**Data Analysis Project Documentation: Bike Sales**

**Overview of the Dataset:**

* The dataset under analysis is an Excel file that contains detailed information about current and prospective customers. It includes demographic details such as age, martial status and bike purchase status.

**Data Acquisition and Backup:**

1. **Download and Initial Setup:**

* Initially downloaded the dataset as an excel file
* Created a backup sheet named ‘Original Data’ to preserve the original dataset. This was done to ensure we had a reference point and to safeguard against any issues that might arise during the analysis process.

**Data Cleaning Steps:** (Sheet = Working Sheet)

1. Duplicate Removal: Deleted Duplicate entries to ensure data accuracy and integrity.
2. Standardization of Marial Status: Replaced ‘M’ and ‘S’ with ‘Married’ and ‘Single’ respectively in the Marital Status column to maintain consistency and to enhance clarity in visualizations.
3. Standardization of Gender: Replaced ‘M’ and ‘F’ with ‘Male’ and ‘Female’
4. Income Formatting: Changed the type of the Income column to currency format in dollars ($) for better readability and analysis.
5. Data Validation: has been applied to the children column, ensuring that entries accurately reflect the number of children, ranging from 0 to 20

**Data Transformation Steps:**

1. Age Range Categorization: To improve data visualization, a new column was created to define age ranges using an IF statement. This categorizes individuals as ‘Adolescent’ for ages below 31, ‘Middle age’ for ages between 31 and 54, and ‘Old’ for individuals above 54.
2. Building Pivot Tables: (Sheet = Pivot Table)

* Average Income Analysis: Created a pivot table to determine the average income of individuals who did not purchase a bike.
* Commute Distance Analysis: Created a second pivot table to analyze if commute distance affects bike purchases.
* Age Impact Analysis: Created a pivot table to asses if range affect bike purchases.

**Exploratory Data Analysis (EDA):** (Sheet = Working sheet)

* To ensure there were no empty cells throughout the dataset, I utilized the Special dialog in Excel to identify any cell that lacked a value.
* Descriptive statistical (Average income of female/Male, Max Income Female/Male, Min Income Female/Male, Number of Female/Male in the dataset)
* Assess the relationship between car ownership and bike purchase behavior. This involved counting occurrences where individuals owned a car and subsequently purchased a bike.

**Results:** (Sheet = Dashboard)

Insights include:

* the average salary of a female who did not buy a bike is 53 440 $
* the average salary of a female who did not buy a bike is 56 208 $
* the analysis reveals that a larger number of individuals who own cars do not purchase bikes compared to those who do not own cars. This indicates a potential trend where car ownership may correlate with a lower likelihood of bike purchase.
* Using Standard deviation to understand how values spread relative to the mean. For example, in case of income and age, it illustrates the range of variation around their respective average, income with 31,085 that indicates that incomes within the dataset vary widely from the average income, and stdev age with 11,36 which means in practical terms, that most individuals in the dataset have ages that typically range from about 11,36 years below the mean age to 11,36 years above the mean age, not a high standard deviation.

**Issues and Resolution:**

* An issue was encountered with the ‘10+’ value in the commute distance column, affecting the visualization order more specifically the X axis. This value was replaced with ’10 Miles Plus’ to ensure proper ordering.
* When attempting to calculate the correlation between the income and the purchase of the bike, a division by zero issue arose. To address this, a new column was created where values of 0 and 1 were assigned to represent ‘no’ and ‘yes’ respectively.

**Final Dataset:**

The cleaned and transformed dataset is ready for detailed analysis, ensuring data integrity and providing actionable insights into customer behaviors and preferences.