

DATA ANALYSIS

- it the science where we inspect, data cleaning, transformation & Modeling for business insights and presenting to the stakeholder

1 EXCEL

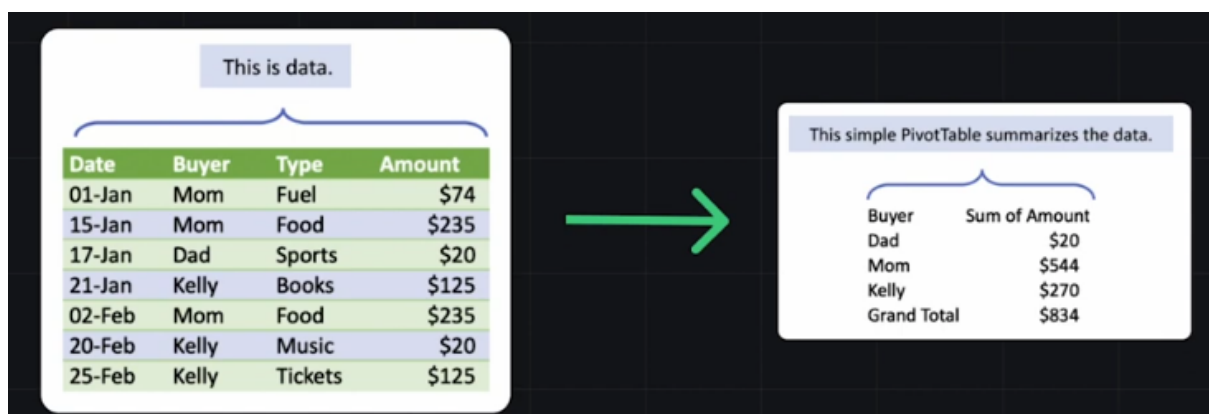
- Pivot Tables
- Function
- XLookup
- Conditional formatting
- Data Visulisation
- Data Cleaning

Process Of Analysis

- Data Requirement
- Data Collection
- Data Processing
- Cleaning
- Analysis
- Communication

1.1 Pivot

Summarises the data to make sense to you without typing any formula's



IMPLEMENTATION

(Kaggle used for obtaining datasets)

Insert —> Pivot table  PIVOT TABLE

	A	B
1	Country	SUM of Revenue
2	Australia	£21,302,059.00
3	Canada	£7,935,738.00
4	France	£8,432,872.00
5	Germany	£8,978,596.00
6	United Kingdom	£10,646,196.00
7	United States	£27,975,547.00
8	Grand Total	£85,271,008.00

1.2 FUNCTION

- You can add ,subtract , multiply, & divide in excel without using any built in functions
- It give the ability to do variety of things like perform mathematical operation

- | | |
|---------|---------------|
| • MAX | • CONCATENATE |
| • MIN | • SUBSITUTE |
| • IF | • SUM |
| • IFS | • SUMIF |
| • LEN | • SUMIFS |
| • LEFT | • COUNT |
| • RIGHT | • COUNTIF |
| • TEXT | • COUNTIFS |
| • TRIM | • DAYS |
| | • NETWORKDAYS |

FUNCTIONS

1.3 XLOOKUP

- **searches for a value in a range or array** and returns a corresponding value from another range. It is a more flexible and improved replacement for **VLOOKUP**, **HLOOKUP**, and **INDEX-MATCH**.
- Searches a range or array and then returns the item corresponding to the first match it find if no match exist, then XLOOKUP can return the closest match
- =XLOOKUP(lookup_value, lookup_array, return_array, [if_not_found], [match_mode], [search_mode])

XLOOKUP

1.4 CONDITIONAL FORMATTING

- It enables you to color the cell depending on the condition, excellent way to visualise the data in a spreadsheet. You can create custom rules
- You can spot significant trend

WE ARE GOING TO LEARN

- Highlight cells
- Creating formula
- Creating and editing rule
- Changing the format

Conditional formatting

1.5 CHARTS

- Create charts using excel data

CHARTS

1.6 DATA CLEANING

- It is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate & Incomplete data.
- When combining multiple data source, there are chances of being duplicated

DATA CLEANING V TRANSFORMATION

- Transformation is the process of converting data from one format to another structure
- Transoframtion can be referred as data wrangling or munging
- Standardise data
- Using proper function to capitalise the data
- Using filters for identifying errors and data & use find me also to replace
- Using Format & cleanup suggestion.

DATA CLEANING

PROJECT : BIKE BUYERS DATASET

The purpose of this repository is to look at what factors may go in to determining whether or not a customer opts to purchase a bike. The dashboard displayed below is interactive upon downloading the Bike Buyers Data Excel Sheet in the main branch.

DATA PREPROCESSING

- Duplicate rows were removed
- Column values were adjusted for the visualizations
 - Marital Status: M/S --> Married/Single
 - Gender: M/F --> Male/Female
- Income decimals removed for a more polished looking income number
- New column created for Age Brackets, using a nested IF statement
 - Varying individual ages with a large range --> 3 age group categories (Young, Middle Aged, Older)
- Reformatted the Commute Distance Column Values for sequential ordering in the visualizations
 - "10+ Miles" --> "More than 10 Miles"

Problem 1, Bar Chart (Marital Status):

Question: How does the count of bike purchases vary among different marital statuses? Are married individuals more likely to purchase bikes?

Step 1: Start the excel sheet.

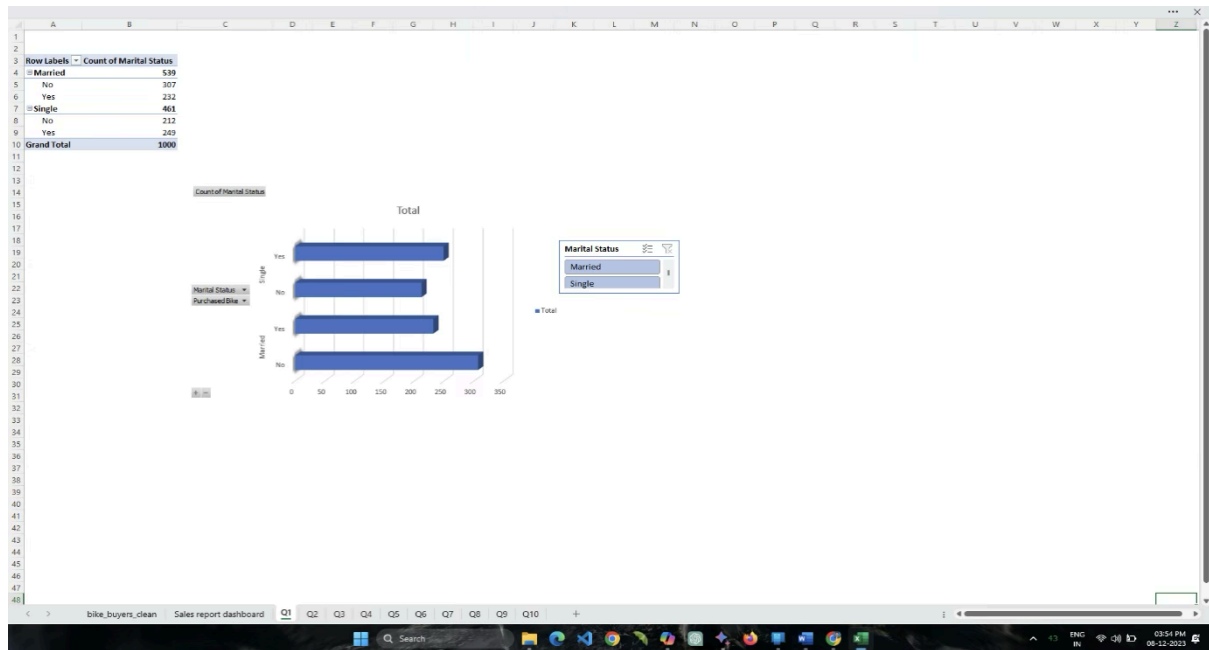
Step 2: Load the value bike buyers clean.

Step 3: Open Excel sheet-file-import the values.

Step 4: Create pivot table insert pivot table open.

Step 5: Select the row & columns, add marital status, purchased bike and fill the Values

Step 6 : Click insert chart



The question ask about the bike buyers material status and theirs purchased behaviour to Analysing and creating by Bar chart creating to visual is the values also It has Filters to View the Exact Analysis of the bike buyers behavior.

Problem 2, Bar Chart (Gender):

Question: Build a bar graph to compare the count of male and female customers. Does gender influence bike purchases, and if so, to what extent?

Step 1: Start the Excel sheet.

Step 2: Load the value bike buyers clean.

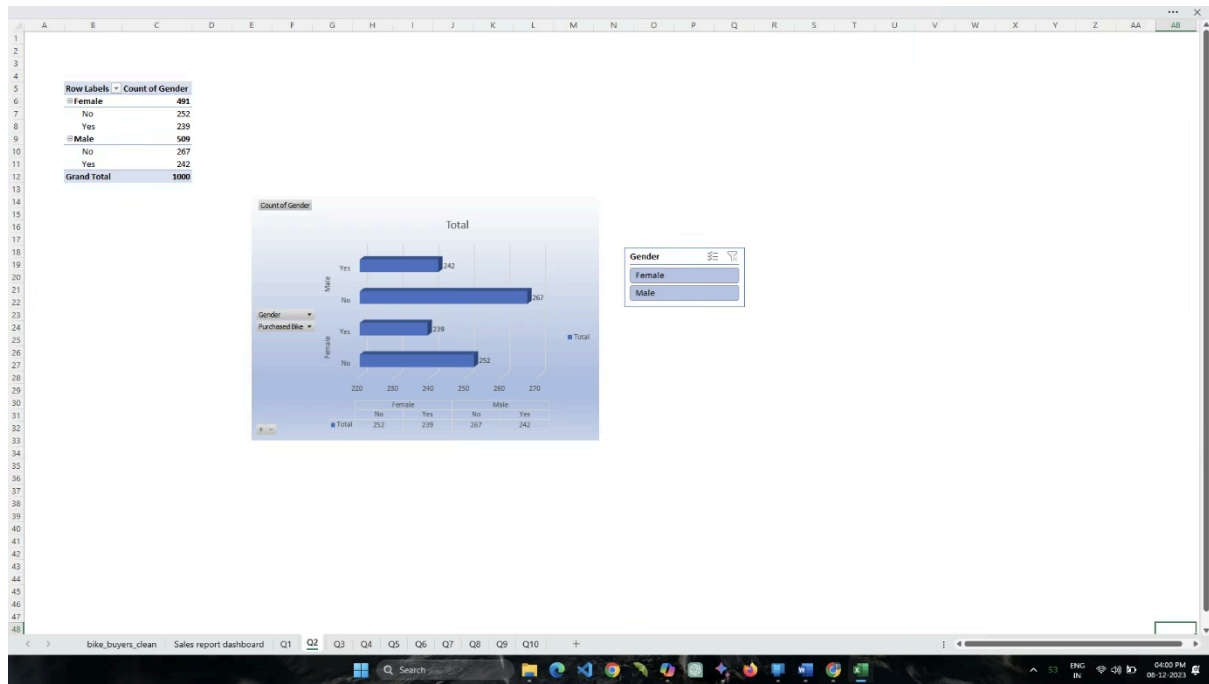
Step 3: Open Excel sheet →file →import the values.

Step 4: Create pivot table insert → pivot table →open.

Step 5: Select the row & columns, Gender, Purchased bike add and fill the values.

Step 6: Click insert chart.

Analysing and visualizing the relationship between bike buyers' Gender and their purchasing behaviour through bar charts and The sheet is equipped with filters that enable a precise analysis of the behaviour of bike buyers.



Problem 3, Histogram (Income):

Question: What is the distribution of income among bike buyers? Are there specific income brackets that show a higher likelihood of bike purchases?

Step 1: Start the Excel sheet.

Step 2: Load the value bike buyers clean.

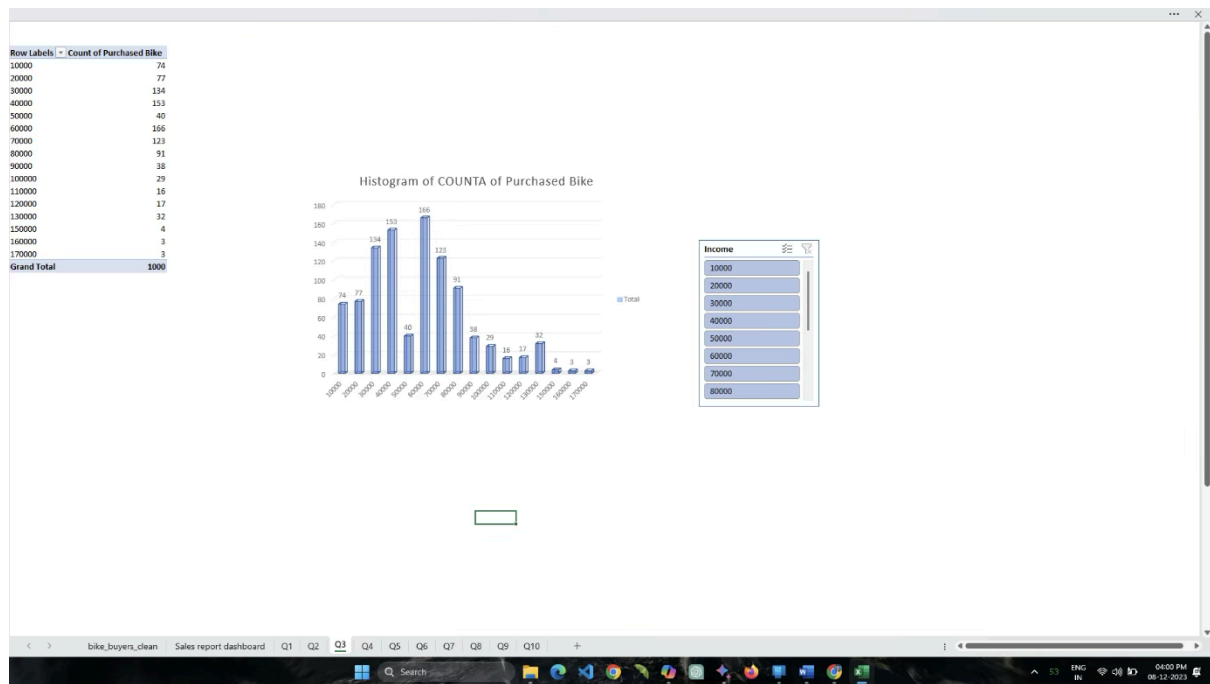
Step 3: Open Excel sheet →file →import the values.

Step 4: Create pivot table insert → pivot table →open.

Step 5: Select the row & columns, Gender, Purchased bike,income add and fill the values.

Step 6: Select the income value → format → number→ currency change.

Step 7: Click insert chart.



I employ a visually appealing histogram chart to depict the income distribution of bike buyers and analyze With filters integrated, the sheet facilitates a meticulous analysis of bike buyer's behaviour.

Problem 4, Histogram (Age):

Question: Create a histogram to understand the age distribution of bike buyers. Are certain age groups more inclined to purchase bikes?

Step 1: Start the Excel sheet.

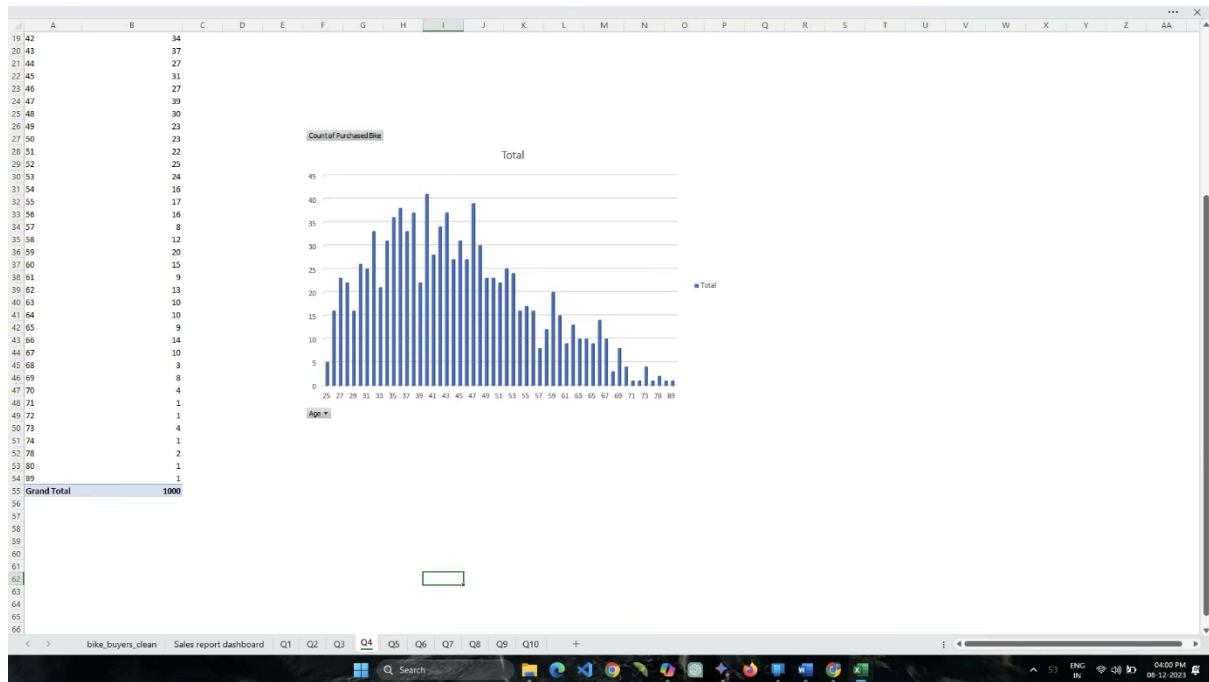
Step 2: Load the value bike buyers clean.

Step 3: Open Excel sheet →file →import the values.

Step 4: Create pivot table insert → pivot table →open.

Step 5: Select the row & columns, Age, Purchased bike add and fill the values.

Step 6: Click insert char



This graphical sheet illustrates the distribution of purchased bikes and the corresponding age demographics of customers.

Problem 5, Box Plot (Income):

Question: Identify outliers in the income distribution of bike buyers. Are there any extreme income values, and how might they impact purchasing behavior?

Step 1: Start the Excel sheet.

Step 2: Load the value bike buyers clean.

Step 3: Open Excel sheet → file → import the values.

Step 4: Create pivot table insert → pivot table → open.

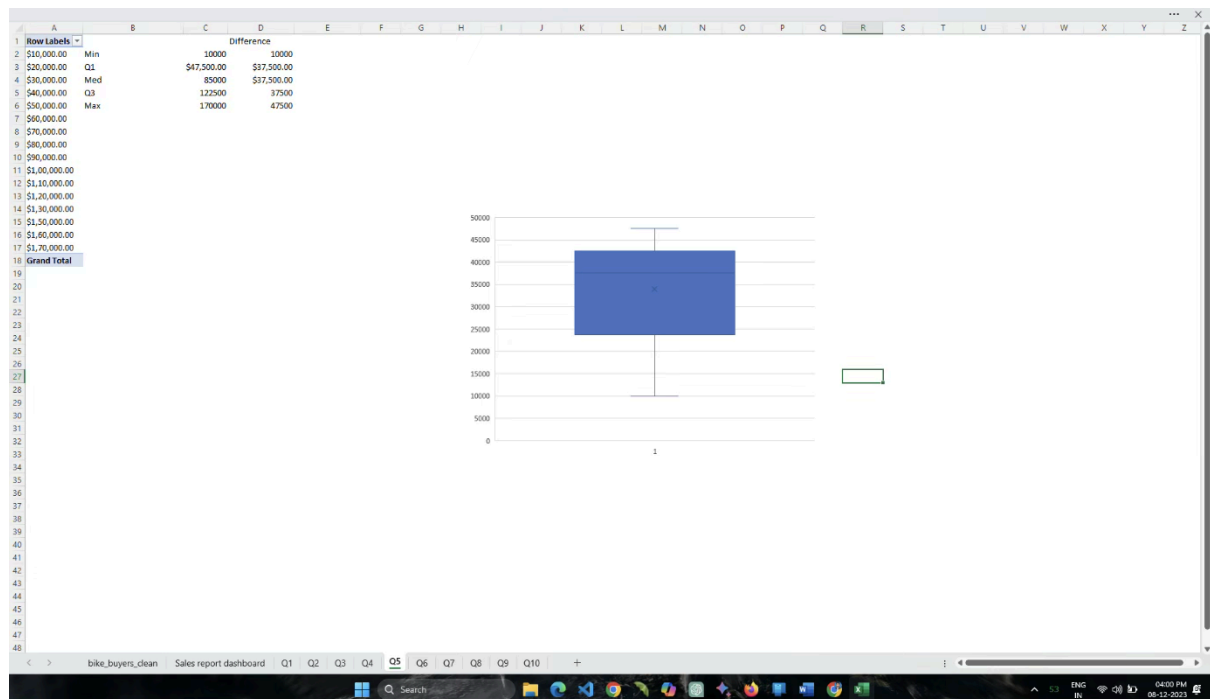
Step 5: Select the row Income add and fill the values.

Step 6: Select the income value select C cell Max, Mini, Median E cell Difference.

Min fx=QUARTILE.INC(A\$2:A\$17,0)..... In the D cell calculate the values.

Difference fx=D3-D4..... In the upper value and lower value. calculate the values.

Step 7: Click insert chart.



Employing box plots to depict the income distribution of bike buyers and their corresponding values, we can effectively illustrate the relationship between purchasing behavior and income levels.

Problem 6, Pie Chart (Region):

Question: Represent the distribution of bike purchases by region using a pie chart. Are there regions where bike purchases are notably higher?

Step 1: Start the Excel sheet.

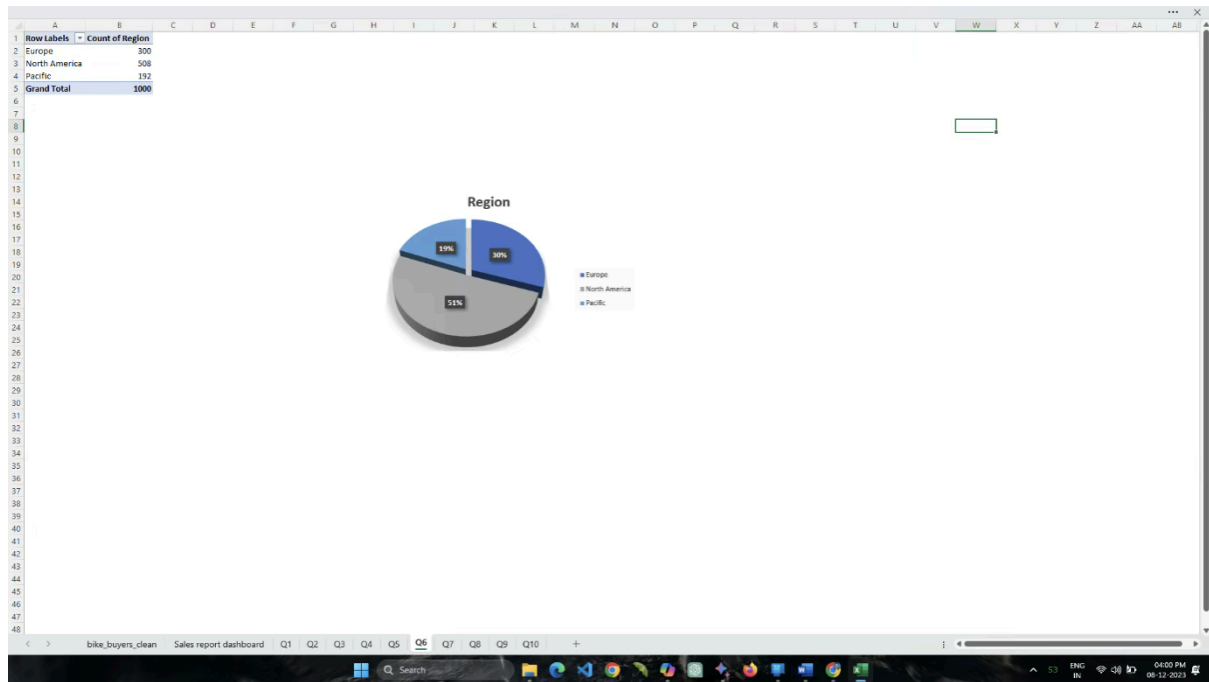
Step 2: Load the value bike buyers clean.

Step 3: Open Excel sheet → file → import the values.

Step 4: Create pivot table insert → pivot table → open.

Step 5: Select the row & columns, Region, Income add and fill the values.

Step 6: Click insert chart.



Professionally visualizing the Europe, North America, Pacific regions leading in bike sales through a pie chart.

Problem 7, Scatter Plot (Income vs. Age):

Question: Create a scatter plot to investigate the relationship between income and age. Do individuals with higher incomes tend to be in specific age groups?

Step 1: Start the Excel sheet.

Step 2: Load the value bike buyers clean.

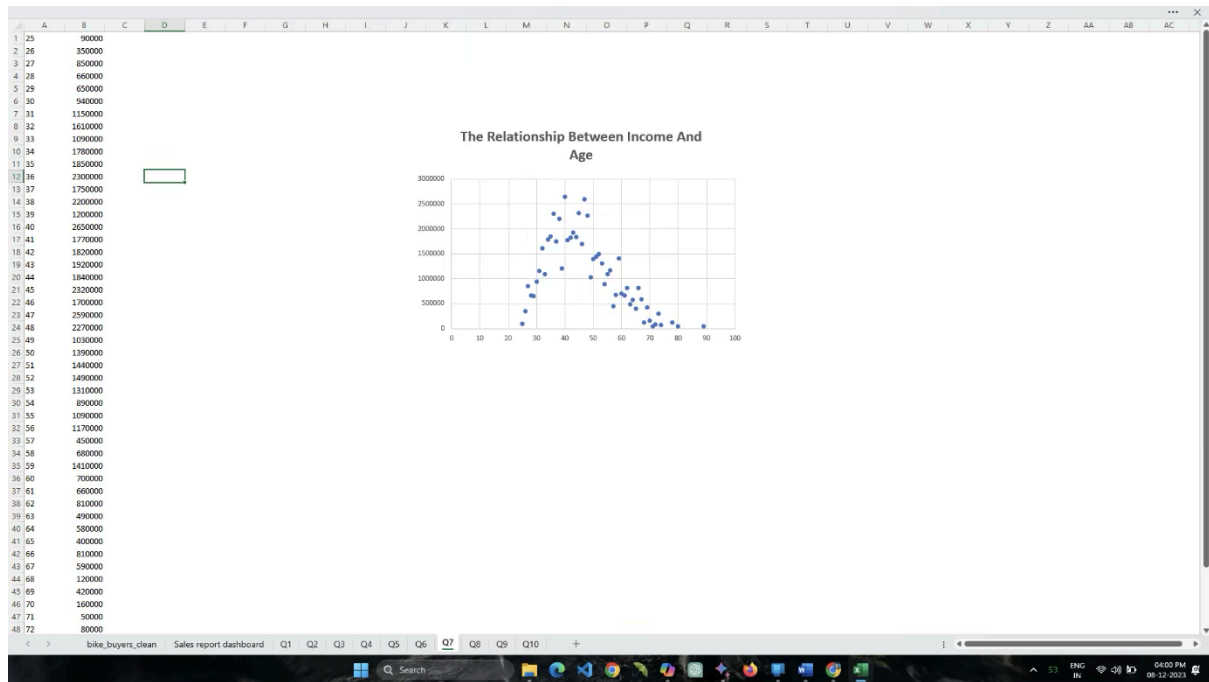
Step 3: Open Excel sheet → file → import the values.

Step 4: Create pivot table insert → pivot table → open.

Step 5: Select the row & columns, Age, Income add and fill the value.

Step 6: Select the row income to value, right click on the age column and create a pivot table group rule (option) show. Keep the values & click ok button.

Step 7: Click insert chart.



Professionally design a scatter plot to analyze the correlation between the age of bike buyers and their corresponding income levels within a specific age group, highlighting trends in higher income.

Problem 8, Stacked Bar Chart (Marital Status & Gender):

Question: How does the distribution of bike purchases differ when considering both marital status and gender simultaneously? Are there notable patterns?

Step 1: Start the Excel sheet.

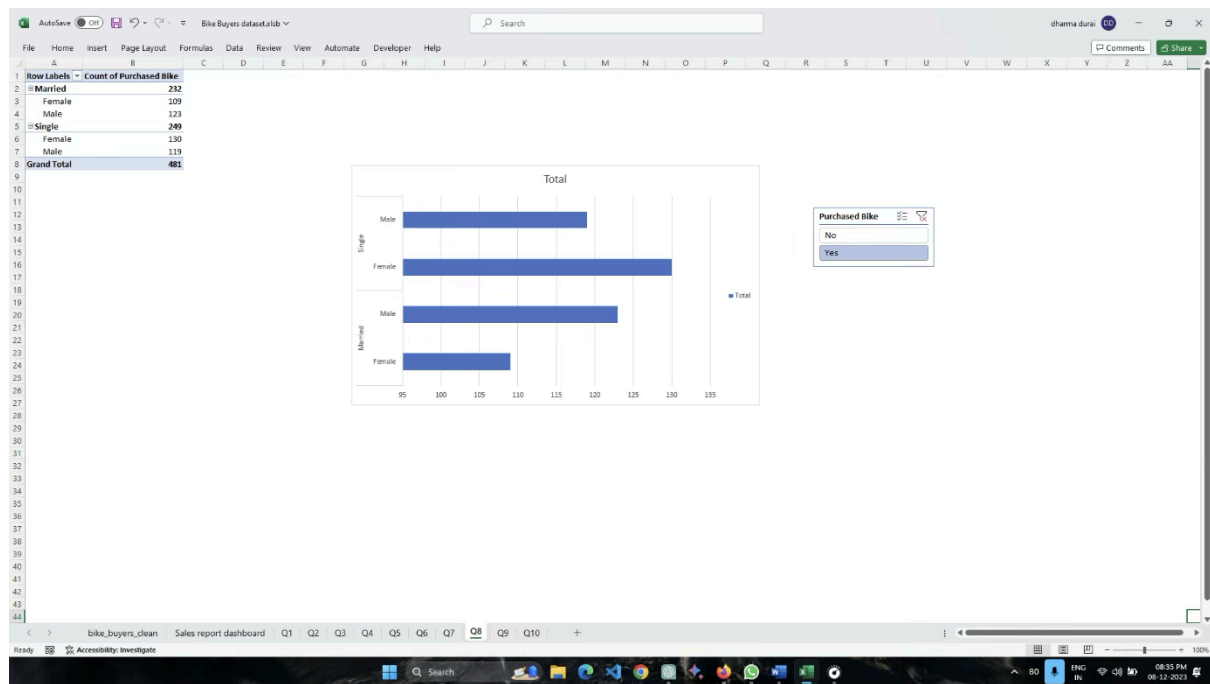
Step 2: Load the value bike buyers clean.

Step 3: Open Excel sheet → file → import the values.

Step 4: Create pivot table insert → pivot table → open.

Step 5: Select the row & columns, Marital status , Gender add and fill the values.

Step 6: Click insert chart.



Professional rearrangement: "Examine the purchasing behavior of bike buyers, observing how it varies based on both marital status and gender."

Problem 9 Correlation Heatmap (Numeric Variables):

Question: Use a heatmap to visualize the correlation matrix between numeric variables. What variables show a strong correlation, and how might this influence purchasing behavior?

Step 1: Start the google sheet.

Step 2: Load the value bike buyers clean.

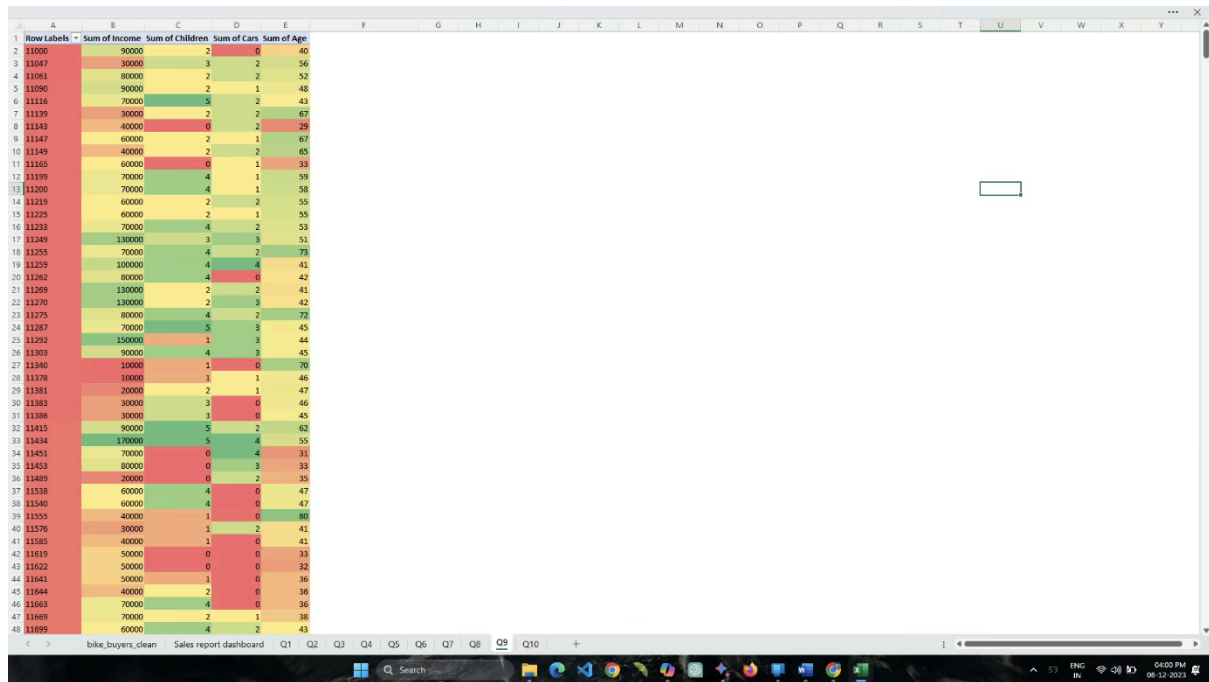
Step 3: Open google sheet → file → import the values.

Step 4: Create pivot table insert → pivot table → open.

Step 5: Select the ID row & columns sum of income, children, cars, age add and fill the values.

Step 6: Click Home & Insert Conditional formatting

Step 7: Click color scales.



heat map is a graphical representation of data where values are depicted using color gradients. It allows users to visualize and analyze data patterns by assigning colors to cells based on their numerical values. Typically, higher values are represented by warmer or more intense colors, such as red, while lower values are associated with cooler or lighter colors, such as green or blue. Heat maps are valuable for identifying trends, patterns, and variations in data, providing an intuitive way to interpret complex information at a glance.

Problem 10 Correlation Heatmap (Numeric Variables):

Question: Create a pair plot for a subset of variables (e.g., Income, Age, Children). Are there clear relationships between these variables, and how might they impact bike purchases?

Step 1: Start the google sheet.

Step 2: Load the value bike buyers clean.

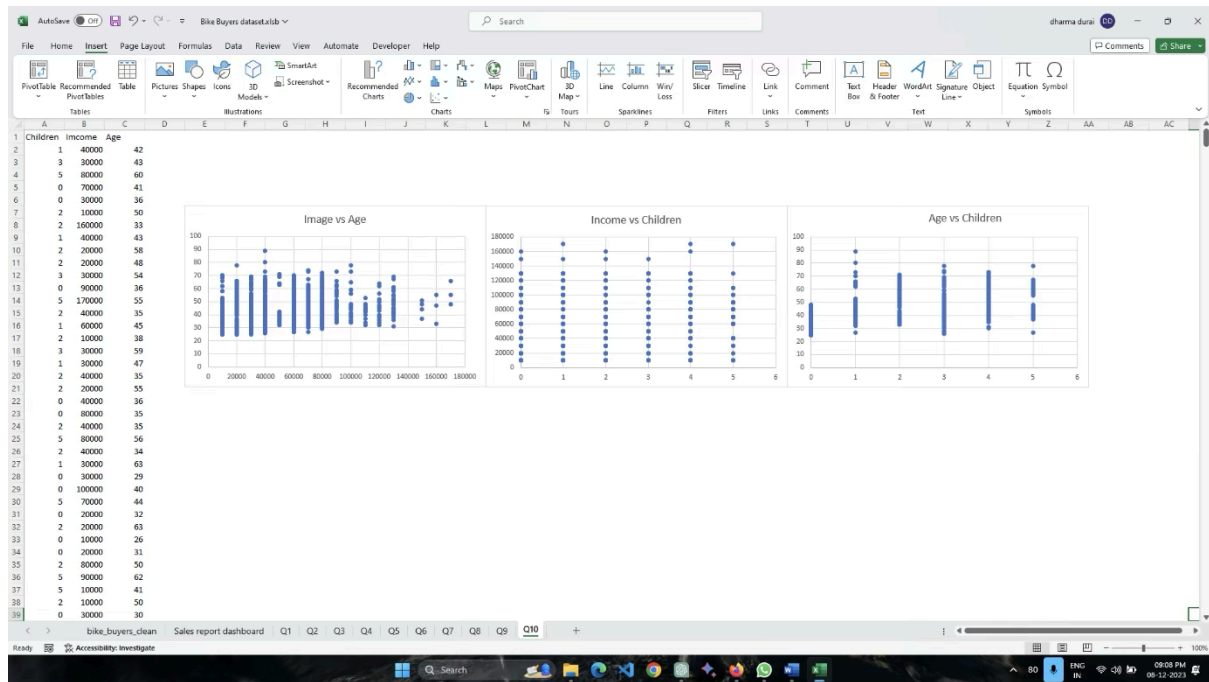
Step 3: Open google sheet →file →import the values.

Step 4: Create pivot table insert → pivot table →open.

Step 5: Select the row & columns of income, children & age

Step 6: Click Home & Insert

Step 7: Click Scatter Plot Chart



A pair plot, also known as a scatterplot matrix, displays scatterplots for pairs of variables in a dataset. In this type of plot, each variable is compared to every other variable, forming a matrix of scatterplots. The diagonal of the matrix typically shows histograms or kernel density plots for each individual variable.

The pair plot is useful for visually assessing relationships and distributions between multiple variables simultaneously, aiding in the identification of patterns, trends, and potential correlations in complex datasets.