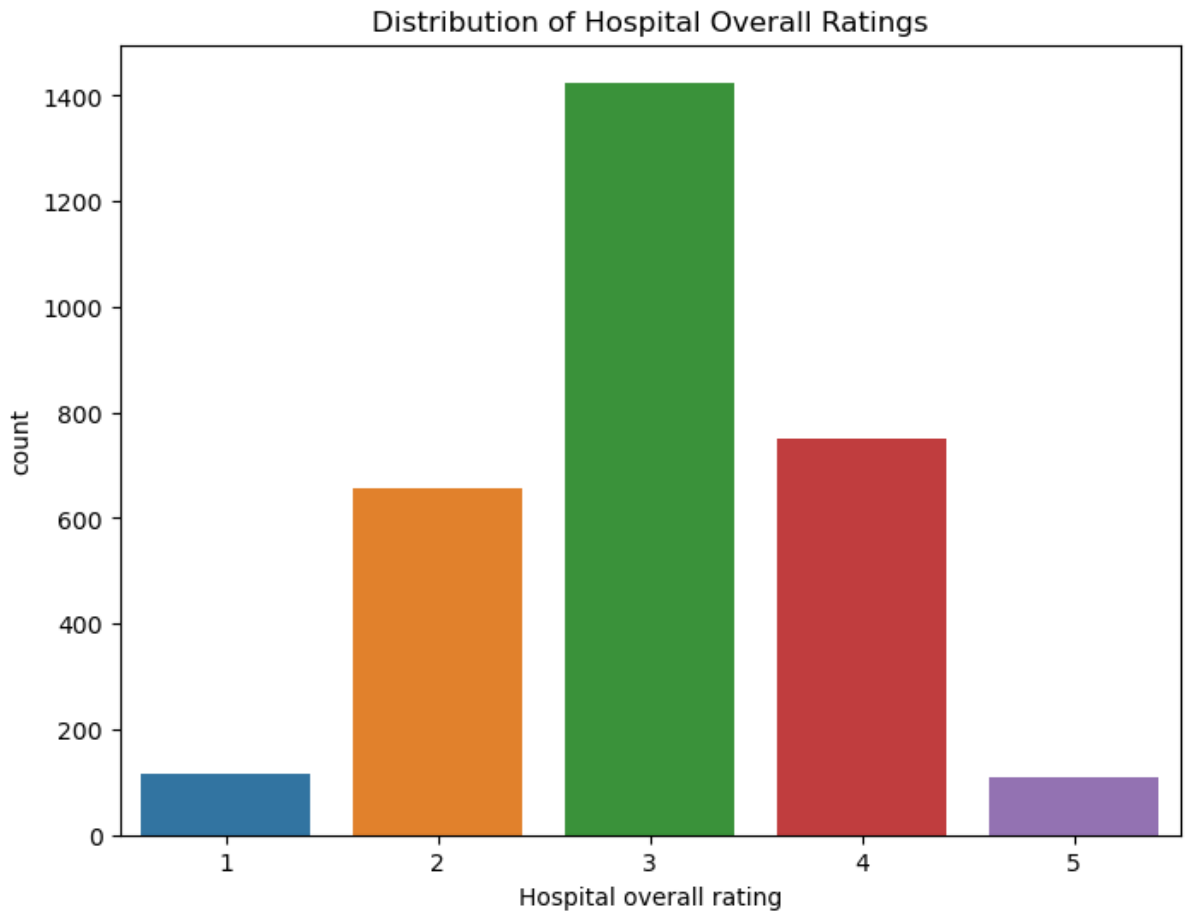
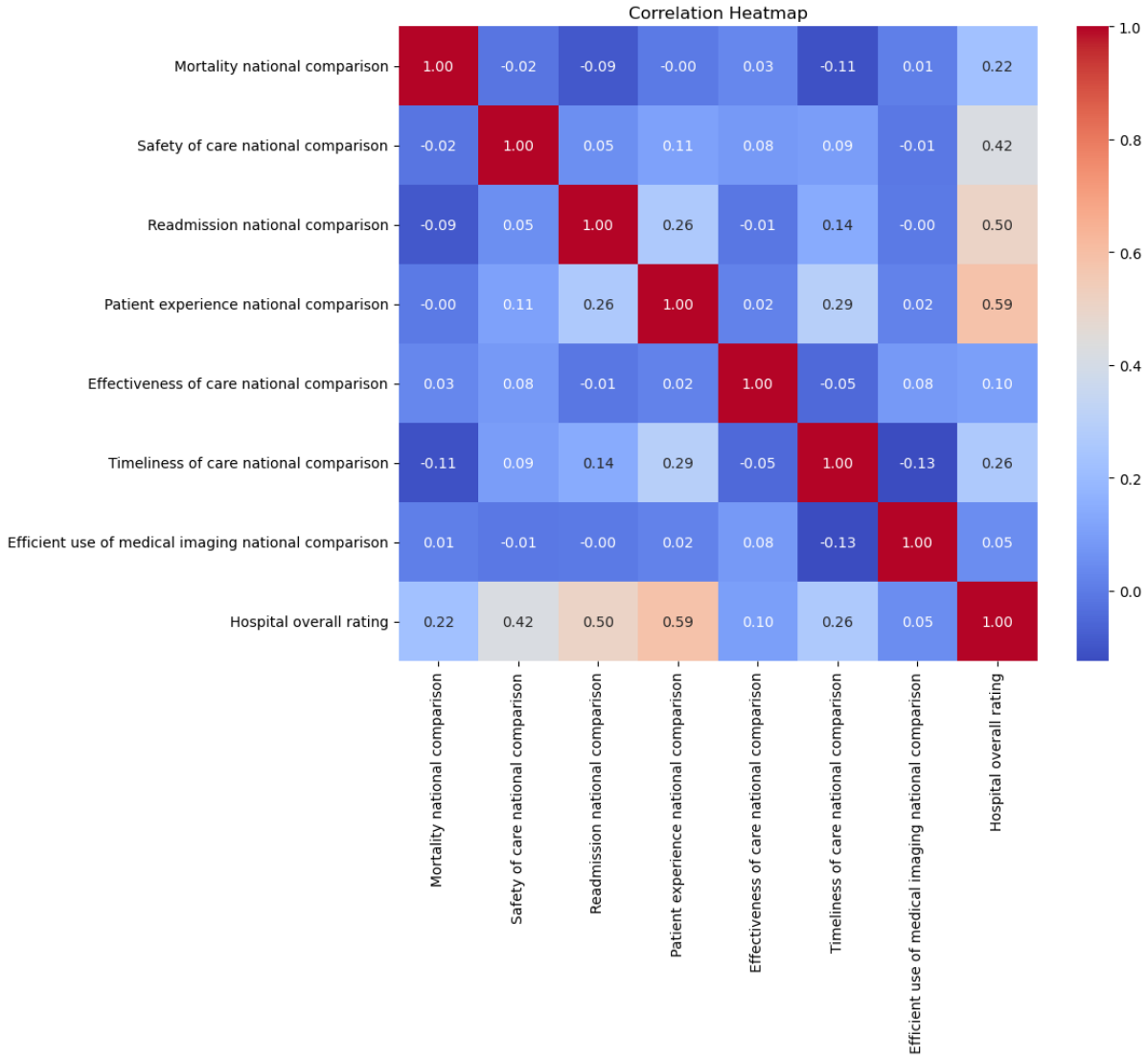
Name: Vimal Gondaliya

Task 1: Understand the data



- The distribution of hospital overall ratings is skewed to the right, meaning that there are more hospitals with lower ratings than hospitals with higher ratings.
- The most common hospital rating is 3 stars, with 18.7% of hospitals receiving this rating.
- There are relatively few hospitals with 1 star (5.4%) or 5 stars (12.2%).
- The data in the chart is based on the Centers for Medicare & Medicaid Services (CMS) Overall Hospital Quality Star Ratings, which are calculated using a variety of measures, including mortality, safety of care, readmission rates, patient experience, effectiveness of care, timeliness of care, and efficient use of medical imaging.

It is important to note that the star ratings are just one measure of hospital quality, and they should not be the only factor you consider when choosing a hospital. Other factors you may want to consider include the hospital's reputation, the expertise of its doctors and nurses, and its location.



- **Hospital overall rating appears to be positively correlated with patient experience, safety of care, and timeliness of care.** This means that hospitals that received higher ratings for patient experience, safety of care, and timeliness of care also tended to receive higher overall ratings.
- **Hospital overall rating appears to be negatively correlated with mortality, readmission, and efficient use of medical imaging.** This means that hospitals that had higher mortality rates, higher readmission rates, and less efficient use of medical imaging also tended to receive lower overall ratings.
- **The strength of the correlations varies.** The strongest correlation is between patient experience and hospital overall rating, followed by safety of care and timeliness of care. The

weakest correlations are between efficient use of medical imaging and hospital overall rating, and between mortality and hospital overall rating.

Task 2: Build machine learning models

Accuracy: The model has a high accuracy of 96.3%, indicating it correctly classifies 96.3% of the observations. This is a very good result, suggesting the model is effective at distinguishing between the two classes.

Precision and Recall: Both precision and recall are above 0.9 for both classes, indicating the model is good at both identifying true positives and avoiding false positives/negatives.

Class-Specific Performance:

Class 0: The model performs even better for class 0 (precision and recall both very high at 0.97 and 0.98), suggesting it can accurately identify negative examples.

Class 1: While still good, the performance for class 1 is slightly lower (precision 0.96, recall 0.91). This means the model might be misclassifying some positive examples as negative more often than vice versa.

Insights on k-Nearest Neighbors (k-NN) Model:

Accuracy: The k-NN model has a moderate accuracy of 85.9%, which is significantly lower than the Logistic Regression model. This suggests the k-NN might not be as effective in learning the underlying patterns in your data.

Precision and Recall: Both precision and recall are lower than the Logistic Regression model, especially for class 1 (positive class). This indicates the k-NN struggles with both identifying true positives and avoiding false positives/negatives compared to Logistic Regression.

Class-Specific Performance:

Class 0: Similar to Logistic Regression, the k-NN model performs better for class 0 (precision and recall are higher at 0.89 and 0.92).

Class 1: The performance for class 1 is significantly worse than the Logistic Regression model (precision 0.77, recall 0.72). This implies the k-NN model is misclassifying many positive examples as negative more frequently.

Insights on Decision Tree Model:

Accuracy: The Decision Tree model has a moderate accuracy of 83.1%, which is lower than the Logistic Regression model but similar to the k-NN model. This suggests the Decision Tree may not be capturing the most important relationships in your data compared to Logistic Regression.

Precision and Recall: Both precision and recall are lower than the Logistic Regression model, especially for class 1 (positive class). While similar to k-NN in overall values, the Decision Tree shows a different pattern with higher precision (0.73) but lower recall (0.65) for class 1. This indicates the Decision Tree is less prone to false positives but still struggles with identifying true positives compared to Logistic Regression.

Class-Specific Performance:

Class 0: Similar to the other models, the Decision Tree performs better for class 0 (precision and recall are higher at 0.87 and 0.90).

Class 1: Similar to k-NN but worse than Logistic Regression, the Decision Tree struggles with class 1 (precision 0.73, recall 0.65). This implies difficulty in correctly classifying positive examples.

Model	Accuracy	Strengths	Weaknesses
Logistic Regression	0.963	- High accuracy for both classes (precision and recall above 0.9)	- Less interpretable compared to Decision Tree
k-NN	0.859	- Moderate accuracy, good for identifying negative class (class 0)	- Lower accuracy for identifying positive class (class 1)
Decision Tree	0.831	- More interpretable due to its tree structure	- Lower accuracy overall compared to Logistic Regression and k-NN