**E-commers project for analysis and finding insights**

In [1]:

**import** pandas **as** pd

In [2]:

url **=** 'https://raw.githubusercontent.com/DataThinkers/Datasets/main/DS/Ecommerce%20Purchases'

**def** read\_data\_git():

ex1 **=** pd.read\_csv(url)

**return** ex1

In [3]:

ecommerce**=**read\_data\_git()

ecommerce.head(1)

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-SI) 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 |

In [4]:

*## Copy the main data*

data**=**ecommerce.copy()

In [5]:

data.head(2)

Out[5]:

|  | **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-SI) 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 |

In [6]:

*## Findings of null value*

data.isnull().sum()

Out[6]:

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

In [7]:

*## Check datatype of each columns*

data.dtypes

Out[7]:

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

In [8]:

*## Number of rows and columns*

data.shape

Out[8]:

(10000, 14)

In [9]:

*## highest and lowest purchase price*

​

data.columns

Out[9]:

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 [10]:

*## Highest purchase price*

​

print('Highest purchase price',data['Purchase Price'].max())

Highest purchase price 99.99

In [11]:

print('Lowest purchase price',data['Purchase Price'].min())

Lowest purchase price 0.0

In [12]:

*## Average purchase price*

print('Average purchase price -',data['Purchase Price'].mean())

Average purchase price - 50.347302

In [13]:

*## How many people have French language*

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

Out[13]:

1097

In [14]:

*## job title contain 'Enginers'*

​

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

Out[14]:

984

In [15]:

*## Find Email of the person with the fllowing IP adress: 132.207.160.22*

​

data.columns

​

Out[15]:

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 [61]:

data[data['IP Address']**==** '132.207.160.22'][['Email','Company']]

Out[61]:

|  | **Email** | **Company** |
| --- | --- | --- |
| **2** | amymiller@morales-harrison.com | Simpson, Williams and Pham |

In [17]:

*## How many people have master card and purchase above 50*

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

​

Out[17]:

405

In [18]:

*## Find Email of person with the following card number:4664825258997302*

​

data[data['Credit Card']**==** 4664825258997302]['Email']

Out[18]:

9992 bberry@wright.net

Name: Email, dtype: object

In [19]:

*## How many people purchase in AM and PM*

​

*## AM*

len(data[data['AM or PM']**==**'AM'])

Out[19]:

4932

In [20]:

*## PM*

​

len(data[data['AM or PM']**==**'PM'])

​

Out[20]:

5068

In [21]:

data['AM or PM'].value\_counts()

Out[21]:

PM 5068

AM 4932

Name: AM or PM, dtype: int64

In [22]:

*## How many people have credit card expire on 2020*

data2**=**ecommerce.copy()

In [23]:

**def** fun():

count**=**0

**for** date **in** data2['CC Exp Date']:

**if** date.split('/')[1]**==**'20':

count**=**count**+**1

print(count)

In [24]:

fun()

988

In [39]:

*## another processure*

​

data2['Ex\_Year'] **=** data2['CC Exp Date'].str[3:5]

In [45]:

data2.head(2)

Out[45]:

|  | **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** | **Ex\_Year** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **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 | 149.146.147.205 | el | 98.14 | 20 |
| **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 | 18 |

In [46]:

*## How many people have credit card expire on 2020*

​

len(data2[data2['Ex\_Year'] **==** '20'])

Out[46]:

988

In [48]:

*## another processure using Lamda function*

​

len(data2[data2['CC Exp Date'].apply(**lambda** x:x[3:]**==**'20')])

Out[48]:

988

In [52]:

*## Top 5 Email provider*

data2['Email\_last'] **=** data2['Email'].apply(**lambda** x:x.split('@')[1])

data2.head(2)

Out[52]:

|  | **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** | **Ex\_Year** | **Email\_last** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **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 | 149.146.147.205 | el | 98.14 | 20 | 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 | Drilling engineer | 15.160.41.51 | fr | 70.73 | 18 | reed.com |

In [59]:

Top\_5 **=** data2['Email\_last'].value\_counts()

Top\_5.head(5)

Out[59]:

hotmail.com 1638

yahoo.com 1616

gmail.com 1605

smith.com 42

williams.com 37

Name: Email\_last, dtype: int64

In [ ]:

​