In [59]:
```
greeting = "Assalam-o-Alaikum!"
print(greeting)
```

```
Assalam-o-Alaikum!
```

Import Libraries

In [60]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

Import Dataset

In [61]:
```
df = pd.read_csv("Salary_Data.csv")
df
```

Out[61]:

	Age	Gender	Education Level	Job Title	Years of Experience	Salary
0	32.0	Male	Bachelor's	Software Engineer	5.0	90000.0
1	28.0	Female	Master's	Data Analyst	3.0	65000.0
2	45.0	Male	PhD	Senior Manager	15.0	150000.0
3	36.0	Female	Bachelor's	Sales Associate	7.0	60000.0
4	52.0	Male	Master's	Director	20.0	200000.0
...	...	...	...	...	...	...
6699	49.0	Female	PhD	Director of Marketing	20.0	200000.0
6700	32.0	Male	High School	Sales Associate	3.0	50000.0
6701	30.0	Female	Bachelor's Degree	Financial Manager	4.0	55000.0
6702	46.0	Male	Master's Degree	Marketing Manager	14.0	140000.0
6703	26.0	Female	High School	Sales Executive	1.0	35000.0

6704 rows × 6 columns

In [62]:
```
df["Gender"] = df["Gender"].replace({"Male": 1, "Female": 0, "Other": 2})
df
```

Out[62]:

	Age	Gender	Education Level	Job Title	Years of Experience	Salary
0	32.0	1.0	Bachelor's	Software Engineer	5.0	90000.0
1	28.0	0.0	Master's	Data Analyst	3.0	65000.0
2	45.0	1.0	PhD	Senior Manager	15.0	150000.0
3	36.0	0.0	Bachelor's	Sales Associate	7.0	60000.0
4	52.0	1.0	Master's	Director	20.0	200000.0
...	...	...	...	...	...	...
6699	49.0	0.0	PhD	Director of Marketing	20.0	200000.0
6700	32.0	1.0	High School	Sales Associate	3.0	50000.0
6701	30.0	0.0	Bachelor's Degree	Financial Manager	4.0	55000.0
6702	46.0	1.0	Master's Degree	Marketing Manager	14.0	140000.0
6703	26.0	0.0	High School	Sales Executive	1.0	35000.0

6704 rows × 6 columns

In [63]:
```
df["Education Level"].unique()
```

Out[63]:
```
array(["Bachelor's", "Master's", 'PhD', nan, "Bachelor's Degree",
       "Master's Degree", 'High School', 'phD'], dtype=object)
```

In [64]:
```
df["Education Level"] = df["Education Level"].replace({"Bachelor's": "Bachelor's Degree", "Master's": "Master's
df
```

Out[64]:

	Age	Gender	Education Level	Job Title	Years of Experience	Salary
0	32.0	1.0	Bachelor's Degree	Software Engineer	5.0	90000.0
1	28.0	0.0	Master's Degree	Data Analyst	3.0	65000.0
2	45.0	1.0	PhD	Senior Manager	15.0	150000.0
3	36.0	0.0	Bachelor's Degree	Sales Associate	7.0	60000.0
4	52.0	1.0	Master's Degree	Director	20.0	200000.0
...	...	...	...	...	...	...
6699	49.0	0.0	PhD	Director of Marketing	20.0	200000.0
6700	32.0	1.0	High School	Sales Associate	3.0	50000.0
6701	30.0	0.0	Bachelor's Degree	Financial Manager	4.0	55000.0
6702	46.0	1.0	Master's Degree	Marketing Manager	14.0	140000.0
6703	26.0	0.0	High School	Sales Executive	1.0	35000.0

6704 rows × 6 columns

In [65]:
```
df["Education Level"].unique()
```

Out[65]:
```
array(["Bachelor's Degree", "Master's Degree", 'PhD', nan, 'High School'],
      dtype=object)
```

In [66]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6704 entries, 0 to 6703
Data columns (total 6 columns):
 #   Column               Non-Null Count  Dtype  
---  ------               --------------  -----  
 0   Age                  6702 non-null   float64
 1   Gender               6702 non-null   float64
 2   Education Level      6701 non-null   object 
 3   Job Title            6702 non-null   object 
 4   Years of Experience  6701 non-null   float64
 5   Salary               6699 non-null   float64
dtypes: float64(4), object(2)
memory usage: 314.4+ KB
```

In [67]:
```
df = df.dropna()
```

In [68]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 6698 entries, 0 to 6703
Data columns (total 6 columns):
 #   Column               Non-Null Count  Dtype  
---  ------               --------------  -----  
 0   Age                  6698 non-null   float64
 1   Gender               6698 non-null   float64
 2   Education Level      6698 non-null   object 
 3   Job Title            6698 non-null   object 
 4   Years of Experience  6698 non-null   float64
 5   Salary               6698 non-null   float64
dtypes: float64(4), object(2)
memory usage: 366.3+ KB
```

In [69]:
```
df["Education Level"].unique()
```

Out[69]:
```
array(["Bachelor's Degree", "Master's Degree", 'PhD', 'High School'],
      dtype=object)
```

In [70]:
```
df["Education Level"] = df["Education Level"].replace({"Bachelor's Degree": 1, "Master's Degree": 2, "PhD": 3,
df
```

```
C:\Users\adil.zubair\AppData\Local\Temp\ipykernel_14124\2257547589.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#ret
urning-a-view-versus-a-copy
  df["Education Level"] = df["Education Level"].replace({"Bachelor's Degree": 1, "Master's Degree": 2, "PhD": 3
, "High School": 0})
```

Out[70]:

	Age	Gender	Education Level	Job Title	Years of Experience	Salary
0	32.0	1.0	1	Software Engineer	5.0	90000.0
1	28.0	0.0	2	Data Analyst	3.0	65000.0
2	45.0	1.0	3	Senior Manager	15.0	150000.0
3	36.0	0.0	1	Sales Associate	7.0	60000.0
4	52.0	1.0	2	Director	20.0	200000.0
...	...	...	...	...	...	...
6699	49.0	0.0	3	Director of Marketing	20.0	200000.0
6700	32.0	1.0	0	Sales Associate	3.0	50000.0
6701	30.0	0.0	1	Financial Manager	4.0	55000.0
6702	46.0	1.0	2	Marketing Manager	14.0	140000.0
6703	26.0	0.0	0	Sales Executive	1.0	35000.0

6698 rows × 6 columns

In [71]:
```
removing = df.drop(columns = "Job Title")
removing
```

Out[71]:

	Age	Gender	Education Level	Years of Experience	Salary
0	32.0	1.0	1	5.0	90000.0
1	28.0	0.0	2	3.0	65000.0
2	45.0	1.0	3	15.0	150000.0
3	36.0	0.0	1	7.0	60000.0
4	52.0	1.0	2	20.0	200000.0
...	...	...	...	...	...
6699	49.0	0.0	3	20.0	200000.0
6700	32.0	1.0	0	3.0	50000.0
6701	30.0	0.0	1	4.0	55000.0
6702	46.0	1.0	2	14.0	140000.0
6703	26.0	0.0	0	1.0	35000.0

6698 rows × 5 columns

In [72]:
```
removing.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 6698 entries, 0 to 6703
Data columns (total 5 columns):
 #   Column               Non-Null Count  Dtype  
---  ------               --------------  -----  
 0   Age                  6698 non-null   float64
 1   Gender               6698 non-null   float64
 2   Education Level      6698 non-null   int64  
 3   Years of Experience  6698 non-null   float64
 4   Salary               6698 non-null   float64
dtypes: float64(4), int64(1)
memory usage: 314.0 KB
```

In [73]:
```
x = removing.drop(columns = "Salary")
x
```

Out[73]:

	Age	Gender	Education Level	Years of Experience
0	32.0	1.0	1	5.0
1	28.0	0.0	2	3.0
2	45.0	1.0	3	15.0
3	36.0	0.0	1	7.0
4	52.0	1.0	2	20.0
...	...	...	...	...
6699	49.0	0.0	3	20.0
6700	32.0	1.0	0	3.0
6701	30.0	0.0	1	4.0
6702	46.0	1.0	2	14.0
6703	26.0	0.0	0	1.0

6698 rows × 4 columns

In [74]:
```
y = removing["Salary"]
```

In [75]:
```
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, train_size = 0.3, random_state = 32)
```

In [76]:
```
from sklearn.linear_model import LinearRegression
lr = LinearRegression()
```

In [77]:
```
lr.fit(x_train, y_train)
```

Out[77]:
```
▾ LinearRegression
LinearRegression()
```

In [81]:
```
c = lr.intercept_
c
```

Out[81]:
```
102729.66518930583
```

In [82]:
```
m = lr.coef_
m
```

Out[82]:
```
array([-2572.2952617 ,  8330.74610233, 13172.77334656,  8990.01940846])
```

In [85]:
```
y_pred_test = lr.predict(x_test)
y_pred_test
```

Out[85]:
```
array([ 58289.3164042 , 141780.52010092,  75333.54665082, ...,
       177366.59345167, 124163.83435203,  89442.12856768])
```

In [87]:
```
pd.DataFrame({"Actual": y_test,
              "Predicted": y_pred_test})
```

Out[87]:

	Actual	Predicted
6443	50000.0	58289.316404
4334	125000.0	141780.520101
4564	38000.0	75333.546651
4375	49000.0	78876.412257
1515	135000.0	120594.940731
...	...	...
3600	115000.0	154294.020141
4637	70000.0	89442.128568
2032	195000.0	177366.593452
4940	105000.0	124163.834352
4568	60000.0	89442.128568

4689 rows × 2 columns

In [ ]:
```

```

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js