

# Coding Challenge Python

DE131 Sameer Pal

## Overall Data Summarization

Summarize key variables in the dataset to build a basic understanding, laying the groundwork for deeper analysis.

```
In [2]: import pandas as pd
import numpy as np

data=pd.read_excel("C:\\Users\\SAMEER\\Downloads\\project3 (1) (1).xlsx")
```

Out[2]:

	Bounces	Exits	Continent	Sourcegroup	Timeinpage	Uniquepageviews	Visits	Bour
0	0	0	OC	(direct)	18	1	0	
1	0	0	N.America	(direct)	4	1	0	
2	0	0	N.America	Others	35	1	0	
3	0	0	N.America	public.tableausoftware.com	70	1	0	
4	0	0	N.America	public.tableausoftware.com	81	1	0	

```
In [45]: data.head()
```

Out[45]:

	Bounces	Exits	Continent	Sourcegroup	Timeinpage	Uniquepageviews	Visits	Bour
0	0	0	OC	(direct)	18	1	0	
1	0	0	N.America	(direct)	4	1	0	
2	0	0	N.America	Others	35	1	0	
3	0	0	N.America	public.tableausoftware.com	70	1	0	
4	0	0	N.America	public.tableausoftware.com	81	1	0	

## Basic Info

```
[3]: data.columns
```

```
[3]: Index(['Bounces', 'Exits', 'Continent', 'Sourcegroup', 'Timeinpage',  
         'Uniquepageviews', 'Visits', 'BouncesNew'],  
        dtype='object')
```

```
[4]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 32109 entries, 0 to 32108  
Data columns (total 8 columns):  
#   Column                Non-Null Count  Dtype  
---  -  
0   Bounces                32109 non-null  int64  
1   Exits                  32109 non-null  int64  
2   Continent              32109 non-null  object  
3   Sourcegroup            32109 non-null  object  
4   Timeinpage             32109 non-null  int64  
5   Uniquepageviews        32109 non-null  int64  
6   Visits                  32109 non-null  int64  
7   BouncesNew             32109 non-null  float64  
dtypes: float64(1), int64(5), object(2)  
memory usage: 2.0+ MB
```

---

## Some Mathematical function

In [35]: `print(data.describe())`

	Bounces	Exits	Timeinpage	Uniquepageviews
count	32109.000000	32109.000000	32109.000000	32109.000000
mean	0.713009	0.906039	73.184746	1.114329
std	0.708215	0.695819	394.441111	0.614880
min	0.000000	0.000000	0.000000	1.000000
25%	0.000000	1.000000	0.000000	1.000000
50%	1.000000	1.000000	0.000000	1.000000
75%	1.000000	1.000000	10.000000	1.000000
max	30.000000	36.000000	46745.000000	45.000000

	Visits	BouncesNew
count	32109.000000	32109.000000
mean	0.906039	0.007130
std	0.730068	0.007082
min	0.000000	0.000000
25%	1.000000	0.000000
50%	1.000000	0.010000
75%	1.000000	0.010000
max	45.000000	0.300000

In [36]: `print(data.isnull().sum())`

```
Bounces      0
Exits        0
Continent    0
Sourcegroup  0
Timeinpage   0
Uniquepageviews 0
Visits       0
BouncesNew   0
dtype: int64
```

## Unique Page Views vs Visits

Analyze if the number of unique page views depends on the total number of visits to the website

### Correlation Coefficient Calculation

In [58]: # 2

```
correlation = data['Visits'].corr(data['Uniquepageviews'])  
print(f"Correlation between visits and unique page views: {correlation}")
```

Correlation between visits and unique page views: 0.8144457070735212

## Scattered Chart : UniquePageViews V/S Visits

In [47]: # 2

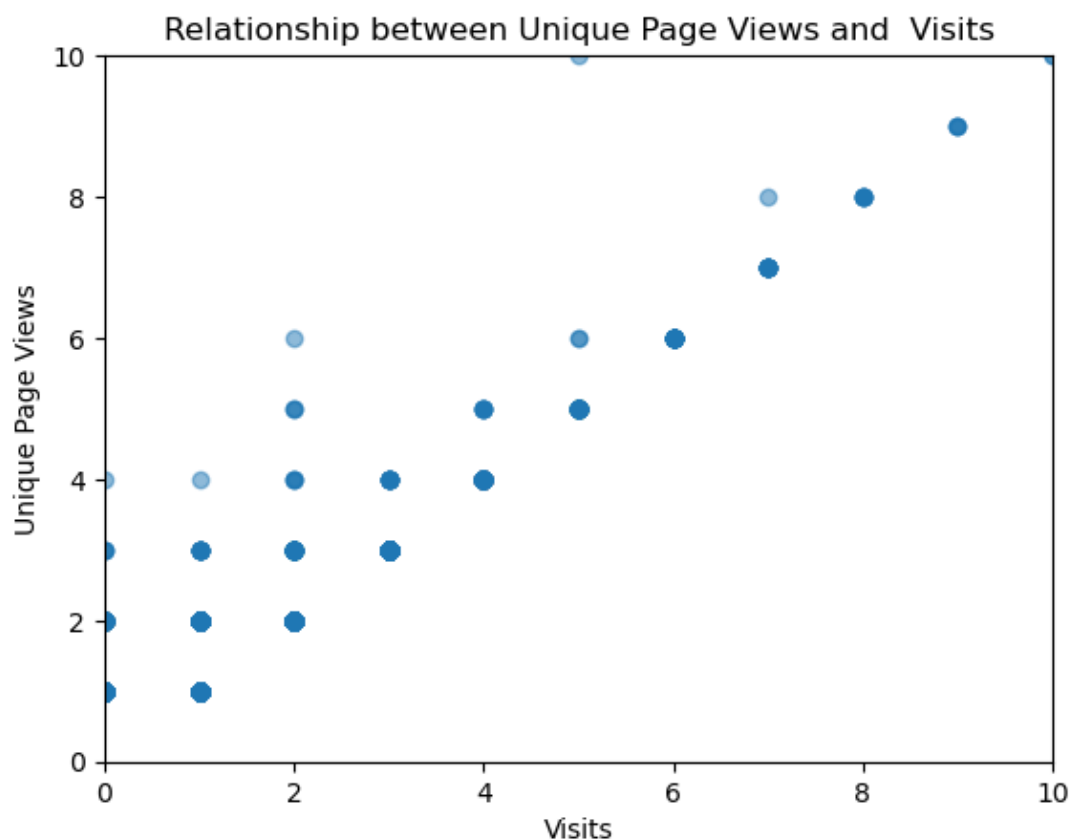
```
import matplotlib.pyplot as plt

plt.scatter(data["Visits"], data['Uniquepageviews'],alpha=[0.5])

plt.title('Relationship between Unique Page Views and Visits')
plt.xlabel('Visits')

plt.xlim(0, 10)
plt.ylim(0,10)

plt.ylabel('Unique Page Views')
plt.show()
```



## Factors Affecting Exits

Identify the factors in the dataset that could contribute to users exiting the site, helping understand why users leave.

### Correlation Coefficient Calculation

In [59]: # 3

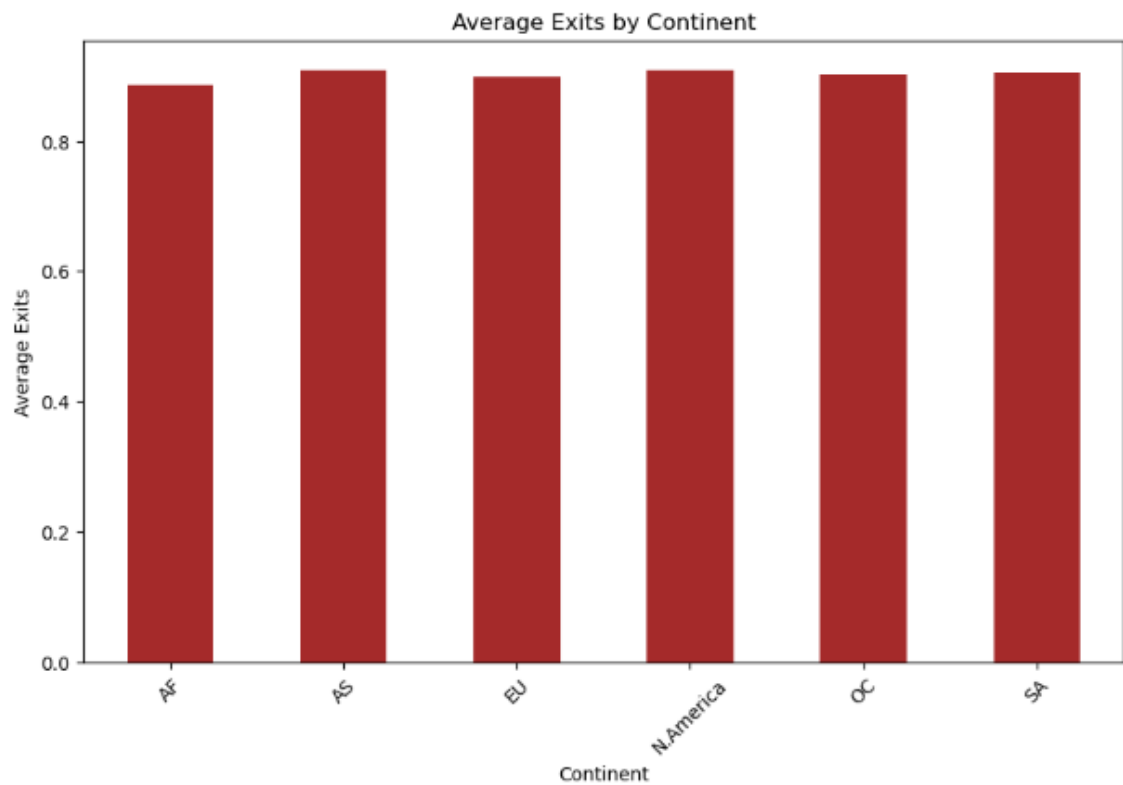
```
numeric_data = data.select_dtypes(include=['number'])
corr_mat = numeric_data.corr()
exit_corr = corr_mat['Exits']
print(exit_corr)
```

```
Bounces      0.824912
Exits         1.000000
Timeinpage   0.001325
Uniquepageviews 0.791129
Visits        0.800979
BouncesNew    0.824912
Name: Exits, dtype: float64
```

---

## Average Exits V/s Continent

```
In [48]: # 3
plt.figure(figsize=(10, 6))
data.groupby('Continent')['Exits'].mean().plot(kind='bar', color='brown')
plt.title('Average Exits by Continent')
plt.xlabel('Continent')
plt.ylabel('Average Exits')
plt.xticks(rotation=45)
plt.show()
```



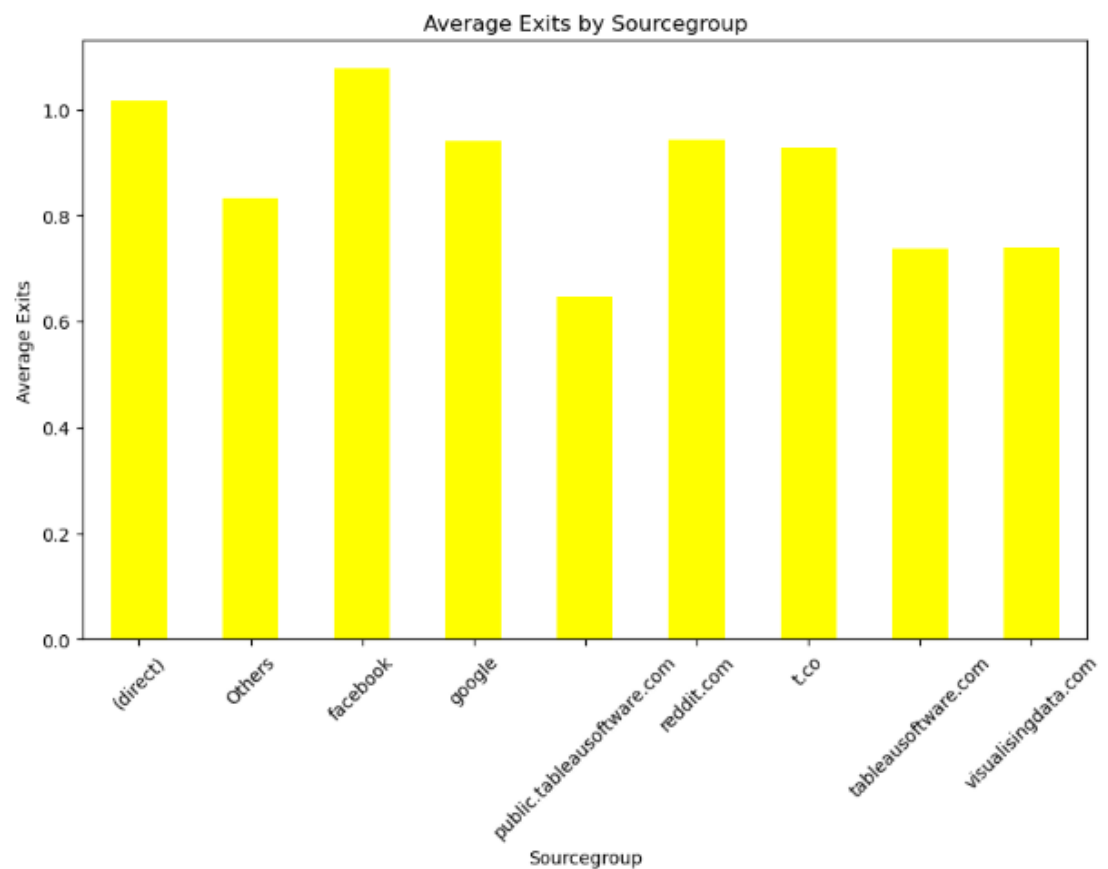
## Average Exits vs Source Group

```
In [49]: # 3
plt.figure(figsize=(10, 6))

data.groupby('Sourcegroup')['Exits'].mean().plot(kind='bar', color='yellow')

plt.title('Average Exits by Sourcegroup')
plt.xlabel('Sourcegroup')

plt.ylabel('Average Exits')
plt.xticks(rotation=45)
plt.show()
```





## Factors Affecting Time on Page

Determine which variables influence the amount of time a visitor spends on a page to improve user engagement

### Correlation Coefficient

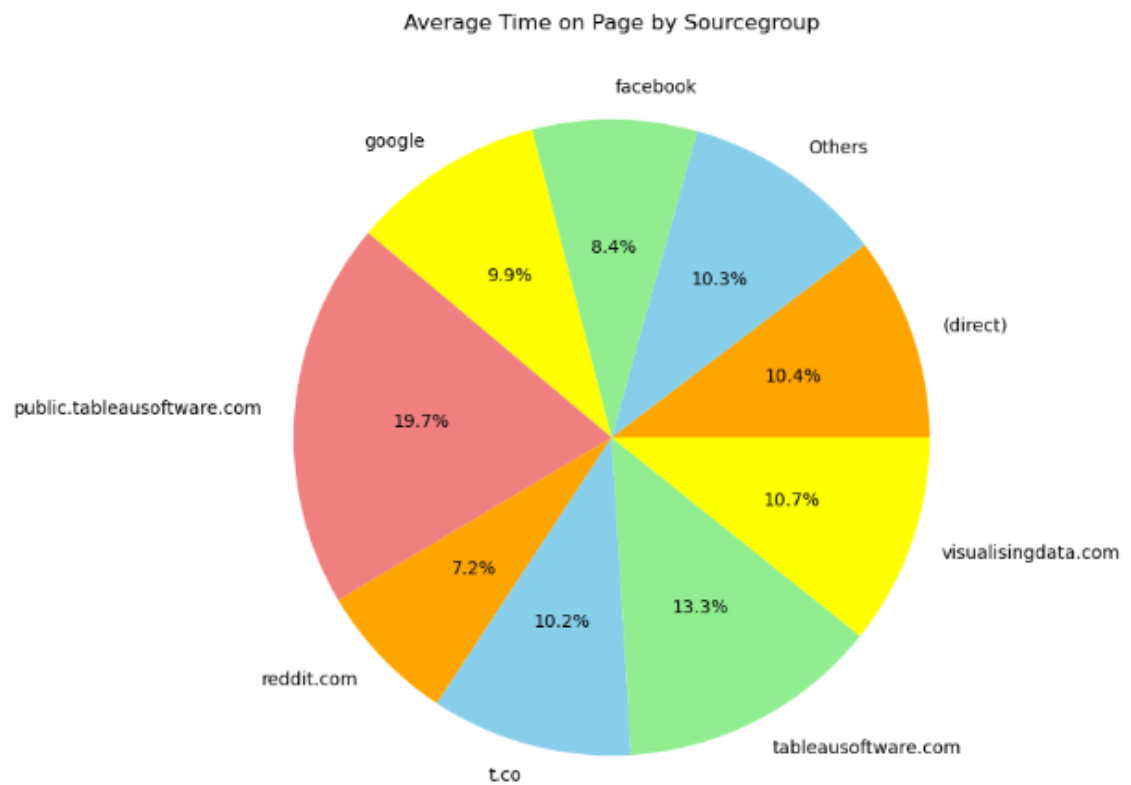
In [52]: # 4

```
numeric_data = data.select_dtypes(include=['number'])
corr_mat = numeric_data.corr()
exit_corr = corr_mat['Timeinpage']
print(exit_corr)
```

```
Bounces          -0.109106
Exits             0.001325
Timeinpage        1.000000
Uniquepageviews   0.114593
Visits            0.066650
BouncesNew        -0.109106
Name: Timeinpage, dtype: float64
```

## Pie Chart for Page By SourceGroup

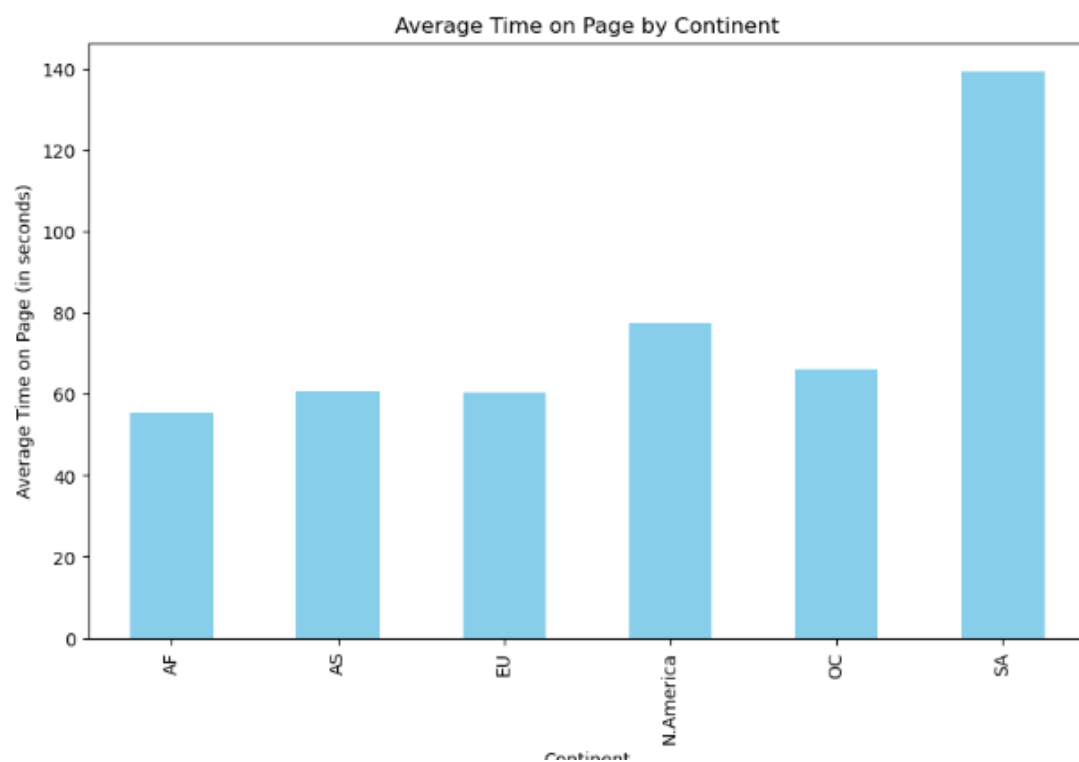
```
In [54]: # 4
plt.figure(figsize=(8, 8))
sourcegroup_avg_time = data.groupby('Sourcegroup')['Timeinpage'].mean()
sourcegroup_avg_time.plot(kind='pie', autopct='%1.1f%%', colors=['orange', 'skyb
plt.title('Average Time on Page by Sourcegroup')
plt.ylabel('')
plt.show()
```



## Average Time on Page By Continent

```
In [53]: # 4
plt.figure(figsize=(10, 6))

data.groupby('Continent')['Timeinpage'].mean().plot(kind='bar', color='skyblue')
plt.title('Average Time on Page by Continent')
plt.xlabel('Continent')
plt.ylabel('Average Time on Page (in seconds)')
plt.show()
```



## Factors Impacting Bounce Rate

Identify factors that impact the bounce rate, helping understand which aspects of the site cause visitors to leave without interaction.

### Correlation Coefficient

sourcegroup

In [52]: # 4

```
numeric_data = data.select_dtypes(include=['number'])
corr_mat = numeric_data.corr()
exit_corr = corr_mat['Timeinpage']
print(exit_corr)
```

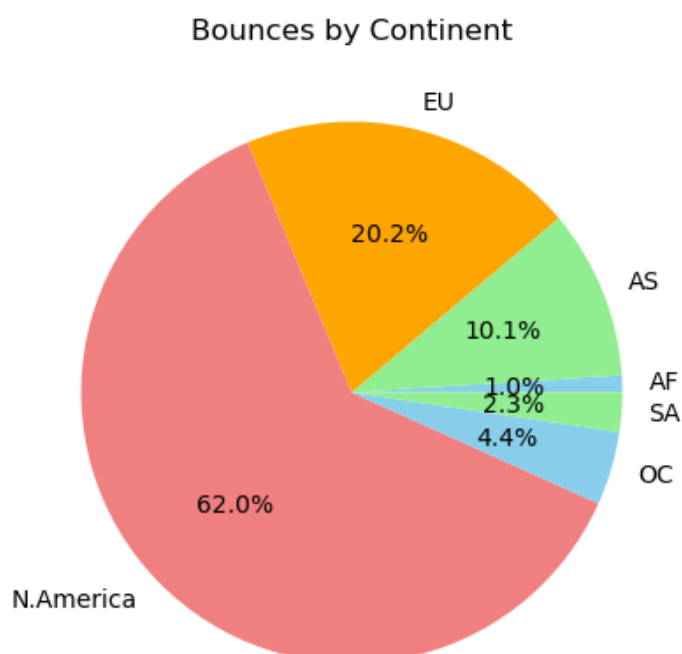
```
Bounces          -0.109106
Exits             0.001325
Timeinpage        1.000000
Uniquepageviews   0.114593
Visits            0.066650
BouncesNew        -0.109106
Name: Timeinpage, dtype: float64
```

---

## Pie Chart for Bounces by Continent

In [60]: # 5

```
plt.figure(figsize=(5, 5))
continent_bounces = data.groupby('Continent')['Bounces'].sum()
continent_bounces.plot(kind='pie', autopct='%1.1f%%', colors=['skyblue', 'lightg
plt.title('Bounces by Continent')
plt.ylabel('')
plt.show()
```



## Average Source Rate By SourceGroup

```
In [57]: # 5
plt.figure(figsize=(10, 6))
data.groupby('Sourcegroup')['Bounces'].mean().plot(kind='bar', color='purple')
plt.title('Average Bounce Rate by Sourcegroup')
plt.xlabel('Sourcegroup')
plt.ylabel('Average Bounce Rate')
plt.show()
```

