

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

In [2]: `df=pd.read_csv(' ../DSML17_ASSIGNMENT/pandas/Ecommerce Purchases')`

Check the head of the DataFrame.

In [3]: `df`

Out[3]:

		Address	Lot	AM or PM	Browser Info	Company	Credit Card	CC Exp Date	CC Security Code	CC Provider	Email	Job	IP Address	Language	Purchase Price
0		16629 Pace Camp Apt. 448\nAlexisborough, NE 77...	46 in	PM	Opera/9.56.(X11; Linux x86_64; sl-Sl) Presto/2...	Martinez- Herman	6011929061123406	02/20	900	JCB 16 digit	pdunlap@yahoo.com	Scientist, product/process development	149.146.147.205	el	98.14
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	Drilling engineer	15.160.41.51	fr	70.73
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	Customer service manager	132.207.160.22	de	0.95
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	Drilling engineer	30.250.74.19	es	78.04
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	Fine artist	24.140.33.94	es	77.82
...	
9995		966 Castaneda Locks\nWest Juliafurt, CO 96415	92 XI	PM	Mozilla/5.0 (Windows NT 5.1) AppleWebKit/5352 ...	Randall-Sloan	342945015358701	03/22	838	JCB 15 digit	iscott@wade-garner.com	Printmaker	29.73.197.114	it	82.21
9996		832 Curtis Dam Suite 785\nNorth Edwardburgh, T...	41 JY	AM	Mozilla/5.0 (compatible; MSIE 9.0; Windows NT ...	Hale, Collins and Wilson	210033169205009	07/25	207	JCB 16 digit	mary85@hotmail.com	Energy engineer	121.133.168.51	pt	25.63
9997		Unit 4434 Box 6343\nDPO AE 28026-0283	74 Zh	AM	Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10_7...	Anderson Ltd	6011539787356311	05/21	1	VISA 16 digit	tyler16@gmail.com	Veterinary surgeon	156.210.0.254	el	83.98
9998		0096 English Rest\nRoystad, IA 12457	74 cL	PM	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_8_8;...	Cook Inc	180003348082930	11/17	987	American Express	elizabethmoore@reid.net	Local government officer	55.78.26.143	es	38.84
9999		40674 Barrett Stravenue\nGrimesville, WI 79682	64 Hr	AM	Mozilla/5.0 (X11; Linux i686; rv:1.9.5.20) Gec...	Greene Inc	4139972901927273	02/19	302	JCB 15 digit	rachelford@vaughn.com	Embryologist, clinical	176.119.198.199	el	67.59

10000 rows × 14 columns

How many rows and columns are there?

In [4]: `df.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]: `df['Purchase Price'].mean()`

Out[5]: 50.347302000000025

What were the highest and lowest purchase prices?

In [6]: `df['Purchase Price'].max()`

Out[6]: 99.99

In [7]: `df['Purchase Price'].min()`

Out[7]: 0.0

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

In [8]: `df[df["Language"]=="en"].count()[0]`

Out[8]: 1098

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

In [9]: `df[df["Job"]=="Lawyer"].count()[0]`

Out[9]: 30

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

In [10]: `df['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]: `df2=df['Job'].value_counts()
df2.head()`

Out[11]:

Interior and spatial designer	31
Lawyer	30
Social researcher	28
Purchasing manager	27
Designer, jewellery	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]: `df2=df[df['Lot']=="90 WT"]
df2['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]: `df2 = df[df['Credit Card']==4926535242672853]
df2["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]: `df2 = df[df['CC Provider']=='American Express']
df2[df2['Purchase Price']>95.00].count()[1]`

Out[14]: 39

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

In [15]: `df[df['CC Exp Date'].str.slice(3)=='25'].count()[1]`

Out[15]: 1033

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

In [16]: `df['Email'].apply(lambda Email:Email.split('@')[1]).value_counts().head()`

Out[16]:

hotmail.com	1638
yahoo.com	1616
gmail.com	1605
smith.com	42
williams.com	37

Name: Email, dtype: int64

Great Job!