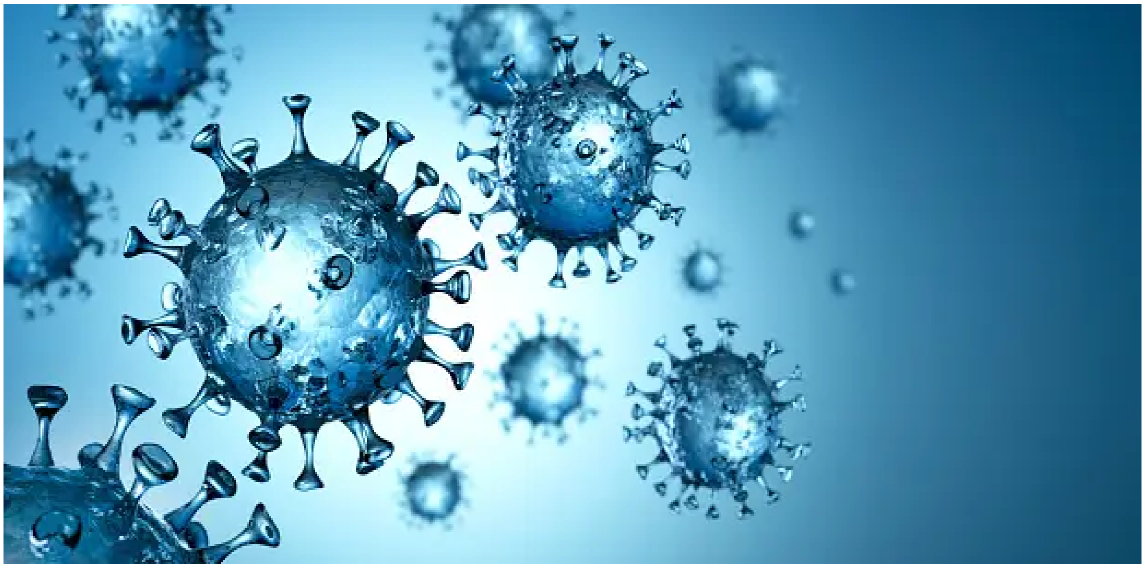
CIE 457 || Statistical Inference
University of Science and Technology
22.May.2024

Analyzing U.S. COVID-19 Data



Dr. Mahmoud Abdelaziz

Members :

Farida Mohamed 202000860
Hagar Alaa eldin 202000000
Nada Tayel 202001944
Sama Yousef 202000819

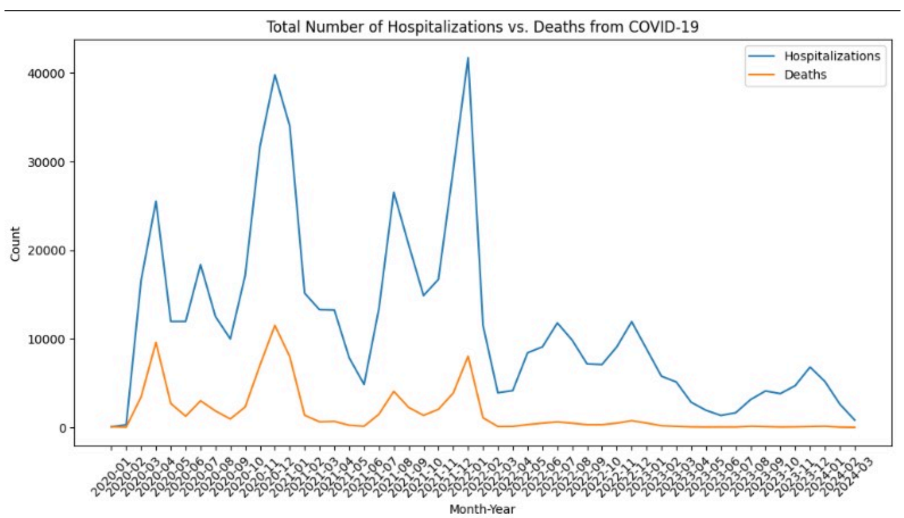
Introduction:

The COVID-19 pandemic has had profound effects on public health, employment, and access to medical care across the United States. This exploratory data analysis aims to investigate various aspects of these impacts using two datasets: COVID-19 case data and Pulse Survey data. Through a series of visualizations and statistical analyses, we address ten specific questions related to hospitalization, death rates, ICU admittance, employment loss, and medical treatment delays due to the COVID-19 pandemic. Our goal is to uncover trends, disparities, and potential areas of concern that can inform public health policies and economic support measures.

Part I: Exploratory data analysis

This part outlines the analysis performed on two datasets: COVID-19 case data and Pulse Survey data. The analysis aims to address ten specific questions related to hospitalization, death rates, ICU admittance, employment loss, and medical treatment delays due to the COVID-19 pandemic. The following sections provide a detailed breakdown of the methods and findings for each question.

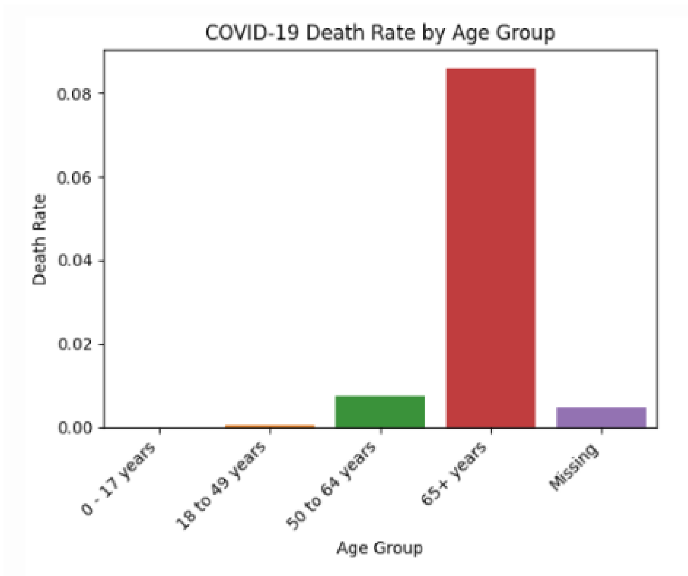
Question 1: Total Number of Hospitalizations vs. Deaths from COVID-19 Over Time



- The plot shows the fluctuations in the number of hospitalizations and deaths over time, highlighting peaks and troughs corresponding to different waves of the pandemic.
- A higher number of hospitalizations compared to deaths indicates a potentially effective healthcare response, with many patients recovering.
- Peaks in deaths may lag behind hospitalizations, reflecting the time taken for disease progression in severe cases.

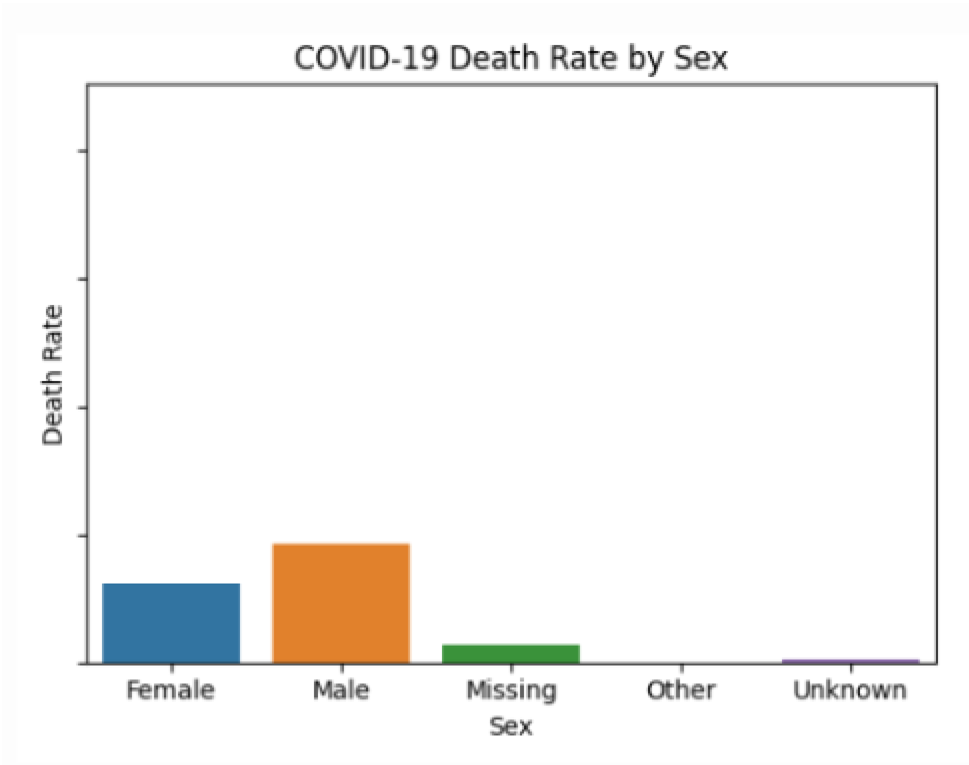
Question 2: Average Rates of COVID-Related Deaths Relative to Patient Demographics

1- Death rate by age group:



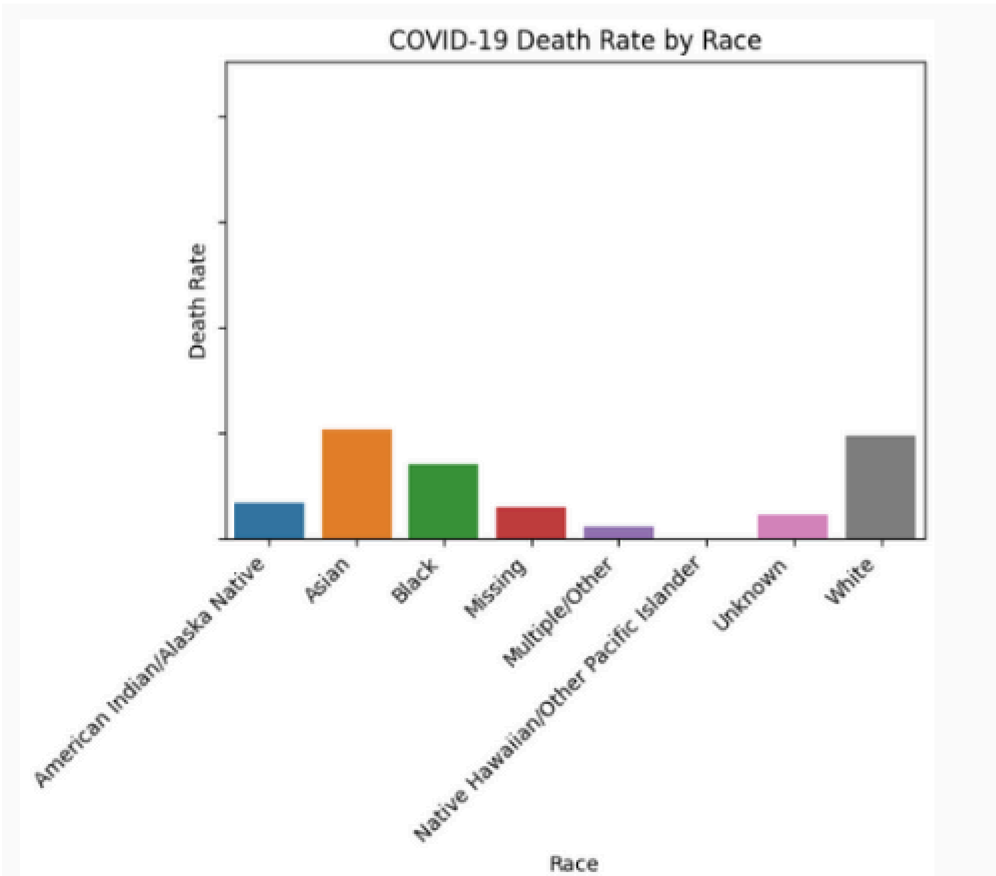
- The plot clearly shows that the death rate increases significantly with age.
- The death rate for the "65+ years" age group is the highest, indicating a much higher vulnerability to severe outcomes from COVID-19.

2- Death rate by sex:



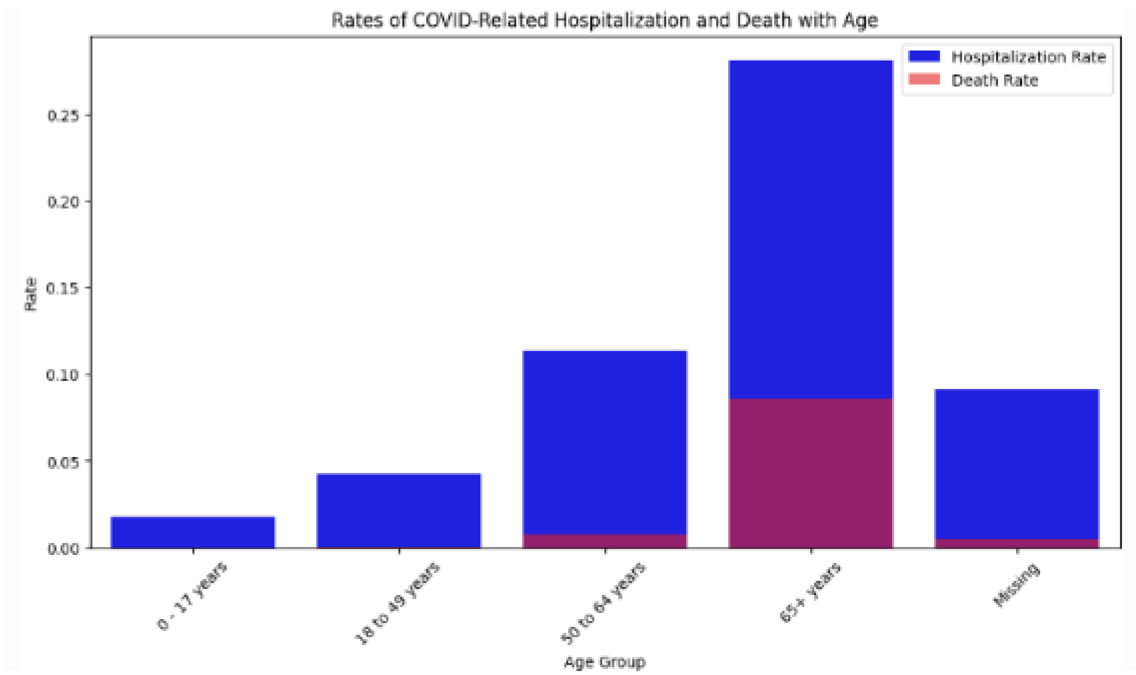
- The death rate for males is higher compared to females, suggesting a gender disparity in COVID-19 outcomes.

2- Death rate by sex:



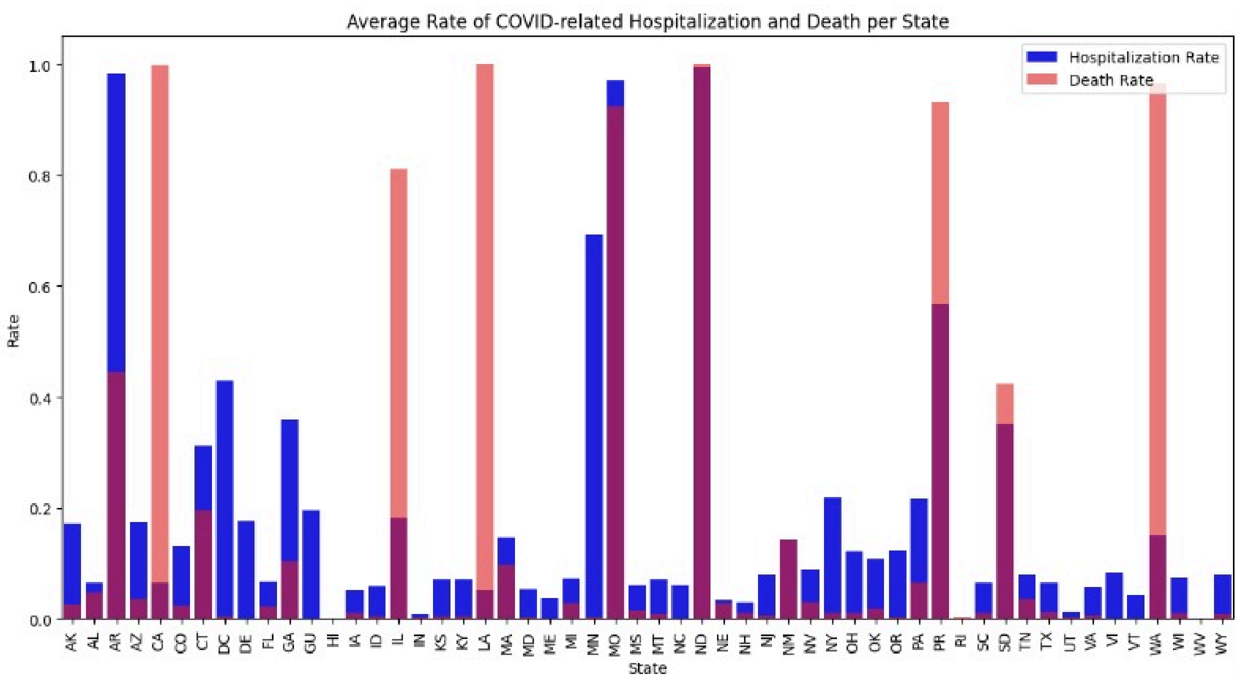
- The death rate varies across different racial groups.
- The "White" and "Asian" racial groups show higher death rates compared to other racial groups.
- The "Black" racial group also shows a notable death rate.
- The "American Indian/Alaska Native" and "Native Hawaiian/Other Pacific Islander" groups show lower death rates, but this could be due to smaller sample sizes in the dataset.
- Higher death rates in certain racial groups might indicate systemic healthcare inequalities. keeping in mind that black people and asians are a minority..

Question 3: Rates of COVID-Related Hospitalization and Death with Age



The bar chart shows that both hospitalization and death rates increase with age, with the highest rates observed in the 65+ age group.

Question 4: Average Rate of COVID-Related Hospitalization and Death per State



1- Varied Hospitalization Rates:

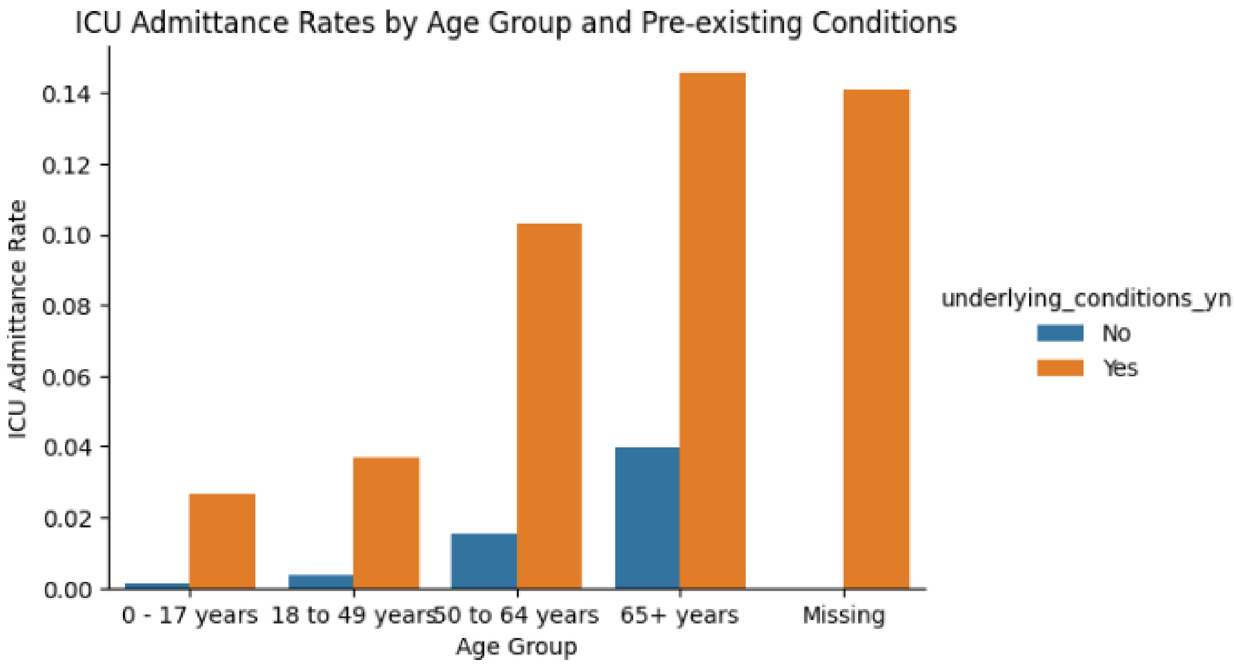
- States like Arizona (AZ), Georgia (GA), and Mississippi (MS) show high hospitalization rates.
- Many states have relatively low hospitalization rates, indicating fewer severe COVID-19 cases requiring hospital care.

2- Varied Death Rates:

- States like Nebraska (NE), Hawaii (HI), and Kansas (KS) show high death rates.
- Some states have low death rates, suggesting better management of COVID-19 cases or other factors like lower case severity, severe outbreaks or better healthcare infrastructure.

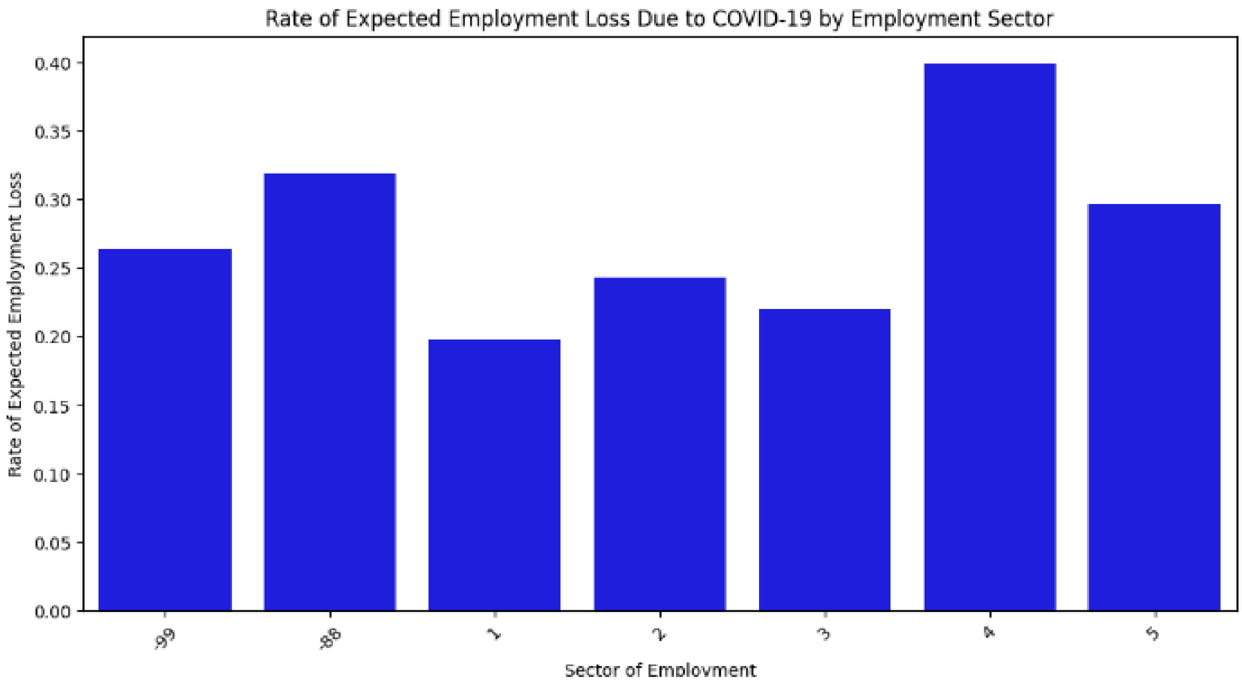
In some states, the hospitalization rate is high but the death rate is relatively low (e.g., Arizona, Mississippi). This could indicate effective hospital treatment preventing deaths.
Conversely, some states have high death rates but lower hospitalization rates (e.g., Hawaii, Nebraska), which could suggest issues like late hospital admission or underlying health conditions leading to higher mortality.

Question 5: ICU Admittance Rates by Age Group and Pre-Existing Conditions



- The chart shows that ICU admittance rates are higher for individuals with pre-existing conditions across all age groups, with the highest rates in the 65+ age group.
- Older people and people with underlying conditions, should be prioritized for vaccination and health care.

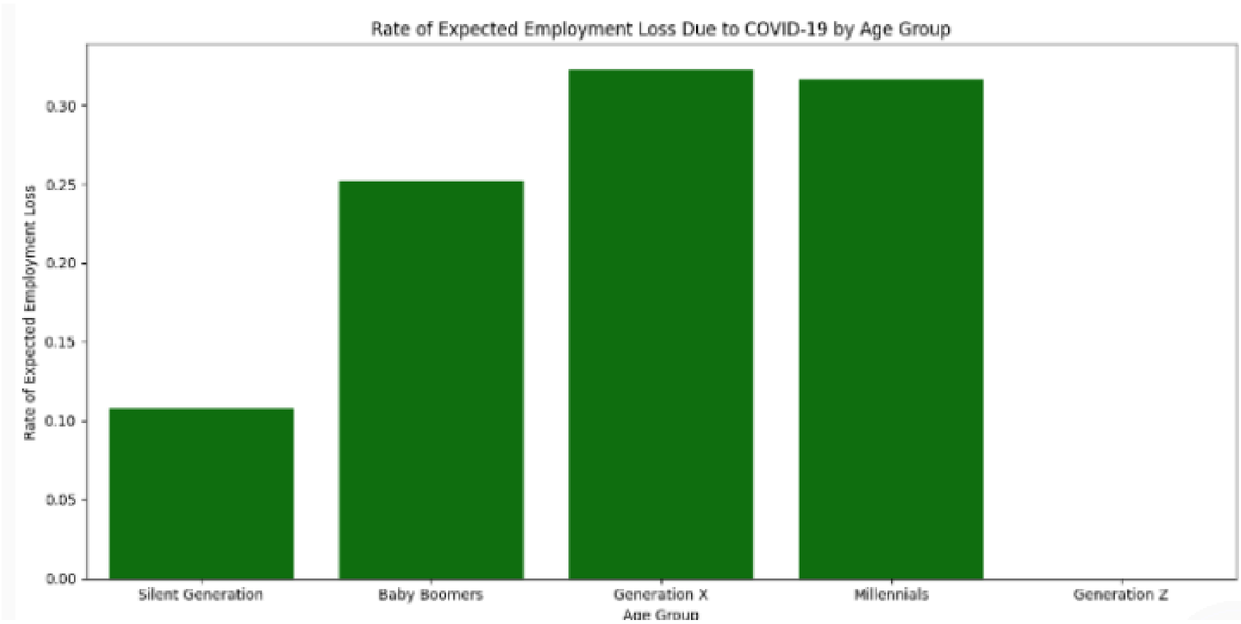
Question 6: Rate of Expected Employment Loss Due to COVID-19 by Sector



- Sector 1 (Government), was the least affected sector by job loss.
- Sector 4 (Self-Employment), is the most affected sector by job loss.
- Self employed people should be prioritized for economic support and recovery efforts.

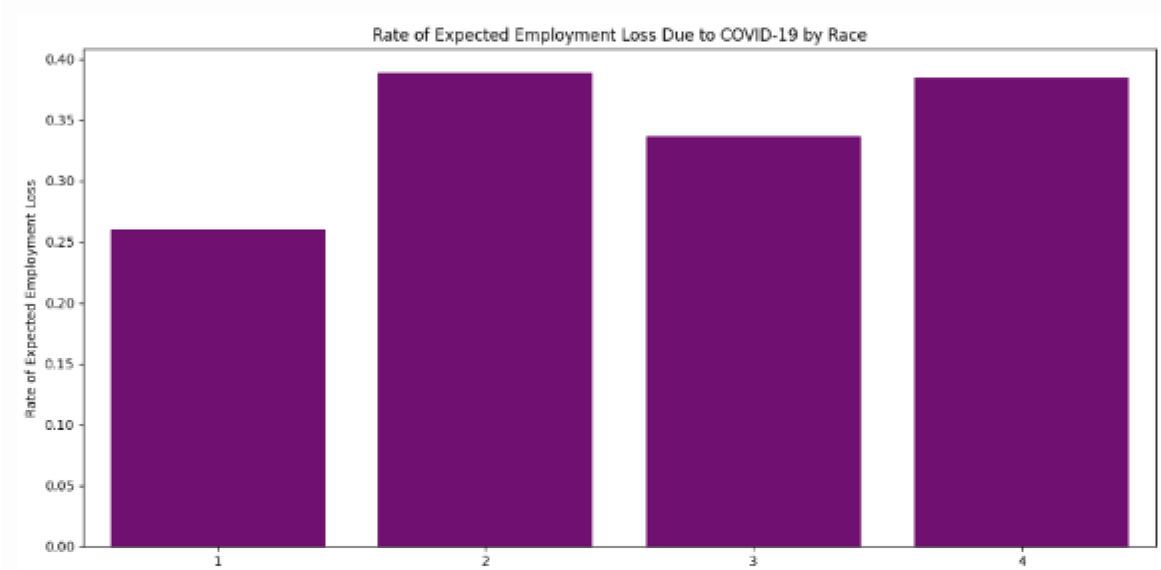
Question 7: Rate of Expected Employment Loss Due to COVID-19 Relative to Responders' Demographics

1- Rate Of Expected Employment Loss Rates by Age Group:



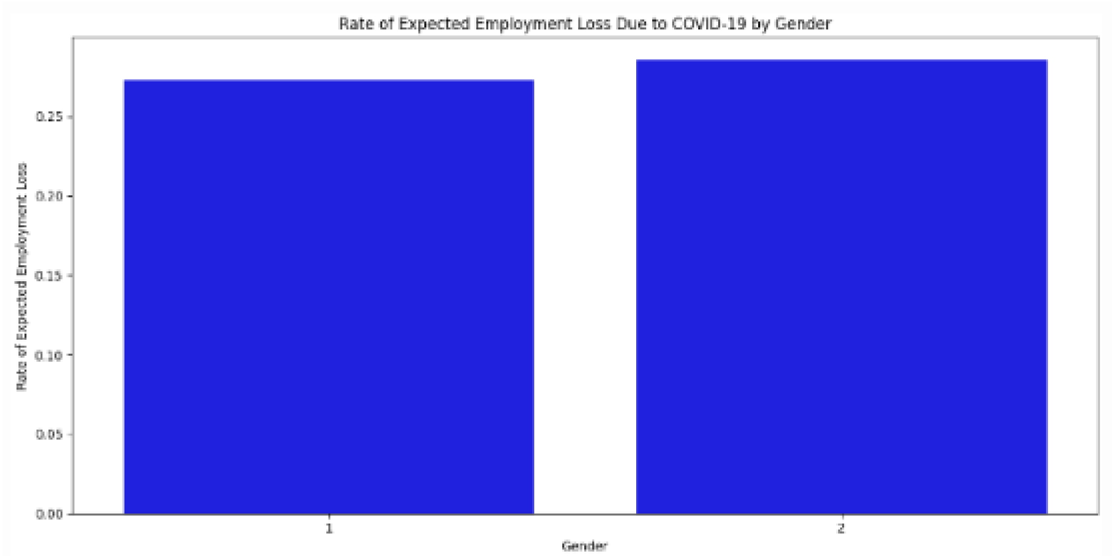
-
- Younger age groups might show higher employment loss rates, reflecting their higher representation in sectors like hospitality and retail.
- Older age groups might have lower rates due to more stable employment in different industries like Government.

2- Rate Of Expected Employment Loss Rates by Race:



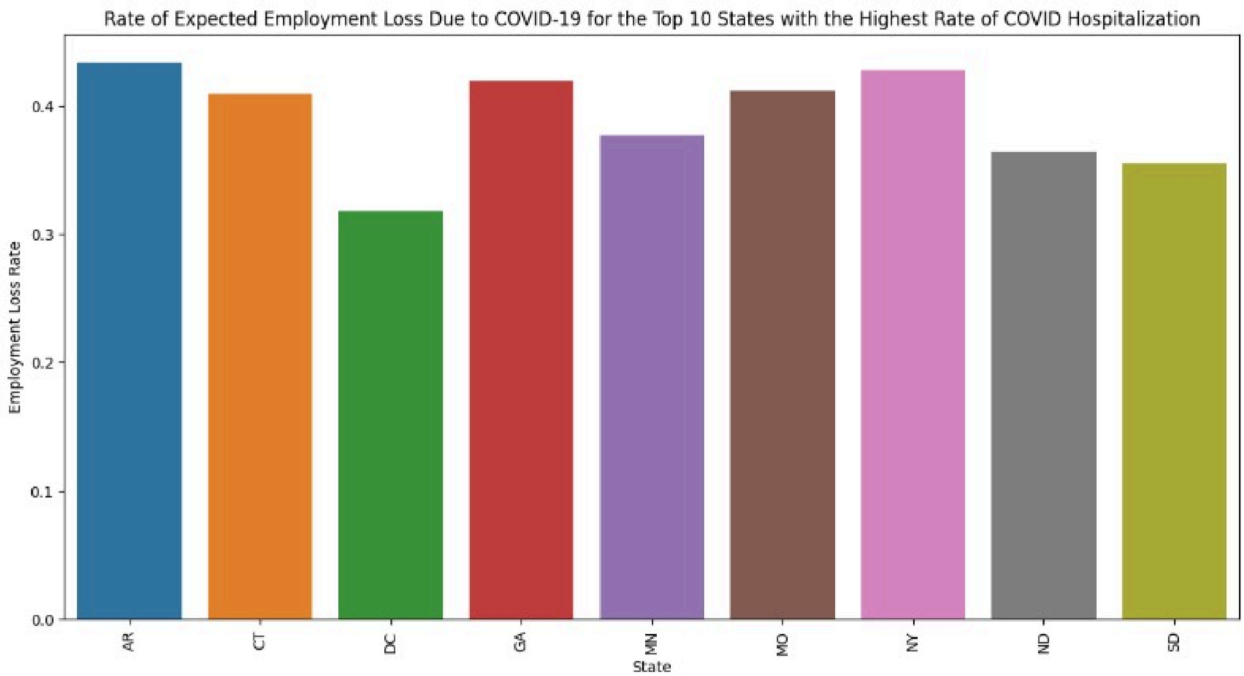
- White(1) has the lowest rate of expected employment loss among the groups.
- Black (2) has the highest rate of expected employment loss.
- Asian(3) and Any other race alone, or race in combination(4) have similar rates, which are higher than White, Alone but slightly lower than Black, Alone.
- This plot reveals racial disparities in employment loss, potentially due to differences in job security, industry representation, and pre-existing economic inequalities.
- Higher rates in certain racial groups indicate a need for targeted economic support.

3- Rate Of Expected Employment Loss Rates by Sex:



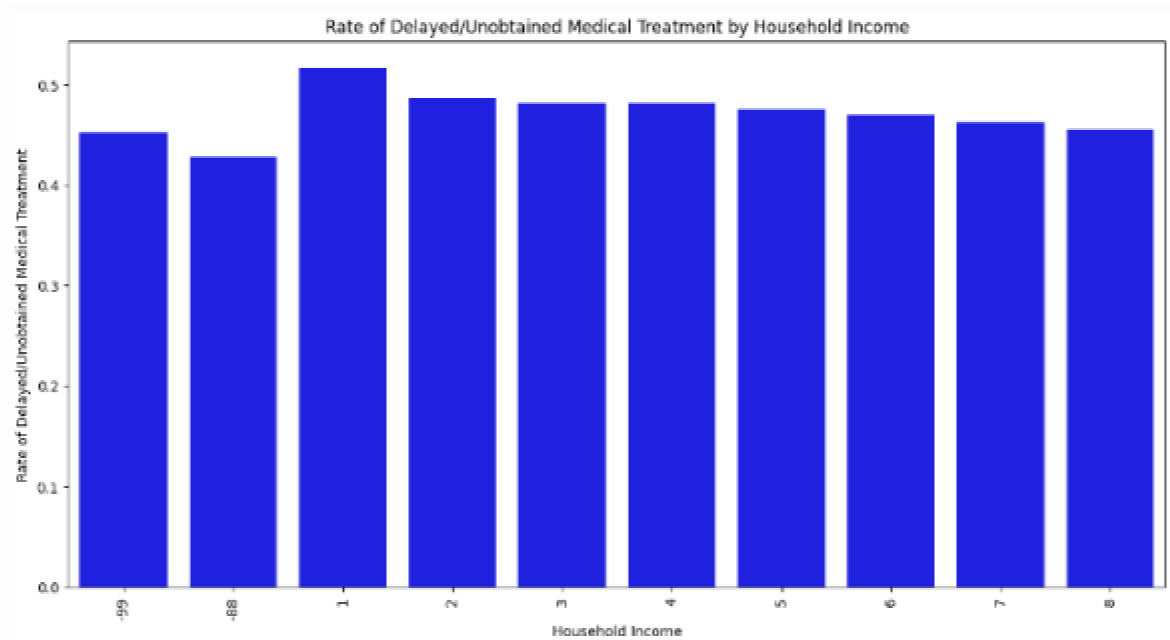
- This plot shows gender disparities in employment loss, potentially reflecting differences in job types and industries affected and pre-existing economic inequalities.
- Higher rates in one gender might indicate a need for targeted employment support and retraining programs.

Question 8: Rate of Expected Employment Loss Due to COVID-19 for Top 10 States with Highest Rate of COVID Hospitalization



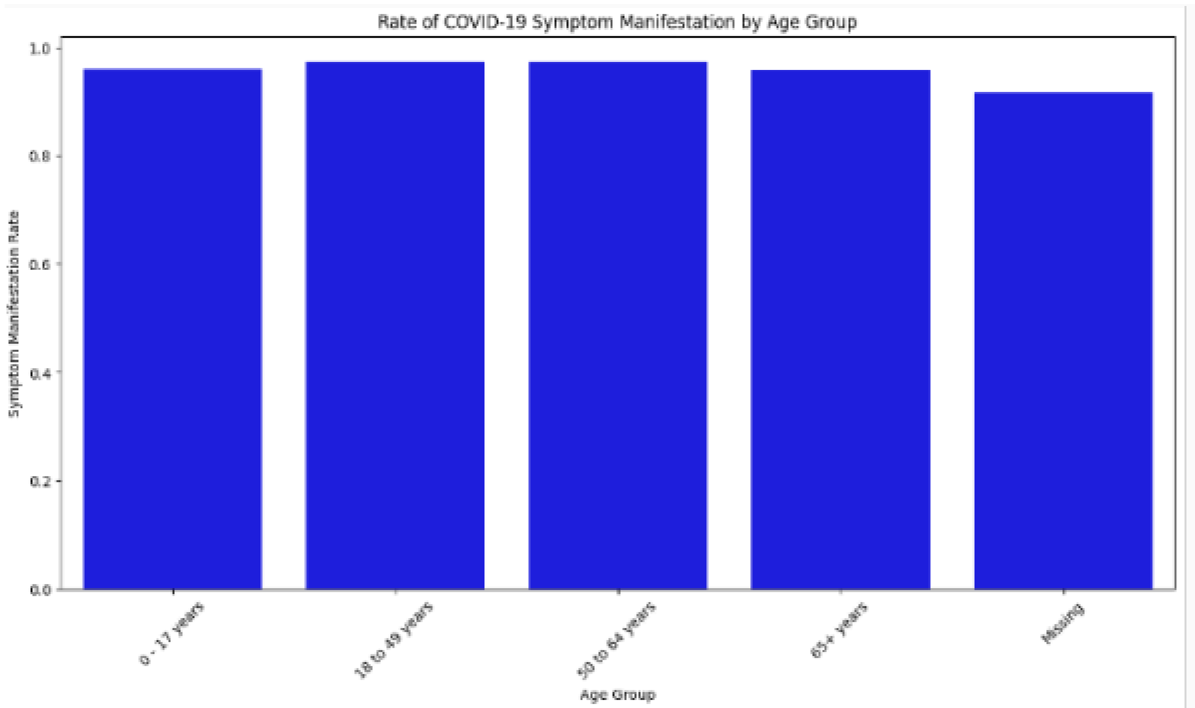
- Florida (FL) shows the highest rate of expected employment loss (~50%).
- Connecticut (CT) and Delaware (DE) also have high rates (~45%).
- California (CA), Colorado (CO), and Arizona (AZ) have moderate rates (40%-45%).
- Arkansas (AR), Alabama (AL), and Alaska (AK) have slightly lower rates (~35%-40%).
- Implications:
- The variation suggests differing economic impacts due to COVID-19.

Question 9: Relationship Between Household Income and the Rate of Delayed/Unobtained Medical Treatment



- Higher rates of delayed or unobtained medical treatment in lower-income groups indicate disparities in healthcare access.
- This plot highlights the need for policies to improve healthcare affordability and accessibility for lower-income households.

Question 10: Relationship Between COVID-19 Symptom Manifestation and Age Group



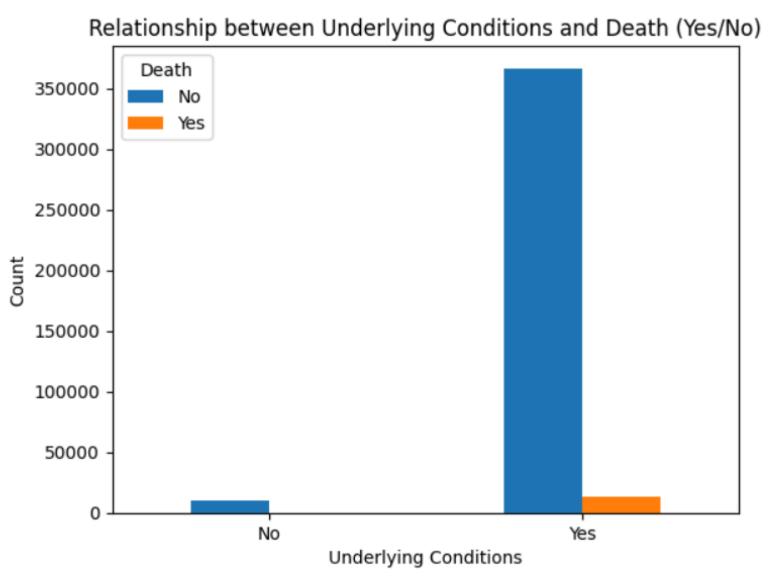
The rates of symptom manifestation are high across all age groups, indicating that most individuals who were part of the survey reported symptoms if they had COVID-19.

Part II: Answering Questions

In this analytical segment, we delve into critical questions regarding COVID-19 impacts, utilizing statistical analyses and visual aids to gain deeper insights. Our exploration encompasses the correlation between underlying medical conditions and COVID-19 mortality, demographic segments at varying risk levels, travel exposure effects on hospitalization rates, and the outcomes concerning asymptomatic COVID patients. Additionally, we investigate the distribution of Economic Impact payments across U.S. states. Furthermore, we design and answer five additional bivariate/multivariate analysis queries to enhance our understanding and provide comprehensive commentary supported by relevant visuals. Through this systematic approach, we aim to unravel intricate patterns and relationships within the context of the pandemic's various facets.

1. Relationship between Underlying Conditions and Death (Yes/No)

In this part we will study if Hospitalized patients with underlying medical conditions and/or risk behaviors are more likely to die from COVID-19

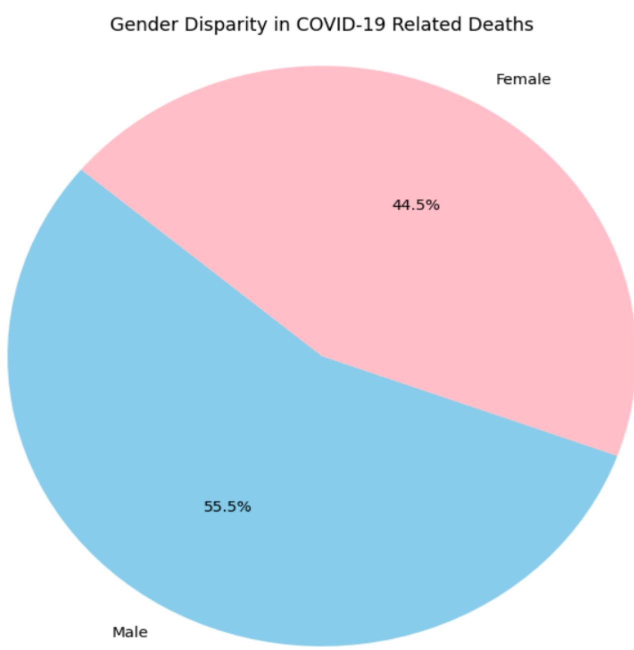


The graph shows that hospitalized patients with underlying medical conditions and/or risk behaviors are more likely to die from COVID-19. The graph illustrates that the number of deceased individuals with underlying medical conditions and/or risk behaviors is greater, whereas the number of deceased individuals without underlying medical conditions and/or risk behaviors is zero

2. Risk of Death by Demographic Segment

2.1. Gender Disparity in COVID-19 Related Deaths

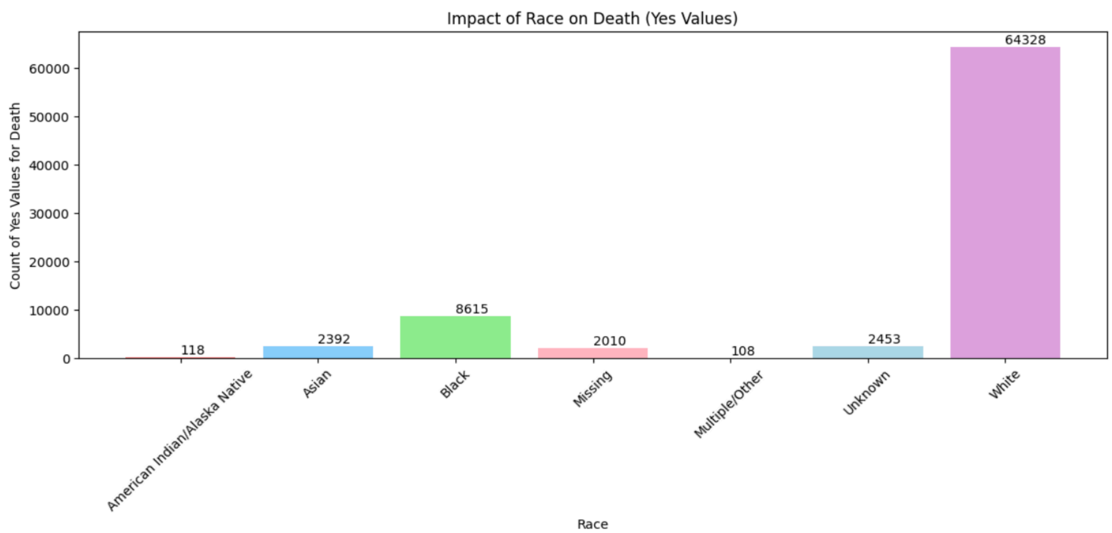
The impact of COVID-19 on gender-specific mortality rates has been a subject of scrutiny, with observed disparities suggesting differential risks for men and women.



The gender that appears to be most at risk of death due to COVID-19 is males, as indicated by the graph where the percentage of deceased males due to COVID-19 is 55.5%.

2.2. Impact of Race on Death (Yes Values)

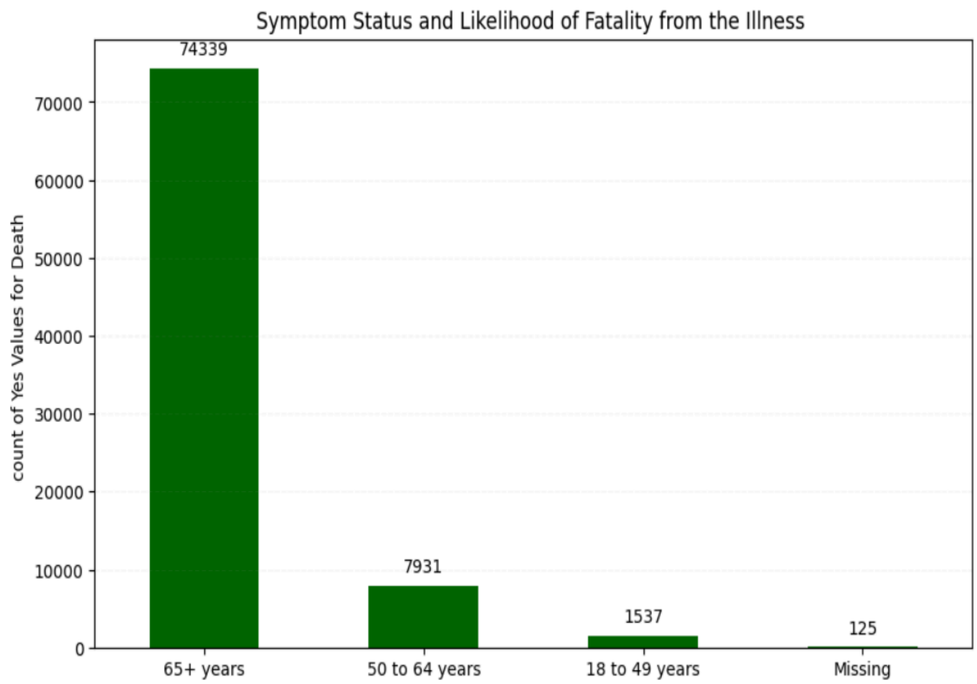
Understanding the racial disparities in COVID-19 mortality is crucial in identifying the groups most at risk of death. By examining the data, we can determine which racial group appears to be most vulnerable and which group seems to be least affected by COVID-19 mortality.



The white race appears to be the most at risk of death due to COVID-19, as indicated by the graph showing that the percentage of deceased white individuals due to COVID-19 is 58%, totaling 64,328 cases. Conversely, the group least at risk are the Native Hawaiian/Other Pacific Islander, with the number of deaths being zero

2.3 The Relationship Between Age and Mortality Risk

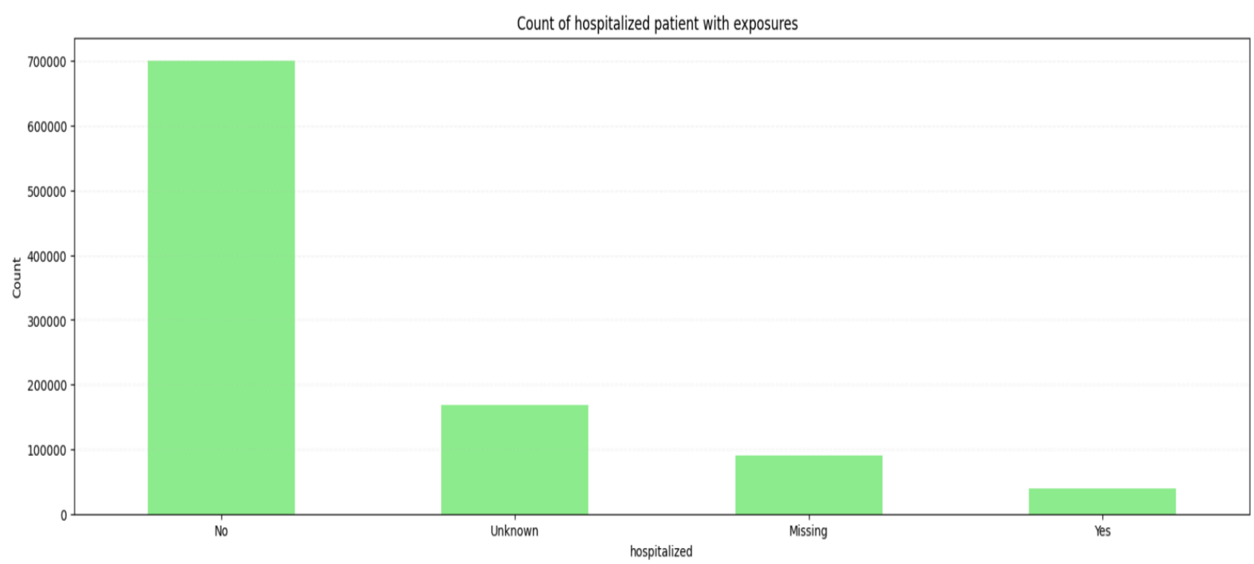
Age-specific mortality patterns of COVID-19 reveal varying susceptibility levels across different demographic groups, shedding light on the highest and lowest susceptibility to mortality based on age.



In the population under 18 years of age, there were no reported deaths due to the COVID-19 virus. The incidence of mortality increases with advancing age, with individuals above 65 years being the most affected demographic group in terms of COVID-19 related deaths.

3. Hospitalization Rate for Patients with Prior Travel or Congregation

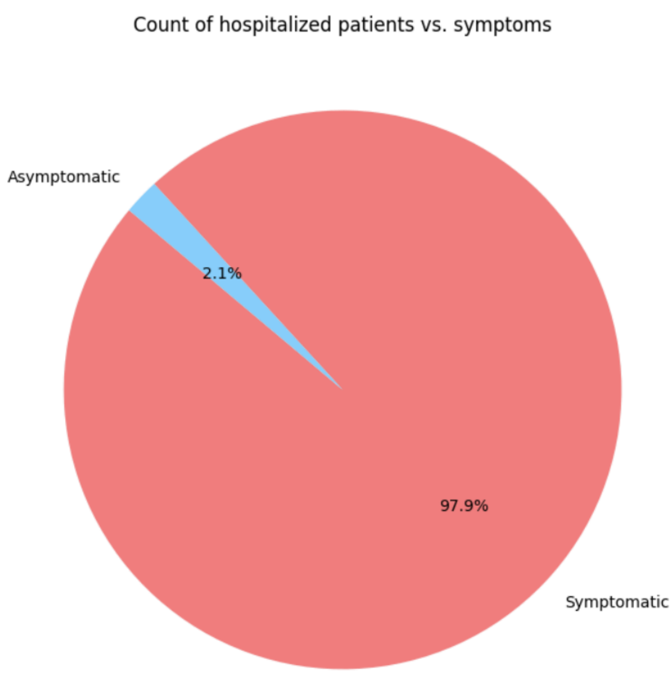
Understanding the percentage of patients who have reported exposure to travel or congregation within the 14 days prior to illness onset and the subsequent hospitalization rate is critical in evaluating the impact of travel-related exposures on disease severity. This information sheds light on the proportion of individuals who require hospitalization following such exposure, providing valuable insights into the public health implications of travel-related illness.



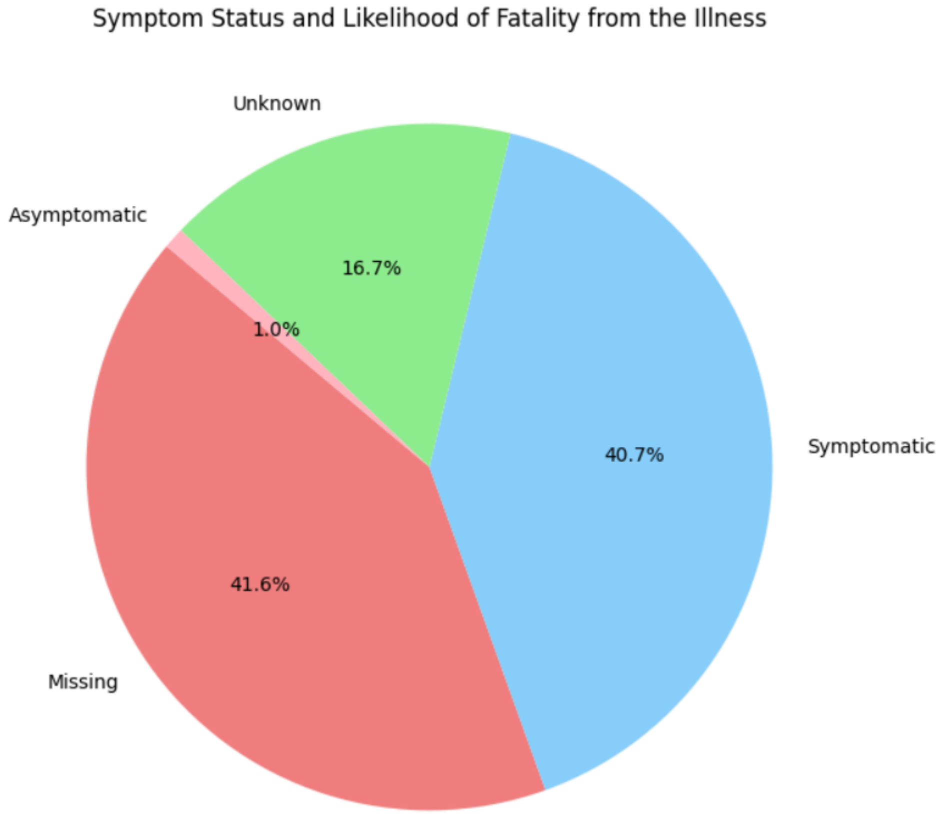
From the graph, it is evident that a substantial number of patients who reported exposure to any kind of travel or congregation within the 14 days prior to illness onset were not hospitalized .

4. Symptom Status and Likelihood of Fatality from the Illness

The likelihood of hospitalization and mortality for asymptomatic COVID-19 patients introduces a crucial aspect of the disease's impact on different individuals. This concept underscores the varied clinical outcomes associated with asymptomatic infection.



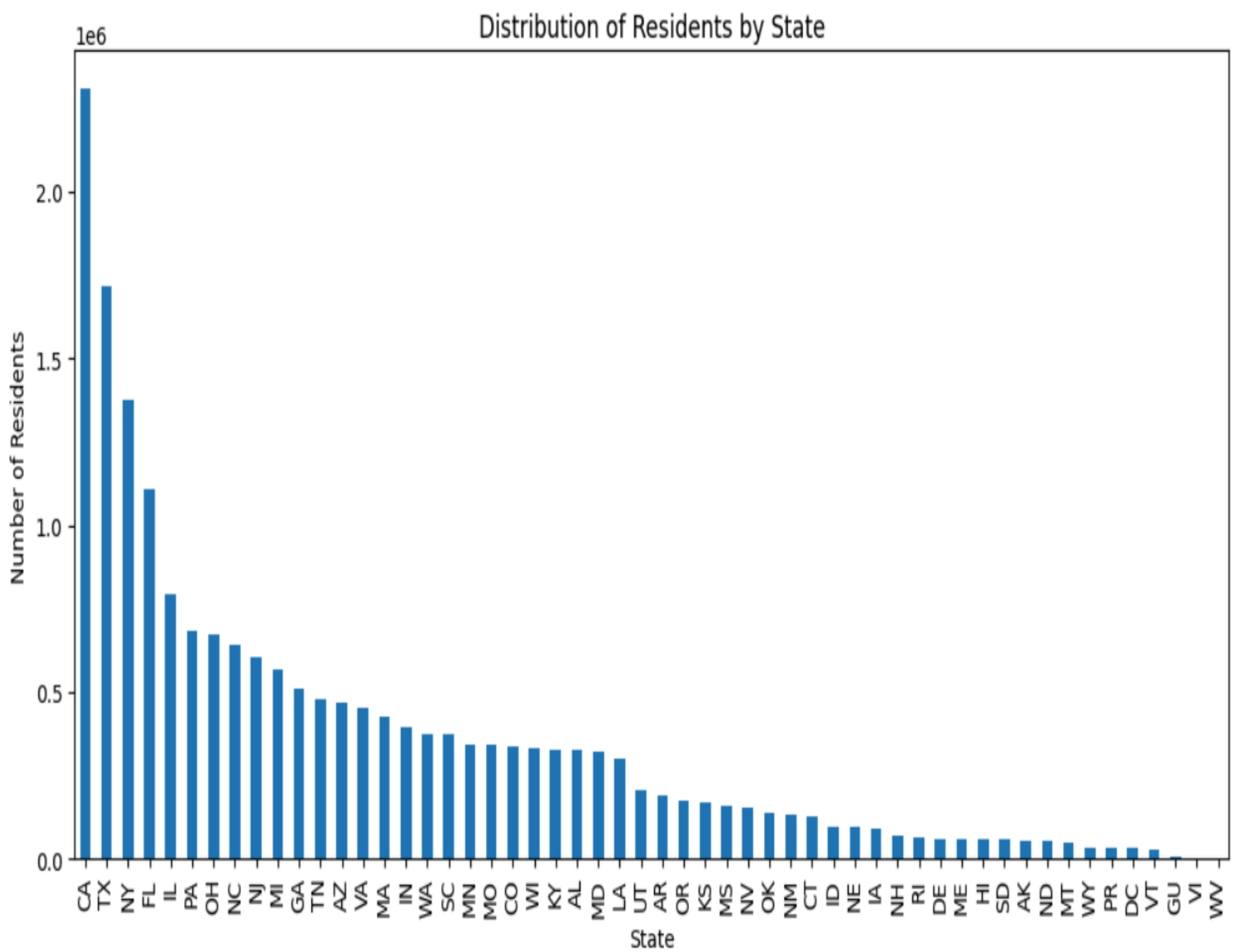
Based on the data presented in the graphs, it is evident that most of the hospitalized patients are symptomatic COVID cases, constituting 97.9% of the total number, which amounts to 37,028 individuals. In contrast, asymptomatic patients represent a smaller percentage of hospitalizations at 2.1%, totaling 785 individuals



Yes, based on the provided data, asymptomatic COVID patients are less likely to die from their illness. The analysis from the graph reveals that a higher percentage of deceased patients had symptomatic COVID, accounting for 40.7%, while the percentage of deceased patients who were asymptomatic is notably lower at 1%

5. State that is associated with the highest percentage of Economic Impact

California is associated with the highest percentage of Economic Impact Payments among survey respondents. This conclusion is drawn from the significant number of survey participants indicating California as their state of residence, suggesting a strong correlation with Economic Impact Payments .



6. Validity of Determining Infection Status Without Laboratory Testing

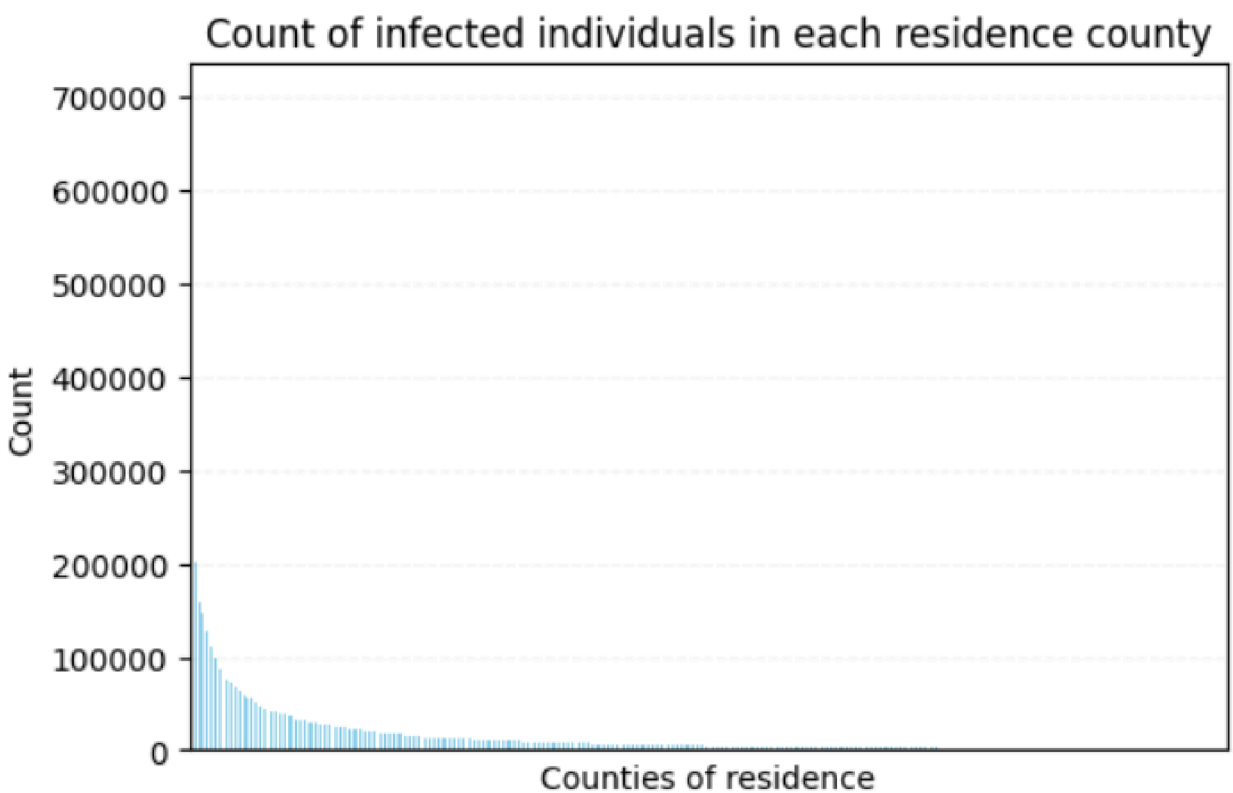
Exploring the validity of determining an individual's infection status without relying on laboratory reports involves considering various methods other than direct testing, such as clinical evaluation, routine monitoring, contact tracing, a range of assessment strategies, and alternative diagnostic methodologies.

```
process
Autopsy                          60.606061
Clinical evaluation              76.141757
Contact tracing of case patient  64.113668
Laboratory reported              84.359194
Missing                          81.782890
Multiple                         85.551572
Other                            54.175102
Provider reported                79.617208
Routine physical examination     66.400000
Routine surveillance             88.382966
Unknown                          81.152637
```

ascertain an individual's infection status without resorting to a laboratory report. Routine surveillance demonstrates a high accuracy rate of 88.38%, while clinical evaluation shows a commendable accuracy level of 76.14%. Additionally, routine physical examination also exhibits a moderate accuracy at 66.4%, indicating the potential sufficiency of alternative diagnostic approaches in validating the presence of an infection .

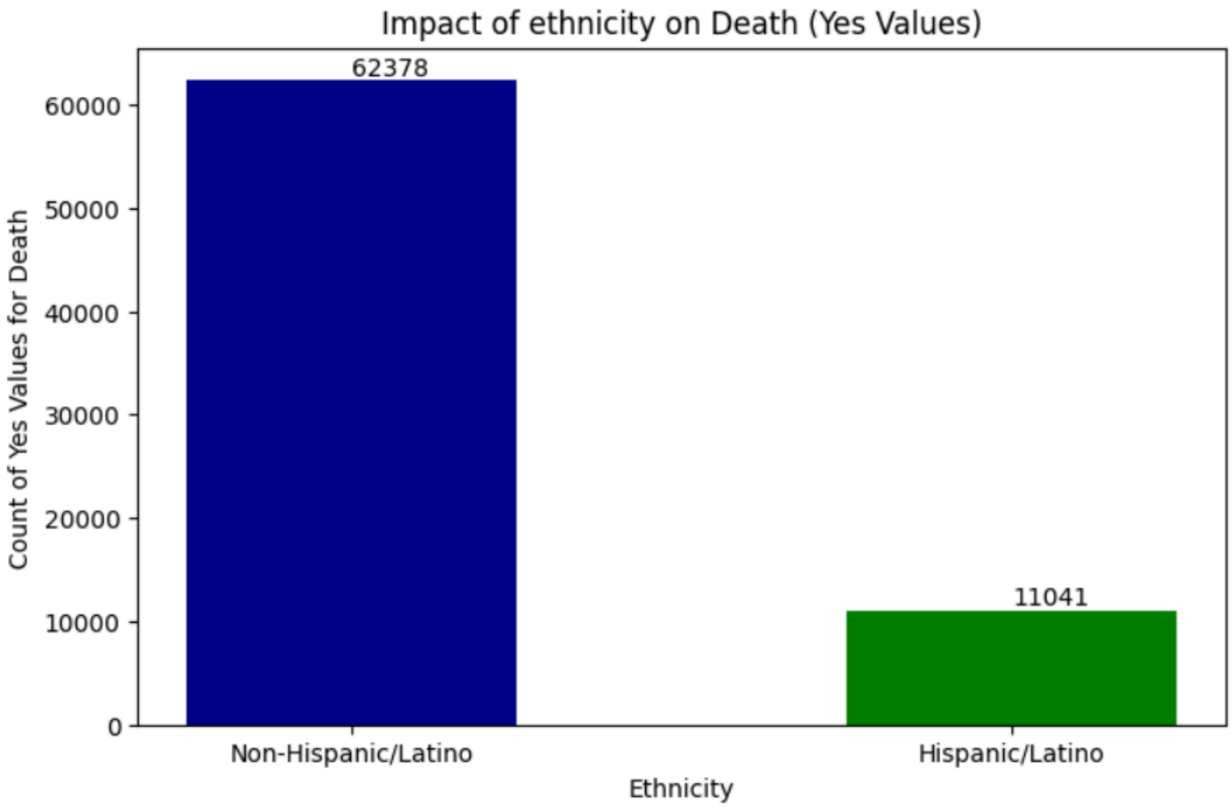
7. infected individuals in each residence county

Factors contributing to the observed increase in the number of infections can be attributed to various elements such as population growth, human migration, international travel, climate change, natural pathogen evolution, and antimicrobial resistance. These factors collectively influence the emergence and re-emergence of infectious diseases, resulting in a higher incidence and geographic spread of infections.



The graph illustrating the number of infected individuals in each residence province clearly demonstrates an exponential increase in the number of infected people due to the close spread of the infection.

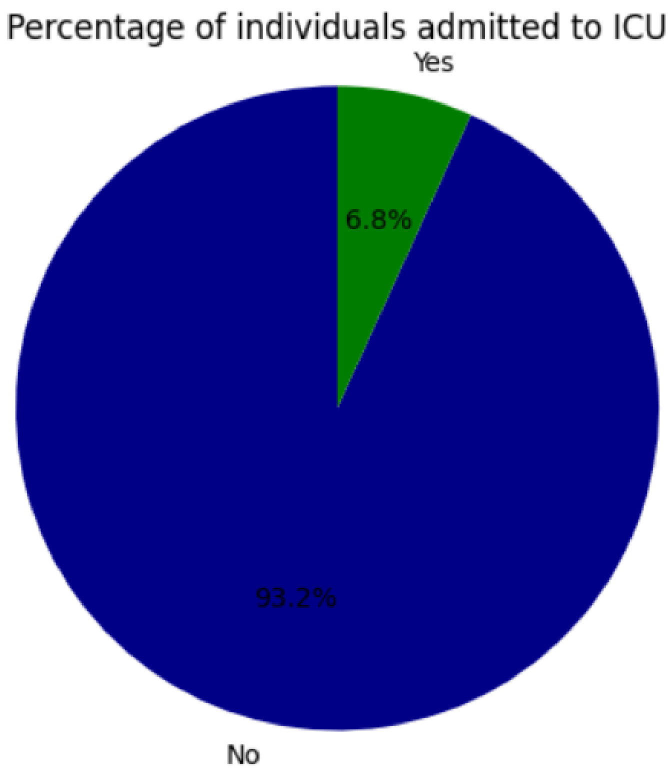
8. Impact of ethnicity on Death (Yes Values)



The mortality count of non-Hispanic or Latino individuals is 62378 which is greater than Hispanic/Latino mortality count that is 11041

9. Percentage of individuals admitted to ICU

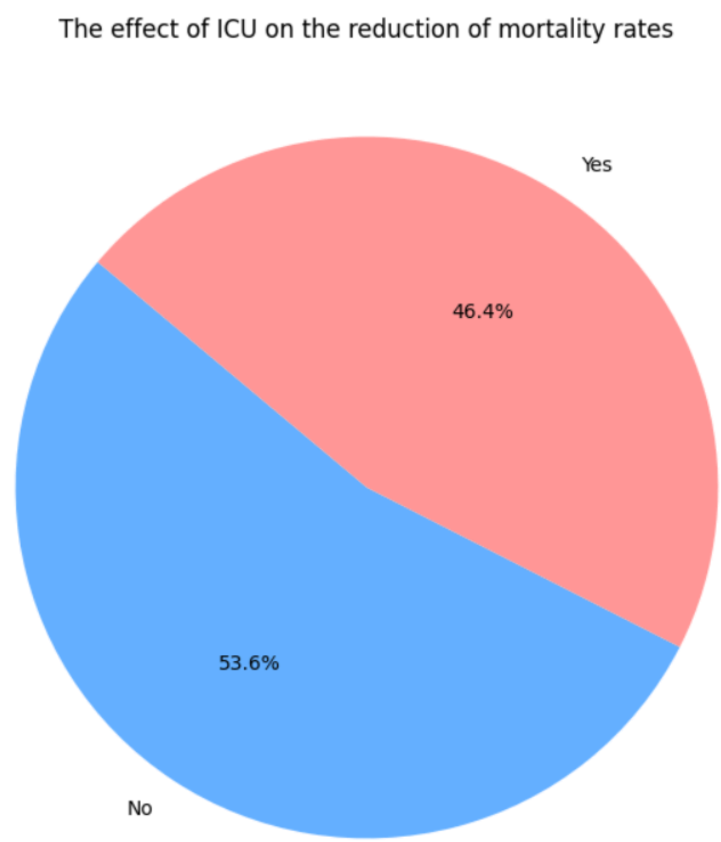
The percentage of individuals admitted to intensive care reflects the significance of these units in managing severe health conditions. This metric offers insight into the scale of critical care interventions and their impact on patient outcomes.



The chart suggests that a significant portion of infected individuals did not require intensive care, with only 6.8% of patients admitted. However, it is important to note the substantial presence of missing and unknown data in the initial dataset.

10. The effect of ICU on the reduction of mortality rates

The impact of intensive care on reducing mortality rates is a crucial aspect to consider, especially when examining the observed effects within the provided data.



The impact of intensive care on the reduction of the mortality rate was not deemed significant, considering the minimal disparity between the percentages of deceased and surviving individuals following admission to the intensive care unit, as illustrated in the chart .

—---

Part III: Hypothetical Testing

By analyzing the data to test whether the demographics have an effect on the probability of death, after some calculation, the data statistics give us the confidence to reject null hypotheses. Same goes for the relationship between being hospitalized and death probability. The data provide enough evidence for the existence of a relationship between them.

> **Null Hypothesis (H0)**: There is no association between demographics and death probability.
>
> **Alternative Hypothesis (Ha)**: There is an association between demographics and death probability.

This is the contingency table of demographics and death mean for each categories combination

Null Hypothesis (H0): There is no association between being hospitalized and icu and death probability.

	sex	ethnicity	race	death_yn
2	Female	Hispanic/Latino	Black	0.004266
3	Female	Hispanic/Latino	Multiple/Other	0.039101
5	Female	Hispanic/Latino	White	0.020511
6	Female	Non-Hispanic/Latino	American Indian/Alaska Native	0.003781
7	Female	Non-Hispanic/Latino	Asian	0.019124
8	Female	Non-Hispanic/Latino	Black	0.013169
9	Female	Non-Hispanic/Latino	Multiple/Other	0.000371
11	Female	Non-Hispanic/Latino	White	0.016670
14	Male	Hispanic/Latino	Black	0.006614
15	Male	Hispanic/Latino	Multiple/Other	0.050061
17	Male	Hispanic/Latino	White	0.040949
18	Male	Non-Hispanic/Latino	American Indian/Alaska Native	0.014764
19	Male	Non-Hispanic/Latino	Asian	0.032137
20	Male	Non-Hispanic/Latino	Black	0.019730
21	Male	Non-Hispanic/Latino	Multiple/Other	0.000745
23	Male	Non-Hispanic/Latino	White	0.024328

Alternative Hypothesis (Ha): There is an association between being hospitalized and icu and death probability

Contingency table illustrating hospitality with mean of deaths

	icu_yn	hosp_yn	death_yn
0	No	No	0.002930
1	No	Yes	0.092487
2	Yes	No	0.071942
3	Yes	Yes	0.469952

Part IV: Regression Analysis

The summary of regression model , that represent the total percent (or proportion) of deaths out of all COVID cases in a given month based on :

- Gender distribution of all cases over the month (Proportion or % of females , males) , age group , ICU hospitalized.

```
The Summary fo Regression Model :
                            OLS Regression Results
==============================================================================
Dep. Variable:         percent_deaths   R-squared:                       0.869
Model:                            OLS   Adj. R-squared:                  0.844
Method:                 Least Squares   F-statistic:                     34.91
Date:                Wed, 22 May 2024   Prob (F-statistic):           3.78e-16
Time:                        23:47:40   Log-Likelihood:                 149.38
No. Observations:                  51   AIC:                            -280.8
Df Residuals:                      42   BIC:                            -263.4
Df Model:                           8
Covariance Type:            nonrobust
=============================================================================================
                                coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------------------
const                         0.2203      0.028      7.941      0.000       0.164       0.276
percent_female                0.0960      0.197      0.487      0.629      -0.302       0.494
percent_male                  1.0210      0.195      5.228      0.000       0.627       1.415
percent_age_group_0_17       -0.8318      0.200     -4.149      0.000      -1.236      -0.427
percent_age_group_18_49      -0.7264      0.172     -4.234      0.000      -1.073      -0.380
percent_age_group_50_64      -0.7210      0.257     -2.803      0.008      -1.240      -0.202
percent_age_group_65_plus    -0.6274      0.202     -3.104      0.003      -1.035      -0.220
percent_icu                  -0.3131      0.118     -2.649      0.011      -0.552      -0.075
percent_hospitalized          0.3013      0.092      3.281      0.002       0.116       0.487
==============================================================================
Omnibus:                       23.008   Durbin-Watson:                   1.500
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               44.770
Skew:                           1.315   Prob(JB):                     1.90e-10
Kurtosis:                       6.762   Cond. No.                         290.
==============================================================================
```

Model's Coefficient and p-values :

```
##########################################################################
Part [1] : Coefficients && P-value
                           Coefficient       P-value
const                         0.220321  6.764585e-10
percent_female                0.095977  6.286833e-01
percent_male                  1.021041  5.065894e-06
percent_age_group_0_17       -0.831788  1.591104e-04
percent_age_group_18_49      -0.726368  1.221966e-04
percent_age_group_50_64      -0.721037  7.634338e-03
percent_age_group_65_plus    -0.627414  3.407418e-03
percent_icu                  -0.313119  1.132373e-02
percent_hospitalized          0.301286  2.085459e-03
##########################################################################
```

Good Predictors are those predictors whose P-value < 0.05 :

```
                           Coefficient       P-value
const                         0.220321  6.764585e-10
percent_male                  1.021041  5.065894e-06
percent_age_group_0_17       -0.831788  1.591104e-04
percent_age_group_18_49      -0.726368  1.221966e-04
percent_age_group_50_64      -0.721037  7.634338e-03
percent_age_group_65_plus    -0.627414  3.407418e-03
percent_icu                  -0.313119  1.132373e-02
percent_hospitalized          0.301286  2.085459e-03
```

Bad predictors are those P-value > = 0.05 :

```
                Coefficient   P-value
percent_female     0.095977  0.628683
```

- **correlation matrix to represent the correlation among Proportions:**

```
Part [3] : Correlation Matrix :
                            const  percent_female  percent_male  \
const                         NaN             NaN           NaN
percent_female                NaN        1.000000      0.734727
percent_male                  NaN        0.734727      1.000000
percent_age_group_0_17        NaN        0.484833      0.521147
percent_age_group_18_49       NaN        0.456467      0.888493
percent_age_group_50_64       NaN        0.644491      0.644954
percent_age_group_65_plus     NaN        0.398002     -0.215762
percent_icu                   NaN       -0.467891     -0.389836
percent_hospitalized          NaN       -0.460844     -0.536161

                            percent_age_group_0_17  percent_age_group_18_49  \
const                                          NaN                      NaN
percent_female                            0.484833                 0.456467
percent_male                              0.521147                 0.888493
percent_age_group_0_17                    1.000000                 0.404148
percent_age_group_18_49                   0.404148                 1.000000
percent_age_group_50_64                  -0.124813                 0.414146
percent_age_group_65_plus                -0.322322                -0.533068
percent_icu                              -0.530496                -0.391035
percent_hospitalized                     -0.698677                -0.584881

                            percent_age_group_50_64  percent_age_group_65_plus  \
const                                           NaN                        NaN
percent_female                             0.644491                   0.398002
percent_male                               0.644954                  -0.215762
percent_age_group_0_17                    -0.124813                  -0.322322
percent_age_group_18_49                    0.414146                  -0.533068
percent_age_group_50_64                    1.000000                   0.363822
percent_age_group_65_plus                  0.363822                   1.000000
percent_icu                    1.000000              0.883650
percent_hospitalized           0.883650              1.000000
################################################################################
```

- **Experiment with different ways to improve the fit and interpretability techniques :**

1) summary of the regression without intercept :

```
                            OLS Regression Results
=======================================================================================
Dep. Variable:          percent_deaths   R-squared (uncentered):                   0.742
Model:                             OLS   Adj. R-squared (uncentered):              0.694
Method:                  Least Squares   F-statistic:                              15.46
Date:                 Wed, 22 May 2024   Prob (F-statistic):                    2.06e-10
Time:                         23:47:40   Log-Likelihood:                          126.00
No. Observations:                   51   AIC:                                     -236.0
Df Residuals:                       43   BIC:                                     -220.6
Df Model:                            8
Covariance Type:             nonrobust
=============================================================================================
                              coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------------------
percent_female             -0.4619      0.288     -1.606      0.116      -1.042       0.118
percent_male                0.1475      0.252      0.585      0.562      -0.361       0.656
percent_age_group_0_17      0.1439      0.248      0.581      0.564      -0.355       0.643
percent_age_group_18_49     0.2227      0.192      1.158      0.253      -0.165       0.611
percent_age_group_50_64    -0.0360      0.379     -0.095      0.925      -0.800       0.728
percent_age_group_65_plus   0.1946      0.271      0.717      0.477      -0.353       0.742
percent_icu                -0.3808      0.184     -2.067      0.045      -0.752      -0.009
percent_hospitalized        0.5689      0.134      4.261      0.000       0.300       0.838
==============================================================================
Omnibus:                       41.313   Durbin-Watson:                   1.257
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              173.322
Skew:                           2.047   Prob(JB):                     2.31e-38
Kurtosis:                      11.050   Cond. No.                         158.
==============================================================================

Notes:
[1] R² is computed without centering (uncentered) since the model does not contain a constant.
```

2) summary of the regression with higher order terms :

```
                            OLS Regression Results
==============================================================================
Dep. Variable:         percent_deaths   R-squared:                       0.961
Model:                            OLS   Adj. R-squared:                  0.951
Method:                 Least Squares   F-statistic:                     98.55
Date:                Wed, 22 May 2024   Prob (F-statistic):           6.07e-25
Time:                        23:47:40   Log-Likelihood:                 180.23
No. Observations:                  51   AIC:                            -338.5
Df Residuals:                      40   BIC:                            -317.2
Df Model:                          10
Covariance Type:            nonrobust
=============================================================================================
                                coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------------------
const                         0.2453      0.018     13.421      0.000       0.208       0.282
percent_female                0.1824      0.111      1.648      0.107      -0.041       0.406
percent_male                  0.5415      0.149      3.628      0.001       0.240       0.843
percent_age_group_0_17       -0.6955      0.122     -5.686      0.000      -0.943      -0.448
percent_age_group_18_49      -0.6109      0.098     -6.227      0.000      -0.809      -0.413
percent_age_group_50_64      -0.3572      0.227     -1.574      0.123      -0.816       0.101
percent_age_group_65_plus    -0.6572      0.120     -5.499      0.000      -0.899      -0.416
percent_icu                   1.3081      0.182      7.206      0.000       0.941       1.675
percent_hospitalized         -0.8122      0.167     -4.870      0.000      -1.149      -0.475
percent_icu_squared          -7.7380      0.798     -9.696      0.000      -9.351      -6.125
percent_hospitalized_squared  3.3572      0.458      7.335      0.000       2.432       4.282
==============================================================================
Omnibus:                        2.453   Durbin-Watson:                   1.421
Prob(Omnibus):                  0.293   Jarque-Bera (JB):                1.761
Skew:                          -0.447   Prob(JB):                        0.414
Kurtosis:                       3.175   Cond. No.                     1.06e+03
==============================================================================
```

3) Summary of the regression without outliers :

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          percent_deaths   R-squared:                       0.911
Model:                             OLS   Adj. R-squared:                  0.894
Method:                  Least Squares   F-statistic:                     52.39
Date:                 Wed, 22 May 2024   Prob (F-statistic):           4.31e-19
Time:                         23:47:40   Log-Likelihood:                 161.82
No. Observations:                   50   AIC:                            -305.6
Df Residuals:                       41   BIC:                            -288.4
Df Model:                            8
Covariance Type:             nonrobust
============================================================================================
                               coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------------
const                        0.2429      0.021     11.684      0.000       0.201       0.285
percent_female               0.1321      0.145      0.909      0.369      -0.161       0.426
percent_male                 0.8898      0.146      6.112      0.000       0.596       1.184
percent_age_group_0_17      -0.8050      0.148     -5.446      0.000      -1.104      -0.507
percent_age_group_18_49     -0.7075      0.126     -5.595      0.000      -0.963      -0.452
percent_age_group_50_64     -0.6778      0.190     -3.572      0.001      -1.061      -0.295
percent_age_group_65_plus   -0.6303      0.149     -4.232      0.000      -0.931      -0.329
percent_icu                 -0.1737      0.090     -1.927      0.061      -0.356       0.008
percent_hospitalized         0.1494      0.072      2.069      0.045       0.004       0.295
==============================================================================
Omnibus:                        3.420   Durbin-Watson:                   1.822
Prob(Omnibus):                  0.181   Jarque-Bera (JB):                2.522
Skew:                           0.318   Prob(JB):                        0.283
Kurtosis:                       3.897   Cond. No.                         288.
==============================================================================
```

Part V : Bonus part

1) predict the likelihood of death due to COVID-19 using any/all of the relevant attributes in the COVID case surveillance dataset:

Logistic Regression :

```
Logistic Regression Classification Report
              precision    recall  f1-score   support

           1       1.00      1.00      1.00        11

    accuracy                           1.00        11
   macro avg       1.00      1.00      1.00        11
weighted avg       1.00      1.00      1.00        11

Accuracy: 1.0
##########################################################################
```

Random Forest classifier :

```
Random Forest Classification Report
              precision    recall  f1-score   support

           1       1.00      1.00      1.00        11

    accuracy                           1.00        11
   macro avg       1.00      1.00      1.00        11
weighted avg       1.00      1.00      1.00        11

Accuracy: 1.0
##########################################################################
```

Neural Network :

The network consists of 128 Layers of Relu , 64 of Relu , 1 Sigmoid , then it was modeled using ADAM optimizer , Binary-Crossentropy loss function .

```
1/1 ━━━━━━━━━━━━━━━━━━━━ 1s 1s/step - accuracy: 0.3125 - loss: 0.6956 - val_accuracy: 0.8750 - val_loss: 0.6679
Epoch 2/20
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 152ms/step - accuracy: 1.0000 - loss: 0.6634 - val_accuracy: 0.8750 - val_loss: 0.6444
Epoch 3/20
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 121ms/step - accuracy: 1.0000 - loss: 0.6326 - val_accuracy: 0.8750 - val_loss: 0.6217
Epoch 4/20
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 48ms/step - accuracy: 1.0000 - loss: 0.6031 - val_accuracy: 0.8750 - val_loss: 0.6003
Epoch 5/20
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 50ms/step - accuracy: 1.0000 - loss: 0.5748 - val_accuracy: 0.8750 - val_loss: 0.5806
Epoch 6/20
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 78ms/step - accuracy: 1.0000 - loss: 0.5479 - val_accuracy: 0.8750 - val_loss: 0.5621
Epoch 7/20
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 48ms/step - accuracy: 1.0000 - loss: 0.5222 - val_accuracy: 0.8750 - val_loss: 0.5445
Epoch 8/20
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 48ms/step - accuracy: 1.0000 - loss: 0.4975 - val_accuracy: 0.8750 - val_loss: 0.5272
Epoch 9/20
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 47ms/step - accuracy: 1.0000 - loss: 0.4737 - val_accuracy: 0.8750 - val_loss: 0.5110
Epoch 10/20
```

Conclusion

The exploratory data analysis of COVID-19 impacts has revealed significant disparities and trends across various demographics and sectors. Key findings include the heightened vulnerability of older age groups and individuals with pre-existing conditions, gender and racial disparities in death and employment loss rates, and the severe economic impacts on specific sectors and states. These insights emphasize the need for targeted public health policies and economic support measures to address the identified disparities and support the most affected populations.

- Limitations:
    - The inherent variability in state-level policies and healthcare infrastructure can affect the comparability of the results.
    - Socioeconomic factors and underlying health conditions not captured in the datasets might influence the observed trends.
    - Future research should aim to address these limitations by incorporating more comprehensive and detailed data, enabling a more nuanced understanding of the pandemic's multifaceted impacts.