

Data Analysis with Python

E-COMMERCE PURCHASE CASE STUDY

Import the file

```
import pandas as pd
import zipfile

# Unzipping and Loading the dataset from Kaggle
with zipfile.ZipFile('C:\\Users\\HP\\Downloads\\Ecommerce Purchases (1).zip', 'r') as zip_ref:
    zip_ref.extractall('Downloads/')

# Load the CSV file after extracting
df = pd.read_csv('C:\\Users\\HP\\Downloads\\Ecommerce Purchases (1).zip')

# Previewing the data
df.head()
```

	Address	Lot	AM or PM	Browser Info	Company	Credit Card	CC Exp Date	CC Security Code	CC Provider	Email	Job	IP Addres
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	149.146.147.20
	9374 Jasmine Spurs			Opera/8.93. (Windows	Fletcher,							

DISPLAY TOP 10 ROWS OF THE DATA SET

	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
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

Problem

check last 10 rows of the data set

df.ta:	df.tail(10)							
	Address	Lot	AM or PM	Browser Info	Company	Credit Card		
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		
9991	PSC 8165, Box 8498\nAPO AP 60327- 0346	50 dA	AM	Mozilla/5.0 (compatible; MSIE 8.0; Windows NT	Snyder	4221582137197481		

check data type of each column.

df.dtypes	-	
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 data set

Problem

how many rows and columns are there in our data set

```
len(df.columns)
: 14
: len(df)
: 10000
```

Highest and Lowest purchase price

Problem

Average purchase price

```
df['Purchase Price'].mean()
np.float64(50.347302)
```

How many people have french 'fr' as their language

Problem

Job title contains engineer



find email of the person with the following ip address: 132.207.160.22

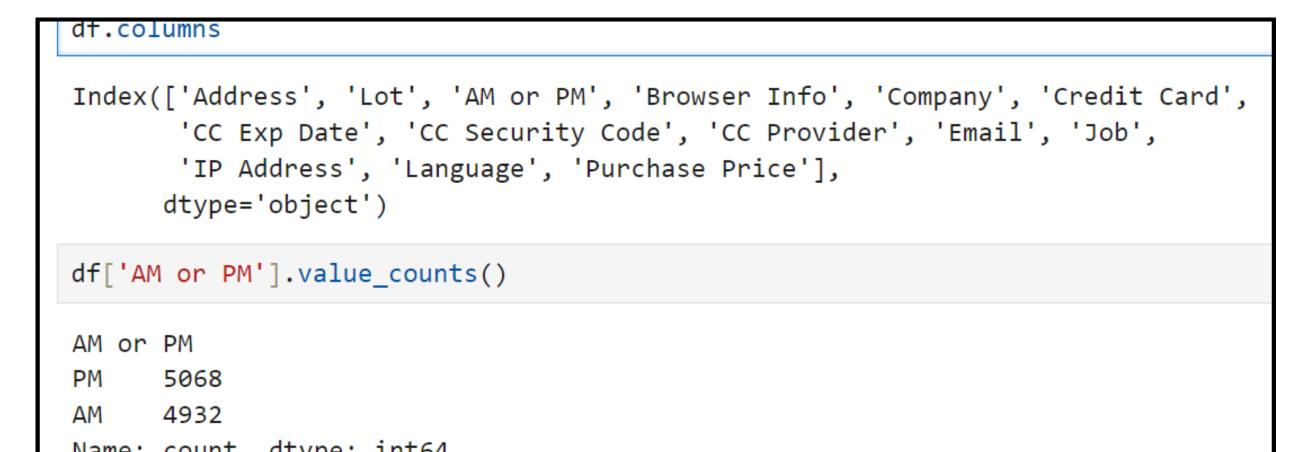
How many people have mastercard as their credit card provider and a purchase above 50

Problem

Find email of the person with the following credit number-4664825258997302

How many people purchase during 'am' and how many people purchase during 'pm'

Problem



How many people have credit card that expires in 2020

Top 5 most popular in email providers (e.g, gmail.com,yahoo.com etc)

Problem

created by

Arpita Chakroborty

arpitachakroborty.official@gmail.com

