# Healthcare Analytics — Analysing Hospitalized Patient Mortality Dataset using MySQL.

PIYUSH CHAUDHARI

# Chapter 1

# Introduction

Hospitalized Patient Mortality, Healthcare data analysis plays a crucial role in understanding patient outcomes, improving care quality, and optimizing resource allocation in medical facilities. In this report, we will delve into a comprehensive analysis of Patient Mortality and healthcare data, focusing on a dataset named " ps_data.csv ".

Using MySQL, we will perform advanced data analysis to address complex questions and provide valuable information for healthcare professionals and administrators.

The analysis presented in this report will not only shed light on patient survival and healthcare utilization patterns but will also highlight opportunities for improving the quality and efficiency of healthcare services. We will examine trends in patient admissions, patient outcomes, and the financial impact on healthcare facilities.

We will use MySQL as our tool for data analysis.

Introduction to SQL:
SQL, or Structured Query Language, is a powerful and standardized programming language used for managing and manipulating relational databases. Developed in the 1970s, SQL has become the de facto standard for working with data stored in a structured format, and it plays a crucial role in modern data management.
SQL is designed to perform a wide range of tasks related to database management, including:

1. Data Retrieval: SQL allows users to retrieve data from a database by using queries to specify what information they need. This is done using the SELECT statement.

2. Data Manipulation: SQL enables users to insert, update, and delete records in a database. The INSERT, UPDATE, and DELETE statements are used for these tasks.

3. Data Definition: SQL is used to define the structure of a database, including creating tables, defining their structure, setting constraints, and establishing relationships between tables.

4. Data Control: SQL provides mechanisms for controlling access to data, ensuring security and maintaining data integrity through permission management.

SQL is known for its simplicity, versatility, and portability. It is not tied to a specific database management system (DBMS), which means you can use SQL with various relational database systems such as MySQL, PostgreSQL, SQL Server, Oracle, and more.

Whether you are a database administrator, a data analyst, a software developer, or someone working with data in any capacity, SQL is an essential tool for effectively managing and extracting information from relational databases. It provides a common language for interacting with data, making it a fundamental skill in the world of data management and analysis**.**

# Chapter 2

# Overview

**Dataset used:**

The source of the data for this project is Kaggle, a platform that hosts various datasets for data science and machine learning. The data consists of web-scraped information from ESPN, a sports media company. The data is available on Kaggle as a CSV file.

https://www.kaggle.com/datasets/mitishaagarwal/patient

**Dataset (Kaggle):** Patient Survival

**Software Used:** MS SQL

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | encounter_id | patient_id | hospital_id | age | bmi | elective_surgery | ethnicity | gender | height | icu_admit_source | icu_id | icu_stay_type | icu_type | pre_icu_los_days | weight |
| 2 | 66154 | 25312 | 118 | 68 | 22.73 | 0 | Caucasian | M | 180.3 | Floor | 92 | admit | CTICU | 0.541666667 | 73.9 |
| 3 | 114252 | 59342 | 81 | 77 | 27.42 | 0 | Caucasian | F | 160 | Floor | 90 | admit | Med-Surg ICU | 0.927777778 | 70.2 |
| 4 | 119783 | 50777 | 118 | 25 | 31.95 | 0 | Caucasian | F | 172.7 | Accident & Emergency | 93 | admit | Med-Surg ICU | 0.000694444 | 95.3 |
| 5 | 79267 | 46918 | 118 | 81 | 22.64 | 1 | Caucasian | F | 165.1 | Operating Room / Recovery | 92 | admit | CTICU | 0.000694444 | 61.7 |
| 6 | 92056 | 34377 | 33 | 19 | | 0 | Caucasian | M | 188 | Accident & Emergency | 91 | admit | Med-Surg ICU | 0.073611111 | |
| 7 | 33181 | 74489 | 83 | 67 | 27.56 | 0 | Caucasian | M | 190.5 | Accident & Emergency | 95 | admit | Med-Surg ICU | 0.000694444 | 100 |
| 8 | 82208 | 49526 | 83 | 59 | 57.45 | 0 | Caucasian | F | 165.1 | Accident & Emergency | 95 | admit | Med-Surg ICU | 0.000694444 | 156.6 |
| 9 | 120995 | 50129 | 33 | 70 | | 0 | Caucasian | M | 165 | Accident & Emergency | 91 | admit | Med-Surg ICU | 0.002083333 | |
| 10 | 80471 | 10577 | 118 | 45 | | 0 | Caucasian | M | 170.2 | Other Hospital | 114 | admit | CCU-CTICU | 0.009027778 | |
| 11 | 42871 | 90749 | 118 | 50 | 25.71 | 0 | | M | 175.3 | Accident & Emergency | 114 | admit | CCU-CTICU | 0.060416667 | 79 |
| 12 | 105427 | 125898 | 77 | 72 | 28.25705249 | 1 | Hispanic | F | 154.9 | Operating Room / Recovery | 113 | admit | Med-Surg ICU | 0.004861111 | 67.8 |
| 13 | 91609 | 78266 | 83 | 80 | 27.3828125 | 1 | Caucasian | F | 160 | Operating Room / Recovery | 95 | admit | Med-Surg ICU | 1.271527778 | 70.1 |
| 14 | 76356 | 41311 | 118 | 48 | | 0 | Caucasian | M | 175.3 | Accident & Emergency | 102 | admit | Neuro ICU | 0.020833333 | |
| 15 | 34030 | 103766 | 118 | 65 | | 1 | Caucasian | M | 172.7 | Operating Room / Recovery | 93 | admit | Med-Surg ICU | 1.045138889 | |
| 16 | 108388 | 98174 | 118 | 81 | 38.18906706 | 1 | Caucasian | M | 172.7 | Operating Room / Recovery | 97 | admit | MICU | 1.376388889 | 113.9 |
| 17 | 120677 | 124688 | 118 | 78 | | 0 | Caucasian | F | 160 | Accident & Emergency | 97 | admit | MICU | 0.000694444 | |
| 18 | 115771 | 71252 | 81 | 30 | 23.38317807 | 0 | Caucasian | M | 193 | Accident & Emergency | 90 | admit | Med-Surg ICU | 0.240972222 | 87.1 |
| 19 | 22471 | 112115 | 118 | 46 | 25.84571744 | 0 | Hispanic | M | 167.6 | Accident & Emergency | 92 | admit | CTICU | 0 | 72.6 |
| 20 | 48056 | 114220 | 118 | 65 | 28.40892909 | 0 | Hispanic | M | 167.6 | Accident & Emergency | 100 | admit | Neuro ICU | 0 | 79.8 |
| 21 | 45954 | 102946 | 118 | 39 | 37.0102781 | 1 | | M | 177.8 | Operating Room / Recovery | 92 | admit | CTICU | 5.163888889 | 117 |
| 22 | 79115 | 88295 | 83 | 79 | 28.37610001 | 1 | Caucasian | F | 170.2 | Operating Room / Recovery | 95 | admit | Med-Surg ICU | 0.89375 | 82.2 |
| 23 | 118620 | 59115 | 118 | 80 | 35.65998942 | 0 | Caucasian | M | 170.2 | Floor | 97 | admit | MICU | 0.000694444 | 103.3 |
| 24 | 21050 | 31257 | 81 | 58 | | 1 | Caucasian | F | 162.6 | Operating Room / Recovery | 90 | admit | Med-Surg ICU | 0.003472222 | |
| 25 | 95460 | 120539 | 118 | 87 | 21.96376348 | 0 | Caucasian | M | 180.3 | Floor | 97 | admit | MICU | 5.046527778 | 71.4 |
| 26 | 7220 | 92453 | 77 | 60 | 29.50995926 | 0 | Caucasian | M | 188 | Accident & Emergency | 113 | admit | Med-Surg ICU | 0.08125 | 104.3 |

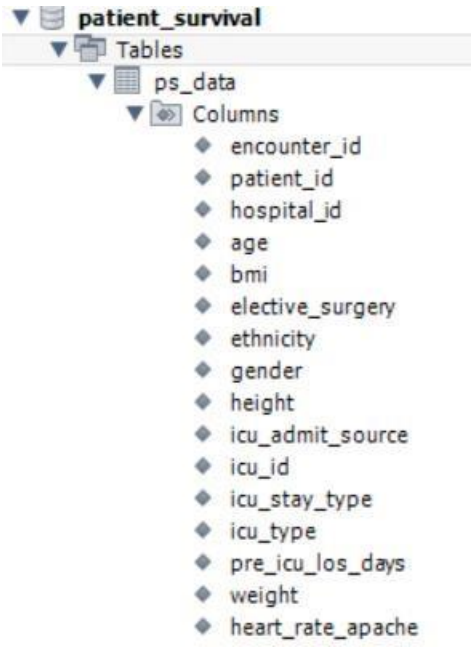
# Chapter 3

# Methodology

**Database Setup:**
We've ps_data.csv files in this dataset. I have framed some questions on this data that I will solve using SQL queries and find answers to them.

First let us create a database by clicking on 'Create a new schema' option. I have created a DB called Patient Survival. We will then use the created database by writing *Patient Survival.*

Then we will import all the ps_data.csv files into tables by right clicking on the Patient Survival database and then on the Table Data Import Wizard ps_data.csv. We've to do this step for all the .csv files. As we've the data in our tables, we will go ahead and answer the questions.




My DB Setup

# Chapter 4

# Implementation

**Analyzing the match results:**

The Analyzing Hospitalized Patient Mortality or Patient Survival results are in the ps_data.csv file that is loaded into the results table. I will first fix the datatype of the date column in results table using the below queries.

```
SELECT * FROM `patient_survival`.`ps_data`;
```

| encounter_id | patient_id | hospital_id | age | bmi | elective_surgery | ethnicity | gender | height | icu_admit_source | icu_id |
|---|---|---|---|---|---|---|---|---|---|---|
| 66154 | 25312 | 118 | 68 | 22.73 | 0 | Caucasian | M | 180.3 | Floor | 92 |
| 114252 | 59342 | 81 | 77 | 27.42 | 0 | Caucasian | F | 160 | Floor | 90 |
| 119783 | 50777 | 118 | 25 | 31.95 | 0 | Caucasian | F | 172.7 | Accident & Emergency | 93 |
| 79267 | 46918 | 118 | 81 | 22.64 | 1 | Caucasian | F | 165.1 | Operating Room / Recovery | 92 |
| 92056 | 34377 | 33 | 19 | | 0 | Caucasian | M | 188 | Accident & Emergency | 91 |
| 33181 | 74489 | 83 | 67 | 27.56 | 0 | Caucasian | M | 190.5 | Accident & Emergency | 95 |
| 82208 | 49526 | 83 | 59 | 57.45 | 0 | Caucasian | F | 165.1 | Accident & Emergency | 95 |
| 120995 | 50129 | 33 | 70 | | 0 | Caucasian | M | 165 | Accident & Emergency | 91 |
| 80471 | 10577 | 118 | 45 | | 0 | Caucasian | M | 170.2 | Other Hospital | 114 |

Q1: How many total deaths occured in the hospital and what was the percentage of the mortality rate?

```
SELECT COUNT(CASE WHEN hospital_death = 1 THEN 1 END) AS total_hospital_deaths,
ROUND(COUNT(CASE WHEN hospital_death = 1 THEN 1 END)*100/COUNT(*),2) AS mortality_rate
FROM `patient_survival`.`ps_data`;
```

| total_hospital_deaths | mortality_rate |
|---|---|
| 110 | 5.88 |

Q2: What was the death count of each ethnicity?

```
SELECT ethnicity, COUNT(hospital_death) as total_hospital_deaths
FROM `patient_survival`.`ps_data`
WHERE hospital_death = '1'
GROUP BY ethnicity
```

| | ethnicity | total_hospital_deaths |
|---|---|---|
| ▶ | Caucasian | 78 |
| | African American | 18 |
| | Hispanic | 5 |
| | | 3 |
| | Native American | 2 |
| | Asian | 4 |

Q3: What was the death count of each gender?

```
SELECT gender, COUNT(hospital_death) as total_hospital_deaths
FROM `patient_survival`.`ps_data`
WHERE hospital_death = '1'
GROUP BY gender
```

| | gender | total_hospital_deaths |
|---|---|---|
| ▶ | M | 54 |
| | F | 56 |

Q4: Comparing the average and max ages of patients who died and patients who survived.

```
SELECT ROUND(AVG(age),2) as avg_age,
MAX(age) as max_age,  hospital_death
FROM `patient_survival`.`ps_data`
WHERE hospital_death = '1'
GROUP BY hospital_death
UNION
SELECT ROUND(AVG(age),2) as avg_age,
MAX(age) as max_age,  hospital_death
FROM `patient_survival`.`ps_data`
WHERE hospital_death = '0'
GROUP BY hospital_death
```

| | avg_age | max_age | hospital_death |
|---|---|---|---|
| ▶ | 66.66 | 89 | 1 |
| | 62.41 | 89 | 0 |

Q5: Comparing the amount of patients that died and survived by each age

```
SELECT age,
    COUNT(CASE WHEN hospital_death = '1' THEN 1 END) as amount_that_died,
    COUNT(CASE WHEN hospital_death = '0' THEN 1 END) as amount_that_survived
FROM `patient_survival`.`ps_data`
GROUP BY age
ORDER BY age ASC;
```

| | age | amount_that_died | amount_that_survived |
|---|---|---|---|
| ▶ | 17 | 0 | 1 |
| | 18 | 0 | 6 |
| | 19 | 0 | 9 |
| | 20 | 0 | 4 |
| | 21 | 0 | 6 |
| | 22 | 0 | 3 |
| | 23 | 1 | 7 |
| | 24 | 0 | 5 |
| | 25 | 0 | 9 |
| | 26 | 1 | 8 |

| | age | amount_that_died | amount_that_survived |
|---|---|---|---|
| | 81 | 2 | 30 |
| | 82 | 0 | 30 |
| | 83 | 5 | 31 |
| | 84 | 4 | 39 |
| | 85 | 2 | 28 |
| | 86 | 2 | 22 |
| | 87 | 0 | 30 |
| | 88 | 1 | 19 |
| | 89 | 1 | 13 |

Q6: Which country had the highest win percentage against Australia?

```
SELECT
   CONCAT(FLOOR(age/10)*10, '-', FLOOR(age/10)*10+9) AS age_interval,
   COUNT(*) AS patient_count
FROM `patient_survival`.`ps_data`
GROUP BY age_interval
ORDER BY age_interval;
```

| | age_interval | patient_count |
|---|---|---|
| ▶ | 10-19 | 16 |
| | 20-29 | 67 |
| | 30-39 | 104 |
| | 40-49 | 170 |
| | 50-59 | 367 |
| | 60-69 | 447 |
| | 70-79 | 401 |
| | 80-89 | 299 |

Q7: Amount of patients above 65 who died vs Amount of patients between 50-65 who died

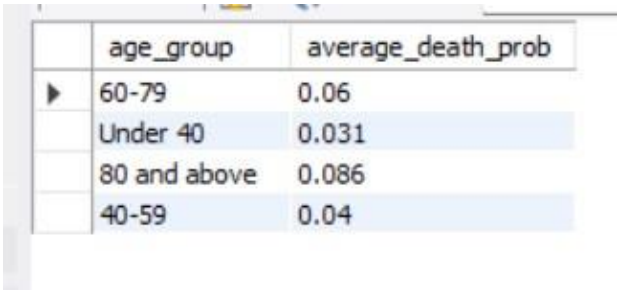
```
SELECT COUNT(CASE WHEN age > 65 AND hospital_death = '0' THEN 1 END) as survived_over_65,
    COUNT(CASE WHEN age BETWEEN 50 AND 65 AND hospital_death = '0' THEN 1 END) as survived_between_50_and_65,
    COUNT(CASE WHEN age > 65 AND hospital_death = '1' THEN 1 END) as died_over_65,
    COUNT(CASE WHEN age BETWEEN 50 AND 65 AND hospital_death = '1' THEN 1 END) as died_between_50_and_65
FROM `patient_survival`.`ps_data`;
```

| | survived_over_65 | survived_between_50_and_65 | died_over_65 | died_between_50_and_65 |
|---|---|---|---|---|
| ▶ | 833 | 583 | 62 | 36 |

Q8: Calculating the average probability of hospital death for patients of different age groups

```
SELECT
   CASE
      WHEN age < 40 THEN 'Under 40'
      WHEN age >= 40 AND age < 60 THEN '40-59'
      WHEN age >= 60 AND age < 80 THEN '60-79'
```

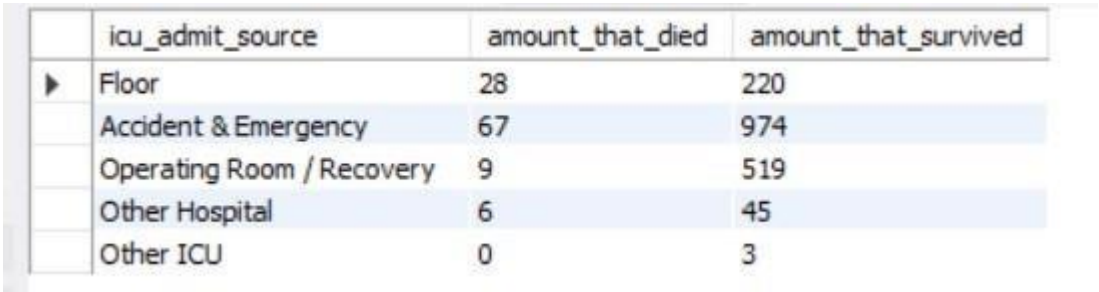
```
        ELSE '80 and above'
    END AS age_group,
    ROUND(AVG(apache_4a_hospital_death_prob),3) AS average_death_prob
FROM `patient_survival`.`ps_data`
GROUP BY age_group;
```



| age_group | average_death_prob |
|---|---|
| 60-79 | 0.06 |
| Under 40 | 0.031 |
| 80 and above | 0.086 |
| 40-59 | 0.04 |

Q9: Which admit source of the ICU did most patients die in and get admitted to?

```
SELECT DISTINCT icu_admit_source,
        COUNT(CASE WHEN hospital_death = '1' THEN 1 END) as amount_that_died,
        COUNT(CASE WHEN hospital_death = '0' THEN 1 END) as amount_that_survived
FROM `patient_survival`.`ps_data`
GROUPBY icu_admit_source;
```



| icu_admit_source | amount_that_died | amount_that_survived |
|---|---|---|
| Floor | 28 | 220 |
| Accident & Emergency | 67 | 974 |
| Operating Room / Recovery | 9 | 519 |
| Other Hospital | 6 | 45 |
| Other ICU | 0 | 3 |

Q10: Average age of people in each ICU admit source and amount that died.

```
SELECT DISTINCT icu_admit_source,
        COUNT(hospital_death) as amount_that_died,
        ROUND(AVG(age),2) as avg_age
FROM `patient_survival`.`ps_data`
WHERE hospital_death = '1'
GROUP BY icu_admit_source
```

| | icu_admit_source | amount_that_died | avg_age |
|---|---|---|---|
| ▶ | Other Hospital | 6 | 56.83 |
| | Floor | 28 | 71.79 |
| | Operating Room / Recovery | 9 | 70.78 |
| | Accident & Emergency | 67 | 64.85 |

Q11: Average age of people in each type of ICU and amount that died.

```
SELECT DISTINCT icu_type,
        COUNT(hospital_death) as amount_that_died,
        ROUND(AVG(age),2) as avg_age
FROM `patient_survival`.`ps_data`
WHERE hospital_death = '1'
GROUP BY icu_type;
```

| | icu_type | amount_that_died | avg_age |
|---|---|---|---|
| ▶ | CCU-CTICU | 17 | 65.12 |
| | Neuro ICU | 13 | 68.54 |
| | Med-Surg ICU | 46 | 67.39 |
| | MICU | 24 | 63.63 |
| | SICU | 5 | 74.40 |
| | CTICU | 5 | 67.20 |

Q12: How many patients are suffering from each comorbidity?

```
SELECT
    SUM(aids) AS patients_with_aids,
    SUM(cirrhosis) AS patients_with_cirrhosis,
    SUM(diabetes_mellitus) AS patients_with_diabetes,
    SUM(hepatic_failure) AS patients_with_hepatic_failure,
    SUM(immunosuppression) AS patients_with_immunosuppression,
    SUM(leukemia) AS patients_with_leukemia,
    SUM(lymphoma) AS patients_with_lymphoma,
    SUM(solid_tumor_with_metastasis) AS patients_with_solid_tumor
FROM `patient_survival`.`ps_data`;
```

| | patients_with_aids | patients_with_cirrhosis | patients_with_diabetes | patients_with_hepatic_failure | patients_with_immunosuppression | patients_with_le |
|---|---|---|---|---|---|---|
| ▶ | 0 | 39 | 480 | 29 | 115 | 15 |

Q13: What is the mortality rate in percentage?

```
SELECT
    COUNT(CASE WHEN hospital_death = 1 THEN 1 END)*100/COUNT(*) AS mortality_rate
FROM `patient_survival`.`ps_data`;
```

| | mortality_rate |
|---|---|
| ▶ | 5.8792 |

Q14: What was the percentage of patients who underwent elective surgery?

```
SELECT
    COUNT(CASE WHEN elective_surgery = 1 THEN 1 END)*100/COUNT(*) AS elective_surgery_percentage
FROM patient_survival.ps_data;
```

| | elective_surgery_percentage |
|---|---|
| ▶ | 25.3340 |

Q15: What was the average weight and height for male & female patients who underwent elective surgery?

```
SELECT
    ROUND(AVG(CASE WHEN gender = 'M' THEN weight END),2) AS avg_weight_male,
    ROUND(AVG(CASE WHEN gender = 'M' THEN height END),2) AS avg_height_male,
    ROUND(AVG(CASE WHEN gender = 'F' THEN weight END),2) AS avg_weight_female,
    ROUND(AVG(CASE WHEN gender = 'F' THEN height END),2) AS avg_height_female
FROM patient_survival.ps_data
WHERE elective_surgery = 1;
```

| avg_weight_male | avg_height_male | avg_weight_female | avg_height_female |
|---|---|---|---|
| 81.78 | 176.93 | 67.54 | 162.37 |

Q16: What were the top 10 ICUs with the highest hospital death probability?

```
SELECT icu_id,
        apache_4a_hospital_death_prob as hospital_death_prob
FROM`patient_survival`.`ps_data`
ORDER BY apache_4a_hospital_death_prob DESC
LIMIT 10;
```

| icu_id | hospital_death_prob |
|---|---|
| 98 | 0.97 |
| 92 | 0.96 |
| 95 | 0.94 |
| 103 | 0.92 |
| 91 | 0.92 |
| 98 | 0.91 |
| 85 | 0.86 |
| 108 | 0.85 |
| 97 | 0.83 |
| 114 | 0.82 |

Q17: What was the average length of stay at each ICU for patients who survived and those who didn't?

```
SELECT
   icu_type,
   ROUND(AVG(CASE WHEN hospital_death = 1 THEN pre_icu_los_days END), 2) AS avg_icu_stay_death,
   ROUND(AVG(CASE WHEN hospital_death = 0 THEN pre_icu_los_days END), 2) AS avg_icu_stay_survived
FROM `patient_survival`.`ps_data`
GROUP BY icu_type
ORDER BY icu_type;
```

| | icu_type | avg_icu_stay_death | avg_icu_stay_survived |
|---|---|---|---|
| ▶ | CCU-CTICU | 2.02 | 0.34 |
| | CTICU | 0.34 | 1.41 |
| | Med-Surg ICU | 0.67 | 0.43 |
| | MICU | 1.62 | 0.89 |
| | Neuro ICU | 1.79 | 0.6 |
| | SICU | 1.52 | 0.41 |

Q18: What was the average BMI for patients that died based on ethnicity? (excluding missing or null values)

```
SELECT
    ethnicity,
    ROUND(AVG(bmi),2) AS average_bmi
FROM `patient_survival`.`ps_data`
WHERE bmi IS NOT NULL
AND hospital_death = '1'
GROUP BY ethnicity;
```

| | ethnicity | average_bmi |
|---|---|---|
| ▶ | Caucasian | 22.67 |
| | African American | 22.48 |
| | Hispanic | 22.63 |
| | | 27.24 |
| | Native American | 23.47 |
| | Asian | 28.73 |

Q19: What was the death percentage for each ethnicity?

```
SELECT
    ethnicity,
    ROUND(COUNT(CASE WHEN hospital_death = 1 THEN 1 END) * 100 / (SELECT COUNT(*) FROM patient_survival.ps_data), 2) AS death_percentage
FROM `patient_survival`.`ps_data`
GROUP BY ethnicity;
```

| | ethnicity | death_percentage |
|---|---|---|
| ▶ | Caucasian | 4.17 |
| | | 0.16 |
| | Hispanic | 0.27 |
| | African American | 0.96 |
| | Asian | 0.21 |
| | Native American | 0.11 |
| | Other/Unknown | 0.00 |

# Chapter 5

# Conclusion

The Patient Mortality dataset has provided valuable insights into patient Mortality or survival and healthcare dynamics. We have answered a series of complex questions, addressing various aspects of patient demographics, and healthcare performance. the data contains related to patients admitted to a hospital.

By analyze patient mortality rates and identify factors that may contribute to patient deaths. By examining the data, we can gain insights into the prevalence of different medical conditions among patients who died, the effectiveness of different types of care, and the impact of demographic factors on patient outcomes.

Overall, the ps_data.csv file provides a valuable resource for researchers and healthcare professionals looking to better understand patient mortality rates and improve patient outcomes. By analyzing this data, we can identify areas for improvement in patient care and develop strategies to reduce patient mortality rates.