COVID-19

Analysis and Prediction

Group 1:

Nguyễn Trọng Tính (1970219)
Vũ Trung Hiếu (1970215)
Võ Tiến Đạt (1970214)
Nguyễn Tấn Nam Anh (1970013)
Nguyễn Tuấn Anh (1770153)

Content

1. Visualize Data
2. Time-Dependent SIR Model
3. Linear Regression Model
4. COVID Impact to Economy

Data source

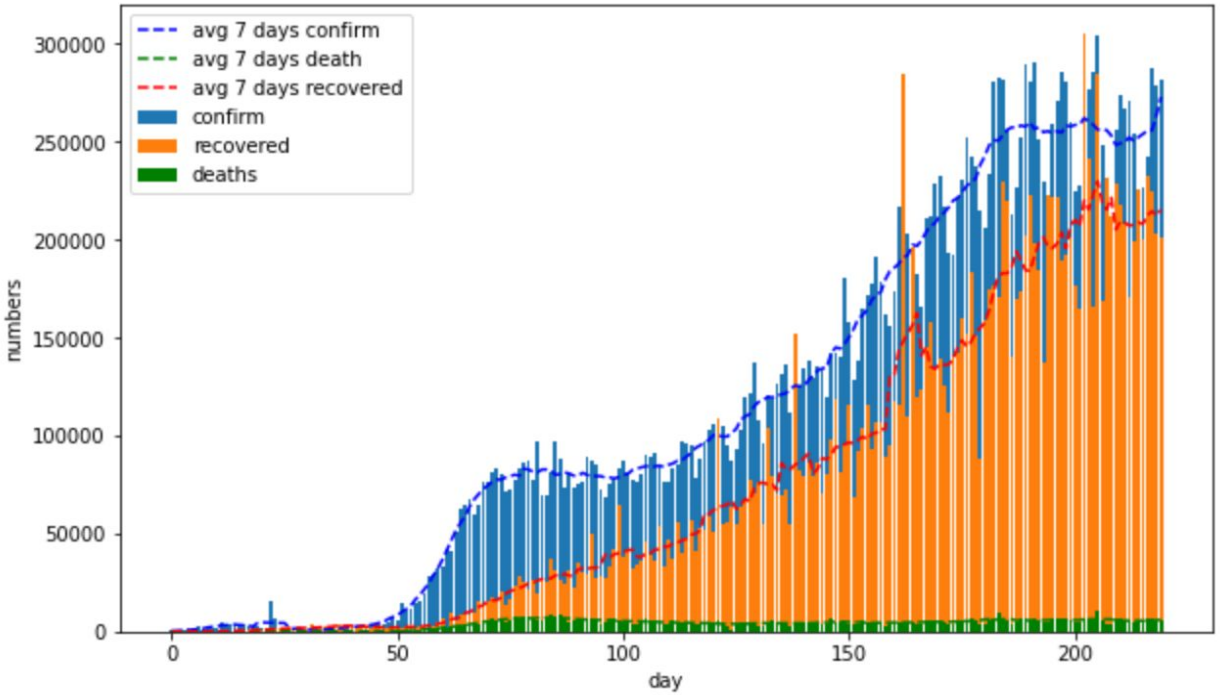
[JHU CSSE COVID-19 Data]

	Province/State	Country/Region	Lat	...	8/26/20	8/27/20	8/28/20
0	NaN	Afghanistan	33.93911	...	38113	38129	38140
1	NaN	Albania	41.15330	...	8927	9083	9195
2	NaN	Algeria	28.03390	...	42619	43016	43403
3	NaN	Andorra	42.50630	...	1098	1098	1124
4	NaN	Angola	-11.20270	...	2332	2415	2471

[5 rows x 224 columns]

1. Visualize Data

In the world

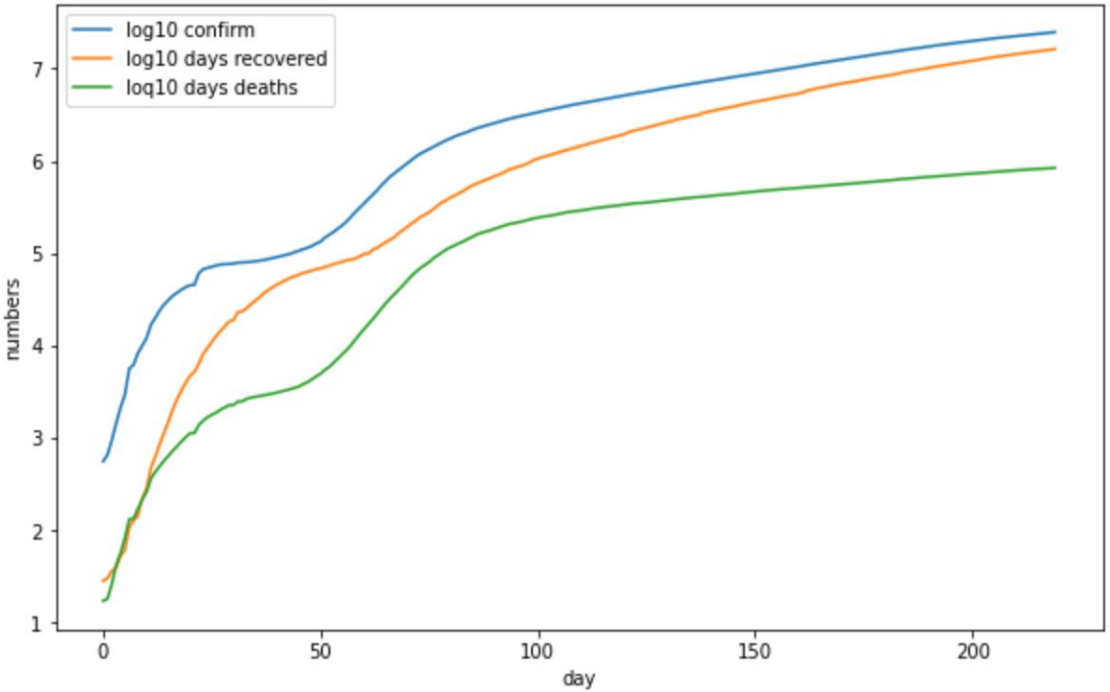


- At the end 1/2020, there are the first case

- 4/2020 - 8/2020, the pandemic start outbreak and spread

1. Visualize Data

The world



There are the two outbreak stages:

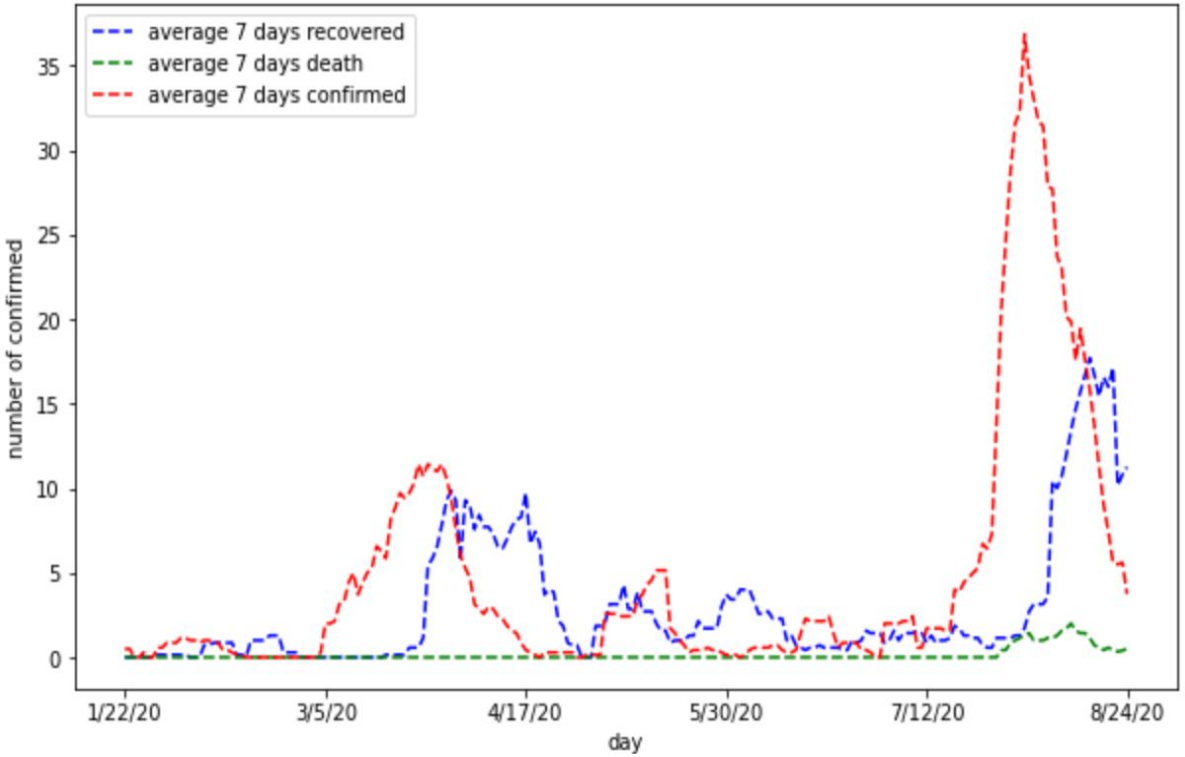
- Asia
- EU, America, Other

1. Visualize Data

Vietnam

Two stages:

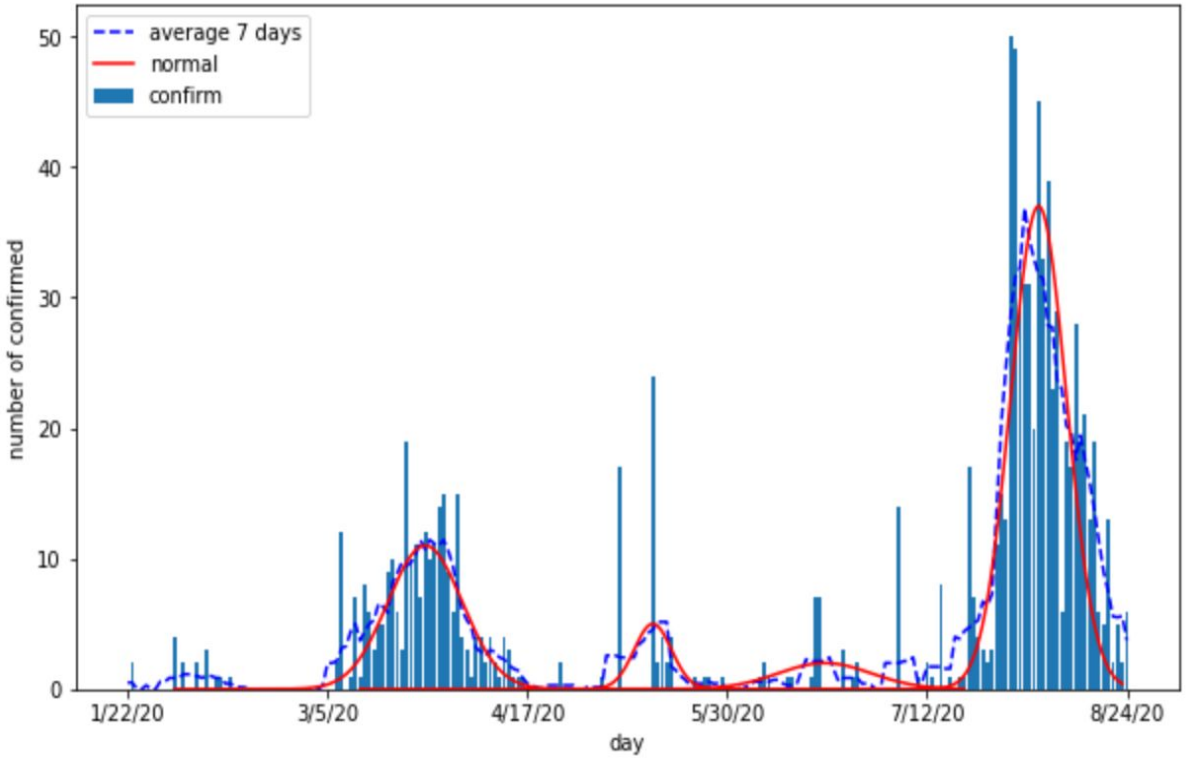
- 3/5 - 4/17: Related to BN17
- 7/15 - Now: welcome international tourist, relax social distance



1. Visualize Data

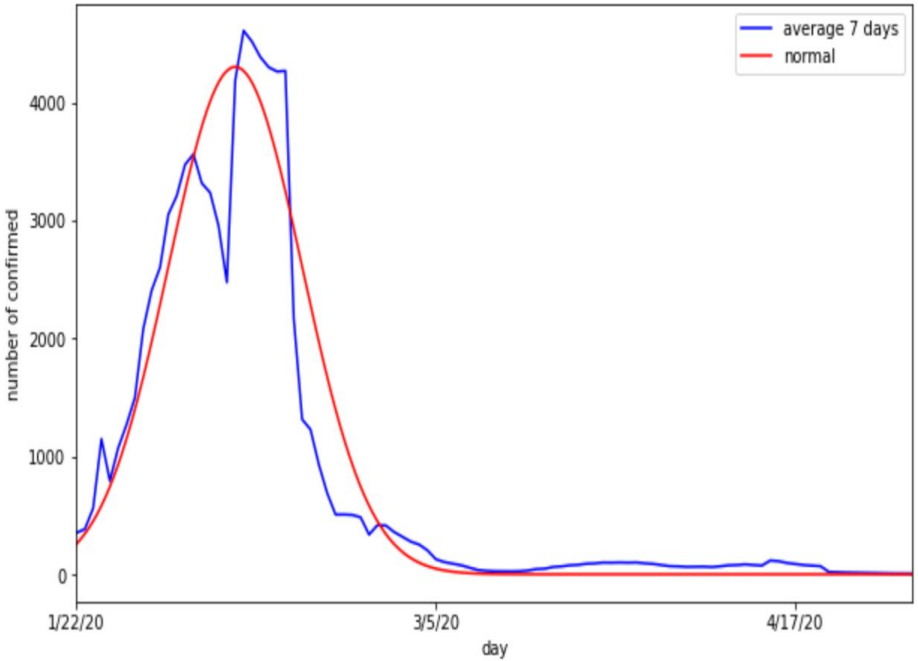
Vietnam

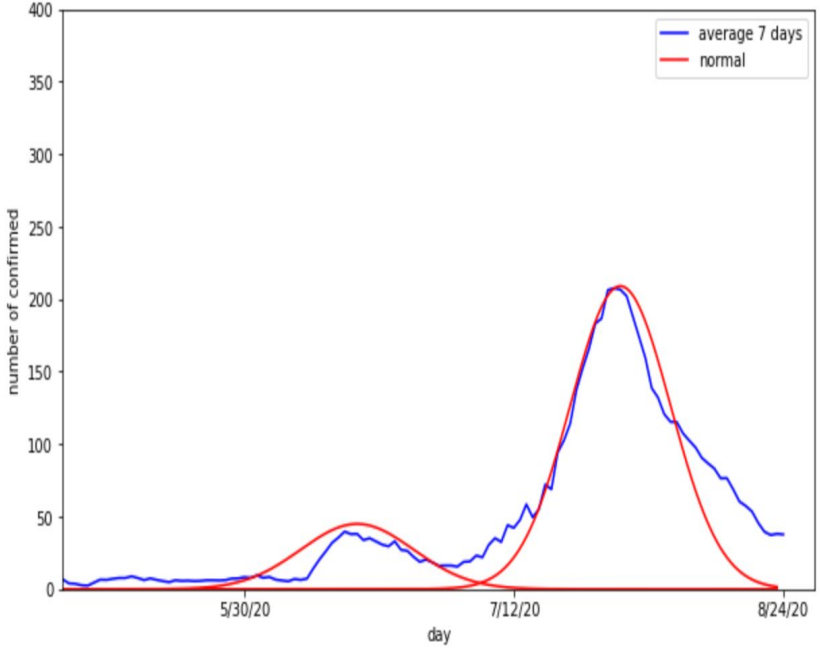
The distribution of new case data is aggregate normal distributions



1. Visualize Data

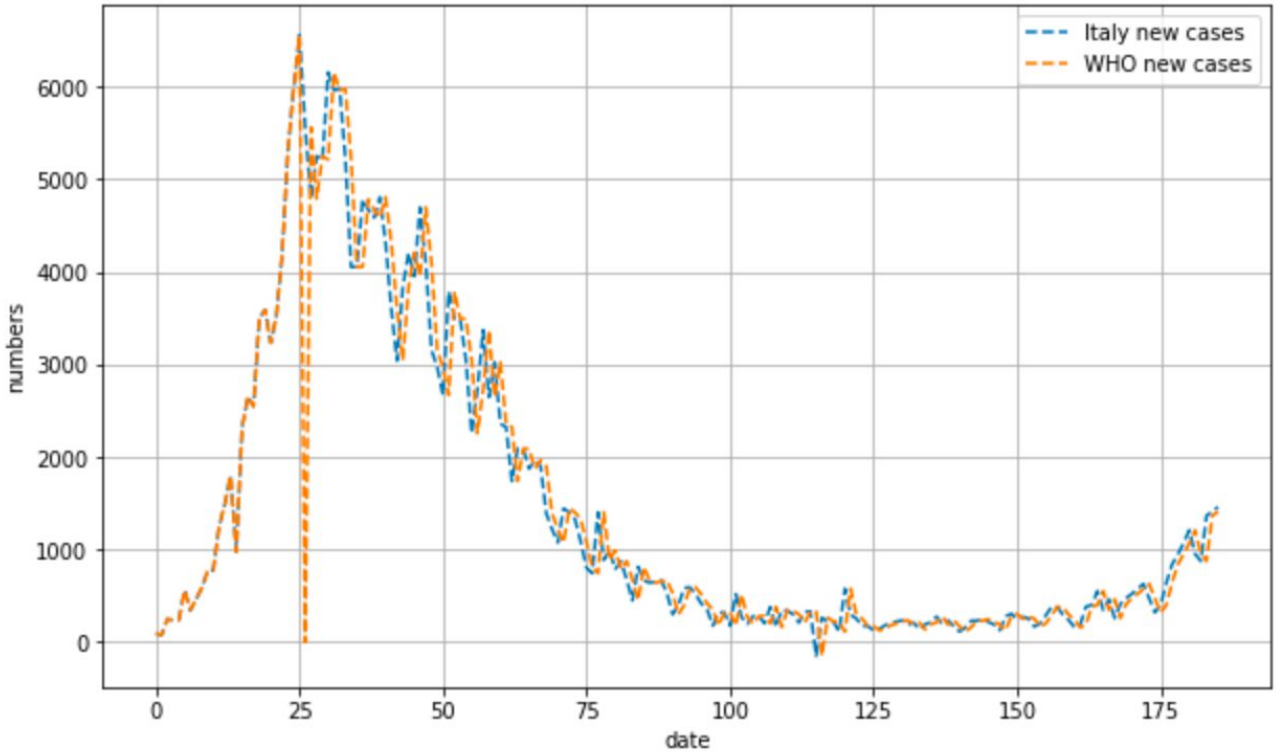
There are some result from China





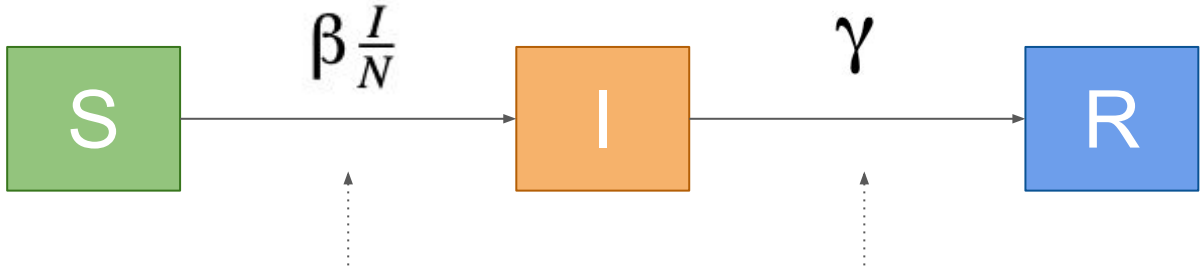
1. Visualize Data

WHO versus Italy Data



2. Time-Dependent SIR Model (1)

- SIR (**S**usceptible, **I**nfected, **R**emoved) is an epidemic model



- We need to determine: **Infection Rate** and **Recover (or Die) Rate**
- Two important assumptions:

    1. The population N is constant: N = S(t) + I(t) + R(t)
    2. Group R is immune.

2. Time-Dependent SIR Model (2)

- Discrete Equations:

$$S(t+1) - S(t) = -\beta(t)\frac{I(t)}{N}S(t)$$
$$I(t+1) - I(t) = \beta(t)\frac{I(t)}{N}S(t) - \gamma(t)I(t)$$
$$R(t+1) - R(t) = \gamma(t)I(t)$$

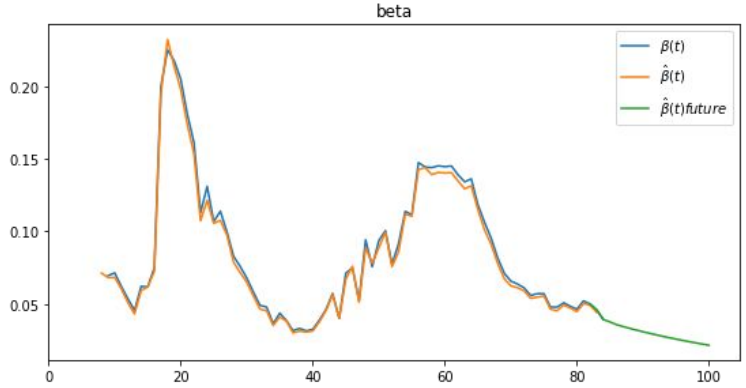
- Parameter Equations:

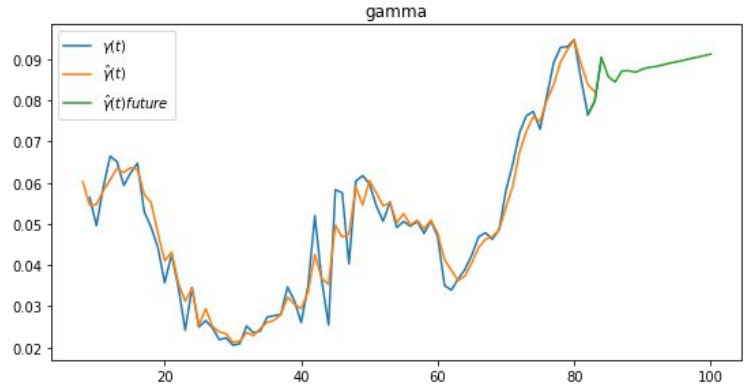
$$\beta(t) = S(t) - S(t+1)\frac{N}{I(t)S(t)}$$
$$\gamma(t) = \frac{R(t+1)-R(t)}{I(t)}$$

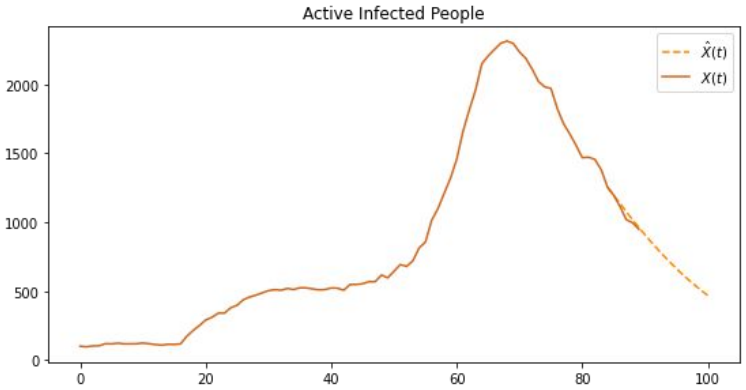
(**β**: Transmission Rate **γ**: Recovery/Death Rate)

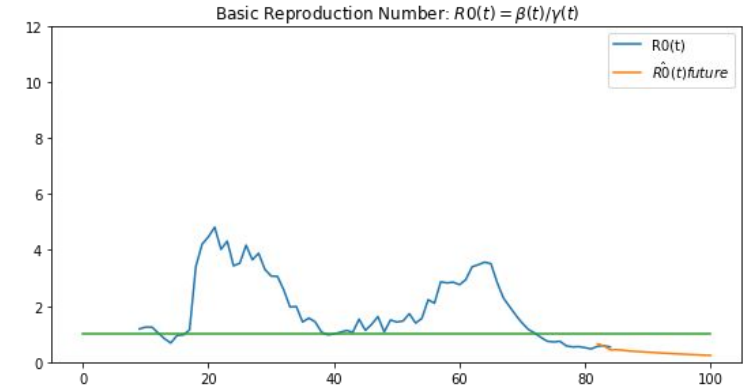
- Use machine learning to estimate **β**(t) and **γ**(t)
- Use the estimated **β** and **γ** to predict the future values of S, I, R group.

2. Time-Dependent SIR Model for China

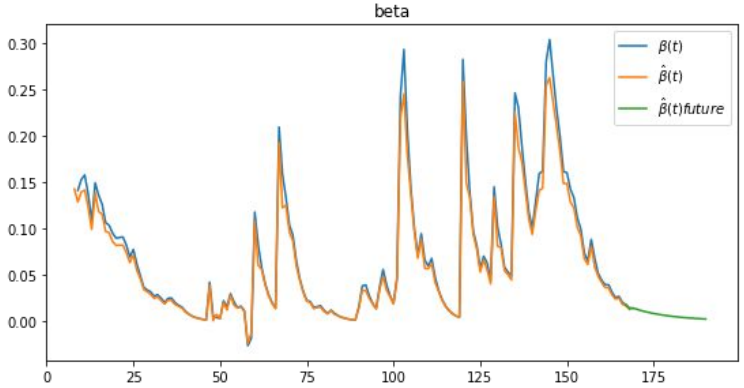


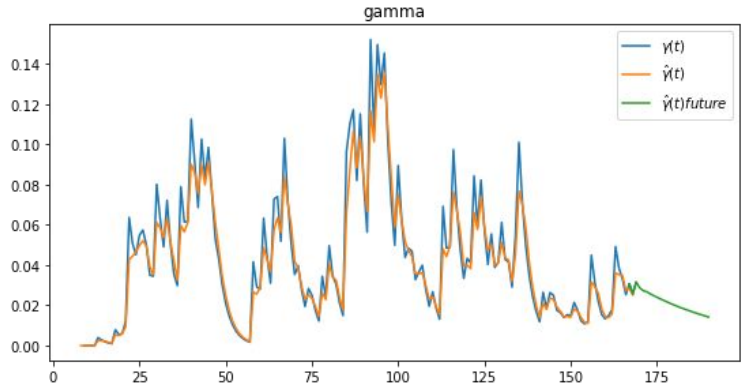


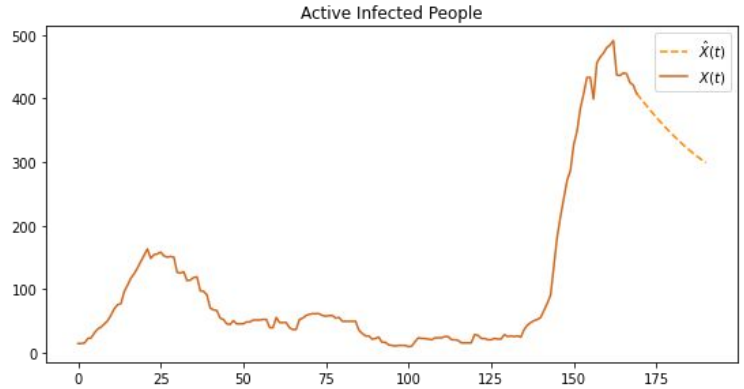


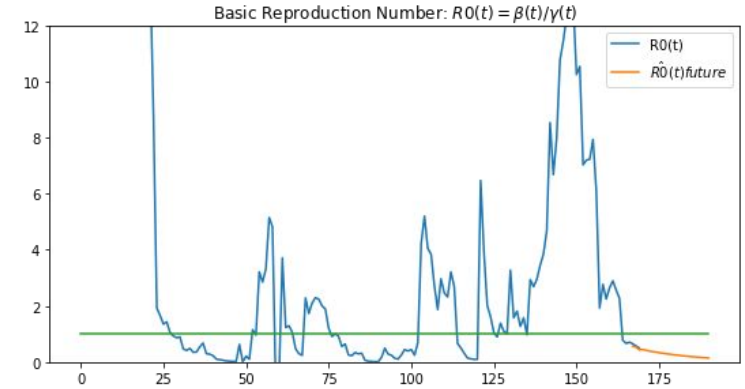


2. Time-Dependent SIR Model for Vietnam









2. Time-Dependent SIR Model: Pros/Cons

Pros:

- Better than SIR with fixed β and γ
- Tracking and predicting $\beta(t)$ and $\gamma(t)$ to give more situation detail

Cons:

- Sensitive with noise
- Have low precision if applying for large area.

3. Linear Regression

A linear function has the form: $\hat{y}(w, x) = w_0 + w_1 x_1 + \ldots + w_p x_p$

Objective of linear regression: $\min_w ||Xw - y||_2^2$

A univariate polynomial function has the form:

$P(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_2 x^2 + a_1 x + a_0$

(n + 1 parameters)

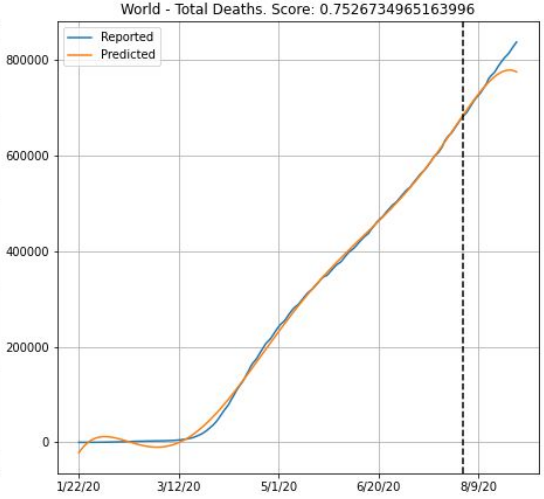
3. Linear Regression

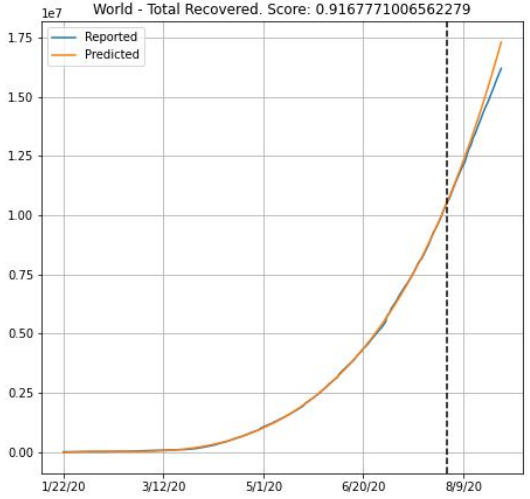
y(x): number of total cases/deaths/recovered cases as of day x since first day of collecting data (Jan 21st 2020 if using JHU CSSE COVID-19 Data).

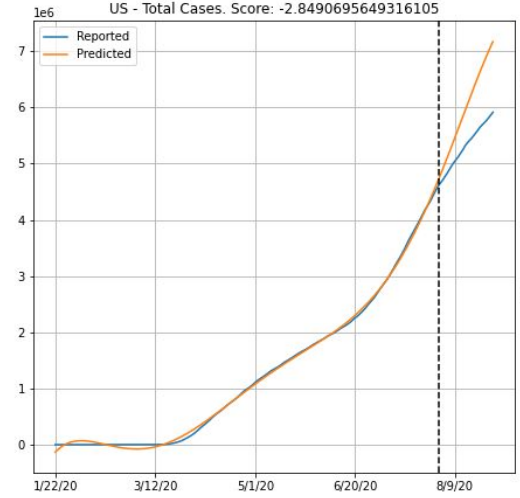
Approach: Fit a n-degree polynomial function to the data before start day, then predict.

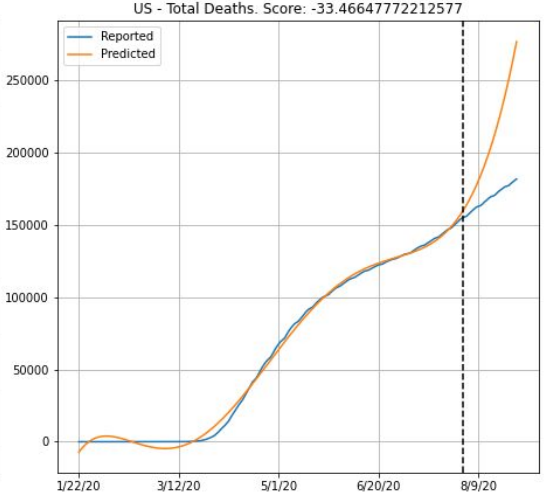
Linear regression model used: n = 6, day start = Aug 1st 2020

Worldwide & US

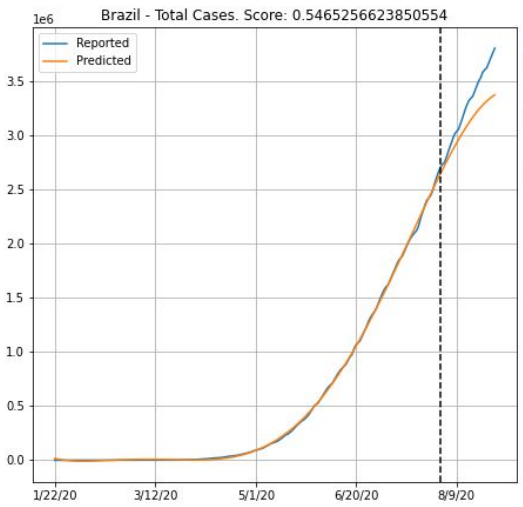


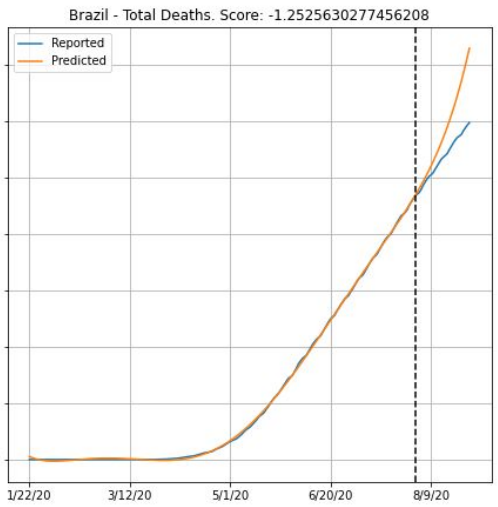


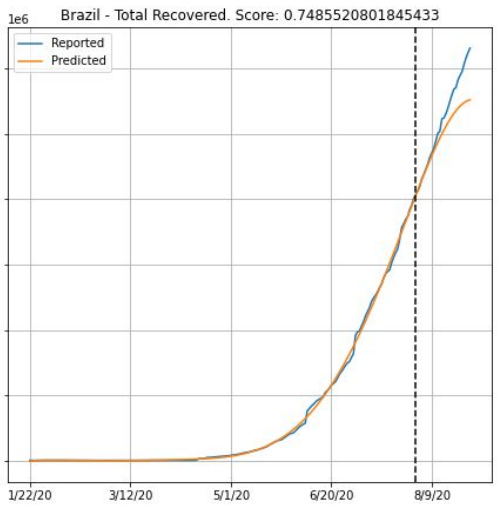


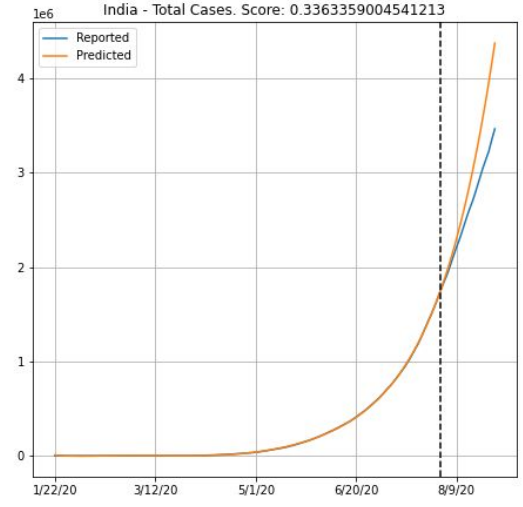


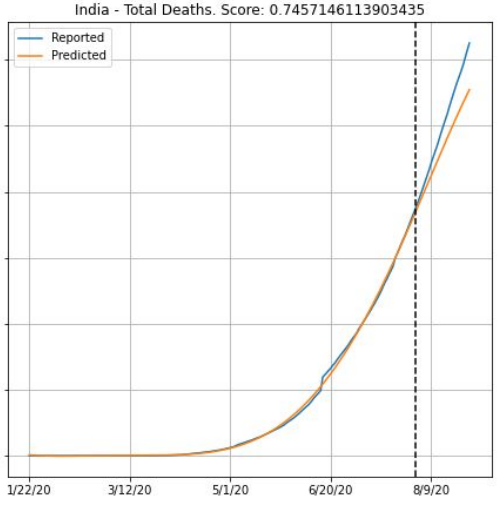
Brazil & India

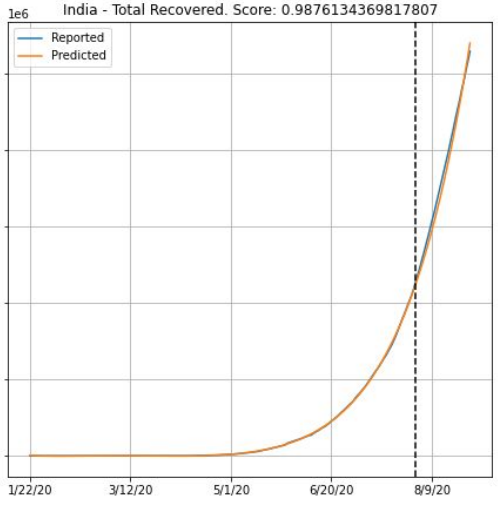




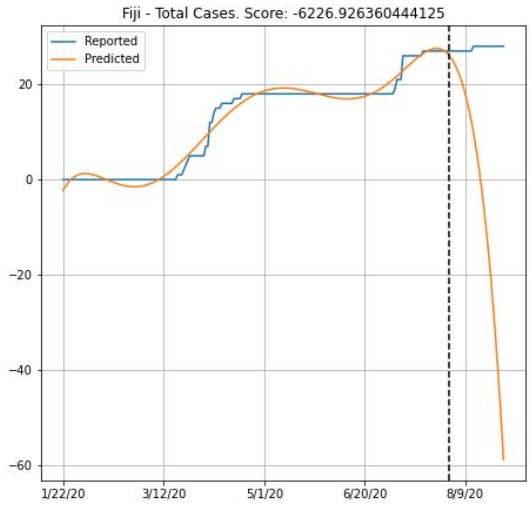


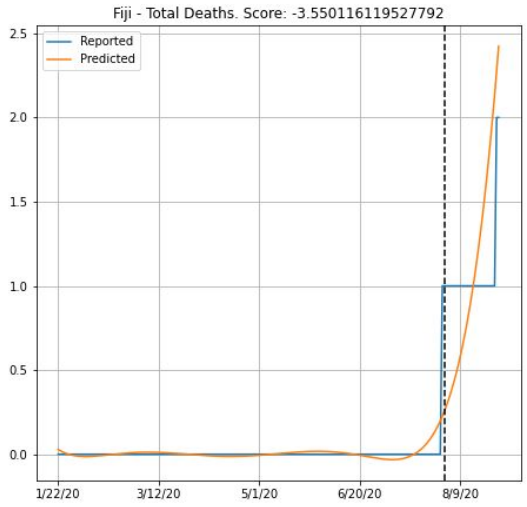


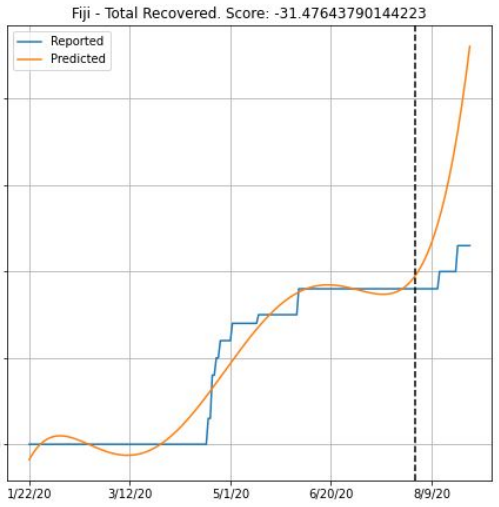


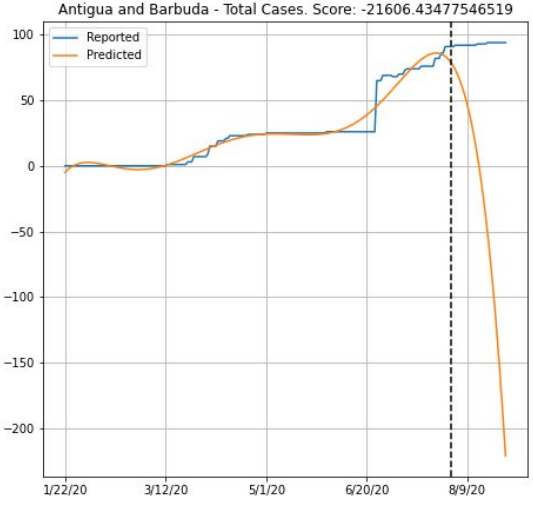


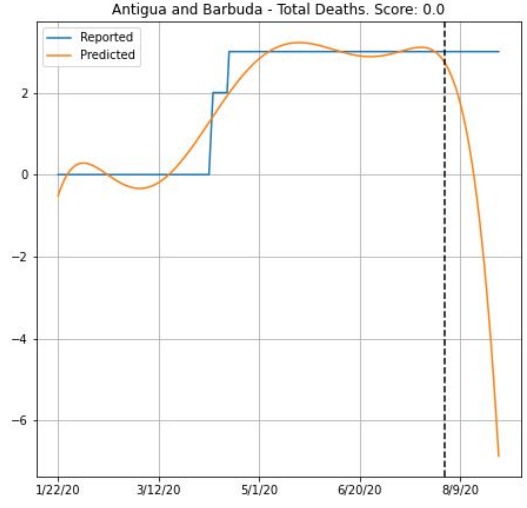
Fiji & Antigua and Barbuda

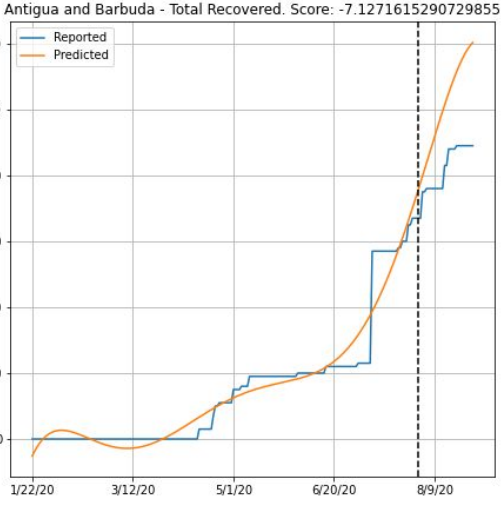












3. Linear Regression

Comments: Why reported data is not actual data

- Tests do not cover all population
- Errors in testing procedures & equipments
- For detected cases: unclear date of infection
- Unclear date of recovery
- Some countries do not report recoveries, e.g UK, Sweden, Netherlands

3. Linear Regression

Pros:

- very simple
- univariate → complexity only scales linearly with n

Cons:

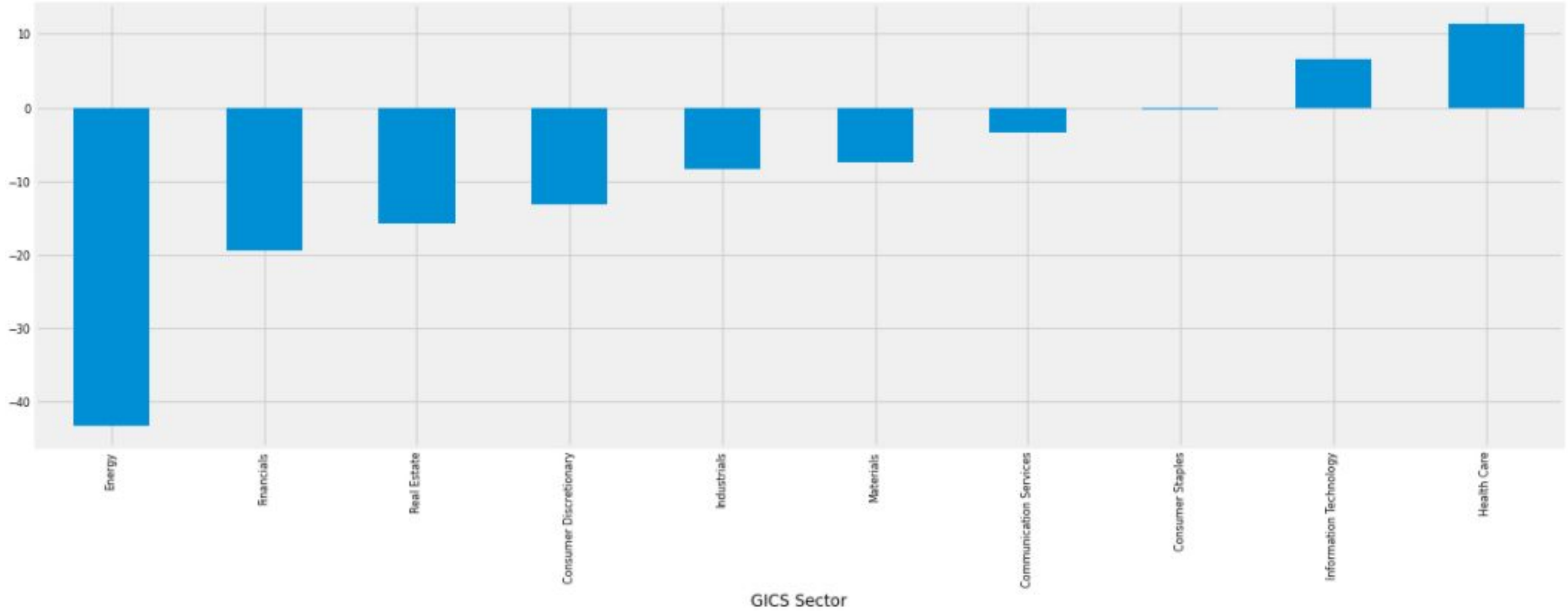
- must search for number of degree
- easy to overfit if n is large or few changes in data

4. COVID Impact to Economy

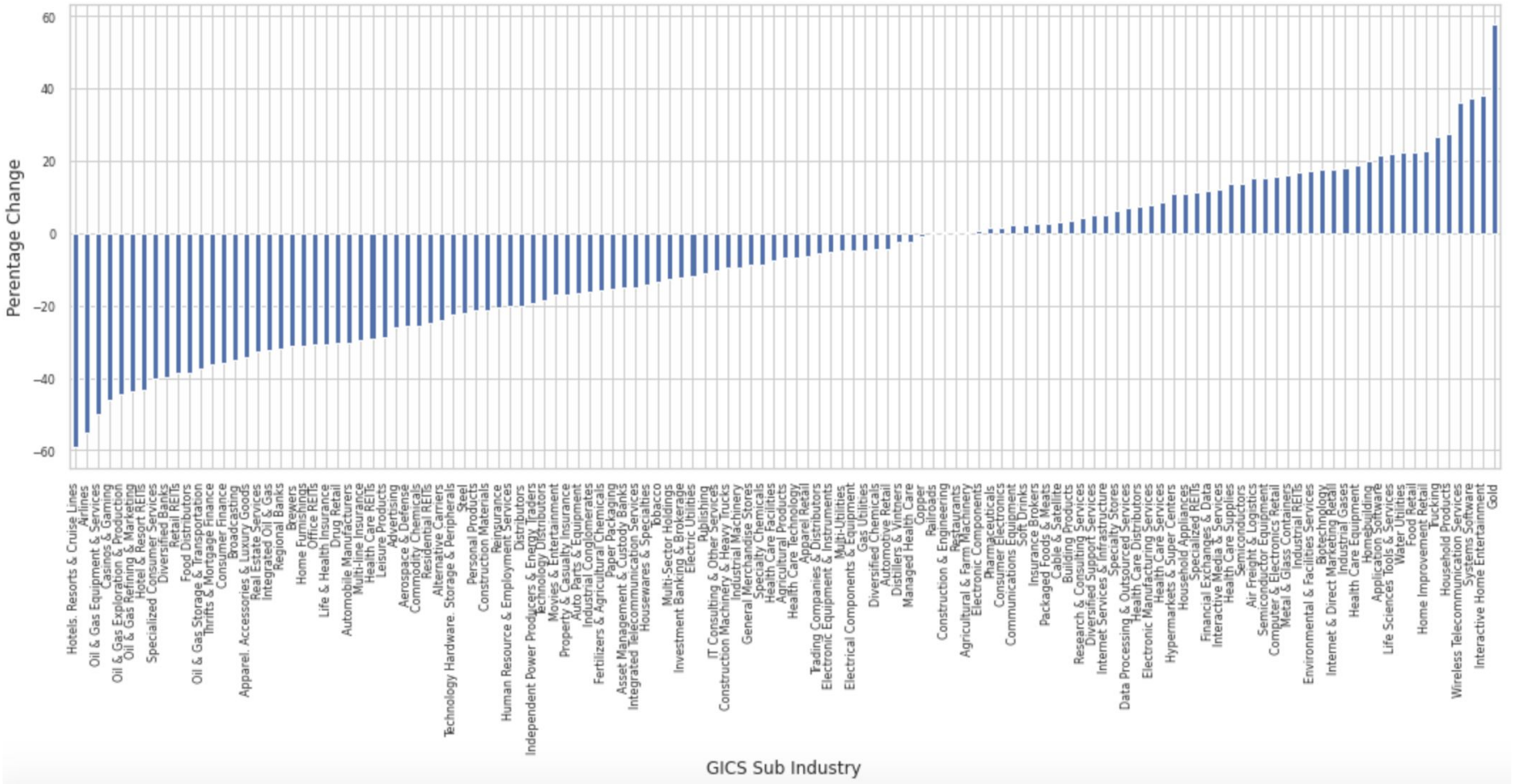
The energy sector was the hardest hit with a 43% average drop in stock prices.

Only Health Care and IT was grow up

Source: S&P 500



4. COVID Impact to Economy



4. COVID Impact to Economy

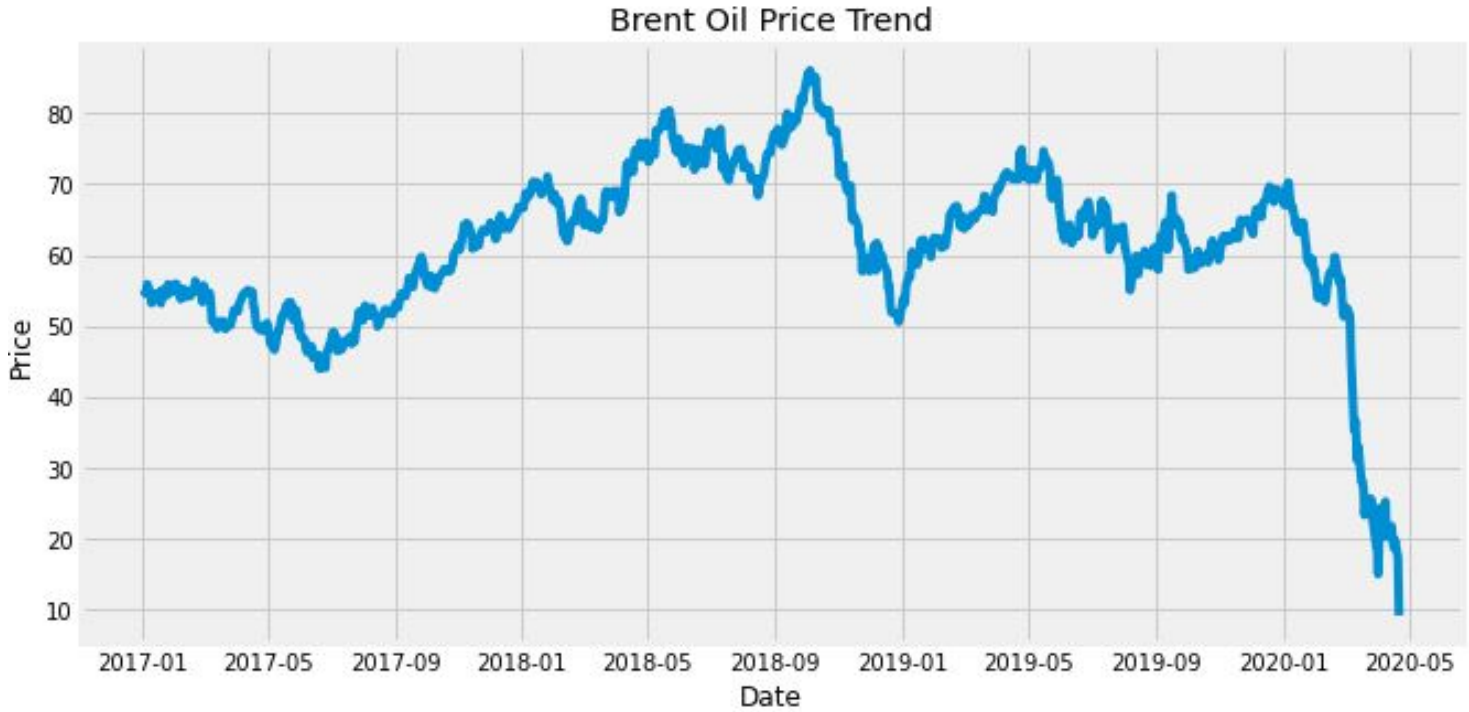
The trend showed:

- Airports plummeting since early day of March;
- India almost stopped all traffic (VABB, VIDP).

4. COVID Impact to Economy

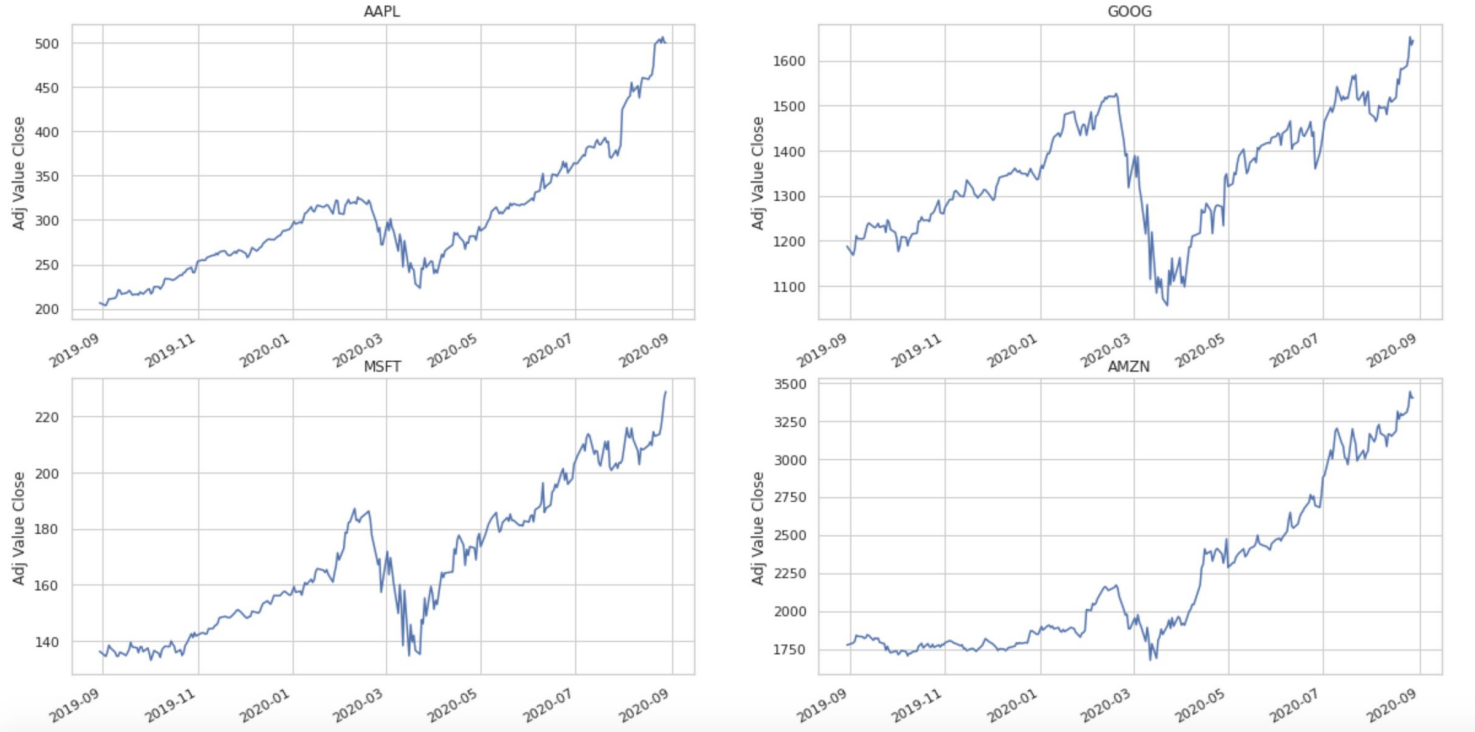
Relationship between World total Covid cases and price.

There is a steep decrease in price from Mid-Jan 2020 to Mar 2020.



4. COVID Impact to Economy

Top company technology grow up value in early 2020



Reference

[1] Yi-Cheng Chen. “A Time-dependent SIR model for COVID-19 with Undetectable Infected Persons”.

[2] Luca Magri and Nguyen Anh Khoa Doan. “First-principles Machine Learning for COVID-19 Modeling”.

[3] Adilmoujahid. “Analyzing the Impact of Coronavirus on the Stock Market using Python, Google Sheets and Google Finance”