In [ ]:

```
#3. Descriptive Statistics - Measures of Central Tendency and variability
Perform the following operations on any open source dataset (e.g., data.csv)
Provide summary statistics (mean, median, minimum, maximum, standard deviatior
with numeric variables grouped by one of the qualitative (categorical) variabl
variable is age groups and quantitative variable is income, then provide summa
the age groups. Create a list that contains a numeric value for each response
Provide the codes with outputs and explain everything that you do in this step

#HR.csv
```

In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [2]:

```
df=pd.read_csv("HR.csv")
df.head()
```

Out[2]:

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	Educ
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Lit
1	49	No	Travel_Frequently	279	Research & Development	8	1	Lit
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Lit
4	27	No	Travel_Rarely	591	Research & Development	2	1	

5 rows × 35 columns

In [3]:
```
df.describe()
```

Out[3]:

	Age	DailyRate	DistanceFromHome	Education	EmployeeCount	EmployeeNu
count	1470.000000	1470.000000	1470.000000	1470.000000	1470.0	1470.0
mean	36.923810	802.485714	9.192517	2.912925	1.0	1024.8
std	9.135373	403.509100	8.106864	1.024165	0.0	602.0
min	18.000000	102.000000	1.000000	1.000000	1.0	1.0
25%	30.000000	465.000000	2.000000	2.000000	1.0	491.2
50%	36.000000	802.000000	7.000000	3.000000	1.0	1020.5
75%	43.000000	1157.000000	14.000000	4.000000	1.0	1555.7
max	60.000000	1499.000000	29.000000	5.000000	1.0	2068.0

8 rows × 26 columns

In [4]:
```
print(df.columns)
```

```
Index(['Age', 'Attrition', 'BusinessTravel', 'DailyRate', 'Department',
       'DistanceFromHome', 'Education', 'EducationField', 'EmployeeCount',
       'EmployeeNumber', 'EnvironmentSatisfaction', 'Gender', 'HourlyRate',
       'JobInvolvement', 'JobLevel', 'JobRole', 'JobSatisfaction',
       'MaritalStatus', 'MonthlyIncome', 'MonthlyRate', 'NumCompaniesWorke
d',
       'Over18', 'OverTime', 'PercentSalaryHike', 'PerformanceRating',
       'RelationshipSatisfaction', 'StandardHours', 'StockOptionLevel',
       'TotalWorkingYears', 'TrainingTimesLastYear', 'WorkLifeBalance',
       'YearsAtCompany', 'YearsInCurrentRole', 'YearsSinceLastPromotion',
       'YearsWithCurrManager'],
      dtype='object')
```

Mean

In [5]:
```
print("The mean of monthly income is :",df.loc[:,"MonthlyIncome"].mean())
```

```
The mean of monthly income is : 6502.931292517007
```

In [6]:
```
print("The mean of age is : ",df.loc[:,"Age"].mean())
```

```
The mean of age is :  36.923809523809524
```

Median

```
In [7]: print("The median of monthly income is : ",df.loc[:,"MonthlyIncome"]
              .median())
```

```
The median of monthly income is :  4919.0
```

```
In [8]: print("The median of age is ", df.loc[:,"Age"].median())
```

```
The median of age is  36.0
```

Mode

```
In [9]: print("The mode of monthly income is ", df.loc[:,"MonthlyIncome"]
              .mode())
```

```
The mode of monthly income is  0    2342
Name: MonthlyIncome, dtype: int64
```

```
In [10]: print("The mode of Age is ", df.loc[:,"Age"].mode())
```

```
The mode of Age is  0    35
Name: Age, dtype: int64
```

Standard deviation

```
In [11]: print("The standard deviation of monthly income is :",
               df.loc[:,"MonthlyIncome"].std())
```

```
The standard deviation of monthly income is : 4707.956783097995
```

```
In [12]: print("The standard deviation of age is :",df.loc[:,"Age"]
               .std())
```

```
The standard deviation of age is : 9.135373489136734
```

Income and age

```
In [13]: array1 = np.array(df["MonthlyIncome"])
         array2 = np.array(df["Age"])
         print("Income", array1)
         print("Age", array2)
```

```
Income [5993 5130 2090 ... 6142 5390 4404]
Age [41 49 37 ... 27 49 34]
```

Maximum income and age

In [14]:
```
print("Maximum income among the employees is ",max(array1))
print("Minimum income among the employees is",min(array1))
```

```
Maximum income among the employees is  19999
Minimum income among the employees is 1009
```

Minimum income and age

In [15]:
```
print("Maximum age among the employees is ",max(array2))
print("Minimum age among the employees is",min(array2))
```

```
Maximum age among the employees is  60
Minimum age among the employees is 18
```

In [16]:
```
df["BusinessTravel"].replace
({"Travel_Rarely":1, "Travel_Frequently":0}, inplace = True)
df["Attrition"].replace({"Yes":1,"No":0}, inplace = True)
df.head()
```

Out[16]:

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	Educa
0	41	1	1	1102	Sales	1	2	Life
1	49	0	0	279	Research & Development	8	1	Life
2	37	1	1	1373	Research & Development	2	2	
3	33	0	0	1392	Research & Development	3	4	Life
4	27	0	1	591	Research & Development	2	1	

5 rows × 35 columns

In [ ]: