



COVID-19 IMPACT ANALYTICS

ON COUNTRY BASIS

MAY 2020

AGENDA

The overall workflow of the story telling and model building



DEFINE THE PROBLEM The covid-19 information and final goal

DATASET SELECTION Input data analysis and processing



DATA COMPRESSION AND CLUSTERING

ESTABLISH THE MODEL Fit the regression model



RESULT ANALYSIS The prediction VS actual cases, dashboards

“THE WHITE SPACE” Uncovered topics in the project



HOW COVID-19 CRISIS IMPACTED

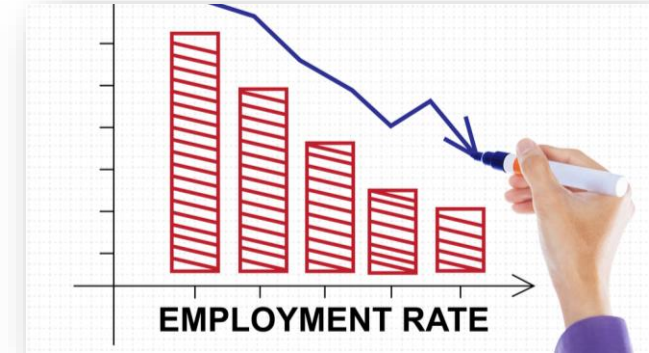
The covid-19 information and final goal of our project



- ✓ The Novel Coronavirus has unprecedented impact on the Global Economy.
- ✓ As of now, there are **3.6 million** cases reported worldwide with the **US** having the highest number as **1.26 million** cases.



- ✓ Our **goal** is to calculate the **performance** of chosen countries in the covid-19 period and thus predicting the impact of COVID-19 on the GDP for the years 2020, 2021 and 2022 via **regression model**.



- The impact is analyzed by taking into consideration following **features**:
- Unemployment rate
 - Urban Density %
 - Population% over 65 years
 - Income per person
 - Cases per million



**Hypothesis:
The Virus may have disparate
impact on different countries.**

DATA SELECTION

Analysis of the input data to create a more reasonable clustering



Step 1: Search for data

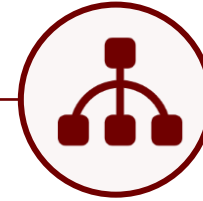
Using five features:

- Unemployment rate
- Urban Density %
- Population% over 65 years
- Income per person
- Cases per million

Step 2: Clean the Data

Once we gathered data from online sources, clean it based on

- Get the time series from 2006 to 2019
- Get the unified countries from five dataset

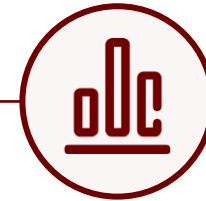


Step 3: Cluster the Data

Use two features to create cluster

- Cluster the data of aged population
- Cluster the date of income

Step 4: Select a training set based on the clustering



DATA COMPRESSION AND CLUSTERING

Clustering by Kmeans and Country selection



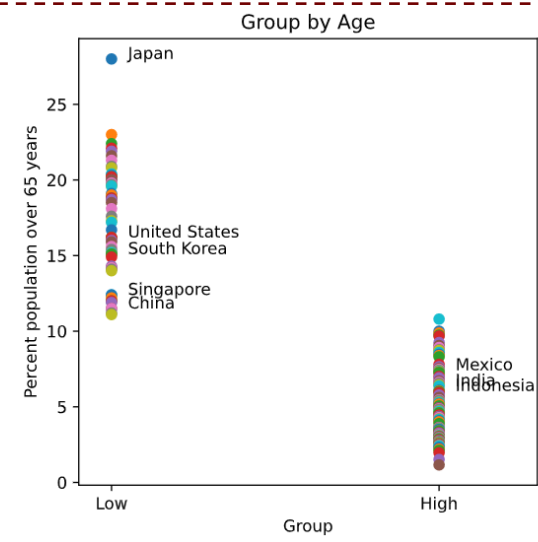
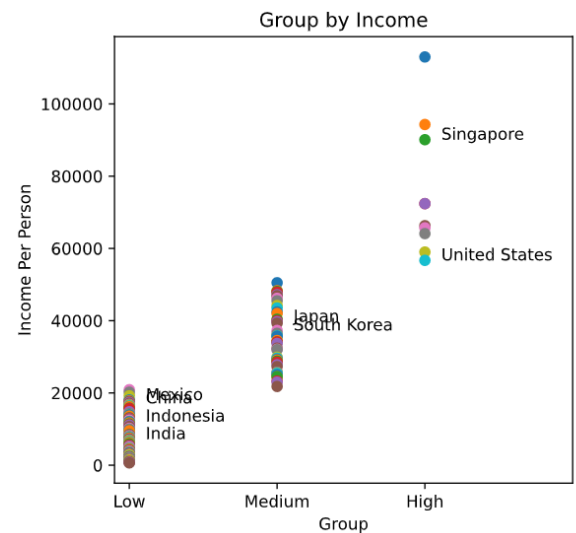
- ✓ Select dataset between 2008-2019.
- ✓ Drop the country with missing data.
- ✓ Create a country list with data available in all dataset.

Population % of age > 65

High elder rate	Low elder rate
Mexico India Indonesia	US South Korea Singapore China

Income per person

High	Medium	Low
US Singapore	Japan South Korea	China India Indonesia Mexico



ESTABLISH THE MODEL

Test and train the regression model to make it qualified in the next step of estimation

$$Y_{with\ impact} = Y_{without\ impact} - \Delta y$$

$$\Delta y = \left[\frac{case}{population}, unemployment, percent\ of\ 65\ yrs\ old, income\ per\ person, urban\ density \right] \times w + b$$

RESULT ANALYSIS

Use the model to calculate the impact. Δy result for the 8 countries

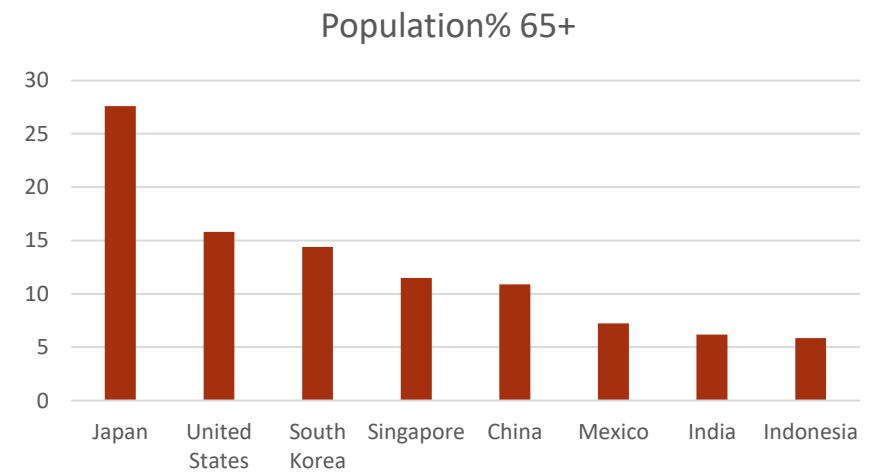
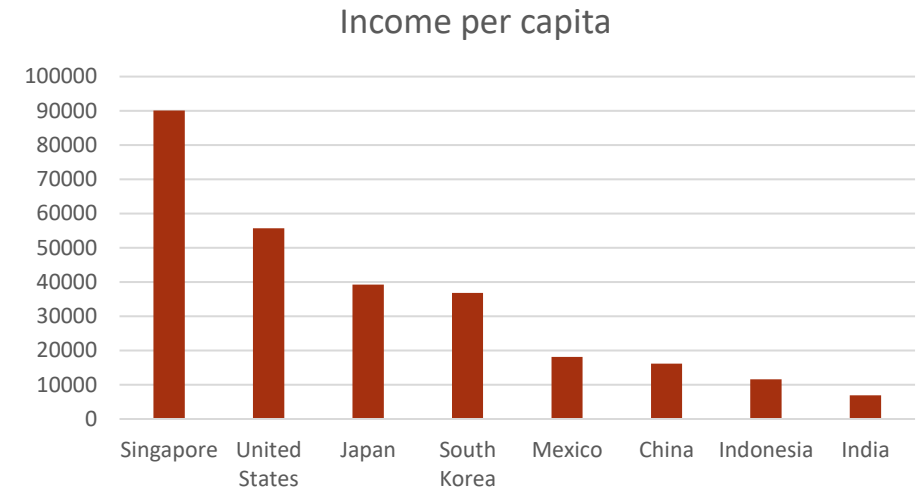
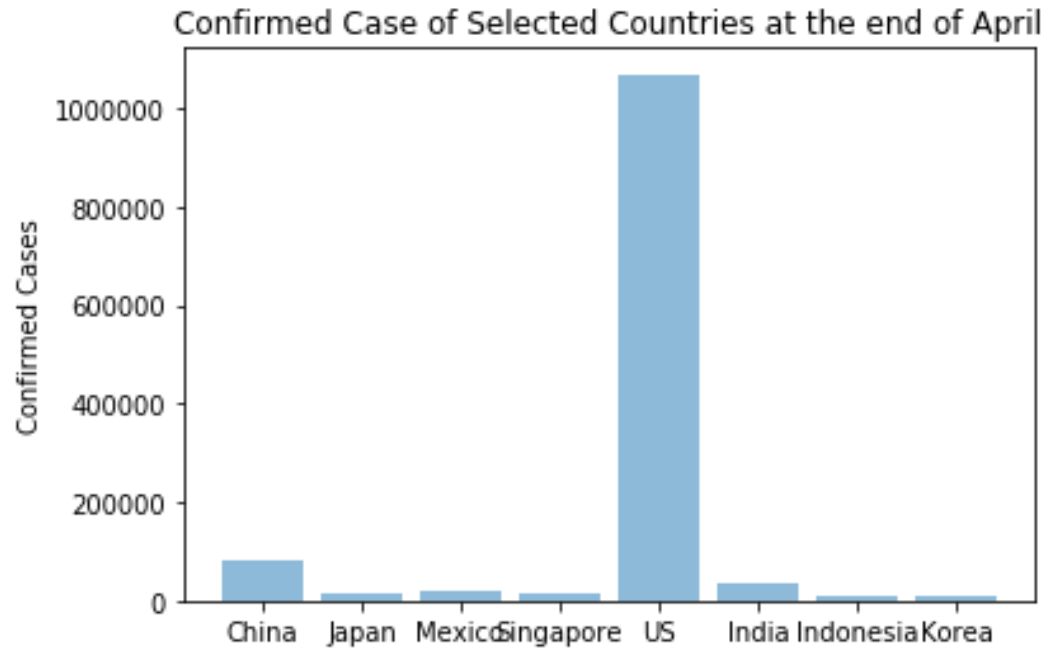
Result Generated from the Model :

China (12808923076923.125, 206442306302.5891, 13015365383225.715, 0.016117069722631147)
Japan (6239890109890.102, -462588923491.29767, 5777301186398.804, -0.07413414584947634)
Mexico (1391973626373.6328, -319929407228.3962, 1072044219145.2366, -0.22983869892842418)
Singapore (382004395604.39453, -686976659127.3943, -304972263522.99976, -1.7983475243537)
United States (18429010989011.0, -1402162571947.5115, 17026848417063.488, -0.07608452633641623)
India (3193516483516.5, 145779226345.9466, 3339295709862.447, 0.045648496601910026)
Indonesia (1304771428571.4219, 136105639829.10918, 1440877068400.531, 0.10431378006041224)
South Korea (1528446153846.1562, 69602300421.0035, 1598048454267.1597, 0.0455379473106445)

- **Based on the model, the impact on GDP from Covid-19 for the eight countries range from 0.34% to 44.99%. (*In order to better analysis the data and keep the accuracy of our model, we decided to remove Singapore from the discussion because its result does not fit in with the others)**
- **China has the least impact on GDP and Singapore has the most impact of GDP**
- **However the impact on GDP does not reflect the five parameters.**

RESULT ANALYSIS

From the prediction, we find that the highest income country with high elder population suffered most from the COVID-19.

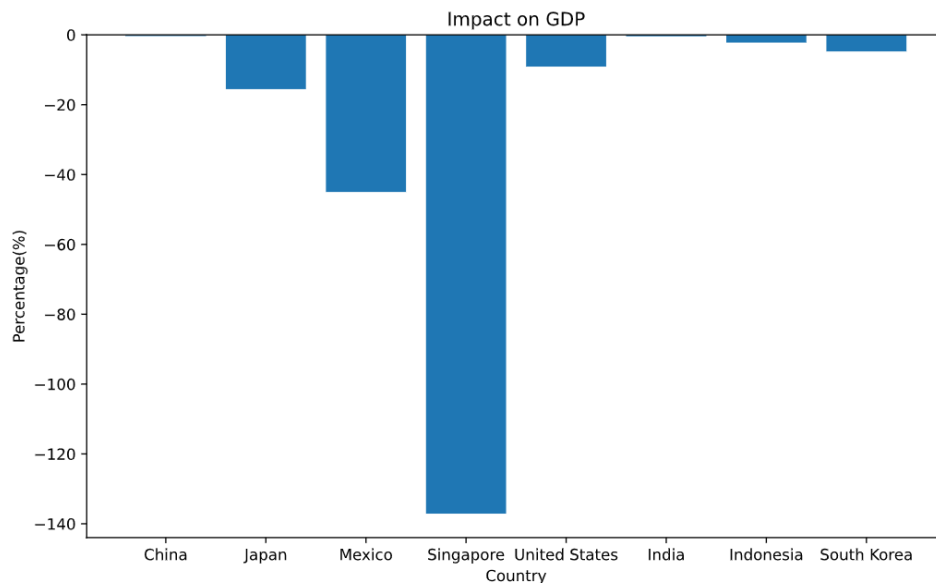


RESULT ANALYSIS

Our prediction on selected countries

Prediction

Using the 5 parameters we choose, model will be able to predict the future impact on GDP.



Test Hypothesis

We used **Norway** similar to **Singapore** to test the prediction model.

Assumed Norway: 10,000 cases in future.

Result: -82.89%

Conclusion: It's hold our hypothesis that **“that the highest income country with high elder population suffered most from the COVID-19.”**

(505536263736.2637, -419055765504.87787, 86480498231.3858, -0.8289331459780255)

“THE WHITE SPACE”

Topics and areas that we haven't covered because of lacking information and valid data sources

01

Model Application

The model does not work well with every country. There are outliers (such as Singapore) that the model cannot predict accurately.

02

Lack of Data

Due to the current resources, we are unable to get an accurate prediction on how many case each country will have.

03

Parameter limitations

We only took the most important five parameters our group thought. There are other factors such as education, politics that can also affect the final result.

04

Fast-Developing Cases

The pandemic is still developing and changing fast in many countries, therefore prediction at the current stage may not reflect the final case.

05

Resource Limitation

Counties with more severe Covid-19 situation tends to have more comprehensive data. Therefore the resource and data can be biased.



THANK YOU!