

COVID-19 Outbreak Case Study

ROHIT BHINGARE

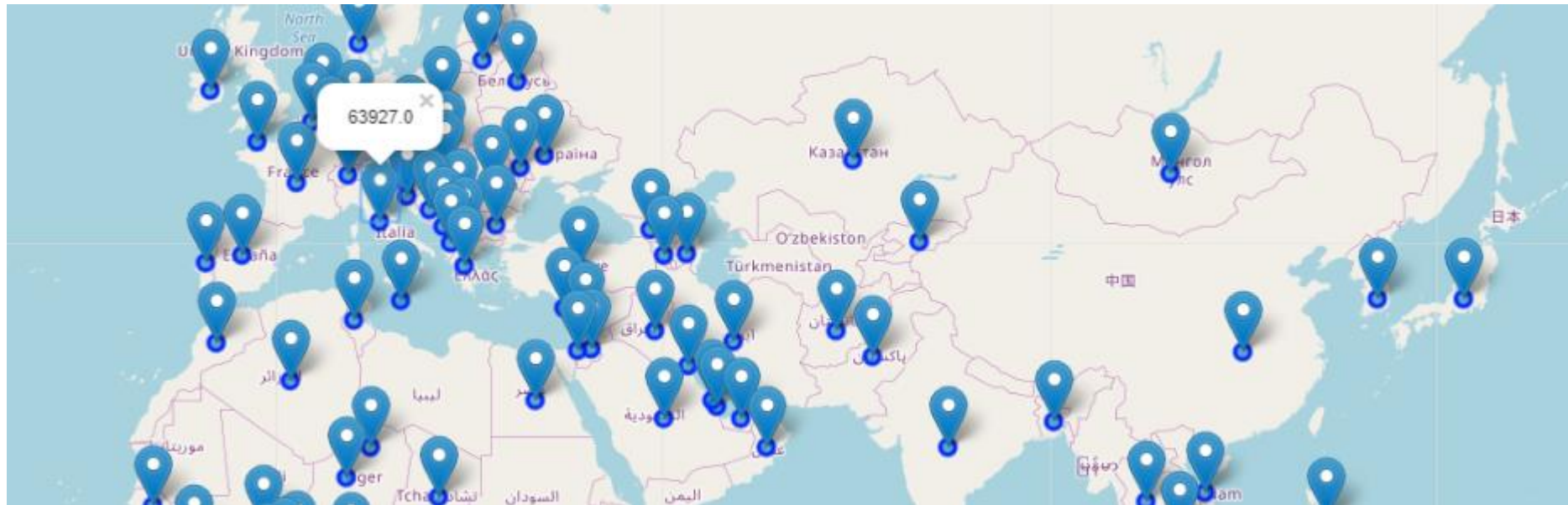
Background of disease

- Coronavirus disease (COVID-19) is an infectious disease caused by a new virus
- Causes respiratory illness (like the flu) with symptoms such as a cough and fever
- Can create difficulty in breathing in more severe cases
- 2019-nCoV (new variant) was first identified in Wuhan, China
- Existing vaccines or treatments are currently not effective
- Virus has shown evidence of human-to-human transmission

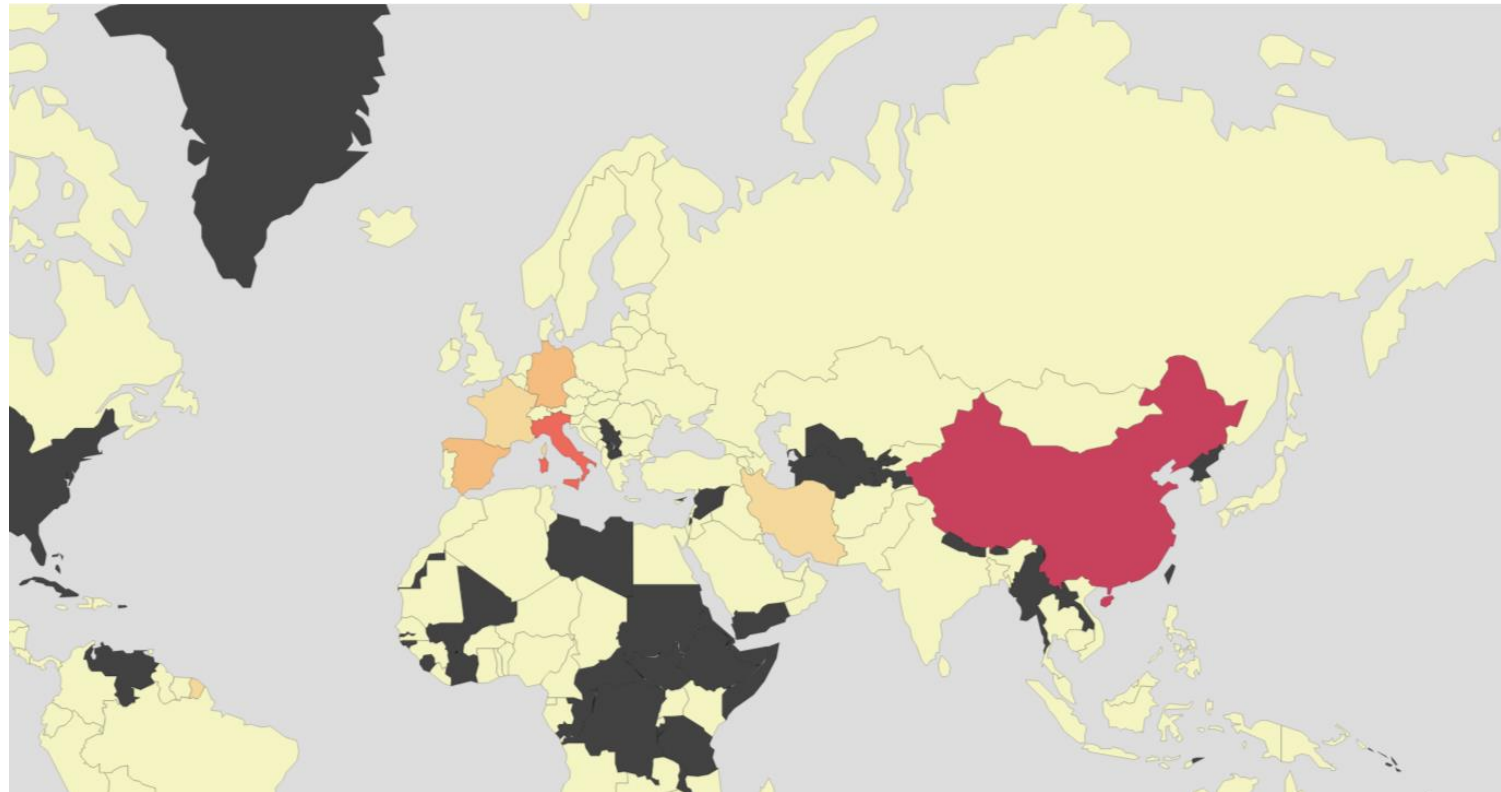
Data acquisition and cleaning

- COVID-19 dataset taken from Kaggle dataset on 22.March.2020
- Snapshot of live statistics taken from <https://www.worldometers.info/coronavirus/>
- Good country index taken from Wikipedia https://en.wikipedia.org/wiki/Good_Country_Index
- Population taken from Wikipedia
[https://en.wikipedia.org/wiki/List_of_countries_by_population_\(United_Nations\)](https://en.wikipedia.org/wiki/List_of_countries_by_population_(United_Nations))
- Nominal GDP taken from Wikipedia
[https://en.wikipedia.org/wiki/List_of_countries_by_GDP_\(nominal\)](https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(nominal))

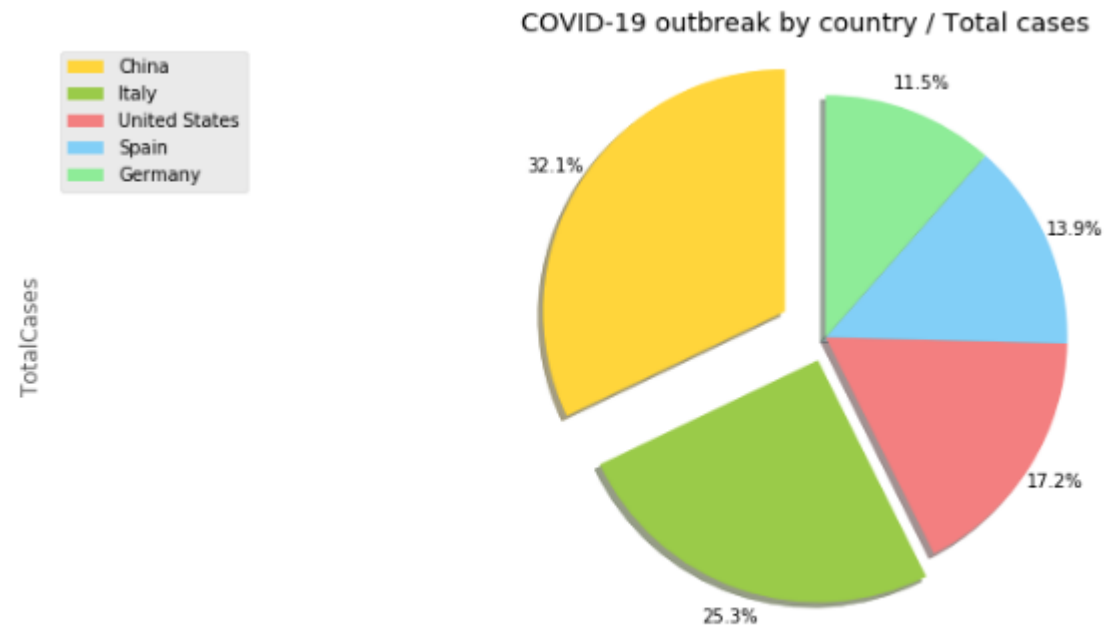
Use world-map to see a country level overview



Use a Choropleth map to see the intensity of outbreak

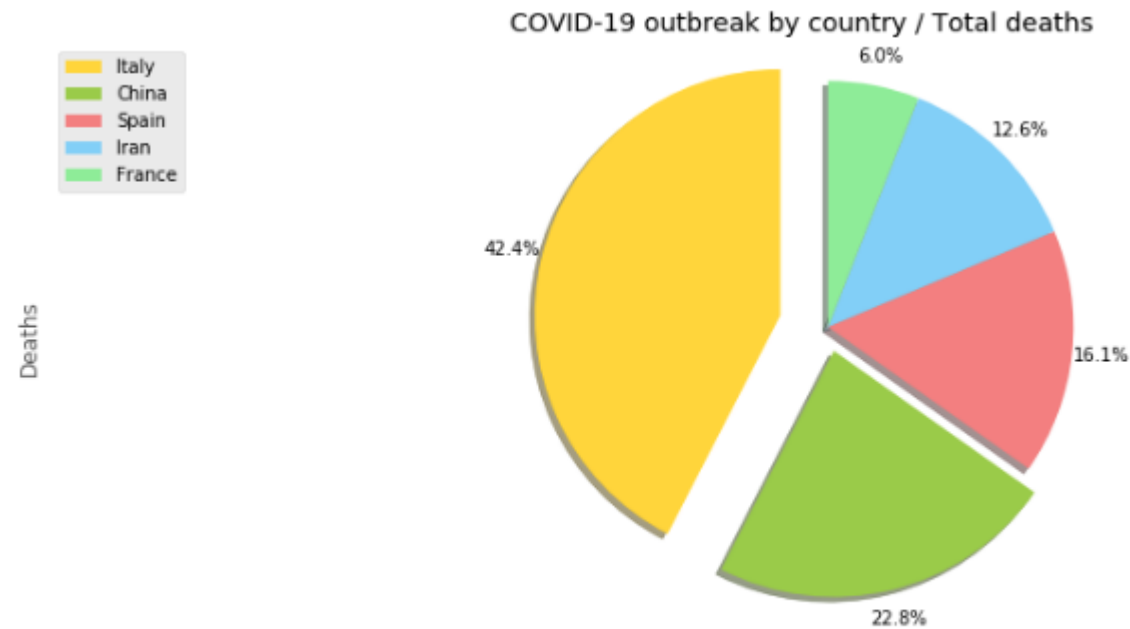


Highest spread countries



We see that China and Italy have the highest spread, followed by USA, Spain and Germany

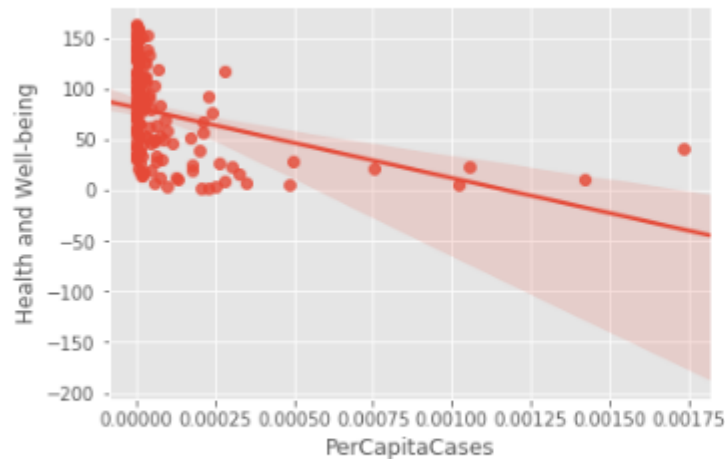
Countries with highest deaths



We see that Italy has the highest reported deaths so far, followed by China, Spain, Iran and France

Economic and Well being index of countries against the outbreak

```
ax = sns.regplot(x='PerCapitaCases', y='Health and Well-being', data=dfcovid19)
```



Observation 2 - Again expected, but we see a inverse linear relationship between economic health and well-being index and per capita cases reported.

Science and Technology index of countries against the outbreak

```
ax = sns.regplot(x='PerCapitaCases', y='Science and Technology', data=dfcovid19)
```



Observation 3 - Similar to 2, we see a inverse linear relationship between Science and Technology index and per capita cases reported.

Conclusion

- China has the highest outbreak (and recoveries per outbreak)
- Italy has the highest deaths (total and per capita)
- Major outbreaks of COVID-19 have occurred in China and Europe
- African nations have the lowest recovery rate (total and per capita) from the outbreak
- Direct correlation between population of country to the total cases that are reported
- Inverse (but weak) correlation trend between the Science and Technology to outbreak per capita
- Inverse (but weak) correlation trend between the Health and Well-being to outbreak per capita
- Due to weak correlations, it is complex to fit when a machine learning model was created

Further extension scope

- Explore the theme of "rate of recovery" across the months for the outspread per country against the good country KPIs.
- Assess the socioeconomic impact of COVID-19 outbreak to financial indices, situation in hospitals as well as job market
- Assess the impact of COVID-19 outbreak to the mortality/illness rates from previous years
- Drill-down on individuals to address relative deaths due to other/known illnesses