

In [1]:

```
import pandas as pd
data = pd.read_csv('https://raw.githubusercontent.com/DataThinkers/Datasets/refs/heads/main/data')
```

Out[1]:

	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 [4]:

```
data.head(10)
```

Out[4]:

	Address	Lot	AM or PM	Browser Info	Company	Credit Card	CC Exp Date	CC Securit Cod
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	90
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	56
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	69
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	38
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	67
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	716
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	71
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	25
8	2129 Dylan Burg\nMichelle, ME 28650	45 JN	PM	Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10_7...	Galloway and Sons	180041795790001	04/24	89
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	93

## 2. Check last 10 Rows of the Dataset

In [7]:

```
data.tail(10)
```

Out[7]:

	Address	Lot	AM or PM	Browser Info	Company	Credit Card	CC Exp Date	CC Security Code
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
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
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 [10]:

```
data.dtypes
```

Out[10]:

```
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 [15]:
```

```
data.isnull().sum()
```

```
Out[15]:
```

```
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 [18]:
```

```
len(data.columns)
```

```
Out[18]:
```

```
14
```

```
In [20]:
```

```
len(data)
```

```
Out[20]:
```

```
10000
```

```
In [22]:
```

```
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 Price

In [25]:

```
data.columns
```

Out[25]:

```
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 [27]:

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

Out[27]:

```
99.99
```

In [29]:

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

Out[29]:

```
0.0
```

## 7. Average Purchase Price

In [32]:

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

Out[32]:

```
50.347302
```

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

In [35]:

```
data.columns
```

Out[35]:

```
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 [43]:

```
len(data[data['Language'] == 'fr'])
```

Out[43]:

1097

In [47]:

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

Out[47]:

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 [50]:

```
data.columns
```

Out[50]:

```
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 [62]:

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

Out[62]:

984

## 10. Find Email of the person with the following IP address : 132.207.160.22

In [65]:

```
data.columns
```

Out[65]:

```
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 [69]:

```
data[data['IP Address'] == '132.207.160.22']
```

Out[69]:

	Address	Lot	AM or PM	Browser Info	Company	Credit Card	CC Exp Date	CC Security Code	CC Provider	
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@mo harrisor

## 11. How many people have Mastercard as their Credit Card Provider and made a Purchase above 50?

In [73]:

```
data.columns
```

Out[73]:

```
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 [83]:

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

Out[83]:

	Address	Lot	AM or PM	Browser Info	Company	Credit Card	CC Exp Date	CC Security Code
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
18	461 Christopher Square\nWest Michaelchester, C...	17 SB	PM	Mozilla/5.0 (X11; Linux i686; rv:1.9.6.20) Gec...	Beard, Abbott and Pena	6011350184276270	12/22	767
31	USNS Alvarado\nFPO AA 27052-1231	26 Lh	PM	Opera/8.84. (X11; Linux i686; sl-SI) Presto/2.9...	Nicholson Group	4614997834548	03/22	909
35	93392 Webb Gardens Apt. 220\nLaurabury, AR 999...	37 om	AM	Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10_7...	Mora Ltd	6011049630969815	09/16	367

	Address	Lot	AM or PM	Browser Info	Company	Credit Card	CC Exp Date	CC Security Code
90	431 Bowen Lights\nFergusonborough, MH 01362	31 tG	AM	Mozilla/5.0 (Windows NT 5.2; it-IT; rv:1.9.1.2...	Copeland- Lee	639032576097	06/18	868
...	...	...	...	...	...	...	...	...
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
9948	95544 Johnson Isle Suite 939\nMichaelberg, RI ...	91 bW	AM	Opera/8.36. (X11; Linux x86_64; sl-SI) Presto/2...	Fox- Peterson	4762924304307	03/17	567
9954	051 Samantha Wells Apt. 328\nScottborough, OR ...	11 vG	AM	Mozilla/5.0 (Macintosh; U; PPC Mac OS X 10_8_2...	Mendoza, Zimmerman and Reilly	30232295374907	05/21	762
9981	087 Larson Roads Apt. 587\nRicestad, WI 98077-...	95 Kw	AM	Opera/9.63. (Windows NT 5.01; en-US) Presto/2.9...	Ortiz, Schultz and Adams	4690863684605	01/19	695
9987	2754 Klein Mission\nJohnsonview, PA 64925-9624	65 Fg	PM	Opera/9.26. (Windows CE; sl-SI) Presto/2.9.188 ...	Curtis LLC	180027305702263	12/24	865

405 rows × 14 columns

In [85]:

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

Out[85]:

405

In [88]:

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

Out[88]:

```
Address      405
Lot          405
AM or PM     405
Browser Info 405
Company      405
Credit Card  405
CC Exp Date  405
CC Security Code 405
CC Provider  405
Email        405
Job          405
IP Address   405
Language     405
```



Purchase Price 405  
dtype: int64

## 12. Find Email of the Person with the following Credit Card Number : 4664825258997302

In [99]:

```
data.columns
```

Out[99]:

```
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 [96]:

```
data[data['Credit Card'] == 4664825258997302]
```

Out[96]:

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

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

In [103]:

```
data.columns
```

Out[103]:

```
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 [105]:

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

Out[105]:

```
AM or PM  
PM      5068  
AM      4932  
Name: count, dtype: int64
```

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

In [108]:

```
data['CC Exp Date']
```

Out[108]:

```
0      02/20
1      11/18
2      08/19
3      02/24
4      10/25
```

```
...
9995   03/22
9996   07/25
9997   05/21
9998   11/17
9999   02/19
```

Name: CC Exp Date, Length: 10000, dtype: object

In [110]:

```
def func():
    count = 0
    for date in data['CC Exp Date']:
        if date.split('/')[1] == '20':
            count = count+1
    print(count)
```

In [114]:

```
func()
```

988

In [118]:

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

Out[118]:

988

## 15. Top 5 Most Popular Email providers(e.g. gmail.com,yahoo.com etc)

In [121]:

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

In [123]:

```
data['temp'] = list1
```

In [127]:

```
data.head(1)
```

Out[127]:

	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	pc

In [135]:

```
data['temp'].value_counts()
```

Out[135]:

```
temp
hotmail.com      1638
yahoo.com        1616
gmail.com        1605
smith.com         42
williams.com     37
...
booker.com        1
woods-allen.biz  1
richards-wilson.com  1
morris-thomas.com  1
wade-garner.com   1
Name: count, Length: 3416, dtype: int64
```

In [143]:

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

Out[143]:

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

In [ ]:

In [ ]: