

```
In [1]: import pandas as pd

In [2]: df = pd.read_csv("Ecommerce Customers-1.csv")

In [3]: df.head()

Out[3]:
```

	Email	Address	Avatar	Avg. Session Length	Time on App	Time on Website	Length of Membership	Yearly Amount Spent
0	mstephenson@fernandez.com	835 Frank Tunnel\nWrightmouth, MI 82180-9605	Violet	34.497268	12.655651	39.577668	4.082621	587.951
1	hduke@hotmail.com	4547 Archer Common\nDiazchester, CA 06566-8576	DarkGreen	31.926272	11.109461	37.268959	2.664034	392.204
2	pallen@yahoo.com	24645 Valerie Unions Suite 582\nCobborough, D...	Bisque	33.000915	11.330278	37.110597	4.104543	487.547
3	riverarebecca@gmail.com	1414 David Thoroughway\nPort Jason, OH 22070-1220	SaddleBrown	34.305557	13.717514	36.721283	3.120179	581.852
4	mstephens@davidson-herman.com	14023 Rodriguez Passage\nPort Jacobville, PR 3...	MediumAquaMarine	33.330673	12.795189	37.536653	4.446308	599.406

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

Out[4]:
```

	Avg. Session Length	Time on App	Time on Website	Length of Membership	Yearly Amount Spent
count	500.000000	500.000000	500.000000	500.000000	500.000000
mean	33.053194	12.052488	37.060445	3.533462	499.314038
std	0.992563	0.994216	1.010489	0.999278	79.314782
min	29.532429	8.508152	33.913847	0.269901	256.670582
25%	32.341022	11.388153	36.349257	2.930450	445.038277
50%	33.082008	11.983231	37.069367	3.533975	498.887875
75%	33.711985	12.753850	37.716432	4.126502	549.313828
max	36.139662	15.126994	40.005182	6.922689	765.518462

```
In [5]: df.info()

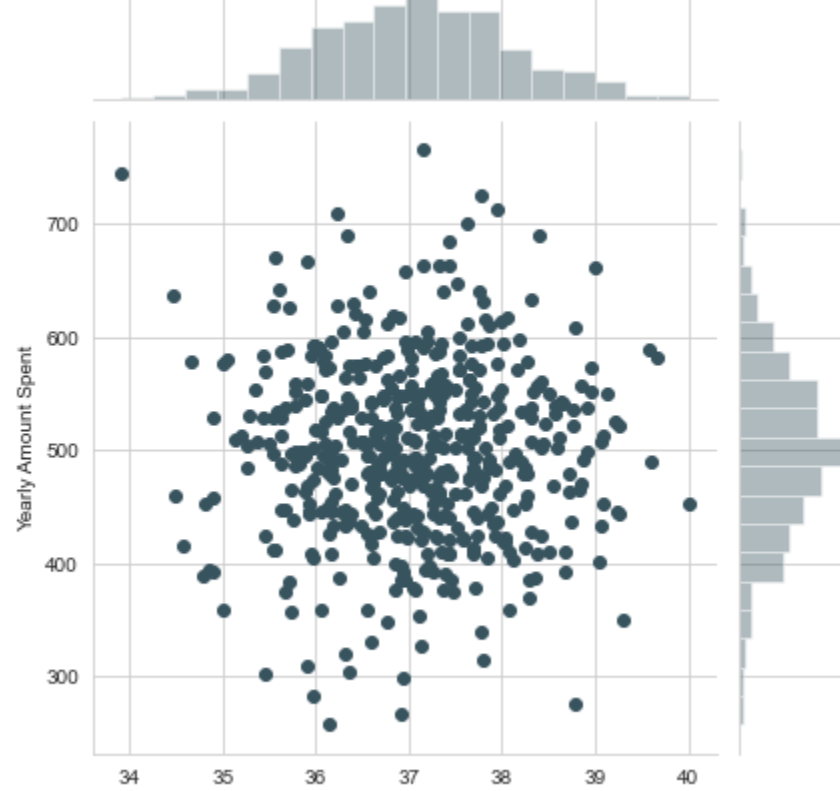
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 500 entries, 0 to 499
Data columns (total 8 columns):
#   Column              Non-Null Count  Dtype
---  --
0   Email                500 non-null    object
1   Address              500 non-null    object
2   Avatar               500 non-null    object
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

In [6]: import seaborn as sns

In [7]: sns.set_palette("GnBu_d")
sns.set_style('whitegrid')

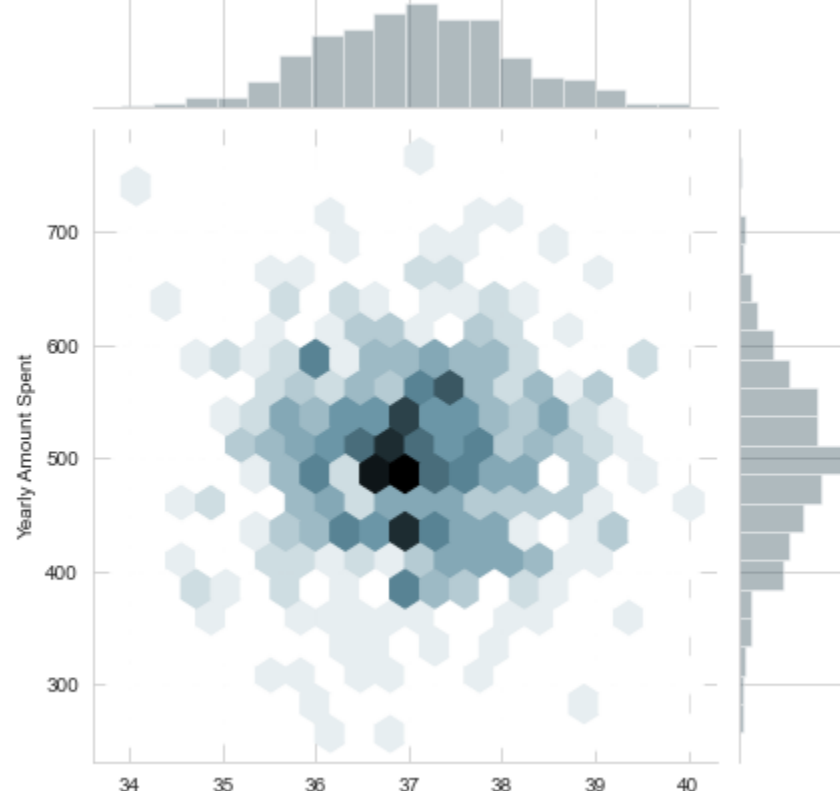
In [8]: # More time on site, more money spent.
sns.jointplot(x='Time on Website',y='Yearly Amount Spent',data=df)

Out[8]: <seaborn.axisgrid.JointGrid at 0x240f39d4a60>
```



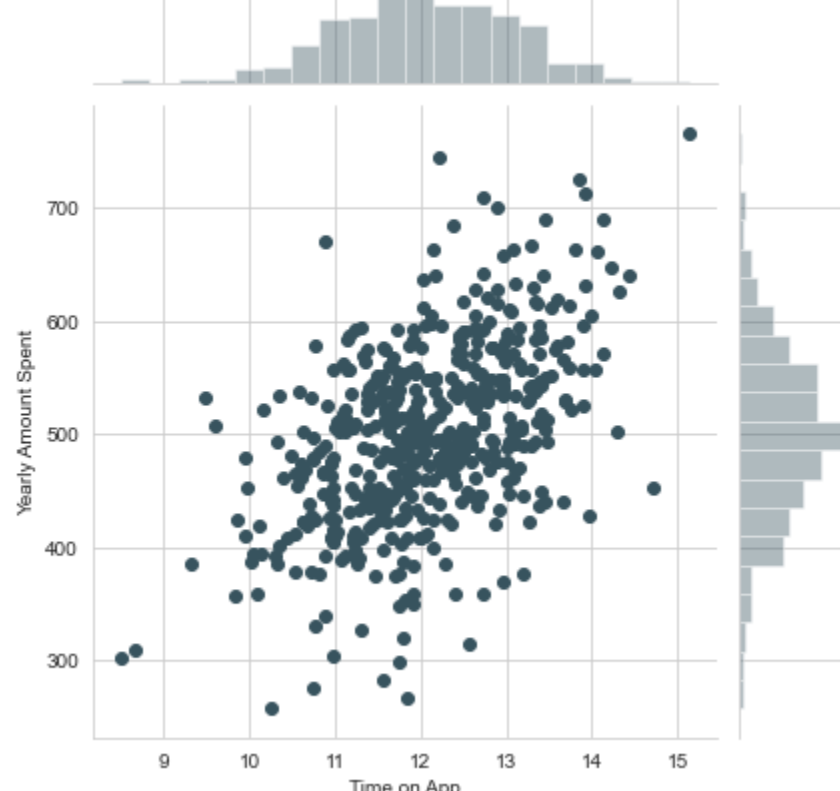
```
In [67]: # More time on site, more money spent.
sns.jointplot(x='Time on Website',y='Yearly Amount Spent',kind='hex',data=df)

Out[67]: <seaborn.axisgrid.JointGrid at 0x1d4e0e92c40>
```



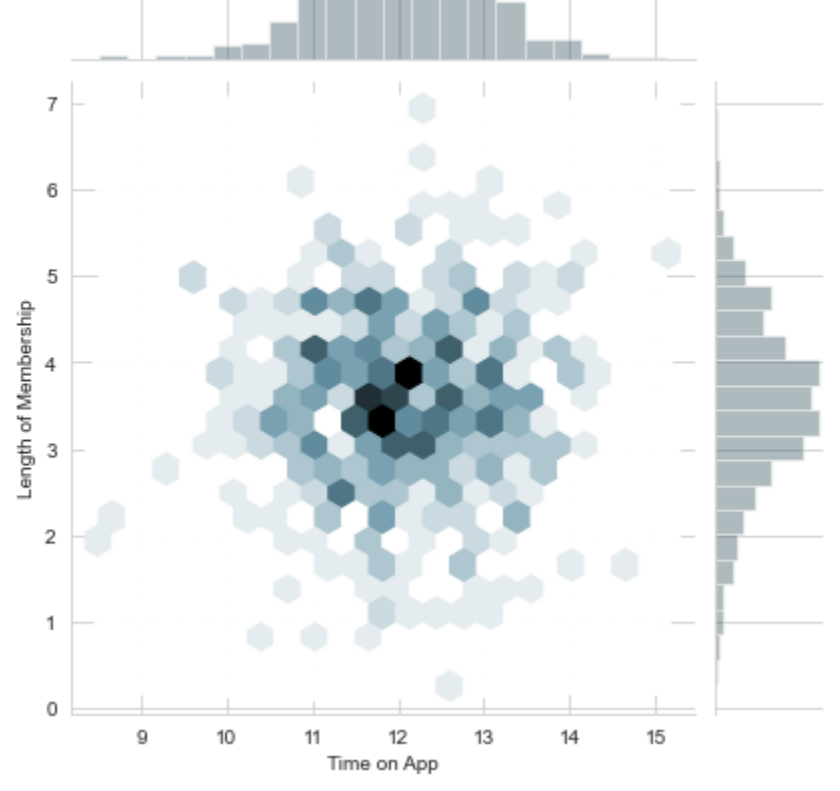
```
In [68]: sns.jointplot(x='Time on App',y='Yearly Amount Spent',data=df)

Out[68]: <seaborn.axisgrid.JointGrid at 0x1d4e0e98b20>
```



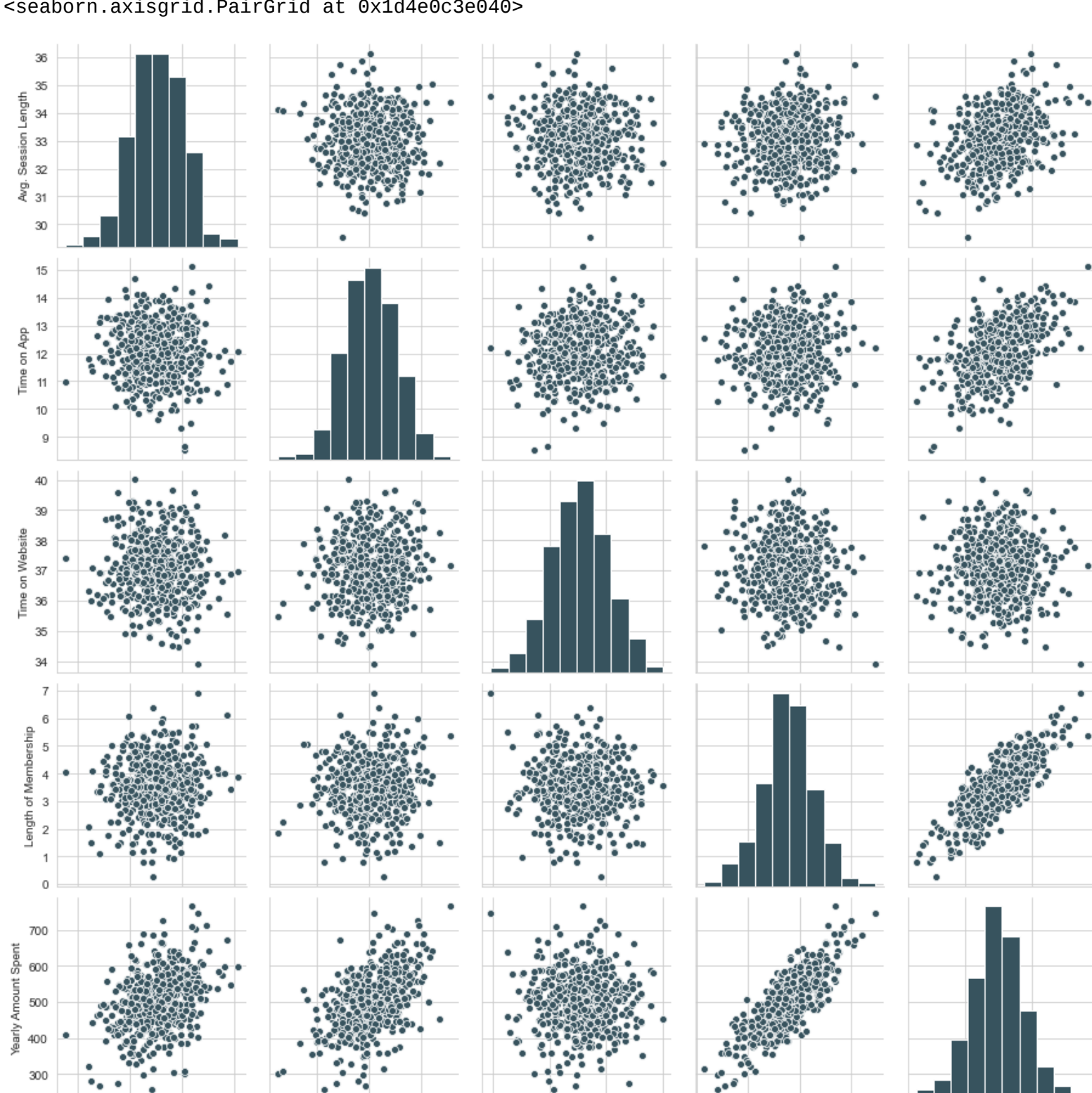
```
In [69]: sns.jointplot(x='Time on App',y='Length of Membership',kind='hex',data=df)

Out[69]: <seaborn.axisgrid.JointGrid at 0x1d4e0f60a30>
```



```
In [70]: sns.pairplot(df)

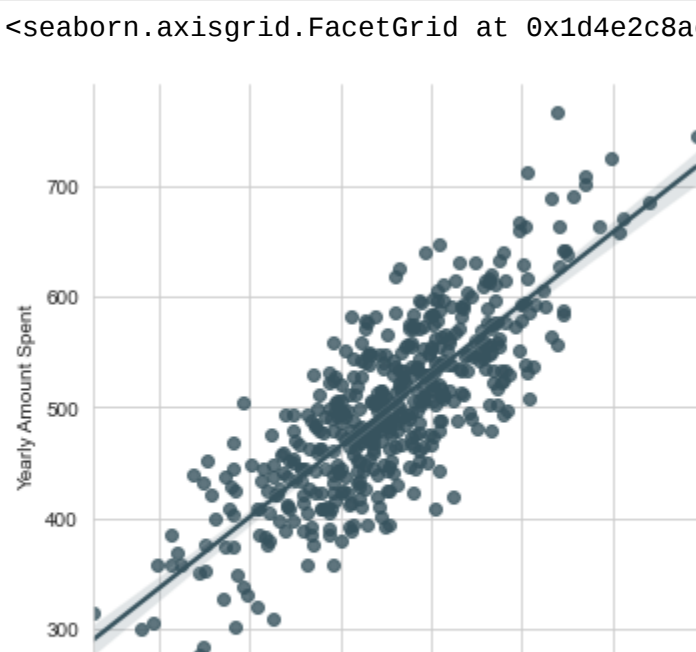
Out[70]: <seaborn.axisgrid.PairGrid at 0x1d4e0c3e040>
```



```
In [71]: # Length of Membership

In [72]: sns.lmplot(x='Length of Membership',y='Yearly Amount Spent',data=df)

Out[72]: <seaborn.axisgrid.FacetGrid at 0x1d4e2c8ae0>
```



```
In [73]: X = df.iloc[:,3:6]

In [74]: y = df['Yearly Amount Spent']

In [75]: from sklearn.model_selection import train_test_split

In [76]: X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.4, random_state=101)

In [77]: from sklearn.linear_model import LinearRegression

In [78]: lm = LinearRegression()

In [79]: lm.fit(X_train,y_train)

Out[79]: LinearRegression()

In [80]: # The coefficients
print('Coefficients: \n', lm.coef_)

Coefficients:
[25.69154034 37.89259966  0.56058149 61.64859402]

In [81]: predictions = lm.predict( X_test)

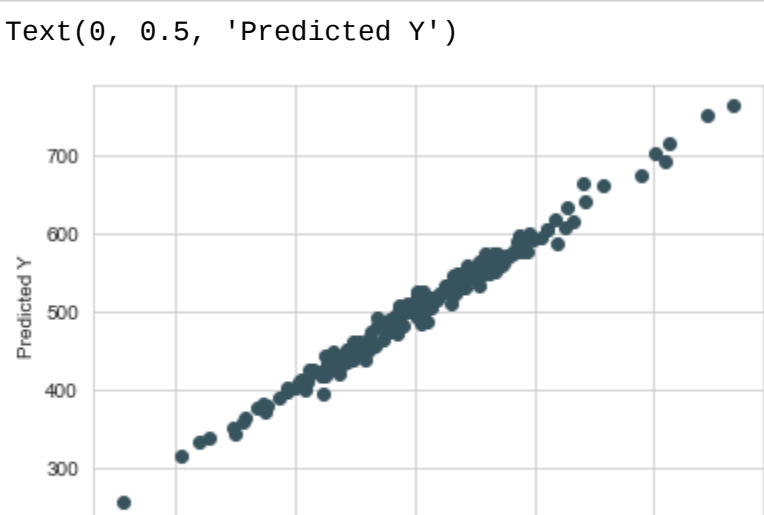
In [82]: from matplotlib import pyplot as plt

In [83]: import numpy as np
predictions = lm.predict( X_test)
np.savetxt("Prediction_LinearRegression_Ecommerce.csv",predictions)
lm.score(X_test,y_test)

Out[83]: 0.9855061240024658

In [84]: plt.scatter(y_test,predictions)
plt.xlabel('Y Test')
plt.ylabel('Predicted Y')

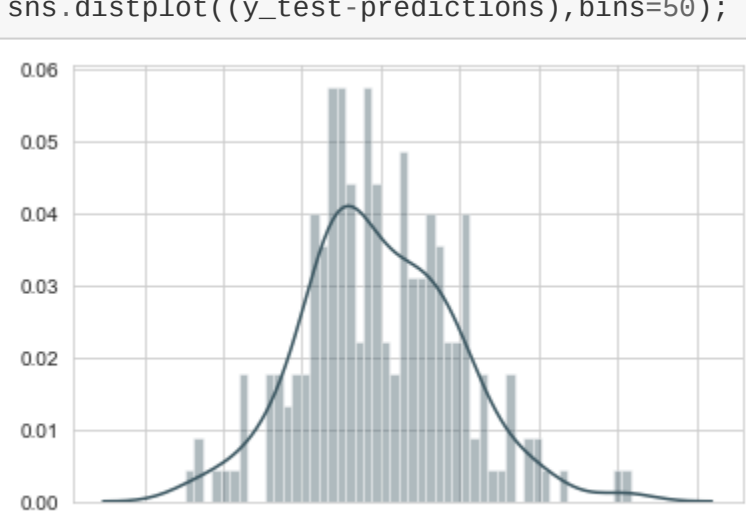
Out[84]: Text(0, 0.5, 'Predicted Y')
```



```
from sklearn import metrics
import numpy as np

print('MAE:', metrics.mean_absolute_error(y_test, predictions))
print('MSE:', metrics.mean_squared_error(y_test, predictions))
print('RMSE:', np.sqrt(metrics.mean_squared_error(y_test, predictions)))

In [85]: sns.distplot((y_test-predictions),bins=50);
```



```
In [86]: coefficients = pd.DataFrame(lm.coef_,X.columns)
coefficients.columns = ['Coefficient']

Out[86]:
```

	Coefficient
Avg. Session Length	25.691540
Time on App	37.892600
Time on Website	0.560581
Length of Membership	61.648594

```
In [87]: lm.score(X_test,y_test)

Out[87]: 0.9855061240024658

In [ ]:
```