**CS5540 - PRINCIPLES OF BIG DATA MANAGEMENT**

**Interim Project Report**

Team Members:

|  |  |  |
| --- | --- | --- |
| **Name** | **ID** | **Email ID** |
| Shiva Kumar Murarishetti | 16296890 | [smpf4@mail.umkc.edu](mailto:smpf4@mail.umkc.edu) |
| Lavanya Gadde | 16298160 | [lavanya.gadde@mail.umkc.edu](mailto:lavanya.gadde@mail.umkc.edu) |
| Sravani Garikapati | 16296889 | [sg324@mail.umkc.edu](mailto:sg324@mail.umkc.edu) |

Hash tags Used for fetching twitter data are: **'coronavirus','chinesevirus','COVID19','COVID-19','novelcoronavirus','lockdown','COVIDIDIOTS','socialdistance','pandemic','WHO','hygenic','quarantine','isolation','sanitizers','contagious','virus','toiletrolles','communityspread','spread','economy','recession','Covid19Out','gocorona',**

After analyzing twitter API attributes:

1. **Discussion of analytics queries that you have implemented or plan to implement.**

* Query to fetch count of Tweeted devices.
* Query to fetch top 10 verified users
* Query to fetch top 10 languages used for Tweets
* Query to fetch top 10 Followers Count based on Location
* Query to fetch top 10 Favorites count.
* Query to fetch top 10 most tweeted dates
* Query to get the top countries where tweets came from
* Query to get the protected status of users
* Retweets based on country
* On which Day More Tweets are posted on
* Timings for more tweets

1. **Distribution of work among teammates.**

As Spark & Hadoop are new to all the three, we sat together sometime. We connected through zoom due to the pandemic of corona most of the time.

1. **Plans for testing your software.**

Not yet finalized.

1. **Plans for visualization of your results.**

Zeppelin/Tableau