

In [157... `import pandas as pd`

In [158... `data = pd.read_csv(r"C:\Users\Pankaj Yadav\OneDrive\Desktop\Python Projects\ecommerce_purchases_a  
data`

Out[158...

	Address	Lot	AM or PM	Browser Info	Company	Credit Card	CC Exp Date	CC Security Code
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
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
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
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
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
...	...	...	...	...	...	...	...	...
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
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
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
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
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

10000 rows × 14 columns



# 1. Display top 10 rows of the dataset

```
In [159... data.head(10)
```

```
Out[159...
```

	Address	Lot	AM or PM	Browser Info	Company	Credit Card	CC Exp Date	CC Security Code
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
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
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
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
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
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
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
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
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
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



# 2. Check last 10 rows of the dataset

```
In [160... data.tail(10)
```

Out[160...

	Address	Lot	AM or PM	Browser Info	Company	Credit Card	CC Exp Date	CC Security Code	P
9990	75731 Molly Springs\nWest Danielle, VT 96934- 5102	93 ty	PM	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_4;...	Pace, Vazquez and Richards	869968197049750	04/24	877	
9991	PSC 8165, Box 8498\nAPO AP 60327- 0346	50 dA	AM	Mozilla/5.0 (compatible; MSIE 8.0; Windows NT ...	Snyder Inc	4221582137197481	02/24	969	
9992	885 Allen Mountains Apt. 230\nWallhaven, LA 16995	40 vH	PM	Mozilla/5.0 (Macintosh; PPC Mac OS X 10_6_5) A...	Wells Ltd	4664825258997302	10/20	431	I
9993	7555 Larson Locks Suite 229\nEllisburgh, MA 34...	72 jg	PM	Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10_8...	Colon and Sons	30025560104631	10/25	629	
9994	6276 Rojas Hollow\nLake Louis, WY 56410-7837	93 Ex	PM	Opera/9.68.(X11; Linux x86_64; sl-SI) Presto/2...	Ritter- Smith	3112186784121077	01/25	1823	
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	
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	
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	
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	A
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	



### 3. Check datatype of each column

In [161...

```
data.dtypes
```

```
Out[161... 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
```

## 4. Check null values in the dataset

```
In [162... data.isnull().sum()
```

```
Out[162... 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
```

## 5. How many rows and columns are there in our dataset

```
In [163... len(data.columns)
```

```
Out[163... 14
```

```
In [164... len(data)
```

```
Out[164... 10000
```

```
In [165... data.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

```

## 6. Highest and lowest purchase prices

```
In [166... data.columns
```

```
Out[166... 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')
```

```
In [167... data['Purchase Price'].max()
```

```
Out[167... 99.99
```

```
In [168... data['Purchase Price'].min()
```

```
Out[168... 0.0
```

## 7. Average purchase price

```
In [169... data['Purchase Price'].mean()
```

```
Out[169... 50.347302
```

## 8. How many people have French 'fr' as their language?

```
In [170... data.columns
```

```
Out[170... 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')
```

```
In [171... data[data['Language'] == 'fr']
```

Out[171...

	Address	Lot	AM or PM	Browser Info	Company	Credit Card	CC Exp Date	CC Security Code	CC Provide
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
19	125 Hall Summit\nBoothton, IL 41721	99 CU	PM	Mozilla/5.0 (compatible; MSIE 7.0; Windows NT ...	Turner- Mckinney	676343504830	02/20	440	VISA 11 digi
53	PSC 9431, Box 7059\nAPO AA 29285-1363	14 qD	AM	Opera/9.34. (X11; Linux x86_64; it-IT) Presto/2...	Higgins, Cardenas and Kennedy	869972604798355	08/17	157	JCB 11 digi
76	49206 Campbell Port\nNorth Cliffordshire, HI 3...	71 iu	PM	Mozilla/5.0 (Macintosh; U; PPC Mac OS X 10_5_9...	Jacobs- Tucker	6011343518820988	01/17	806	Voyage
82	493 Smith Valleys Suite 004\nNew Madelineville...	35 ls	PM	Mozilla/5.0 (iPod; U; CPU iPhone OS 4_0 like M...	Robinson, Johnston and Valdez	4351359627548412	06/17	937	VISA 11 digi
...	...	...	...	...	...	...	...	...	...
9941	43757 Brown Lodge\nValerieberg, PR 88518	68 DG	AM	Mozilla/5.0 (compatible; MSIE 5.0; Windows NT ...	Jones, Williams and Dunn	6011508474487291	04/26	906	Mastercard
9947	32827 Carl Inlet\nSouth Nicole, NY 38081-6636	40 Dp	PM	Mozilla/5.0 (Windows; U; Windows 95) AppleWebK...	Pruitt- Flores	3158564208754951	03/17	386	Mastercard
9951	518 Brown Dam\nGarciaside, IN 33057	71 wq	AM	Mozilla/5.0 (compatible; MSIE 8.0; Windows NT ...	Crosby, Vasquez and Ballard	5256758441931287	12/17	146	VISA 11 digi
9977	02182 Keith Expressway\nEast Shannon, CT 20578...	34 RL	AM	Mozilla/5.0 (compatible; MSIE 9.0; Windows NT ...	Deleon, Jacobson and Benton	4186094003664688	06/21	397	JCB 11 digi
9980	6061 Dave Lights\nPhillipsview, UT 40500	25 Mv	AM	Opera/8.58. (Windows NT 5.0; it-IT) Presto/2.9....	Cortez- Frye	6011944199756993	12/24	755	JCB 11 digi

1097 rows × 14 columns



```
In [172...] len(data[data['Language'] == 'fr'])
```

```
Out[172...] 1097
```

```
In [173...] data[data['Language'] == 'fr'].count()
```

```
Out[173...] Address      1097
Lot          1097
AM or PM     1097
Browser Info 1097
Company      1097
Credit Card  1097
CC Exp Date  1097
CC Security Code 1097
CC Provider  1097
Email        1097
Job          1097
IP Address   1097
Language     1097
Purchase Price 1097
dtype: int64
```

## 9. Job title contains Engineer

```
In [174...] data.columns
```

```
Out[174...] 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')
```

```
In [175...] data['Job'].str.contains('Engineer')
```

```
Out[175...] 0      False
1      False
2      False
3      False
4      False
...
9995    False
9996    False
9997    False
9998    False
9999    False
Name: Job, Length: 10000, dtype: bool
```

```
In [176...] len(data[data['Job'].str.contains('Engineer', case=False)])
```

```
Out[176...] 984
```

## 10. Find email of the person with the following ip address: 132.207.160.22

```
In [177...] data.columns
```

```
Out[177... 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')
```

```
In [178... data[data['IP Address'] == '132.207.160.22']['Email']
```

```
Out[178... 2    amymiller@morales-harrison.com  
      Name: Email, dtype: object
```

## 11. How many people have mastercard as their credit card provider and made a purchase above 50?

```
In [179... data.columns
```

```
Out[179... 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')
```

```
In [180... len(data[(data['CC Provider'] == 'Mastercard') & (data['Purchase Price'] > 50)])
```

```
Out[180... 405
```

## 12. Find the email of the person with the following credit card number: 4664825258997302

```
In [181... data.columns
```

```
Out[181... 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')
```

```
In [182... data[data['Credit Card'] == 4664825258997302]['Email']
```

```
Out[182... 9992    bberry@wright.net  
      Name: Email, dtype: object
```

## 13. How many people purchase during AM and How many people purchase during PM?

```
In [183... data.columns
```

```
Out[183... 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')
```

```
In [184... data['AM or PM'].value_counts()
```



```
Out[184...] AM or PM
PM      5068
AM      4932
Name: count, dtype: int64
```

## 14. How many people have a credit card that expires in 2020?

```
In [185...] data.columns
```

```
Out[185...] 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')
```

```
In [186...] CC_20 = data[data['CC Exp Date'].str.endswith('20')]
```

```
In [187...] count = CC_20.shape[0]
```

```
In [188...] print("People with credit cards expiring in 2020:", count)
```

People with credit cards expiring in 2020: 988

```
In [189...] # so, 988 people's credit card will expire in 2020
```

```
In [190...] data.columns
```

```
Out[190...] 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')
```

```
In [191...] len(data[data['CC Exp Date'].apply(lambda x:x[3:] == '20')])
```

```
Out[191...] 988
```

## 15. Top 5 most popular email providers

```
In [192...] data.head(2)
```

```
Out[192...]
      Address  Lot  AM or PM  Browser Info  Company  Credit Card  CC Exp Date  CC Security Code  CC Provider
0  16629 Pace Camp Apt. 46 in PM  Opera/9.56. (X11; Linux x86_64; sl-SI) Presto/2... Martinez-Herman 6011929061123406 02/20 900 JCB 16 digit f
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 ai
```


```
In [193... list1 = []
for email in data['Email']:
    list1.append(email.split('@')[1])
```

```
In [194... data['temp'] = list1
```

```
In [195... data.head(1)
```

```
Out[195...
```

	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	pdu

◀  ▶

```
In [196... data['temp'].value_counts().head(5)
```

```
Out[196... temp
hotmail.com    1638
yahoo.com     1616
gmail.com     1605
smith.com      42
williams.com   37
Name: count, dtype: int64
```

```
In [197... data['Email'].apply(lambda x:x.split('@')[1]).value_counts().head()
```

```
Out[197... Email
hotmail.com    1638
yahoo.com     1616
gmail.com     1605
smith.com      42
williams.com   37
Name: count, dtype: int64
```

```
In [ ]:
```