

In [1]: `pip install wordcloud`

```
Requirement already satisfied: wordcloud in c:\users\cafeb\anaconda3\lib\site-packages (1.9.3)
Requirement already satisfied: pillow in c:\users\cafeb\anaconda3\lib\site-packages (from wordcloud) (9.2.0)
Requirement already satisfied: numpy>=1.6.1 in c:\users\cafeb\anaconda3\lib\site-packages (from wordcloud) (1.21.5)
Requirement already satisfied: matplotlib in c:\users\cafeb\anaconda3\lib\site-packages (from wordcloud) (3.5.2)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\cafeb\anaconda3\lib\site-packages (from matplotlib->wordcloud) (4.25.0)
Requirement already satisfied: packaging>=20.0 in c:\users\cafeb\anaconda3\lib\site-packages (from matplotlib->wordcloud) (21.3)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\cafeb\anaconda3\lib\site-packages (from matplotlib->wordcloud) (1.4.2)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\cafeb\anaconda3\lib\site-packages (from matplotlib->wordcloud) (2.8.2)
Requirement already satisfied: cycler>=0.10 in c:\users\cafeb\anaconda3\lib\site-packages (from matplotlib->wordcloud) (0.11.0)
Requirement already satisfied: pyparsing>=2.2.1 in c:\users\cafeb\anaconda3\lib\site-packages (from matplotlib->wordcloud) (3.0.9)
Requirement already satisfied: six>=1.5 in c:\users\cafeb\anaconda3\lib\site-packages (from python-dateutil->matplotlib->wordcloud) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

In [2]: `import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
from sklearn.model_selection import train_test_split
from sklearn.linear_model import PassiveAggressiveRegressor`

```
In [3]: df=pd.read_csv('Instagram data.csv',encoding = 'latin1')
```

In [4]: df

Out[4]:

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits	Follows	Caption
0	3920	2586	1028	619	56	98	9	5	162	35	2	Here are some of the most important data visua... #finance #money #
1	5394	2727	1838	1174	78	194	7	14	224	48	10	Here are some of the best data science project... #healthcare #hea
2	4021	2085	1188	0	533	41	11	1	131	62	12	Learn how to train a machine learning model an... #data #datascien
3	4528	2700	621	932	73	172	10	7	213	23	8	Here's how you can write a Python program to d... #python #pythonpro
4	2518	1704	255	279	37	96	5	4	123	8	0	Plotting annotations while visualizing your da... #datavisualization
...
114	13700	5185	3041	5352	77	573	2	38	373	73	80	Here are some of the best data science certifi... #datascience #da
115	5731	1923	1368	2266	65	135	4	1	148	20	18	Clustering is a machine learning technique use... #machinelearning #

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits	Follows	Caption
116	4139	1133	1538	1367	33	36	0	1	92	34	10	Clustering music genres is a task of grouping ... #machinelearning #
117	32695	11815	3147	17414	170	1095	2	75	549	148	214	Here are some of the best data science certifi... #datascience #dai
118	36919	13473	4176	16444	2547	653	5	26	443	611	228	175 Python Projects with Source Code solved an... #python #pythonpro

119 rows × 13 columns

```
In [5]: df.isnull().sum()
```

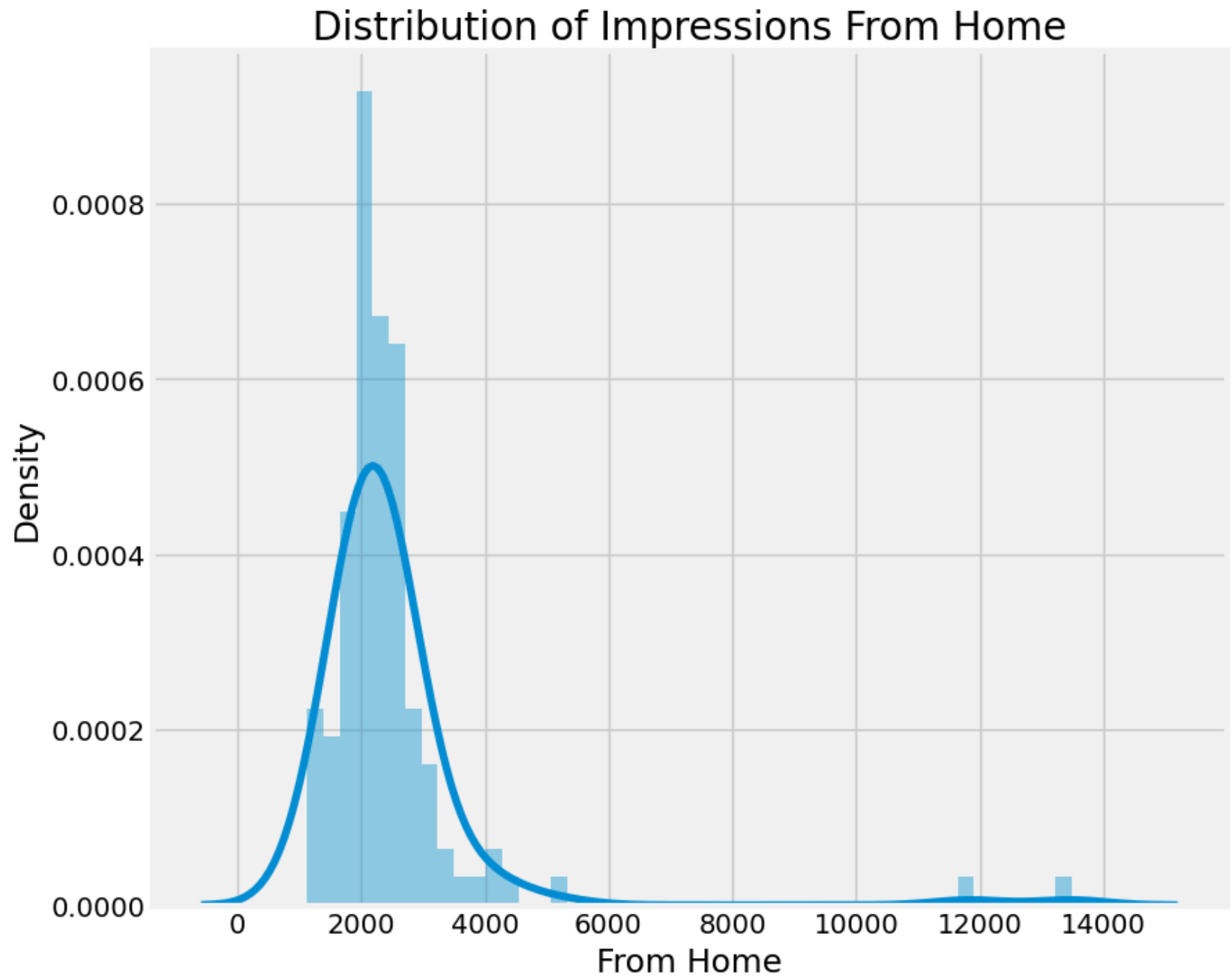
```
Out[5]: Impressions      0
        From Home      0
        From Hashtags  0
        From Explore   0
        From Other     0
        Saves          0
        Comments       0
        Shares         0
        Likes          0
        Profile Visits 0
        Follows        0
        Caption        0
        Hashtags       0
        dtype: int64
```

```
In [6]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 119 entries, 0 to 118
Data columns (total 13 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Impressions           119 non-null    int64
1   From Home             119 non-null    int64
2   From Hashtags         119 non-null    int64
3   From Explore          119 non-null    int64
4   From Other            119 non-null    int64
5   Saves                 119 non-null    int64
6   Comments              119 non-null    int64
7   Shares               119 non-null    int64
8   Likes                 119 non-null    int64
9   Profile Visits        119 non-null    int64
10  Follows               119 non-null    int64
11  Caption               119 non-null    object
12  Hashtags              119 non-null    object
dtypes: int64(11), object(2)
memory usage: 12.2+ KB
```

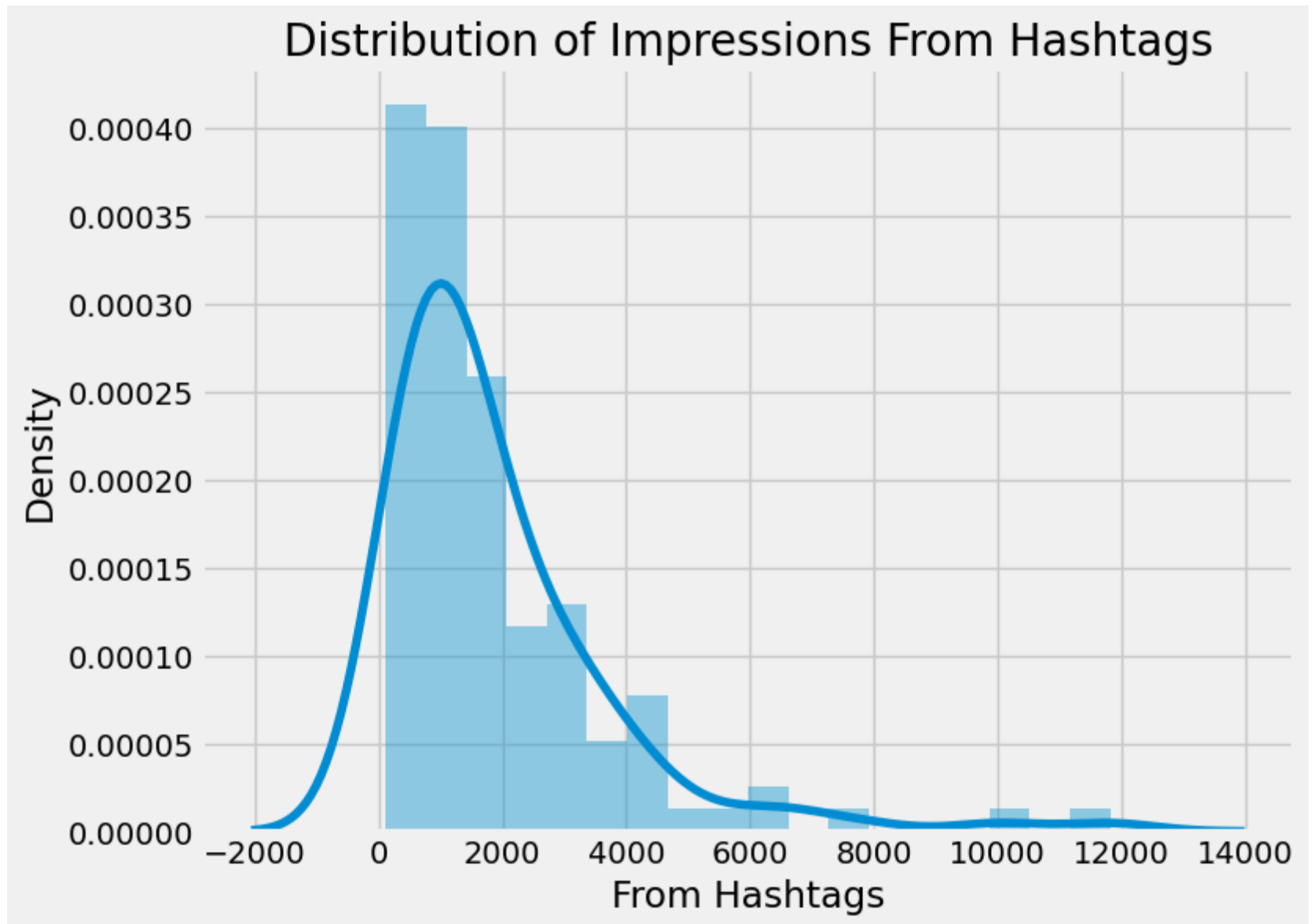
```
In [7]: plt.figure(figsize=(10, 8))
plt.style.use('fivethirtyeight')
plt.title("Distribution of Impressions From Home")
sns.distplot(df['From Home'])
plt.show()
```

C:\Users\cafeb\anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
warnings.warn(msg, FutureWarning)



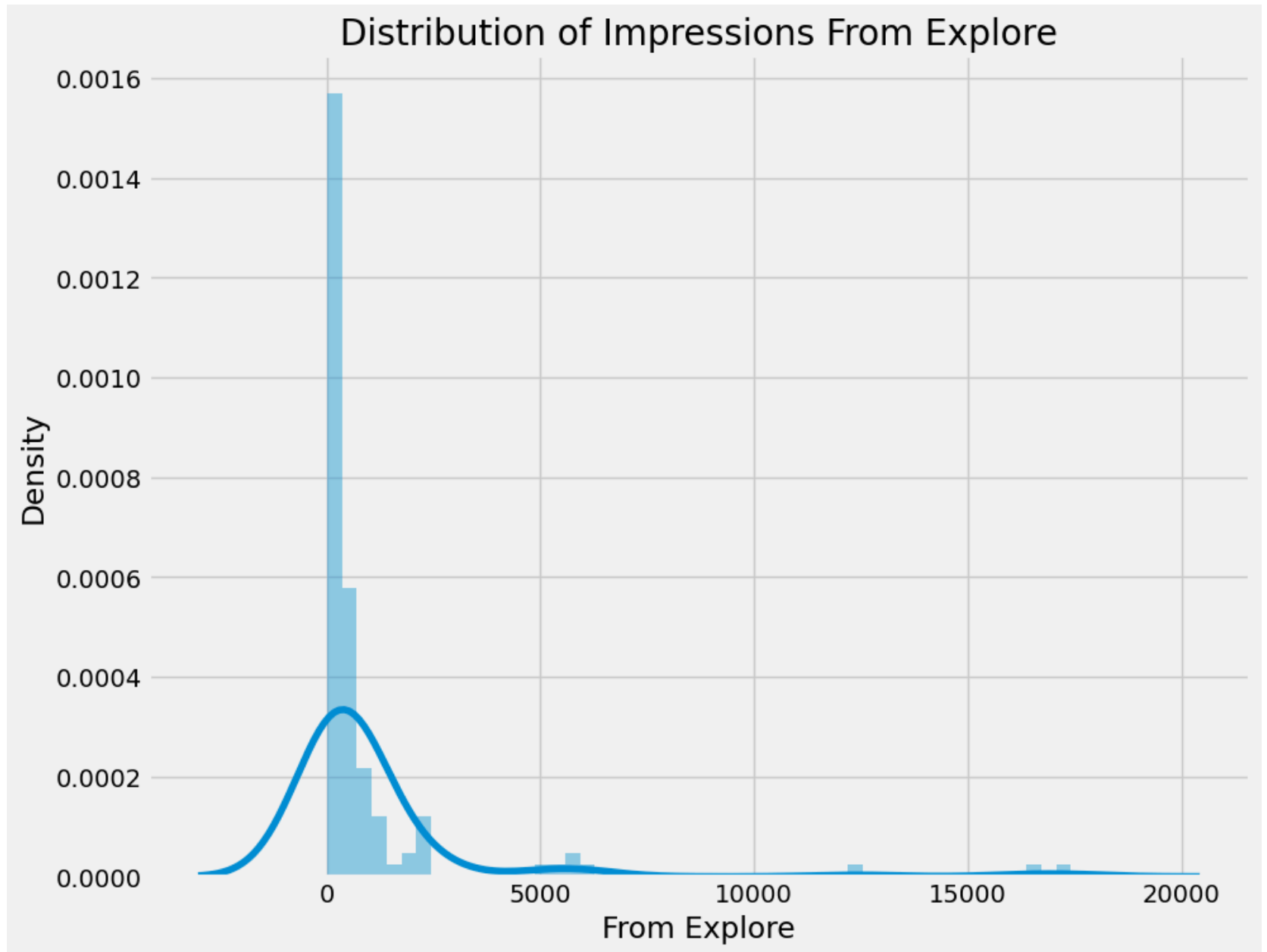

```
In [8]: plt.figure(figsize=(8,6))  
plt.title('Distribution of Impressions From Hashtags')  
sns.distplot(df['From Hashtags'])  
plt.show()
```

C:\Users\cafeb\anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
warnings.warn(msg, FutureWarning)

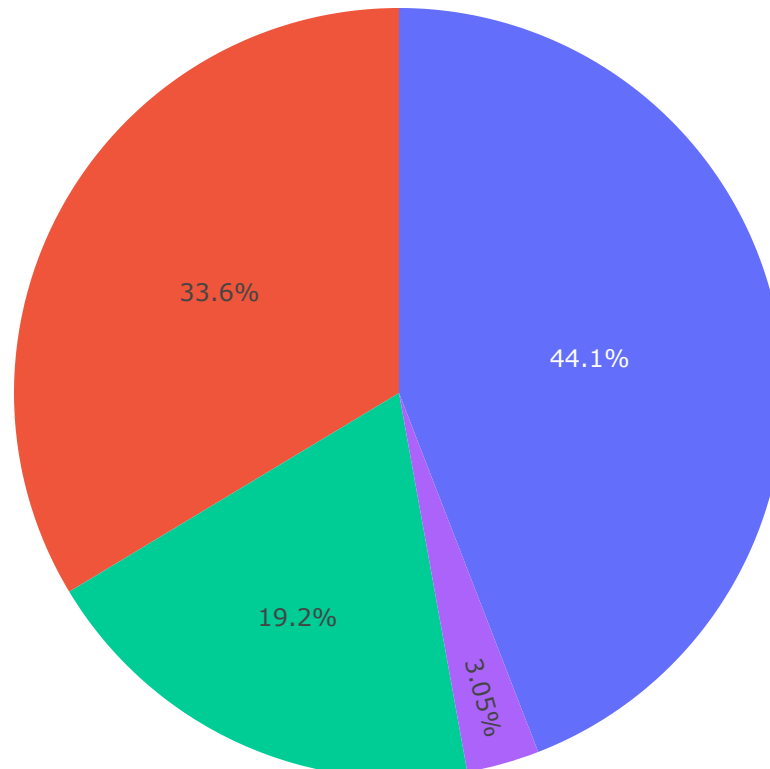


```
In [9]: plt.figure(figsize=(10, 8))  
plt.title("Distribution of Impressions From Explore")  
sns.distplot(df['From Explore'])  
plt.show()
```

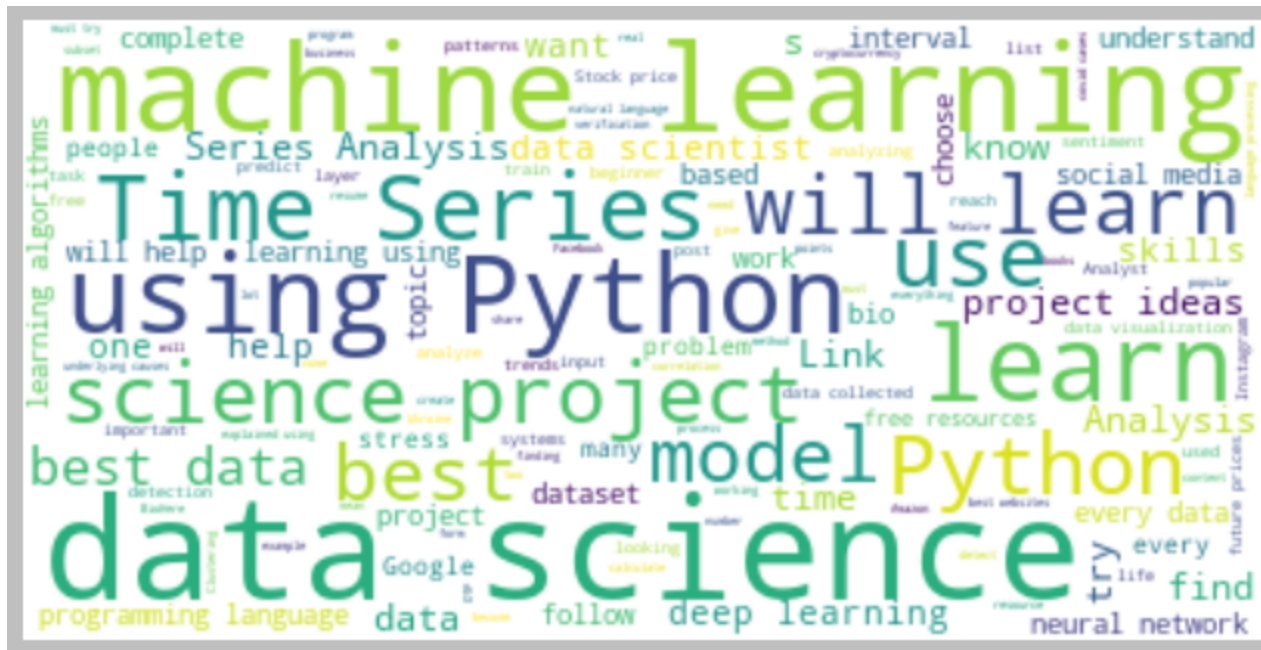
C:\Users\cafeb\anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
warnings.warn(msg, FutureWarning)



```
In [10]: home=df['From Home'].sum()  
hashtag= df['From Hashtags'].sum()  
explore=df['From Explore'].sum()  
other=df['From Other'].sum()  
labels= ['From Home','From Hashtag','From Explore','From Other']  
values=[home,hashtag,explore,other]  
fig=px.pie(df,values=values ,names=labels)  
fig.show()
```



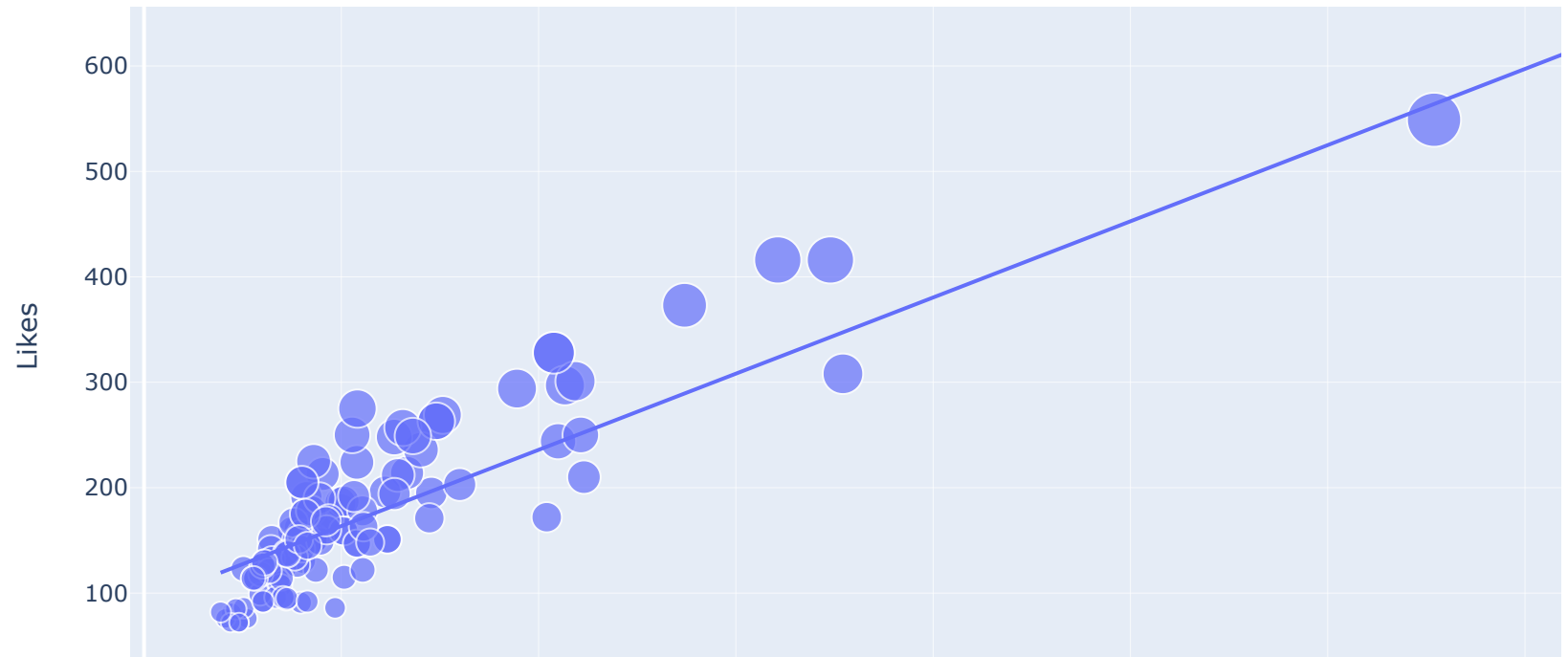
```
In [11]: text= "".join(i for i in df.Caption)
stopword=set(STOPWORDS)
wordcloud=WordCloud(stopwords=stopword,background_color='white').generate(text)
plt.style.use('classic')
plt.figure(figsize=(10,8))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.show()
```



```
In [12]: text= "".join(i for i in df.Hashtags)
stopwords=set(STOPWORDS)
wordcloud = WordCloud(stopwords=stopwords, background_color="white").generate(text)
plt.figure(figsize=(12,10))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```

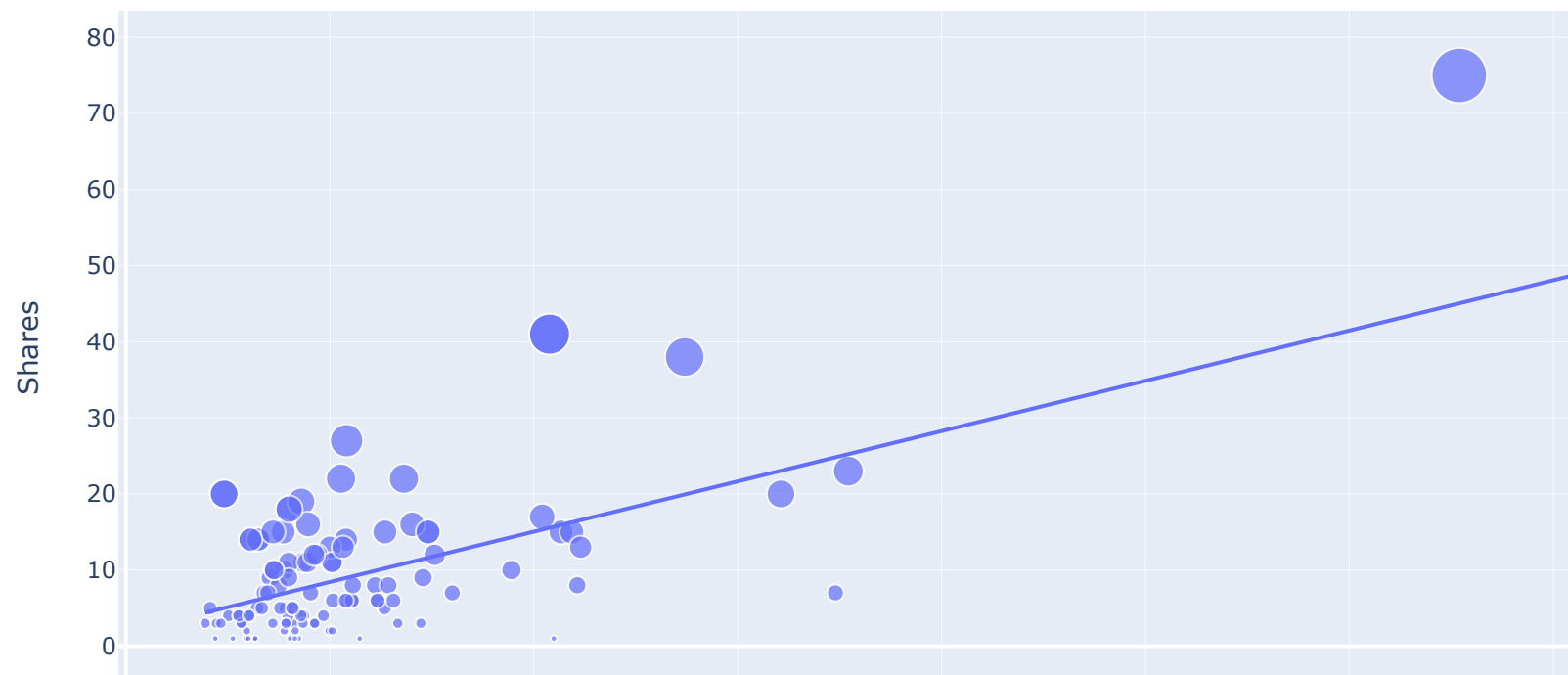
```
In [13]: figure= px.scatter(data_frame= df, x="Impressions", y="Likes", size="Likes", trendline="ols",title='Relation:  
figure.show()
```

Relationship Between Likes and Impress

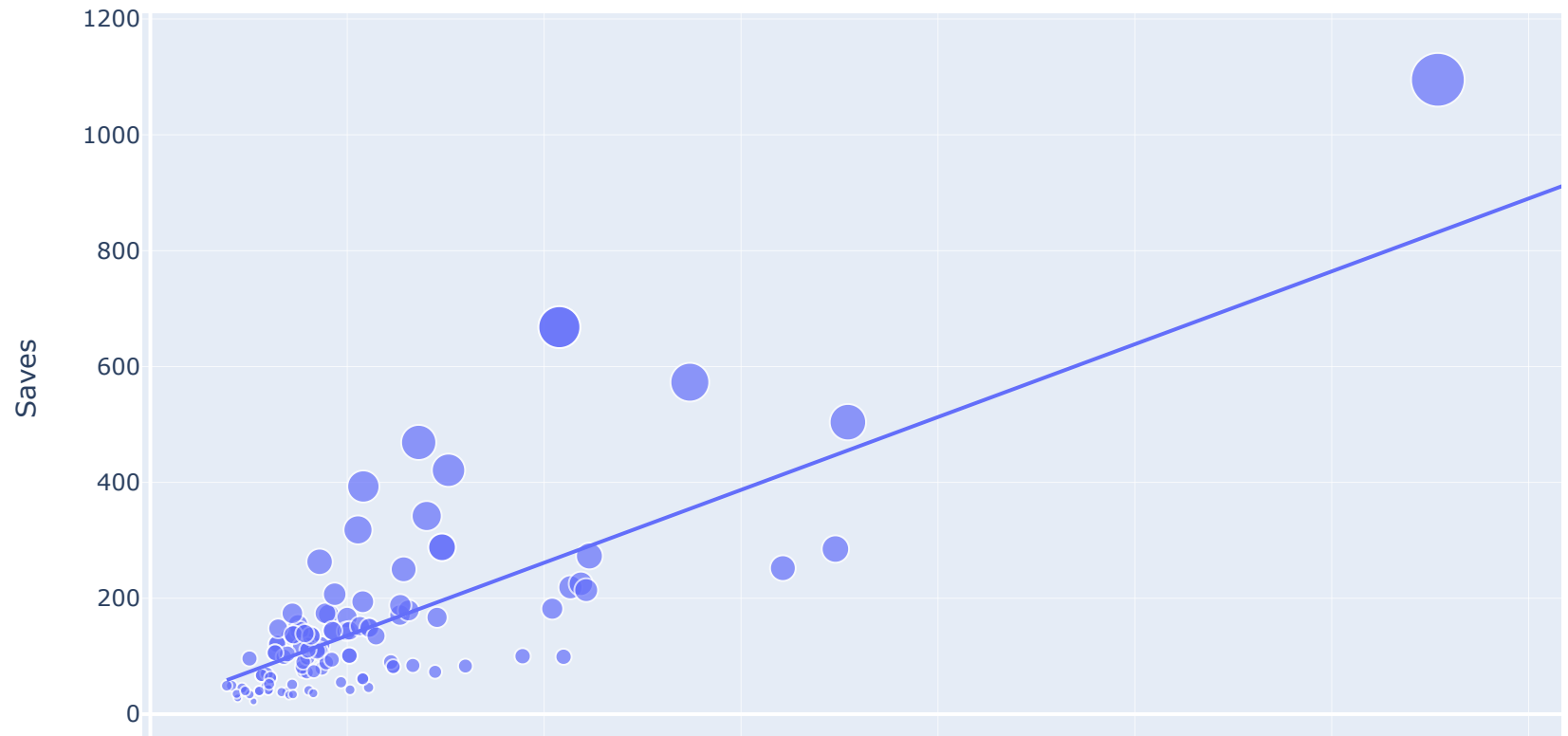



```
In [14]: figure= px.scatter(data_frame=df,x="Impressions", y="Shares",size="Shares",trendline="ols",title="relationship  
figure.show()
```

relationship between shares and impression



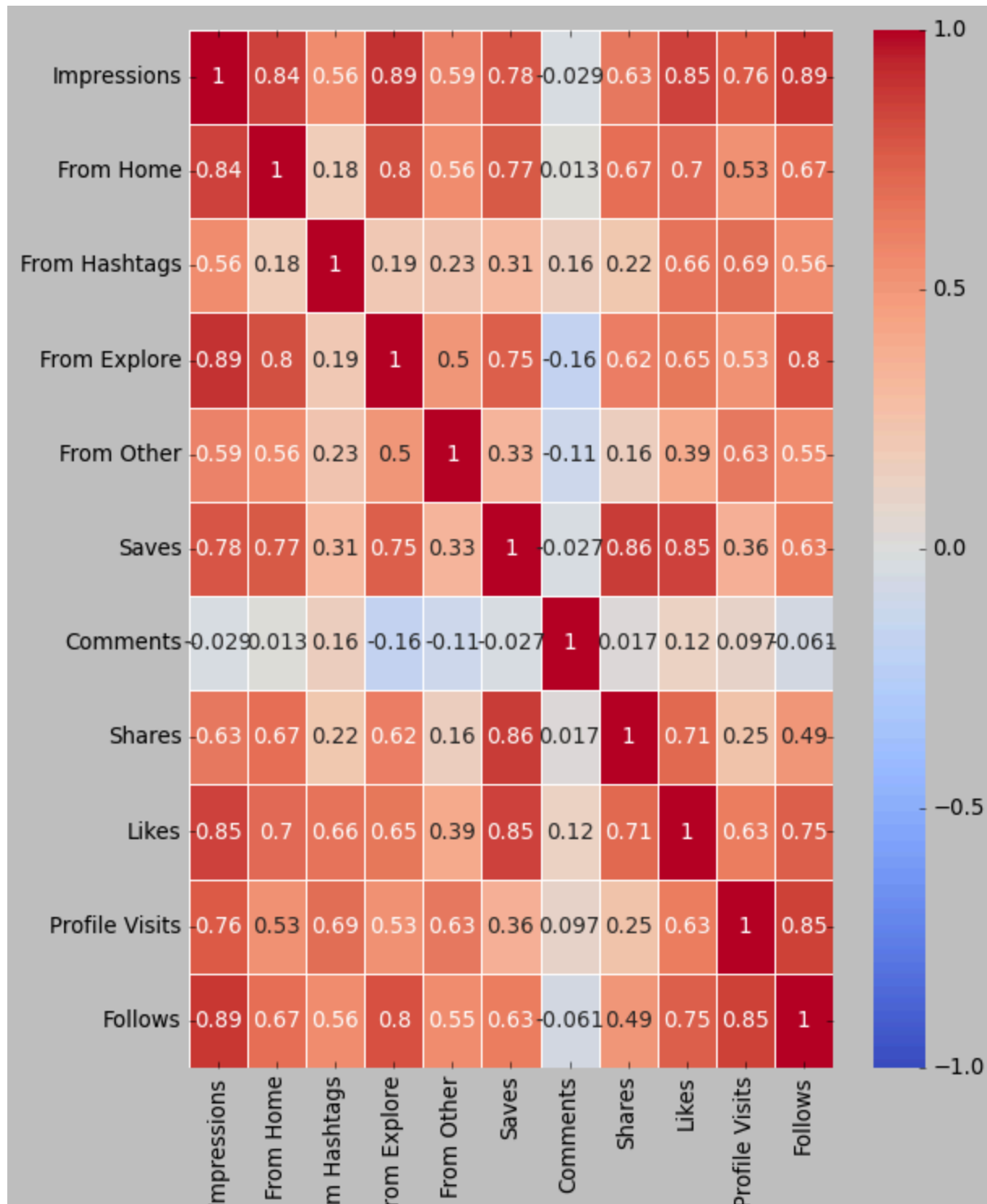
```
In [15]: figure= px.scatter(data_frame=df, x="Impressions", y="Saves", size="Saves", trendline="ols")  
figure.show()
```



```
In [16]: correlation=df.corr()  
  
print(correlation["Impressions"].sort_values(ascending=False))
```

```
Impressions      1.000000  
From Explore     0.893607  
Follows          0.889363  
Likes            0.849835  
From Home        0.844698  
Saves            0.779231  
Profile Visits   0.760981  
Shares           0.634675  
From Other       0.592960  
From Hashtags    0.560760  
Comments         -0.028524  
Name: Impressions, dtype: float64
```

```
In [17]: plt.figure(figsize=(8,10))  
sns.heatmap(correlation, annot=True, cmap='coolwarm', vmin=-1, vmax=1, linewidths=0.5)  
plt.show()
```

Fro
F

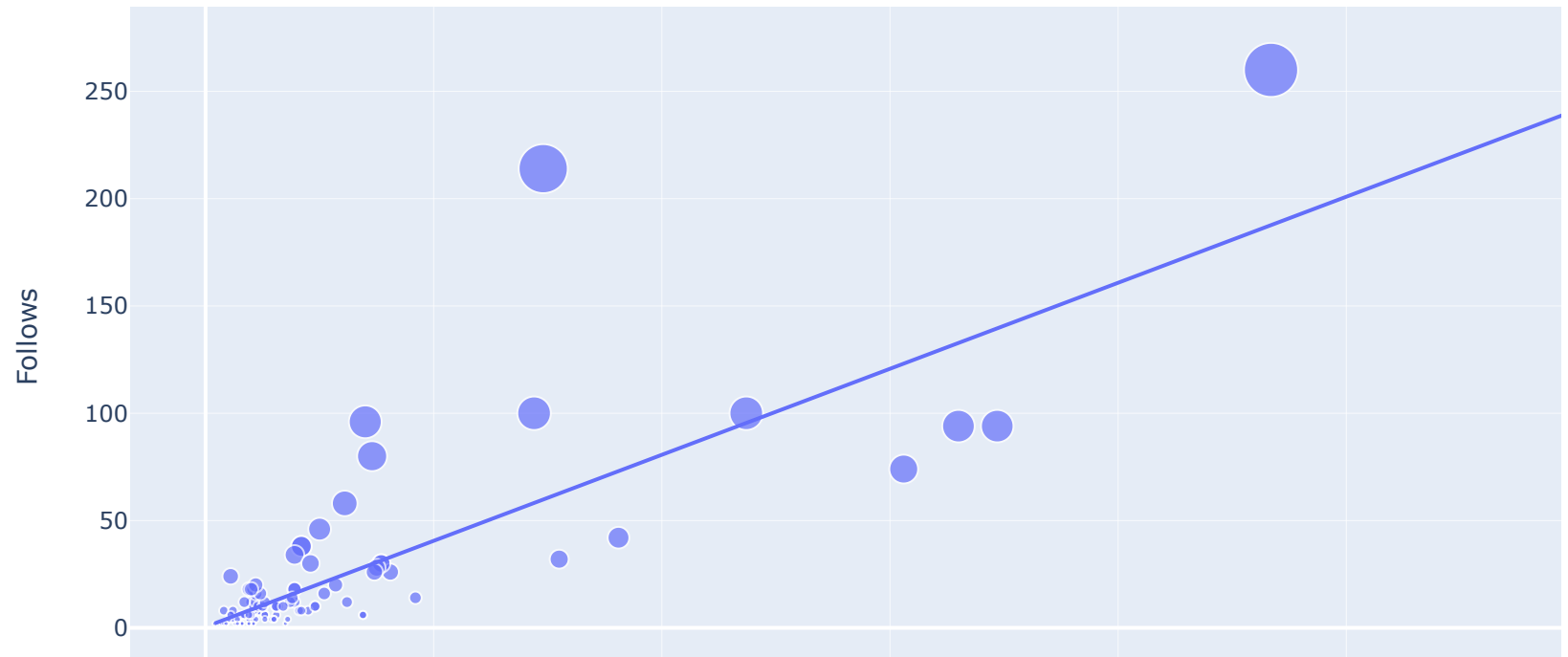
Analyzing Conversion Rate

```
In [18]: conversion_rate = (df["Follows"].sum() / df["Profile Visits"].sum()) * 100  
print(conversion_rate)
```

```
41.00265604249668
```

```
In [19]: figure= px.scatter(data_frame=df,x="Profile Visits", y="Follows",size="Follows",trendline="ols",title = "Relationship Between Profile Visits and Followers Gained")  
figure.show()
```

Relationship Between Profile Visits and Followers Gained



Instagram Reach Prediction Model


```
In [20]: x=np.array(df[['Likes', 'Saves', 'Comments', 'Shares', 'Profile Visits', 'Follows']])  
y=np.array(df['Impressions'])
```

```
In [21]: xtrain,xtest,ytrain,ytest=train_test_split(x,y,test_size=0.2, random_state=42)  
model=PassiveAggressiveRegressor()  
model.fit(xtrain,ytrain)  
model.score(xtest,ytest)
```

Out[21]: 0.8203807510667341

```
In [22]: #prediction  
# Features = [['Likes', 'Saves', 'Comments', 'Shares', 'Profile Visits', 'Follows']]  
features=np.array([[282.0, 233.0, 4.0, 9.0, 165.0, 54.0]])  
model.predict(features)
```

Out[22]: array([9413.55203892])

In []: