Assignment 1 – Pandas and Matplotlib

This assignment requires more individual learning - you are encouraged to check out the panda's documentation to find functions or methods you might not have used yet, or ask questions on Stack Overflow and tag them as pandas and python related. And of course, the discussion forums are open for interaction with your peers and the course staff.

Dataset Description:

Coronavirus disease 2019 (COVID-19) is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The disease was first identified in 2019 in Wuhan, China, and has since spread globally, resulting in the 2019–20 coronavirus pandemic. Common symptoms include fever, cough and shortness of breath. Muscle pain, sputum production and sore throat are less common. The rate of deaths per number of diagnosed cases is on average 3.4%, ranging from 0.2% in those less than 20 to approximately 15% in those over 80 years old.

Data Source: 2019 Novel Coronavirus COVID-19 Kaggle Dataset repository

File naming convention

MM-DD-YYYY.csv in UTC.

Field description

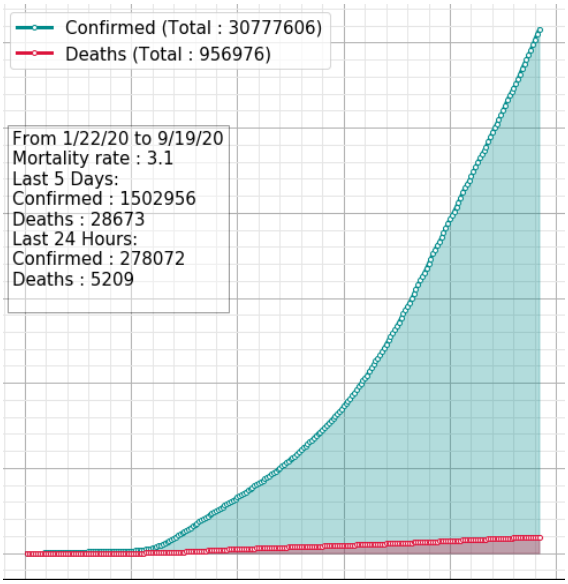
- Province/State: China - province name; US/Canada/Australia/ - city name, state/province name; Others - name of the event (e.g., "Diamond Princess" cruise ship); other countries - blank.
- Country/Region: country/region name conforming to WHO (will be updated).
- Last Update: MM/DD/YYYY HH:mm (24 hour format, in UTC).
- Confirmed: the number of confirmed cases.
- Deaths: the number of deaths.
- Recovered: the number of recovered cases.
- Active: Active cases on given date
- WHO Region: Continent

Part 1: General analysis of Data

- Check for missing values and simply count the sum.
- Detect outliers and simply remove it.
- Group the data into country wise and continent wise and save into useful data frames.
- Show the total number of confirmed cases, deaths reported, recovered and active cases around the globe and mortality rate[1].
- Show continent wise confirmed cases, recovered cases, deaths and active cases and mortality rate.
- Show country wise confirmed cases, recovered cases, deaths and active cases and mortality rate.
- Show top 10 countries w.r.t confirmed cases and deaths.

Part 2: Visualization

- Create a graph shows the confirmed, deaths, recovered and active cases trends over the world.



- Covid-19 spread trends w.r.t different continents i.e. Asia, Europe, America etc (same as above)
- Pandemic spread trend in China in each month. - Bar chart
- Pandemic spread comparison of different continents-Line chart
- Pandemic confirmed and deaths percent split in different countries – Pie chart
- Mortality rate of each continent – Scatter plot
- Confirmed cases in china and outside china in first 50 days- Bar chart and line chart
- Total tests done in Pakistan vs China till Aug 2020 – Bar chart

[1] Mortality rate = (deaths/confirmed) *100