Overview

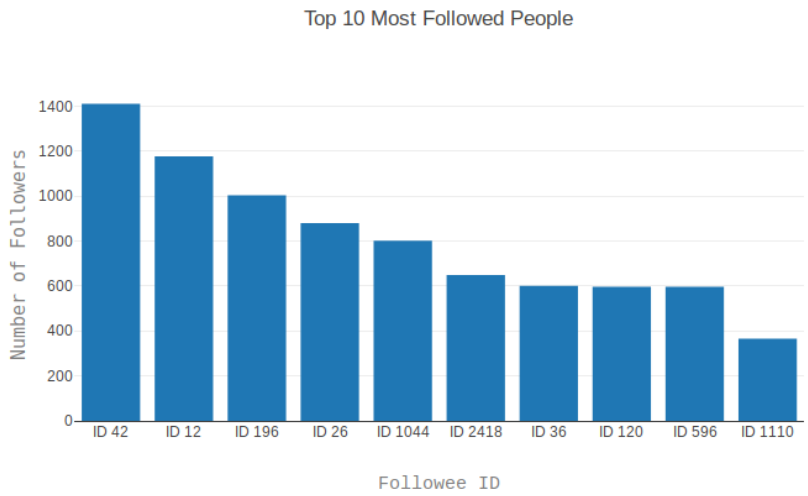
For simplicity, the recommendation system is divided into three parts:

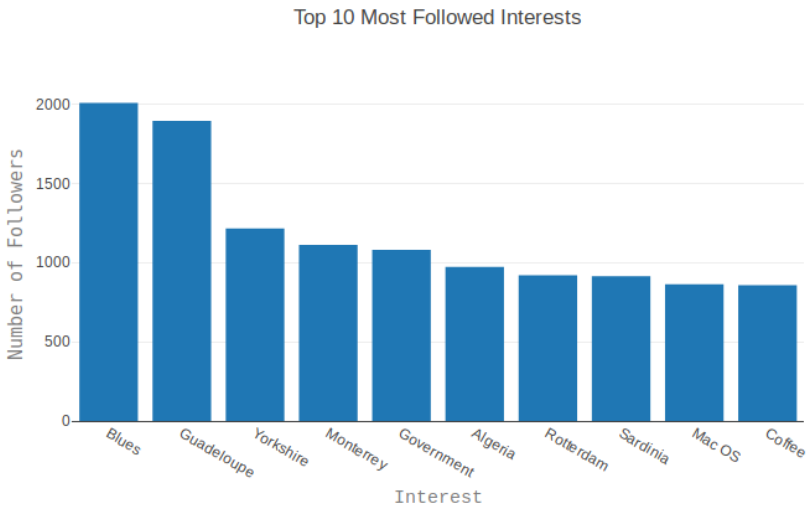
- *Exploratory:* This part consists of 4 graphs, for data exploration and summary, that will be discussed later.

- *People to follow:* A code that suggests people a user might follow if we input the ID of someone it's already following.

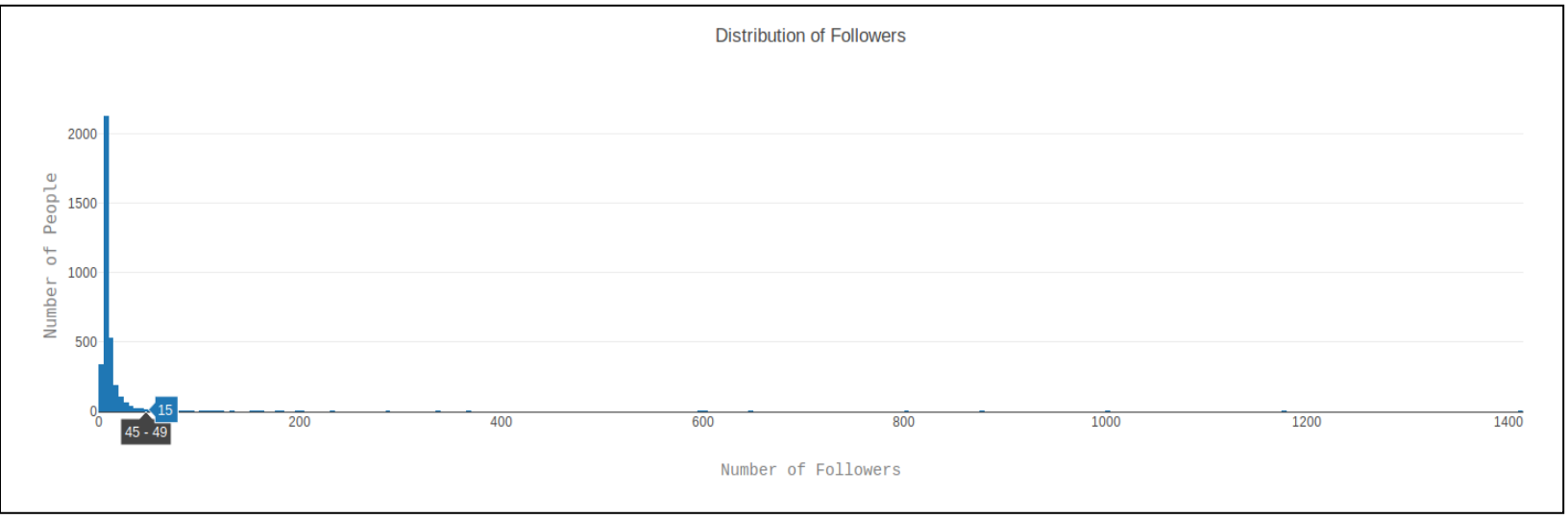
- *Interests to follow:* A code that suggests interests a person my follow based on an interest he/she is already following.

Exploration



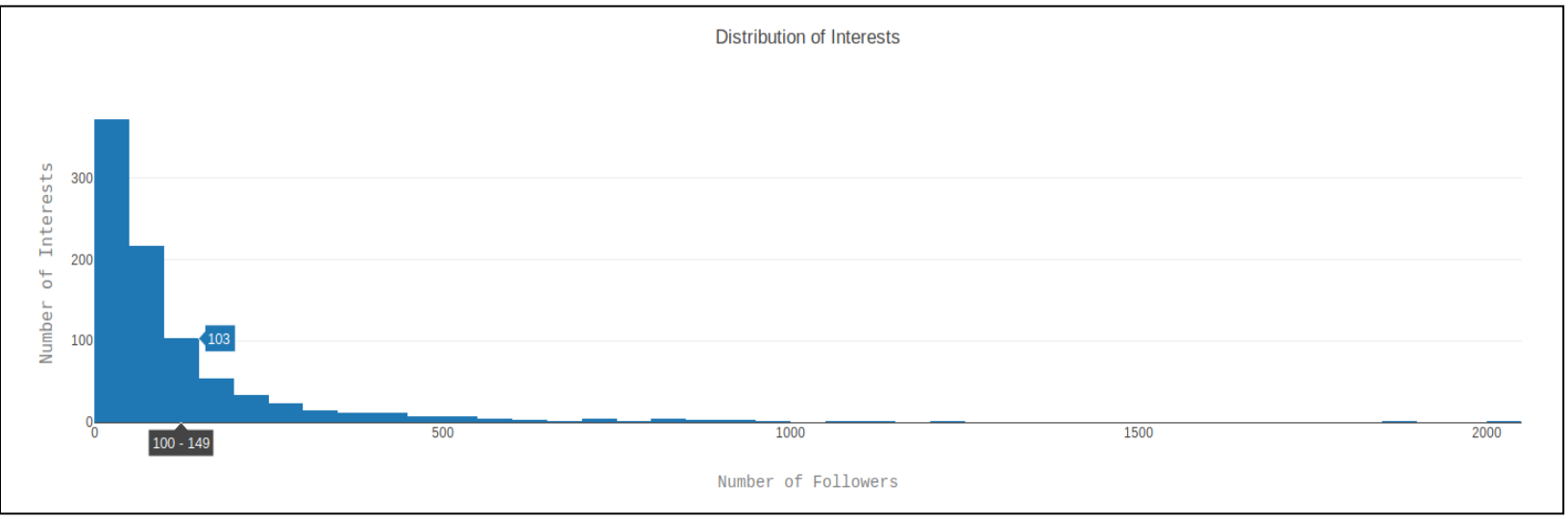


From the above two graphs, we get a quick glimpse at the top 10 followed Users and Interests along with their total number of followers.



From the Distribution of Followers graph above we see that most of the people have less than 50 followers.

And from the Distribution of Interests graph below we see that most of Interests have 100 or less followers.



People to Follow

It works on the logic of pandas' correlation with function.

We used the pivot_table() function in pandas to convert the data frame into a table with Follower IDs as Index and Followee IDs as columns. If a person is following another person, then the respective cell gets assigned a value of 1 otherwise 0 as depicted below.

For example: ID 2 follows ID 4 hence the respective cell has a value of 0, however ID 2 does not follow ID 10, so the respective cell has a value of 0.

Followee ID	2	4	6	8	10	12	14	16	18	20	...	7388	7390	7392	7394	7396	7398	7400	7402	7404	7406
Follower ID																					
2	0.0	1.0	1.0	0.0	0.0	1.0	0.0	1.0	0.0	1.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	1.0	0.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	1.0	1.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	1.0	1.0	1.0	0.0	1.0	1.0	1.0	1.0	0.0	1.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	1.0	1.0	1.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

5 rows × 3595 columns

To recommend people to follow to a user, the social media website team just need to enter the ID of a person the user is following, and the system will suggest the top suggestion.

For example, if a user is already following ID 24, we find the ID of people that have the highest correlation with ID 24 and present them for suggestions.

```
Enter a User ID the person is already following: 24

If a user is following ID # 24 then he might follow the following people as well:

    Followee ID  Correlation  Number of Followers
0            28     0.770462                   27
1             2     0.763901                   30
2            38     0.748349                   34
3            50     0.721421                   28
4            34     0.714451                   21
```

Interests to Follow

This code works on the same logic of people to follow code.

The pivot table for interests looks like the image below.

Interest	Abruzzo	Acapulco	Acura	Adana	Adelaide	Adoption	Afghanistan	Africa	Agra	Agritourism	...	Yamanashi	Yemen	Yokohama	Yorkshire
User ID															
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
12	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
22	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0

5 rows × 883 columns

Interests suggestion:

```
Enter an Interest the person is already following: blues


If a user is following the interest " Blues " then he might follow the following interests as well:

      Interest  Correlation  Number of Followers
0  Guadeloupe     0.224997                 1895
1  Tamaulipas     0.185242                  733
2   Monterrey     0.184425                 1112
3      Mac OS     0.182646                  864
4       Tokyo     0.179607                  817
```

Steps to run the dash_display.py code:

1. In a terminal, open and run dash_display.py to activate the dashboard
2. Visit 127.0.0.1:8050 on your web browser to view the dashboard

Future Data Dumps

For future analysis, the website could provide us with data related to user posts to suggest interest or people to follow.