

# AIDM7330 Individual Assignment

## Exploratory Data Visualisation Of The HongKong Covid-19 Data Report

**Student ID:** 20448589

**Student Name:** Luo YiFeng

The data was from HK Government Open Data: <https://data.gov.hk/sc-data/dataset/hk-dh-chpsebcddr-novel-infectious-agent/resource/dae21c96-34b3-40aa-b111-aac24abde893>

The reason why I want to analyse the data of the coronavirus cases in Hong Kong is because in recently the number of the coronavirus cases in Hong Kong has entered a rising period that so called the fourth wave of coronavirus. Therefore, I want to analyse the past data by my own hand and to see the actual situation in HongKong.

I got the data from HK Government Open Data. After I downloaded the data, I immediately feel it is a bad dataset, and it really need to be cleaned first than I can get to analyse the data in it.

It is a 5702x10 dimensions dataset, the datatype in this data was basically all object originally, which means I need to convert them into the datatype I can use, I cleaned them by each column in the dataframe. The dataset dose provided a tons of information that I can use. The columns of this dataset are as below.

```
Index(['Case no.', 'Report date', 'Date of onset', 'Gender', 'Age',
      'Name of hospital admitted', 'Hospitalised/Discharged/Deceased',
      'HK/Non-HK resident', 'Case classification*', 'Confirmed/probable'],
      dtype='object')
```

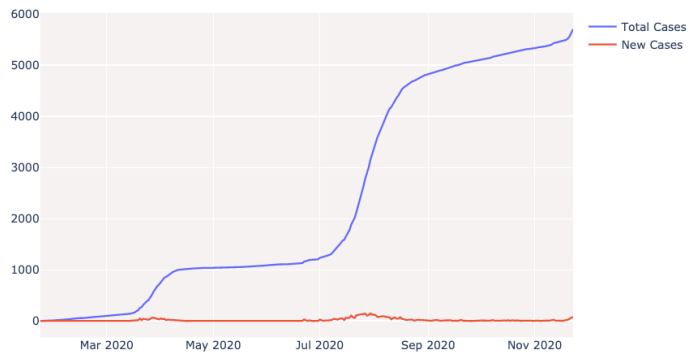
I can analyse the trend of the Covid-19 Cases in Hong Kong. And each the relationship between age and this virus. So I identify these questions below for me to answer:

- How the Covid-19 cases are rising in Hong Kong?
- What is the amount/percentage of Hospitalised/Discharged/Deceased?
- What is the Amount/Percentage of Gender ?
- What is the Amount/Percentage of HK/ Non-HK Resident ?
- What is the Amount/Percentage of Case classification ?
- Produce a Age group(the original dataset not have), See the Age Group Wise Distribution

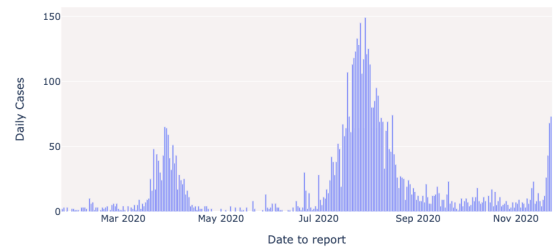
### How the Covid-19 cases are rising in Hong Kong?

Based on the plot, I found that the first two cases in Hong Kong was in Jan 23 2020, and it have several turning date, which are Mar 29 2020, Jul 9 2020 and Nov 17 2020. And May 27 2020 and Jul 30 2020 have the most daily cases reported, which are 65 cases and 149 cases.

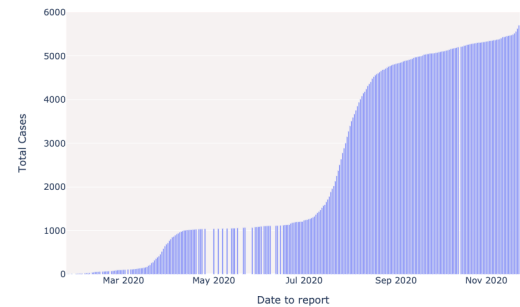
Trend of Coronavirus Cases in HongKong(China)



New Coronavirus Cases in HongKong(China) per day



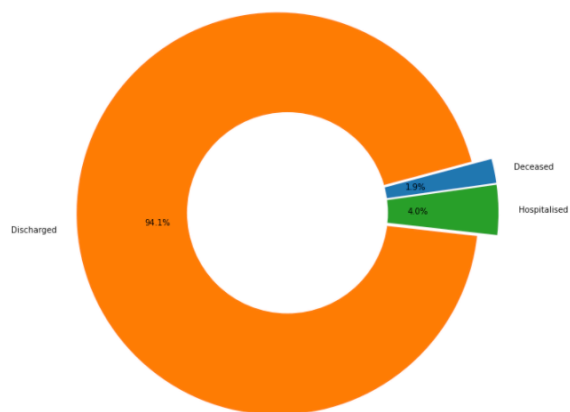
Total Coronavirus Cases in HongKong(China) per day



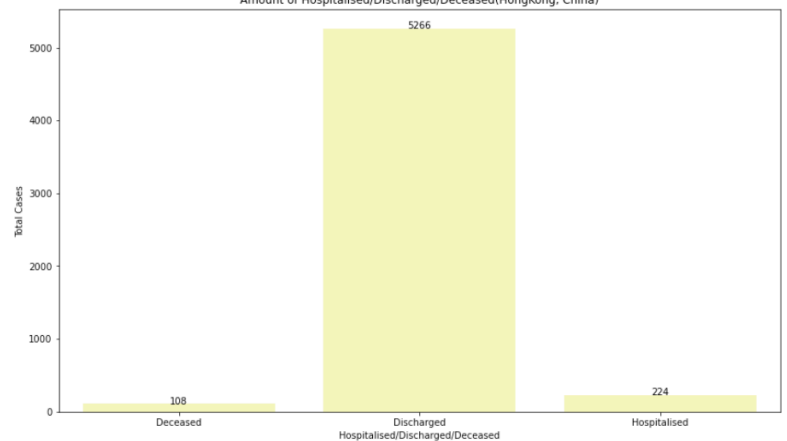
## What is the amount/percentage of Hospitalised/Discharged/Deceased?

By the time to Nov 23 2020, the total cases in Hong Kong was 5701, And There are 5266(94.1%) people was cured, but sadly there are 108(4.0%) people died, and 224(1.9%) people still in hospital.

HongKong(China) - The Percentage of Hospitalised/Discharged/Deceased

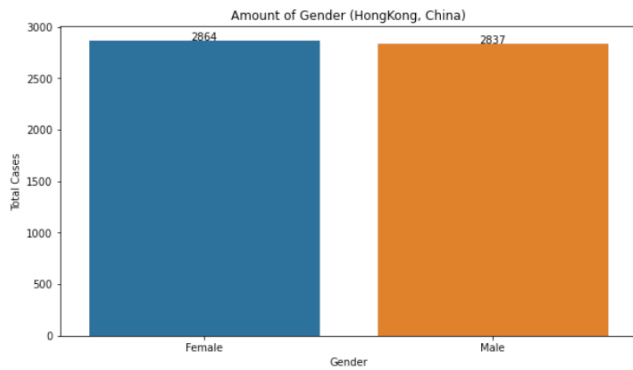


Amount of Hospitalised/Discharged/Deceased(HongKong, China)



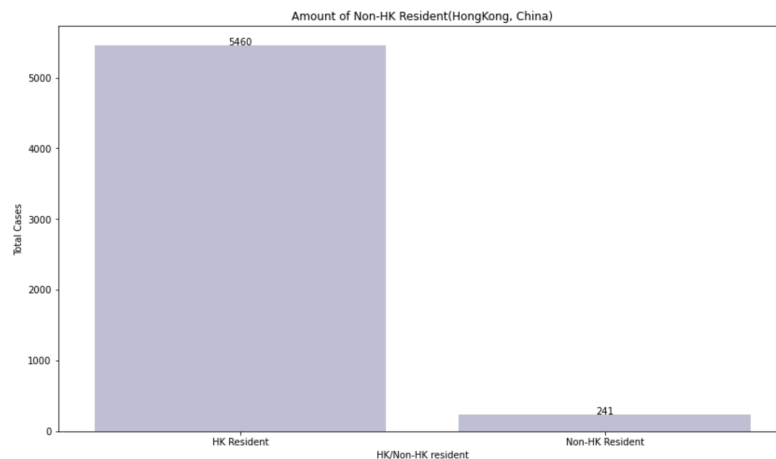
## What is the Amount/Percentage of Gender ?

There are 2864 Females and 2837 Males in the cases which is surprisingly almost half half.



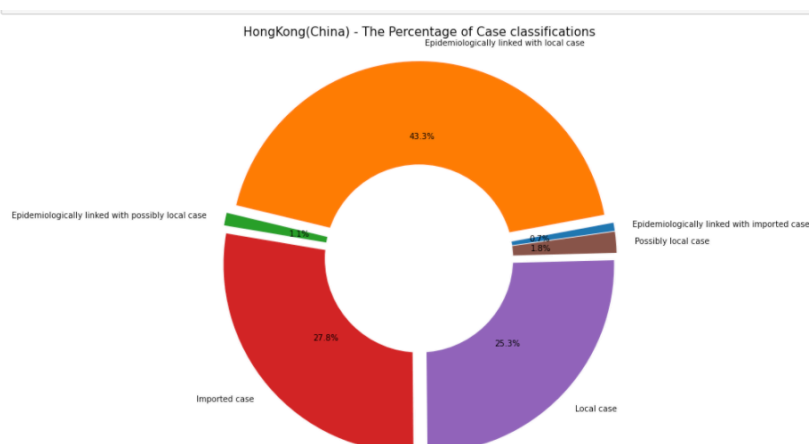
## What is the Amount/Percentage of HK/ Non-HK Resident ?

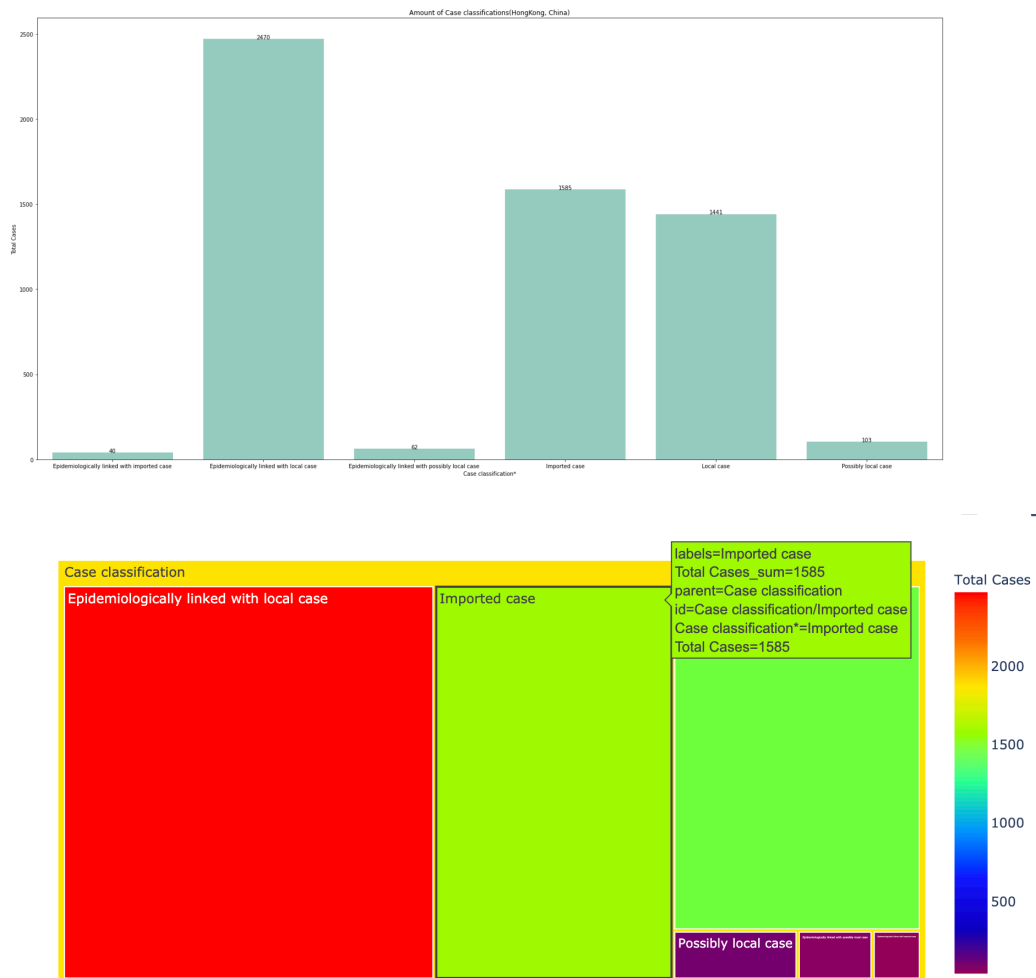
The HK/Non-HK Resident percentage:



## What is the Amount/Percentage of Case classification ?

The Percentage of Case Classification:





The biggest group is epidemiologically linked with local case and Imported case.

## Produce a Age group(the original dataset not have), See the Age Group Wise Distribution

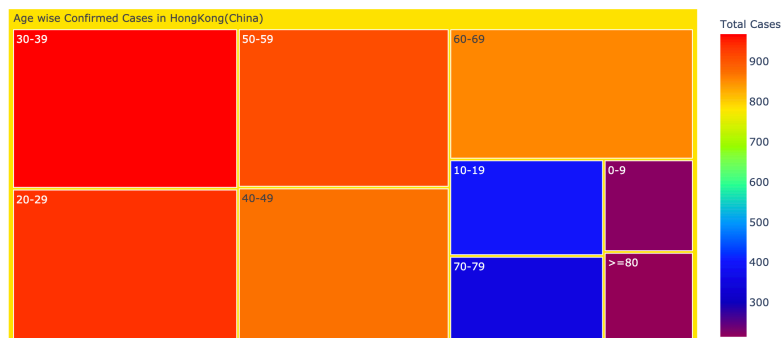
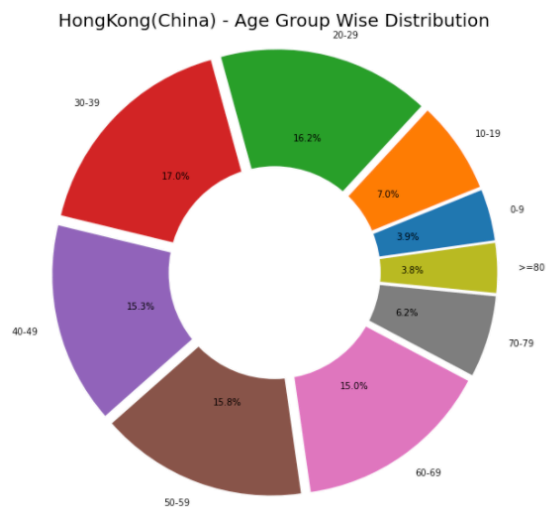
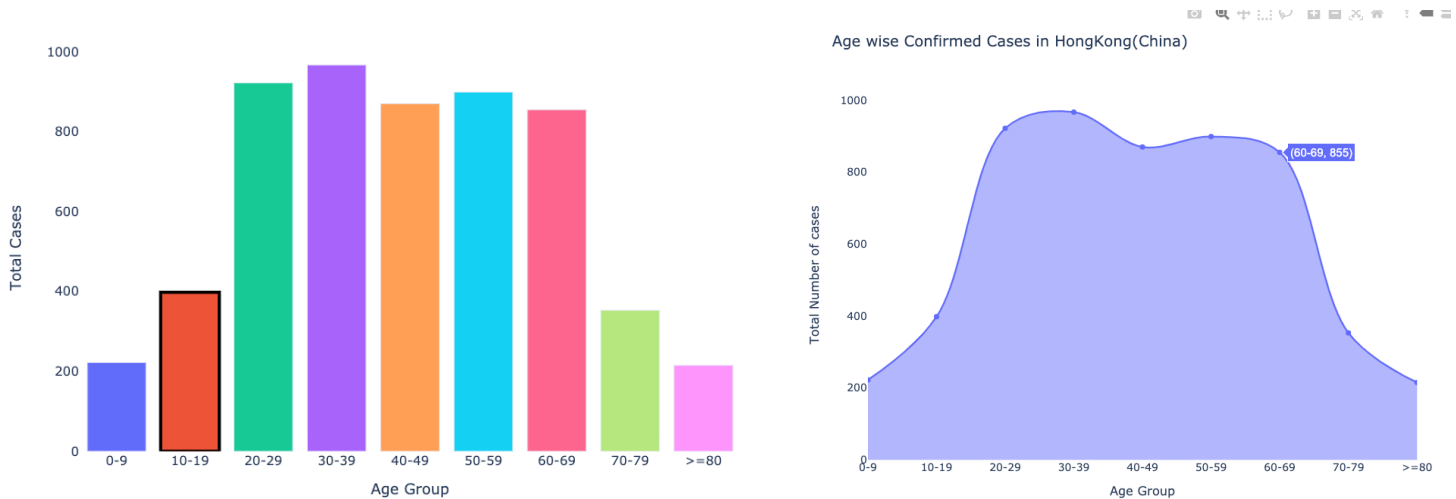
I used the pandas cut method to make a new column which is called Age Group, The DataFrame looks like this:

Age Group	Total Cases
0-9	222
10-19	398
20-29	922
30-39	967
40-49	870
50-59	899
60-69	855
70-79	353
>=80	215

The Age Group Wise Distribution plots are in below:

The plots shows people in the age between 20-29, 30-39, 50,59 have the most number cases.

The age between 30-39 have the most number of cases which is 967.



## What is the relationship between Gender and Age in these cases?

Pivot table

Age Group	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	>=80
Gender									
Female	99	184	423	520	472	477	412	158	119
Male	123	214	499	447	398	422	443	195	96

