PREDICTING THE DAILY DEATHS DUE TO COVID-19

Jasleen Singh

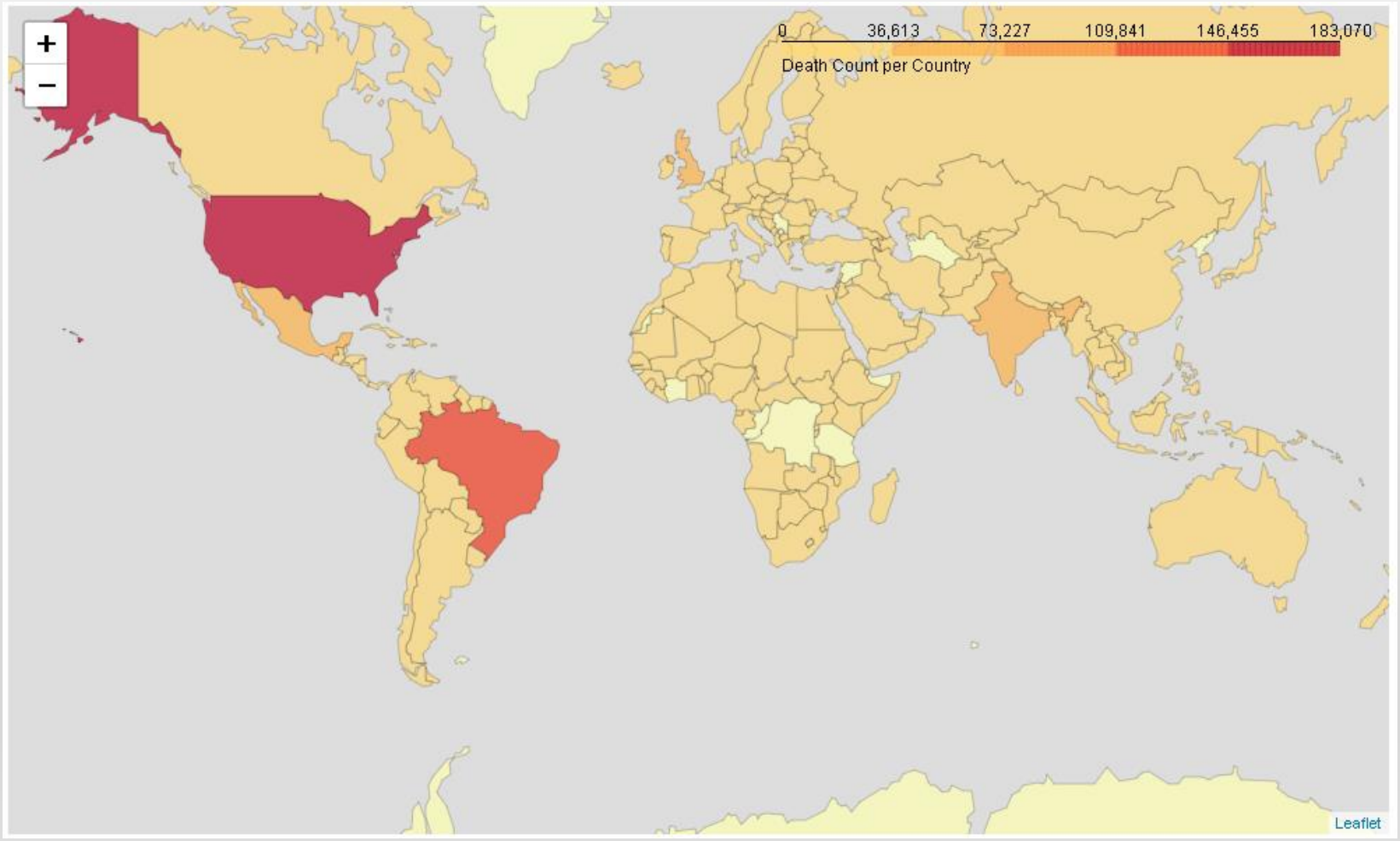
PREDICTING THE DAILY DEATHS DUE TO COVID-19 WILL BE VALUABLE FOR

- WHO (World Health Organization)
- Country administration
- Citizens

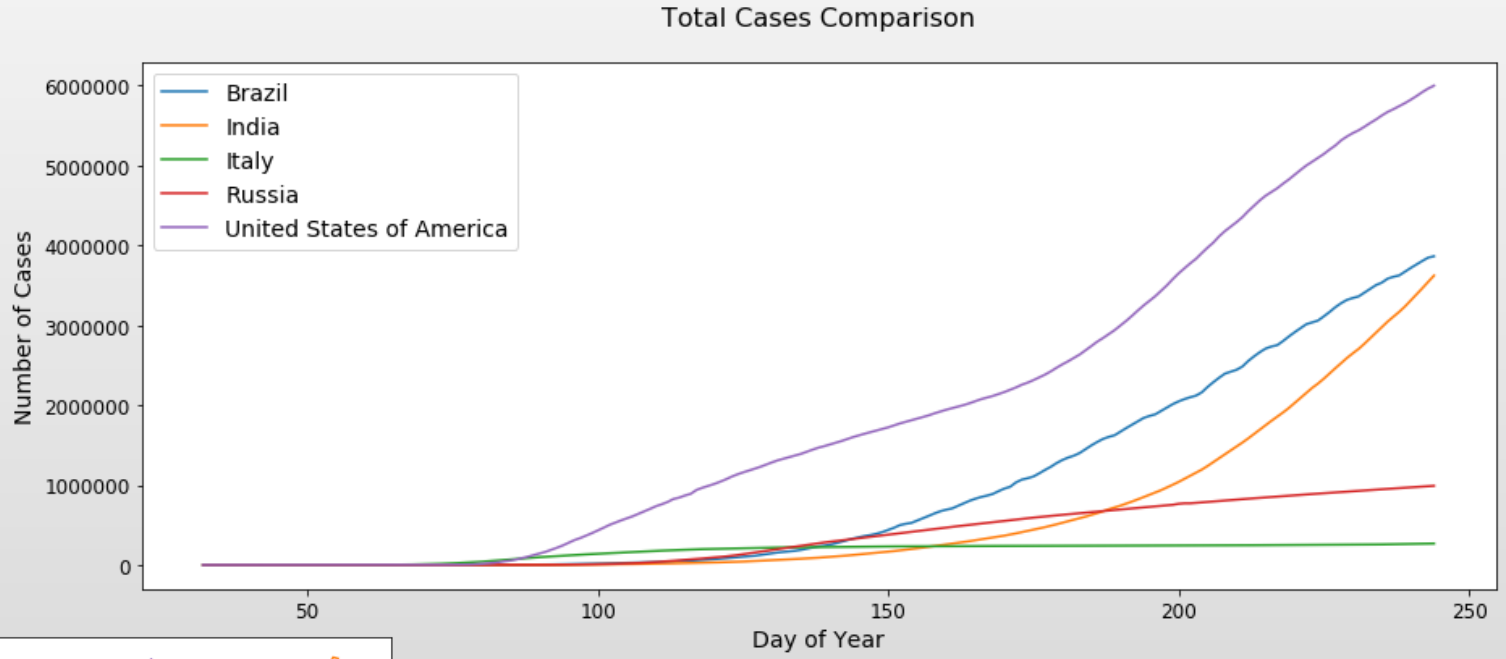
DATA ACQUISITION AND CLEANING

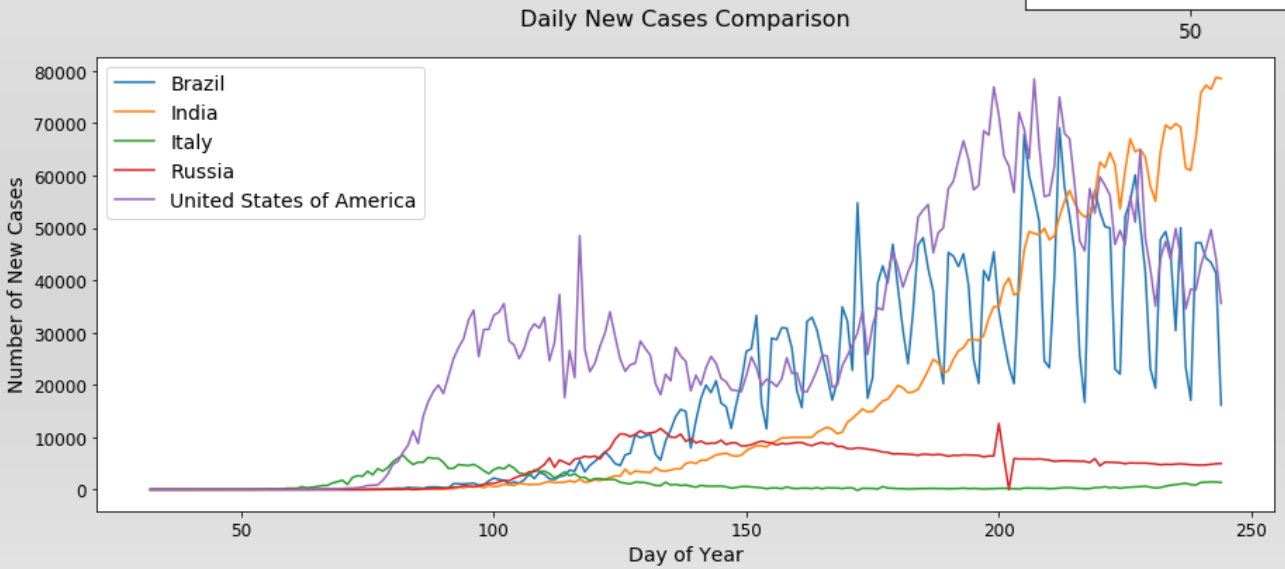
- Data: https://covid.ourworldindata.org/data/owid-covid-data.csv
- Meta-data: https://covid.ourworldindata.org/data/owid-covid-codebook.csv
- In total, 42,770 rows in the raw dataset.
- Used the data from 1st February, 2020 till 31st August, 2020

DATA VISUALIZATION OVER WORLD MAP

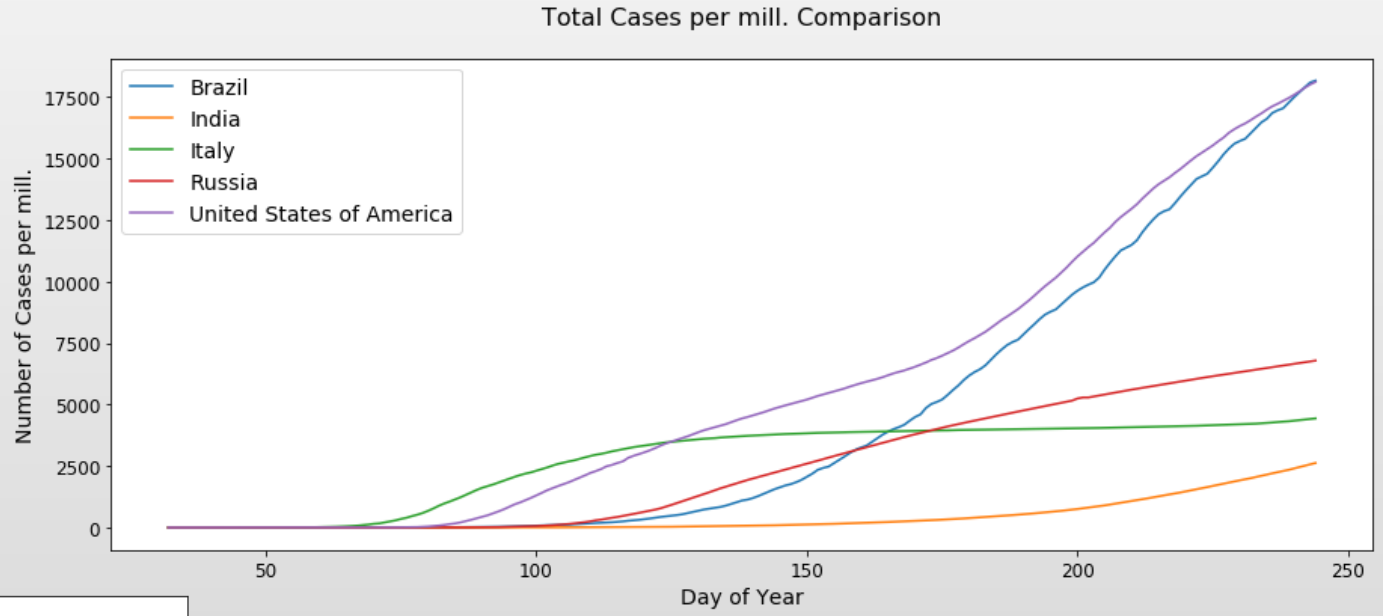


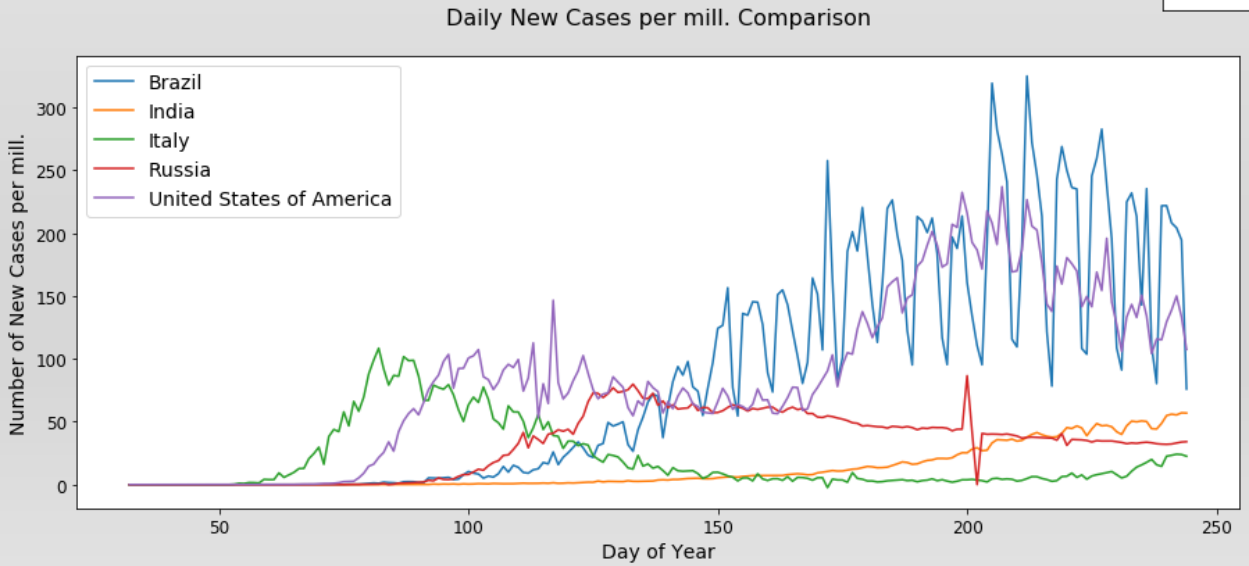
TIME-SERIES ANALYSIS OF TOTAL CASES AND NEW CASES



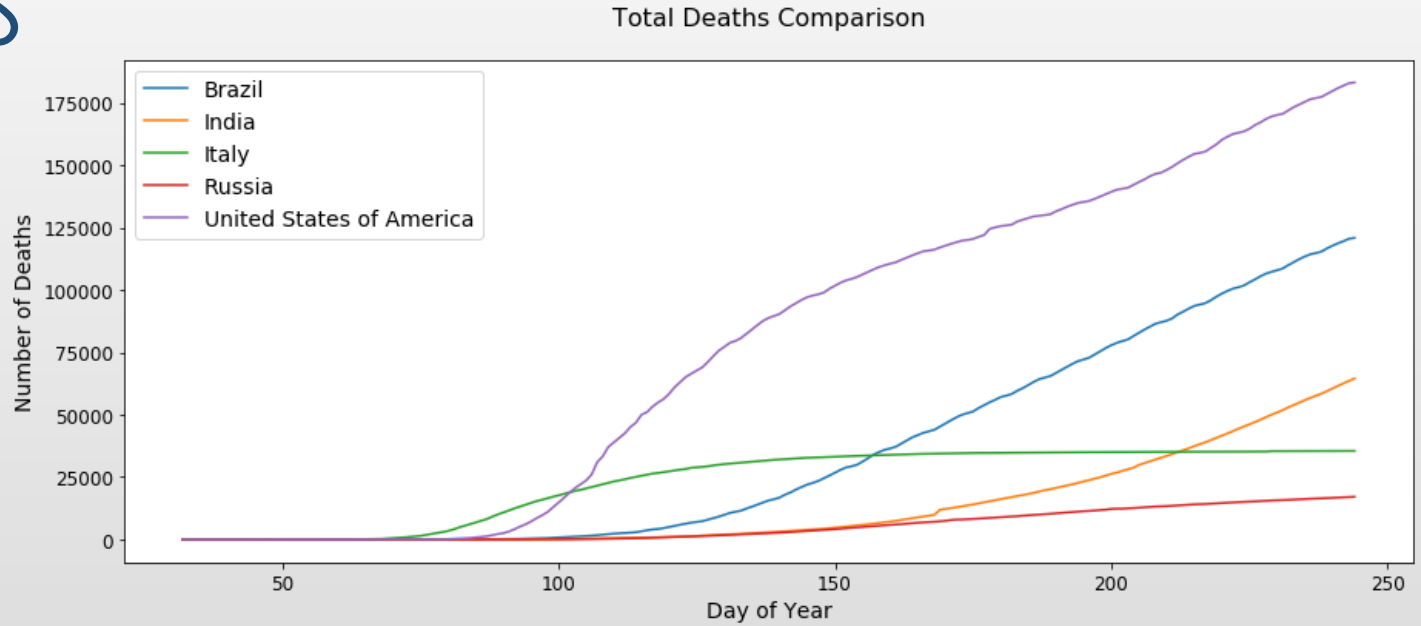


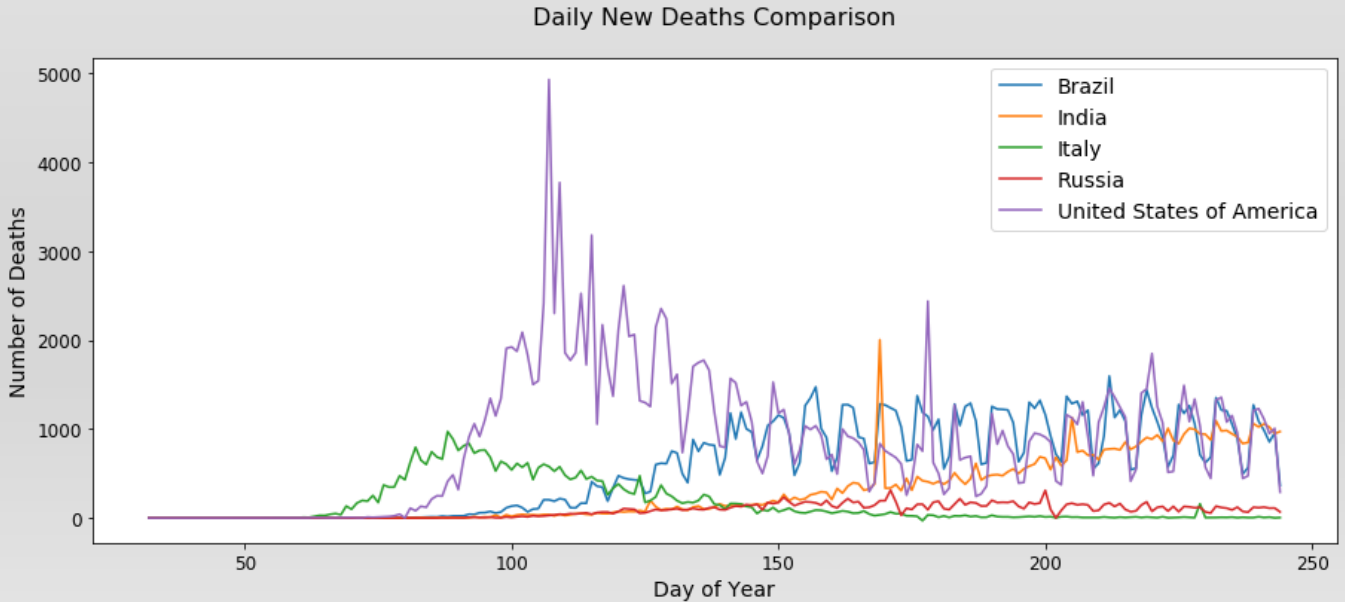
TIME-SERIES ANALYSIS OF TOTAL CASES AND NEW CASES PER MILLION OF POPULATION



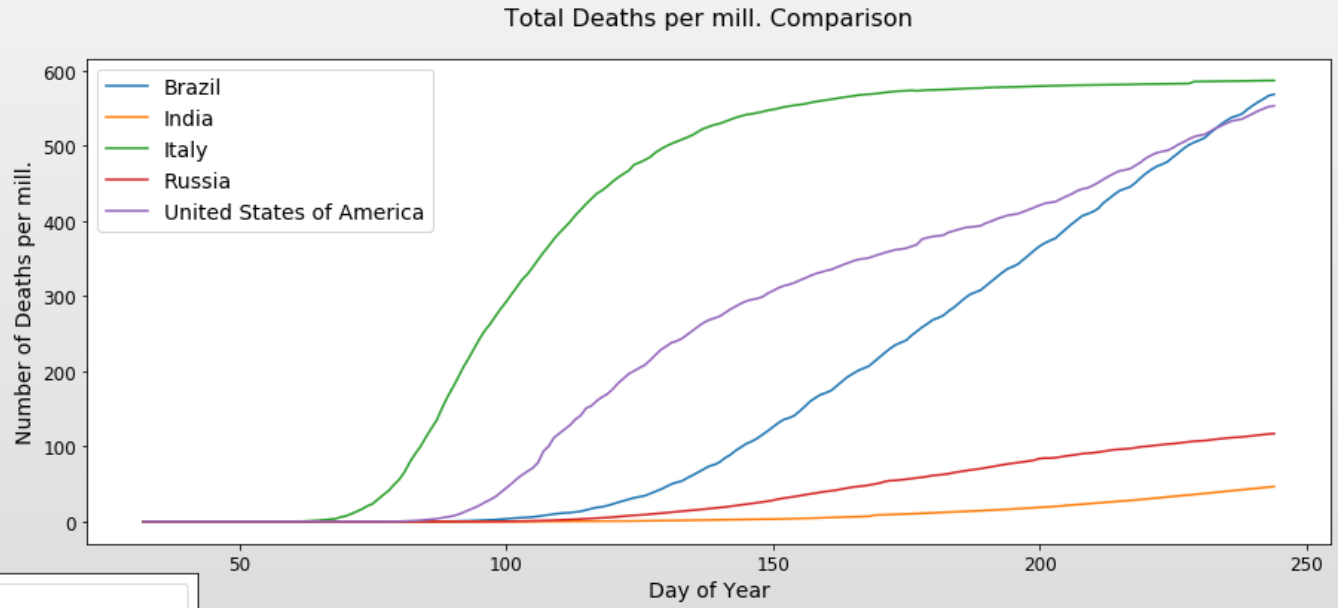


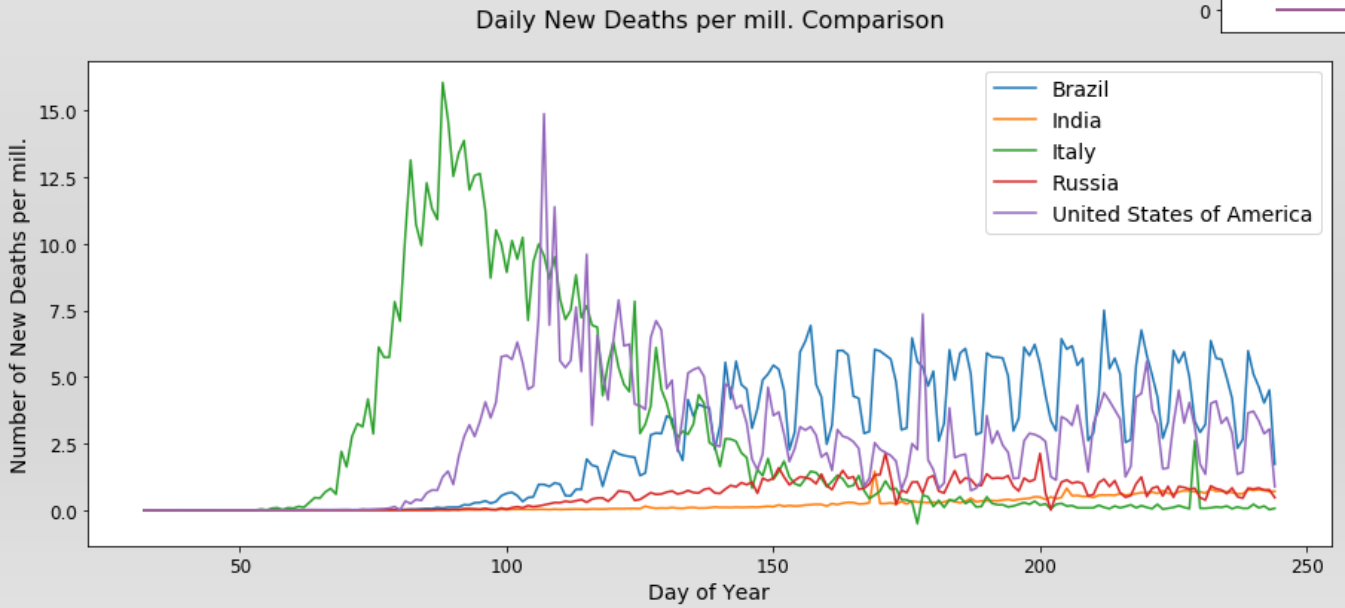
TIME-SERIES ANALYSIS OF TOTAL DEATHS AND NEW DEATHS



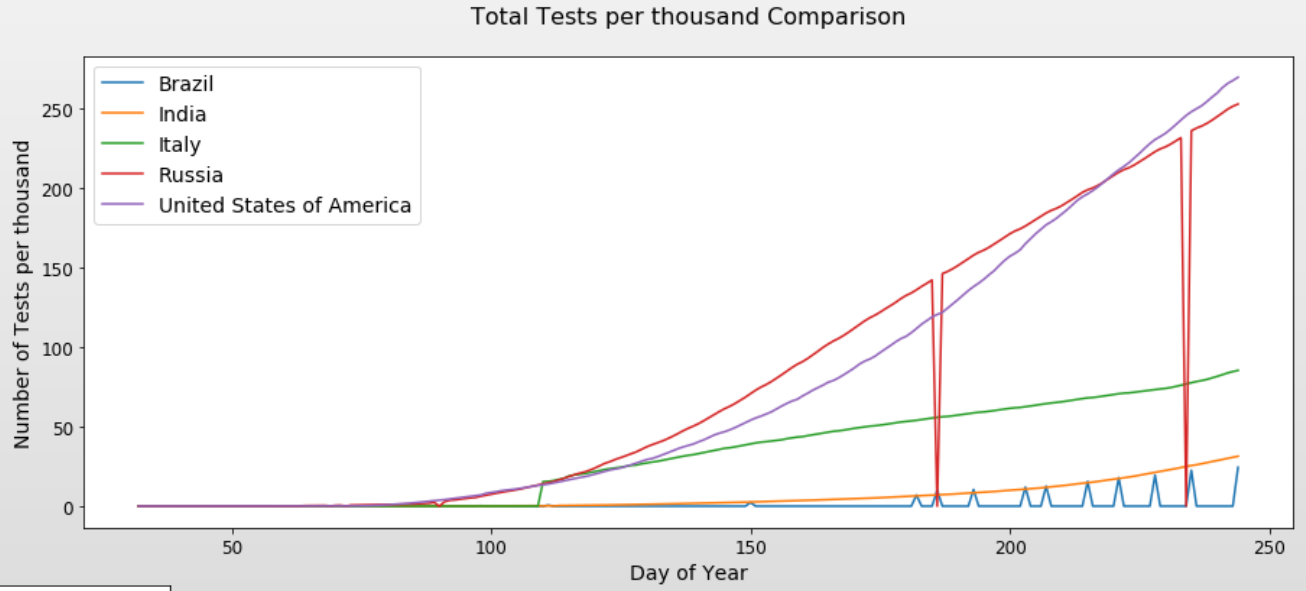


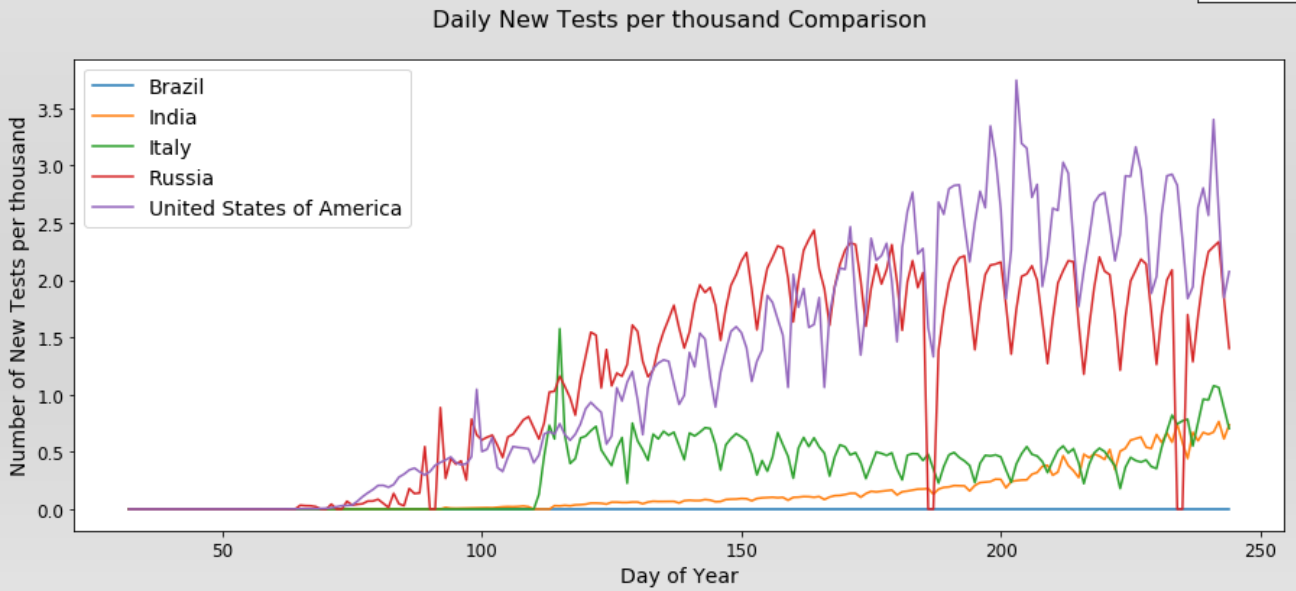
TIME-SERIES ANALYSIS OF TOTAL DEATHS AND NEW DEATHS PER MILLION OF POPULATION





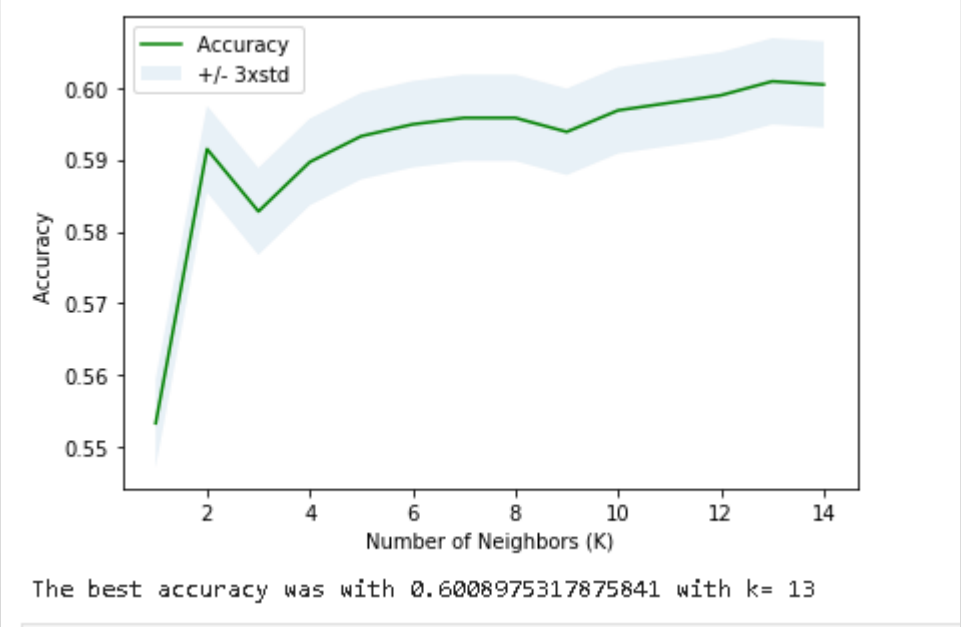
TIME-SERIES ANALYSIS OF TOTAL TESTS AND NEW TESTS PER THOUSAND OF POPULATION





MACHINE LEARNING MODELS

- K Nearest Neighbour (KNN)
- Decision Tree
- Support Vector Machine
- Logistic Regression



CONCLUSION

- KNN performed best so-far as per the analysis and testing
- More number of records required for countries like North Korea, China where data collection is not precisely done
- Proper data collection needs to be done for countries which does not provide data on daily basis.

Algorithm	Accuracy	Jacard	F1-score
KNN	0.60	0.48	0.55
Decision Tree	0.59	0.43	0.50
SVM	0.60	0.38	0.47
Logistic Regression	0.59	0.38	0.47