

Covid Data Analysis Project – Stage I

Task – 2 (Team)

TEAM MEMBERS:

1. Ankitha Chavan
2. Ashrita Dasari
3. Deepika Dasaroju
4. Lakshmi Manogna Bobbili
5. Venkatesh Pulibandla

COVID DATA SETS

1. COVID County Population Dataset

Name	Definition	Data type	Possible Values	Required?
County FIPS	Specific code for the county	Integer	1001,1003,1005 etc.	Yes
County Name	Name of the County	String	Barbour County, Baldwin County etc.	Yes
State	Name of the State	String	AL,AK etc.	Yes
Population	Number of people.	Integer	6893,1183 etc.	Yes

2. COVID County Confirmed Cases Dataset:

Name	Definition	Data Type	Possible values	Required?
countyFIPS	Code for that particular county	Integer	starts from 1000 for all odd value	Yes
County Name	Name of the county	String	Can be of any name depending on the county (Multiple names possible)	Yes
State	Name of the state	String	Can be of any name depending on the state	Yes
StateFIPS	Code for that particular state	Integer	starts from 1 alphabetically for all the states	Yes
Date	No. of confirmed cases on that date	Integer	from 0 to a max of the number of people in that county	Yes

3. COVID County Confirmed Death cases Dataset:

Name	Definition	Data Type	Possible Values	Required?
CountyFIPS	Code for that particular county	Integer	starts from 1000 for all odd value	Yes
CountyName	Name of the county	String	Can be of any name depending on the county (Multiple names possible)	Yes
State	Name of the state	String	Can be of any name depending on the state	Yes
StateFIPS	Code for that particular state	Integer	starts from 1 alphabetically for all the states	Yes
Date	No. of confirmed death on the particular data	Integer	from 0 to a max of the number of people in that county	Yes

From these data sets we can see the COVID daily patterns along the given timeframe. We can analyze and find the factors impacting the growth and spread of COVID. Some basic factors are population, ratio of COVID cases and the population, Ratio of COVID Death to population, Ratio of COVID Deaths to the Cases.

When joined with enrichment data, we can see how COVID-19 has affected those fields. For example, if we consider hospitals, as COVID patients started flooding into hospitals, there was a shortage in hospital facilities. Another example, since the rise of COVID-19, there have been multiple lockdowns, and the employment rate started to fluctuate drastically.

Like the above examples, we can analyze different factors that have been affected by COVID-19.

Initial Hypothesis:

- Regions with more population could be more affected by COVID and has more cases and deaths since the spread will be easier due to more number of people roaming closely in the area
- Regions with more socio gathering events (Like Florida) could have more number of cases
- There could have been raise in COVID cases after special events like Thanks Giving, New Year etc.
- With release of proper medication, social distancing and vaccination, COVID cases should be declined to a minimal value in the recent times.

REFERENCES:

1. <https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/>
2. <https://data.census.gov/cedsci/table?q=dp&tid=ACSDP1Y2018.DP05>
3. <https://data.census.gov/cedsci/table?q=dp&tid=ACSDP1Y2018.DP05>
4. <https://www.bls.gov/cew/downloadable-data-files.htm>
5. <https://www.kaggle.com/unanimad/us-election-2020>
6. <https://protect-public.hhs.gov/pages/hospital-utilization>
7. <https://analystanswers.com/what-is-a-data-dictionary-a-simple-thorough-overview/>
8. <https://bulldogjob.com/news/449-how-to-write-a-good-readme-for-your-github-project>