1 Import Libraries and Datasets

In [ ]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

2 Exploratory Data Analysis

In [ ]:
```
#Read file into python
```

In [ ]:
```
my_data1=pd.read_excel(r"C:\Users\PRASHANT\Documents\datasets\Lab1data.xlsx")
```

In [ ]:
```
#Shape of dataset
my_data1.shape
```

In [ ]:
```
#Names of columns
my_data1.columns
```

In [ ]:
```
#First 2 rows
my_data1.head(2)
```

Categorical and Continuous

In [ ]:
```
#Types of variables
my_data1.dtypes
```

Univariate analysis of Continuous Variables

In [ ]:
```
#Statistical description of continuous variables
my_data1.describe()
```

Visualization

In [ ]:
```
#Continuous variable distribution
my_data1['Age'].plot.hist(bins=6)
```

In [ ]:
```
#Detection of Outliers
```

In [ ]:
```
my_data1['Age'].plot.box()
```

Univariate analysis of Categorical Variables

In [ ]:
```
#Create frequency table
```

In [ ]:
```
my_data1['Gender'].value_counts()
```

Visualization

In [ ]:
```
#Create a bar plot for Gender frequencies
my_data1['Gender'].value_counts().plot.bar()
```

In [ ]:
```
#Create percentage from frequencies
```

In [ ]:
```
#Create percentages of males and females
my_data1['Gender'].value_counts()/len(my_data1['Gender'])
```

In [ ]:
```
#Create percentages of males and females
(my_data1['Gender'].value_counts()/len(my_data1['Gender'])).plot.bar()
```

Bivariate Analysis: Continuous vs Continuous Variable

Visualization

In [ ]:
```
#Scatter Plot
```

In [ ]:
```
my_data1.plot.scatter("Age","Total")
```

In [ ]:
```
#Check relation after correlation
```

In [ ]:
```
my_data1.corr(numeric_only=True)
```

In [ ]:
```
#Find Correlation between two Continuous Variables
my_data1['Age'].corr(my_data1['Interview'])
```

Bivariate Analysis Catgeorical vs Continuous Variable

Like mean age of males vs mean age of females

In [ ]:
```
#GROUP BY
```

In [ ]:
```
my_data1.groupby("Gender")["Total"].mean()
```

Visualization

In [ ]:
```
my_data1.groupby("Gender")["Total"].mean().plot.bar()
```

Bivariate Analysis Catgeorical vs Categorical Variable

In [ ]:
```
pd.crosstab(my_data1['Gender'],my_data1["Results"])
```