AICP Internship Task Week 5

In [1]:

```
import pandas as pd
import plotly.express as px
import plotly.graph_objects as go
import plotly.io as pio
```

In [2]:

```
pio.templates.default = "plotly_white"
```

In [3]:

```
data = pd.read_csv("Instagram data.csv")
```

Q.1: Show column names and have a look at their info.

In [4]:

```
print("Column Names:")
print(data.columns)
print("\nInfo:")
print(data.info())
```

```
Column Names:
Index(['Unnamed: 0', 'S.No', 'USERNAME', 'Caption', 'Followers',
'Hashtags',
       'Time since posted', 'Likes'],
      dtype='object')

Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 8 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   Unnamed: 0         100 non-null    int64
 1   S.No               100 non-null    int64
 2   USERNAME           100 non-null    object
 3   Caption            94 non-null     object
 4   Followers          100 non-null    int64
 5   Hashtags           100 non-null    object
 6   Time since posted  100 non-null    object
 7   Likes              100 non-null    int64
dtypes: int64(4), object(4)
memory usage: 6.4+ KB
None
```

Q.2: Show the descriptive statistics of the data.

```
In [5]: print("\nDescriptive Statistics:")
        print(data.describe())
```

```
Descriptive Statistics:
       Unnamed: 0        S.No   Followers       Likes
count  100.000000  100.000000   100.00000   100.00000
mean     8.940000   16.240000   961.96000    46.48000
std      6.639064    7.384286  1014.62567    55.08698
min      0.000000    1.000000    11.00000     8.00000
25%      4.000000   10.750000   252.75000    19.00000
50%      8.000000   16.500000   612.00000    29.00000
75%     12.250000   22.250000  1197.00000    46.00000
max     26.000000   30.000000  4496.00000   349.00000
```

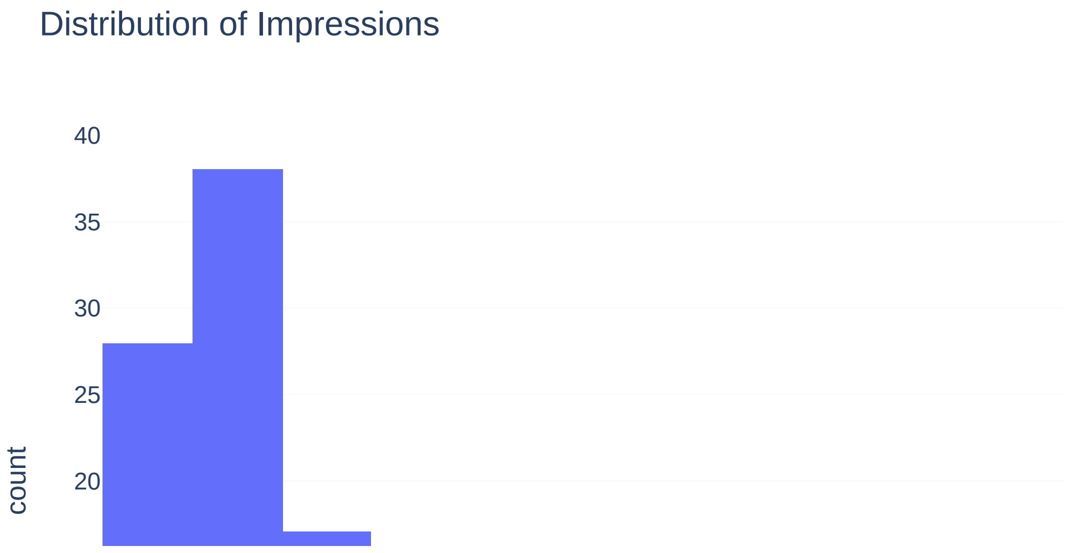
Q.3: Check if your data contains any missing values

```
In [6]: print("\nMissing Values:")
        print(data.isnull().sum())
```

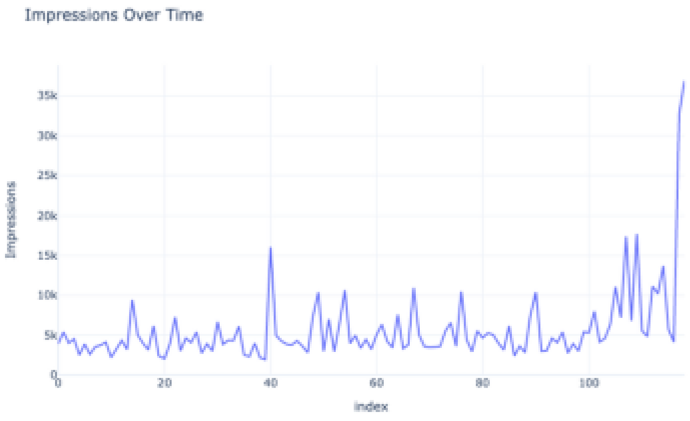
```
Missing Values:
Unnamed: 0           0
S.No                 0
USERNAME             0
Caption              6
Followers            0
Hashtags             0
Time since posted    0
Likes                0
dtype: int64
```

Q.4: When you start exploring your data, always start by exploring the main feature of your data. For example, as we are working on a dataset based on Instagram Reach, we should start by exploring the feature that contains data about reach. In our data, the Impressions column contains the data about the reach of an Instagram post. So let’s have a look at the distribution of the Impressions:

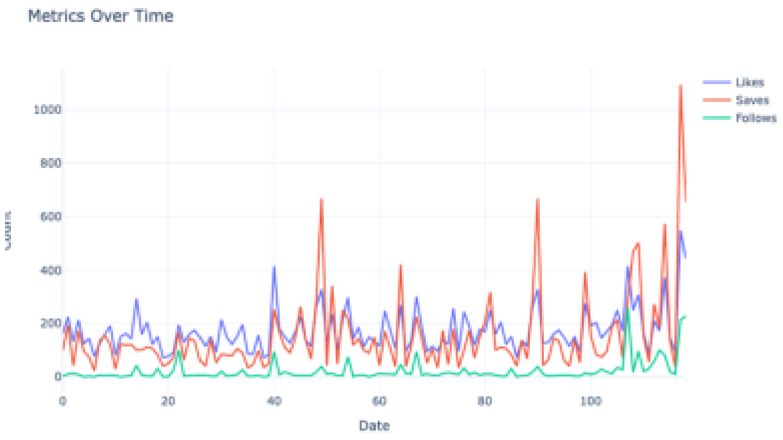
```
In [9]: fig = px.histogram(data, x="Likes", title="Distribution of Impressi
        fig.show()
```



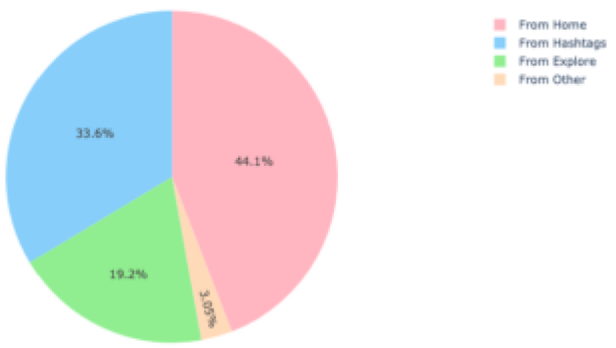
Q.5: Have a look at the number of impressions on each post over time as shown below



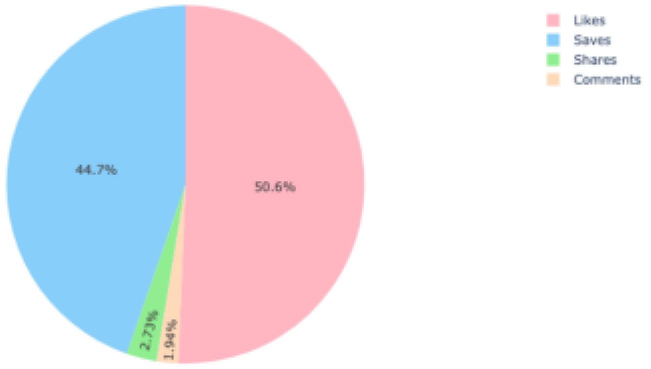
Q.6: Have a look at all the metrics like Likes, Saves, and Follows from each post over time as shown below.



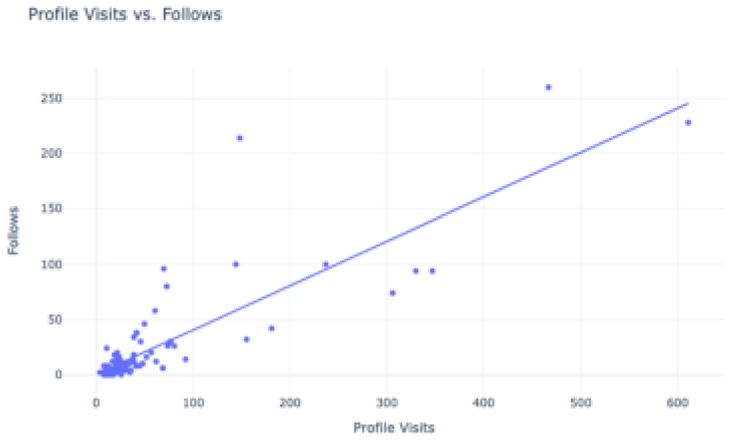
Q.7: Have a look at the distribution of reach from different sources as shown below



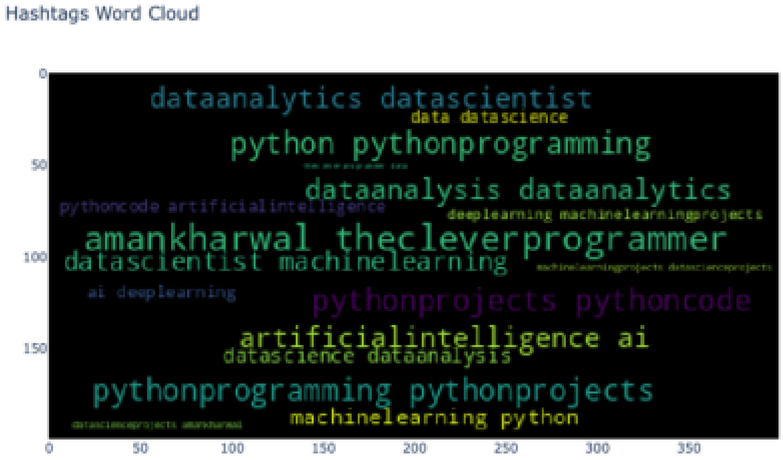
Q.8: Have a look at the distribution of engagement sources as shown below



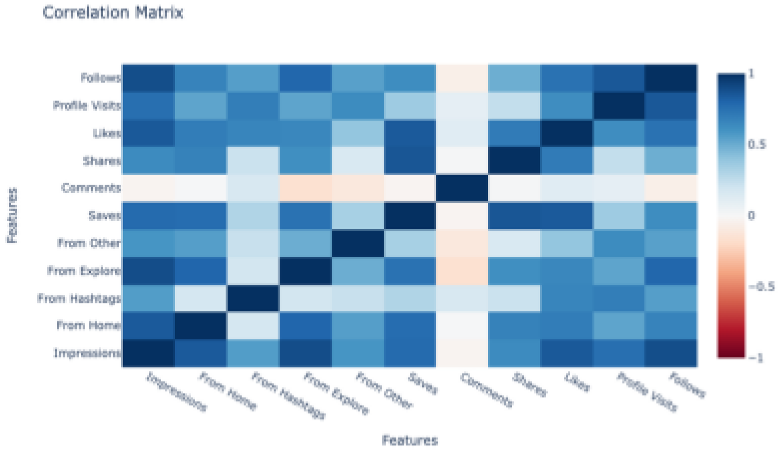
Q.9: Have a look at the relationship between the number of profile visits and follows as shown below:



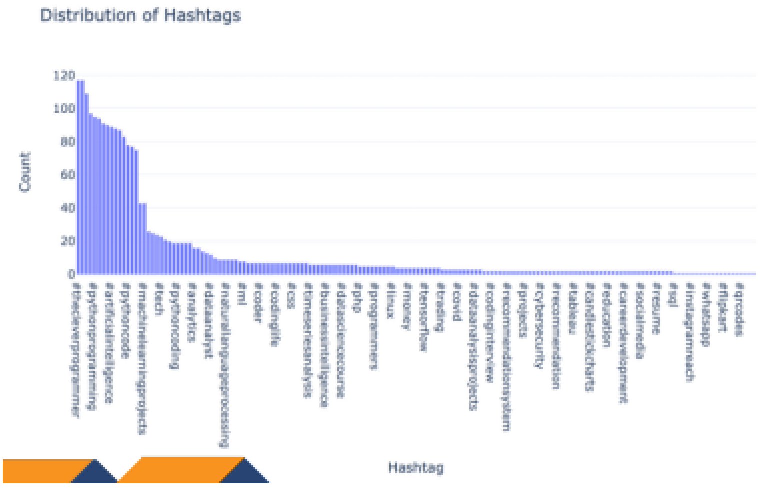
Q.10: Have a look at the type of hashtags used in the posts using a wordcloud as shown below:



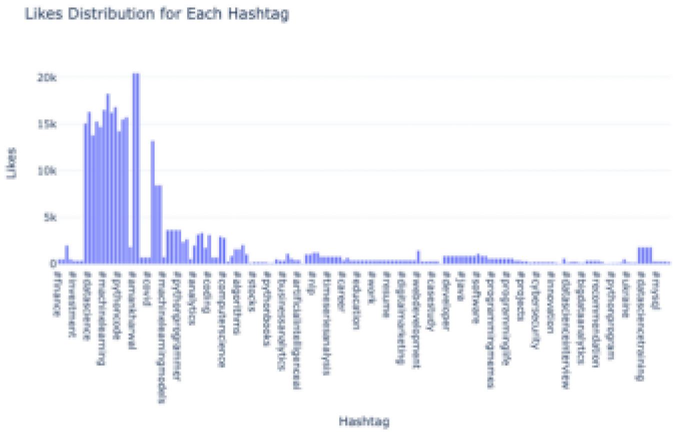
Q.11: Have a look at the correlation between all the features as shown below

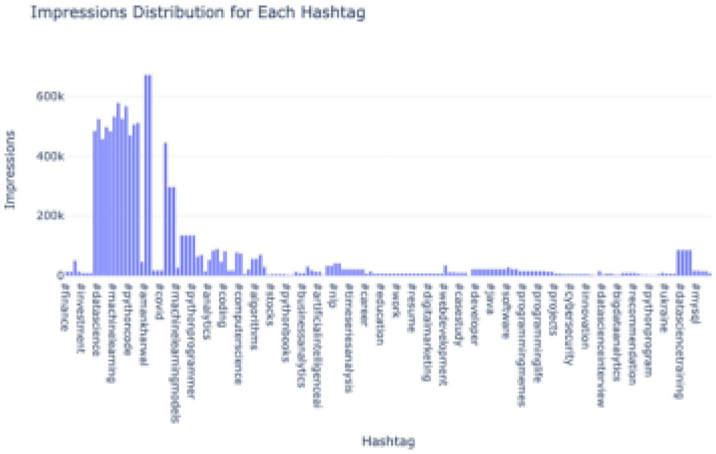


Q.12: Havea look at the distribution of hashtags to see which hashtag is used the most in all the posts as shown below:



Q.13: Have a look at the distribution of likes and impressions received from the presence of each hashtag on the post as shown below:


Likes Distribution for Each Hashtag
Likes
20k
15k
10k
5k
#finance
#investment
#datascience
#machinelearning
#pythoncode
#amankharwal
#covid
#machinelearningmodels
#pythonprogrammer
#analytics
#coding
#computerscience
#algorithms
#stocks
#pythonbooks
#businessanalytics
#artificialintelligenceai
#nlp
#timeseriesanalysis
#career
#education
#work
#resume
#digitalmarketing
#webdevelopment
#casestudy
#developer
#java
#software
#programmingmemes
#programminglife
#projects
#cybersecurity
#innovation
#datascienceinterview
#bigdataanalytics
#recommendation
#pythonprogram
#ukraine
#datasciencetraining
#mysql
Hashtag


Impressions Distribution for Each Hashtag
Impressions
600k
400k
200k
#finance
#investment
#datascience
#machinelearning
#pythoncode
#amankharwal
#covid
#machinelearningmodels
#pythonprogrammer
#analytics
#coding
#computerscience
#algorithms
#stocks
#pythonbooks
#businessanalytics
#artificialintelligenceai
#nlp
#timeseriesanalysis
#career
#education
#work
#resume
#digitalmarketing
#webdevelopment
#casestudy
#developer
#java
#software
#programmingmemes
#programminglife
#projects
#cybersecurity
#innovation
#datascienceinterview
#bigdataanalytics
#recommendation
#pythonprogram
#ukraine
#datasciencetraining
#mysql
Hashtag