

PROJECT  
INSTABOT  
(PART – II)

PLOTS AND EXPLANATION

By :

PAWAN KR. MISHRA

### **Question 1:**

Now your friend has followed a lot of different food bloggers, he needs to analyze the habits of these bloggers.

- a.) From the list of Instagram handles you obtained when you searched 'food' in previous project. Open the first 10 handles and find the top 5 which have the highest number of followers

### **Explanation:**

#### **Given information:**

On searching for 'food' we get list of handles , from that list we should open the profile of first 10 handles and get the top 5 handles which are having maximum count of followers.

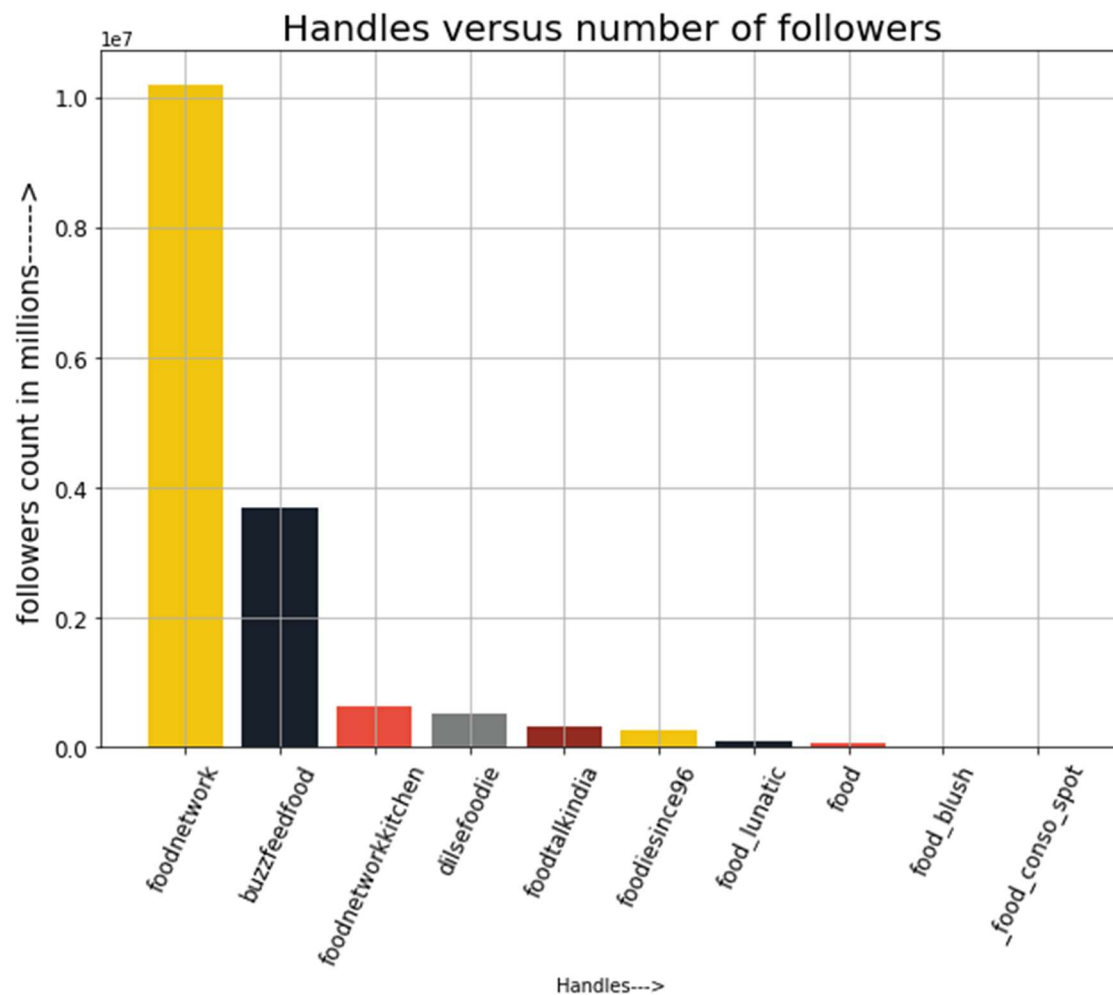
#### **Steps performed to get the Top 5 handles with highest number of followers:**

- After searching for 'food' we get list of Instagram handles that posts contents related to food.
- From that list I have visited the first 10 handles and stored their count of followers in a dictionary with handle names as key and count of follower as value.
- Since the followers are in thousands or millions so they were represented by alphabet so I converted them into integer value so that comparison would be easy.
- In this step I have sorted the dictionary on the count of number of followers i.e., value and then extracted the first 5 handles that has highest count of followers.

## Tables and Graphs:

### Top 5 handles with highest number of followers

Username	Followers
foodnetwork	10200000
buzzfeedfood	3700000
foodnetworkkitchen	615000
dilsefoodie	509000
foodtalkindia	297000



**b) Now Find the number of posts these handles have done in the previous 3 days.**

**Explanation:**

**Given Information:**

The top 5 handles that we have collected in previous questions we are asked to count the number of posts each handle has made in past 3 days.

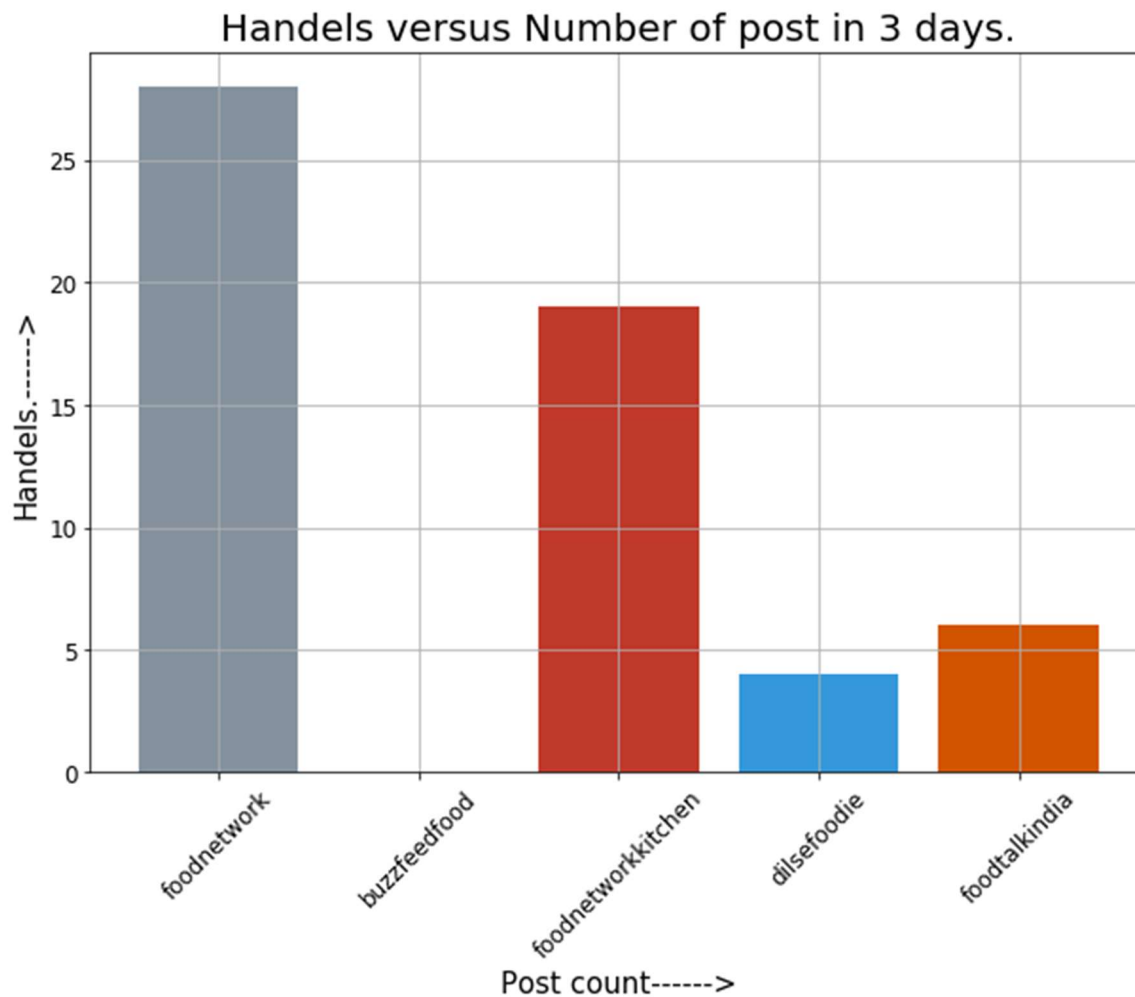
**Steps performed to count the number of posts in last 3 days:**

- For counting the posts done in last 3 days I have visited each post using automation and find out the post details.
- The post can be posted few seconds, few minutes, few hours ago so I have introduced the checks for the same and counted it.
- The second check I have imposed is on the days, if the number of days is greater than 3 then it is older thus not counting that post.
- The above mentioned steps are repeated for entire handles in top 5 handles and respective post count are stored in list.

## Tables and Graphs:

### Number of posts by Top 5 handles in last 3 days

Username	No. of Posts
foodnetwork	28
buzzfeedfood	0
foodnetworkkitchen	19
dilsefoodie	4
foodtalkindia	6



c) Depict this information using a suitable graph.

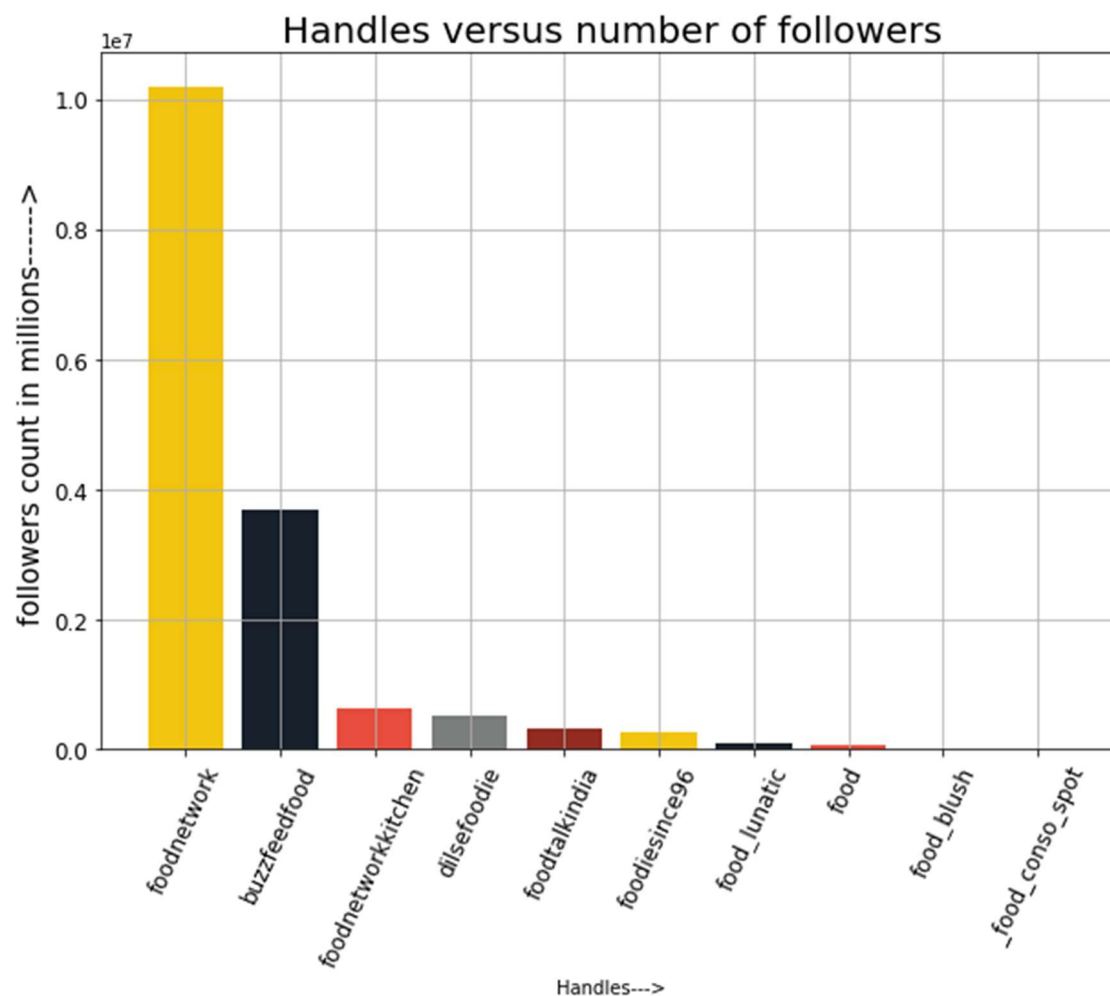
### Explanation:

#### Given Information:

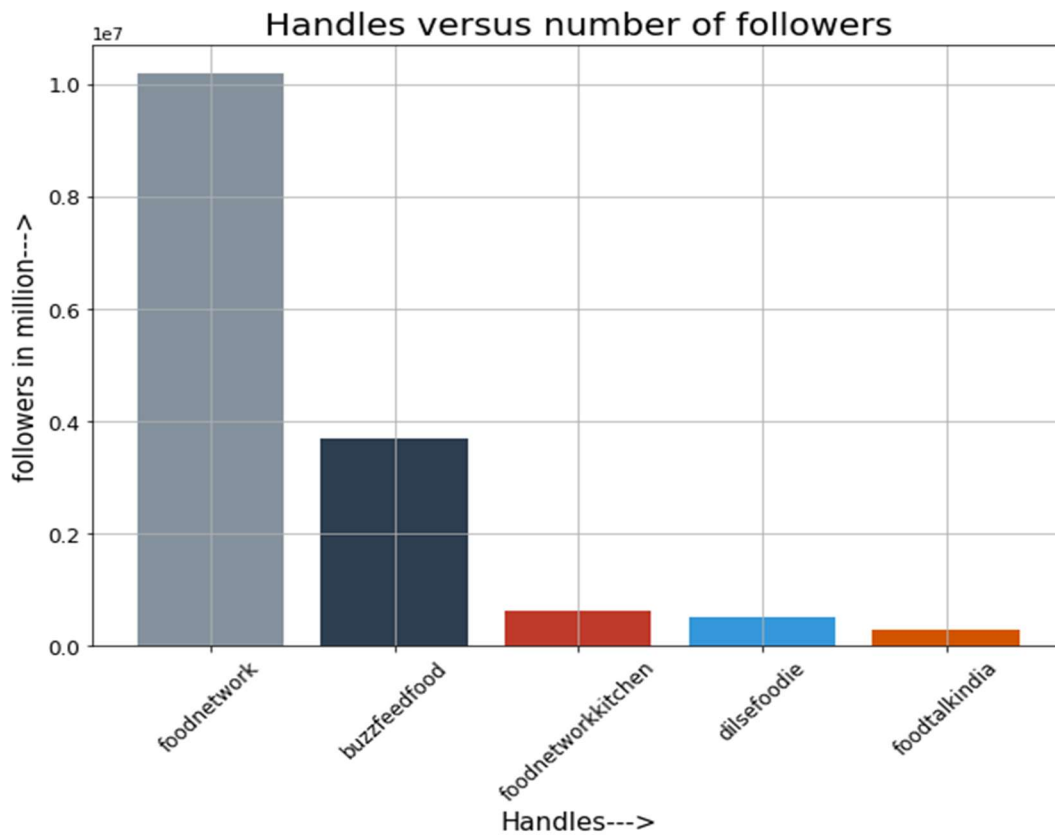
The facts and figures collected above I have to depict this information using graphs.

#### Graphs:

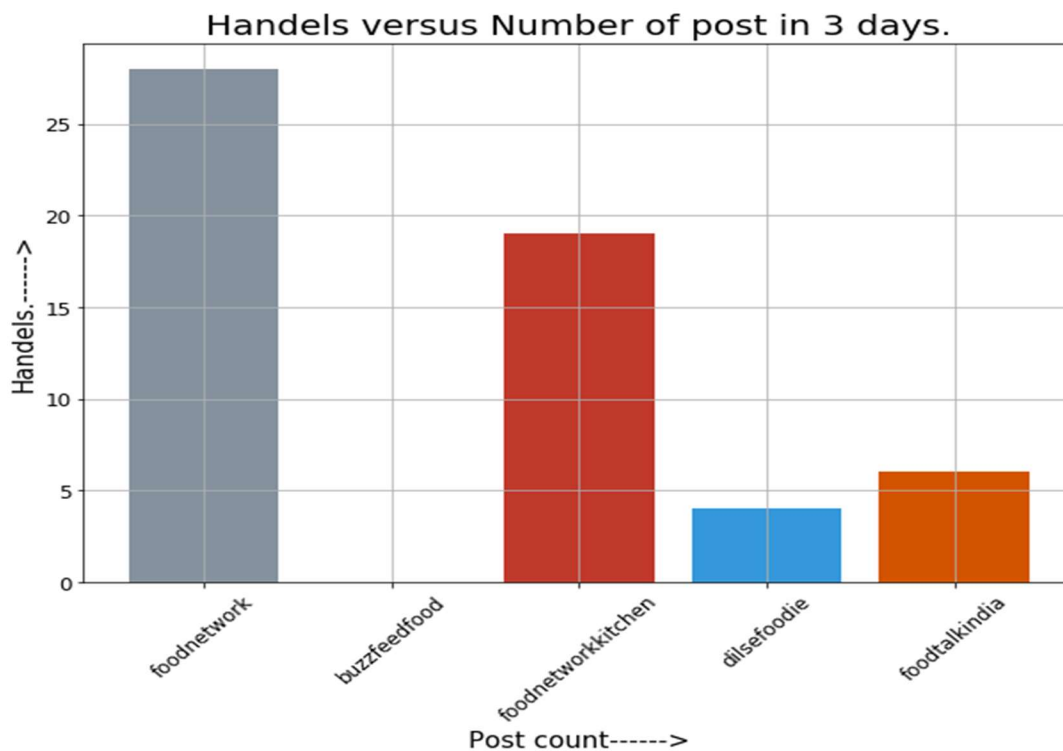
Graph-1: First 10 handles and their followers.



Graph 2: Top 5 handles with maximum number of followers.



Graph -3: Top 5 handles and their posts count of 3 days.



## **Question 2:**

Your friend also needs a list of hashtags that he should use in his posts.

- a. Open the 5 handles you obtained in the last question, and scrape the content of the first 10 posts of each handle.
- b. Prepare a list of all words used in all the scraped posts and calculate the frequency of each word.
- c. Create a csv file with two columns : the word and its frequency.
- d. Now, find the hashtags that were most popular among these bloggers.
- e. Plot a Pie Chart of the top 5 hashtags obtained and the number of times they were used by these bloggers in the scraped posts.

## **Explanation:**

### **Given Information:**

My friend needs popular hashtags that are used by top 5 handles in order to get that I have to traverse over entire content of first 10 posts of each handle and find out which are most frequent used words.

### **Steps performed to get the frequency of each word in content and to get popular hashtags:**

To get the popular words I need to scrape the entire content of top 10 posts, so to store the content and hashtags I have used different arrays.

- Now I will iterate each and every content and note the frequency of each word used in content.
- For noting the frequency I have used dictionary.
- After getting the frequency of each word I have stored the all the words and their respective frequency in words\_frequency.csv file separately.



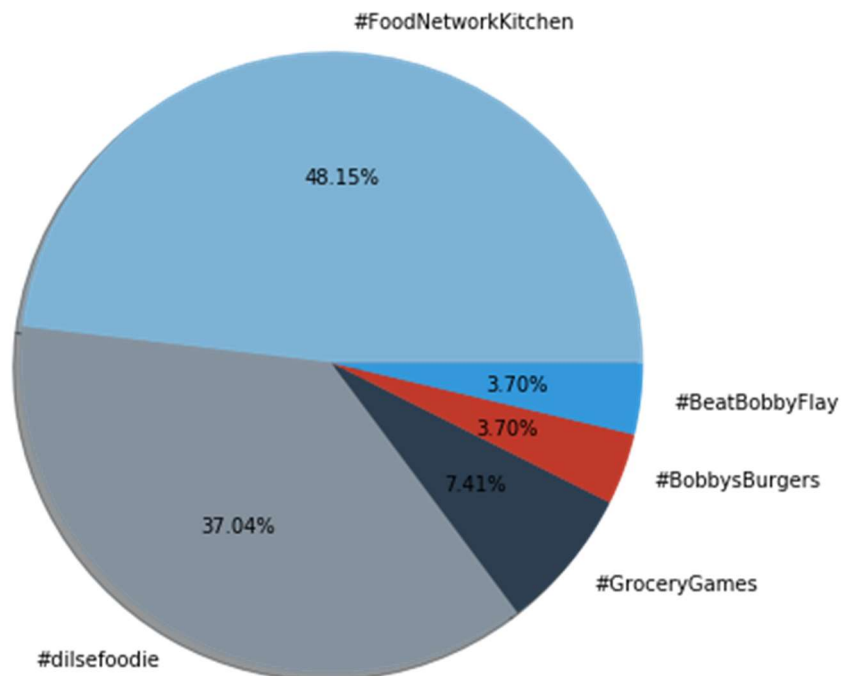
- In order to get the popular hashtags I will iterate over hashtags array and count the frequency of each and every hashtags.
- After getting the frequency I have saved it in hashtags\_frequency.csv file.
- After getting the frequency of hashtags I have the plotted the pie chart for top 5 hashtags.

### Tables and Graphs:

**Top 5 hashtags are:**

Hashtag	Frequency
#FoodNetworkKitchen	13
#dilsefoodie	10
#GroceryGames	2
#BobbysBurgers	1
#BeatBobbyFlay	1

### **Hashtags with number of times they were used.**



### **Question 3:**

**You need to also calculate average followers : likes ratio for the obtained handles.**

**Followers : Likes ratio is calculated as follows :**

- a. Find out the likes of the top 10 posts of the 5 handles obtained earlier.**
- b. Calculate the average likes for a handle.**
- c. Divide the average likes obtained from the number of followers of the handle to get the average followers: like ratio of each handle**
- d. Create a bar graph to depict the above obtained information**

### **Explanation:**

#### **Given Information:**

Here we are asked to calculate the average followers: like ratio for each of the top 5 handles. To calculate the average followers: like ratio following mentioned steps need to be performed:

1. Get the total likes on first 10 posts of handle then divide it by 10 to get the average like for particular handle.
2. Divide the count of followers with the average likes obtained to get the average followers: like ratio.

#### **Steps performed to get the average followers: like ratio for each handle:**

- In order to get the sum of likes we need to iterate over top 10 posts and sum up all likes for handle.
- I have created an array to store the total of likes for each handle, later converting it to NumPy array to perform mathematical operations easily.
- Since the like array is now NumPy array dividing it by 10 will divide each and every element with 10 will give us average like.
- Now dividing the count of followers with average like will result average followers: like ratio for each handle respectively.
- The average followers: likes for each handle is depicted using graphs.

## Tables and Graphs:

Average likes for handle.

Handle Name	Average likes.
foodnetwork	118477.6
buzzfeedfood	150132.7
foodnetworkkitchen	12317.9
dilsefoodie	109500.1
foodtalkindia	1998.7

Followers:likes ratio for handle.

Handle Name	Followers:like.
foodnetwork	86.09222334010816
buzzfeedfood	24.6448641768249
foodnetworkkitchen	50.008524180258
dilsefoodie	4.657529993123294
foodtalkindia	148.59658778205832

**Handles versus average followers:like ratio**

