Ecommerce Purchases Exercise

In this Exercise you will be given some data about some purchases done through Amazon! Just go ahead and follow the directions and answer the questions and complete the tasks.

Also note that all of these questions can be answered with one line of code. ____ ** Import pandas and read in the Ecommerce Purchases csv file and set it to a DataFrame called ecom. **

In [1]:

```
#import the necessary libraries which are numpy and pandas
import pandas as pd
import numpy as np
```

In [2]:

```
#read the data from file with name "Ecommerce Purchases" in csv format using the pandas library to print data i
ecom = pd.read_csv("Ecommerce Purchases")
```

Check the head of the DataFrame.

In [3]:

```
#Reading the first 5 lines of dataframe
ecom.head()
```

Out[3]:

	Address	Lot	AM or PM	Browser Info	Company	Credit Card	CC Exp Date	CC Security Code	CC Provider	Email
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	pdunlap@yahoo.com
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	anthony41@reed.com
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	amymiller@morales-harrison.com
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	brent16@olson-robinson.info
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	christopherwright@gmail.com

** How many rows and columns are there? **

In [4]:

```
#printing the general information of datafrome by the below command
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 [5]:
```
#Printing the Average Purchase Price which is equal to Mean of Purchase Price Values
ecom['Purchase Price'].mean()
```

Out[5]:
```
50.34730200000025
```

** What were the highest and lowest purchase prices? **

In [6]:
```
#Printing the lowest purchase price value available in series "Purchase Price"
ecom['Purchase Price'].min()
```

Out[6]:
```
0.0
```

In [7]:
```
#Printing the highest purchase price value available in series "Purchase Price"
ecom['Purchase Price'].max()
```

Out[7]:
```
99.99
```

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

In [8]:
```
#Let's check the count of people who have registered English with code "en" as their Language.
ecom[ecom.Language == 'en'].shape[0]
```

Out[8]:
```
1098
```

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

In [9]:
```
#Let's check number of people who are Lawyer by their Profession
ecom[ecom.Job == 'Lawyer'].shape[0]
```

Out[9]:
```
30
```

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

In [10]:
```
#Let's have a review on count of people made purchase during the AM and also during PM seperately.
ecom['AM or PM'].value_counts()
```

Out[10]:
```
PM    5068
AM    4932
Name: AM or PM, dtype: int64
```

** What are the 5 most common Job Titles? **

In [11]:
```
#Let's print the 5 most common Job Titles present in the Dataset of "Ecommerce Purchases"
ecom['Job'].value_counts().head(5)
```

Out[11]:
```
Interior and spatial designer        31
Lawyer                               30
Social researcher                    28
Designer, jewellery                  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 [12]:
```
#Printing the Purchase Price of the Transaction made for the Purchase that came from Lot: "90 WT"
ecom[ecom['Lot']== "90 WT"]['Purchase Price']
```

Out[12]:
```
513    75.1
Name: Purchase Price, dtype: float64
```

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

In [13]:
```
#Printing the Email Id of the Employee or Person or Customer with Credit Card Number:4926535242672853
ecom[ecom['Credit Card']==4926535242672853]['Email']
```

Out[13]:
```
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 [14]:
```
#Let's count the number of people who keep Credit Card of American Express and have made a purchase above $95
ecom[(ecom['CC Provider']=='American Express') & (ecom['Purchase Price']>95)].count()
```

Out[14]:

```
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
```

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

In [15]:

```
#Let's print the sum of number of people or count whose credit card will expire in year 2025
sum(ecom["CC Exp Date"].apply(lambda x: x[3:]=='25'))
```

Out[15]:

```
1033
```

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

In [16]:

```
#Spliting up the Column Email Entries on symbol '@'.
#Then, Printing the top 5 email hosts used by the users all over the world by evaluating the data of "Ecommerce
ecom["Email"].apply(lambda x: x.split('@')[1]).value_counts().head(5)
```

Out[16]:

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

Great Job! - Mini Project 1 Finished !!!

Submitted By: Raunak Choudhary

Contact: raunakchoudhary17@gmail.com

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js