Computing for Business Analytics (BANA-6620)

Created on: 24th September 2021

Insights about dataset

I will be using four datasets which is about Covid-19 Vaccination drive throughout the world. I got the Dataset from Kaggle.com

(1st & 2nd Dataset https://www.kaggle.com/gpreda/covid-world-vaccination-progress ,

3rd Dataset https://www.kaggle.com/gpreda/all-covid19-vaccines-tweets?select=vaccination_all_tweets.csv ,

4th Dataset https://www.kaggle.com/sudalairajkumar/undata-country-profiles).

The very first dataset (country_vaccinations.csv) is about country wise daily vaccination drive which was started in different months in respective countries until September 2021. This dataset incudes total vaccinations per date and country, number of people vaccinated, number of people fully vaccinated, daily vaccination, total vaccination percent, people vaccinated percent, daily vaccinations per million, vaccines scheme (the combination of vaccines used by a country).

Second dataset (country_vaccinations_by_manufacturer.csv) gives insights about the Total number of vaccinations manufactured by various Vaccine companies date wise from January 2021 to September 2021 in each country.

Third dataset (vaccination_all_tweets.csv) tell us about the attitude of population in view of different vaccination drives according to the tweets related to vaccinations. This dataset was picked up from tweepy Python package. This will be used as a text data analysis, which is the main focus of our group.

Forth dataset (country_profile_variables.csv) is about the demographics and economy of each country. I will use this to find what influenced these vaccination programmes, and these programmes success.

Statistical, Python techniques and Expected outcomes

Will be using various libraries, plots and techniques:

1. Numpy
2. Matplots
3. Pandas
4. Descriptive Statistics
5. Text data analysis
6. Sentiment Analysis (3 different ways)
7. Exploratory Data Analytics
8. Data Visualization
9. Plotly
10. Bar graphs
11. Histograms
12. Correlations

13. Heat maps
14. Scatter Plots
15. Boxplots
16. Dendrogram

After Explorative Data Analysis we would be able to answer questions like:

Which country's vaccine drive was success?
What percent population of a certain country got vaccinated?
What vaccination schemes (combination of vaccines) are used and in which countries?

What manufacturer had produced the most amount of vaccine and sold to what country?

What was the relationship between vaccination evolution and sentiment toward vaccinations?

What are the factors that influence vaccinations?
What vaccination schemes (combination of vaccines) are used and in which countries?

What country has vaccinated more people?

What country has immunized the largest percent from its population?

How progress of vaccination programmes around the World (or in a specific country) is received by the public, as reflected in the tweets about all vaccines.