In [2]:
```python
# Import Libraries

import pandas as pd
import numpy as np

from sklearn.linear_model import LinearRegression
import joblib     # for saving the model

# Read the data

data = pd.read_csv("Loan_Repayment_Training_Data.csv")

data.head(25)
```

Out[21]:

	Age	Annual Income (LPA)	Borrowed Amount (LPA)	Credit Score	Loan Tenure (months)	Employment Status	Existing Loans	Marital Status	Education Level	Repayment Probability
0	59	9.44	4.22	787	89	Self-Employed	1	Married	Graduate	0.179
1	49	1.00	1.18	656	163	Self-Employed	0	Married	Graduate	0.199
2	35	4.48	4.38	561	67	Salaried	0	Married	Graduate	0.111
3	63	15.81	1.98	658	61	Salaried	1	Married	High School	0.262
4	28	25.73	24.49	741	112	Salaried	1	Married	Post-Graduate	0.043
5	41	11.22	8.84	759	14	Self-Employed	3	Married	Graduate	0.082
6	59	14.62	5.47	835	34	Salaried	1	Married	Graduate	0.246
7	39	19.62	29.31	590	173	Self-Employed	0	Married	Graduate	0.010
8	43	5.51	7.47	656	20	Self-Employed	1	Married	Graduate	0.190
9	31	14.93	7.61	582	93	Self-Employed	3	Single	Post-Graduate	0.146
10	31	6.08	5.21	566	70	Salaried	1	Single	Graduate	0.184
11	44	18.92	10.28	637	77	Self-Employed	1	Married	Graduate	0.152
12	56	13.21	15.66	736	161	Salaried	2	Married	Graduate	0.133
13	60	14.39	20.18	777	126	Salaried	2	Single	Graduate	0.115
14	44	5.74	3.46	605	46	Salaried	1	Married	Post-Graduate	0.168
15	23	17.66	21.68	619	170	Self-Employed	0	Single	Graduate	0.093
16	42	28.68	22.88	522	229	Salaried	0	Married	Post-Graduate	0.025
17	22	3.94	2.58	721	89	Self-Employed	1	Single	High School	0.191
18	44	9.28	13.12	626	120	Self-Employed	3	Single	Post-Graduate	0.018
19	64	13.85	11.28	866	59	Self-Employed	1	Married	Graduate	0.181
20	50	18.29	13.80	740	64	Salaried	1	Married	High School	0.213
21	58	7.52	7.83	682	231	Unemployed	2	Married	Graduate	0.113

	Age	Annual Income (LPA)	Borrowed Amount (LPA)	Credit Score	Loan Tenure (months)	Employment Status	Existing Loans	Marital Status	Education Level	Repayment Probability
22	22	1.00	1.00	635	86	Salaried	0	Single	Graduate	0.110
23	41	22.87	12.69	550	43	Self-Employed	1	Married	Graduate	0.160
24	53	18.08	25.40	563	224	Salaried	3	Married	Post-Graduate	0.000

In [2]:
```python
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 10 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   Age                    10000 non-null  int64
 1   Annual Income (LPA)    10000 non-null  float64
 2   Borrowed Amount (LPA)  10000 non-null  float64
 3   Credit Score           10000 non-null  int64
 4   Loan Tenure (months)   10000 non-null  int64
 5   Employment Status      10000 non-null  object
 6   Existing Loans         10000 non-null  int64
 7   Marital Status         10000 non-null  object
 8   Education Level        10000 non-null  object
 9   Repayment Probability  10000 non-null  float64
dtypes: float64(3), int64(4), object(3)
memory usage: 781.4+ KB
```

In [4]:
```python
# Define independent(X) & Dependent(y) Variables

X = data[['Age','Annual Income (LPA)','Borrowed Amount (LPA)','Credit Score','Loan Tenure (months)','Employ
y = data['Repayment Probability']

# Converting non-numeric columns into numeric columns

categorical_col = ["Employment Status",'Marital Status','Education Level']


X = pd.get_dummies(X, columns =categorical_col)

X.head(20)
```

Out[41]:

	Age	Annual Income (LPA)	Borrowed Amount (LPA)	Credit Score	Loan Tenure (months)	Existing Loans	Employment Status_Salaried	Employment Status_Self-Employed	Employment Status_Unemployed	Marital Status_Married	Statu
0	59	9.44	4.22	787	89	1	False	True	False	True	
1	49	1.00	1.18	656	163	0	False	True	False	True	
2	35	4.48	4.38	561	67	0	True	False	False	True	
3	63	15.81	1.98	658	61	1	True	False	False	True	
4	28	25.73	24.49	741	112	1	True	False	False	True	
5	41	11.22	8.84	759	14	3	False	True	False	True	
6	59	14.62	5.47	835	34	1	True	False	False	True	
7	39	19.62	29.31	590	173	0	False	True	False	True	
8	43	5.51	7.47	656	20	1	False	True	False	True	
9	31	14.93	7.61	582	93	3	False	True	False	False	
10	31	6.08	5.21	566	70	1	True	False	False	False	
11	44	18.92	10.28	637	77	1	False	True	False	True	
12	56	13.21	15.66	736	161	2	True	False	False	True	
13	60	14.39	20.18	777	126	2	True	False	False	False	
14	44	5.74	3.46	605	46	1	True	False	False	True	
15	23	17.66	21.68	619	170	0	False	True	False	False	
16	42	28.68	22.88	522	229	0	True	False	False	True	
17	22	3.94	2.58	721	89	1	False	True	False	False	
18	44	9.28	13.12	626	120	3	False	True	False	False	
19	64	13.85	11.28	866	59	1	False	True	False	True	

In [5]:
```python
# Initiaslise the train the linear regression model

model = LinearRegression()

model.fit(X,y)

# Save the trained model

joblib.dump(model,"Repayment_Model.pkl")

print("The model has been saved")
```

The model has been saved

In [9]:
```python
# Import the test dataset

new_data = pd.read_csv("Loan_Repayment_Test_Data_NoLabel.csv")


# Converting the text columns into numerical

text_columns = ["Employment Status",'Marital Status','Education Level']

new_data_enc = pd.get_dummies(new_data, columns =text_columns)


#Load the trained model

career247 = joblib.load("Repayment_Model.pkl")


# Make the prediction using the model

new_data["Prediction %"] = ((career247.predict(new_data_enc))*100).round(2)

new_data.head(20)
```

Out[9]:

	Age	Annual Income (LPA)	Borrowed Amount (LPA)	Credit Score	Loan Tenure (months)	Employment Status	Existing Loans	Marital Status	Education Level	Prediction %
0	35	5.56	7.84	647	221	Salaried	2	Married	Graduate	9.02
1	47	16.97	14.23	856	39	Self-Employed	2	Single	Graduate	23.46
2	25	8.77	5.36	596	30	Salaried	1	Married	Post-Graduate	13.49
3	22	9.64	10.40	618	128	Salaried	2	Single	Post-Graduate	9.66
4	58	13.50	14.40	765	239	Salaried	0	Married	Graduate	15.73
5	51	27.69	34.41	731	120	Salaried	2	Married	Graduate	9.42
6	62	18.68	26.27	713	205	Salaried	1	Married	Graduate	8.79
7	60	6.63	3.84	727	231	Salaried	2	Married	Post-Graduate	15.34
8	50	1.97	1.00	642	50	Salaried	1	Married	Post-Graduate	15.22
9	65	14.59	19.96	593	26	Self-Employed	4	Married	Graduate	5.26
10	59	7.31	2.61	856	118	Salaried	1	Married	Graduate	25.30
11	39	10.16	6.05	586	49	Salaried	2	Married	Graduate	11.60
12	24	10.53	7.27	501	137	Self-Employed	0	Single	Graduate	7.11
13	47	11.11	14.48	627	191	Salaried	1	Married	Graduate	8.06
14	29	10.28	9.44	780	220	Salaried	1	Single	Post-Graduate	17.56
15	49	8.17	7.66	498	240	Self-Employed	0	Single	Graduate	4.03
16	27	1.00	1.18	615	225	Unemployed	1	Married	Graduate	10.18
17	62	13.49	20.20	790	90	Unemployed	0	Single	Post-Graduate	17.27
18	25	16.95	24.58	573	52	Self-Employed	0	Single	Post-Graduate	6.17
19	51	4.46	1.26	718	87	Unemployed	0	Married	Graduate	19.82

```python
new_data.to_excel("Samraat_Prediction.xlsx", index="False")
```