COVID-19 and Socioeconomic Status and Comorbidities Across the United States

Project 1: Exploratory Data Analysis

Bilal Ismail
Daiana Spataru
Mohammad Mohaiminul Islam

2023-07-27

UTOR-VIRT-DATA-PT-06-20
23-U-LOLC-MWTH(B)

Executive Summary

Industry: Public Health

Project Goal: Explore the relationship of socioeconomic status and underlying health conditions on COVID-19 metrics in the USA.

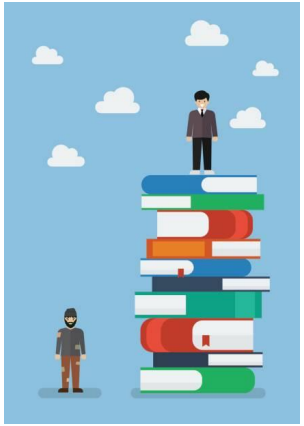
Socioeconomic Status: refers to the class of an individual or group and can be categorized based on a combination of education, income, and occupation.

Comorbidities: the presence of another illness which may or may not influence further health complications. For example, a person is more likely to get heart disease if they have a co-morbid condition, such as diabetes or hypertension.

Q1: Do areas of lower income tend to have higher case fatality rates than that of higher income areas?

Q2: Do states with a lower average education ranking have a higher case fatality rate?

Q3: What is the comorbidity that leads to the highest case fatality rate when coupled with COVID-19 infection?

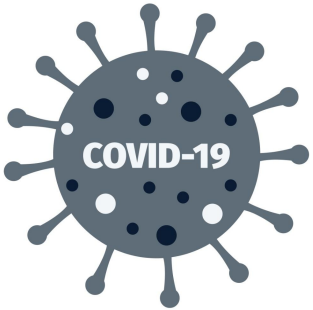


Data Collection, Cleanup, and Exploration

Personal Income
Poverty Rate
(access through APIs)

School Ranking
(accessed as CSV)

Underlying health conditions
(accessed as CSV)



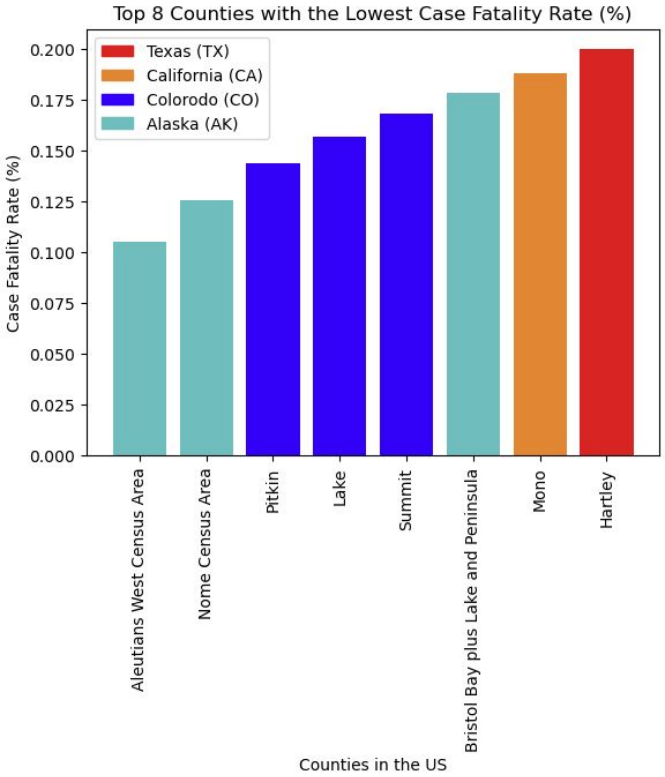
COVID-19 cases and deaths
(accessed as CSV)

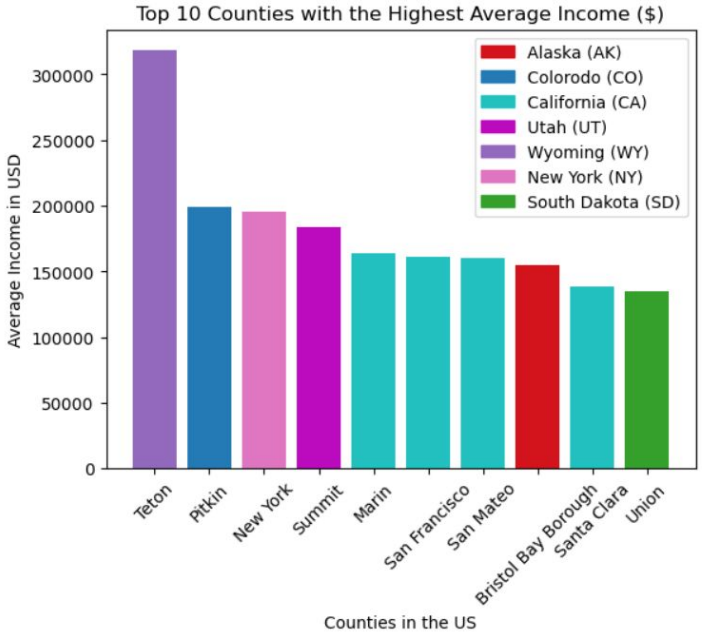
Q1: What relationship exists between income rates across county and state with case fatality rates, if any?

Datasets

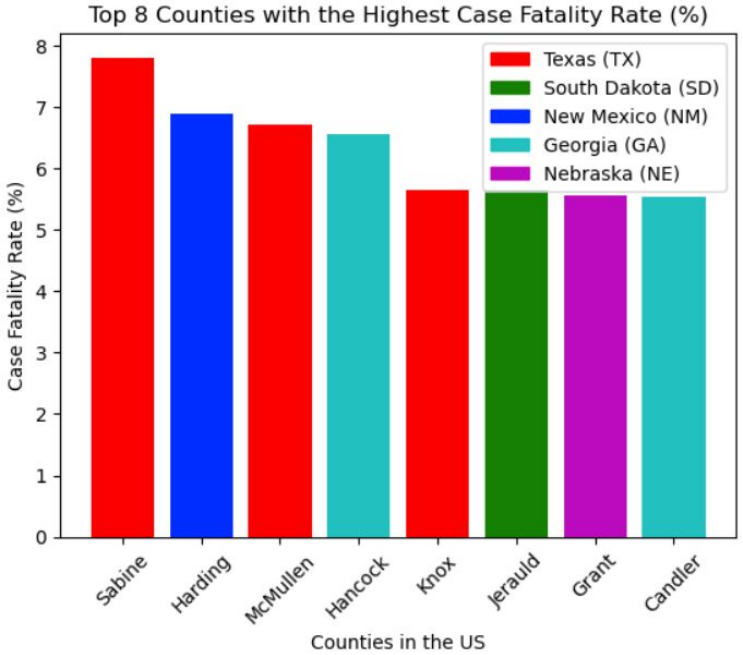
Dataset Name	Link	Used for
Small Area Income and Poverty Estimates: Small Area Income and Poverty Estimates: State and County	https://api.census.gov/data/timeseries/poverty/saipe.html	Poverty vs CFR
Detailed income and employment statistics by state, county, and metropolitan area	https://apps.bea.gov/API/signup/	Average Income vs CFR
Cumulative daily COVID-19 cases and deaths for each County and State in US	https://github.com/nytimes/covid-19-data	Average Income vs CFR Poverty vs CFR

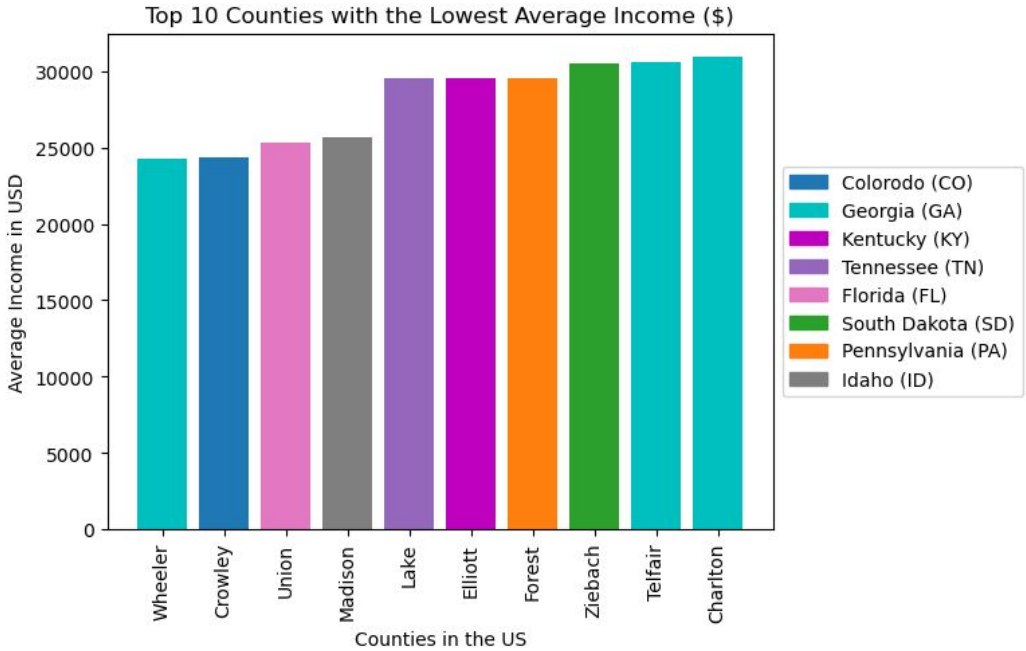
Average Income (USD) and Case Fatality Rates (%) of US Counties



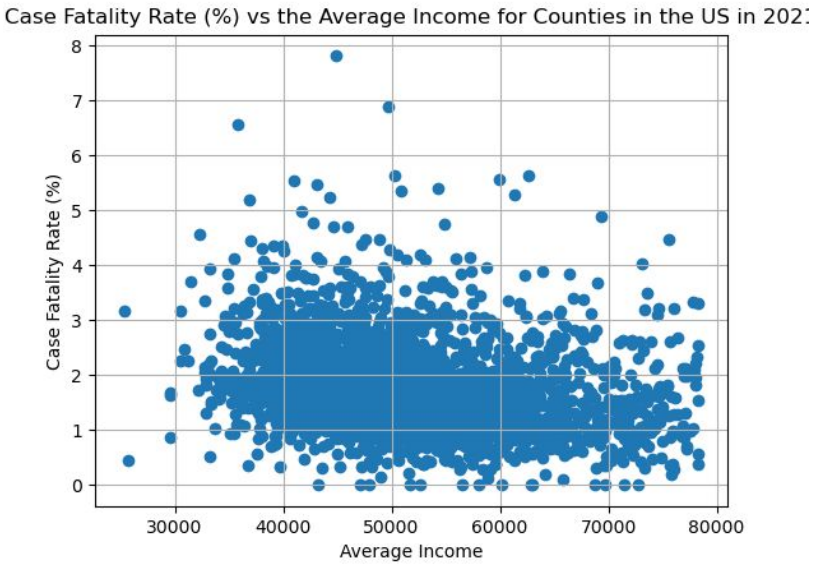


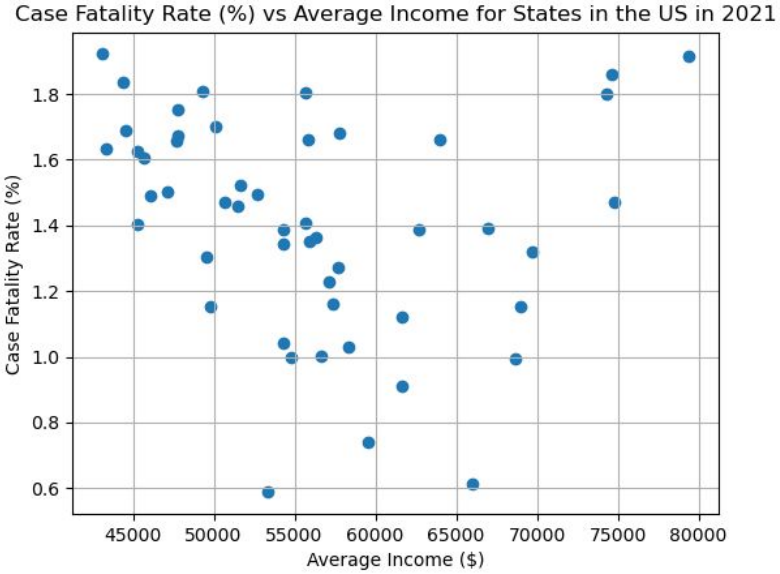
Average Income (USD) and Case Fatality Rates (%) of US Counties





Relationship between County and State Average Income (USD) and CFR (%)



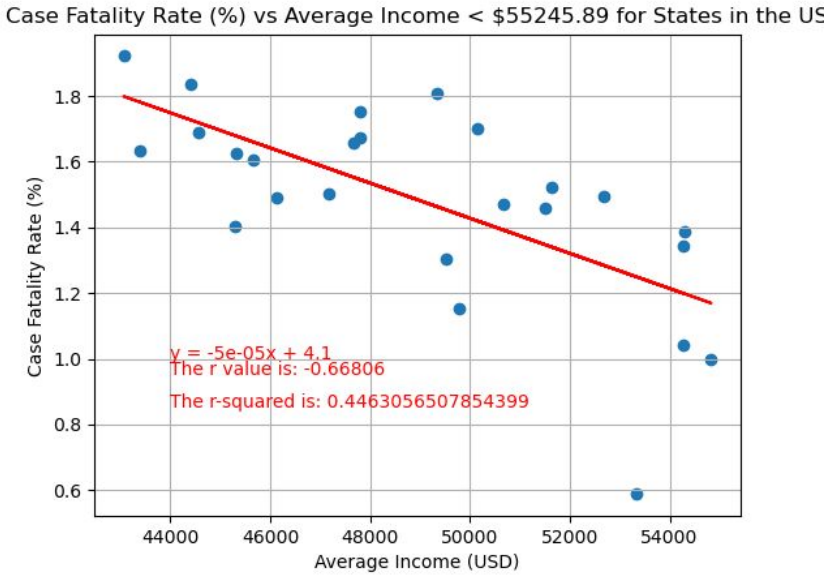


Correlation value = -0.25

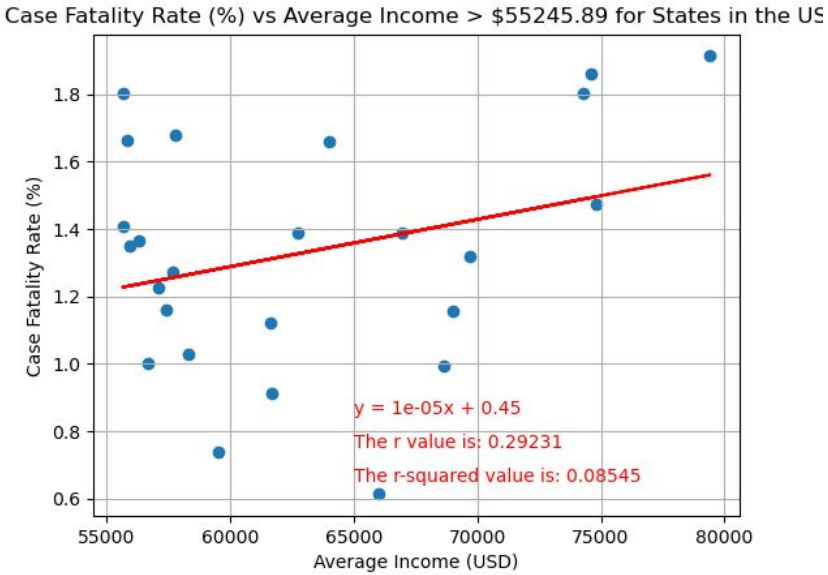
Correlation value = -0.18.

The state median income is: USD$55,245.89
T-test result: The pvalue is: 2.2347e-11 < 0.05

T-test for State Average Income and CFR

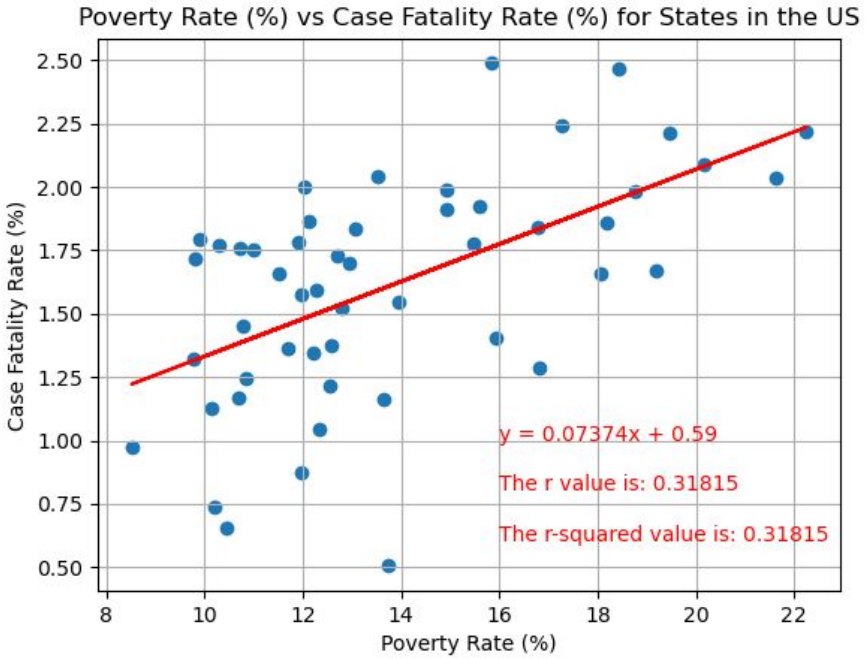


There is a moderate negative association with poverty rate and CFR.



There is a weak positive association with average income higher than the median income in the US and CFR, however, because of the low R^2 value, the distribution of this relationship around the linear regression may not be significant.

Relationship between State Poverty Rate (%) and State CFR (%)



Summary of Analysis & Conclusion

- Two metrics were analyzed against the case fatality rate of COVID-19 in the US: (1) Average Household Income, and (2) Poverty Rate. The following are observations from the data analysis:
  - Average household income in the states seems to play a larger role on COVID-19 CFR when areas are below the median household income in the US based on the dataset that was analyzed.
  - The CFR for States in the US with average household incomes greater than the median household income did not show a significant relationship
  - The general trendline for poverty rate vs the CFR for the States in the US shows a weak positive correlation
-

Limitations and Next Steps

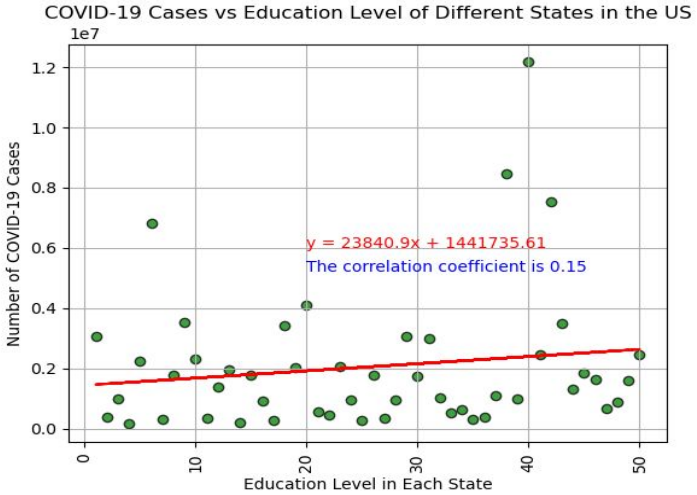
- Limitations:
  - Datasets were analyzed mainly at the State level: not granular enough to capture the social gradient between the upper and lower classes in each State
  - With respect to the median income and poverty rates against COVID-19, the data only looks at a snapshot in time, and does not consider the changing nature of reactions to the pandemic (non-pharmaceutical interventions)
  - The only metric used to analyze the impact of COVID-19 on different areas was the case fatality rate. Another metric to consider would be hospitalization rates as they may tell more about the impact of the virus in certain areas.
  - The COVID-19 case fatality rate does not represent the full impact of COVID-19. Deaths related to other health conditions are not considered in the analysis and may tell a different story.
- Next Steps:
  - Analyze the same information by zip code and see if the results are more what the team expected
  - Look more at population density distributions against case fatality rate
  - Hospital ranking
  - Look at the relationship between average quality of healthcare and population density against COVID-19 metrics

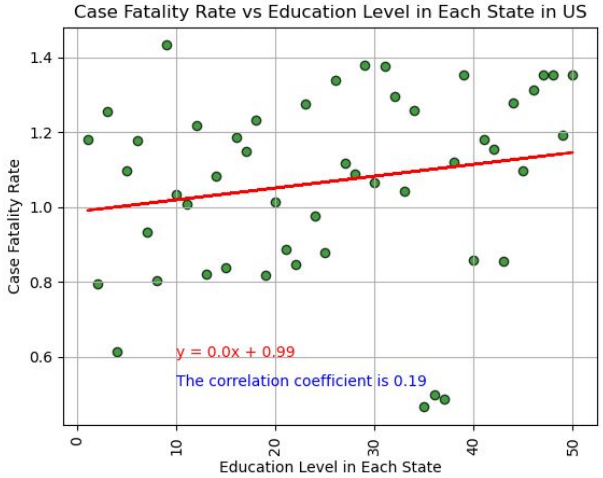
Q2: Do states with a lower average education ranking have a higher case fatality rate?

Datasets Used

1. Using the dataset of us-states (https://github.com/nytimes/covid-19-data) because it has the exact number of COVID-19 case in each states of US, that makes our study easy to ready and pull whatever info we need for our study.
2. Using the dataset from WalletHub (https://wallethub.com/edu/e/states-with-the-best-schools/5335) to get the best and worst education level in each state of US, that helps us a lot in our study since we are trying to see the relation between COVID and education level.

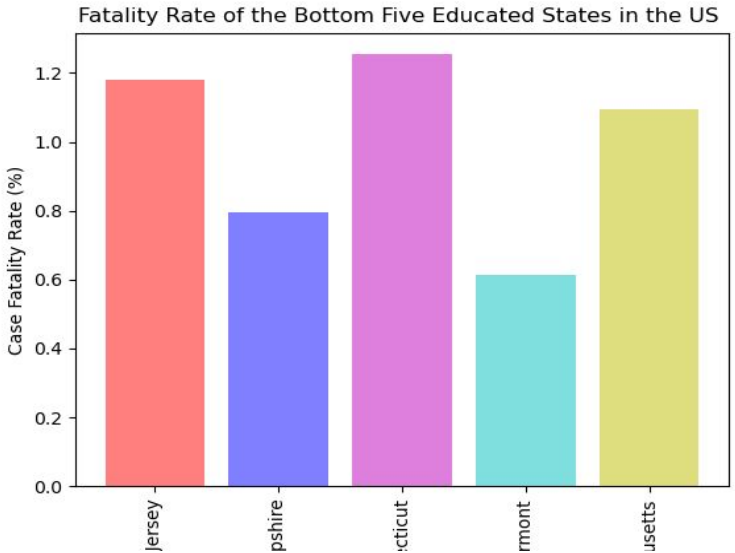
The relationship between education level and COVID-19 cases and case fatality rate in the United States

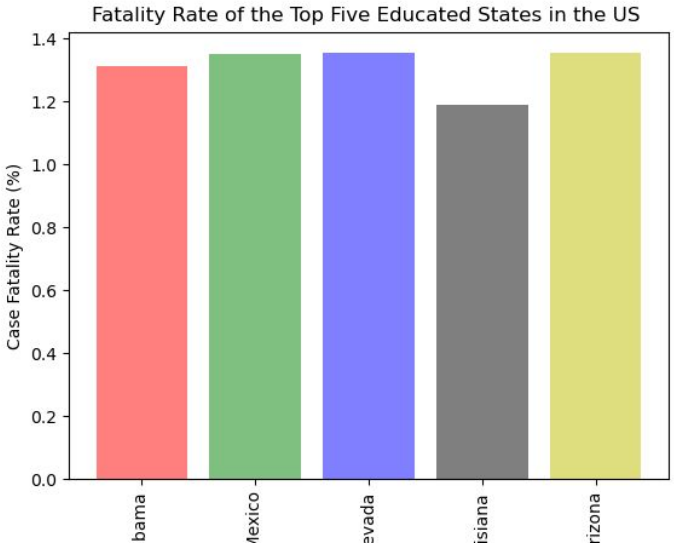




Case Fatality Rate of the Top and Bottom Ranked States in Terms of Their Education Rank

- To support our study we create 2 bar plots.
- The goal of these to bar is to see if the lowest ranked educated state have the highest case fatality rate and vice versa.





Conclusion & Next Steps

- This study showed us a weak relationship between education rank and COVID-19 case fatality rates in the different States in the US.
- Based on the bar graph, the top five educated states actually had higher case fatality rates than the lowest educated states
  - It would be interesting to look at the data for population of each of the schools in the states at a more granular level

Q3: What is the comorbidity that leads to the highest case fatality rate when coupled with COVID-19 infection?

Datasets and Cleaning

Datasets used:

- Covid-19 Data(Includes current Active cases, Deaths,Recovery)
- Covid-19 Vaccination data (Vaccine numbers, Fully Vaccinated)
- Covid-19 deaths with underlying diseases (What underlying diseases cost most deaths In Covid-19)

Questions for First Data set

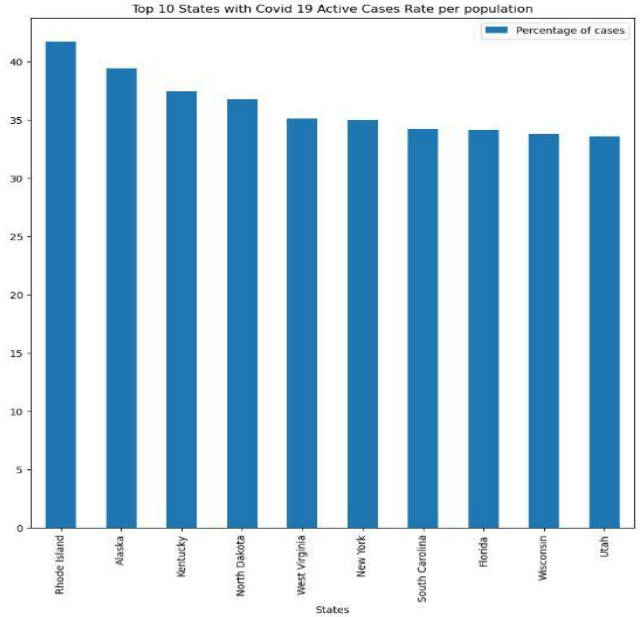
Q1. Which states are more affected by active cases,Covid Deaths?

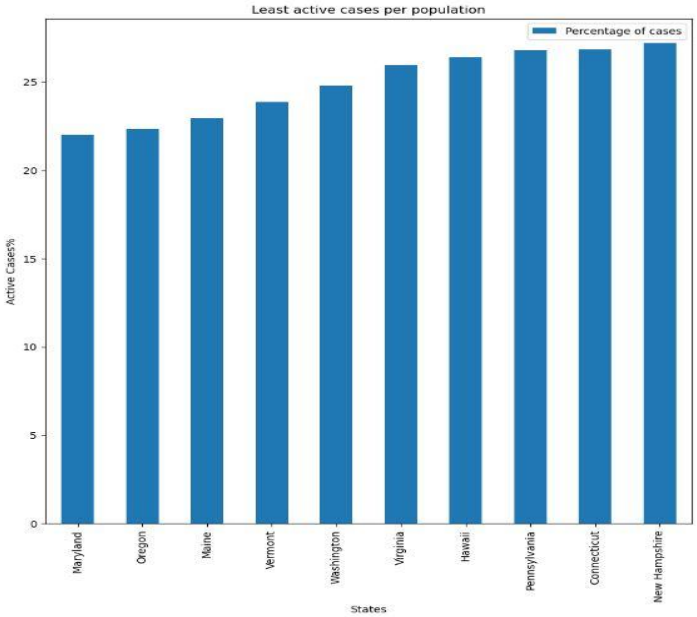
Q2. Which State are least affected by active cases,Covid Deaths?

Q3. Which States are most covid recovery states?

Active Case percentage per States(Most and Least)

Most affected states with active cases are:

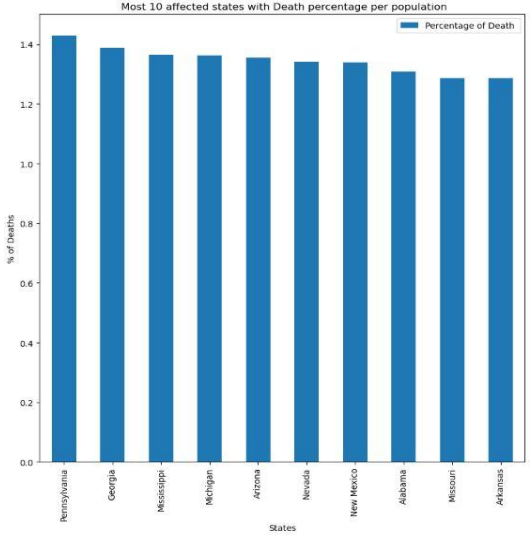


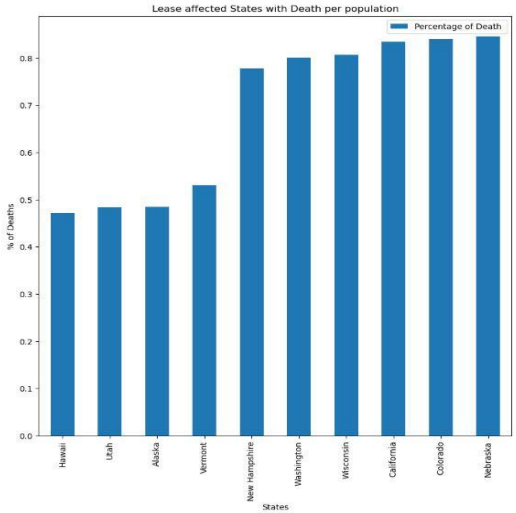


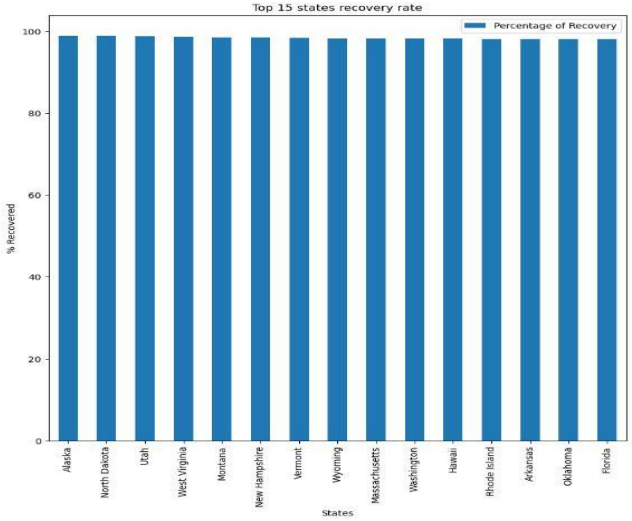
The most percentage of active cases are: Rhode Island, Alaska, Kentucky, North Dakota and New Virginia
The least percentage of active cases are: Maryland, Oregon, Maine, Vermont, Washington

Most and least affected Covid-19 Death percentage per Stases

Most affected, least affected and most recovery rates per states







Most death rates Top 5:Pennsylvania, Georgia, Mississippi,Michigan,Arizona
Least Death rates 5 states: Hawaii, Utah, Alaska, Vermont, New hampshire
Recovery Rate 5 states: Alaska, North Dakota, Utah, West Virginia, Montana

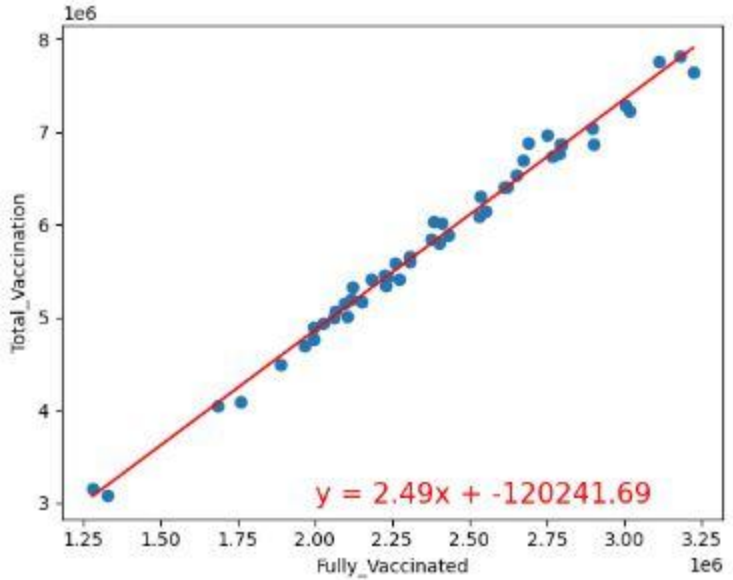
Vaccinations Dataset

Q1. What is the vaccine number for 5000000 people to be fully vaccinated?

Vaccination Regression

Process:

- Process the dataset to calculate the total number of vaccinations and fully vaccinated people
- The goal of Run a linear regression on the data to predict the answer



Total vaccin needed for 5000000 is 12329139.83826615

Total number of vaccine needed is

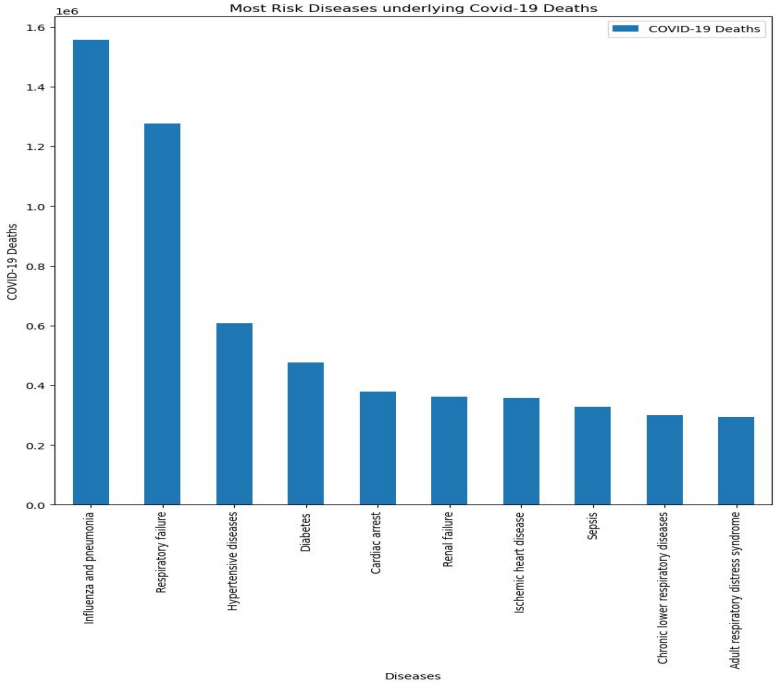
12329140

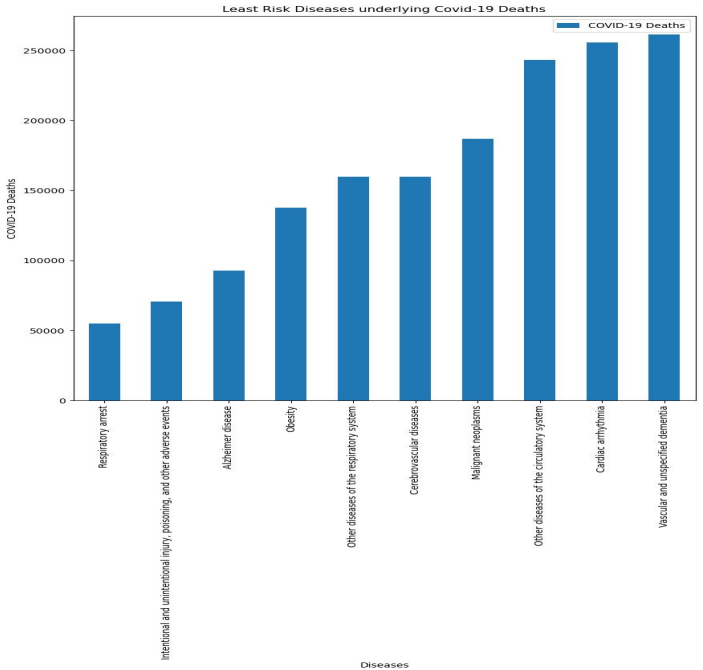
Underlying diseases

Q1. Which age group is most affected by covid 19?

Q2. Which patients are at high and low risk for covid-19 deaths?

Most and least Risky patients for Covid-19 Deaths



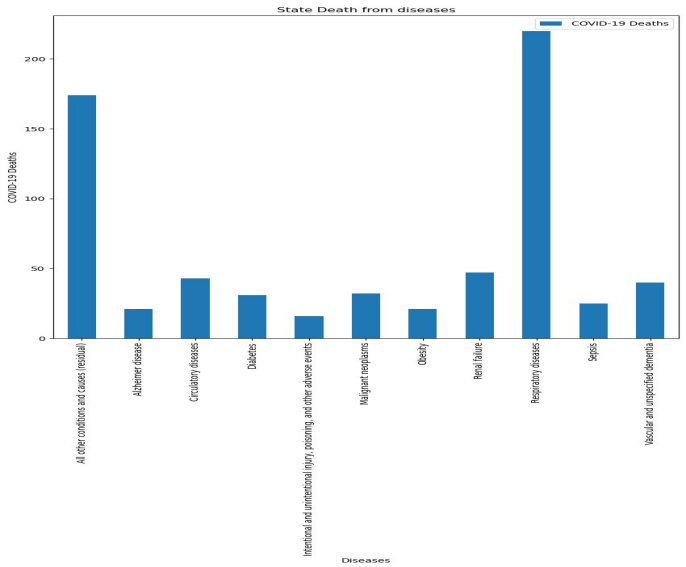


The patient died due to covid-19 (high-risk): Influenza and pneumonia,Respiratory Failure, Hypertensive diseases, Diabetes, Cardiac Arrest
Least Risky patients:Respiratory arrest,injury-poisoning, alzheimer,obesity,other respiratory diseases

Program to find Condition group from any State

In [70]:
```
loop_on=True

while loop_on:
    state=input("Which State you want to check the most covid death reason").capitalize()
    user_input=input("Do you want to check more state?if yes press any button, if no please write no").lower()
    df1=analytic_data.loc[analytic_data["State"]==state]
    df2=df1.groupby("Condition Group").max()

    if user_input=="no":
        loop_on=False

        df2.plot(kind="bar", y="COVID-19 Deaths",figsize=(10,10),xlabel="Diseases", ylabel="COVID-19 Deaths", title="State Death
        plt.show()
    else:


        df2.plot(kind="bar", y="COVID-19 Deaths",figsize=(10,10),xlabel="Diseases", ylabel="COVID-19 Deaths", title="State Death
        plt.show()
```

```
Which State you want to check the most covid death reasonalaska
Do you want to check more state?if yes press any button, if no please write nono
```



Results and Findings

- Most covid active cases percentage are in Rhode Island, Alaska, Kentucky, North Dakota and New Virginia states
- The least percentage of active cases are: Maryland, Oregon, Maine, Vermont, Washington
- The most recovered states are Alaska, North Dakota, Utah, West Virginia, Montana
- Vaccination number to fully vaccinated 5000000 people is 12329140
- Highly risky patients prone to death is Influenza and pneumonia, Respiratory Failure, Hypertensive diseases, Diabetes, Cardiac Arrest

Citations

- https://apps.bea.gov/api/_pdf/bea_web_service_api_user_guide.pdf
- https://www.census.gov/data/tables/time-series/demo/popest/2020s-counties-total.html
- https://github.com/nytimes/covid-19-data
- https://thefactfile.org/us-states-counties/
- https://gist.github.com/rogerallen/158359
- https://journals.lww.com/academicmedicine/fulltext/2004/12000/health_disparities_based_on_socioeconomic.4.aspx
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7314918/
- https://wallethub.com/edu/e/states-with-the-best-schools/5335
- https://nces.ed.gov/pubs2018/2018052/tables/table_02.asp
- https://data.cdc.gov/NCHS/Conditions-Contributing-to-COVID-19-Deaths-by-Stat/hk9y-quqm
- https://www.kaggle.com/datasets/paultimothymooney/usa-covid19-vaccinations
- https://www.kaggle.com/discussions/general/246407