Data Incubator Capstone project Presentation

Project Title:

Constructing a model to predict the number of COVID-19 infected cases based on some factors

Prepared by

Aklilu z Welegebrael

- Objective
  - Construct a model to predict the number of infected cases
  - Examine the trend of infected and deaths people along time
  - To identify risk factors which could inflate the number of infected cases
- Motivation of the study
  - Currently COVID-19 is a serious pandemic disease as a result many people is infected, hospitalized and dying
  - It is spreading rapidly through out the world from time to time

Introduction

- ☞Corona virus Disease 2019 (COVID-19) is the newly headache infectious disease of the world.
- ☞Plenty of people have been infected, hospitalized and dying from time to time.
- ☞It was identified in Hubei province of china in December 2019 for the first time
- ☞As a result of this pandemic disease, many individual become unemployed and the economy of every country is inflonced

➢ Data structure

- In this study two data sets will be used:

  - ✓ World-wide COVID-19 record for about 206 countries
  - ✓ US COVID-19 record for counties of each US states which has experienced COVID-19 dieseases

❑ Exploratory Data Analysis

❖Table 1.1 describes the summary statistics of Covid-19 in the worldwide data set

✓ There are about 1,873,274 infected and 118,853 dead individuals.

✓ The mean of infected and dead individuals are 174.404 and 11.06 respectively

✓The standard deviation of infected and dead individuals are 11.065 and 87.275 respectively

Table 1.1:**Summary statistics of the world-wide COVID-19 data**

	Observation	Mean	Std	min	max
Infected Cases	1873274	174.404	1324.235	0.0	35527.0
Deaths	118853	11.065	87.275	0.0	2087.0

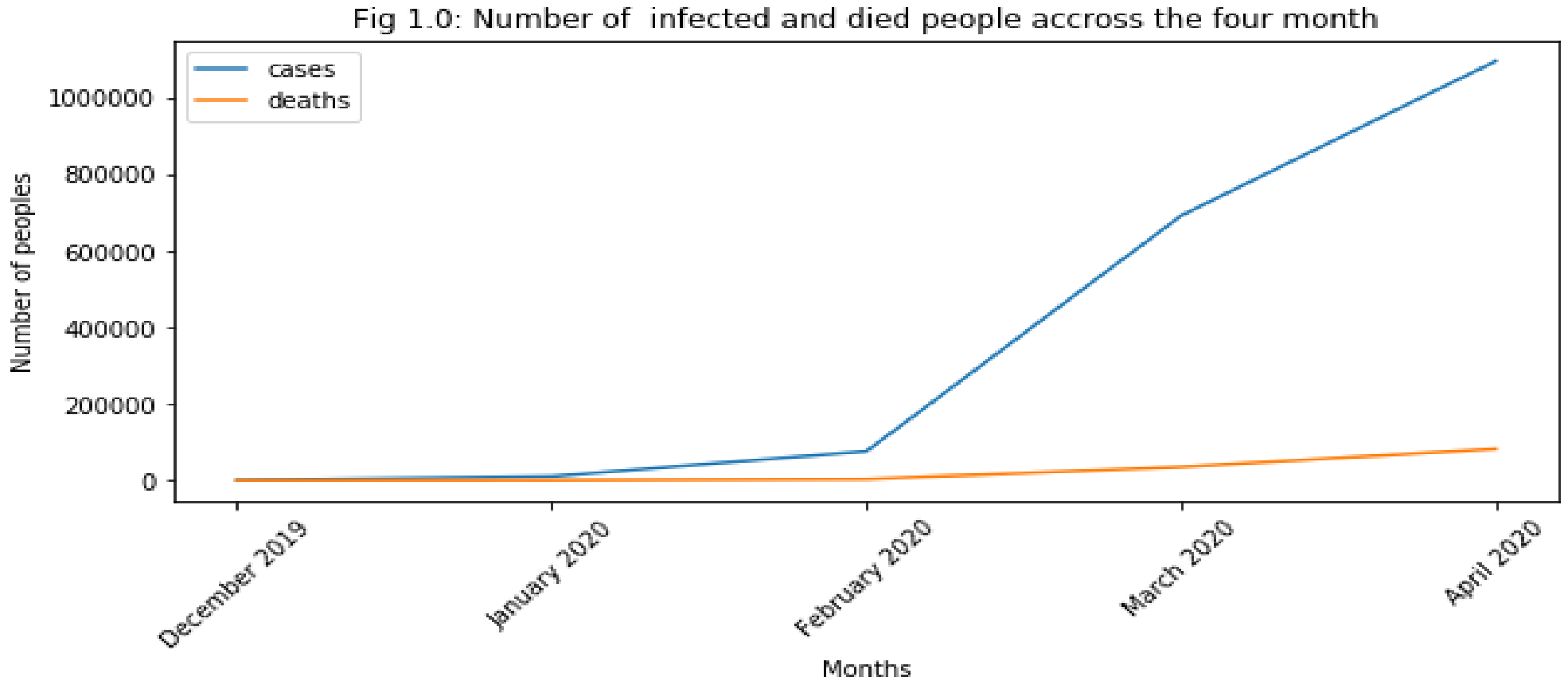
Table 1.2: Distribution infected and dead individuals in each months

	Dec-2019	Jan-2020	Feb-2020	Mar-2020	Apr-2020
Infected Cases	27	9799	75377	691939	1096132
Deaths	0	213	2708	34350	81582

❖Table 1.2 describes the distribution of infected and dead individuals across each month.

✓The number of cases and dead individuals increasing a cross the months from December to April.

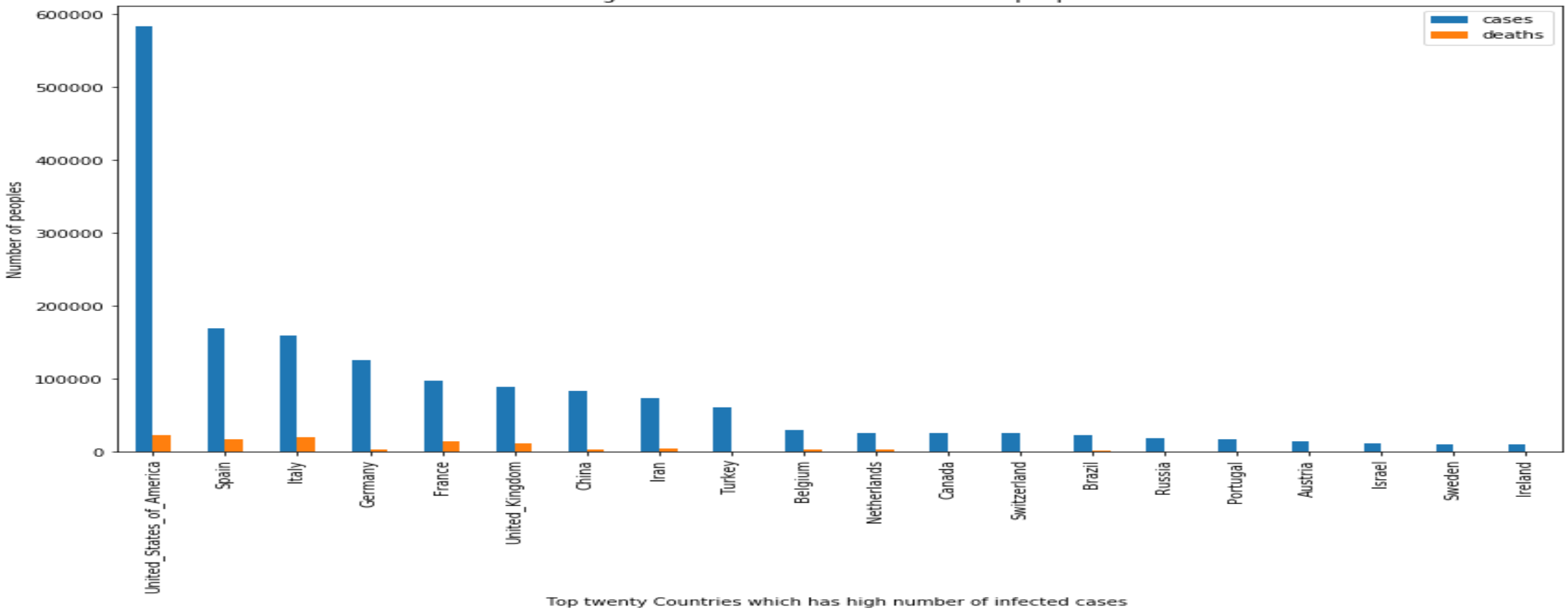
❖Fig 1.0, the number of infected individuals is rapidly increasing from January till present.



❖Fig 1.2 describes the distribution of infected and deaths on highly affected twenty countries.

- ✓ United states, Spain, Italy, Germany and France are the first, second, third, fourth and fifth severely affected respectively

Fig 1.2: Number of infected and dead people



❖Figure 1.3 summarizes the bar chart of tweeny countries which has a least infected individuals.

- ✓ Yemen, Papua_New_Guinea, Anguilla, Biritish_vergin_islands and Bonaire, Saint Eustatius and Saba are the list affected countries.

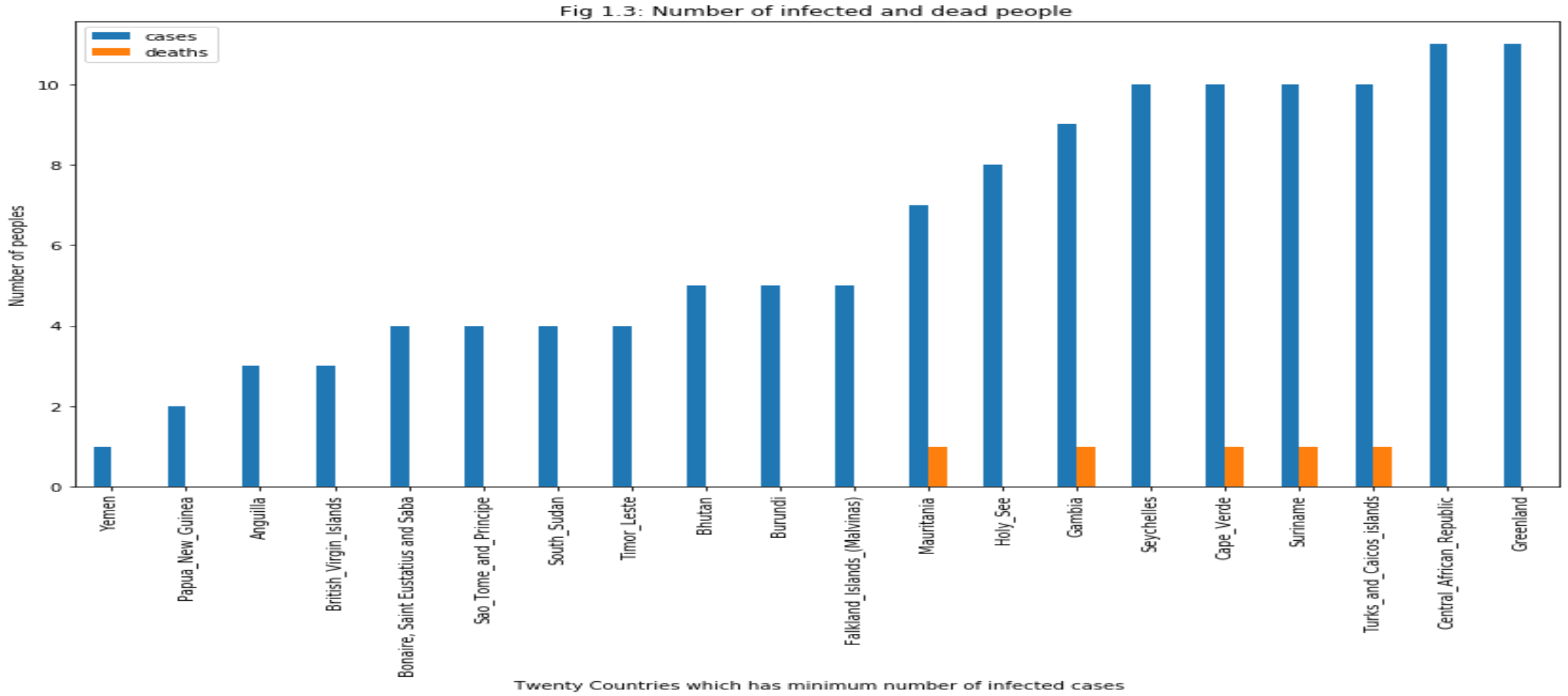


Table 2.1: Summary statistics of the Us COVID-19 data

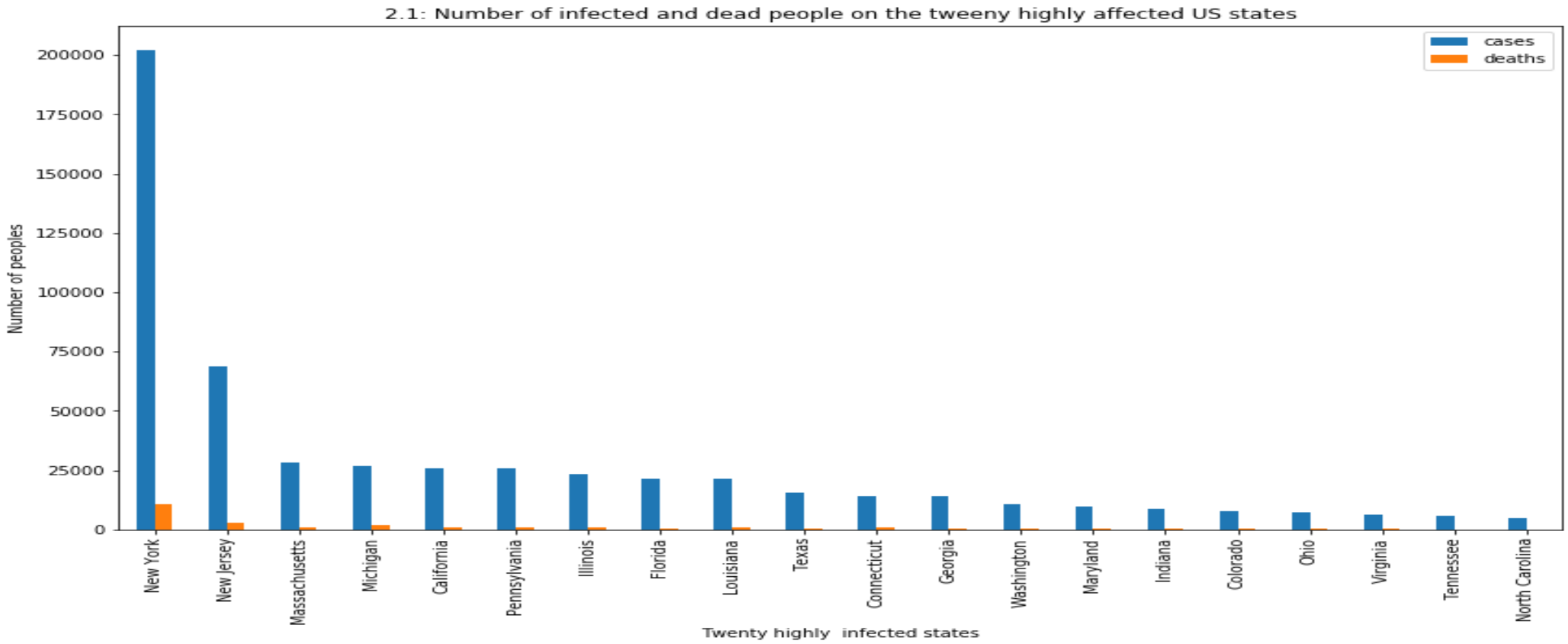
	New York City	New Jersey	Massachusetts	Michigan	California
Infected Cases	202208	68824	28163	26845	25758
Deaths	10834	2805	957	1767	778

❖From Table 2.1 above , we can summarize that :

✓Number of infected cases of the five highly affected states in US are: New York City, New-Jersey, Massachusetts, Michigan and California are the first , second, third, fourth and fifth ranked states respectively.

❖Figure 2.1 bar chart summarizes the first twenty highly affected states in US.

- ✓ It supported for the above Table 2.1



❖Statistical methodology

- ✓ In this study in addition some of the Exploratory data analysis some Statistical models will be used to analysis the data such as:

  Correlation analysis, simple or multiple linear regression