Capstone 2 Hospital Readmission Prediction for Diabetes Patients- Milestone Report

The Centers for Medicare & Medicaid Services, CMS which is part of the Department of Health and Human Services (HHS) has created many programs to improve the quality of care of patient as the healthcare system moves toward value-based care. Hospital Readmission Reduction Program (HRRP), which is one of them, reduces reimbursement to hospitals with above average readmissions as penalization. This enforces the hospitals under this program to take steps to create interventions to provide additional assistance to patients with increased risk of readmission.

One patient population that is at increased risk of hospitalization and readmission is that of diabetes. Diabetes is a medical condition that affects approximately 1 in 10 patients in the United States. So it would be beneficial to use predictive modeling from data science to help identify patients with a likelihood of hospital readmission.

Dataset:

Data is available in UCI machine learning repository (link). The data consists of over 100000 hospital admissions from patients with diabetes from 130 US hospitals between 1998 and 008.

Data Wrangling:

The Centers for Medicare & Medicaid Services, CMS which is part of the Department of Health and Human Services (HHS) has created many programs to improve the quality of care of patient as the healthcare system moves toward value-based care. Hospital Readmission Reduction Program (HRRP), which is one of them, reduces reimbursement to hospitals with above average readmissions. For those hospitals which are currently penalized under this program, one solution is to create interventions to provide additional assistance to patients with increased risk of readmission. I propose to use predictive modeling from data science to help identify patients with a risk for hospital readmission.

Datasets that are available for this project are 1) Diabetic data with all the details of the patients getting admitted and 2) IDS Mapping that has mapping values for some of the columns from diabetic data

Data Visualization and Inferential Statistics

Exploratory data analysis is done to get a basic sense about the data. Since there are numerous variables, finding out any correlation will help in reducing dimensionality of data.

Below given visualization will help understand the dataset better.
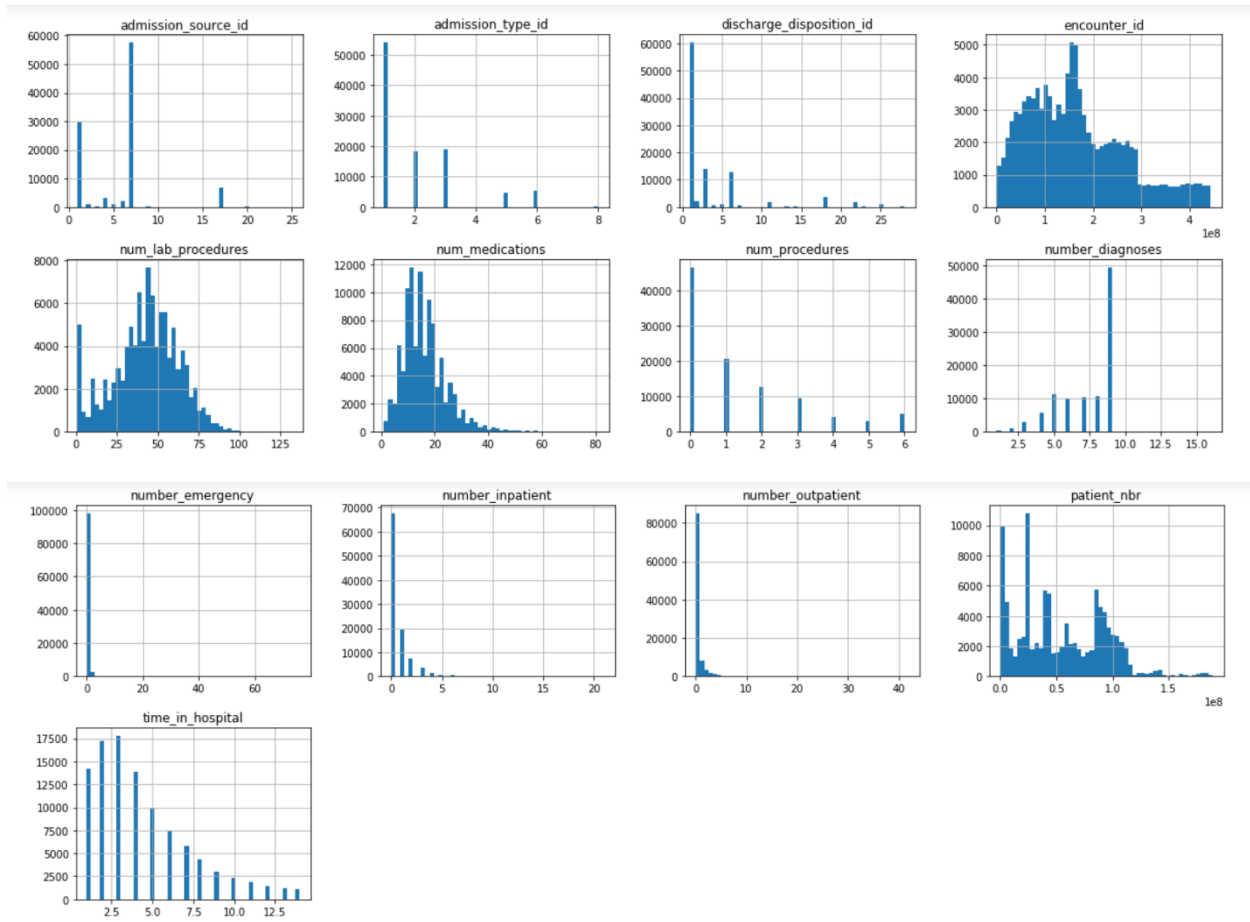
It is interesting to note that the number of lab procedures and number of medications taken, follow a bell curve in the population getting admitted to the hospital. It would be worth the effort to check for influence.
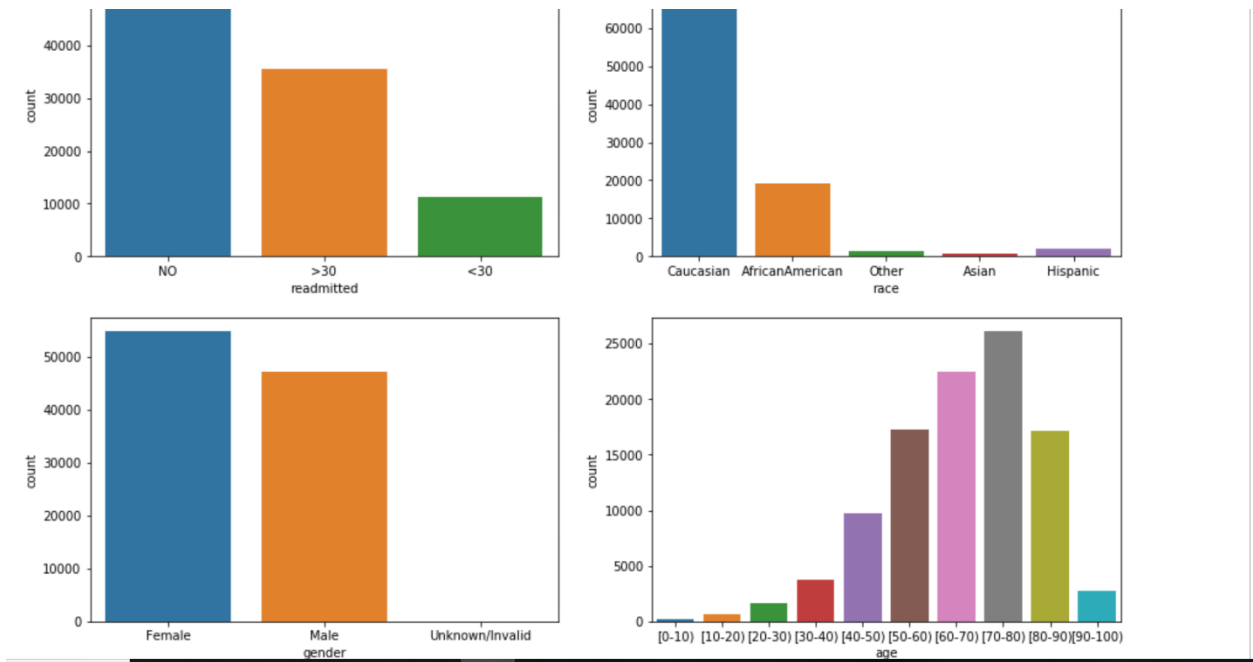
Readmission based on 'race' does not indicate that a specific race has more chances of readmission, as it could also be just reflective of ethnic distribution of population in that city. It is interesting to note that female population has more cases of re-admission. As expected, age pattern follows a bell curve.
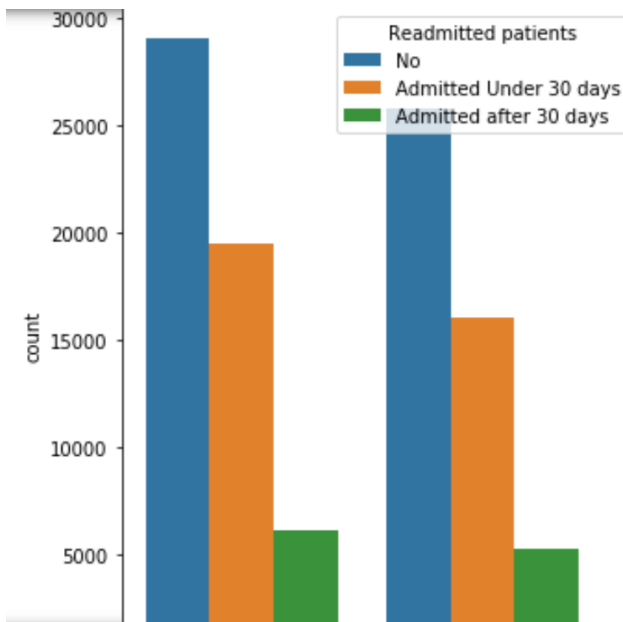
It can be noted that readmission categorization follows similar pattern across both the genders.

Readmissions by Age
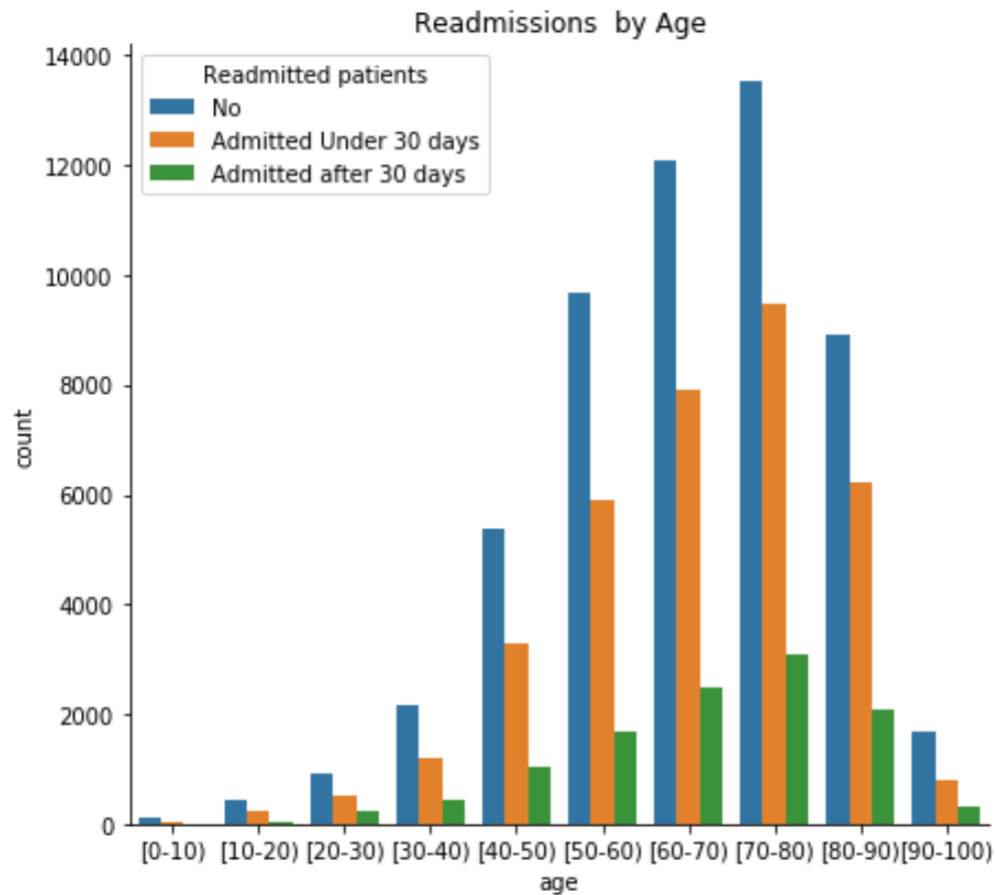
All the 3 categories 'Admitted Under 30 days', 'Admitted after 30 days' and ' No readmissions'
follow a bell curve peaking at 70-80s

From the exploratory data analysis and statistical inferences drawn, though some trends can be
drawn, no significant enough information is seen to proceed with the modeling. As next step,
feature importance can be used to order the feature variables according to their significance,
which would help the prediction.