**1. Acquired the Raw Data of Bike Buyers**  
  
A screenshot of a computer

Description automatically generated  
  
**2.Data Cleaning:**

i) Removed Duplicates:

A screenshot of a computer

Description automatically generated  
  
ii) Converted Marital Status “M” to “Married” and “S” to “Single”, similarly converted Gender “M” to “Male” and “F” to “Female” for more clarity in Understanding the data

(Note: Used the find and replace tool to convert Marital Status, and Used substitute function to replace Genders to display my knowledge of various excel functionalities)  
A screenshot of a computer

Description automatically generated  
  
For Gender formula: =IF(C2="F",SUBSTITUTE(C2,"F","Female"),SUBSTITUTE(C2,"M","Male"))

A screenshot of a computer

Description automatically generated

iii) Did Feature Engineering To divide age into age brackets (If age<31 then “Adolescent”, If age>=31 then “Middle Age”, If age>55 then “Old”)

Formula: =IF(L2>55,"Old",IF(L2>=31,"Middle Age",IF(L2<55,"Adolescent","Invalid")))

A screenshot of a computer

Description automatically generated

**3.Data Analysis:**

i) Average income of customer who have/not have purchased a bike by gender

A screenshot of a computer

Description automatically generated

ii) Count of purchase based on customer commute distance  
Note: While visualizing this I ran into a scenario where a commute distance value was “10 Miles+” and while visualizing was not getting arranged in ascending order. So I Used the find the replace tool and changed the value to “More than 10 Miles” which enabled me to place it at the last according to order.

A screenshot of a graph

Description automatically generated

iii) Distribution of Bike Purchases by Age Bracket

A screenshot of a computer

Description automatically generated

**4.Dashboard:**  
A screenshot of a computer dashboard

Description automatically generated