In [3]:
```
import numpy as np
import pandas as pd
df = pd.read_csv("Mall_Customers.csv")
df
```

Out[3]:

	CustomerID	Genre	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40
...	...	...	...	...	...
195	196	Female	35	120	79
196	197	Female	45	126	28
197	198	Male	32	126	74
198	199	Male	32	137	18
199	200	Male	30	137	83

200 rows × 5 columns

Mean

In [4]:
```
df.mean()
```

Out[4]:
```
CustomerID                100.50
Age                        38.85
Annual Income (k$)         60.56
Spending Score (1-100)     50.20
dtype: float64
```

In [5]:
```
df.loc[:,'Age'].mean()
```

Out[5]:
```
38.85
```

In [6]:
```
df.mean(axis=1)[0:4]
```

Out[6]:
```
0    18.50
1    29.75
2    11.25
3    30.00
dtype: float64
```

Median

In [7]:
```
df.median()
```

Out[7]:
```
CustomerID                100.5
Age                        36.0
Annual Income (k$)         61.5
Spending Score (1-100)     50.0
dtype: float64
```

In [8]:
```
df.loc[:,'Age'].median()
```

Out[8]:
```
36.0
```

In [9]:
```
df.median(axis=1)[0:4]
```

Out[9]:
```
0    17.0
1    18.0
2    11.0
3    19.5
dtype: float64
```

Mode

In [10]:
```
df.mode()
```

Out[10]:

	CustomerID	Genre	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Female	32.0	54.0	42.0
1	2	NaN	NaN	78.0	NaN
2	3	NaN	NaN	NaN	NaN
3	4	NaN	NaN	NaN	NaN
4	5	NaN	NaN	NaN	NaN
...	...	...	...	...	...
195	196	NaN	NaN	NaN	NaN
196	197	NaN	NaN	NaN	NaN
197	198	NaN	NaN	NaN	NaN
198	199	NaN	NaN	NaN	NaN
199	200	NaN	NaN	NaN	NaN

200 rows × 5 columns

In [11]:
```
df.loc[:,'Age'].mode()
```

Out[11]:
```
0    32
dtype: int64
```

In [12]:
```
df.min()
```

Out[12]:
```
CustomerID                1
Genre                Female
Age                      18
Annual Income (k$)       15
```

```
Spending Score (1-100)         1
dtype: object
```

In [13]:
```
df.loc[:,'Age'].min(skipna = False)
```

Out[13]:
```
18
```

In [14]:
```
df.max()
```

Out[14]:
```
CustomerID                200
Genre                    Male
Age                        70
Annual Income (k$)        137
Spending Score (1-100)     99
dtype: object
```

In [15]:
```
df.loc[:,'Age'].max(skipna = False)
```

Out[15]:
```
70
```

Standerd Deviation

In [17]:
```
df.std()
```

Out[17]:
```
CustomerID                57.879185
Age                       13.969007
Annual Income (k$)        26.264721
Spending Score (1-100)    25.823522
dtype: float64
```

In [18]:
```
df.loc[:,'Age'].std()
```

Out[18]:
```
13.969007331558883
```

In [19]:
```
df.std(axis=1)[0:4]
```

Out[19]:
```
0    15.695010
1    35.074920
2     8.057088
3    32.300671
dtype: float64
```

GroupBy

In [20]:
```
df.groupby(['Genre'])['Age'].mean()
```

Out[20]:
```
Genre
Female    38.098214
Male      39.806818
Name: Age, dtype: float64
```

In [35]:
```
df_u = df.rename(columns = {'Annual Income (k$)':'Income'}, inplace=False)
(df_u.groupby(['Genre']).Income.mean())
```

Out[35]:
```
Genre
Female    59.250000
Male      62.227273
Name: Income, dtype: float64
```

In [36]:
```
df_u
```

Out[36]:

	CustomerID	Genre	Age	Income	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40
...	...	...	...	...	...
195	196	Female	35	120	79
196	197	Female	45	126	28
197	198	Male	32	126	74
198	199	Male	32	137	18
199	200	Male	30	137	83

200 rows × 5 columns

In [38]:
```
df_u.groupby(['Genre']).Income.mean()
```

Out[38]:
```
Genre
Female    59.250000
Male      62.227273
Name: Income, dtype: float64
```

Sklearn

In [39]:
```
from sklearn import preprocessing
enc = preprocessing.OneHotEncoder()
enc_df = pd.DataFrame(enc.fit_transform(df[['Genre']]).toarray())
enc_df
```

Out[39]:

	0	1
0	0.0	1.0
1	0.0	1.0
2	1.0	0.0
3	1.0	0.0
4	1.0	0.0
...	...	...
195	1.0	0.0

	0	1
196	1.0	0.0
197	0.0	1.0
198	0.0	1.0
199	0.0	1.0

200 rows × 2 columns

In [40]:
```
df_encode = df_u.join(enc_df)
df_encode
```

Out[40]:

	CustomerID	Genre	Age	Income	Spending Score (1-100)	0	1
0	1	Male	19	15	39	0.0	1.0
1	2	Male	21	15	81	0.0	1.0
2	3	Female	20	16	6	1.0	0.0
3	4	Female	23	16	77	1.0	0.0
4	5	Female	31	17	40	1.0	0.0
...	...	...	...	...	...	...	...
195	196	Female	35	120	79	1.0	0.0
196	197	Female	45	126	28	1.0	0.0
197	198	Male	32	126	74	0.0	1.0
198	199	Male	32	137	18	0.0	1.0
199	200	Male	30	137	83	0.0	1.0

200 rows × 7 columns

In [ ]:
```
```