Data Analysis: Hospital Readmission Rates

Within this paper I will cover my processes of data acquisition, analytical methodologies, outcomes, the standardization of the workflow, and potential recommendations drawn from the analysis.

DATA ACQUISITION

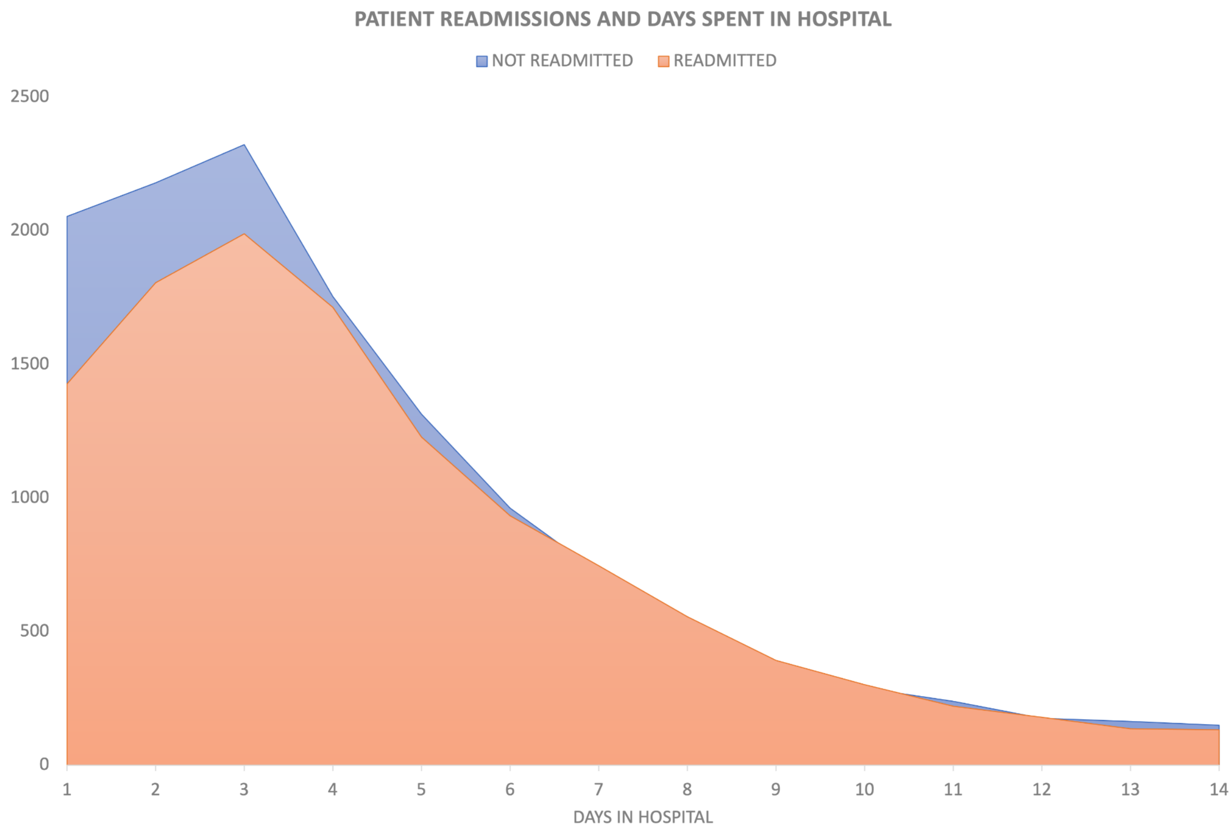
This project analyzes a dataset which contains ten years of patient data regarding hospital readmissions. This dataset contains 25,000 records and was acquired from the data repository website Kaggle. I chose this dataset to simulate similar analyses which may be utilized within a healthcare setting. For this project I chose to work with the data in Excel, although this project could be recreated within another data analysis tool such as Python or Tableau.

ANALYTICAL PROCESS

The first step I take when approaching a new dataset is get to know it. My first few thoughts are to find out; what is the size of the dataset, do I understand all parts of the data, how clean is the data, and how can this data provide meaningful insights? The first few can be rather easy, just some summary statistics and research (this dataset was quite clean and just contained some missing data), although the last question typically takes more thought, especially without a specific direction. Given that the dataset encompasses various health-related aspects, including the patients' age, hospital stay duration, primary diagnosis, glucose test results, and more, it presents a natural opportunity to explore the health features that might contribute to higher patient readmission rates.

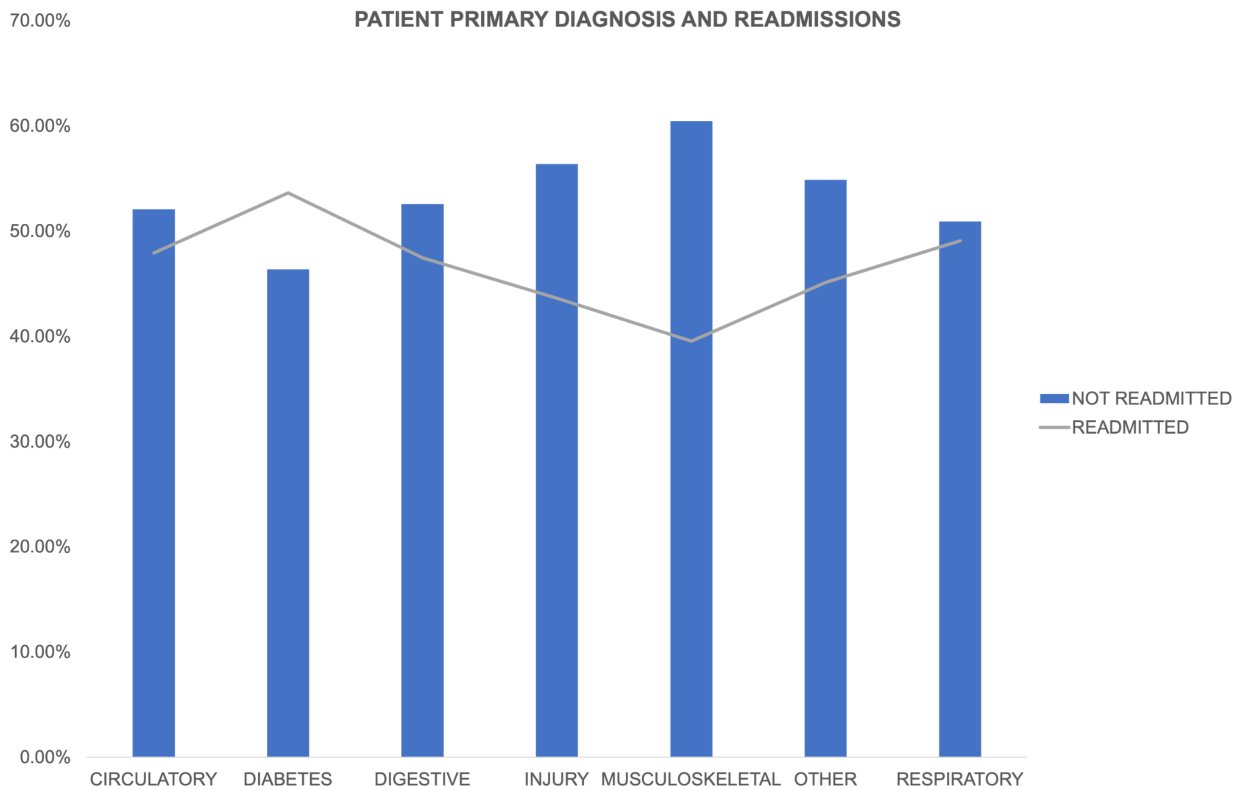
With the general purpose of the analysis set into place, I then began some exploratory data analysis (EDA) to find any trend(s) in the data which may be correlated to higher patient readmission. I start by comparing each feature (or column) of the dataset to the target feature (the target feature being readmission). This allowed me to both analyze and visualize potential trends within the dataset. An example was this visualization of hospital stay duration (in days) compared to patient readmission (see Figure 1).

Figure 1.



Through analyzing the dataset's features, the patient's primary diagnosis was a notable feature for correlation with readmission rates (see Figure 2).

Figure 2.

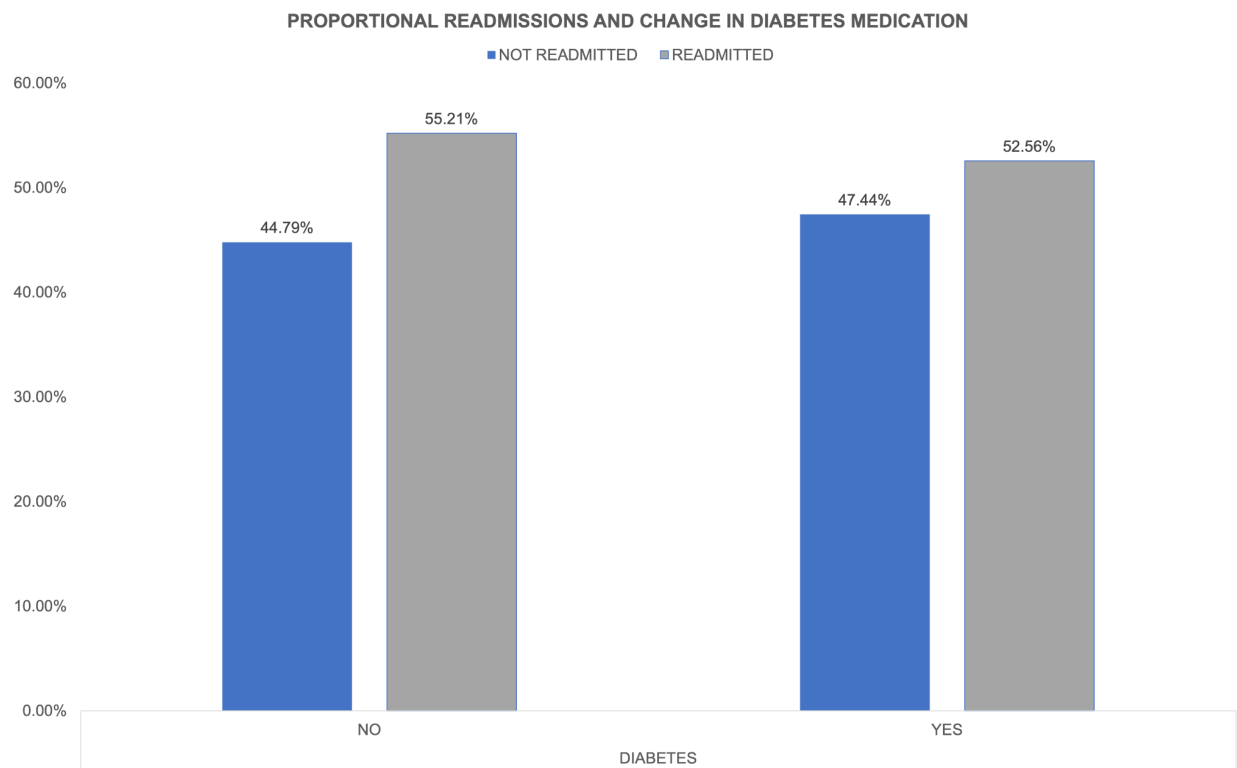


This bar/line chart highlights that diabetes is the primary diagnosis associated with a higher proportion of readmissions. Building on this observation, I further explored a deeper analysis of patients with diabetes to uncover additional meaningful insights.

ANALYSIS

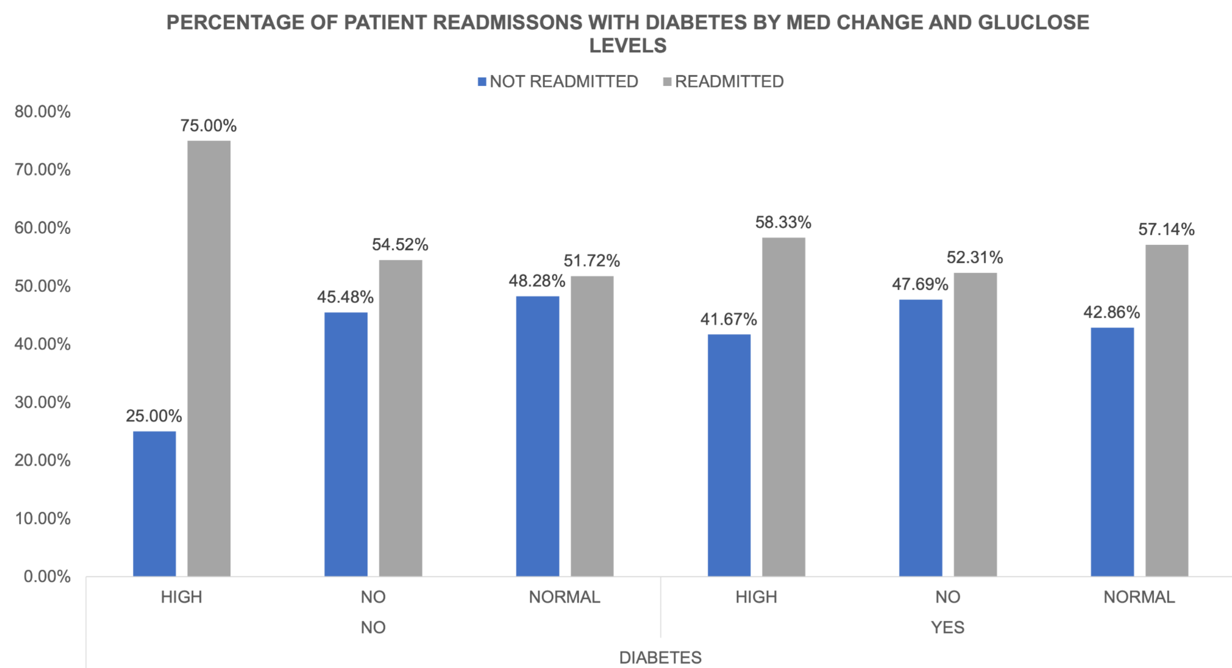
With this core insight, I targeted features that not only have correlation with readmission but also with diabetes. Initially, I created a visualization of patients with a primary diagnosis of diabetes, categorizing them based on whether their diabetes medication was changed during their hospital stay (using "NO" for no medication change and "YES" for medication change). The analysis revealed that, while the two groups were relatively similar, there was a slightly higher readmission rate associated with patients who did not have their diabetes medication changed (see Figure 3).

Figure 3.



To delve deeper into the analysis, I introduced another aspect: the patient's glucose test results. The dataset recorded glucose levels as "HIGH" (> 200), "NORMAL" for normal levels, or "NO" if the test was not performed. By visualizing this information in the chart below (see Figure 4), there are proportionally more readmissions for patients who have high glucose levels, whether their diabetes medication was changed or not.

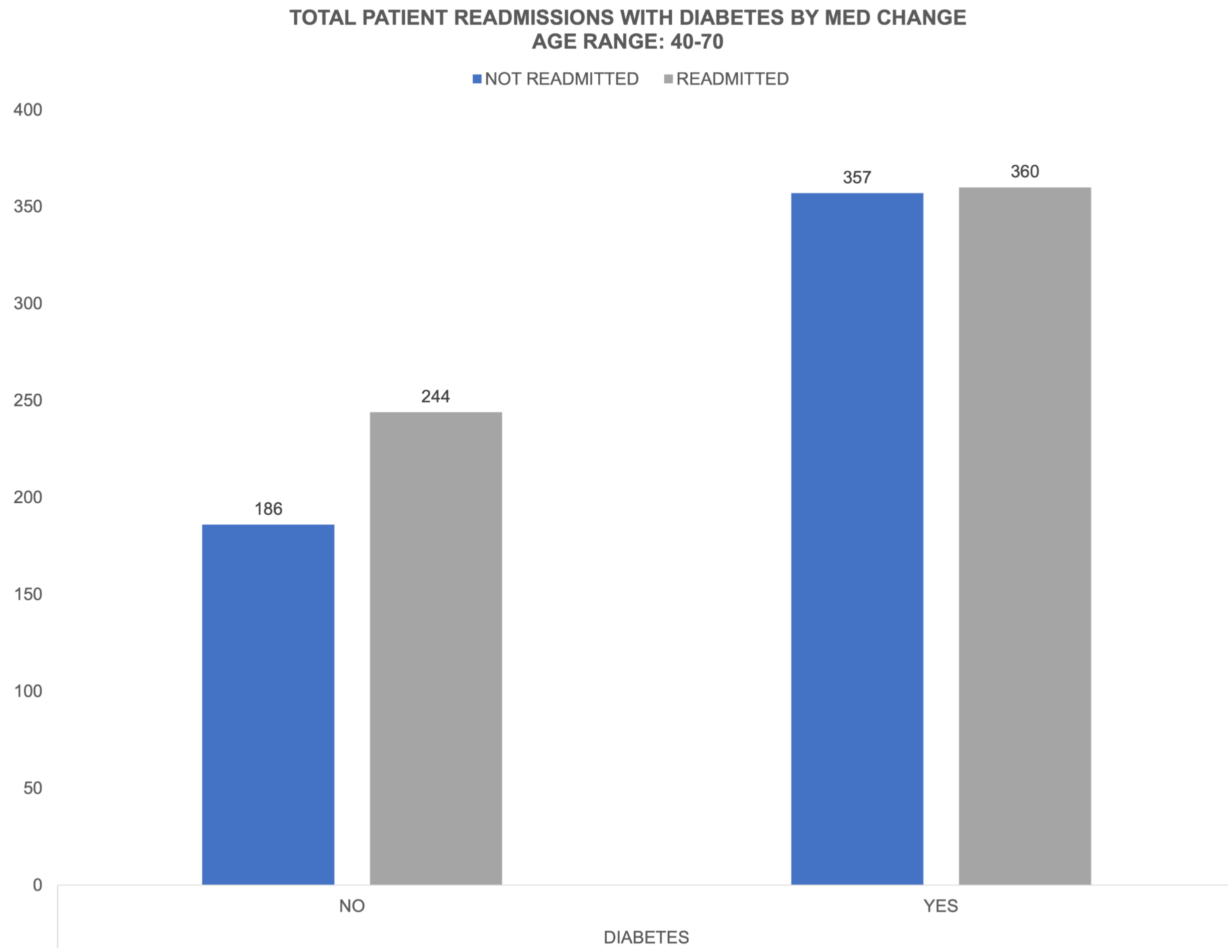
Figure 4.



There are also about 15% higher readmissions for patients whose glucose levels were normal but had their diabetes medication changed. Although, the data here is quite disproportionately represented, with 94% of patients not receiving a glucose test, thus this feature may not be useful for this analysis (more data from patients receiving glucose tests are needed).

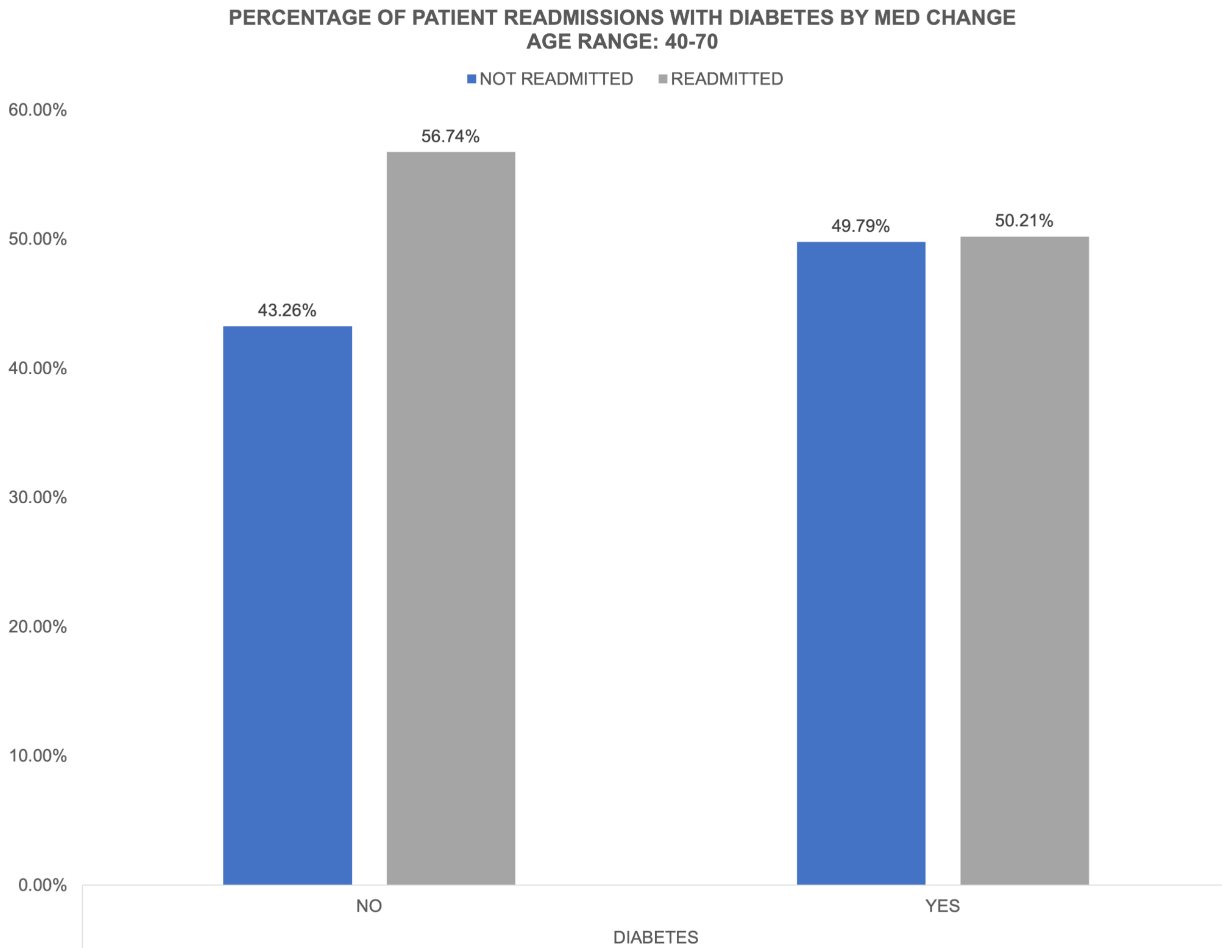
Next, I incorporated the patients' age groups into the analysis. The dataset categorized age groups in a ten-year range, starting from 40 and going up to 100. I initially focused on patients aged 40 to 70 who had diabetes and both did and did not experience a change in their diabetes medication. To gain a comprehensive understanding of the data, I examined both the total count of patients (Figure 5a, representing the total count of patients) and the percentage of patients (Figure 5b, representing the percentage of patients).

Figure 5a.



I discovered that patients between the ages of 40 and 70 who experienced no medication changes, exhibit higher readmission rates compared to those in the same age group who did have medication changes.

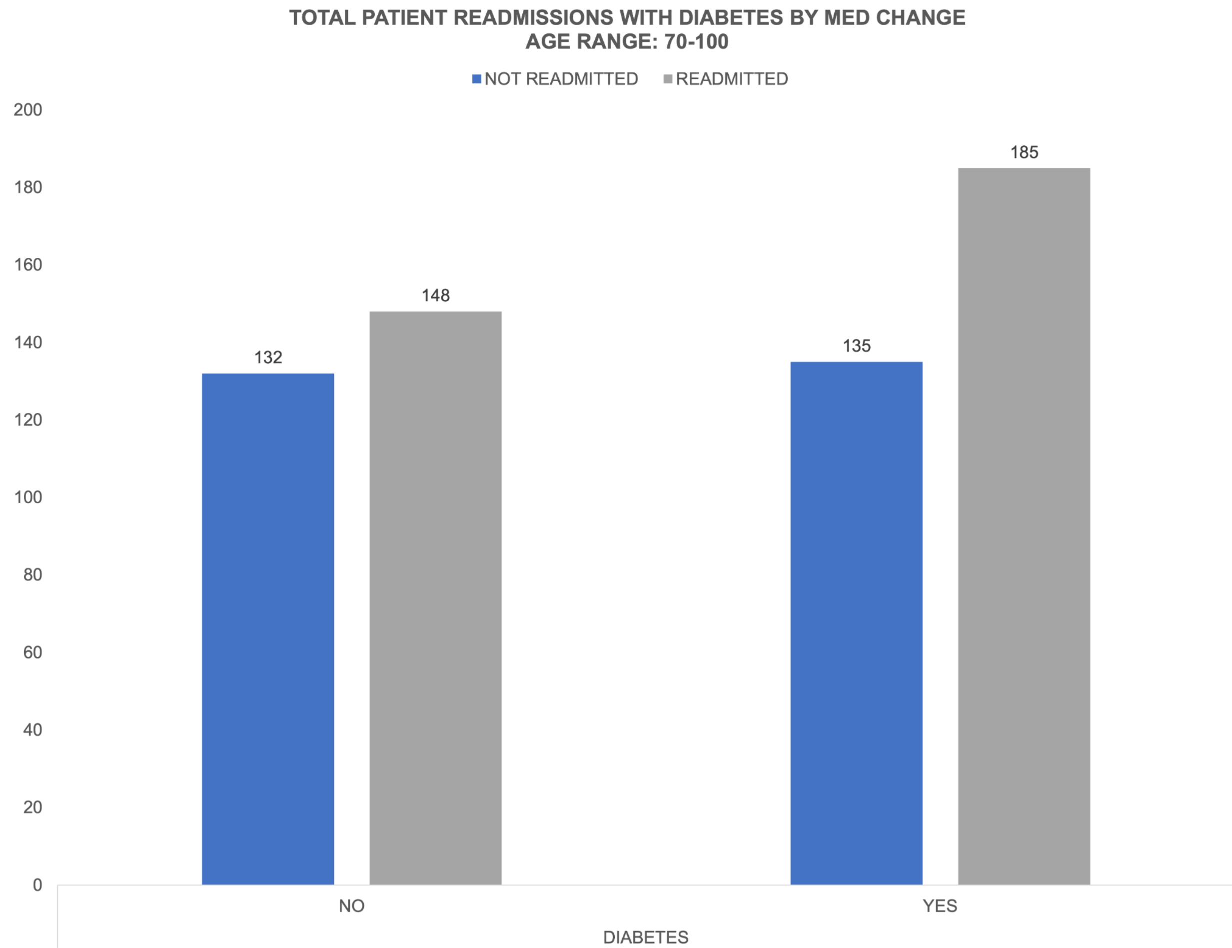
Figure 5b.



Upon analysis, I observed that nearly 13% more patients within the age group of 40-70, who experienced no medication changes, were readmitted to the hospital compared to those who did have medication changes. In contrast, patients in the same age group who underwent medication changes displayed almost equal rates of readmission and non-readmission.

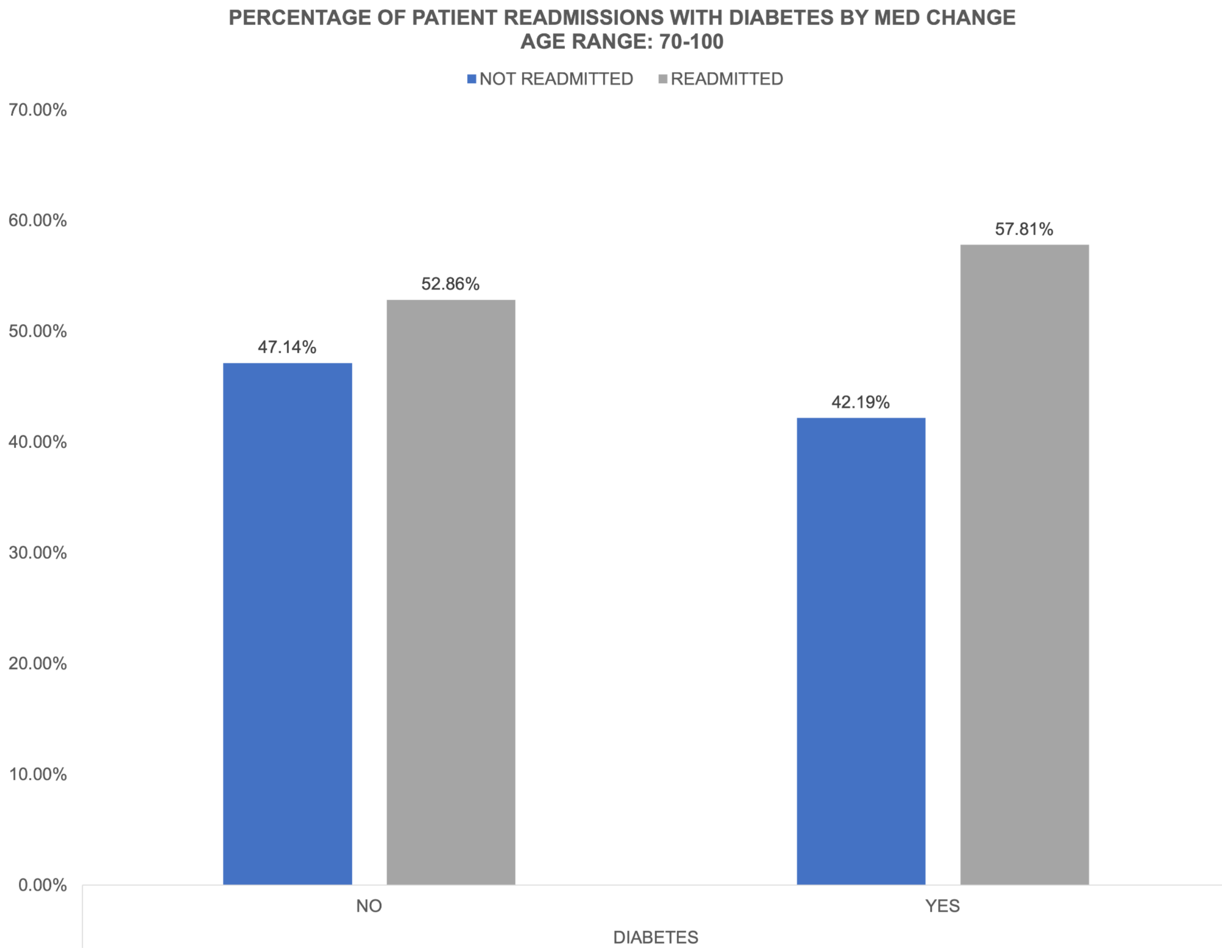
Next, I examined the subsequent age group, 70-100, for patients with diabetes who both experienced and did not experience changes in their diabetes medication. I further explored this data using the total count (Figure 6a) and percentage of patients (Figure 6b).

Figure 6a.



The data analysis revealed that patients in the age group 70-100 who experienced medication changes exhibited higher readmission rates compared to those in the same age group who did not undergo diabetes medication changes.

Figure 6b.



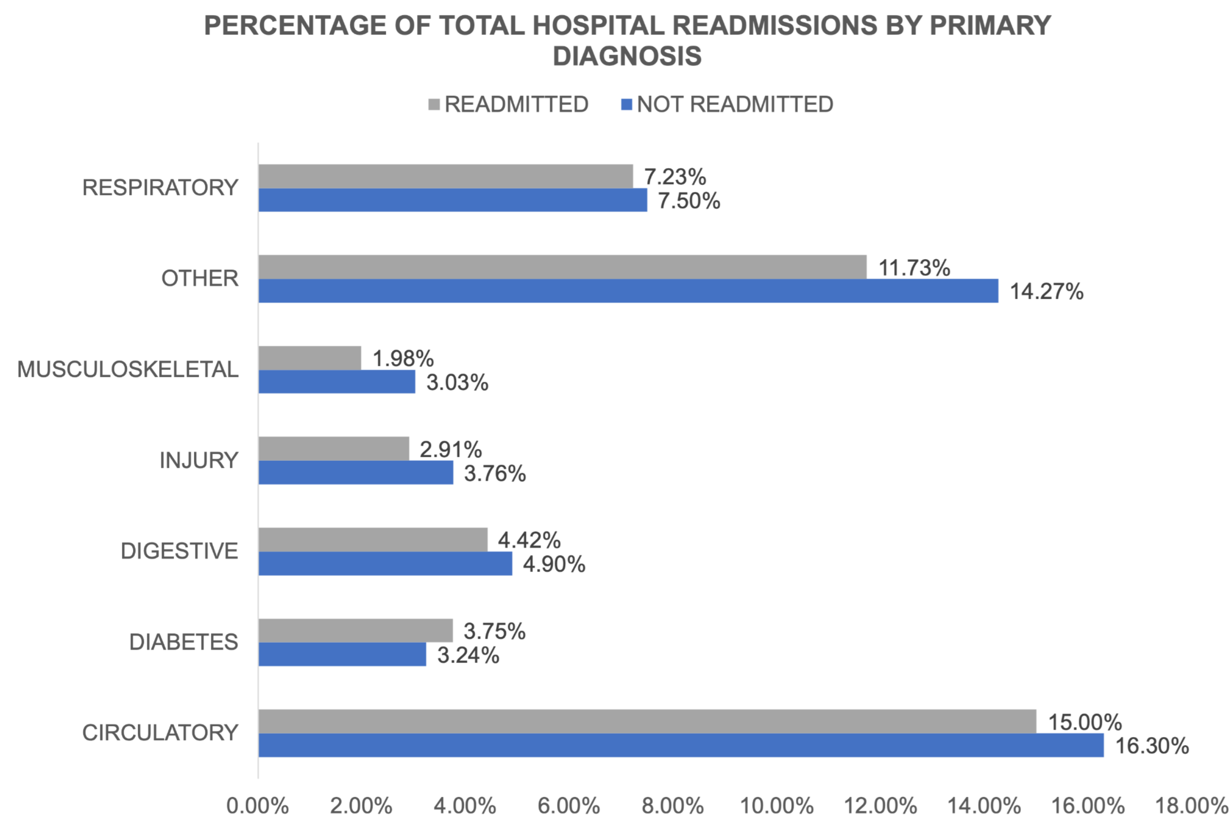
Based on the proportional values, it was observed that nearly 15% more patients in the age group 70-100, who had medication changes, experienced readmission to the hospital. Conversely, patients within the same age group who did not undergo medication changes showed a 5% higher readmission rate. I noticed that this older age group exhibits a trend that contrasts with the younger age group.

DATA OUTCOMES

Overall, the data indicates a correlation between hospital readmission and patients with a primary diagnosis of diabetes. Further analysis of this patient population reveals slightly higher readmission rates among those whose diabetes medication remained unchanged. Specifically, within the age group of 40-70, the data demonstrates a 13.48% higher readmission rate for patients whose diabetes medication was not changed compared to those who were not readmitted. Conversely, in the age group of 70-100, the data shows a 15.62% higher readmission rate for patients whose diabetes medication was changed compared to those who were not readmitted.

It was imperative for me to emphasize that this diabetes population is quite disproportionate to the hospital's overall patients (see Figure 7).

Figure 7.



The focus of this analysis was on the population of patients with a primary diagnosis of diabetes, which constituted 7% of the dataset, while patients with a primary diagnosis of Circulatory conditions accounted for 31% of the dataset.

Overall, there remains more space for further analysis, particularly with valuable insights from qualified healthcare experts. This analysis represents only a single facet of a multifaceted and intricate dataset, offering numerous opportunities for exploration from diverse perspectives.

RECOMMENDATIONS

Based on the dataset, one recommendation is for the appropriate healthcare staff to investigate the reasons behind higher readmission rates among patients aged 40-70 with a primary diagnosis of diabetes, who did not have changes in their diabetes medication. Additionally, another recommendation is for the healthcare staff to explore the factors contributing to higher readmission rates among patients aged 70-100 with a primary diagnosis of diabetes, who had changes in their diabetes medication.

WORKFLOW

I adopt a consistent workflow for every data analysis project, as outlined in this paper. This approach, though simple, ensures that I organize my workflow efficiently and effectively. My standardized workflow consists of the following steps:

- Acquire the data
- Ensure data integrity
- Understand the data
- Formulate meaningful questions about the data
- Explore the data for trends relevant to the questions
- Analyze the data with identified trends
- Present the results of the analysis

By following this structured approach, I can approach data analysis systematically, hopefully leading to more reliable and insightful outcomes.