```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

```python
from google.colab import files
uploaded = files.upload()
df = pd.read_csv('customers.csv')
df.info()
```

Choose Files customers.csv
- **customers.csv**(text/csv) - 87360 bytes, last modified: 7/11/2022 - 100% done
Saving customers.csv to customers.csv

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 500 entries, 0 to 499
Data columns (total 8 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   Email                 500 non-null    object
 1   Address               500 non-null    object
 2   Avatar                500 non-null    object
 3   Avg. Session Length   500 non-null    float64
 4   Time on App           500 non-null    float64
 5   Time on Website       500 non-null    float64
 6   Length of Membership  500 non-null    float64
 7   Yearly Amount Spent   500 non-null    float64
dtypes: float64(5), object(3)
memory usage: 31.4+ KB
```

```python
df.head()
```

	Email	Address	Avatar	Avg. Session Length	Time on App
0	mstephenson@fernandez.com	835 Frank Tunnel\nWrightmouth, MI 82180-9605	Violet	34.497268	12.65565
1	hduke@hotmail.com	4547 Archer Common\nDiazchester, CA 06566-8576	DarkGreen	31.926272	11.10946
2	pallen@yahoo.com	24645 Valerie Unions Suite 582\nCobbborough, D...	Bisque	33.000915	11.33027
		1414 David			

```
df.corr()
```

	Avg. Session Length	Time on App	Time on Website	Length of Membership	Yearly Amount Spent
Avg. Session Length	1.000000	-0.027826	-0.034987	0.060247	0.355088
Time on App	-0.027826	1.000000	0.082388	0.029143	0.499328
Time on Website	-0.034987	0.082388	1.000000	-0.047582	-0.002641
Length of Membership	0.060247	0.029143	-0.047582	1.000000	0.809084

```
X = df.select_dtypes(exclude = 'object').drop('Yearly Amount Spent', axis=1)
y = df['Yearly Amount Spent']


from sklearn.model_selection import train_test_split


X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=2)


from sklearn.linear_model import LinearRegression
lm = LinearRegression()
lm.fit(X_train, y_train)

    LinearRegression()


lm.coef_

    array([25.73355935, 38.74177284,  0.74764502, 61.43626237])


pred = lm.predict(X_test)


from sklearn import metrics
print('MAE', metrics.mean_absolute_error(y_test, pred))
print('MSE', metrics.mean_squared_error(y_test, pred))
print('RMSE', np.sqrt(metrics.mean_squared_error(y_test, pred)))
print('r2Score', metrics.r2_score(y_test, pred))

    MAE 7.546199505771423
    MSE 91.0436844294419
    RMSE 9.54168142569442
    r2Score 0.9859800372444311


df_coef = pd.DataFrame(lm.coef_, X_train.columns, columns = ['Coefficient'])


df_coef
```

	Coefficient
Avg. Session Length	25.733559
Time on App	38.741773
Time on Website	0.747645
Length of Membership	61.436262

✓ 0s   completed at 19:16   ● ✕