```
In [5]: import pandas as pd
```

```
In [6]: df=pd.read_csv('Ecommerce Purchases')
```

```
In [7]: df.head(10)
```

	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	pro
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	
5	7502 Powell Mission Apt. 768\nTravisland, VA 3...	21 XT	PM	Mozilla/5.0 (Macintosh; U; PPC Mac OS X 10_8_5...	Silva-Anderson	30246185196287	07/25	7169	Discover	ynguyen@gmail.com	
6	93971 Conway Causeway\nAndersonburgh, AZ 75107	96 Xt	AM	Mozilla/5.0 (compatible; MSIE 7.0; Windows NT ...	Gibson and Sons	6011398782655569	07/24	714	VISA 16 digit	olivia04@yahoo.com	
7	260 Rachel Plains Suite 366\nCastroberg, WV 24...	96 pG	PM	Mozilla/5.0 (X11; Linux i686) AppleWebKit/5350...	Marshall-Collins	561252141909	06/25	256	VISA 13 digit	phillip48@parks.info	Ev
8	2129 Dylan Burg\nNew Michelle, ME 28650	45 JN	PM	Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10_7...	Galloway and Sons	180041795790001	04/24	899	JCB 16 digit	kdavis@rasmussen.com	
9	3795 Dawson Extensions\nLake Tinafort, ID 88739	15 Ug	AM	Mozilla/5.0 (X11; Linux i686; rv:1.9.7.20) Gec...	Rivera, Buchanan and Ramirez	4396283918371	01/17	931	American Express	qcoleman@hunt-huerta.com	

```
In [8]: 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
```

```
In [11]: df.isna().sum()
```

```
Out[11]: Address              0
         Lot                  0
         AM or PM             0
         Browser Info         0
         Company              0
         Credit Card          0
         CC Exp Date          0
         CC Security Code     0
         CC Provider          0
         Email                0
         Job                  0
         IP Address           0
         Language             0
         Purchase Price       0
         dtype: int64
```

check datatype of each column

```
In [15]: df.dtypes
```

```
Out[15]: Address               object
         Lot                   object
         AM or PM              object
         Browser Info          object
         Company               object
         Credit Card            int64
         CC Exp Date           object
         CC Security Code       int64
         CC Provider           object
         Email                 object
         Job                   object
         IP Address            object
         Language              object
         Purchase Price       float64
         dtype: object
```

print number of rows and columns

```
In [19]: df.shape[0]
```

Out[19]: 10000

```
In [21]: len(df.columns)
```

Out[21]: 14

```
In [22]: df.columns
```

Out[22]: Index(['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'],
      dtype='object')

print highest and lowest purchase price average price

```
In [23]: df.describe()
```

Out[23]:

	Credit Card	CC Security Code	Purchase Price
count	1.000000e+04	10000.000000	10000.000000
mean	2.341374e+15	907.217800	50.347302
std	2.256103e+15	1589.693035	29.015836
min	6.040186e+10	0.000000	0.000000
25%	3.056322e+13	280.000000	25.150000
50%	8.699942e+14	548.000000	50.505000
75%	4.492298e+15	816.000000	75.770000
max	6.012000e+15	9993.000000	99.990000

```
In [24]: df['Purchase Price'].max()
```

Out[24]: 99.99

```
In [25]: df['Purchase Price'].min()
```

Out[25]: 0.0

```
In [26]: df['Purchase Price'].mean()
```

Out[26]: 50.347302

how many people have fr as their language

```
In [28]: df['Language'].unique()
```

Out[28]: 9

```
In [35]: df['individuals']=1
```

```
In [36]: df.head(10)
```

	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	pr
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	
5	7502 Powell Mission Apt. 768\nTravisland, VA 3...	21 XT	PM	Mozilla/5.0 (Macintosh; U; PPC Mac OS X 10_8_5...	Silva-Anderson	30246185196287	07/25	7169	Discover	ynguyen@gmail.com	
6	93971 Conway Causeway\nAndersonburgh, AZ 75107	96 Xt	AM	Mozilla/5.0 (compatible; MSIE 7.0; Windows NT ...	Gibson and Sons	6011398782655569	07/24	714	VISA 16 digit	olivia04@yahoo.com	
7	260 Rachel Plains Suite 366\nCastroberg, WV 24...	96 pG	PM	Mozilla/5.0 (X11; Linux i686) AppleWebKit/5350...	Marshall-Collins	561252141909	06/25	256	VISA 13 digit	phillip48@parks.info	Ev
8	2129 Dylan Burg\nNew Michelle, ME 28650	45 JN	PM	Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10_7...	Galloway and Sons	180041795790001	04/24	899	JCB 16 digit	kdavis@rasmussen.com	
9	3795 Dawson Extensions\nLake Tinafort, ID 88739	15 Ug	AM	Mozilla/5.0 (X11; Linux i686; rv:1.9.7.20) Gec...	Rivera, Buchanan and Ramirez	4396283918371	01/17	931	American Express	qcoleman@hunt-huerta.com	

```
In [40]: len(df[df['Language']=='fr'])          #len method can help the best

Out[40]: 1097
```

job title contains engineer (Very important)

```
In [41]: df.columns

Out[41]: Index(['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', 'individuals'],
        dtype='object')
```

```
In [59]: len(df[df['Job'].str.contains('engineer',case=False)])

Out[59]: 984
```

Find Email of the person with ip address 132.207.160.22

```
In [60]: df.columns

Out[60]: Index(['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', 'individuals'],
        dtype='object')
```

```
In [63]: df[df['IP Address']=='132.207.160.22']['Email']          #only shows the email colummn

Out[63]: 2    amymiller@morales-harrison.com
         Name: Email, dtype: object
```

how many people have cc provider is matercard and purchase >50

```
In [67]: len(df[(df['CC Provider']=='Mastercard') &(df['Purchase Price']>50)])
```

Out[67]: 405

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

```
In [73]: print("In AM:" + str(len(df[df['AM or PM']=='AM'])))
         print("In PM:" + str(len(df[df['AM or PM']=='PM'])))
```

```
In AM:4932
In PM:5068
```

```
In [74]: df['AM or PM'].value_counts()
```

Out[74]: AM or PM
         PM   5068
         AM   4932
         Name: count, dtype: int64

How many people have credit card that expires in 2020

```
In [75]: df.head()
```

Out[75]:

	Address	Lot	AM or PM	Browser Info	Company	Credit Card	CC Exp Date	CC Security Code	CC Provider	Email	Job
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	Scientist, product/process development
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
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
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
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

```
In [77]: len(df[df['CC Exp Date'].str.contains('/20',case=False)])
```

Out[77]: 988

How many people have credit card that expires in 2020 using a lambda function

```
In [80]: df[df['CC Exp Date'].apply(lambda x:x[3:]=='20')].count()
```

Out[80]:
```
Address            988
Lot                988
AM or PM           988
Browser Info       988
Company            988
Credit Card        988
CC Exp Date        988
CC Security Code   988
CC Provider        988
Email              988
Job                988
IP Address         988
Language           988
Purchase Price     988
individuals        988
dtype: int64
```

The top 5 email providers

```
In [86]: my_list=[]
         for email in df['Email']:
             my_list.append(email.split('@')[1])
```

```
In [88]: df['email_provider']= my_list
```

```
In [89]: df.head()
```

Out[89]:

	Address	Lot	AM or PM	Browser Info	Company	Credit Card	CC Exp Date	CC Security Code	CC Provider	Email	Job
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	Scientist, product/process development
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
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
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
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

```
In [90]: df['email_provider'].value_counts().head()
```

Out[90]:
```
email_provider
hotmail.com    1638
yahoo.com      1616
gmail.com      1605
smith.com        42
williams.com     37
Name: count, dtype: int64
```

by using lambda function

```
In [93]: df['Email'].apply(lambda x : x.split('@')[1]).value_counts().head()
```

```
Out[93]: Email
         hotmail.com      1638
         yahoo.com        1616
         gmail.com        1605
         smith.com          42
         williams.com       37
         Name: count, dtype: int64
```

In [ ]: