**Ms-Excel-Project.**

## **Introduction**

* In this project, a 'bike purchase' dataset is presented.

the aim is to **clean, transform, analyze and visualize the dataset** in order to **draw insights** that will **guide the bike seller** in **making informed marketing decisions** hence improving his/her bike sales.

## **Purpose of Project**

**1)**. Identify which demographic of the identifiers purchased bikes more compared to their counterparts.

* This will grant the bike seller insight on which demographic to market more to in order to maximize bike sales.

**2)**. Identify which demographics are underexplored and have a potential to purchase bikes at a high rate.

* This will guide the bike seller in diverting marketing targets from demographics that are least interested in bike purchase to demographics that have a greater potential to purchase more bikes.

## **Dataset Overview**

* The **dataset is obtained** from the open source plattform **'Kaggle'** and is accessible to the general public for use.
* Check the number of Row and Columns

## **Dataset Cleaning**

* The first step in data analysis is often data cleaning. In this case, I proceeded to clean the dataset after data overview.
* Copy and paste the dataset into a new excel sheet, namely, 'working sheet’. The **process of data cleaning and data transformation takes place on the working sheet**. This allows for retention of the raw data in its original form at the original excel sheet.
* The first step in data cleaning is **removing any duplicates** from the data set.**Method**:
* Check each column for any inconsistencies or in accuracies.

## **Dataset Transformation**

* **Check each column for any data type,data value or data format transformation required**.
* In **'Marital status' column**,replace **'M'** data values with **'Married'** and also replace **'S'** data values with **'Single'**.This is because other users may not be familiar with the given abbreviations.It is best to use univerasally known words. **Proccess**: select column and use the find and replace function available at the home tab.
* In **'Gender' column**,replace **'M'** data values with **'Male'** and also replace **'F'** data values with **'Female'**.This is because other users may not be familiar with the given abbrevations.It is best to use commonly known words. **Process** : select column and use the find and replace function available at the home tab.
* **Convert** the **'Income' column** to **currency form**. **Process**: >select column >number > currency. In addittion,make the column data values only whole numbers.
* In the **'Commute' column**, **replace the '10+ miles'** data values with **'More than 10 miles'**.Use the **find and replace function**.This transformation will enable efficient sorting of the data in the column as excel will now recognize the 'More than 10 miles' data values as the greatest in the column.
* Since the **'Age column'** contains **alot of distinct data values**,it is best to create a **new column namely 'Age bracket'** , this will ensure that our visualization later on is not crowded. **Process**: >insert column(right of 'age' column).
* **Group the distinct age data values into ranges**: **'Old' is when x>54,'Middle age' is when x>=31 and 'Adolescent' is when x<31**.**Use the nested IF function to transform the data in 'Age' column i.e =IF(L2>54,"Old",IF(L2>31,"Middle age",IF(L2<31,"Adolescent","Invalid)))**.Write your code on the 'Age bracket' column first cell,then fill the rest of the column cells with the formula.**Thus a new column with only three distict values will have been created thus making visualization easier**.

## **Dataset Analysis**

* Does the level of income of an identifier make a difference in their decision of whether or not to make a bike purchase?
* Does the commute distance of an identifier make a difference in their decision of whether or not to make a bike purchase?
* Which age group is more interested in bike purchase?