Bridge The Gap 2 (Data Science) ¶

Atharva Taras (TE A - 73)

In [1]:
```
import pandas as pd
```

1. Check the head of the DataFrame.

In [2]:
```
data = pd.read_csv(r'A:\Python Projects\College Practicals/Ecommerce Purchas
data.head()
```

Out[2]:

	Address	Lot	AM or PM	Browser Info	Company	Credit Card	CC Exp Date	CC Security Code	Provi
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 d
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	Masterc
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 d
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	Disco
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	Din Clu Ca Blanc

2. How many rows and columns are there?

In [3]:
```
print(f'There are {data.shape[0]} Rows and {data.shape[1]} Columns in this d
```

```
There are 10000 Rows and 14 Columns in this dataset.
```

3. What is the average Purchase Price?

In [4]:
```
avg = sum(data['Purchase Price'])/len(data['Purchase Price'])
print(f'Average Purchase Price is {round(avg, 2)} $')
```

```
Average Purchase Price is 50.35 $
```

4. What were the highest and lowest purchase prices?

In [5]:
```
print('Highest Purchase Price - {}\nLowest Purchase Price - {}'.format(max(d
                                                                       min(d
                                                                      ))
```

```
Highest Purchase Price - 99.99
Lowest Purchase Price - 0.0
```

5. How many people have English 'en' as their Language of choice on the website?

In [6]:
```
print('{} people have English "en" as their choice.'.format(data['Language']
```

```
1098 people have English "en" as their choice.
```

6. How many people have the job title of "Lawyer"

In [7]:
```
print('{} people are lawyers.'.format(data['Job'].str.count('Lawyer').sum())
```

```
30 people are lawyers.
```

7. How many people made the purchase during the AM and how many people made the purchase during PM ?

In [8]:
```
print('{} people made purchase during AM and {} made purchase during PM'.for

```

```
4932 people made purchase during AM and 5068 made purchase during PM
```

8. What are the 5 most common Job Titles?

In [9]:
```
print('These are the 5 most common job titles - \n\n{}'.format(data['Job'].v
```

```
These are the 5 most common job titles -

Interior and spatial designer    31
Lawyer                           30
Social researcher                28
Purchasing manager               27
Designer, jewellery              27
Name: Job, dtype: int64
```

9. Someone made a purchase that came from Lot: "90 WT" , what was the Purchase Price for this transaction?

In [10]:
```
price = data[(data['Lot'] == '90 WT')]['Purchase Price'].item()
print('Purchase price for lot 90 WT is ${}'.format(price))
```

```
Purchase price for lot 90 WT is $75.1
```

10. What is the email of the person with the following Credit Card Number: 4926535242672853

In [11]:
```
mail = data[(data['Credit Card'] == 4926535242672853)]['Email'].item()
print('Email of that person is {}'.format(mail))
```

```
Email of that person is bondellen@williams-garza.com
```

11. How many people have American Express as their Credit Card Provider and made a purchase above $95 ?

In [12]:
```
tmp = data[((data['CC Provider'] == 'American Express') & (data['Purchase Pr
print(f'{tmp.shape[0]} people use American Express and have purchase price a
```

```
39 people use American Express and have purchase price above $95
```

12. How many people have a credit card that expires in 2025?

In [13]:
```
i = 0

for date in data['CC Exp Date']:
    if date.endswith('/25'):
        i += 1

print(f'{i} cards expire in 2025')
```

```
1033 cards expire in 2025
```

13. What are the top 5 most popular email providers/hosts (e.g. gmail.com, yahoo.com, etc...)

In [14]:
```
tmplist = []

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

tdf = pd.DataFrame(tmplist)
print('The top 5 email providers are -\n')

for i in range(5):
    print(i+1, tdf[0][i])
```

```
The top 5 email providers are -

1 yahoo.com
2 reed.com
3 morales-harrison.com
4 olson-robinson.info
5 gmail.com
```