Tweet Analytics

Group ID: Group035

Members

1. Ananta Patil
2. Mahesh Mehendale
3. Harsh Ajit Khajgiwale

In [1]:

```
!pip install pymongo
!pip install matplotlib
!pip intall pandas
!pip install numpy
```

Defaulting to user installation because normal site-packages is not wr
iteable
Requirement already satisfied: pymongo in /home/harsh/.local/lib/pytho
n3.8/site-packages (3.12.0)
WARNING: You are using pip version 21.1.3; however, version 21.2.4 is
 available.
You should consider upgrading via the '/usr/bin/python3 -m pip install
--upgrade pip' command.
Defaulting to user installation because normal site-packages is not wr
iteable
Requirement already satisfied: matplotlib in /home/harsh/.local/lib/py
thon3.8/site-packages (3.4.3)
Requirement already satisfied: kiwisolver>=1.0.1 in /home/harsh/.loca
l/lib/python3.8/site-packages (from matplotlib) (1.3.2)
Requirement already satisfied: cycler>=0.10 in /home/harsh/.local/lib/
python3.8/site-packages (from matplotlib) (0.10.0)
Requirement already satisfied: pillow>=6.2.0 in /usr/lib/python3/dist-
packages (from matplotlib) (7.0.0)
Requirement already satisfied: python-dateutil>=2.7 in /usr/lib/python
3/dist-packages (from matplotlib) (2.7.3)
Requirement already satisfied: pyparsing>=2.2.1 in /home/harsh/.local/
lib/python3.8/site-packages (from matplotlib) (2.4.7)
Requirement already satisfied: numpy>=1.16 in /home/harsh/.local/lib/p
ython3.8/site-packages (from matplotlib) (1.21.0)
Requirement already satisfied: six in /usr/lib/python3/dist-packages
 (from cycler>=0.10->matplotlib) (1.14.0)
WARNING: You are using pip version 21.1.3; however, version 21.2.4 is
 available.
You should consider upgrading via the '/usr/bin/python3 -m pip install
--upgrade pip' command.
ERROR: unknown command "intall" - maybe you meant "install"
Defaulting to user installation because normal site-packages is not wr
iteable
Requirement already satisfied: numpy in /home/harsh/.local/lib/python
3.8/site-packages (1.21.0)
WARNING: You are using pip version 21.1.3; however, version 21.2.4 is
 available.
You should consider upgrading via the '/usr/bin/python3 -m pip install
--upgrade pip' command.

In [2]:

```
from pymongo import MongoClient
import json
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
```

In [3]:

```
#Connecting to MonogDB
client = MongoClient("mongodb://localhost:27017/")
```

In [4]:

```python
#Database name => IrGroup10
mydatabase = client["IrGroup10"]
```

In [5]:

```python
#Selecting collection => consumer_data
mycollection = mydatabase["consumer_data"]
```

In [6]:

```python
#Total records in consumer_data
mycollection.count()
```

```
<ipython-input-6-b253b998c5ee>:2: DeprecationWarning: count is depreca
ted. Use estimated_document_count or count_documents instead. Please n
ote that $where must be replaced by $expr, $near must be replaced by
$geoWithin with $center, and $nearSphere must be replaced by $geoWithi
n with $centerSphere
  mycollection.count()
```

Out[6]:

```
21266
```

In [7]:

```python
cursor = mycollection.find()
```

In [8]:

```python
tweets_feed = {}
for record in cursor:
    convertedDict = json.loads(record['tweets'])
    tweets_feed[convertedDict['tweet_id']] = []
    tweets_feed[convertedDict['tweet_id']].append(convertedDict['user_id'])
    tweets_feed[convertedDict['tweet_id']].append(convertedDict['user_name'])
    tweets_feed[convertedDict['tweet_id']].append(convertedDict['user_screen_name']
    tweets_feed[convertedDict['tweet_id']].append(convertedDict['user_followers_cou
    tweets_feed[convertedDict['tweet_id']].append(convertedDict['user_location'])
    tweets_feed[convertedDict['tweet_id']].append(convertedDict['tweet_text'])
    tweets_feed[convertedDict['tweet_id']].append(convertedDict['tweet_lang'])
```

In [9]:

```python
df = pd.DataFrame.from_dict(tweets_feed)
```

In [10]:

```python
df = df.transpose()
```

In [11]:

```python
df.columns=['user_id','user_name','user_screen_name','user_followers_count','user_l
```

In [12]:

```
df.index.name= 'tweet_id'
```

In [13]:

```
df.head()
```

Out[13]:

tweet_id	user_id	user_name	user_screen_name	user_followers_cou
1433450460833275915	1347685363763077121	ɪ.ᴠɪᴠɪ 🖤	EscobarIvivi	2
1433450461181403145	981555457	litö 🐞	scallywagcarlos	218
1433450461529395202	4724605634	SchoolFinds	school_finds	
1433450462511050753	865991759247093760	Daniel cracknall	MLeCraquepot	42
1433450468748013569	17290173	Zoraya Suarez	ZorayaSuarezPNC	11

In [14]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 21266 entries, 1433450460833275915 to 1433479724089679886
Data columns (total 7 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   user_id               21266 non-null  object
 1   user_name             21266 non-null  object
 2   user_screen_name      21266 non-null  object
 3   user_followers_count  21266 non-null  object
 4   user_location         13036 non-null  object
 5   tweet_text            21266 non-null  object
 6   tweet_lang            21266 non-null  object
dtypes: object(7)
memory usage: 1.3+ MB
```

Number of unique users and Unique locations of tweets

In [15]:

```
len(df['user_id'].unique()),len(df['user_location'].unique())
```
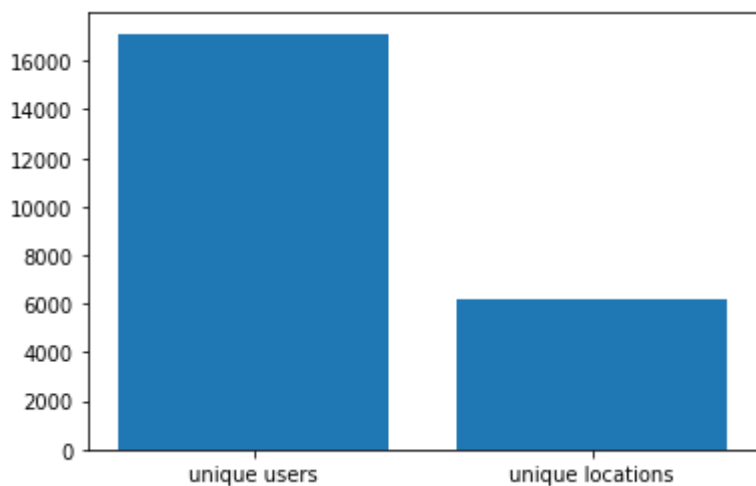
Out[15]:

```
(17122, 6212)
```

In [16]:

```python
x = np.array(["unique users", "unique locations"])
y = np.array([len(df['user_id'].unique()),len(df['user_location'].unique())])
plt.bar(x,y)
```

Out[16]:

```
<BarContainer object of 2 artists>
```


Top 20 active locations (country,state and city wise) on twitter

In [17]:

```python
top_20_active_locations = df.groupby(by='user_location').count().sort_values(by='us
```

In [18]:

```
top_20_active_locations.plot.bar()
```

Out[18]:

```
<AxesSubplot:xlabel='user_location'>
```

```
/home/harsh/.local/lib/python3.8/site-packages/matplotlib/backends/b
ackend_agg.py:240: RuntimeWarning: Glyph 3349 missing from current f
ont.
  font.set_text(s, 0.0, flags=flags)
/home/harsh/.local/lib/python3.8/site-packages/matplotlib/backends/b
ackend_agg.py:240: RuntimeWarning: Glyph 3399 missing from current f
ont.
  font.set_text(s, 0.0, flags=flags)
/home/harsh/.local/lib/python3.8/site-packages/matplotlib/backends/b
ackend_agg.py:240: RuntimeWarning: Glyph 3376 missing from current f
ont.
  font.set_text(s, 0.0, flags=flags)
/home/harsh/.local/lib/python3.8/site-packages/matplotlib/backends/b
ackend_agg.py:240: RuntimeWarning: Glyph 3379 missing from current f
ont.
  font.set_text(s, 0.0, flags=flags)
/home/harsh/.local/lib/python3.8/site-packages/matplotlib/backends/b
ackend_agg.py:240: RuntimeWarning: Glyph 3330 missing from current f
ont.
  font.set_text(s, 0.0, flags=flags)
/home/harsh/.local/lib/python3.8/site-packages/matplotlib/backends/b
ackend_agg.py:240: RuntimeWarning: Glyph 3611 missing from current f
ont.
  font.set_text(s, 0.0, flags=flags)
/home/harsh/.local/lib/python3.8/site-packages/matplotlib/backends/b
ackend_agg.py:240: RuntimeWarning: Glyph 3619 missing from current f
ont.
  font.set_text(s, 0.0, flags=flags)
/home/harsh/.local/lib/python3.8/site-packages/matplotlib/backends/b
ackend_agg.py:240: RuntimeWarning: Glyph 3632 missing from current f
ont.
  font.set_text(s, 0.0, flags=flags)
/home/harsh/.local/lib/python3.8/site-packages/matplotlib/backends/b
ackend_agg.py:240: RuntimeWarning: Glyph 3648 missing from current f
ont.
  font.set_text(s, 0.0, flags=flags)
/home/harsh/.local/lib/python3.8/site-packages/matplotlib/backends/b
ackend_agg.py:240: RuntimeWarning: Glyph 3607 missing from current f
ont.
  font.set_text(s, 0.0, flags=flags)
/home/harsh/.local/lib/python3.8/site-packages/matplotlib/backends/b
ackend_agg.py:240: RuntimeWarning: Glyph 3624 missing from current f
ont.
  font.set_text(s, 0.0, flags=flags)
/home/harsh/.local/lib/python3.8/site-packages/matplotlib/backends/b
ackend_agg.py:240: RuntimeWarning: Glyph 3652 missing from current f
ont.
  font.set_text(s, 0.0, flags=flags)
/home/harsh/.local/lib/python3.8/site-packages/matplotlib/backends/b
ackend_agg.py:240: RuntimeWarning: Glyph 3618 missing from current f
ont.
  font.set_text(s, 0.0, flags=flags)
/home/harsh/.local/lib/python3.8/site-packages/matplotlib/backends/b
```
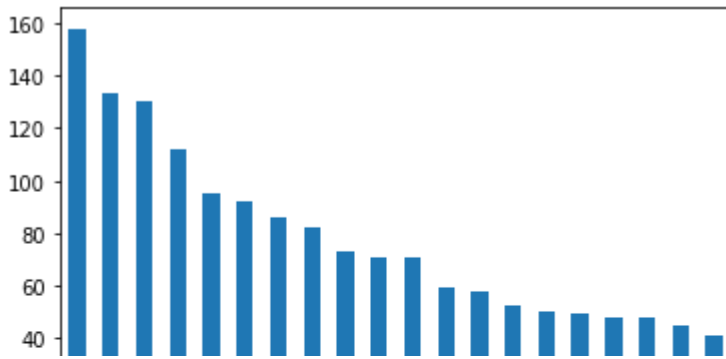
```
ackend_agg.py:203: RuntimeWarning: Glyph 3349 missing from current f
ont.
  font.set_text(s, 0, flags=flags)
/home/harsh/.local/lib/python3.8/site-packages/matplotlib/backends/b
ackend_agg.py:203: RuntimeWarning: Glyph 3399 missing from current f
ont.
  font.set_text(s, 0, flags=flags)
/home/harsh/.local/lib/python3.8/site-packages/matplotlib/backends/b
ackend_agg.py:203: RuntimeWarning: Glyph 3376 missing from current f
ont.
  font.set_text(s, 0, flags=flags)
/home/harsh/.local/lib/python3.8/site-packages/matplotlib/backends/b
ackend_agg.py:203: RuntimeWarning: Glyph 3379 missing from current f
ont.
  font.set_text(s, 0, flags=flags)
/home/harsh/.local/lib/python3.8/site-packages/matplotlib/backends/b
ackend_agg.py:203: RuntimeWarning: Glyph 3330 missing from current f
ont.
  font.set_text(s, 0, flags=flags)
/home/harsh/.local/lib/python3.8/site-packages/matplotlib/backends/b
ackend_agg.py:203: RuntimeWarning: Glyph 3611 missing from current f
ont.
  font.set_text(s, 0, flags=flags)
/home/harsh/.local/lib/python3.8/site-packages/matplotlib/backends/b
ackend_agg.py:203: RuntimeWarning: Glyph 3619 missing from current f
ont.
  font.set_text(s, 0, flags=flags)
/home/harsh/.local/lib/python3.8/site-packages/matplotlib/backends/b
ackend_agg.py:203: RuntimeWarning: Glyph 3632 missing from current f
ont.
  font.set_text(s, 0, flags=flags)
/home/harsh/.local/lib/python3.8/site-packages/matplotlib/backends/b
ackend_agg.py:203: RuntimeWarning: Glyph 3648 missing from current f
ont.
  font.set_text(s, 0, flags=flags)
/home/harsh/.local/lib/python3.8/site-packages/matplotlib/backends/b
ackend_agg.py:203: RuntimeWarning: Glyph 3607 missing from current f
ont.
  font.set_text(s, 0, flags=flags)
/home/harsh/.local/lib/python3.8/site-packages/matplotlib/backends/b
ackend_agg.py:203: RuntimeWarning: Glyph 3624 missing from current f
ont.
  font.set_text(s, 0, flags=flags)
/home/harsh/.local/lib/python3.8/site-packages/matplotlib/backends/b
ackend_agg.py:203: RuntimeWarning: Glyph 3652 missing from current f
ont.
  font.set_text(s, 0, flags=flags)
/home/harsh/.local/lib/python3.8/site-packages/matplotlib/backends/b
ackend_agg.py:203: RuntimeWarning: Glyph 3618 missing from current f
ont.
  font.set_text(s, 0, flags=flags)
```

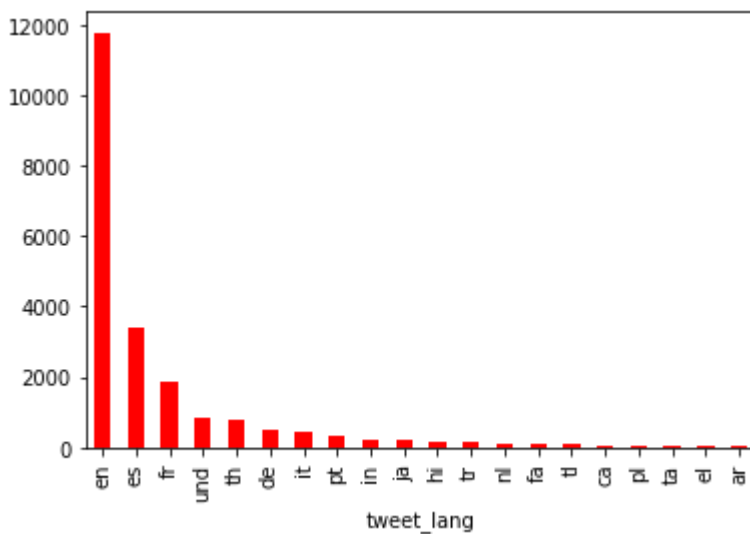# Top 20 Langauages used on Twitter across globe

In [19]:

```
top_20_lang_used_on_twitter = df.groupby(by='tweet_lang').count().sort_values(by='u
```

In [20]:

```
top_20_lang_used_on_twitter.plot.bar(color='red')
```

Out[20]:

```
<AxesSubplot:xlabel='tweet_lang'>
```


Top 20 user names followed on twitter

In [21]:

```
top_20_followed_users = df.sort_values(by='user_followers_count', ascending=False)[
```

In [22]:

```python
plt.figure(figsize=(15,4))
plt.bar(top_20_followed_users['user_name'],top_20_followed_users['user_followers_co
plt.xticks(rotation=90)
```

Out[22]:

```
([0, 1, 2, 3, 4, 5, 6, 7],
 [Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, '')])
```