

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

customers=pd.read_csv('Ecommerce Customers.csv')

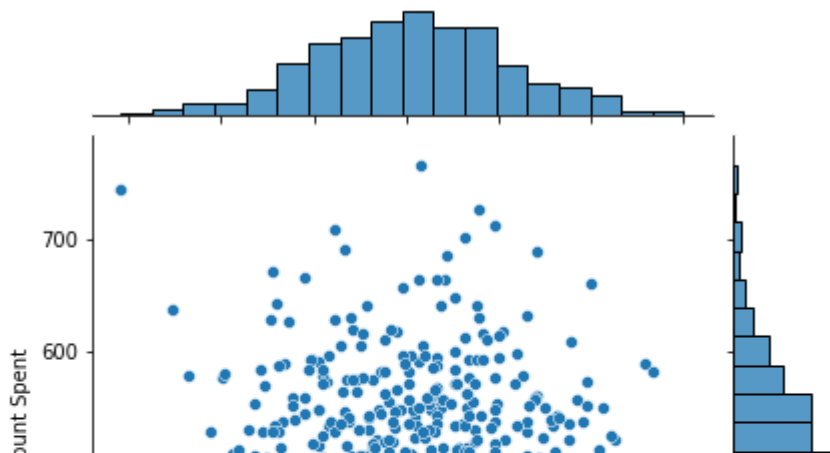
customers.head()
```

| | Email | Address | Avatar | Avg. Session Length | Time on Website |
|---|-------------------------------|--|------------------|---------------------------|-----------------|
| 0 | mstephenson@fernandez.com | 835 Frank Tunnel\nWrightmouth, MI 82180-9605 | Violet | 34.497268 | 12.65561 |
| 1 | hduke@hotmail.com | 4547 Archer Common\nDiazchester, CA 06566-8576 | DarkGreen | 31.926272 | 11.10941 |
| 2 | pallen@yahoo.com | 24645 Valerie Unions Suite 582\nCobbborough, D... | Bisque | 33.000915 | 11.33021 |
| 3 | riverarebecca@gmail.com | 1414 David Throughway\nPort Jason, OH 22070-1220 | SaddleBrown | 34.305557 | 13.71751 |
| 4 | mstephens@davidson-herman.com | 14023 Rodriguez Passage\nPort Jacobville, PR 3... | MediumAquaMarine | 33.330673 | 12.79511 |



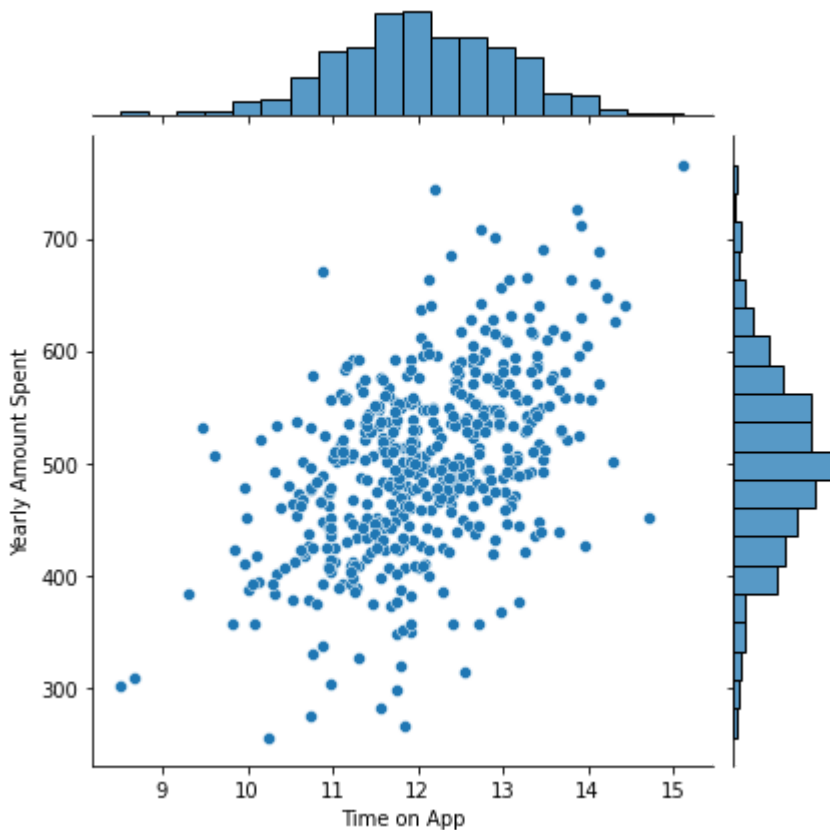
```
sns.jointplot(data=customers,x="Time on Website",y="Yearly Amount Spent")
```

<seaborn.axisgrid.JointGrid at 0x7f507a54f690>



```
sns.jointplot(data=customers,x="Time on App",y="Yearly Amount Spent")
```

<seaborn.axisgrid.JointGrid at 0x7f50703ca0d0>



```
y_=customers["Yearly Amount Spent"]
x_=customers[["Length of Membership"]]
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=101)
```

```
plt.scatter(x_train, y_train, color='red') # plotting the observation line
```

```
plt.plot(x_train, lm.predict(x_train), color='blue') # plotting the regression line
```

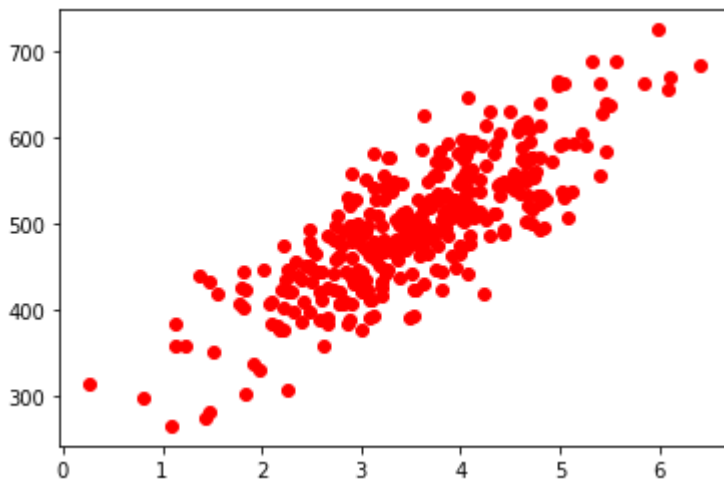
```
plt.title("linera model plot") # stating the title of the graph
```

```
plt.xlabel("Length of Membership") # adding the name of x-axis
plt.ylabel("Yearly Amount Spent") # adding the name of y-axis
plt.show() # specifies end of graph
```

```
-----
NameError                                Traceback (most recent call last)
<ipython-input-7-26a62265acd9> in <module>()
      1 plt.scatter(x_train, y_train, color='red') # plotting the observation line
      2
----> 3 plt.plot(x_train, lm.predict(x_train), color='blue') # plotting the regression
line
      4
      5 plt.title("linera model plot") # stating the title of the graph
```

NameError: name 'lm' is not defined

SEARCH STACK OVERFLOW



```
y=customers["Yearly Amount Spent"]
x=customers[["Avg. Session Length", "Time on App", "Time on Website", "Length of Membership"]]
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=101)
```

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

```
predictions=lm.predict(x_test)
```

```
from sklearn import metrics
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)))
```

```
cdf=pd.DataFrame(lm.coef_,x.columns,columns=["Coeff"])
```

cdf

```
plt.scatter(predictions,y_test)
```

```
sns.distplot(y_test-predictions,bins=50)
```

```
import numpy as np
from scipy.stats import entropy
from math import log, e
import pandas as pd
```

```
import timeit
```

```
def entropy1(labels, base=None):
    value,counts = np.unique(labels, return_counts=True)
    return entropy(counts, base=base)
```

```
labels = [0,1,1,1,1,1]
```

```
print(entropy1(labels))
```

