Ecommerce Purchases Exercise

This Exercise is gotten from Pierian Data and it has Fake Data about some purchases done through Amazon!

In [7]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
```

Check the head of the DataFrame.

In [9]:
```
ecom=pd.read_csv('Ecommerce Purchases.csv')
```

In [11]:
```
ecom.head()
```

Out[11]:

	Address	Lot	AM or PM	Browser Info	Company	Credit Card	CC Exp Date	CC Security Code	CC Provider	
0	16629 Pace Camp Apt. 448\nAlexisborough, NE 77...	46 in	PM	Opera/9.56. (X11; Linux x86_64; sl-SI) Presto/2...	Martinez-Herman	6011929061123406	02/20	900	JCB 16 digit	
1	9374 Jasmine Spurs Suite 508\nSouth John, TN 8...	28 rn	PM	Opera/8.93. (Windows 98; Win 9x 4.90; en-US) Pr...	Fletcher, Richards and Whitaker	3337758169645356	11/18	561	Mastercard	
2	Unit 0065 Box 5052\nDPO AP 27450	94 vE	PM	Mozilla/5.0 (compatible; MSIE 9.0; Windows NT ...	Simpson, Williams and Pham	675957666125	08/19	699	JCB 16 digit	
3	7780 Julia Fords\nNew Stacy, WA 45798	36 vm	PM	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_8_0 ...	Williams, Marshall and Buchanan	6011578504430710	02/24	384	Discover	bren
4	23012 Munoz Drive Suite 337\nNew Cynthia, TX 5...	20 IE	AM	Opera/9.58. (X11; Linux x86_64; it-IT) Presto/2...	Brown, Watson and Andrews	6011456623207998	10/25	678	Diners Club / Carte Blanche	christ

How many rows and columns are there?

In [14]:
```
ecom.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 14 columns):
 #   Column            Non-Null Count  Dtype 
---  ------            --------------  ----- 
 0   Address           10000 non-null  object
 1   Lot               10000 non-null  object
```

```
 2   AM or PM          10000 non-null  object
 3   Browser Info      10000 non-null  object
 4   Company           10000 non-null  object
 5   Credit Card       10000 non-null  int64
 6   CC Exp Date       10000 non-null  object
 7   CC Security Code  10000 non-null  int64
 8   CC Provider       10000 non-null  object
 9   Email             10000 non-null  object
 10  Job               10000 non-null  object
 11  IP Address        10000 non-null  object
 12  Language          10000 non-null  object
 13  Purchase Price    10000 non-null  float64
dtypes: float64(1), int64(2), object(11)
memory usage: 1.1+ MB
```

What is the average Purchase Price?

In [16]:
```
ecom['Purchase Price'].mean ()
```

Out[16]:
```
50.34730200000025
```

What were the highest and lowest purchase prices?

In [17]:
```
ecom['Purchase Price'].max()
```

Out[17]:
```
99.99
```

In [18]:
```
ecom['Purchase Price'].min()
```

Out[18]:
```
0.0
```

How many people have English 'en' as their Language of choice on the website?

In [30]:
```
ecom[ecom['Language']=='en']['Language'].count()
```

Out[30]:
```
1098
```

How many people have the job title of "Lawyer" ?

In [58]:
```
ecom[ecom['Job']=='Lawyer'].count()
```

Out[58]:
```
Address             30
Lot                 30
AM or PM            30
Browser Info        30
Company             30
Credit Card         30
CC Exp Date         30
CC Security Code    30
CC Provider         30
Email               30
Job                 30
IP Address          30
Language            30
Purchase Price      30
dtype: int64
```

How many people made the purchase during the AM and how many people made the purchase during PM ?

In [54]:
```
ecom['AM or PM'].value_counts()
```

Out[54]:
```
PM    5068
AM    4932
```

```
Name: AM or PM, dtype: int64
```

What are the 5 most common Job Titles?

In [66]:
```
ecom['Job'].value_counts().head(5)
```

Out[66]:
```
Interior and spatial designer        31
Lawyer                               30
Social researcher                    28
Purchasing manager                   27
Research officer, political party    27
Name: Job, dtype: int64
```

Someone made a purchase that came from Lot: "90 WT" , what was the Purchase Price for this transaction?

In [72]:
```
ecom[ecom['Lot']== '90 WT']['Purchase Price']
```

Out[72]:
```
513    75.1
Name: Purchase Price, dtype: float64
```

What is the email of the person with the following Credit Card Number: 4926535242672853

In [79]:
```
ecom[ecom['Credit Card']==4926535242672853]['Email']
```

Out[79]:
```
1234    bondellen@williams-garza.com
Name: Email, dtype: object
```

How many people have American Express as their Credit Card Provider *and* made a purchase above $95 ?

In [88]:
```
ecom[(ecom['CC Provider']=='American Express') & (ecom['Purchase Price']> 95)].count
```

Out[88]:
```
Address             39
Lot                 39
AM or PM            39
Browser Info        39
Company             39
Credit Card         39
CC Exp Date         39
CC Security Code    39
CC Provider         39
Email               39
Job                 39
IP Address          39
Language            39
Purchase Price      39
dtype: int64
```

How many people have a credit card that expires in 2025?

In [95]:
```
#sum(ecom['CC Exp Date'].apply(lambda exp: exp[3:]== '25'))
            # or
ecom[ecom['CC Exp Date'].apply(lambda exp: exp[3:]== '25')].count()
```

Out[95]:
```
Address             1033
Lot                 1033
AM or PM            1033
Browser Info        1033
Company             1033
Credit Card         1033
CC Exp Date         1033
CC Security Code    1033
CC Provider         1033
Email               1033
Job                 1033
IP Address          1033
```

```
Language            1033
Purchase Price      1033
dtype: int64
```

What are the top 5 most popular email providers/hosts (e.g. gmail.com, yahoo.com, etc...)

In [119…

```
ecom['Email'].apply(lambda email: email.split('@')[1]).value_counts().head(5)
```

Out[119…

```
hotmail.com     1638
yahoo.com       1616
gmail.com       1605
smith.com         42
williams.com      37
Name: Email, dtype: int64
```