Exploration and modeling of COVID-19 Data

Part 1: description of proposed datasets

Dataset chosen-

1. Johns Hopkins covid-19 dataset

Source: <https://github.com/CSSEGISandData/COVID-19>

Variables:

|  |  |
| --- | --- |
| Province\_State | String |
| Country\_Region | String |
| Last\_Update | ###### |
| Lat | Num |
| Long\_ | Num |
| Confirmed | Num |
| Deaths | Num |
| Recovered | Num |
| Active | Num |
| FIPS | Num |
| Incident\_Rate | Num |
| People\_Tested | Num |
| People\_Hospitalized | Num |
| Mortality\_Rate | Num |
| UID | Num |
| ISO3 | String |
| Testing\_Rate | Num |
| Hospitalization\_Rate | Num |

* This dataset provides extensive information about confirmed, deaths and active cases and many other additional information. All the information is broken down to countries and states and provinces as well. I have highlighted the variables in red that I will be using for my data exploration and modeling.
* I will be using data about China and US for modeling

1. Acaps\_covid19\_government\_measures\_dataset

Source: <https://www.acaps.org/covid19-government-measures-dataset>

This dataset includes governments’ measures worldwide in response to coronavirus. These researched information falls into five categories: social distancing, movement restrictions, public health measures, social and economic measures and lockdowns. This dataset contains rich information of countries, and regions and dates of when government measures were taken. I’d be using information about china and united states from this dataset.

Variables:

|  |  |
| --- | --- |
| ID | Numeric |
| COUNTRY | Text |
| ISO | Text |
| ADMIN\_LEVEL | Text |
| PCODE | Text |
| REGION | Text |
| LOG\_TYPE | Text |
| CATEGORY | Text |
| MEASURE\_TYPE | Text |
| TARGETED\_POP\_GROUP | Text |
| COMMENTS | Text |
| NON\_COMPLIANCE | Text |
| DATE\_IMPLEMENTED | Date (DD/MM/YYYY) |
| SOURCE | Text |
| SOURCE\_TYPE | Text |
| LINK | Text |
| ENTRY\_DATE | Date (DD/MM/YYYY) |
| ALTERNATIVE SOURCE | Text |