# Analysis of the impact of coronary disease on COVID-19 mortality in middle-aged people.

Bharath Gopal (TF00410)

## Abstract

The COVID-19 pandemic has brought to light the intricate relationship between the viral infection and cardiovascular complications, underscoring the dire consequences of this deadly combination. COVID-19 is a contagious, viral disease which, most often, causes respiratory issues. They target the lungs and the respiratory system which can affect other parts of the body such as the heart. According to the Harvard Medical School, people with pre-existing cardiac diseases are more likely to die from COVID-19 complications when compared to other who do not have these diseases.

The purpose of this project is to analyze data related to conditions contributing to COVID-19 deaths with a focus on cardiac issues using data analysis methods and study any possible relations between them.

## Introduction

Heart disease is a leading cause of death among the population in the US. A large variety of factors such as cholesterol levels, BMI, smoking history, and other factors can lead to heart disease. According to data from the Mayo Clinic, there is limited data related to heart attacks in young people. A 10-year study by the Framingham Heart Study shows the prevalence of heart attacks in men and women in different age ranges. For men, the prevalence of heart attacks was 12.9 in 1000 for ages 30 to 34 and 28.2 in 1000 for ages 35 to 44. Similarly, for women, the number was 2.2 in 1000 for ages 30 to 34 and 5.2 in 1000 for ages 35 to 44. Women have a higher risk of unrecognized myocardial infarction. It is essential to analyze these factors to identify the risk of heart failure early to provide the required treatment.

The COVID-19 pandemic has introduced the complex relationship between viral infections and cardiac issues. According to Vidal-Perez et al., 2022, cardiac arrhythmias pose a substantial risk to COVID-19 patients. Medication used for treatment of the infection can also cause adverse reactions. While the primary focus on the study of COVID-19 was its effect on the respiratory system, evidence shows effects on the other systems, particularly the cardiovascular system. COVID-19 aggravates existing cardiovascular issues and is frequently complicated by myocardial infarction, or strokes. COVID-19-induced myocarditis is a significant issue, the exact pathogenesis of which is still uncertain. Pre-existing cardiovascular disease is an important risk factor for patients with COVID-19 and commonly results in serious outcomes. Studies also point to signs that COVID-19 results in damage to the cardiovascular system.

Another study investigating the factors influencing COVID-19 outcomes among patients in Brazil analyses various factors and attempts to investigate what some of the most influential factors are and how they can affect the outcome of the patient. The study highlights the impact

of comorbidities such as obesity, diabetes, heart disease, etc. The authors found that the co-occurrence of heart disease and diabetes was strongly associated with COVID-19 mortality. Various Machine Learning models were used to analyze the data along with comorbidity analysis. Some of the ML models that were used included logistic regression, support vector machine, random forest model, and XGBoost. XGBoost performed the best among these models. The findings of the study contribute valuable insights into the factors influencing COVID-19 mortality among hospitalized patients (Passarelli-Araujo, et al., 2022).

A recent study from Germany in 2022 analyzed the impact of the COVID-19 pandemic on mortality rates among patients suffering from heart failure using machine learning models. Data was collected from 86 German Helios hospitals between January 2016 to August 2020 identifying patients with cases of heart failure as their main discharge diagnosis. The ML models developed included random forest, gradient boosting machine, single-layer neural network, and extreme gradient boosting, which were compared to logistic regression out of which the extreme gradient boosting model performed the best. The model calculated the expected mortality rate for 2019 and 2020 and compared them to the observed mortality rate. The study ultimately found no significant increase in overall mortality above the predicted mortality rate with respect to pandemic phases (König et al., 2022).

Previous studies in this field include analysis to predict the risk of heart disease in patients over the age of 40 with a prior diagnosis of prediabetes or diabetes. This study used 5 machine learning models to predict the risk of heart disease. Out of the 5 machine learning models used, the Random Forest model performed the best. Another report used data from the National Health and Nutrition Examination Survey (NHANES) dataset. The study analyzed people aver the age of 18 who completed the health questionnaire and had laboratory and physical exam data. The study uses the machine learning model XGBoost.

## Methods

Dataset: Conditions Contributing to COVID-19 Deaths, by State and Age, Provisional 2020-2023

This dataset contains over 600,000 records with 14 features. The dataset shows health conditions and contributing causes mentioned in conjunction with deaths involving coronavirus disease 2019 (COVID-19) by age group and jurisdiction of occurrence.

License: Public Domain U.S. Government

Source Link: https://www.cdc.gov/nchs/nvss/vsrr/covid_weekly/

The dataset required further cleaning. Since most of the fields in the ‘Condition’ column were irrelevant to the project, the dataset needed to be filtered down further. 8 conditions were selected – Hypertensive diseases, Cerebrovascular diseases, Cardiac Arrest, Ischemic Heart Disease, Cardiac Arrhythmia, Heart Failure, Other Diseases of the Circulatory System and COVID-19. This brought down the size of the dataset from 621,000 rows down to around 200,000 rows. Next the columns, ‘Data As Of’, ‘Flag’ and ‘Number of Mentions’ were removed since they did not provide enough relevant information about the dataset.

A discrepancy that was noticed in this dataset was that the columns 'Start Date', and 'End Date' columns were of type object. These 2 columns were converted to type 'datetime64'. The data in the dataset have be grouped by three different grouping methods – 'By Total', 'By Year', and 'By Month'. This causes redundant data to be repeated. Similarly, the 'State' column contains all the states in the United States as well as the total for the entire country for that specific month and year. Removing the 'United States' field in the 'State' can reduce this redundancy. A sample of data for the state of Maryland was extracted. This subset of data contained 3000 rows of data which can be filtered down further by selecting the relevant grouping method. Further the 'Age Group' column consisted of 10 different ranges: 0-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75-84, 85+, 'Not Stated', and 'All Ages'. This can help us further filter the data for middle-aged and younger populations.

## Results

For visualization, 'seaborn' and 'matplotlib' was used. An initial plot of the number of COVID-19 deaths in 2020, based on the previous condition shows a significant disparity. Patients with Hypertensive Diseases had almost double the number of deaths when compared to the next highest condition which was a tie between Cardiac Arrest and Ischemic Heart Disease. The 'Cerebrovascular disease' condition had the lowest number of COVID-19 deaths in 2020 compared to other circulatory diseases.

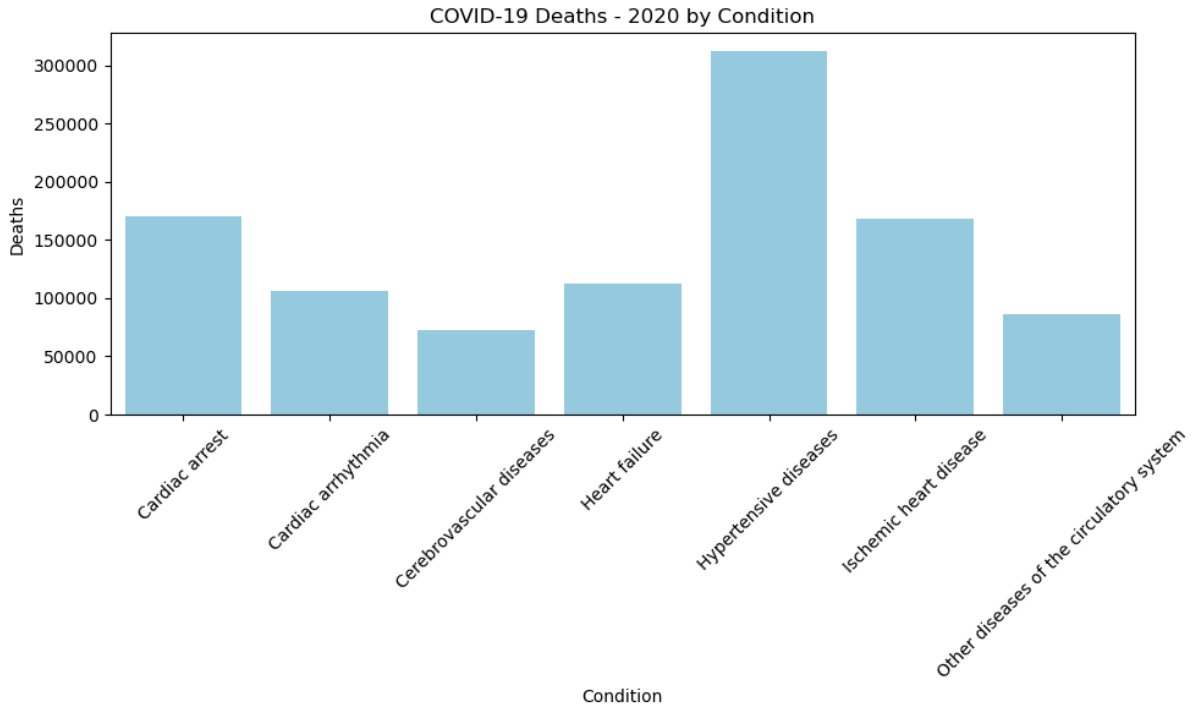


Fig. 1. A bar-plot showcasing the number of COVID-19 deaths in the year 2020 based on the condition.

We go back to the sample of data from the state of Maryland. A plot of COVID-19 deaths in Maryland in 2020 shows a spike April and May, when the pandemic began initially. Another spike is seen in December which could be cause due to people gathering for the holidays or due to the colder winter season. Continuing with the same pattern, a plot of COVID-19 deaths

in Maryland in 2021 shows a spike in January which could again be caused for the same reasons.

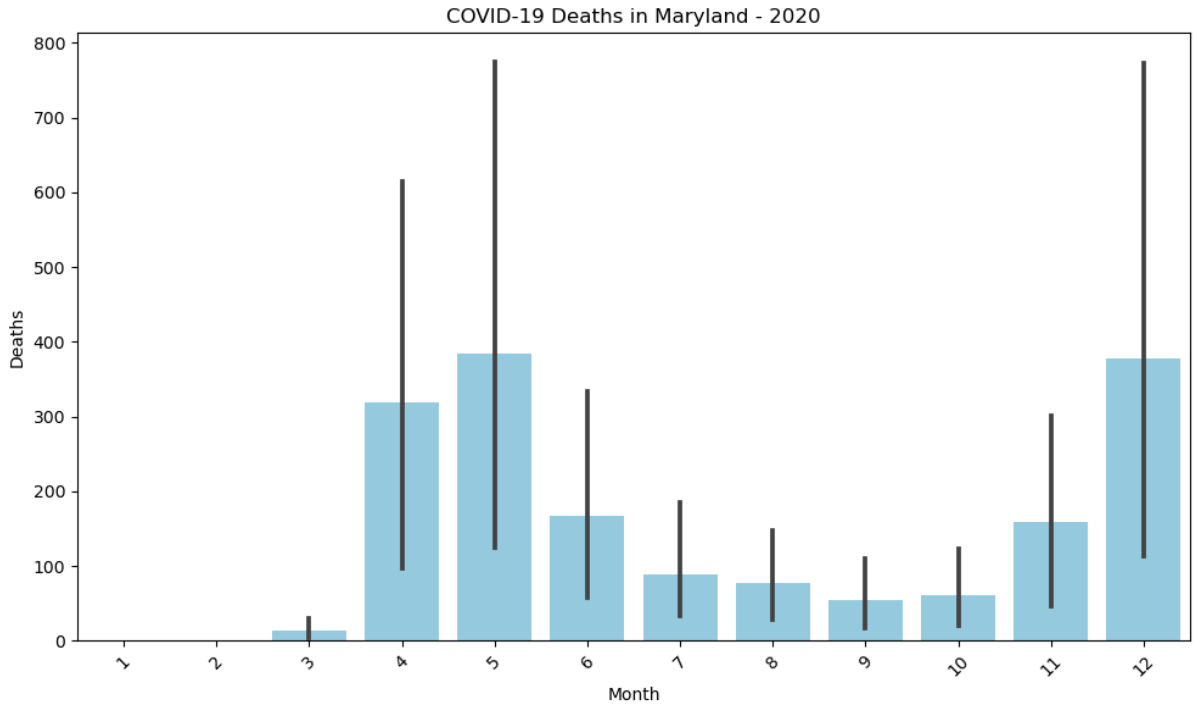


Fig. 2. A bar plot showing the number of COVID-19 deaths in Maryland for the year 2020

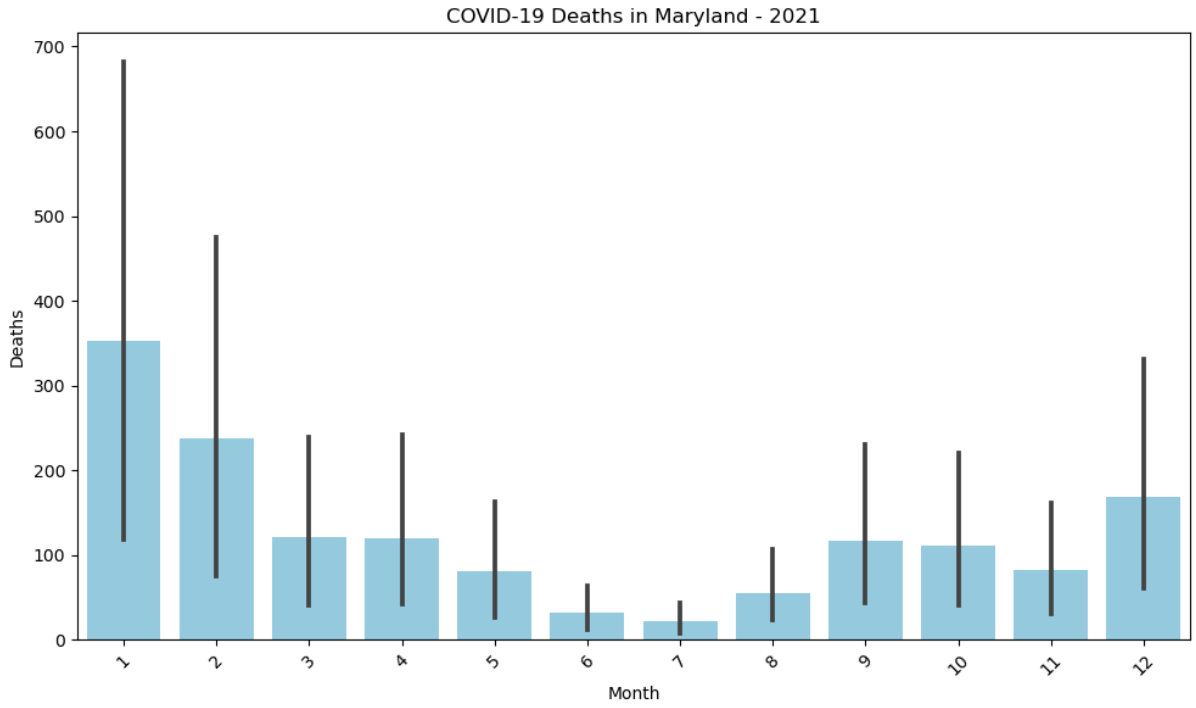


Fig. 3. A bar plot showing the number of COVID-19 deaths in Maryland for the year 2021

An analysis of the age distribution shows a direct correlation between age and the number of deaths. Higher the age range, larger the number of deaths. Sample data from the states of New York and Maryland shows similar patterns. Sample data from California, on the other hand,

has the highest rate of COVID-19 deaths in the 75-84 year range, followed by the 55-64 year range. The highest age range, 85+ years, in California comes third in the COVID-19 death rate plot which is skewed from the national trend.

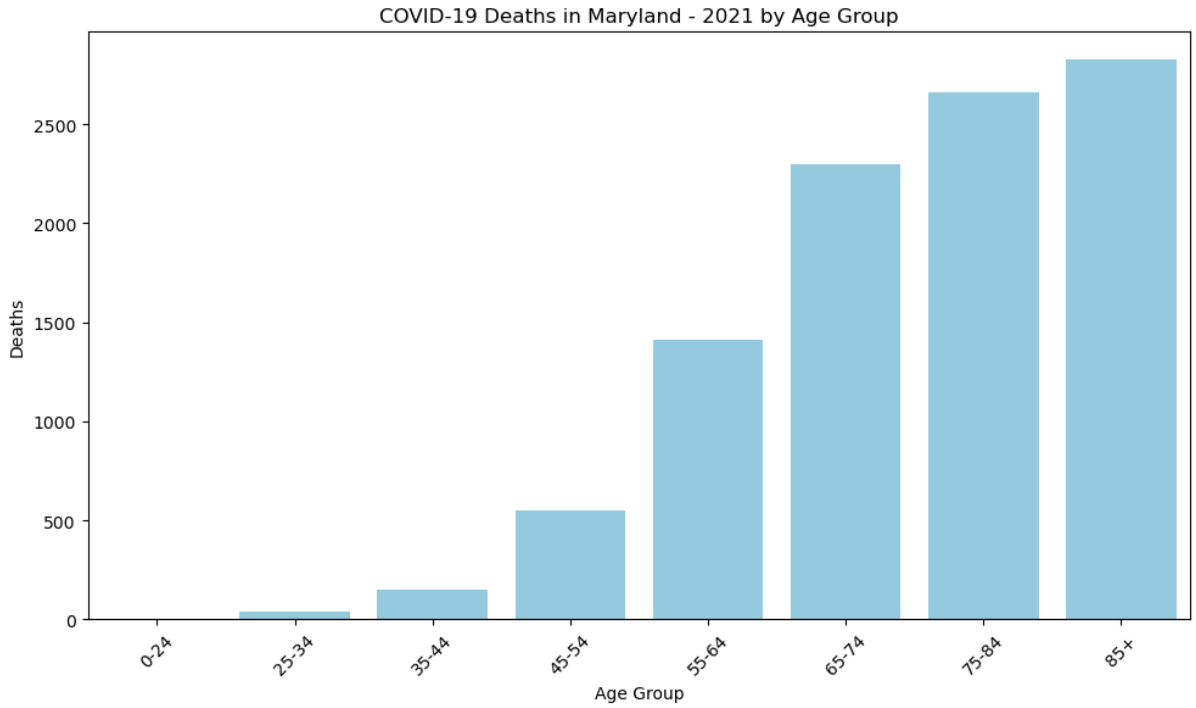


Fig. 4. A bar plot showing the number of COVID-19 deaths in Maryland for the year 2021 based on the age group

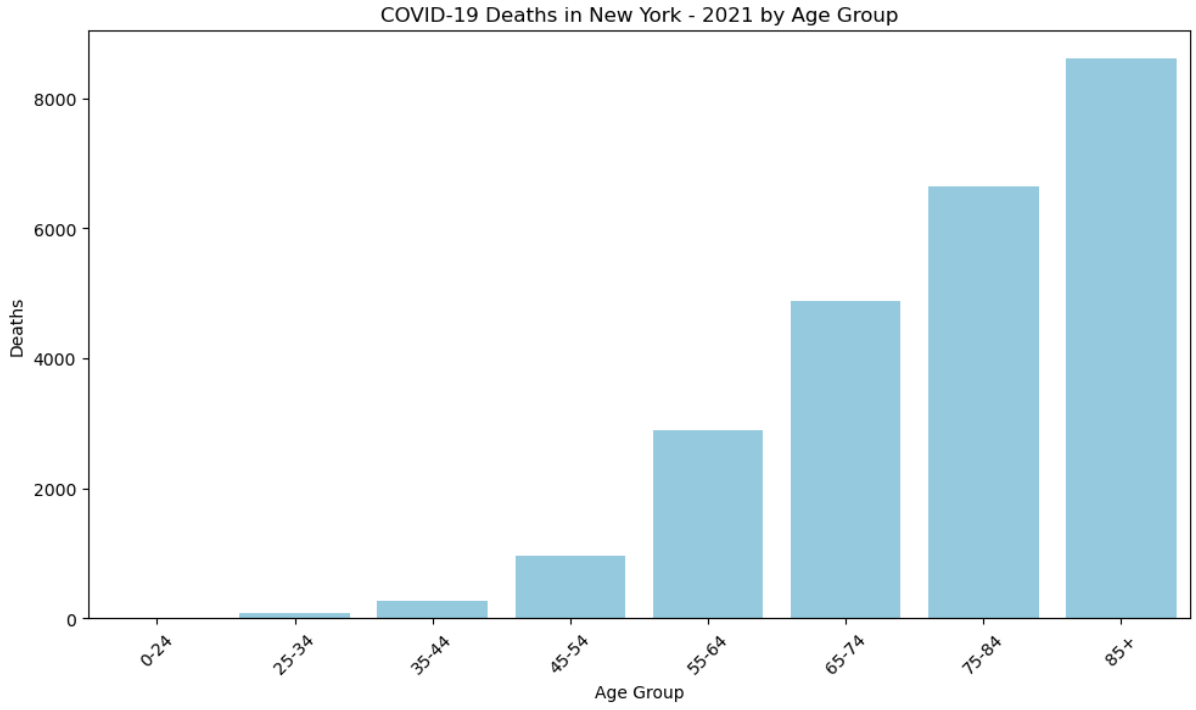


Fig. 5. A bar plot showing the number of COVID-19 deaths in New York for the year 2021 based on the age group

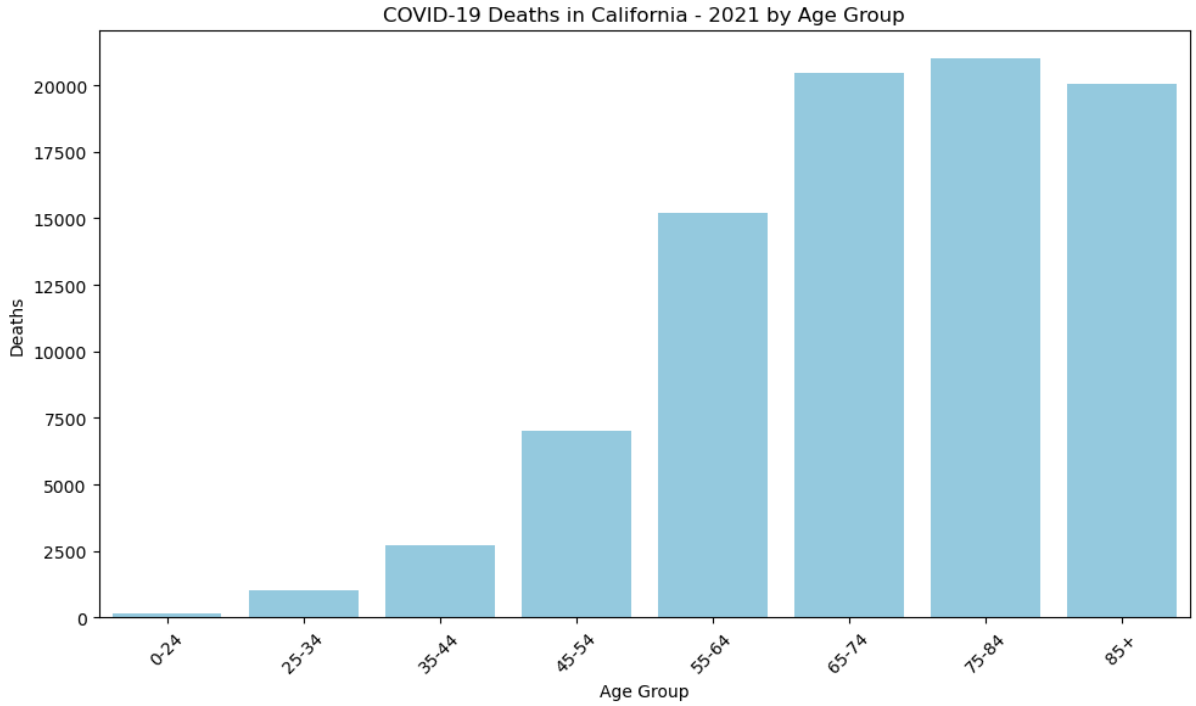


Fig. 6. A bar plot showing the number of COVID-19 deaths in California for the year 2021 based on the age group

A bar plot of the COVID-19 deaths by state accumulates data from the entire dataset and groups it by the state. The result of the plot shows California with the highest number of deaths, followed by Texas while, Vermont and Alaska had the lowest number of deaths. Another plot showcasing the time-period highlights January 2021 with the highest number of COVID-19 deaths through the 3-year period. The lowest since the start of the pandemic was during June of 2023.

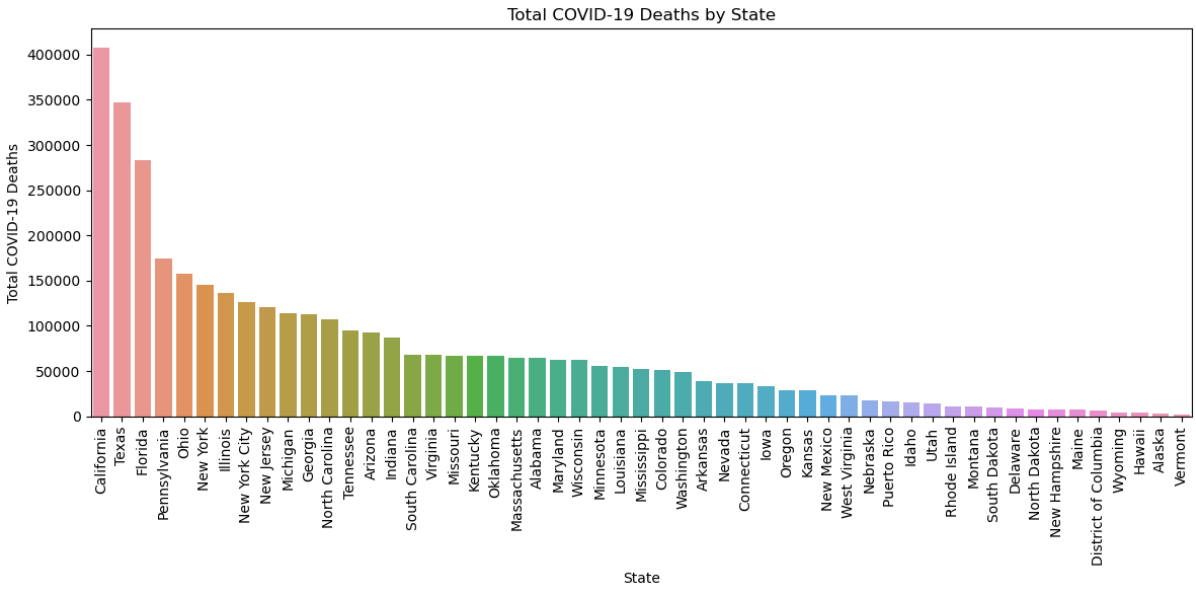


Fig. 7. A bar plot showing the number of COVID-19 deaths according to the state.

Finally, throughout the entire dataset, Hypertensive Diseases have the highest number of COVID-19 deaths, while Cerebrovascular Diseases have the lowest number of COVID-19

deaths. Taking in age as a factor show the 85+ year age range has the highest frequency for each condition except for Cardiac Arrest where the 75-84 year range comes first, followed by the 65-74 year range, and the 85+ year range coming third.

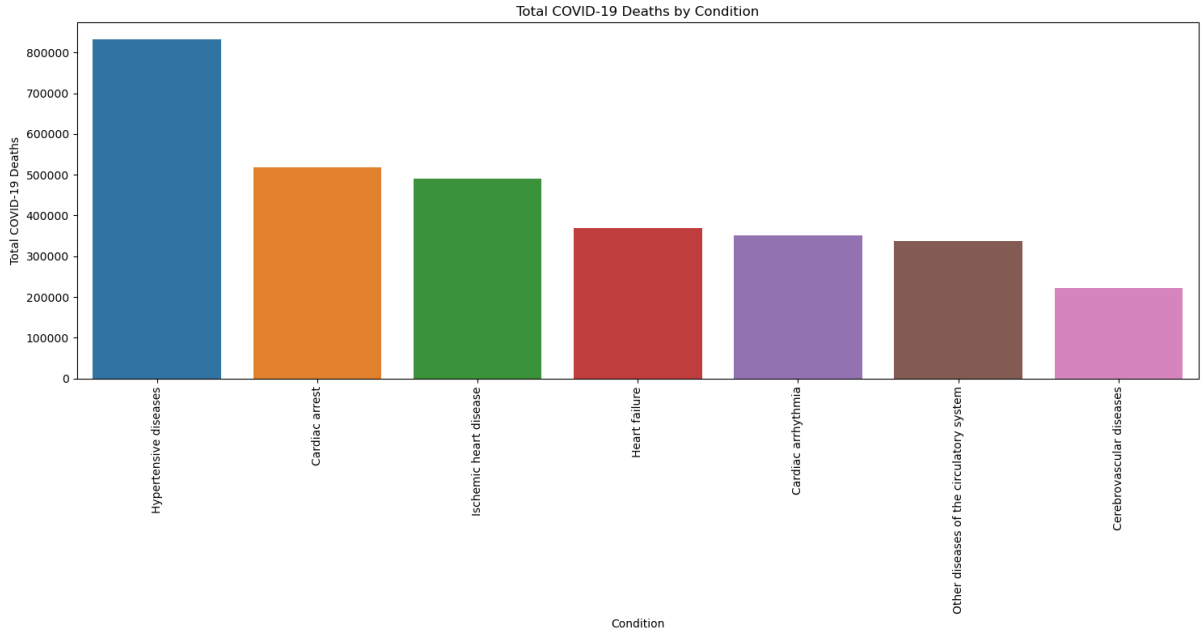


Fig. 8. A bar plot showing the number of COVID-19 deaths according to the condition.

## Related Studies

Various studies have been conducted to understand the association between cardiovascular diseases and COVID-19. A study conducted in 2022 attempted to analyze the effect of cardiovascular disease on patients affected by COVID-19 and assess the risk of heart failure in patients that have recovered from COVID-19. Data was extracted from MEDLINE and Scopus databases and included studies that reported patients that have recovered from COVID-19 and experienced heart failure after their recovery. The study highlighted a 90% increased risk of heart failure within nine months of a COVID-19 infection. Limited analysis due to treatment strategies as well as bias during data collection are some of the potential limitations of this paper. Additionally, the underestimated impact of heart failure and other cardiovascular conditions can affect the quality of the data that was collected. The study concludes that patients that have previously survived COVID-19 are at a higher risk of developing heart failure in the long term, but further studies need to be performed to identify the exact factors that cause the increased risk. (Zuin et al., 2022)

A study from Indonesia, conducted in August of 2021, attempted to answer a similar question. A subset of the population affected by COVID-19 face severe symptoms due to pre-existing conditions. Some of these conditions include chronic kidney disease, cerebrovascular disease, diabetes, etc. The study went through literature from sources such as PubMed, SCOPUS, EuropePMC, Cochrane Central Database and medRxiv and searched for studies with the terms ‘Heart Failure’ and ‘COVID-19’. From a total pool of 204 papers, this study narrowed down the data to 21,640 patients from 18 studies. The majority of the patients from these studies were male and COVID-19 patients with a history of heart failure had a higher risk of hospitalization and increased mortality when compared to patients without a history of heart failure. Some of

the limitations of the study is the existence of bias due to the focus on specific conditions, as well as potential overlap in the data from studies conducted in the same region. The authors conclude by saying that patients with heart failure are at a higher risk of severe outcomes from COVID-19. The meta-analysis also showcases the importance of prioritizing COVID-19 vaccinations for those patients that suffer from heart failure. (Yonas et al., 2022)

A study from China performed analysis on the risk factors associated with cardiovascular disease and COVID-19. Quantitative meta-analysis was conducted on research papers from databases such as PubMed, MedRxiv, Scopus, Elsevier, Web of Science, ScienceDirect, Cochrane Library and Embase. Certain keywords were used to search for relevant studies. Some of these keywords include 'COVID-19', 'coronavirus disease 2019', 'cardiovascular disease', 'heart disease', 'heart failure', etc. For the final selection, a total of 203 studies were selected containing around 24 million patients. These studies span all across the world – 81 from Europe, 54 from North America, 61 from Asia, 2 from Australia, and 5 from different countries. The results of the meta-analysis showed a significant association with pre-existing cardiovascular disease and adverse outcomes in COVID-19 patients. The authors acknowledge certain limitations to the paper. At the time that the paper was published, concrete data on treatment and drug data was not available. This prevented the authors from studying the effects of treatment on patients with a pre-existing cardiovascular disease who have COVID-19. The authors also acknowledge that studies with larger sample sizes can provide additional, more detailed information about this subject. The study concludes that pre-existing cardiovascular disease is an independent risk factor that results in patients with COVID-19 having severe outcomes, usually resulting in hospitalizations and in some cases death. (Xu et al., 2022)

## Limitations and Future Work

The dataset contains a large amount of non-descriptive data, which makes it difficult to create a machine learning model to predict COVID-19 deaths based on the presence of a specific cardiovascular disease. Another limitation is the lack of datasets from different sources which can cause bias in the analysis. A more detailed dataset which provides information about the patients such as their medical history as well as factors such as socio-economic conditions can help provide us with a more robust analysis of the data.

## Conclusion

The COVID-19 pandemic has highlighted the importance of understanding the relation between COVID-19 and cardiovascular diseases. Analysis on the data shows disparities in the number of COVID-19 deaths based on different cardiovascular diseases. The analysis also shows that patients with a history of cardiovascular disease are at a higher risk of hospitalization or even death in some cases, due to COVID-19, when compared to patients with no history of cardiovascular disease. Future research can indicate the exact reason for the relation between COVID-19 and cardiovascular disease as well as suggest possible treatment methods to protect vulnerable patients and improve public health outcomes in the future.

GitHub Link: https://github.com/BGopal1/Data606

## References


Vidal-Perez, R., Brandão, M., Pazdernik, M., Kresoja, P., Carpenito, M., Maeda, S., Casado-Arroyo, R., Muscoli, S., Pöss, J., Fontes-Carvalho, R., & Vazquez-Rodriguez, J. M. (2022). Cardiovascular disease and COVID-19, a deadly combination: A review about direct and indirect impact of a pandemic. World Journal of Clinical Cases, 10(27), 9556-9572. https://doi.org/10.12998/wjcc.v10.i27.9556

Passarelli-Araujo, H., Passarelli-Araujo, H., Urbano, M. R., & Pescim, R. R. (2022). Machine learning and comorbidity network analysis for hospitalized patients with COVID-19 in a city in Southern Brazil. Smart Health (Amsterdam, Netherlands), 26, 100323. https://doi.org/10.1016/j.smhl.2022.100323

König, S., Pellissier, V., Leiner, J., Hohenstein, S., Ueberham, L., Kuhlen, R., Hindricks, G., & Bollmann, A. (2022). Expected and observed in-hospital mortality in heart failure patients before and during the COVID-19 pandemic: Introduction of the machine learning-based standardized mortality ratio at Helios hospitals. Clinical Cardiology, 45(1), 75-82. https://doi.org/10.1002/clc.23762

National Center for Health Statistics. Conditions Contributing to COVID-19 Deaths, by State and Age, Provisional 2020-2023. Date accessed 6 March 2024. Available from https://data.cdc.gov/d/hk9y-quqm.

Wang, Y., Hou, R., Ni, B., Jiang, Y., & Zhang, Y. (2023). Development and validation of a prediction model based on machine learning algorithms for predicting the risk of heart failure in middle-aged and older US people with prediabetes or diabetes. Clinical Cardiology, 46(10), 1234-1243. https://doi.org/10.1002/clc.24104

Cao, J., Zhang, L., Ma, L., Zhou, X., Yang, B., & Wang, W. (2022). Study on the risk of coronary heart disease in middle-aged and young people based on machine learning methods: A retrospective cohort study. PeerJ, 10. https://doi.org/10.7717/peerj.14078

Gulati, R., MD, PhD, Behfar, A., MD, PhD, Narula, J., MD, PhD, Kanwar, A., Lerman, A., MD, Cooper, L., MD, & Singh, M., MD, MPH (2020, January). Acute Myocardial Infarction in Young Individuals. Mayo Clinic Proceedings. https://www.mayoclinicproceedings.org/article/S0025-6196(19)30421-5/fulltext

Zuin, M., Rigatelli, G., Roncon, L., Pasquetto, G., & Bilato, C. Risk of incident heart failure after COVID-19 recovery: A systematic review and meta-analysis. Heart Failure Reviews, 1-6. https://doi.org/10.1007/s10741-022-10292-0

Yonas, E., Alwi, I., Pranata, R., Huang, I., Lim, M. A., Gutierrez, E. J., Yamin, M., Siswanto, B. B., & Virani, S. S. (2021). Effect of heart failure on the outcome of COVID-19 — A meta analysis and systematic review. The American Journal of Emergency Medicine, 46, 204-211. https://doi.org/10.1016/j.ajem.2020.07.009

Xu, J., Xiao, W., Liang, X. et al. A meta-analysis on the risk factors adjusted association between cardiovascular disease and COVID-19 severity. BMC Public Health 21, 1533 (2021). https://doi.org/10.1186/s12889-021-11051-w