DS 710 Final Project Part 3

Submitted by Your Name

Research question:

This project explores the tweets related to word given by user such as NASA, Machine Learning, Artifical Intelligence etc. Furthermore, I got some impactful insights from recent tweets such as getting hashtags which are most trending among all tweets with their number of occurence in descending order. I have extracted number of tweets categorized by their types. These visualizations are very efficient when it comes to find trending items or topics across the globe. We can predict meaningful trait's density as well as concentration in tweets.

Data collection:

This fabulous model is designed by collecting data from Twitter via the Twitter API using the Tweepy library. First user gives a query word and a json file containig recent tweets related to that word is fetched out from tweeter. this model is fetching out 300 recent tweets and stores in json file. then for hashtag-extraction I have filtered out tweets that does not contain any hashtag using Pandas library. here, i have created scenario for the word "Machine Learning". In below base code, we got top five hashtag names with its number of occurance.

In [1]:

```
## call Python data processing code

import warnings
warnings.filterwarnings('ignore')

from analyze_data_project_part3 import *

## call Python visualization code

from visualize_part3 import plot_hashtags, plot_tweettype
```

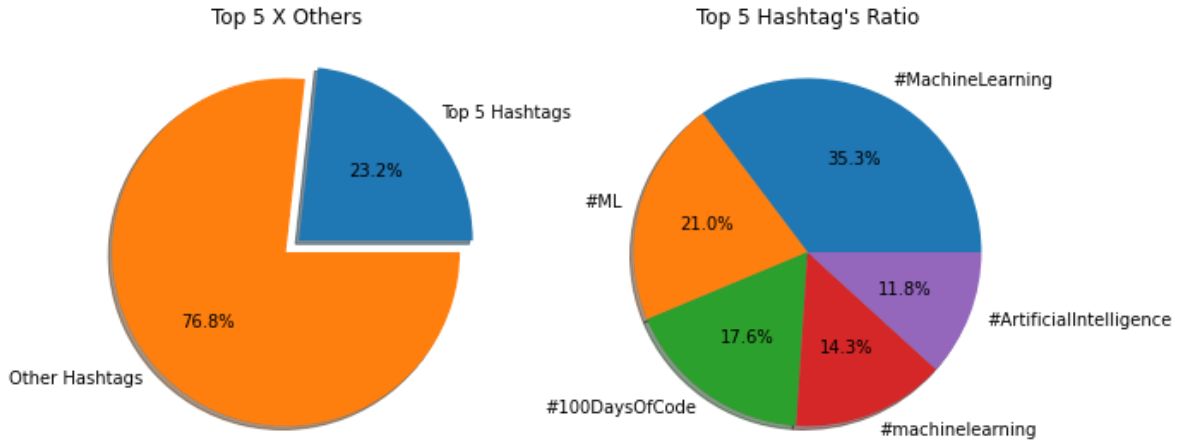
```
Top 5 Hashtags:
MachineLearning : 42
ML : 25
100DaysOfCode : 21
machinelearning : 17
Trading : 14
```

Discussion:

I have visualized pie graph which shows Proportional percentage of top 5 hashtags versus rest of other hashtags. then I have visualized pie graph representing the proportion of top 5 hashtag's occurence which is showing comparison of tweet trends in a very beutiful manner. I have animated these graphs by using shadow effect and fragment popping out from the pie. By these visualisations I hypothesize that top 5 trending hashtags are occupying 23.2% of total. Moreover,"#MachineLearning" is the most trended hashtage in all tweets.

In [2]:

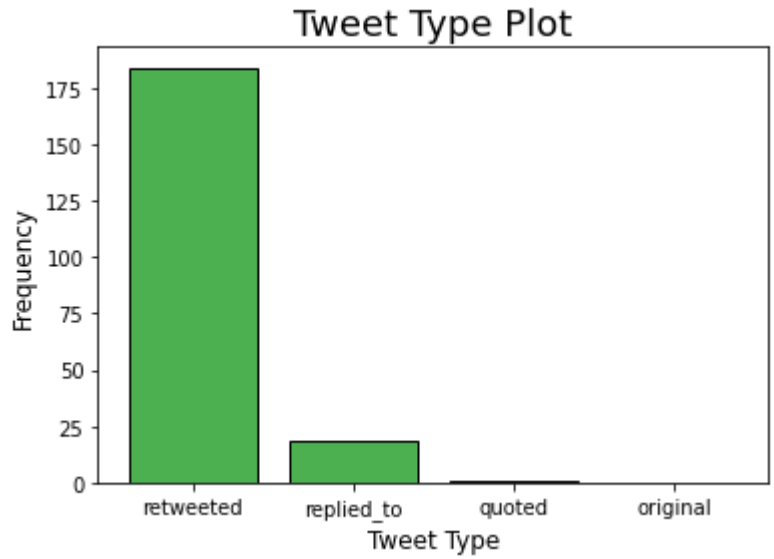
```
plot_hashtags();
```



In Tweet type plot bar graph, we are having tweet type on X-axis and Frequency on Y-axis. The most type tweets are retweeted tweets which are around 175 out 300 and there is not a single orignal tweet in present dataset as of now.

In [3]:

```
plot_tweettype();
```



Conclusion:

In this data analytics model,I have more focused on the basement situational scenario which can be more extended or used towards predicting recent market trends and the ideological direction towards which market is moving in.I can also find out how much the particular item is spread out across the globe in terms of depth as well as breadth.We also get to know the most closest hashtags from any given word.