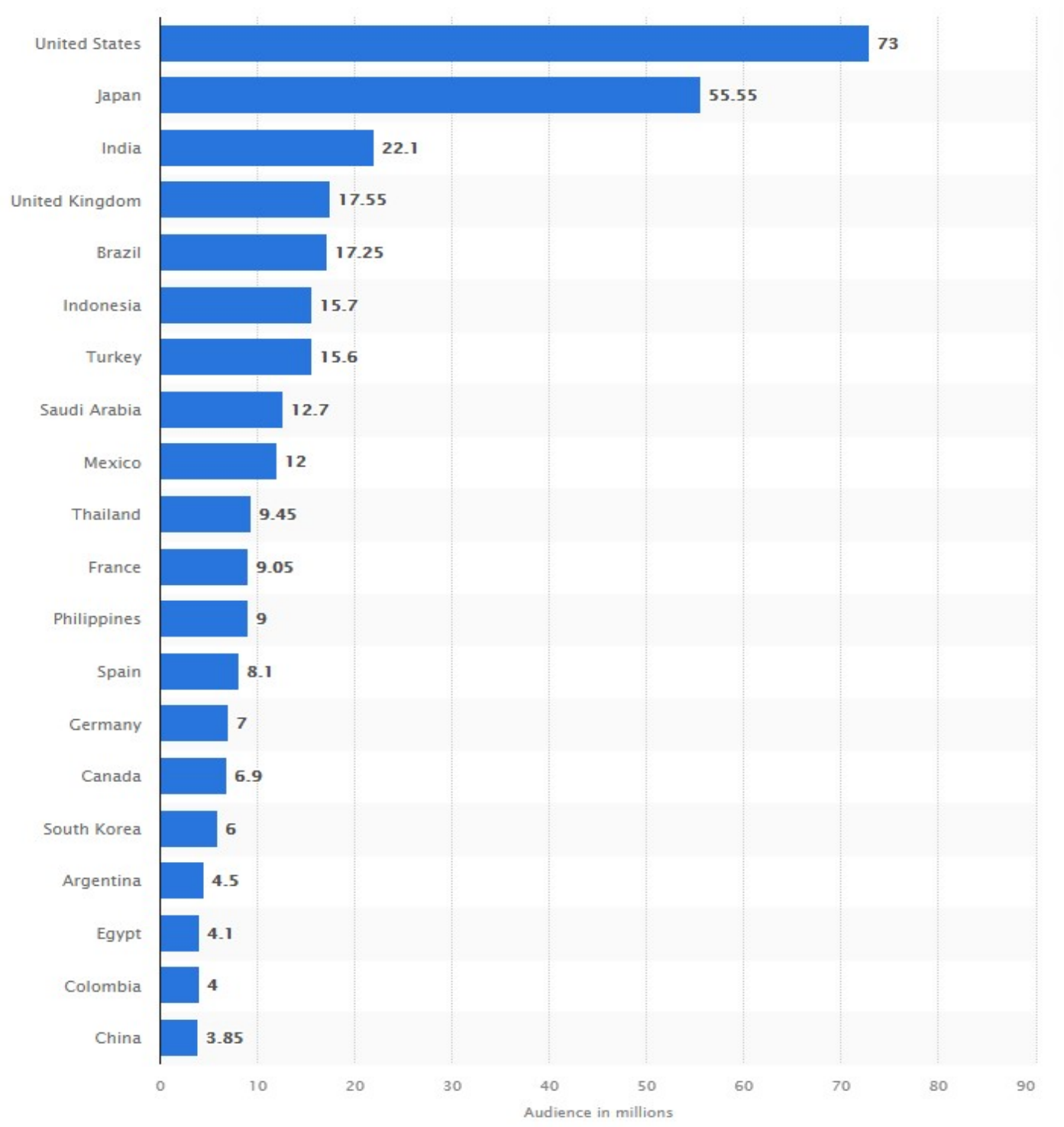
Project Proposal

Prepared by

Name
Mohammad suliman Alowibdi

Question/needs:

Studying the analytical data of government agencies to ensure the speed of response to inquiries and reduce the time and effort expended for that through the use of statistical functions that ensure, God willing, sorting out what is reported on the authority's account on the Twitter platform, providing computer-technical suggestions, and determining how to deal with those tweets quickly.



The chart show active users only (https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/)

Data description:

Big data is everywhere. Period. In the process of running a successful Governments in today's day and age, Twitter.com is a gold mine of data. Unlike other social platforms, almost every user's tweets are completely public and pullable. This is a huge plus if you're trying to get a large amount of data to run analytics on. Twitter data is also pretty specific. Twitter's API allows you to do complex queries like pulling every tweet about a certain topic within the last twenty minutes, or pull a certain user's non-retweeted tweets.We can also target users that specifically live in a certain location, which is known as spatial data. Another application of this could be to map the areas on the globe where the account has been mentioned the most.As we can see, Twitter data can be a large door into the insights of the general public, and how they receive a topic. That, combined with the openness and the generous rate limiting of Twitter's API, can produce powerful results.

In [1]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.ensemble import RandomForestClassifier
import warnings
warnings.simplefilter(action='ignore', category=FutureWarning)
warnings.simplefilter(action='ignore', category=UserWarning)
sns.set(style="white")
%matplotlib inline
```

In [3]:
```
df=pd.read_csv("MOI-tweets.csv")
df.head()
```

Out[3]:

	Unnamed: 0	id_str	retweet_count	favorite_count	replyCount	quoteCount	user.screen_name	datenew	created_at	datecreatetweetold	
0	3	1448997995475881988	5223	7173	1126.0	740.0	MOISaudiArabia	2021-10-15	Fri Oct 15 13:03:37 +0000 2021	16 days 10:56:23	لصحية في
1	55	1441521487970639877	4436	8179	957.0	405.0	MOISaudiArabia	2021-09-24	Fri Sep 24 21:54:39 +0000 2021	37 days 02:05:21	ر..التحرش
2	66	1440718303383539712	2080	7538	440.0	125.0	MOISaudiArabia	2021-09-22	Wed Sep 22 16:43:05 +0000 2021	39 days 07:16:55	امنين \n\n#
3	36	1435302078092619777	2218	3587	1660.0	251.0	MOISaudiArabia	2021-09-07	Tue Sep 07 18:00:56 +0000 2021	54 days 05:59:04	ي المملكة
4	78	1443908772410318849	2360	2954	1120.0	81.0	MOISaudiArabia	2021-10-01	Fri Oct 01 12:00:52 +0000 2021	30 days 11:59:08	/ : 10 / 10

49054 rows × 18 columns

As seen above the rows are 49,054 and it has more than 10 features:

1.	id_str	9.	text
2.	retweet_count	10.	datecreateuserold
3.	favorite_count	11.	user.created_at
4.	replyCount	12.	engagements
5.	quoteCount	13.	day
6.	user.screen_name	14.	month
7.	datenew	15.	year
8.	created_at	16.	url
9.	datecreatetweetold	17.	

Tools:

I will be planning to use deep learning model and library. I will be using Mitplot, pandas and numpy library for visualization and calculation. On top of my head those are the tools I can think of. However, going through the model I may come up with another approaches.

MVP Goal:

The importance of this project is related to the large increase in Twitter users in the Kingdom of Saudi Arabia, and what this requires is the importance of interacting with that increase through the use of technology in sorting and classifying all tweets and identifying the important ones to deal with it quickly. By doing that it Applied contribution to improving the development efforts of man and place in the Kingdom of Saudi Arabia and in light of the evaluation results of the impact of those interactions with social networks .

Thanks

- Dr. Mariam elmasry
- Dr. Chiekh Alloul

for your time