Stage I - Data and Project Understanding

Task 1:

* Readme file

Task 2:

* Variable dictionary for COVID-19 data
* According to Covid-19 dataset [such as case and deaths] representation which is in column form.
* We applied the melt function to convert “column form” to “row form” due to the easy joining function as compared to column join operation.

**Number of Cases**

| Name | Definition | Data Type | Possible values | Required? |
| --- | --- | --- | --- | --- |
| countyFIPS | County code | int64 | 0,1001, 56045 | yes |
| County Name | County name | object | Autauga County,  Bibb County | yes |
| State | State name | object | California, North Carolina | yes |
| StateFIPS | State code | int64 | 1, 2, 3, 56 | yes |
| Date | Date of recorded cases | object | 2020-01-22,... | yes |
| Cases | Number of cases on specific date | int64 | 0,1,2,3, ... | yes |

**Number of Deaths**

| Name | Definition | Data Type | Possible values | Required? |
| --- | --- | --- | --- | --- |
| countyFIPS | County code | int64 | 0,1001, 56045 | yes |
| County Name | County name | object | Autauga County,  Bibb County | yes |
| State | State name | object | California, North Carolina | yes |
| StateFIPS | State code | int64 | 1, 2, 3, 56 | yes |
| Date | Date of recorded deaths | object | 2020-01-22,... | yes |
| Death | Number of deaths on specific date | int64 | 0, 1, 2, 3,... | yes |

**Name of county**

| Name | Definition | Data Type | Possible values | Required? |
| --- | --- | --- | --- | --- |
| countyFIPS | County Code | int64 | 0,1001, 56045 | yes |
| State | State Name | object | California, North Carolina | yes |
| population | Number of people | int64 | 22394, 12589 | yes |

* Preliminary intuitions on data:
  + Cases:
    - This data set shows the number of Covid-19 cases in each state and county on a specific date. This is helpful in showing the behavior of Covid-19 in terms of geographical location and rate of spreading.
  + Deaths:
    - This data set shows the number of deaths caused by Covid-19 in each state and county on a specific date. This data can be helpful to see how lethal Covid-19 is and to see which geographical locations and conditions impact the mortality rate of covid. When used with the data set of cases, we can also see how quickly people die from the virus.
  + Counties:
    - This data set shows important information about a county such as the state it is in and its population. When used with the other data sets, we can see how dangerous Covid-19 is in proportion to population size and density over time.

Task 3:

* Team notebook