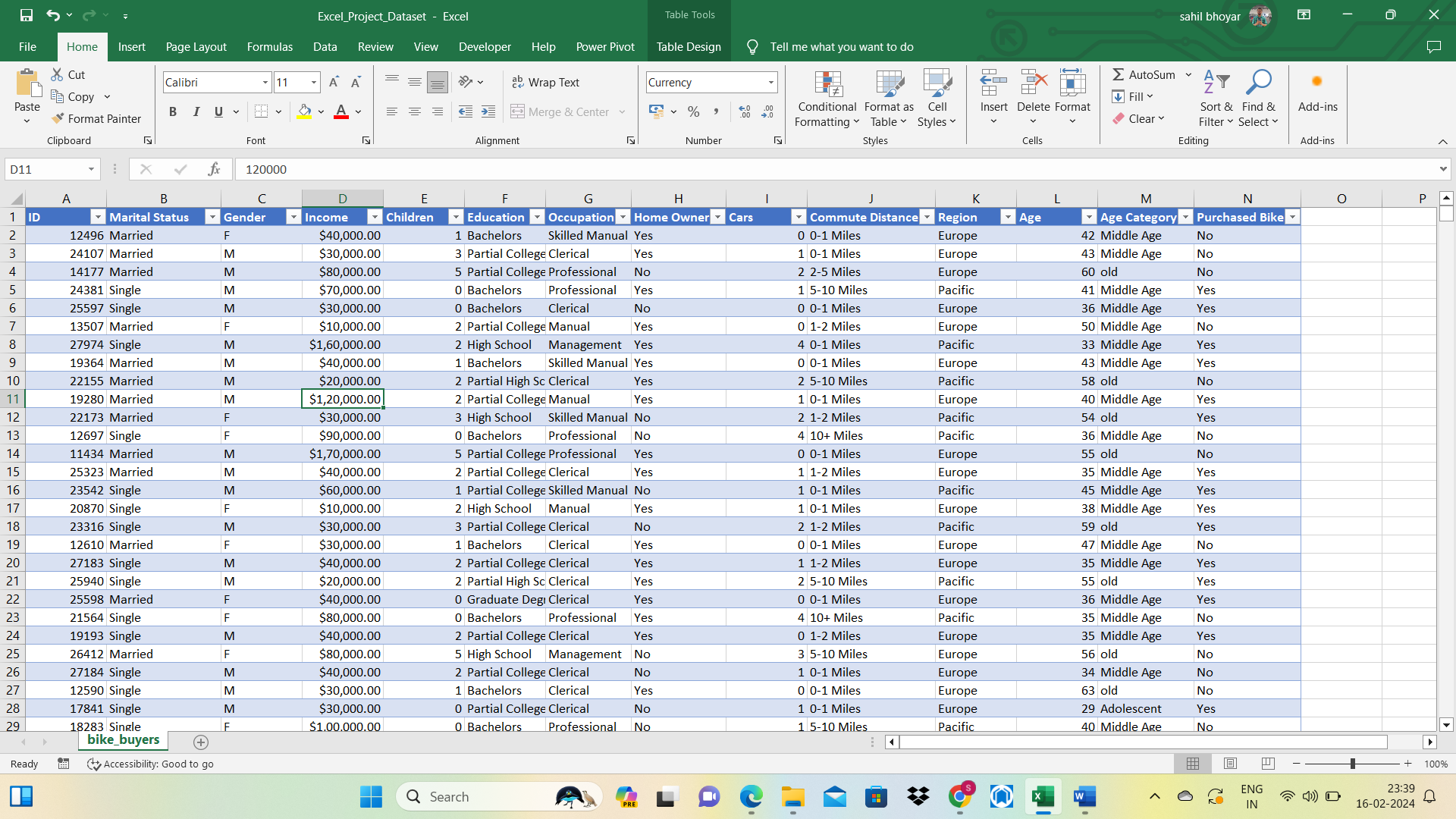
* There were no missing values but 26 duplicates in the data set.
* The income column was not correctly formatted as a currency but as a text converted into currency format.
* Creating Age Categories by the data cleaning and formatting process was to create age categories based on the age of the customers. used **IF condition.**

nested IF conditions to divide the customers into three age categories:

▪ Adolescent (age <= 30)

▪ Middle-Aged (30 < age < 54)

▪ Old (age >= 54).

After cleaning the data the clean data shows as :

**2.1 Data Aggregation using Pivot Tables**

1. Calculated and presented the count and percentage of bike purchases by various dimensions using pivot tables.

2. Filtered and analyzed the pivot tables by applying slicers and timelines for different criteria.

3. Designed and built pivot tables to measure bike purchases across multiple dimensions.

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

**3.1 Data Visualization using Charts**

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

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

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

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.

The occupation-wise income shows, that most professionals have higher incomes compared to others. Management occupation is second after the professional.

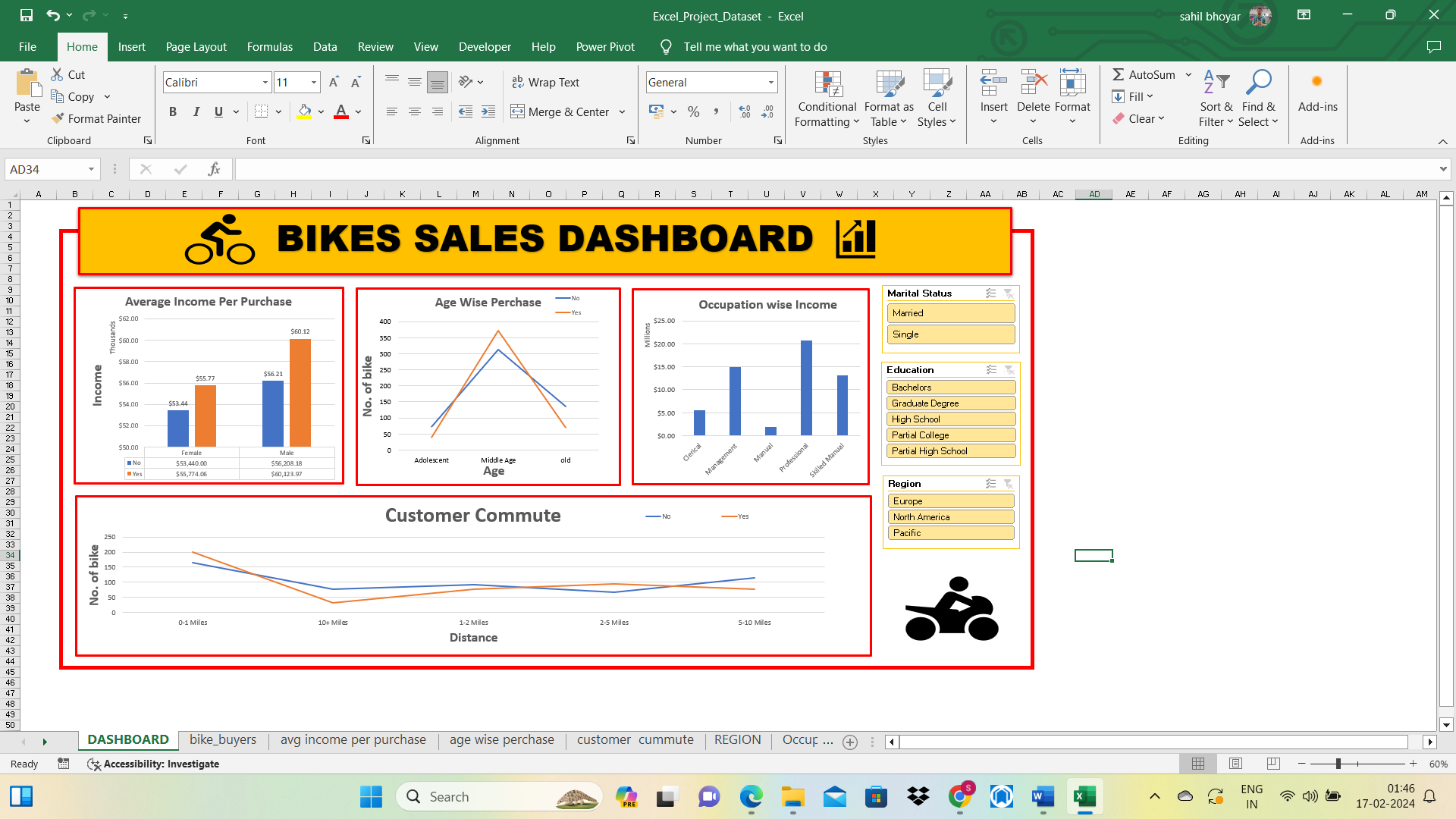
These findings have some implications for the bike sales and marketing strategies.

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