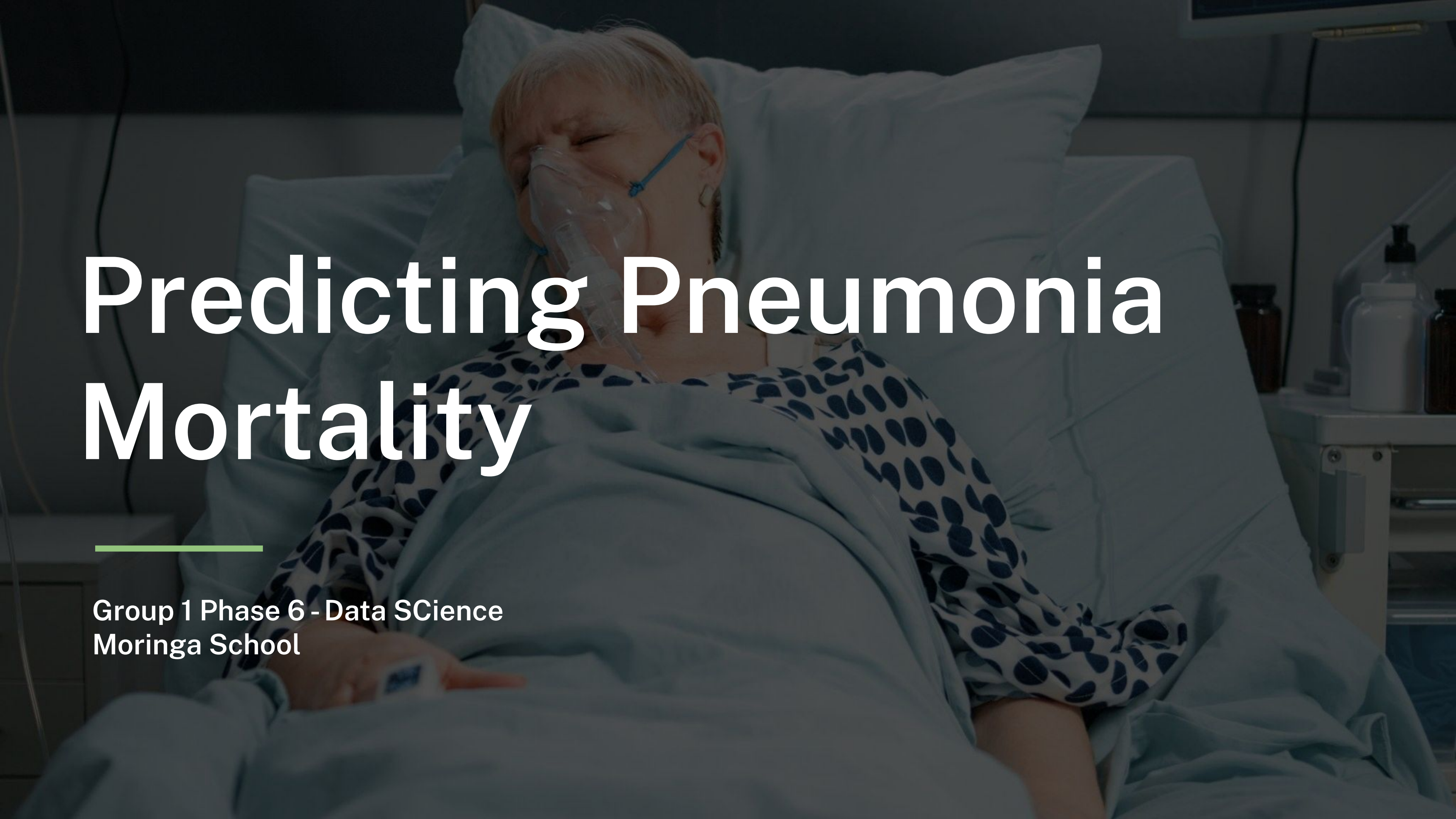


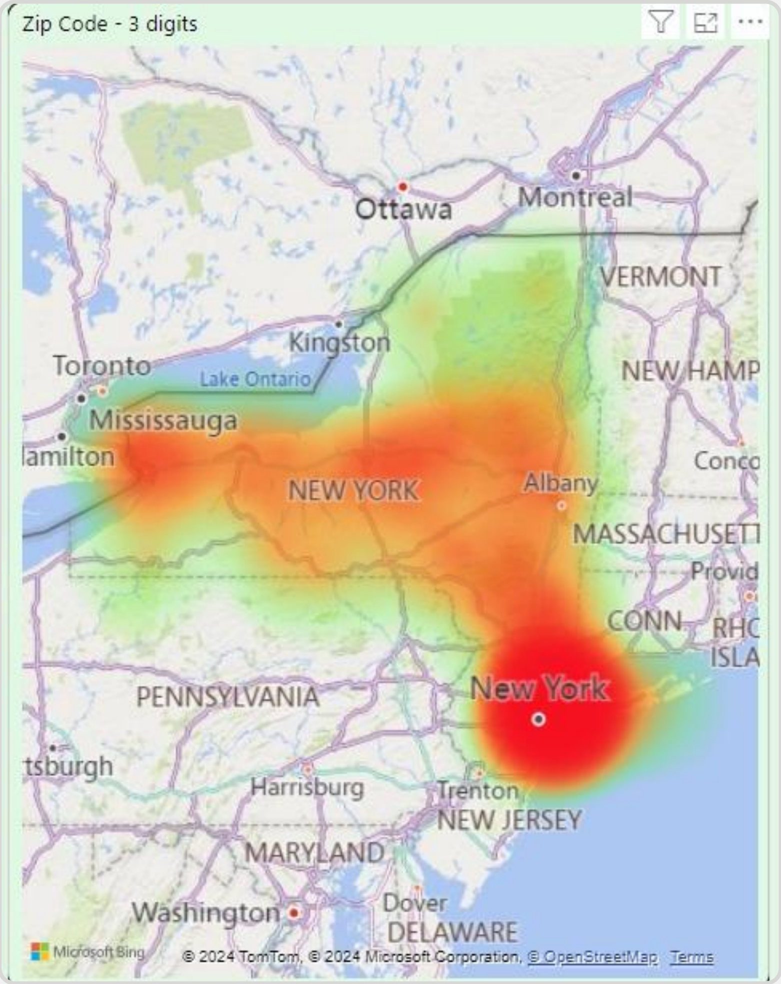
# Predicting Pneumonia Mortality

Group 1 Phase 6 - Data SCience
Moringa School

## Background

Pneumonia remains a major global health issue, leading to **millions of hospitalizations** and **significant mortality rates**, particularly among children under five and the elderly.

- According to the World Health Organization, **pneumonia accounts for 15% of all deaths** of children under five years old, killing approximately **800,000 children annually**.
- In Africa, the situation is even more severe, with sub-Saharan Africa accounting for more than half of the global pneumonia deaths in children under five. **Limited healthcare access, low vaccination coverage,** and high HIV/AIDS prevalence contribute to the high mortality rates.
- Developing predictive models for in-hospital mortality can help in the **early identification of high-risk patients**, leading to timely interventions, better resource allocation, and improved clinical outcomes.




# Problem statement

- Healthcare providers and administrators in New York State aim to **improve patient outcomes and reduce in-hospital mortality** among pneumonia patients.
- In-hospital mortality occurs when patients succumb to their condition during their hospital stay, and healthcare providers must predict which **patients are at the highest risk** of mortality
- By identifying these high-risk patients in advance, healthcare providers in New York can **implement targeted interventions** and **allocate resources more effectively** to improve patient care and survival rates.

## Objectives

| | |
|---|---|
| **Identify High-Risk Populations** | Determine demographic and clinical factors associated with higher pneumonia hospitalization rates. |
| **Evaluate Quality of Care** | Assess metrics such as length of stay, patient disposition, and mortality rates to gauge the quality of pneumonia care. |
| **Assess Economic Burden** | Investigate the costs associated with pneumonia hospitalizations and identify opportunities for cost reduction. |
| **Reduce Readmissions** | Identify factors contributing to pneumonia-related readmissions and suggest strategies to minimize them. |

# About the project

This project focuses on the **healthcare industry**, specifically epidemiology and public health, aiming to **provide insights for healthcare providers, public health officials, policymakers, and researchers.**

It aims to help us understand the **factors influencing pneumonia** in order to assist the said groups develop **targeted interventions** to reduce its impact.

The project also aims to formulate insights on how lead to **better-targeted prevention strategies,** more efficient resource allocation, and improved patient outcomes, ultimately r**educing pneumonia-related hospitalizations and deaths.**

**The project uses data from the Statewide Planning and Research Cooperative System(SPARCS) spanning between 2009 and 2016

## Key Features

**Patient Characteristics**

Information about patients such as age, gender, race, and ethnicity.

**Diagnoses**

Codes and descriptions of the primary and secondary diagnoses recorded during the hospital stay.

**Treatments and Procedures**

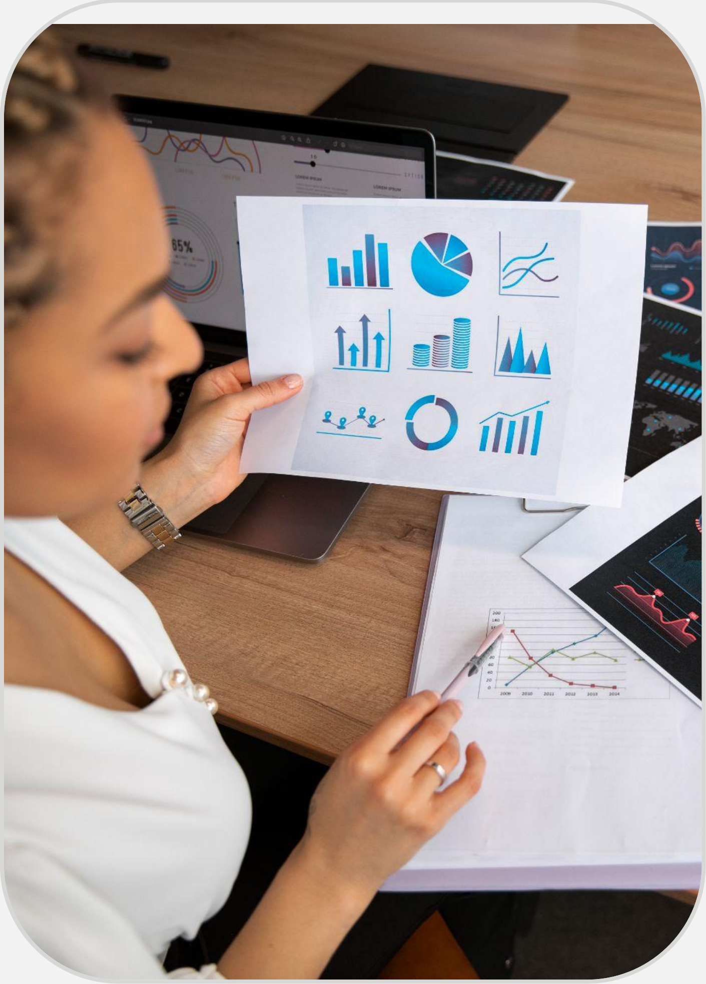
Details on the treatments and medical procedures performed on the patients.

**Services**

Information on the healthcare services provided during the hospital stay.

**Charges and Costs**

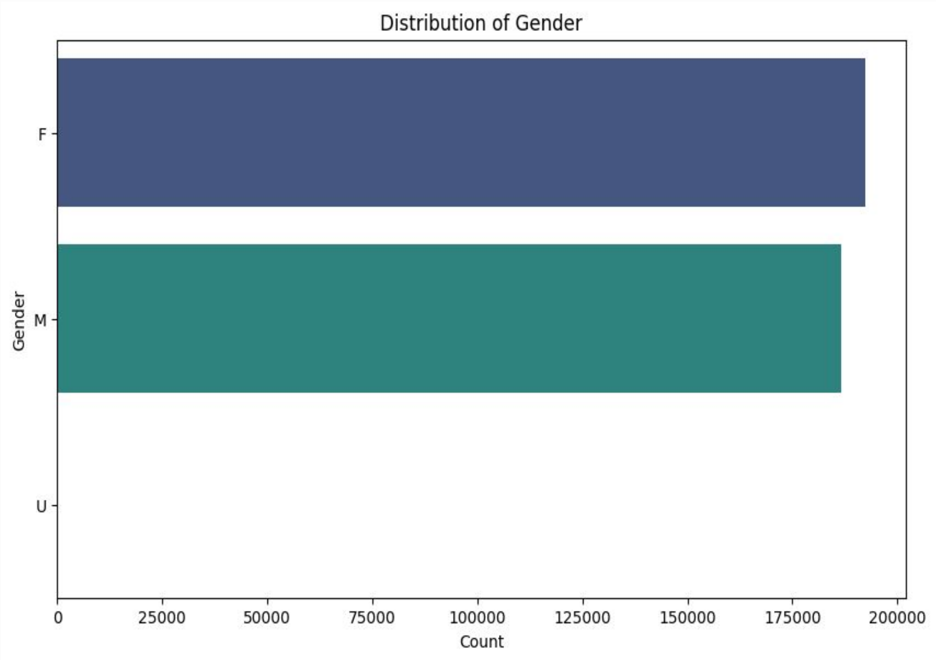
Financial data related to the total charges and costs incurred during the hospital stay.




# EDA

# Identify High-Risk Populations

# Distribution of Gender

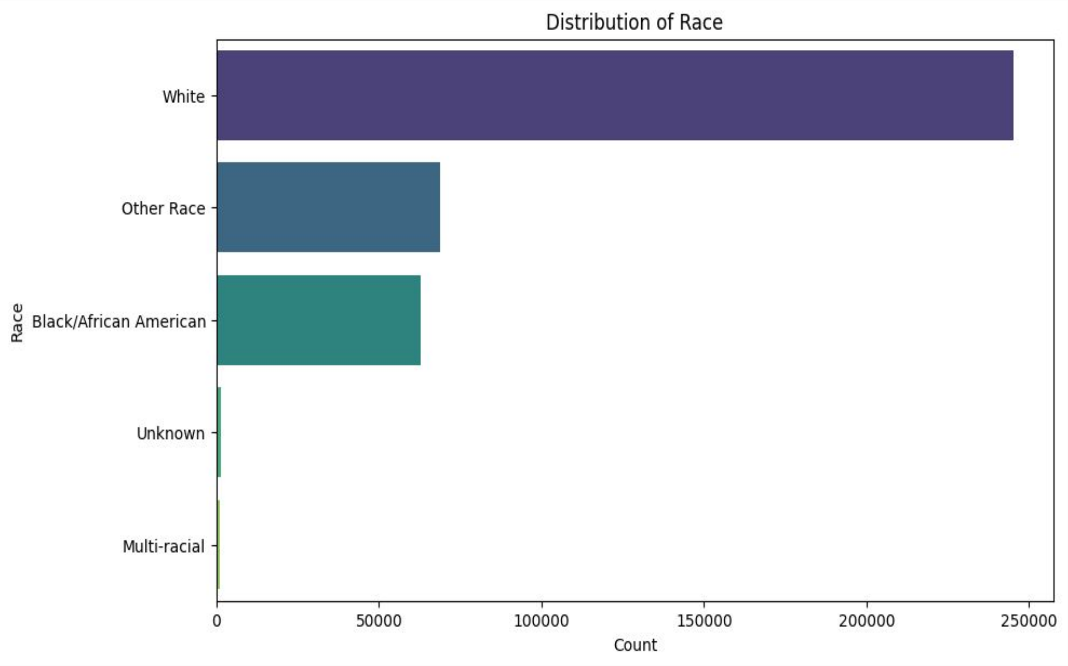



F stands for Female, M stands for Male and U stands for Unknown

The dataset shows that females have a higher representation compared to males, with 192,456 females and 186,581 males.

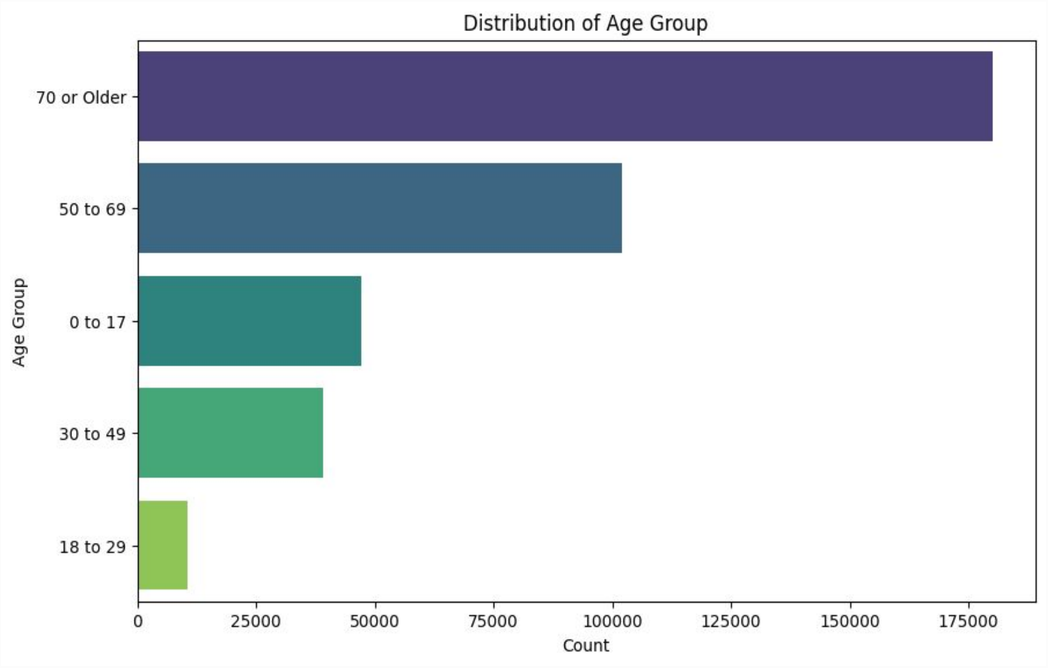
The difference between the number of females and males is 5,875. The very small number of unknown entries (3) indicates that gender data is largely available and well-documented.

# Distribution by Race



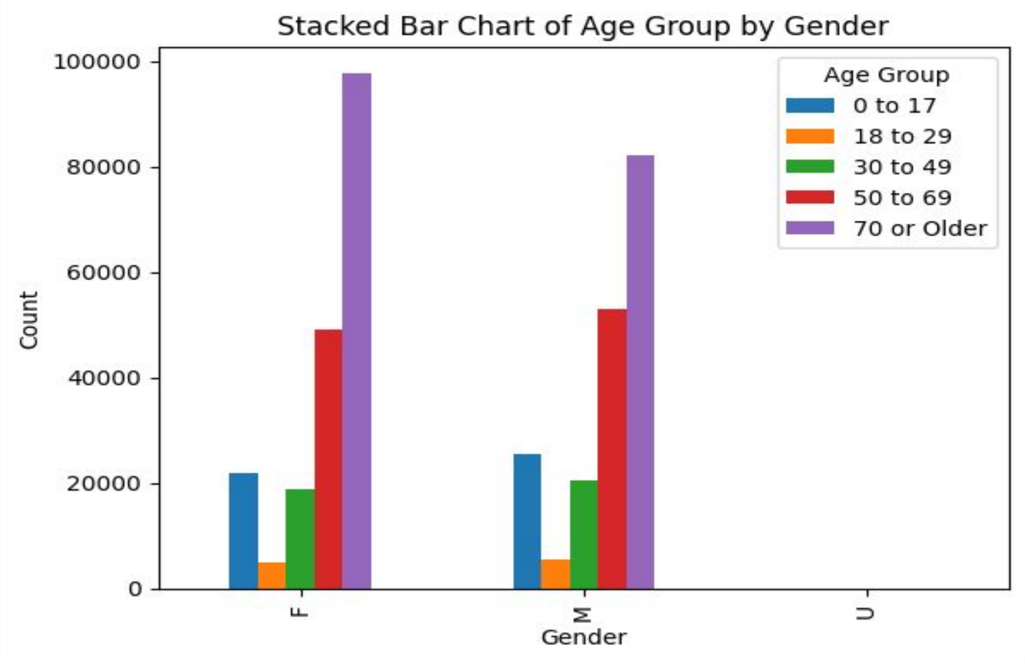

The most dominant race in the dataset is White, followed by Other Race, Black/African American, Unknown, and finally Multi-Race.

# Distribution by Age Group




The most prevalent age group in our data set is 70 or older

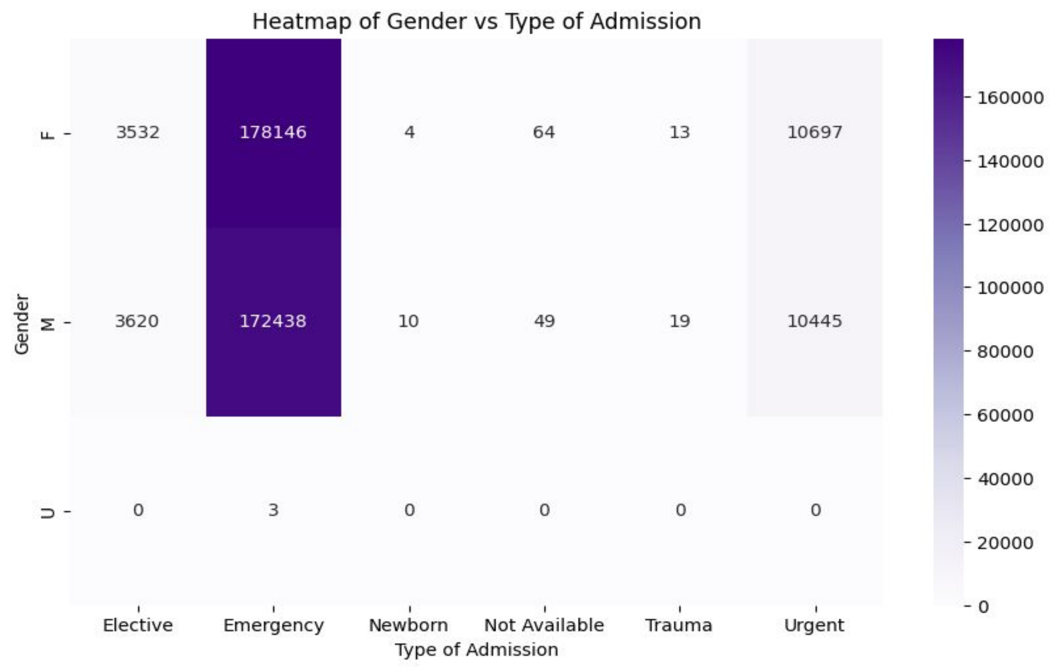
# Age group by Gender




The age distribution shows that both genders have the highest representation in the "70 or Older" category, with females (97,749) and males (82,243) being predominant.

In the younger age groups, males and females are more evenly distributed, but females have higher counts across all age groups compared to males, except for the "0 to 17" category where males slightly exceed females.

# Gender by Type of Admission

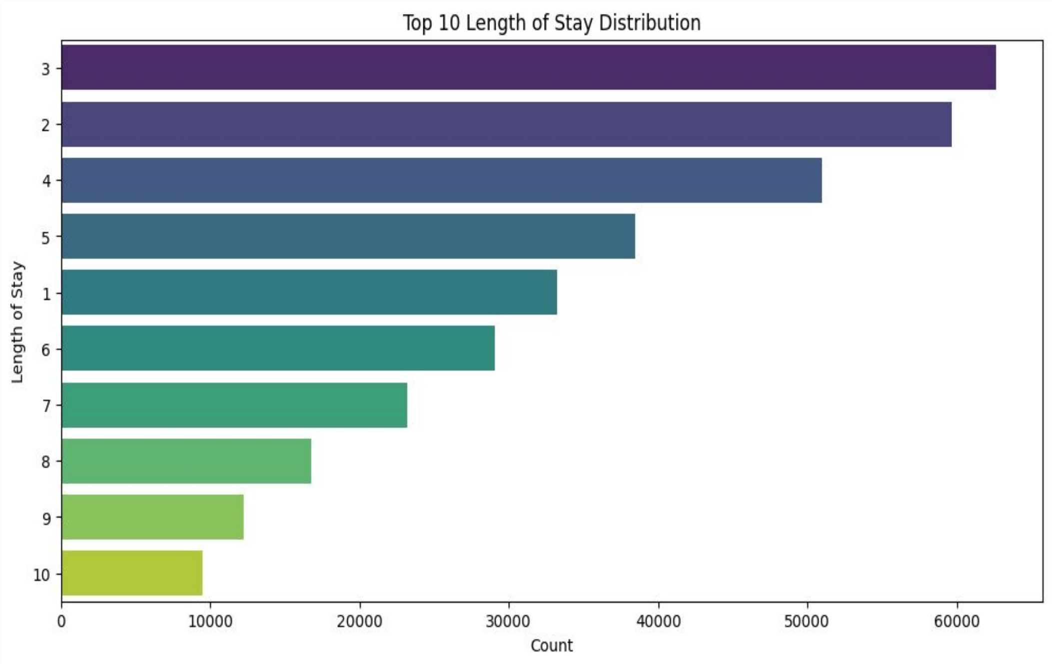



**Emergency Admissions:**
The number of emergency admissions for females (178,058) is slightly higher than for males (172,353).

**Elective Admissions:**
The number of elective admissions is quite similar between females (3,531) and males (3,620).

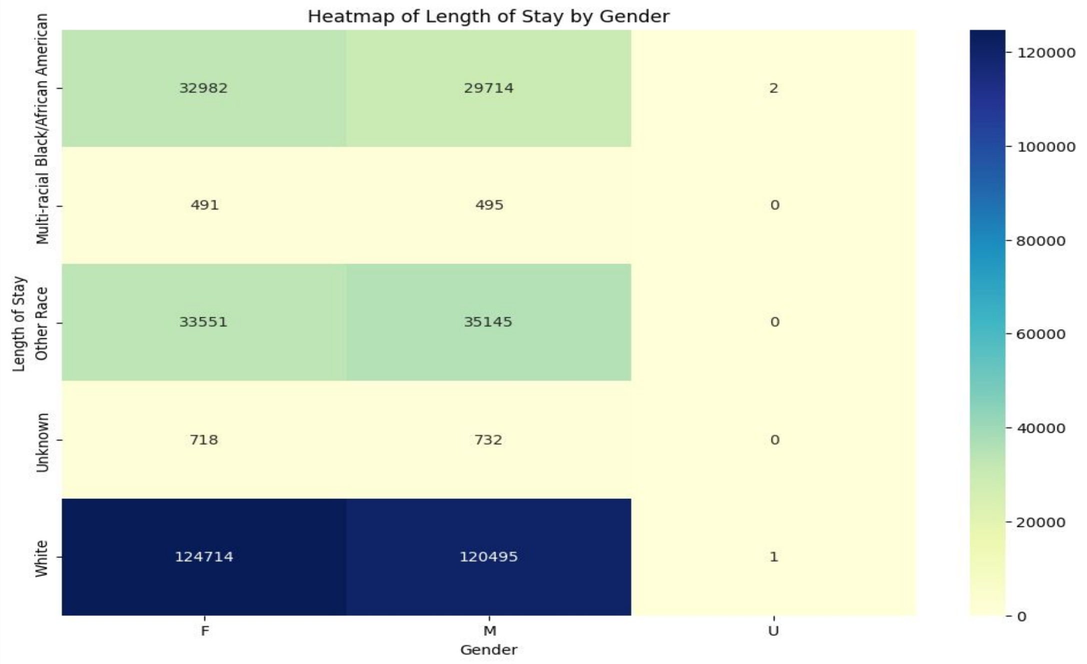
# Evaluate Quality of Care

# Assessing Top 10 Length of Stay




The top 10 lengths of stay predominantly range from 1 to 10 days, with shorter stays being much more common. In contrast, the bottom 10 lengths of stay, which are significantly longer, are rare, indicating that extended stays are exceptional in the dataset.

# Assessing Length of Stay by Gender

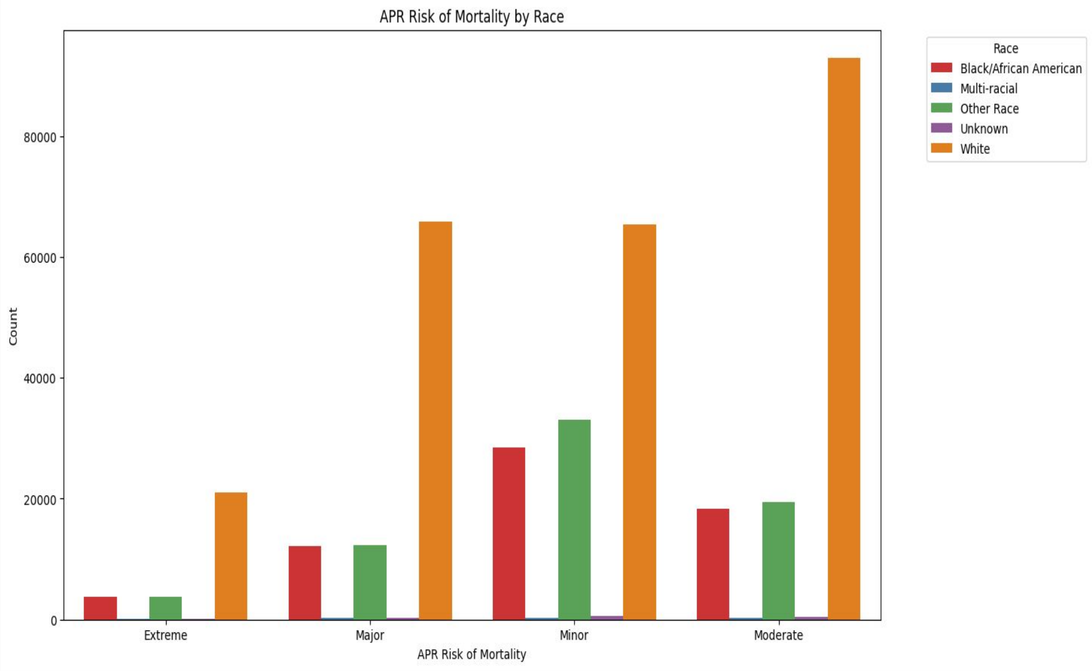



Female and Male Patients: Both genders have similar median and 75th percentile lengths of stay,

The length of stay ranges from 1 to 120 days for both genders.

Unknown Gender: The mean length of stay for unknown gender is notably higher than that of known genders

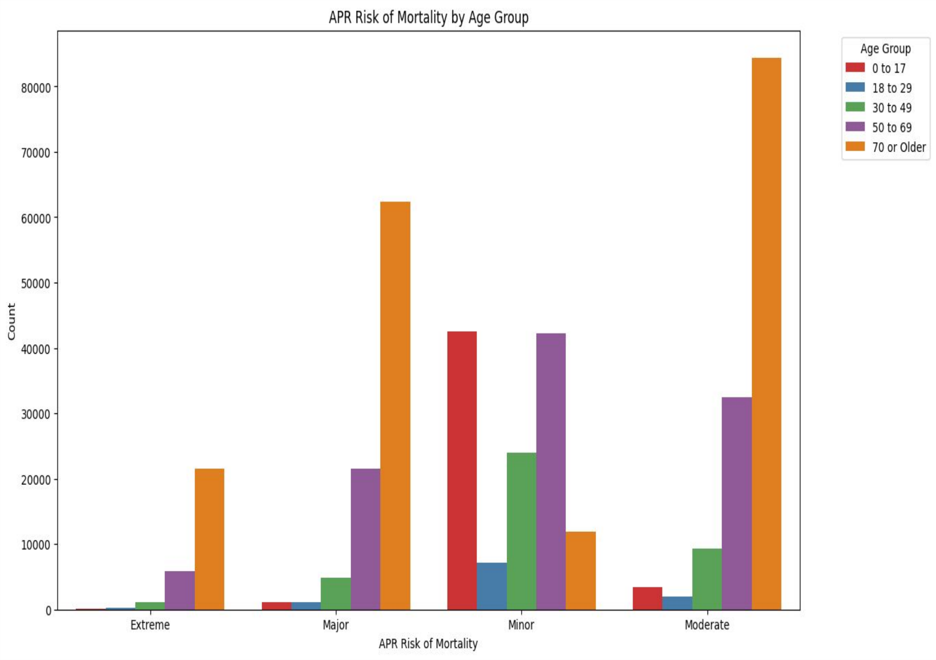
# Risk of Mortality by Race



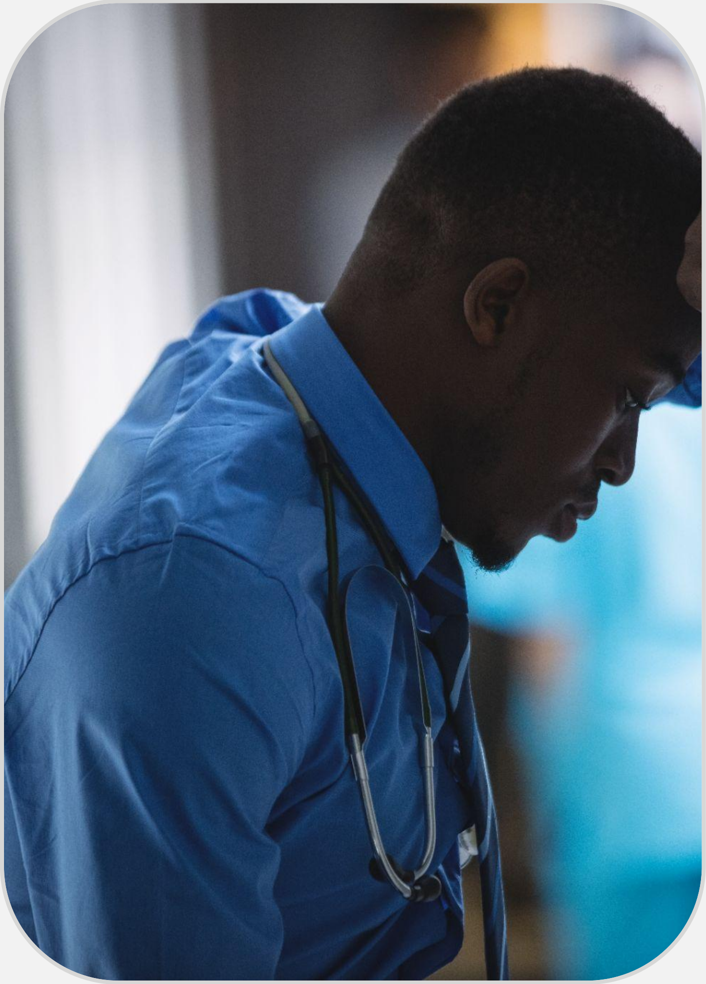

Extreme Risk: The number of patients at extreme risk is notably low across all races, with the highest counts seen in the White population (21,067) and the lowest in Multi-racial (94).

Major Risk: The White population has the highest count (65,867), followed by the Black/African American group (12,106). The Multi-racial group has a lower count (299).

# Mortality rate by Age Group

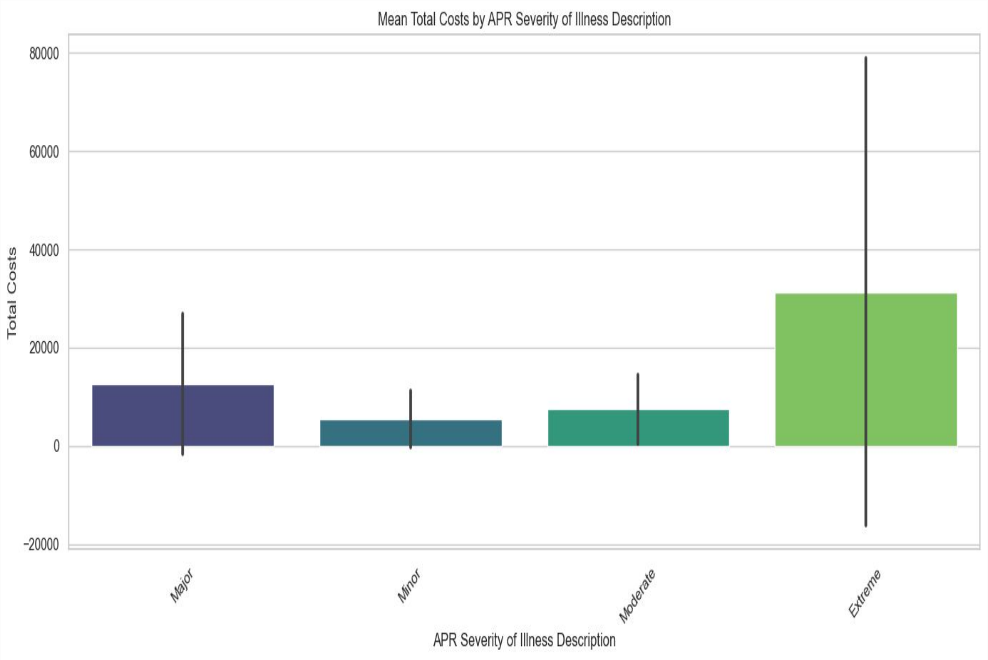



Extreme Risk: This risk level is most prevalent among the elderly (70 or Older) with 21,542 cases.
Major Risk: The highest counts are observed in the 70 or Older age group (62,306), followed by the 50 to 69 group (21,572)
Minor Risk: This risk is most common among the 0 to 17 age group (42,588) and significantly drops in older age groups
Moderate Risk: The highest counts are in the 70 or Older group (84,364) and 50 to 69 group (32,440)



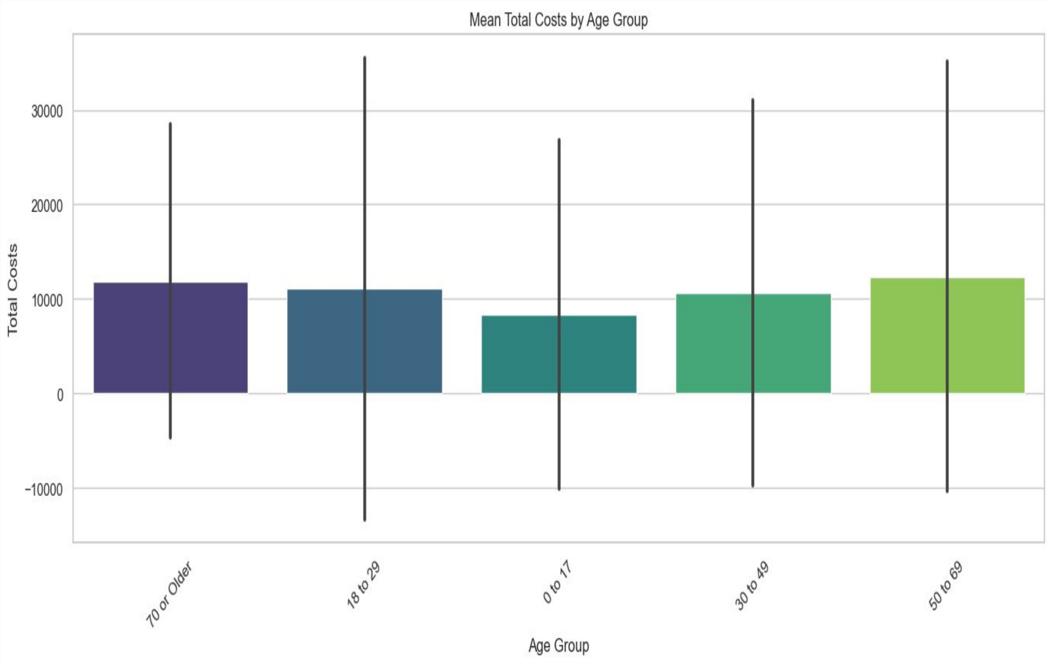
# Assessing the economic burden

# Assessing the economic burden

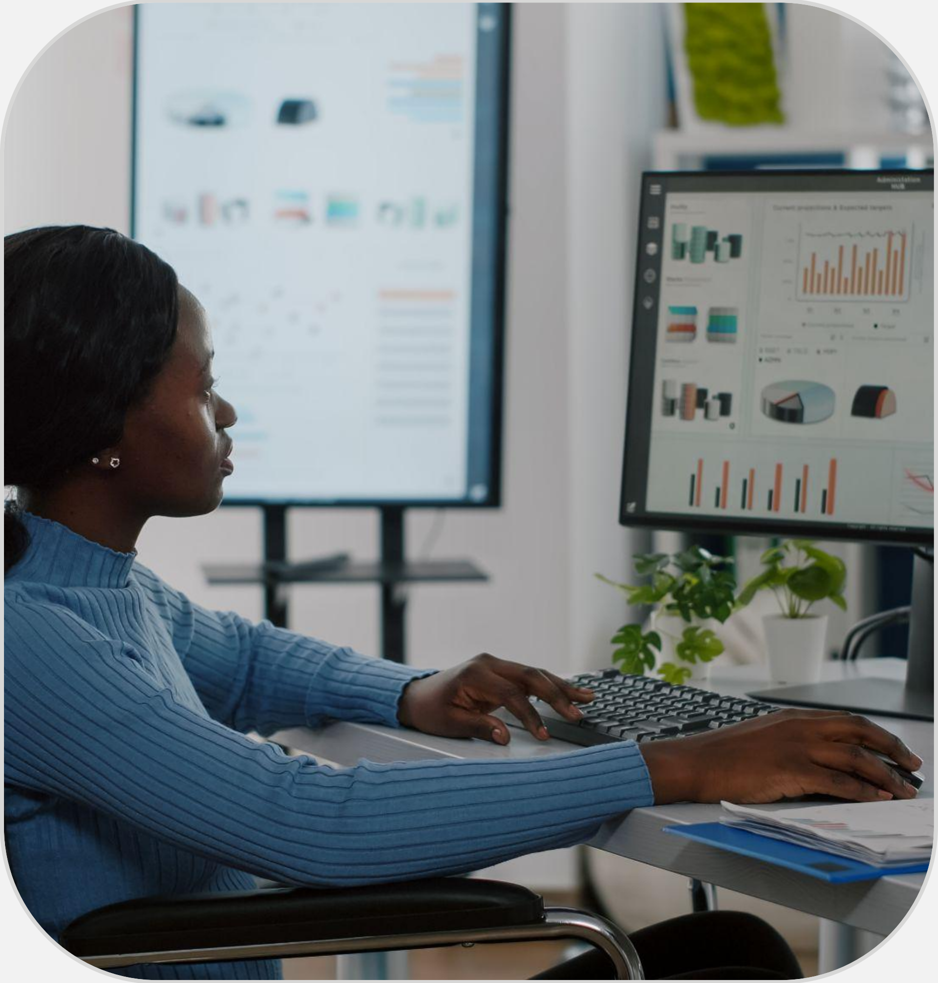



As the severity of pneumonia increases, the total cost of treatment also increases

# Total Costs by Age Groups




The bar chart presents the mean total costs by age group
70 or Older: Mean total cost is approximately 12,000.
18 to 29: Mean total cost is approximately 11,000.
0 to 17: Mean total cost is approximately 7,000.
30 to 49: Mean total cost is approximately 8,000.



# Modeling

| MODEL | METRIC | CLASS 0 | CLASS 1 | CLASS 2 | CLASS3 | ACCURACY |
|---|---|---|---|---|---|---|
| Logistic Regression (Baseline) | Precision | 0.63 | 0.55 | 0.71 | 0.81 | 0.71 |
| | Recall | 0.75 | 0.67 | 0.52 | 0.87 | |
| | F1 - Score | 0.68 | 0.61 | 0.60 | 0.84 | |
| Logistic Regression - Tuned | Precision | 0.63 | 0.55 | 0.71 | 0.81 | 0.70 |
| | Recall | 0.75 | 0.67 | 0.52 | 0.87 | |
| | F1 - Score | 0.68 | 0.61 | 0.60 | 0.84 | |
| Ensemble Methods | Precision | 0.70 | 0.63 | 0.69 | 0.85 | 0.71 |
| | Recall | 0.62 | 0.68 | 0.67 | 0.84 | |
| | F1 -Score | 0.66 | 0.65 | 0.68 | 0.85 | |

| | | | | | | |
|---|---|---|---|---|---|---|
| Feed Forward Neural Network | Precision | 0.68 | 0.59 | 0.64 | 0.84 | 0.71 |
| | Recall | 0.61 | 0.59 | 0.66 | 0.84 | |
| | F1 - Score | 0.64 | 0.59 | 0.65 | 0.84 | |
| Multi-Layer Perceptron | Precision | 0.68 | 0.60 | 0.65 | 0.85 | 0.71 |
| | Recall | 0.62 | 0.63 | 0.66 | 0.82 | |
| | F1 - Score | 0.65 | 0.61 | 0.65 | 0.83 | |
| Multi-Layer Perceptron - Tuned | Precision | 0.70 | 0.61 | 0.68 | 0.85 | 0.72 |
| | Recall | 0.63 | 0.67 | 0.65 | 0.85 | |
| | F1 - Score | 0.66 | 0.64 | 0.66 | 0.85 | |
| Artificial Neural Network | Precision | 0.69 | 0.59 | 0.65 | 0.84 | 0.70 |
| | Recall | 0.58 | 0.62 | 0.65 | 0.82 | |
| | F1 - Score | 0.63 | 0.61 | 0.68 | 0.83 | |

## Conclusion

| | |
|---|---|
| **APR Severity of Illness Code** | The most significant predictor of APR Risk of Mortality<br>Patients with more severe conditions have higher mortality risk |
| **Age** | Crucial factor in determining mortality risk<br>Patients aged 70 and above have significantly higher mortality risk<br>Need to prioritize close monitoring and care for older patients<br>Ensure proper emergency activation mechanisms are in place |
| **Hemodialysis** | Patients undergoing hemodialysis have increased mortality risk<br>Highlights need for careful monitoring and management of patients with renal failure or related conditions |
| **Severity of Illness** | Patients classified as having major or moderate severity levels should be monitored closely<br>Description of illness, particularly major and moderate, plays a significant role in predicting mortality risk |

## Recommendations

| | |
|---|---|
| **Enhanced Monitoring and Treatment** | Implement advanced monitoring and treatment strategies for patients with higher APR Severity of Illness Codes (Extreme, Major, Moderate)<br>Deploy resources for intensive care units and specialized treatment plans |
| **Age-Specific Care Plans:** | Create age-specific care plans, focusing on preventive measures and early interventions for patients aged 70 and older<br>Conduct regular health assessments for elderly patients to identify potential risks early and address them promptly |
| **Comprehensive Care for Renal Patients:** | Develop comprehensive care programs for hemodialysis patients<br>Include regular check-ups, nutritional support, and access to specialized care |
| **Improve Transition Care**: | Enhance discharge planning and follow-up care for patients discharged to home or self-care<br>Ensure patients have access to necessary resources and support |

# Meet the team

Jubilant Mutuku

Ivan Wawire

Millicent Muthomi

Maureen Muriithi

Wilson Wachira

Lydia Masabarakiza