Descriptive Analysis of Global Event - COVID 2019
COMP 8157 Advanced Database Topics
Guide: Dr. Pooya Moradian Zadeh

Sankalp Desai
ID: 105197946
University of Windsor
desai12r@uwindsor.ca

Maulik Tailor
ID: 105183659
University of Windsor
tailo115@uwindsor.ca

Harshal Vakharia
ID: 105190680
University of Windsor
vakharih@uwindsor.ca

Deepti Bamel
ID: 110008446
University of Windsor
bamel@uwindsor.ca

Abstract

We investigate the task of analysing the impact of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), colloquially known as the coronavirus pandemic, on global economy and compare the severity with other viruses in the history. Furthermore, we derive some important facts by finding co-relation between the trends in the economical and pandemic data.

1 Introduction

In December 2019, there were several case of patients with pneumonia with unknown causes was connected to wholesale seafood market in Wuhan, China. Human airway epithelial cells were used to isolate a novel coronavirus, named 2019-nCoV, which formed a clade within the subgenus sarbecovirus, Orthocoronavirinae subfamily[1].

The World Health Organization (WHO) declared the outbreak to be a Public Health Emergency of International Concern on 30 January 2020 and recognized it as a pandemic on 11 March 2020[2]. As of 4 April 2020, more than 1.18 million cases of COVID-19 have been reported in more than 200 countries and territories[3].

This pandemic has a powerful effect on global economy. In our project, we analyse the economical impact of CVOID-19 on economy and also do the comparison of various factors affected by it with other viruses in the history.

2 Project Details and Methodology

2.1 Definitions

Data Source: The sources from where raw data can be collected for that database or the project. For example, News Articles, World Health Organization, Data repositories etc. are data sources for this project. These can vary as data collected from the internet or surveyed information, it can be structured or unstructured, numeric or textual, video, audio or image data.

Database: The database is a storage unit where we can store our collected data. It is generally managed using a Database Management System (DBMS). There are multiple types of database depending on the structure of data to be stored and on the requirement of the project. Generally, for structured data, we use tabular databases and for unstructured data, a document database is the preferred one.

Database Replica: The replica of the database is as the name suggests the replication of the database we design. A replica will have all the data stored on the main database, the purpose of Database replication is to increase the availability of the data on multiple systems even if the database server fails, the replica would still be up and running.

BI tool: Business Intelligence tool is the one where one creates their custom graphs on the data extracted from the database. A BI tool has the feature to connect to a database and this database can be on server or on the local machine same as the BI tool. BI tool helps in creating dashboards and helps represent the data into a story format.

DBMS: DBMS is an acronym for Database Management System, DBMS is a software which allows to manage the data within a database. It also provides us with rules and the query processing language to deal with the database and perform the CRUD operations.

2.2 Specifications

Various specification of our project are:

- **Database:** Microsoft SQL server
- **BI tool:** Tableau, R

- **Data preparation:** R, python
- **Data Filtration:** Microsoft Excel

2.3 Architecture

As shown in *Figure 1*, the data has been collected form various data-sources example text articles, CSV files, data collected form website, etc. The sample database replication is shown in the figure which is used to improve the availability of data. The database is connected to Business Intelligence (BI) tool to perform analysis task and visualizations.

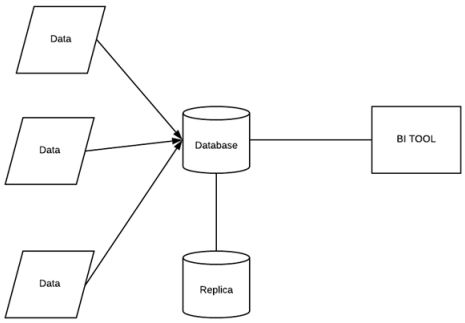

Figure 1: System Architecture

2.4 Platform

We have used the concept of RDMS. For creating graphs and data preprocessing, we have made use of Business Intelligence tools like PowerBI. For generating animated plots we used ggplot2 in R.

2.5 Design

As shown in *Figure 2*, we created a database name Pandemic which has four tables: GDP, Comparison, Cvoid-19, SARS.
GDP: contains the data of GDP of each country for Q3 and Q4 of 2019.
Comparison: It contains comparison of death, affected countries, number of cases and fatality ratio of different viruses.
Cvoid-19: Contains data of Covid-19 cases.
SARS: Contains data of SARS cases.

3 Experimental Setup

3.1 Implementation Details

There are 3 phases of implementation of this project. The first phase focuses on data collection and preparation, second focus on database design and facts are derived in last phase using data.

Figure 2: System Design

3.1.1 Phase 1

Data Collection: Collecting the data in their rawest form from multiple data sources is known as data collection. We have collected data from the internet, news articles and multiple data repositories. We used the internet as our primary source to collect the data as most of the data used in the project was collected from World Health Organization, Kaggle Datasets, CDC about covid-19 and other viruses particularly SARS. We used news articles to collect data for the comparison table of the database and data repository of OECD for the GDP data.
Data Preparation: Some of our data was unstructured as it was in textual format, we converted those data into tabular format and stored it as a flat file (csv file). The task here was to prepare the data so that it can be stored into a tabular database. Also, there was some data filtration required as some of the data was in NaN format, so it was replaced by the mean values so that the overall data was not skewed on one side.

3.1.2 Phase 2

Database Design: The database used here is a tabular database as most of the data was structured and the one which was not was converted to a structured format to store in this database.

3.1.3 Phase 3

Fact derivation: Fact Derivation was the process of deriving some hidden insights from the data we collected. This was mostly done on the BI tools by making graphs to fulfill the purpose of descriptive analytics.

3.2 Challenges

In this analysis tasks we faced three challenges during implementation. As the data is unstructured and comes from different sources like text documents like news articles, web-page and CSV

files from different data source, the first challenge was data collection. Second challenge was data filtration as the data was in different format. We converted textual data into numeric data and also parsed the flat files to store data collectively in relational database. The third challenge was handling limitation of authentic data and it's completeness.

3.3 Data Sources

Following are our data sources:

Coronavirus:
https://www.coronatracker.com/
https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset

SARS:
https://www.who.int/csr/sars/country/en/

GDP Q4 and Q3 of countries:
https://data.oecd.org/gdp/quarterly-gdp.htm

Comparison:
https://www.businessinsider.com/coronavirus-compared-to-sars\-swine-flu-mers-zika-2020-3
https://www.cdc.gov/
https://www.who.int/

3.4 Results and Findings

3.4.1 Outputs

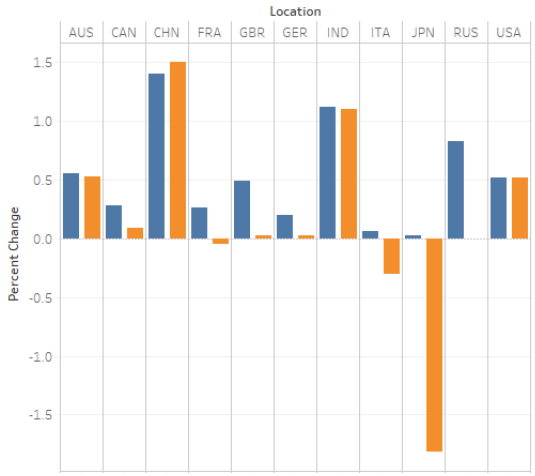

Figure 3: GDP Q3 and Q4 2019

As shown in the *Figure 3* we can see that the Gross Domestic Product(GDP) of Quarter 4 of 2019 is lesser then Quarter 3 of 2019 and we predict that the GDP of Quarter 1 of 2020 will plummet even more due to the pandemic.

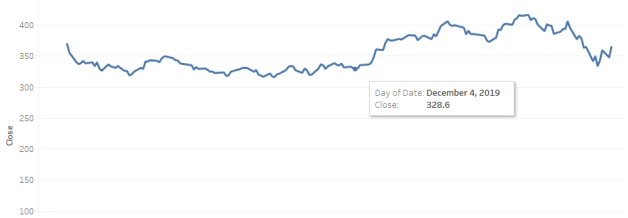

Figure 4: Tencent Holdings Ltd. stocks from August, 2019 - March, 2020

As shown in *Figure 4*, there is elevation in stock prices for Tencent Holdings Ltd. in the time period of pandemic. By this we can conclude that during the lock-down in the countries, people chooses the online games as their mode of passing time. The stocks are predicted to increases if the pandemic situation exacerbates and lock-down continues.

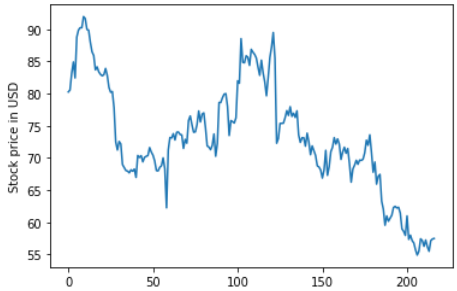

Figure 5: Stock Prices for Match Group, Inc. (MTCH) from August 2019 to March 2020

As shown in *Figure 5*, The stock prices for Match Group, Inc. (Owns multiple dating applications) has decreased since the lock-down period due to social distancing. *Figure 6* shows the impact of different viruses comparing their spread in different countries.

In *Figure 7*, the red graph is total number of cases with time, Yellow graph is number of deaths with time and Green graph is number of patients recovered with time for SARS. *Figure 8* shows the cumulative cases for Covid-19 over time till March 09, 2020 in multiple countries. We can see the sudden spike for Iran. From further analysis of news articles we found that Iran's coronavirus outbreak began in Qom, the holy city in which thousands of pilgrims arrive daily. The government's first acknowledgement of the virus was on February 19, when officials reported two deaths in Qom[4].

Figure 9 shows the merged plot of the number of

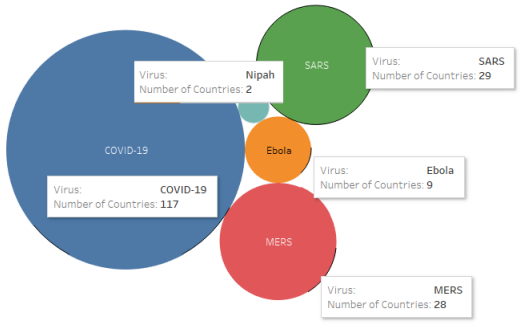

Figure 6: Magnitude of impact of other virus

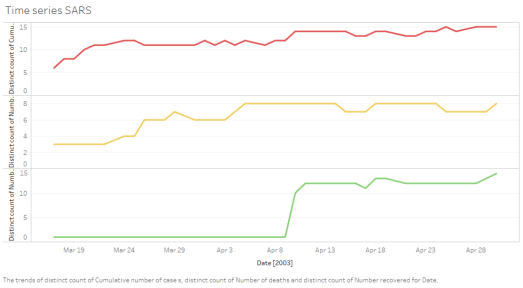

Figure 7: SARS Effect Outline

confirmed cases, recovered cases and deaths with respects to different countries. *Figure 10* and *Figure 11* shows the number of death cases and number of recovered cases with time respectively for different different countries.

3.4.2 Key Insights

- COVID-19 cases can be bifurcated into 4 categories mild, moderate, severe and critical
- As of February 20 the 80% of total cases were mild/ moderate, whereas 14% were severe and 6% were critical
- Common symptoms - Cold, Flu, Fever, Short-Breath
- Cold, Flu and Fever are the symptoms which come up within the first 2-14 days of contact of the virus
- 87.39% of the reported cases had fever, only 43.78% of the cases admitted to the hospital had a fever but eventually the number went up to 87.39%.
- 17.3% of the cases suffered from shortness of breath on day 5 after coming in contact with the virus

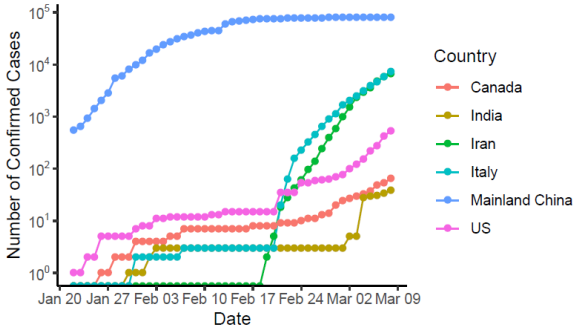

Figure 8: Cumulative Cases for Covid-19

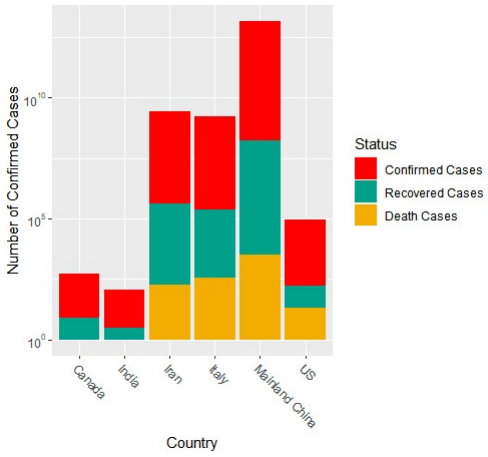

Figure 9: COVID-19 Confirmed, recovered and Death cases

- Expected inflation rate for 2020 was 24% which was 9% higher then past year, but this going to aggravate due to hoardings

4 Conclusion

Data can be very vague in the collection phase, so Data Filtration is an important or a mandatory task to make it certain and ready for storage and analysis purposes. Sometimes in some scenarios Data is not enough to derive accurate information and in some cases just the sample of data would be enough to get to our goal. In this project we were dealing with multiple data sources and thus, data filtration was important as different data sources had data in different formats. Also, for the comparison of viruses we did not require the time series of each and every virus as the final numbers were comprehensive enough to present the comparison. We could work on both structured and unstructured data together to achieve our goal but either we should convert structured into unstructured or vice-versa so that database selection is easy and we can use a single way of CRUD op-

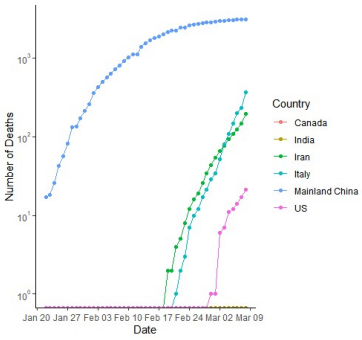

Figure 10: Cumulative death over time for Covid-19

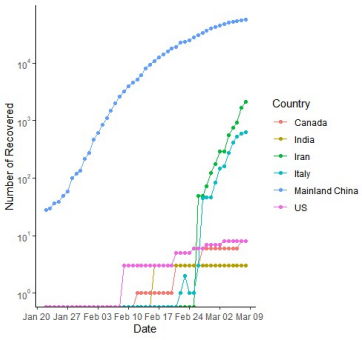

Figure 11: Cumulative recovered cases over time for Covid-19

erations throughout our project.

Coronavirus or Covid-19 is a pandemic which has catastrophic effects on the majority of demographics as well as businesses and particularly food, tourism, travel but on the other hand there has been some upsides in the businesses of retail, pharmacy. Although the number of cases of Covid-19 are much more than other pandemics and viruses, the mortality rate of the pandemic is much less in comparison. Also, through the time series of the number of cases we can see that for many countries after a period of time the curve flattens.

References

[1] C. P. E. R. E. Novel *et al.*, "The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (covid-19) in china," *Zhonghua liu xing bing xue za zhi= Zhonghua liuxingbingxue zazhi*, vol. 41, no. 2, p. 145, 2020.

[2] W. H. Organization *et al.*, "Statement on the second meeting of the international health regulations (2005) emergency committee regarding the outbreak of novel coronavirus (2019-ncov)," 2005.

[3] "2019–20 coronavirus pandemic." https://en.wikipedia.org/wiki/2019%E2%80%9320_coronavirus_pandemic. Accessed: 2010-09-30.

[4] "The coronavirus has caused a full breakdown in iran, with an unknown death toll, infected leaders, and massive burial pits visible from space." https://www.businessinsider.com/iran-coronavirus-covid19-deaths\-cases-updates-2020-3. Accessed: 2010-09-30.