Initial Exploratory Data Analysis

David K. Kirui

January 30, 2020

1 README

For this project, I am using data provided by the Center for Clinical and Translational Research at Virginia Commonwealth Univeristy. The data is publicly accessible and can be accessed by clicking here: Diabetes Data

My preliminary machine learning analysis and model building can be found in the document in my github entitled rough_draft.pdf.

2 Days Spent in the Hospital

Figure 1 displays the frequency of time that patients spent in the hospital in days among patients who readmitted within 30 days and those who were not. This figure suggests that patients who are not readmitted to the hospital within a month have shorter initial hospital visits - with the majority of non-readmitted patients having initial stays of 5 days of less.

By contrast, a greater proportion of patients who are later readmitted to the hospital have longer hospital stays. This suggests that the shorter the initial visit, the less likely patients are to be readmitted in the short term. However, additional analysis is need to prove that this association is more than coincidental.

3 Age Group

Figure 2 displays the frequency of patients who experienced short-term readmission among the various age groups in the data. This figure suggests that older patients are readmitted to the hospital more often than younger patients. Patients who are between 60 and 79 seem to be the age group that experiences short term readmission the most, while diabetic patients under 20 experience the least short term readmission.

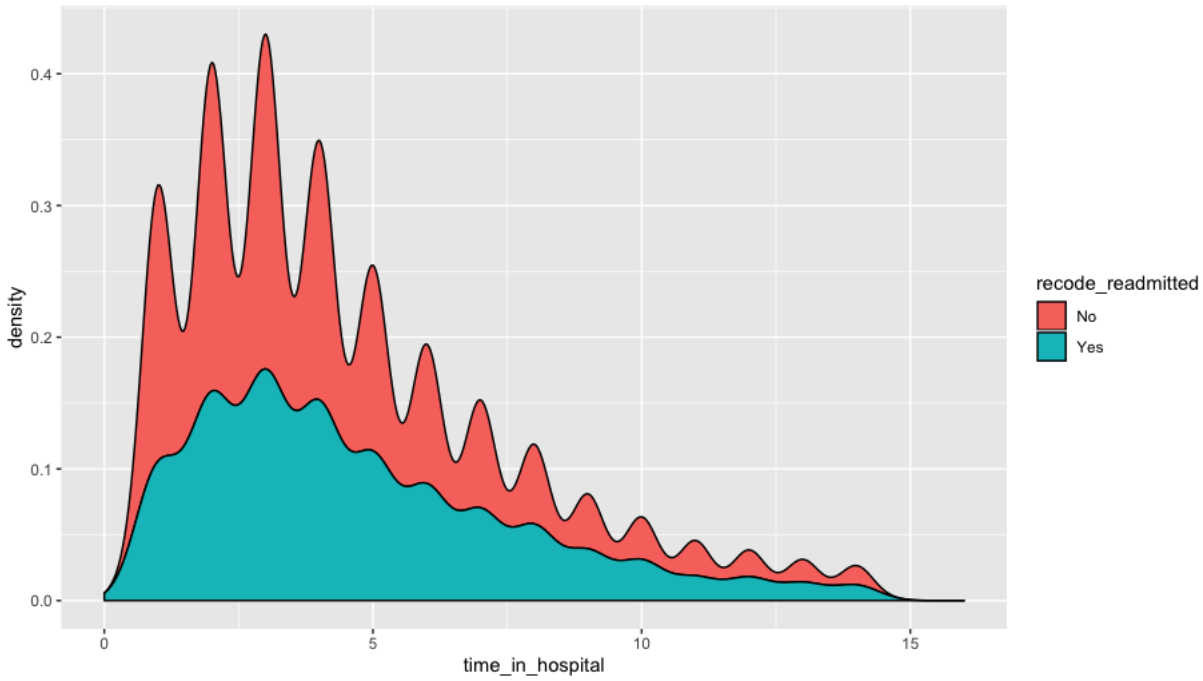

Figure 1: Frequency of Time Spent in the Hospital by Readmittance Status

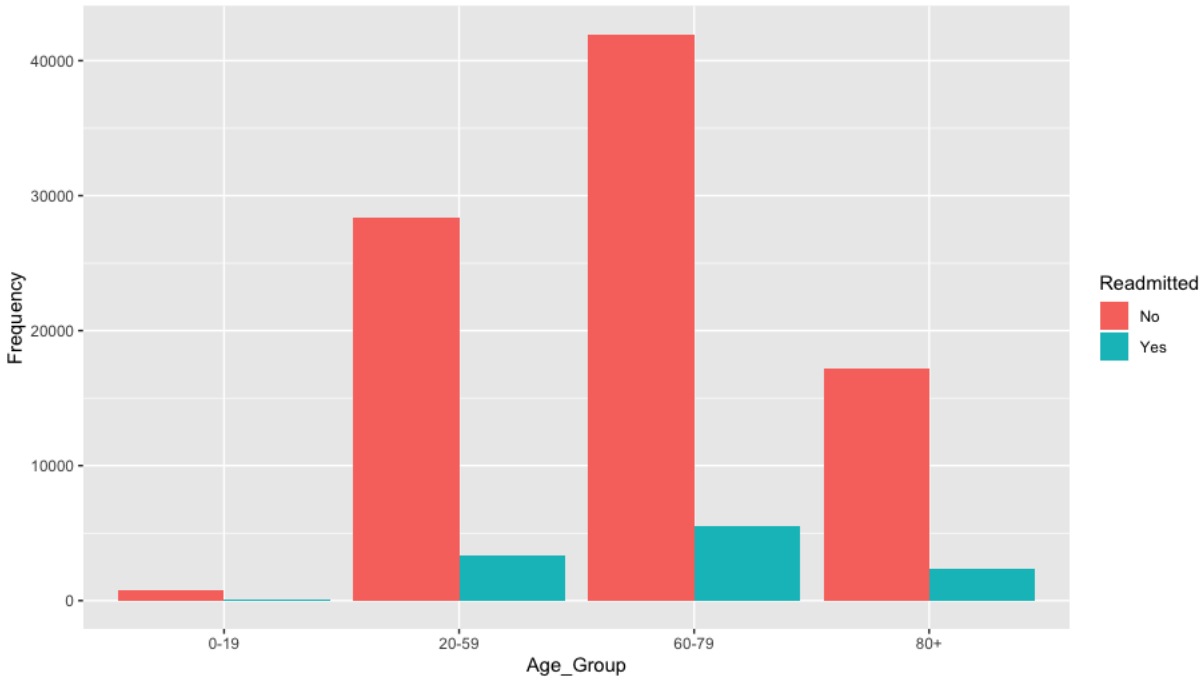

Figure 2: Frequency of Time Spent in the Hospital by Readmittance Status