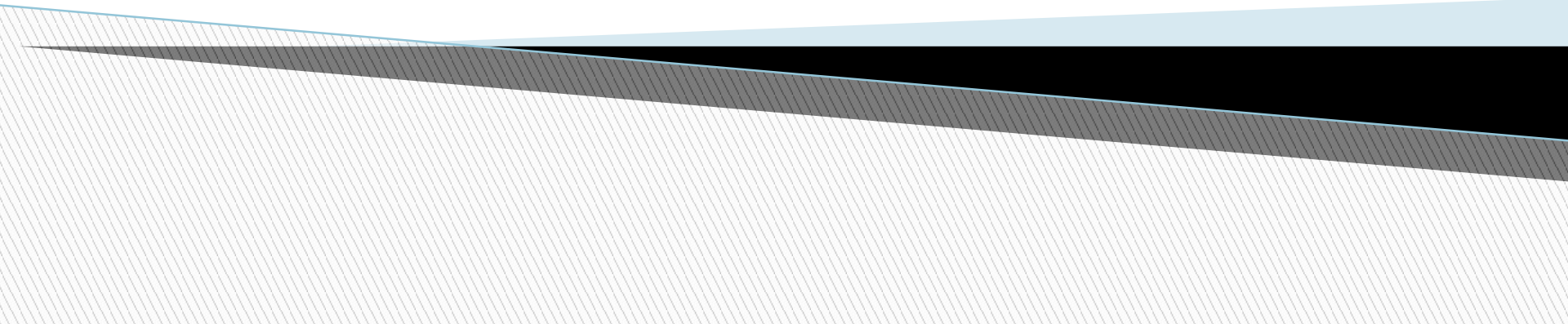# Prediction of diabetes readmission
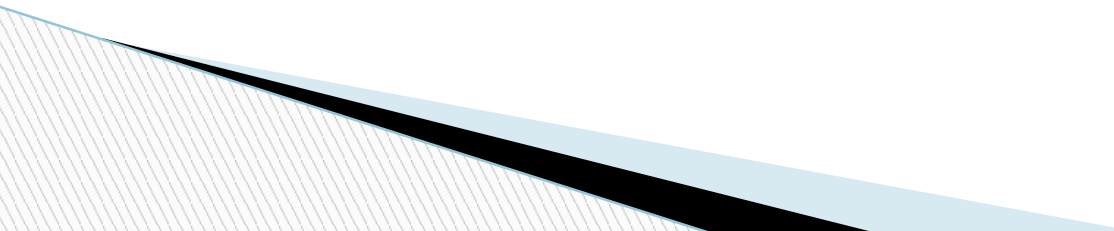
by
Shyam Sunder Reddy alijarla

Introduction

The data set represents 10 years (1999-2008) of clinical care at 130 US hospitals and integrated delivery networks. It includes over 50 features representing patient and hospital outcomes. Information was extracted from the database for encounters that satisfied the following criteria:

- It is an inpatient encounter (a hospital admission).
- It is a diabetic encounter, that is, one during which any kind of diabetes was entered to the system as a diagnosis.
- The length of stay was at least 1 day and at most 14 days.
- Laboratory tests were performed during the encounter.
- Medications were administered during the encounter.
- The data contains such attributes as patient number, race, gender, age, admission type, time in hospital, medical specialty of admitting physician, number of lab test performed, HbA1c test result, diagnosis, number of medication, diabetic medications, number of outpatient, inpatient, and emergency visits in the year before the hospitalization, etc.

Problem Definition

It is important to know if a patient will be readmitted in some hospital. The reason is that you can change the treatment, in order to avoid a readmission.

In this database, you have 3 different outputs:

- No readmission
- A readmission in less than 30 days
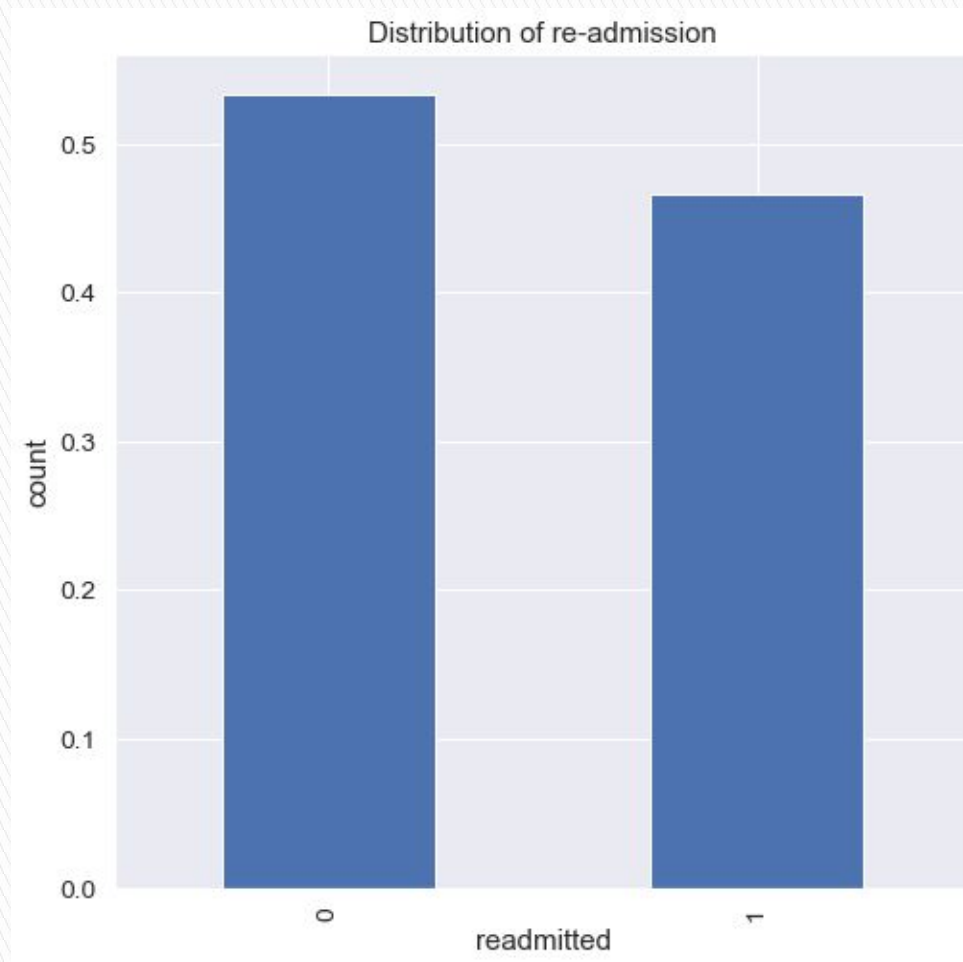- A readmission in more than 30 days

Features

- patient_nbr
- race
- gender
- age
- weight
- admission_type_id
- discharge_disposition_id
- admission_source_id
- time_in_hospital
- payer_code
- medical_specialty
- num_lab_procedures
- num_procedures

- num_medications
- number_outpatient
- number_emergency
- number_inpatient
- encounter_id
- diag_1
- diag_2
- diag_3
- number_diagnoses
- max_glu_serum
- A1Cresult
- metformin
- repaglinide

Features

- nateglinide
- chlorpropamide
- glimepiride
- acetohexamide
- glipizide
- glyburide
- tolbutamide
- pioglitazone
- rosiglitazone
- acarbose
- miglitol
- troglitazone
- tolazamide
- examide
- citoglipton
- insulin
- glyburide-metformin
- glipizide-metformin
- glimepiride-pioglitazone
- metformin-rosiglitazone
- metformin-pioglitazone
- change
- diabetesMed
- readmitted

# INSIGHTS

Distribution of re-admission

readmitted

1. The number of patients appearing for readmission in the hospitals is slightly less than the ones not appearing for the same.
2. The target variable is fairly balanced.
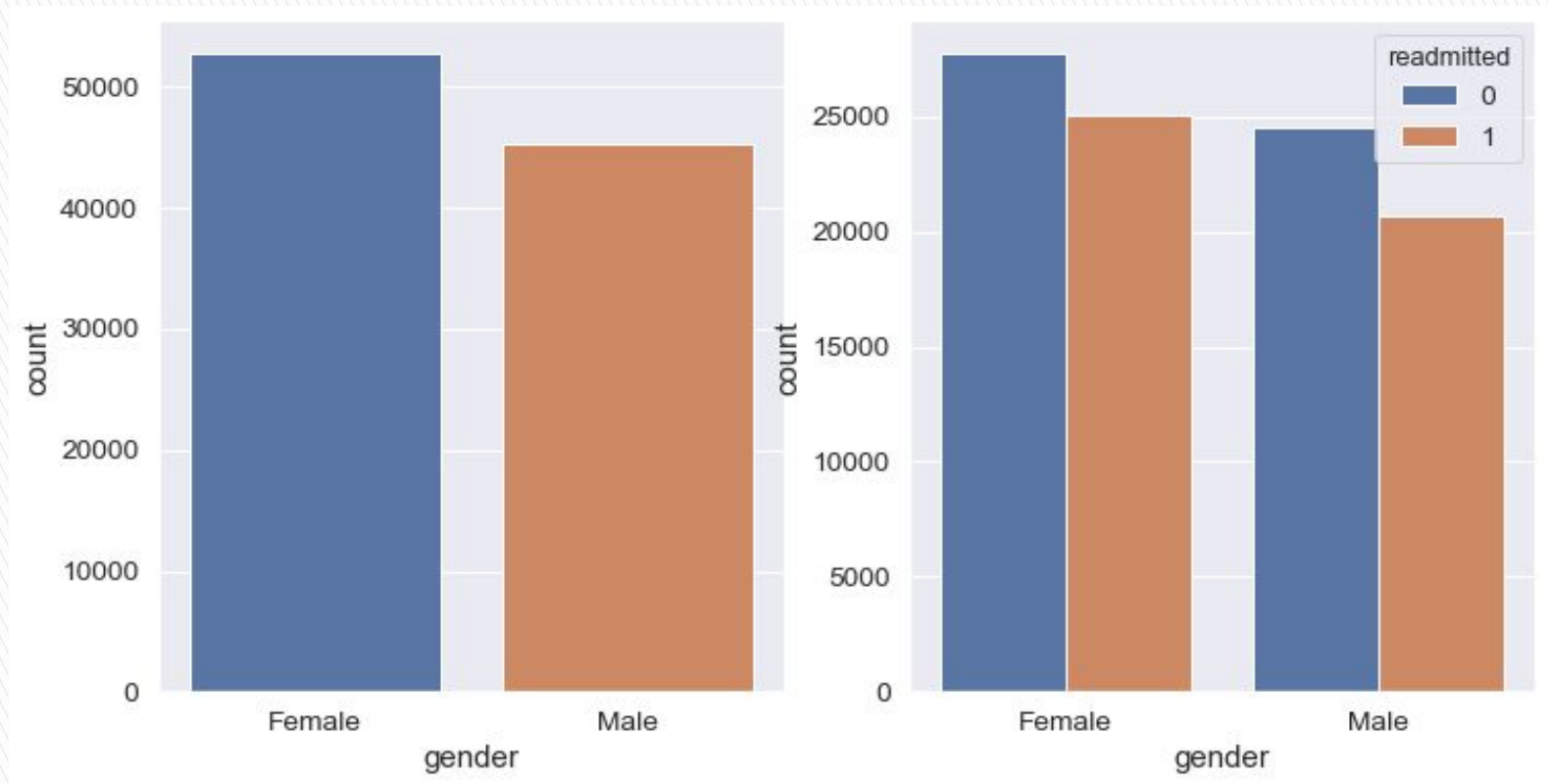
gender vs. readmitted

1. The number of females are more than the number of males.
2. Percentage of males getting readmitted is approximately 45% of total males.
3. Percentage of females getting readmitted is approximately 47% of total females.
4. Percentage of females getting readmitted are more than that of males.
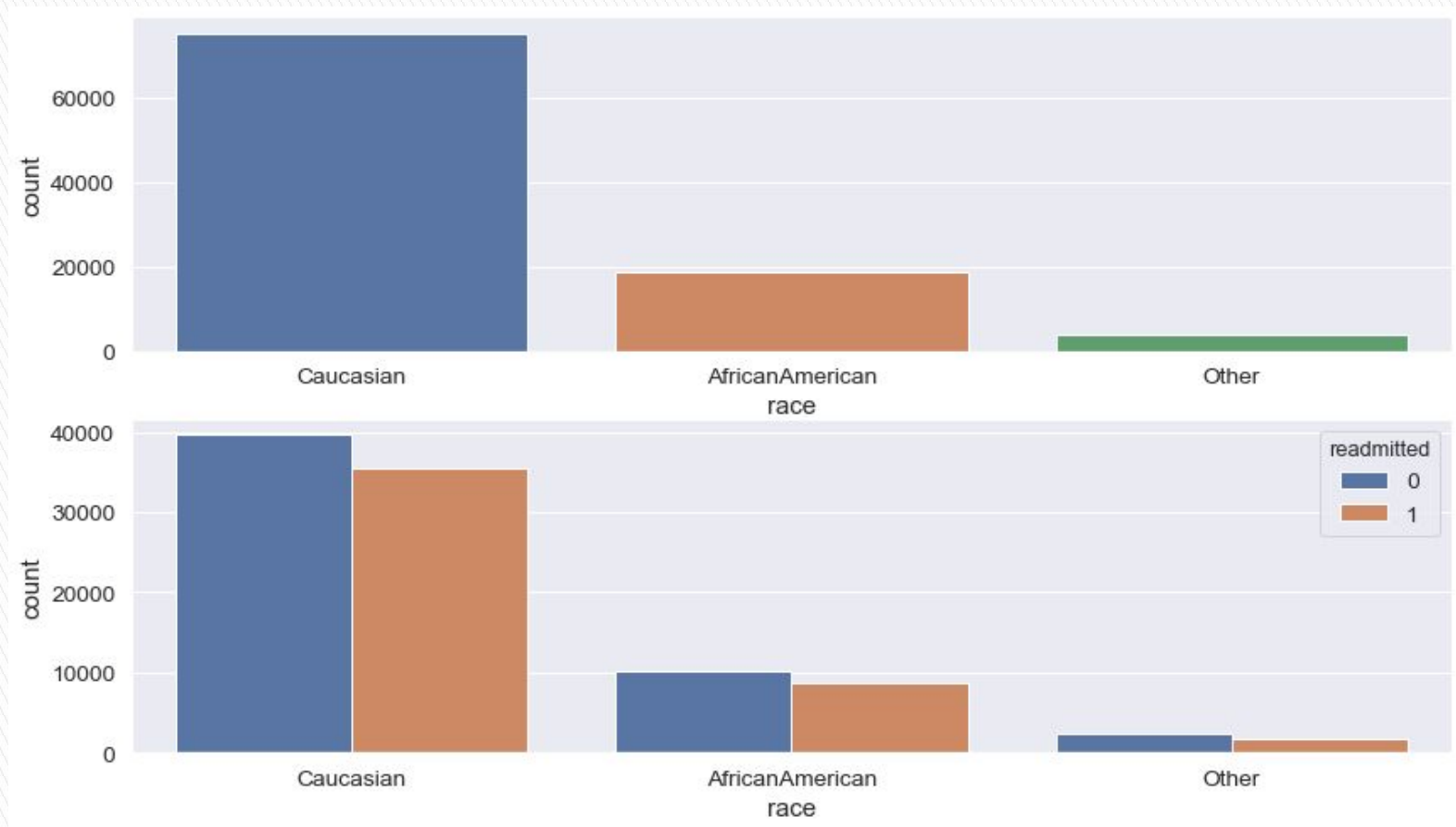
race vs. readmitted

1. The total number of Caucasians are significantly more than all other races.
2. The percentage of readmissions as well as non-readmissions are more for Caucasians than any other race.
3. The race of Asians are least in number.
4. Approximately 90% of the Caucasians gets readmitted whereas 83% of the AfricanAmerican gets readmitting.
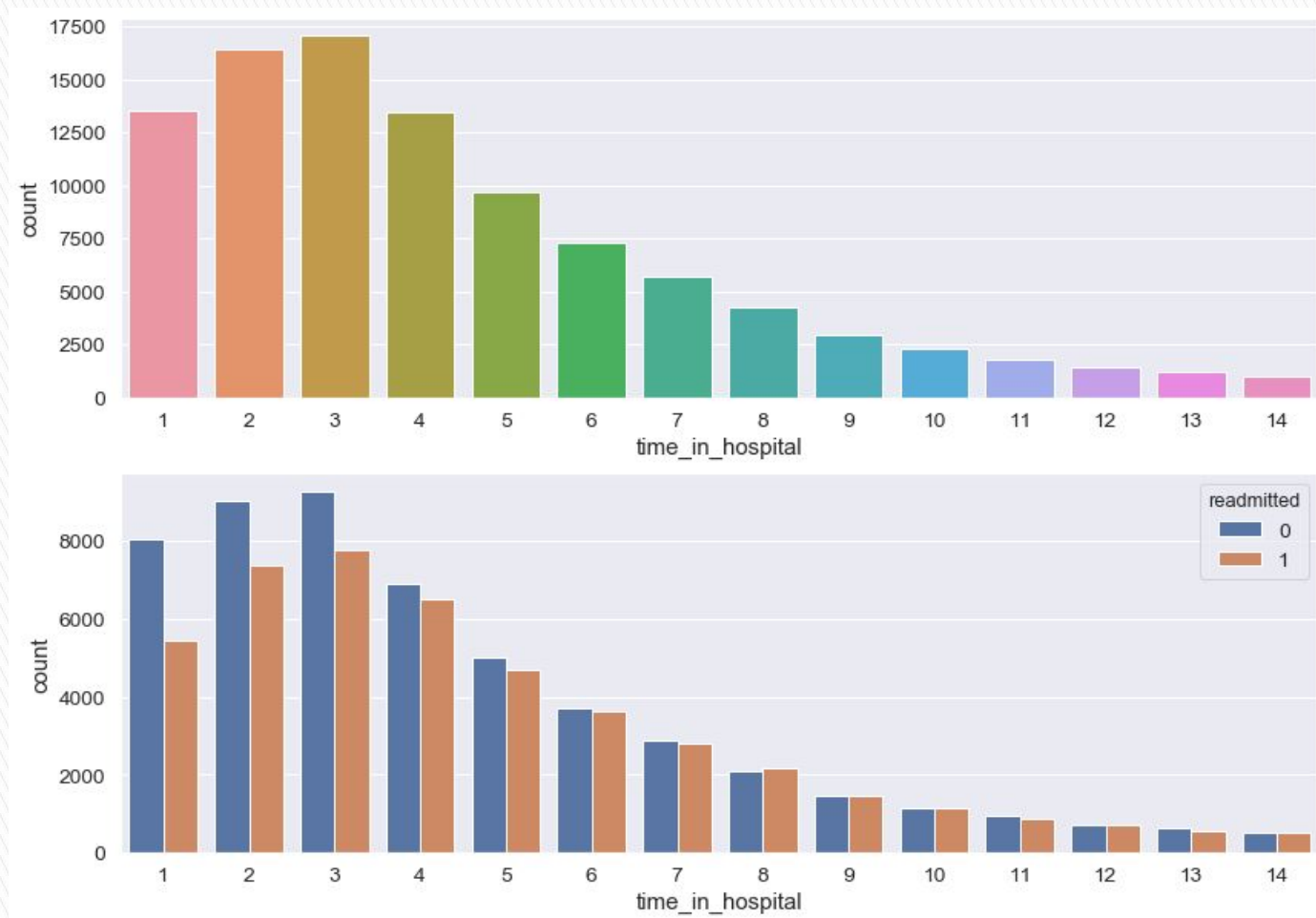
time_in_hospital vs. readmitted

1. The average number of days of admission is around 4 days.

2. It is most likely that people readmitting those who were admitted for more than 7days in there first admission.

3. Percentage of patients readmitting who have been admitted for 1day is very less compare to other patients.
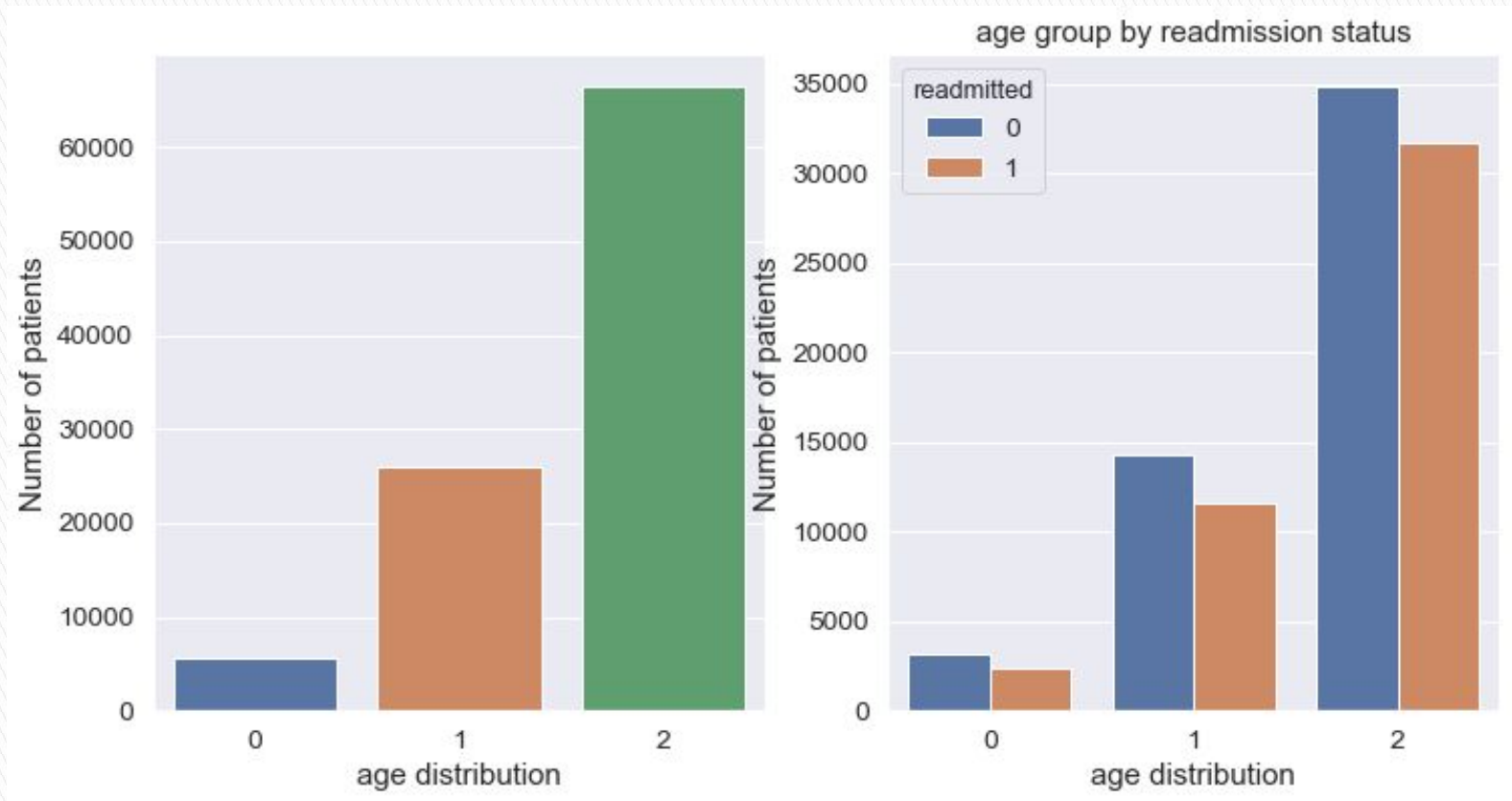
age vs. readmitted

1. Patients with age more than 60 years are more in number.
2. Around 69% of the patients above 60 years are readmitting.
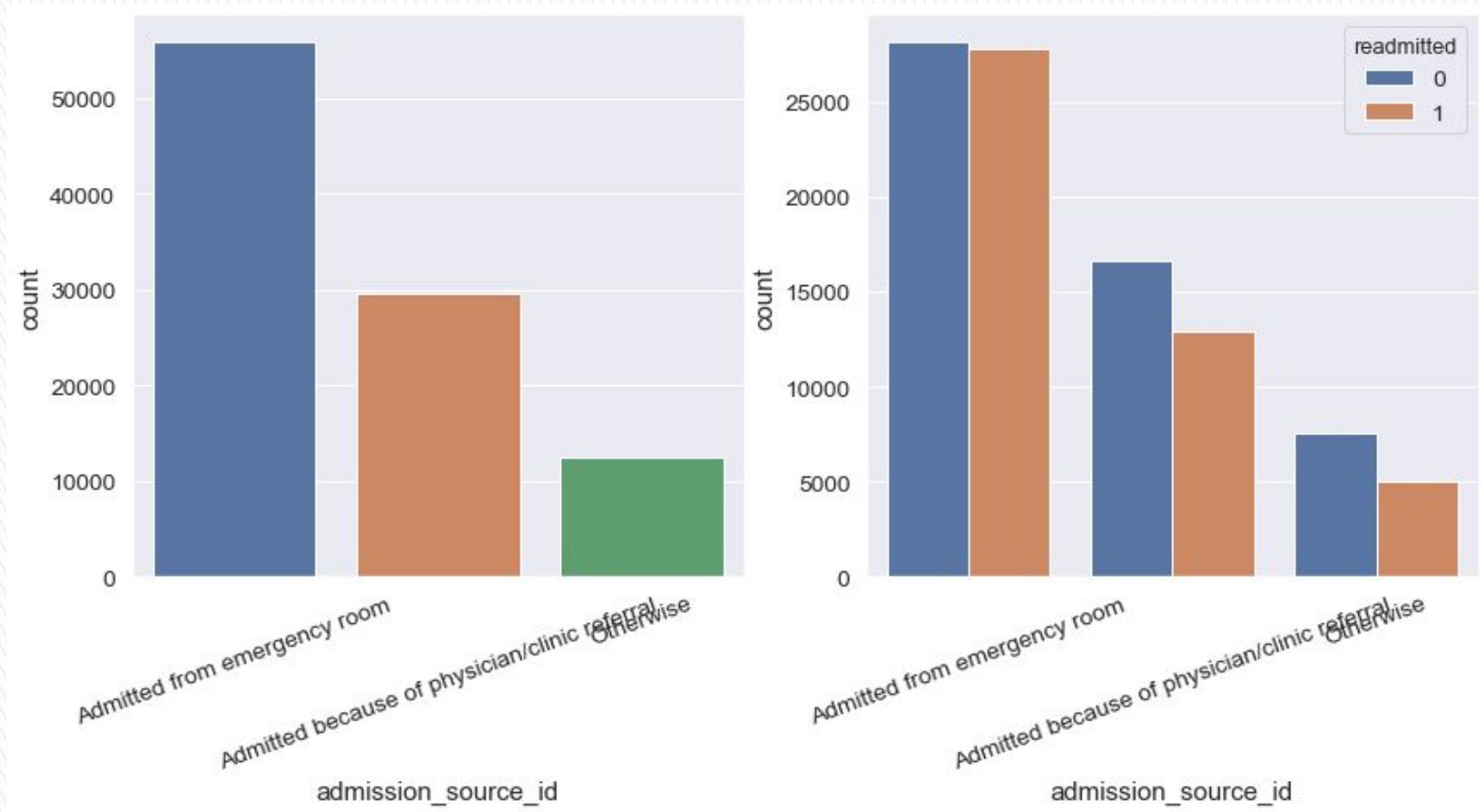
admission_source vs. readmitted

1. Patients Admitted from emergency room are more in number than those admitted from any other admission source.

2. Over 98.6% of the patients readmitted are those who admitted from emergency room.
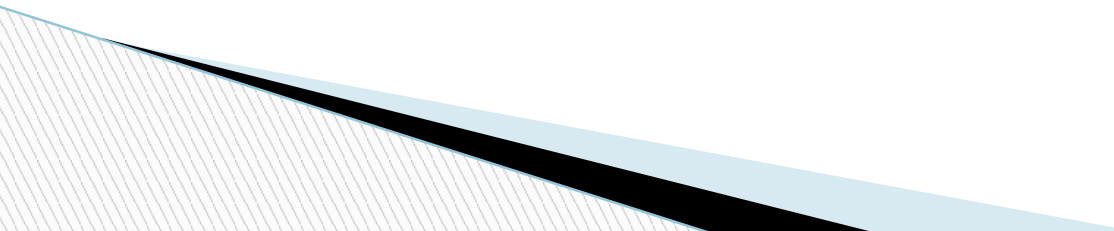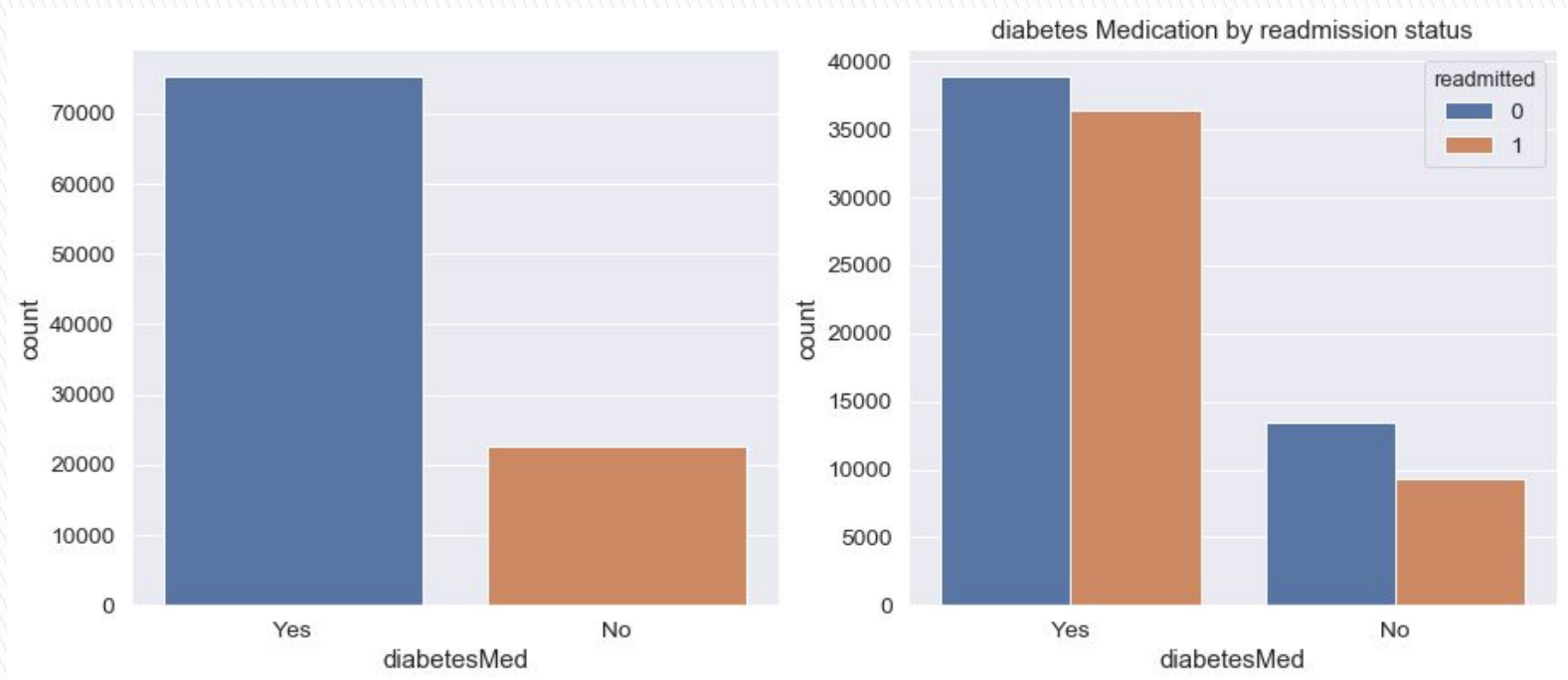
diabetesMed vs. readmitted

1. From the total data around 76% took medications and in that 95% are readmitted.
2. Those who are didn't take medications, around 67% are readmitted.

Challenges

- Missing Values
- Convert Data to Correct Types and Feature Engineering
- Significant features selection

Recommendations to firm

- Having records with the observed weight of the patients
- Maintaining quality of data
- Addition of Body Mass Index (BMI)

Basic Models

	model	train_accuracy	test_accracy
0	Logistic Regression	0.656462	0.648801
1	GaussianNB	0.640394	0.633519
2	K-Nearest Neighbors	0.787125	0.579605
3	Decision Tree	1.000000	0.543037
4	Random Forest	0.998941	0.605846
5	AdaBoost(Logistic)	0.654319	0.646897
6	AdaBoost(Random Forest)	1.000000	0.621497
7	AdaBoost(GaussianNB)	0.493916	0.498235
8	AdaBoostClassifier	0.662934	0.655118

```python
train = []
test = []
def model_fit(algo,X_train,y_train,X_test,y_test):
  algo.fit(X_train,y_train)
  y_train_pred=algo.predict(X_train)
  y_train_prob=algo.predict_proba(X_train)[:,1]
  print("auc score of train data",roc_auc_score(y_train,y_train_prob))
  train.append(roc_auc_score(y_train,y_train_prob))

  y_test_pred=algo.predict(X_test) #ML
  y_test_prob=algo.predict_proba(X_test)[:,1] #  in statistics
  print("auc score of test data",roc_auc_score(y_test,y_test_prob))
  test.append(roc_auc_score(y_test,y_test_prob))
  print("confusion_matrix",confusion_matrix(y_test,y_test_pred))
  print("classification_report")
  print(classification_report(y_test,y_test_pred))
  fpr,tpr,thresholds=roc_curve(y_test,y_test_prob)
  plt.plot(fpr,tpr)
  plt.plot(fpr,fpr)
  plt.xlabel("FPR")
  plt.ylabel("TPR")
  plt.title("ROC_AUC CURVE")
```

Thank you