

# PM566 Final Project

Sherry Shen

## 1. Introduction

### 1.1 Background

When talking about 2020, "COVID-19", a novel coronavirus firstly found in Wuhan, China, in the end of 2019 (McIntosh, 2020), must be the most popular word that appears on the Internet, newspapers and daily conversations. Since COVID-19 has been discovered and claimed as human pandemic by Chinese government, there are more than 36 millions of people being confirmed with COVID-19 and about 1 million deaths by now (2020). COVID-19 is definitely raising a huge concern of public health around the world.

### 1.2 Problem Statement

There are many opinions about the origin of COVID-19. Some suggests that the sequence of COVID-19 is similar to bat virus (Shin, 2020), while others do not support that idea. There are also many unclear characteristics about this new virus. Therefore, this report will mainly focus on discovering main and critical characteristics of COVID-19, including incubation period, COVID-19 distribution of age, gender.

### 1.3 Data Description

This data is obtained from Kaggle: Novel Corona Virus 2019 Data set. These COVID-19 case data are collected from 20th January to 28th February, 2020, which is the early pandemic period of COVID-19. This data set consists of 10 qualitative attributes, 7 quantitative attributes and 3 other attributes (Table 1.1).