In [111…
```
import pandas as pd
import numpy as np
```

In [112…
```
data=pd.read_csv('train_data.csv')
```

In [113…
```
data.head(3)
```

Out[113]:

	MMM-YY	Emp_ID	Age	Gender	City	Education_Level	Salary	Dateofjoining	LastWorkingDate
0	2016-01-01	1	28	Male	C23	Master	57387	2015-12-24	NaN
1	2016-02-01	1	28	Male	C23	Master	57387	2015-12-24	NaN
2	2016-03-01	1	28	Male	C23	Master	57387	2015-12-24	2016-03-11

In [114…
```
data.isnull().sum()
```

Out[114]:
```
MMM-YY                     0
Emp_ID                     0
Age                        0
Gender                     0
City                       0
Education_Level            0
Salary                     0
Dateofjoining              0
LastWorkingDate        17488
Joining Designation        0
Designation                0
Total Business Value       0
Quarterly Rating           0
dtype: int64
```

Droping null values and unwanted column for model building

In [116…
```
data.isnull().sum().sum()
```

Out[116]:
```
17488
```

In [117…
```
data.drop('LastWorkingDate',axis=1,inplace=True)
```

In [118…
```
data.isnull().sum().sum()
```

Out[118]:
```
0
```

In [119…
```
from sklearn.preprocessing import LabelEncoder
```

In [120…
```
LE=LabelEncoder()
```

In [121…
```
data['Gender']=LE.fit_transform(data['Gender'])
data['Education_Level']=LE.fit_transform(data['Education_Level'])
```

In [122…
```
data.head()
```

Out[122]:

	MMM-YY	Emp_ID	Age	Gender	City	Education_Level	Salary	Dateofjoining	Joining Designation	Desi
0	2016-01-01	1	28	1	C23	2	57387	2015-12-24	1	
1	2016-02-01	1	28	1	C23	2	57387	2015-12-24	1	
2	2016-03-01	1	28	1	C23	2	57387	2015-12-24	1	
3	2017-11-01	2	31	1	C7	2	67016	2017-11-06	2	
4	2017-12-01	2	31	1	C7	2	67016	2017-11-06	2	

In [124…
```
x=data[['Age','Gender','Education_Level']]
y=data[['Salary']]
```

In [125…
```
from sklearn.model_selection import train_test_split
```

In [126…
```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.30)
```

In [127…
```
from sklearn.linear_model import LinearRegression
```

In [128…
```
model=LinearRegression()
```

In [129…
```
model.fit(x_train,y_train)
```

Out[129]:
```
LinearRegression()
```

In [147…
```
model.predict([[28.0,0.0,2.0]])[0][0].round(0)
```

```
C:\ProgramData\Anaconda3\lib\site-packages\sklearn\base.py:450: UserWarning: X doe
s not have valid feature names, but LinearRegression was fitted with feature names
  warnings.warn(
```

Out[147]:
```
61409.0
```

In [148…
```
y_pred=model.predict(x_test)
```

In [149…
```
pd.DataFrame(np.c_[x_test,y_test,y_pred],columns=["Age","Gender","Education_Level",
```

Out[149]:

	Age	Gender	Education_Level	Original_salary	predicted_salary
0	36.0	0.0	0.0	108997.0	65533.158957
1	32.0	0.0	2.0	43282.0	65375.187575
2	25.0	1.0	0.0	23745.0	53809.076998
3	41.0	0.0	2.0	68830.0	74298.748435
4	26.0	1.0	0.0	39559.0	54800.583760

In [150…
```
from sklearn.metrics import mean_squared_error
```

In [151…
```
model.coef_
```

Out[151]:
```
array([[ 991.50676215, -817.50757574, 1904.0278335 ]])
```

In [154…
```
import joblib
```

In [157…
```
joblib.dump(model,"employee salary prediction.pkl")
```

Out[157]:
```
['employee salary prediction.pkl']
```

In [178…
```
model=joblib.load("employee salary prediction.pkl")
```

In [ ]: