REDUCING DIABETES READMISSIONS

CYNTHIA PEDRASA

MODULE 5 PROJECT

TABLE OF CONTENTS

PROBLEM:

- Who among the hospitalized Diabetic patients are at risk for 30-day hospital readmissions?
- How can we risk stratify and rank patient/encounters with the highest probability of getting readmitted within 30 days?
- Which features are important predictors of 30-day readmission in patients with Diabetes?
- Are HbA1c result, changes in treatment, and glycemic factors significant contributors to readmissions?

HOSPITAL READMISSIONS ARE ASSOCIATED WITH UNFAVORABLE PATIENT OUTCOMES AND HIGH FINANCIAL COSTS

- The Medicare Payment Advisory Commission (MedPAC) reported that in 2005, 17.6% of hospital admissions resulted in readmissions within 30 days of discharge, 11.3% within 15 days, and 6.2% within 7 days. https://www.ncsl.org/documents/health/medicare_hospital_readmissions_and_ppaca.pdf
- Diabetes is one of the most frequently treated condition in US Hospitals with 20.3% readmission rate

https://www.hcup-us.ahrq.gov/reports/statbriefs/sb153.pdf

FEDERAL & STATE PROGRAMS

The Centers for Medicare & Medicaid Services (CMS) includes readmission measures for conditions and procedures that significantly affect the lives of large numbers of Medicare patients.

Hospital Readmissions Reduction Program (HRRP) is a Medicare value-based purchasing program that reduces payments to hospitals with excess readmissions. The Affordable Care Act established the Hospital Readmissions Reduction Program to improve the quality of care while reducing costs. The program incentivizes hospitals to improve communication and care coordination efforts, and to better engage patients and caregivers, with respect to post-discharge planning.

Bundled Payments for Care Improvement Advanced (BPCIA). An alternative payment model to incentivize financial accountability, care redesign, data analysis and feedback, provider engagement, and patient engagement with bundled payments, care redesign activities, and accountability for performance on quality measures. This program links reimbursement or payment to the quality of care provided during a specific episode period

FEDERAL & STATE PROGRAMS

Comprehensive Care for Joint Replacement (CJR) model is an episode payment model that uses bundled payments for clinical episodes focused on lower extremity joint replacements

Medicare Shared Savings Program (MSSP) is a voluntary program that encourages groups of doctors, hospitals, and other health care providers to come together as an Accountable Care Organization (ACO) to give coordinated, high quality care to their Medicare patients. This program showed greater reductions in readmission rates.

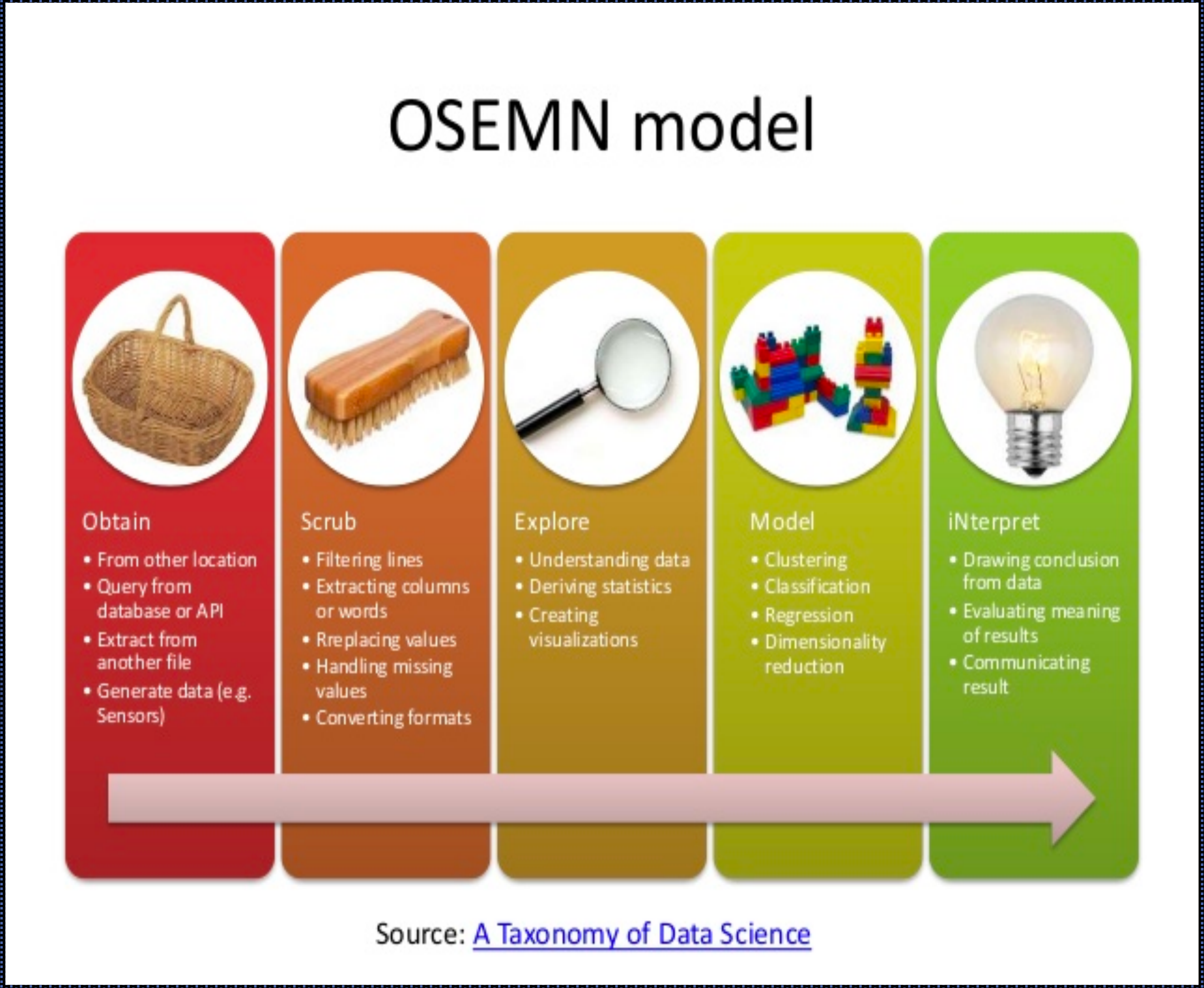
Delivery System Reform Incentive Payment (DSRIP) Program is the main mechanism by which New York State will implement the Medicaid Redesign Team (MRT) Waiver Amendment. DSRIP´s purpose is to restructure the health care delivery system by reinvesting in the Medicaid program, with the primary goal of reducing avoidable hospital use by 25% over 5 years. Up to $6.42 billion dollars are allocated to this program with payouts based upon achieving predefined results in system transformation, clinical management and population health.

BUSINESS DRIVERS

- Pinpoint patients with high readmission risk to reduce the occurrences of preventable hospital readmissions and avoidable admissions.
- Machine Learning out-predicts common approaches to readmission risk stratification by rendering more precise and complete views into patient predispositions by using patient characteristics and other comorbidity index computations in enabling the patient level predictions
- Improve resource utilization and increase operational efficiency
- Improve hospital rating based on lower readmission rate and increased patient satisfaction
- A positive financial return is expected from the readmission and avoidable admission reduction rate. Revenue generation by decreasing the hospital's excess readmission ratio that reduces payments for hospitals whose 30-day readmission rates are high relative to other facilities

DATA SCIENCE PROCESS

- **Obtain**-Requirements and information gathering on the problem.
- **Scrub-** Pre-processing our data (removing nulls, outliers, normalization, feature selection)
- **Explore**-Understand the cohort characteristics and impactful predictors
- **Model**-Build and tune the model
- **iNterpret** - and communicate results to stakeholders.

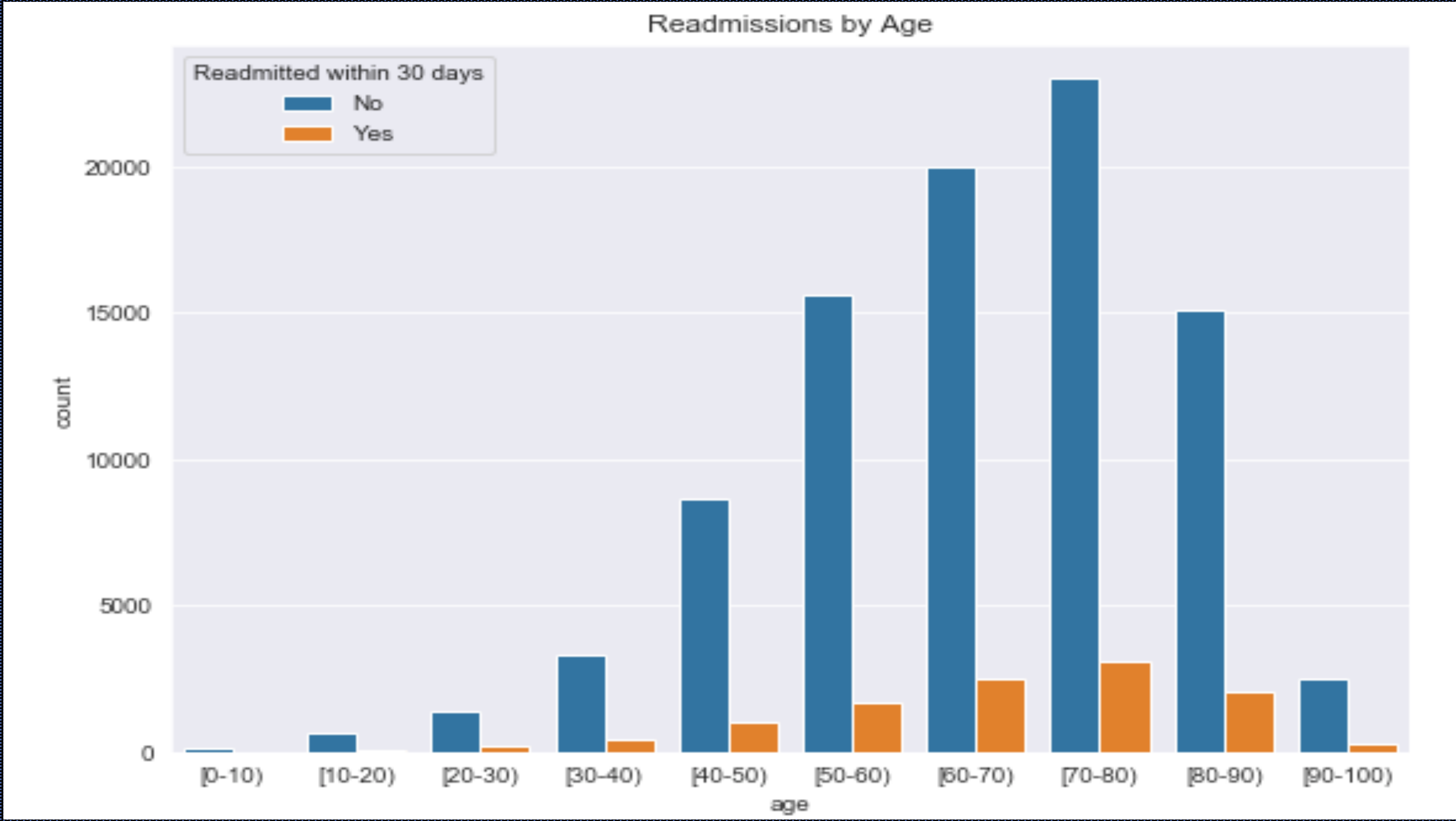


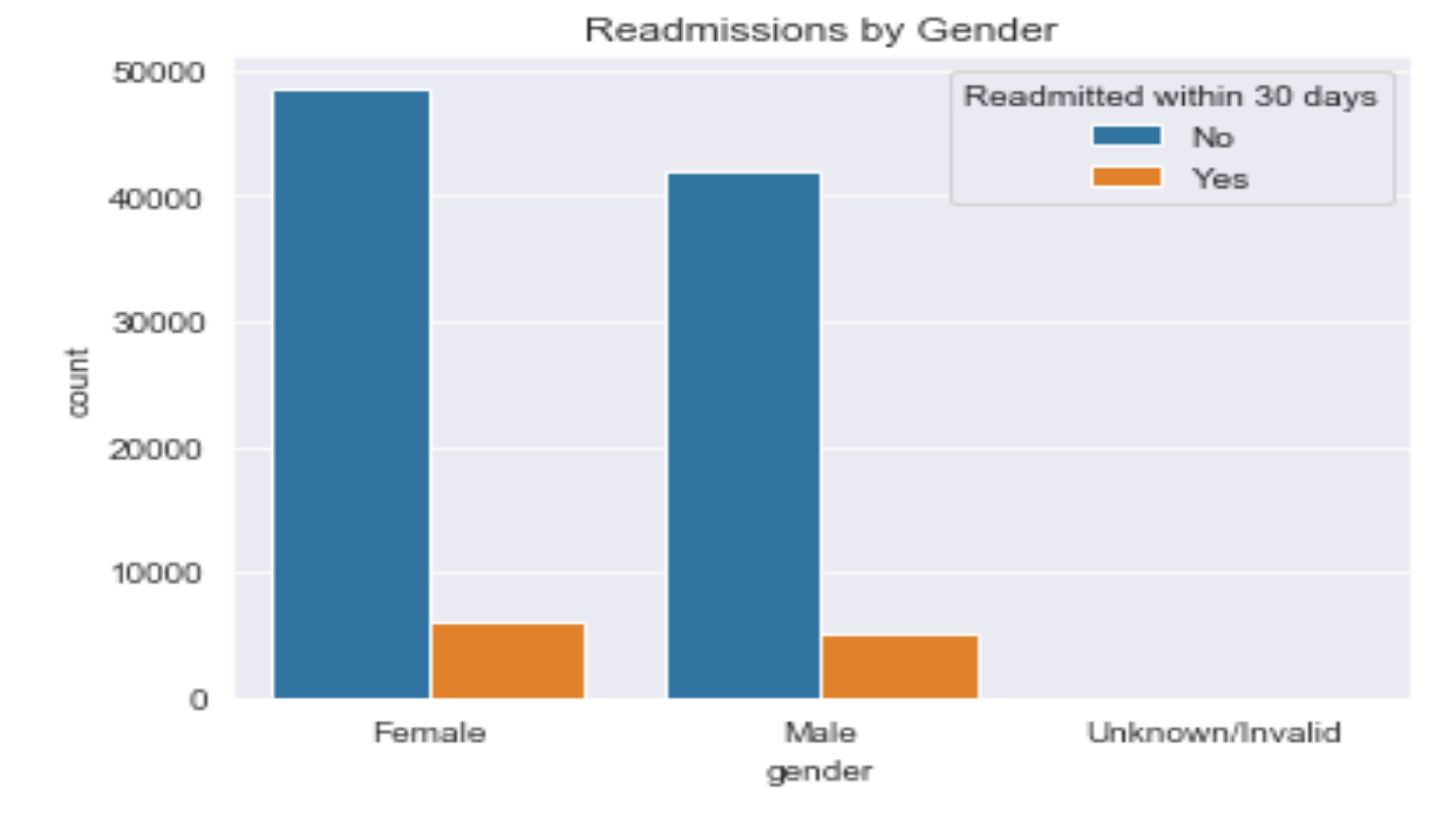
HOSPITAL READMISSIONS DIABETES DATA SET

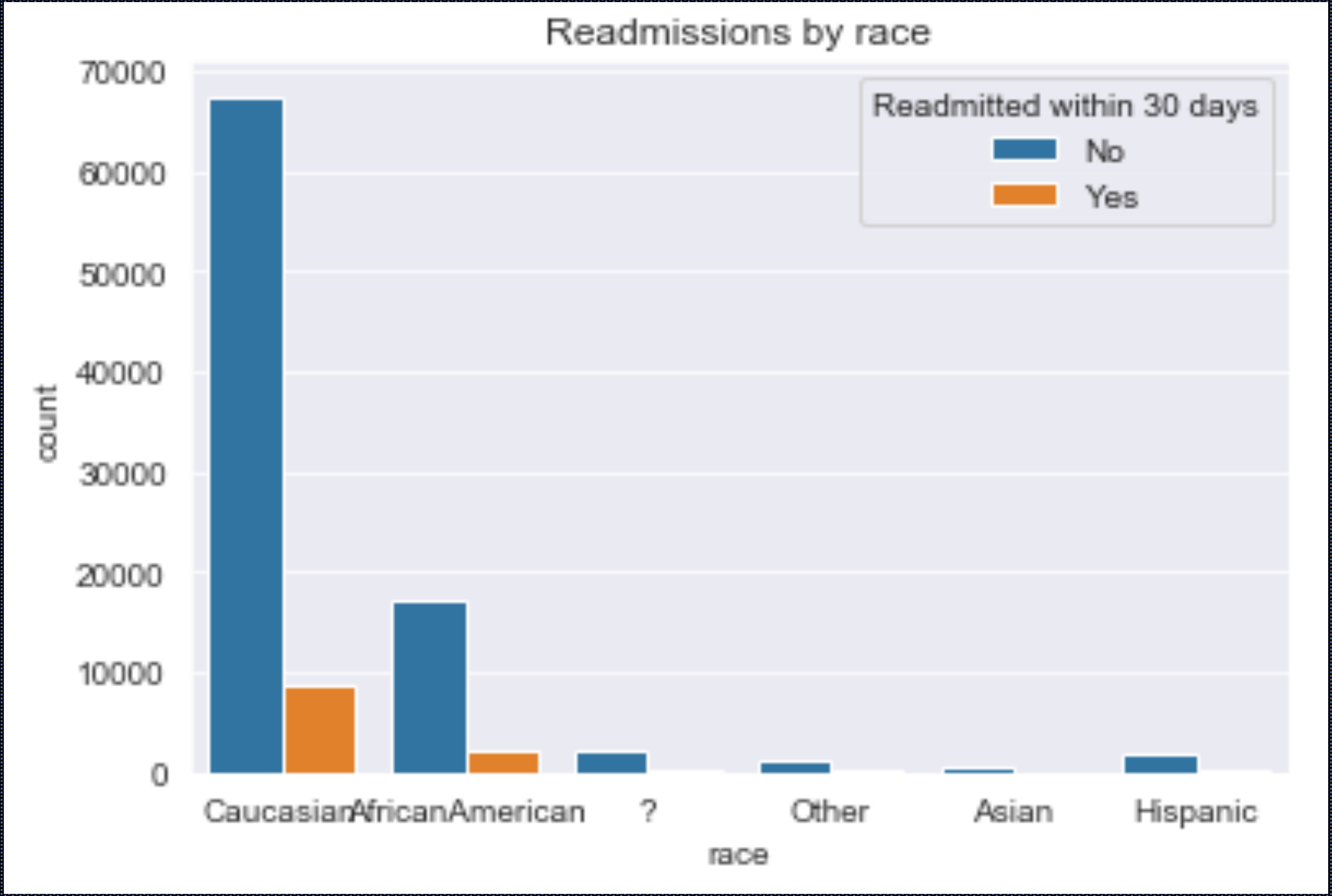
The dataset represents 10 years (1999-2008) of clinical care at 130 US hospitals and integrated delivery networks. It includes over 50 features representing patient and hospital outcomes. Information was extracted from the database for encounters that satisfied the following criteria.

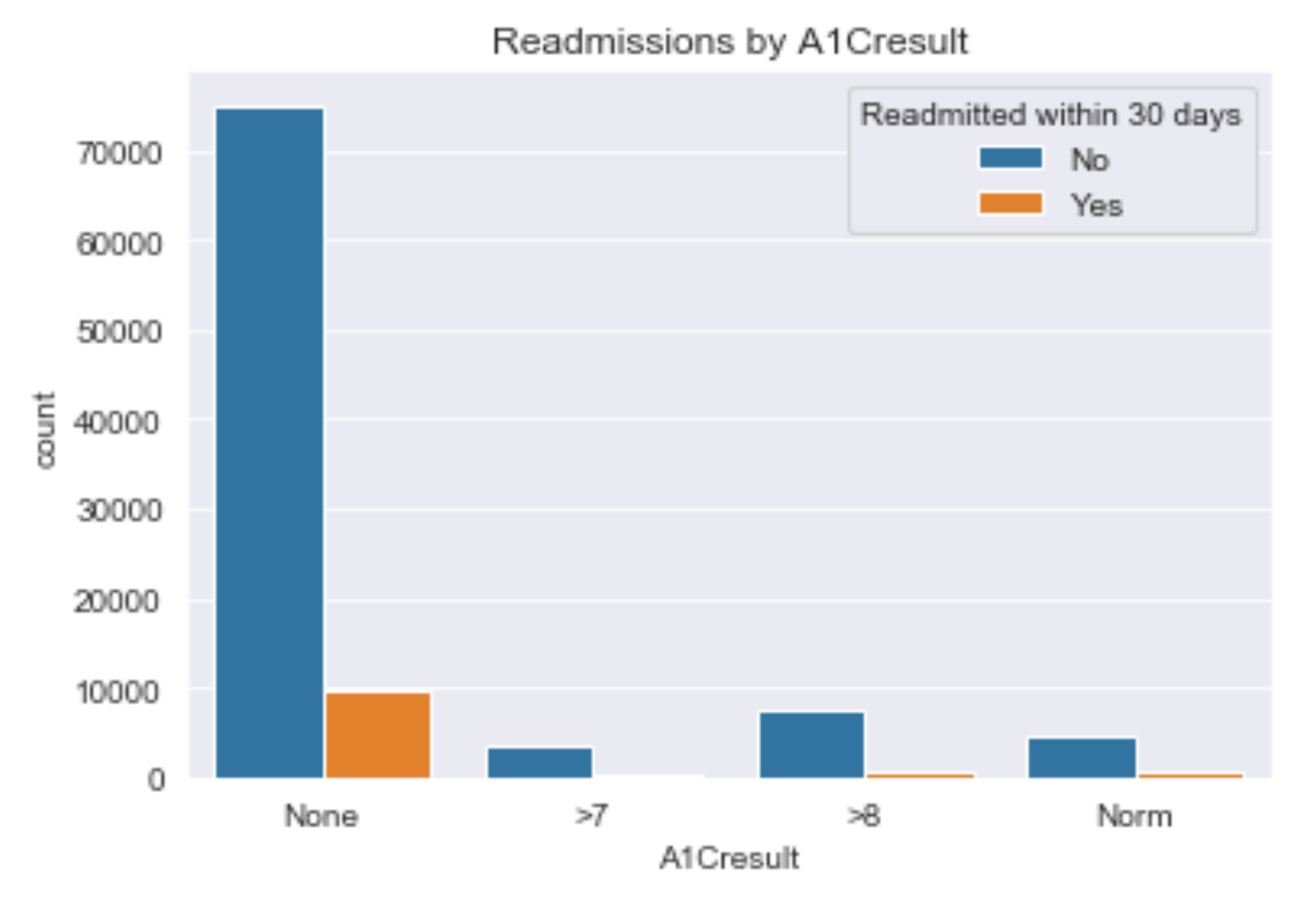
(1) It is an inpatient encounter (a hospital admission).
(2) It is a diabetic encounter, that is, one during which any kind of diabetes was entered to the system as a diagnosis.
(3) The length of stay was at least 1 day and at most 14 days.
(4) Laboratory tests were performed during the encounter.
(5) Medications were administered during the encounter.

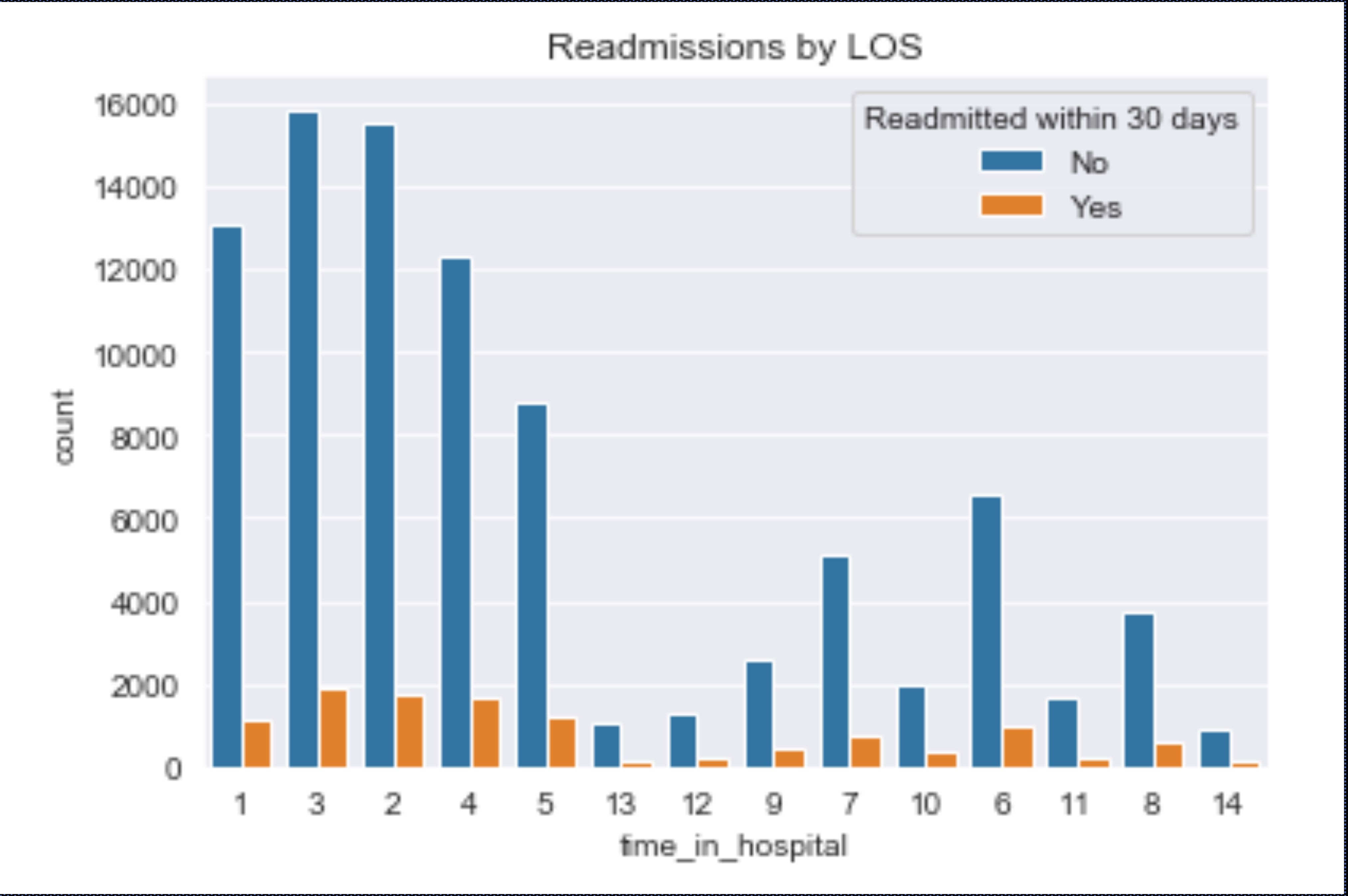
https://archive.ics.uci.edu/ml/datasets/diabetes+130-us+hospitals+for+years+1999-2008


Readmissions by Age
Readmitted within 30 days
No
Yes
count
20000
15000
10000
5000
0
[0-10)
[10-20)
[20-30)
[30-40)
[40-50)
[50-60)
[60-70)
[70-80)
[80-90)
[90-100)
age

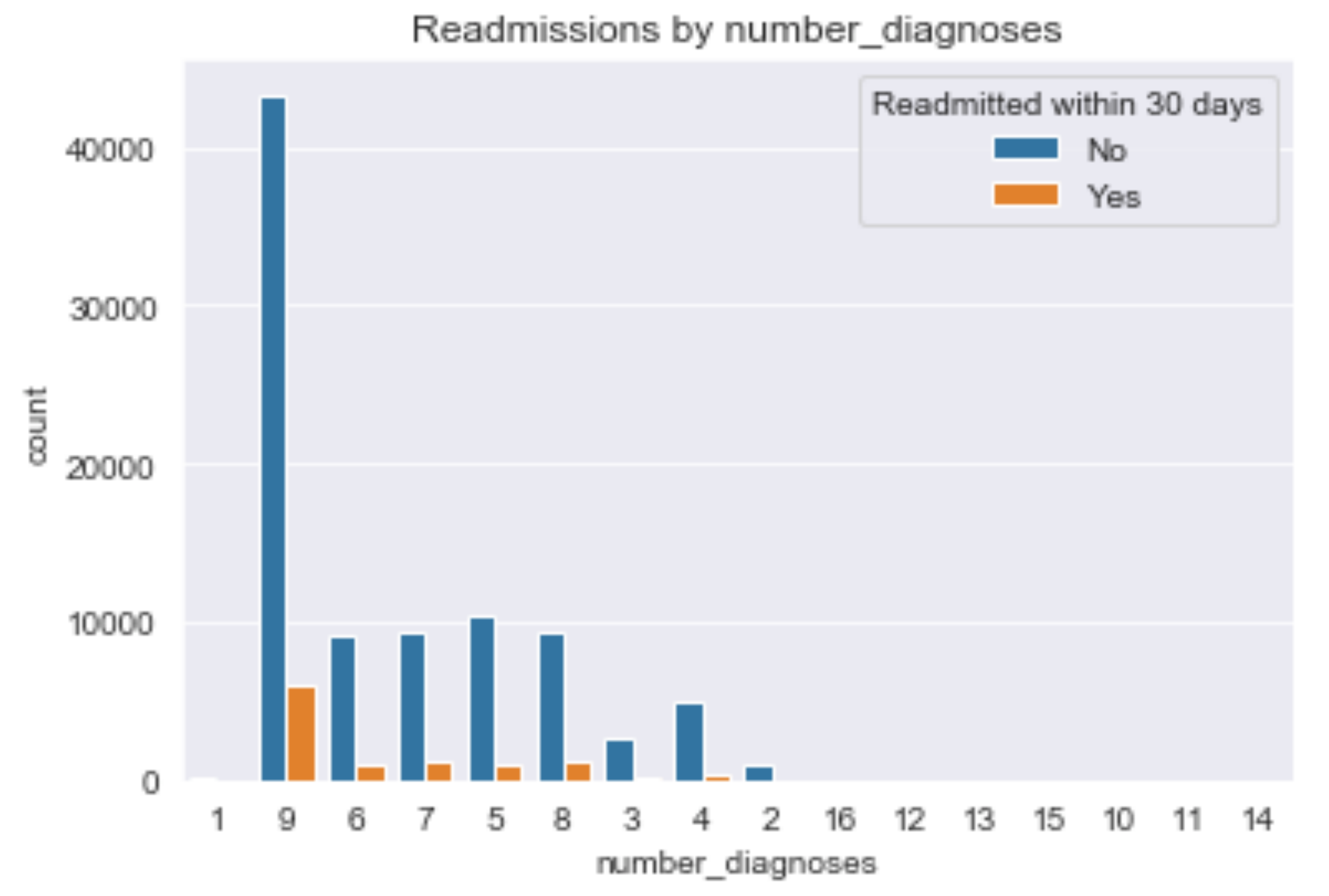

Readmissions by Gender
Readmitted within 30 days
No
Yes
50000
40000
30000
20000
10000
0
count
Female
Male
Unknown/Invalid
gender


Readmissions by race
70000
60000
50000
40000
30000
20000
10000
0
count
Readmitted within 30 days
No
Yes
Caucasian
AfricanAmerican
?
Other
Asian
Hispanic
race


Readmissions by A1Cresult
Readmitted within 30 days
No
Yes
70000
60000
50000
40000
30000
20000
10000
0
count
None
>7
>8
Norm
A1Cresult



Readmissions by LOS
Readmitted within 30 days
No
Yes
count
16000
14000
12000
10000
8000
6000
4000
2000
0
1
3
2
4
5
13
12
9
7
10
6
11
8
14
time_in_hospital


Readmissions by number_diagnoses
Readmitted within 30 days
No
Yes
count
40000
30000
20000
10000
0
1
9
6
7
5
8
3
4
2
16
12
13
15
10
11
14
number_diagnoses

HISTORICAL DATA EXPLORATION

- Frequency of Attributes contributing to readmission
- Characteristics of readmitted patients
- Prevalence of Readmissions
- Historical Readmission Rates
- Establish baseline/Benchmark
- Build Models - train and test

DATA SET ATTRIBUTES

- Encounter ID - Unique identifier of an encounter
- Patient number - Unique identifier of a patient
- Race Values: Caucasian, Asian, African American, Hispanic, and other
- Gender Values: male, female, and unknown/invalid
- Age Grouped in 10-year intervals: 0, 10, (10, 20), …, 90, 100)
- Weight - Weight in pounds
- Admission type - Integer identifier corresponding to 9 distinct values, for example, emergency, urgent, elective, newborn, and not available
- Discharge disposition - Integer identifier corresponding to 29 distinct values, for example, discharged to home, expired, and not available
- Admission source - Integer identifier corresponding to 21 distinct values, for example, physician referral, emergency room, and transfer from a hospital

DATA SET ATTRIBUTES

- Time in hospital - Integer number of days between admission and discharge
- Payer code - Integer identifier corresponding to 23 distinct values, for example, Blue Cross/Blue Shield, Medicare, and self-pay Medical
- Medical specialty - Integer identifier of a specialty of the admitting physician, corresponding to 84 distinct values, for example, cardiology, internal medicine, family/general practice, and surgeon
- Number of lab procedures - Number of lab tests performed during the encounter
- Number of procedures - Numeric Number of procedures (other than lab tests) performed during the encounter
- Number of medications - Number of distinct generic names administered during the encounter

DATA SET ATTRIBUTES

- Number of outpatient visits - visits of the patient in the year preceding the encounter
- Number of emergency visits - Number of emergency visits of the patient in the year preceding the encounter
- Number of inpatient visits - Number of inpatient visits of the patient in the year preceding the encounter
- Diagnosis 1 The primary diagnosis (coded as first three digits of ICD9);
- Diagnosis 2 Secondary diagnosis (coded as first three digits of ICD9);
- Diagnosis 3 Additional secondary diagnosis (coded as first three digits of ICD9); Number of diagnoses Number of diagnoses entered to the system 0%

DATA SET ATTRIBUTES

- Glucose serum test result - Indicates the range of the result or if the test was not taken. Values: “>200,” “>300,” “normal,” and “none” if not measured
- A1c test result - Indicates the range of the result or if the test was not taken. Values: “>8” if the result was greater than 8%, “>7” if the result was greater than 7% but less than 8%, “normal” if the result was less than 7%, and “none” if not measured.
- Change of medications - Indicates if there was a change in diabetic medications (either dosage or generic name). Values: “change” and “no change”
- Diabetes medications - Indicates if there was any diabetic medication prescribed. Values: “yes” and “no”

DATA SET ATTRIBUTES

- 24 features for medications: metformin, repaglinide, nateglinide, chlorpropamide, glimepiride, acetohexamide, glipizide, glyburide, tolbutamide, pioglitazone, rosiglitazone, acarbose, miglitol, troglitazone, tolazamide, examide, sitagliptin, insulin, glyburide-metformin, glipizide-metformin, glimepiride- pioglitazone, metformin-rosiglitazone, and metformin- pioglitazone, the feature indicates whether the drug was prescribed or there was a change in the dosage. Values: “up” if the dosage was increased during the encounter, “down” if the dosage was decreased, “steady” if the dosage did not change, and “no” if the drug was not prescribed
- Dependent/Outcome Variable
- Readmitted Days to inpatient readmission. Values: “<30” if the patient was readmitted in less than 30 days, “>30” if the patient was readmitted in more than 30 days, and “No” for no record of readmission

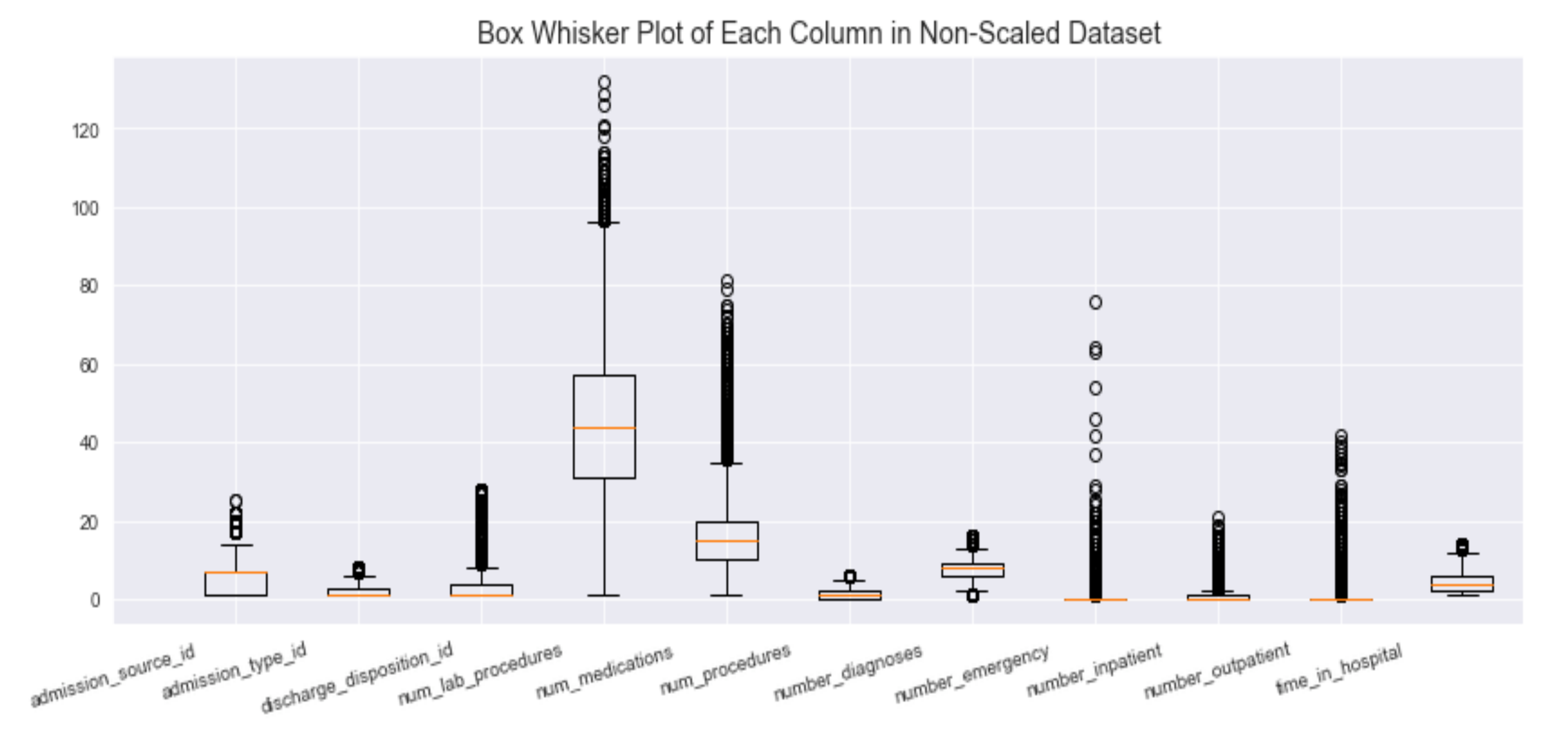
DATA PRE-PROCESSING

- Missing values
- Aggregations
- Cardinality – attributes with same values
- Normality, Multicollinearity checks
- Outliers removal
- ICD-10 mapping
- Feature creation
- Feature Subset selection
- Standardization
- Sampling

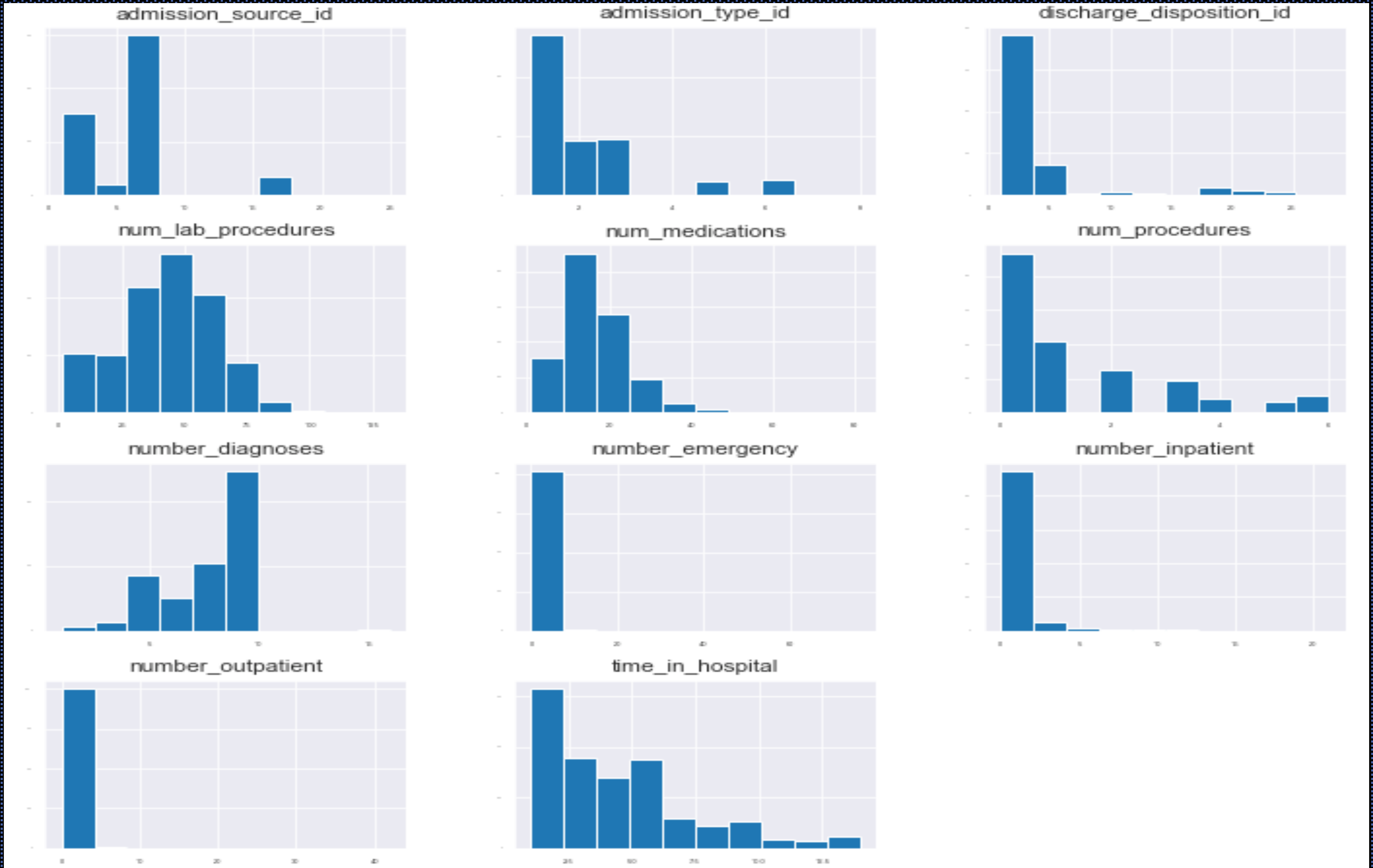
DATA PRE-PROCESSING

- Target impactable patients with diabetes who are at risk for 30-day hospital readmissions
- Excluded Discharge disposition:
  - deceased /transfers/ hospice
  - left against medical advice
- Removed duplicate patients
  - Index visit

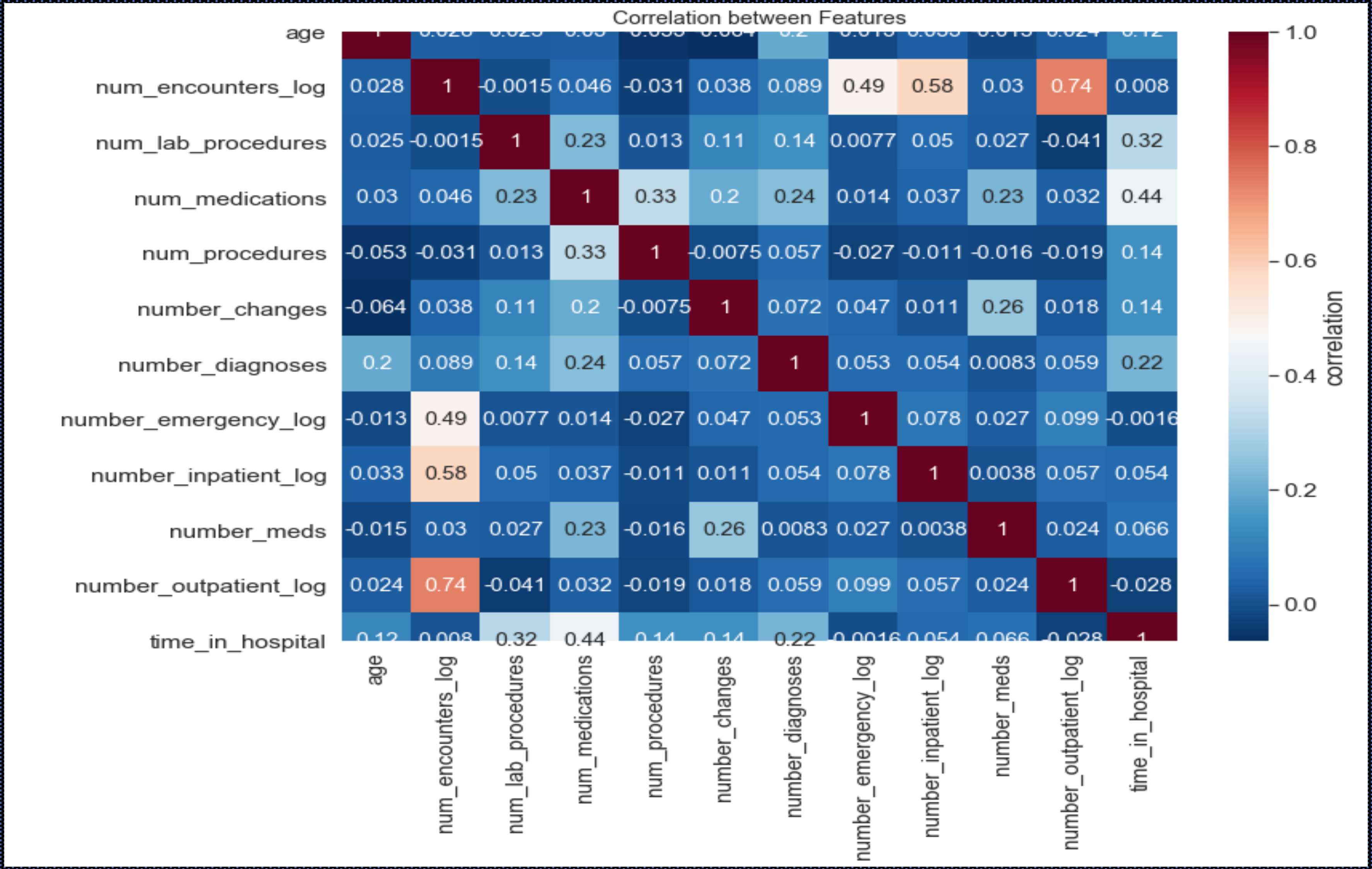
DATA VISUALIZATION



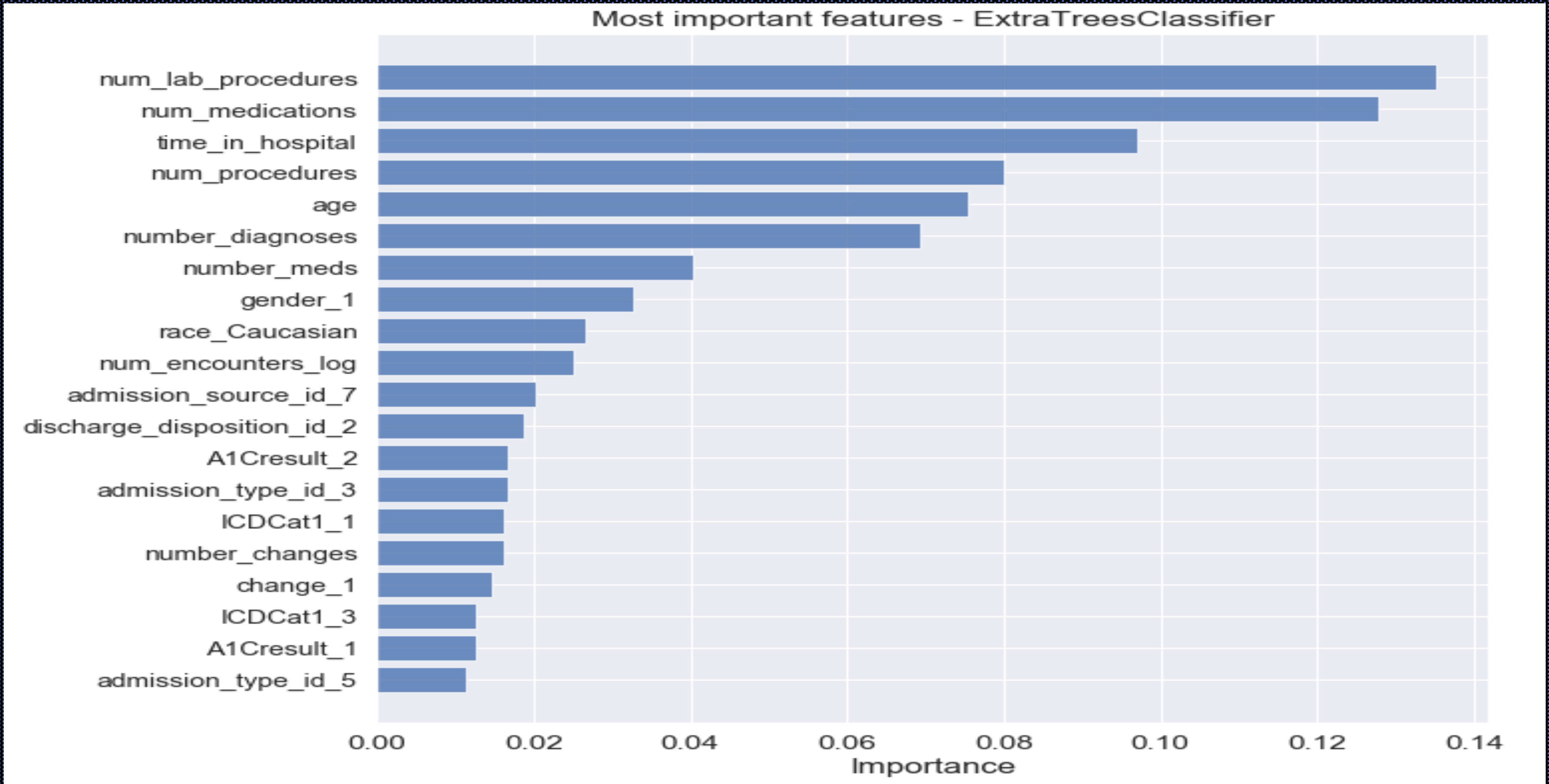
NORMALITY CHECK



MULTICOLLINEARITY



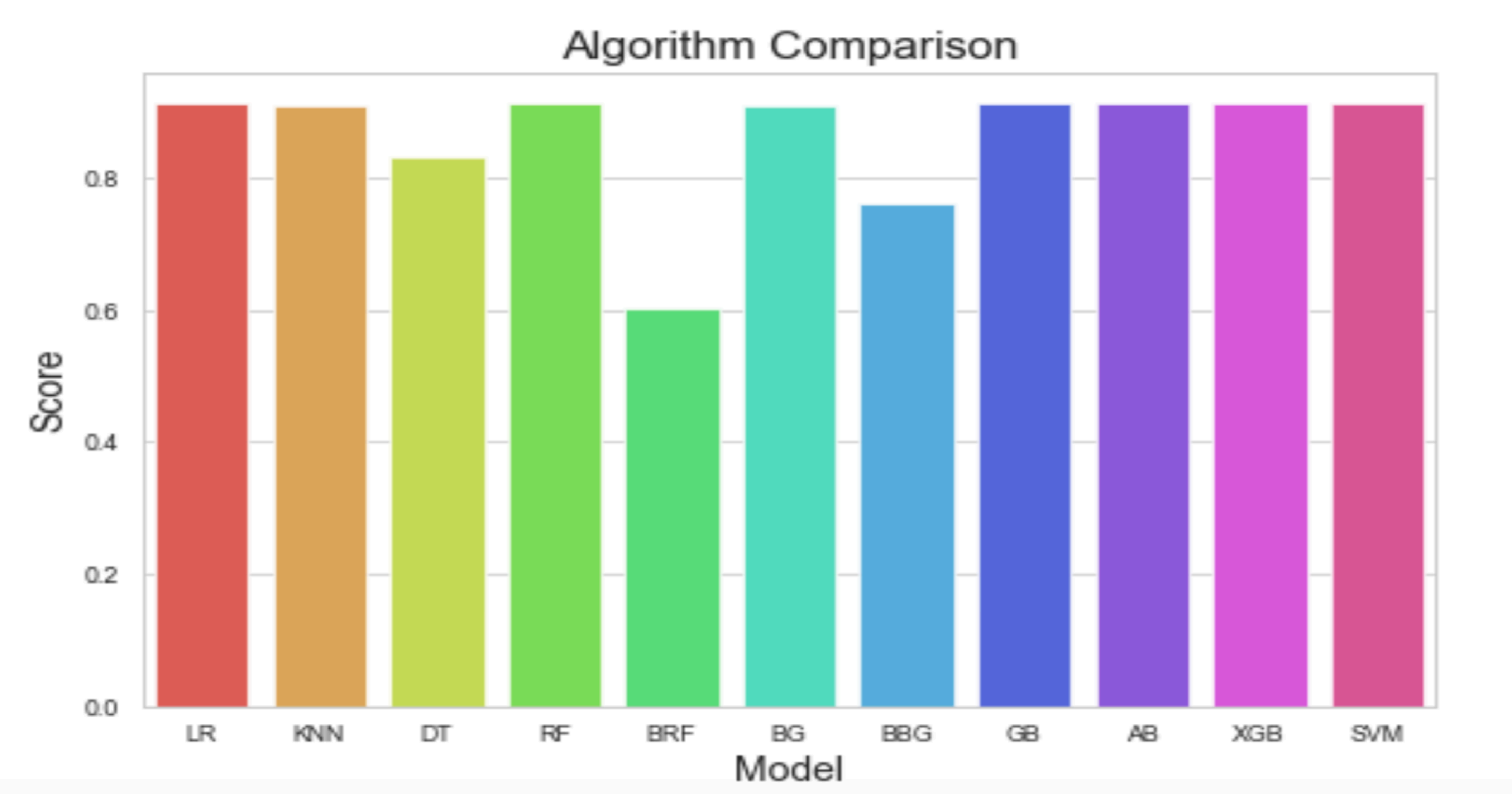
FEATURE SELECTION



PERFORMANCE METRICS

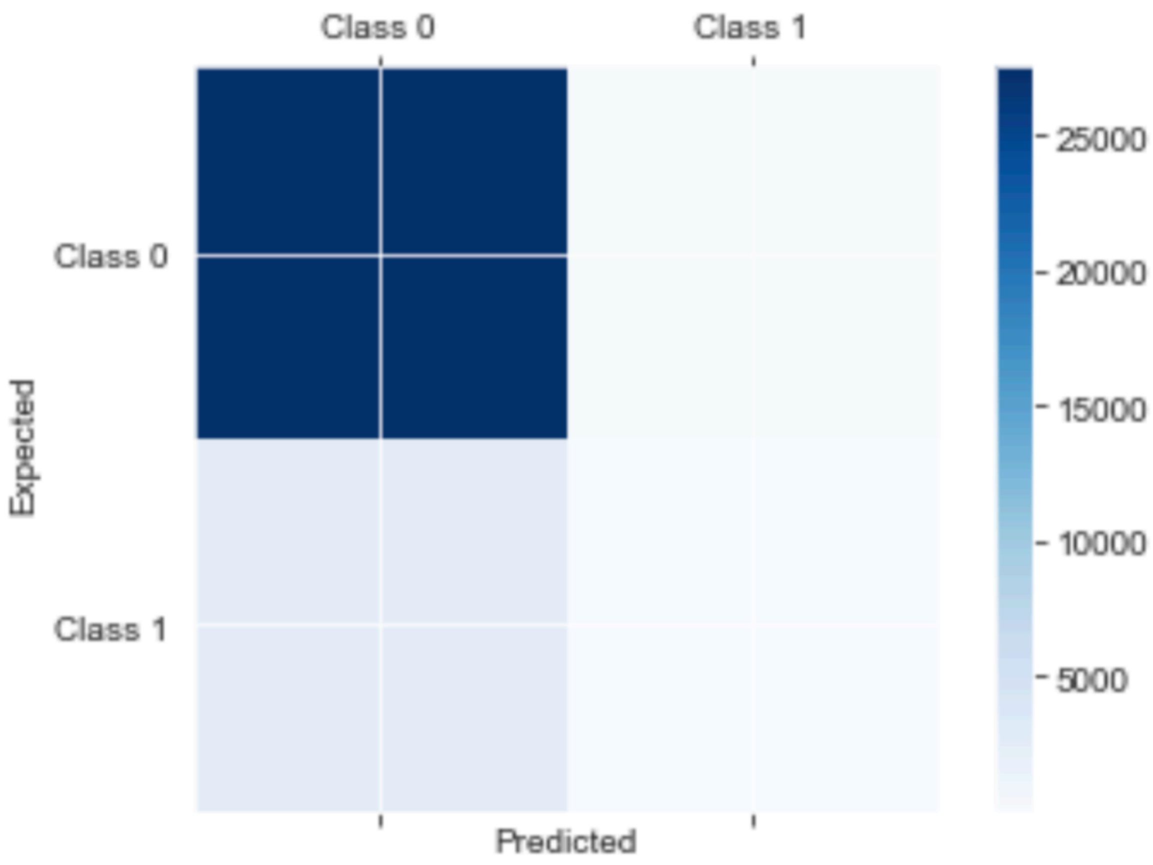
- Confusion Matrix - a table showing correct predictions and types of incorrect predictions.
- Precision - proportion of + identifications that are correct (TP/(TP+FP))
- Recall - proportion of actual positives identified correctly (TP/(TP+FN))
- Receiver operating characteristics curve - diagnostic ability of a binary classifier

BASELINE MODEL ACCURACY



PROBLEM WITH ACCURACY

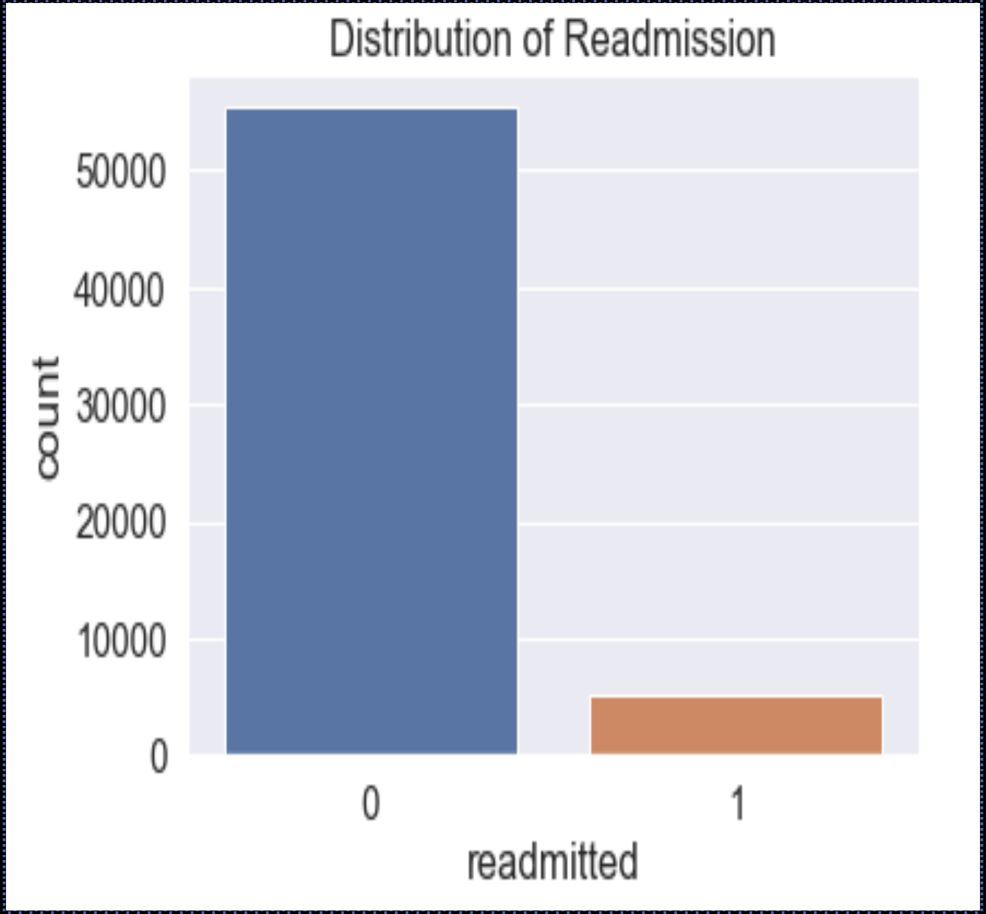
```
Confusion matrix:
 [[27502   187]
 [ 2547    32]]
```




```
Accuracy: 90.97 %
Precision: 0.1461187214611872
Recall: 0.012407910042652 19
F1-score: 0.022873481057898 5
R2: -0.15884038594015704
MSE: 0.0903264173384432 4
Prevalence: 0.007235364080877494
```

```
--------------------------------------------------------------

Classification Report:
              precision    recall  f1-score   support

          NO       0.92      0.99      0.95     27689
         YES       0.15      0.01      0.02      2579

    accuracy                           0.91     30268
   macro avg       0.53      0.50      0.49     30268
weighted avg       0.85      0.91      0.87     30268

--------------------------------------------------------------
```

IMPORTANT CLASS – MINORITY



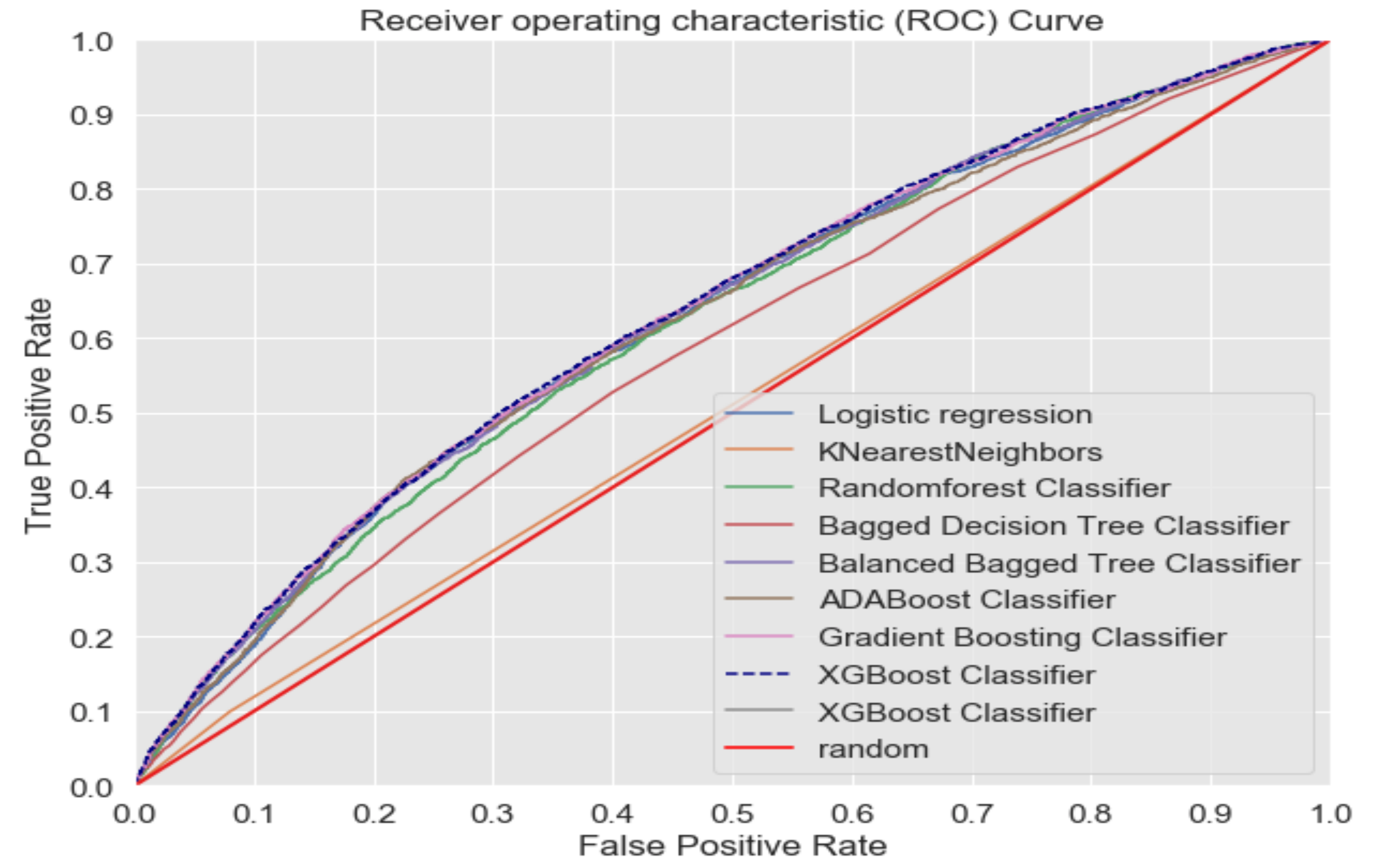
Imbalanced Class

- Readmitted (minority) class is important
- Resample the training set
- Repeated Stratified Kfold Cross validation
- Resample with different ratios
- Ensemble different resampled datasets
- Use the right evaluation metrics

MODEL SCORES

Algorithm	Accuracy Initial	ROC AUC KFold10	ROC AUC Repeated StratKfold	Tuned ROC AUC	Tuned PR AUC	Execution time Accuracy/ROC/ RepeatStratKFold
Logistic Regression()	91.47%	0.6192	0.6234	0.626	0.133	00:00:03s/4s/12
KNNeighbors()	91.04%	0.5158	0.5161	0.515	0.149	00:00:11s/10s/32
Decision Tree()	82.94%	0.5127	0.5138	0.608	0.128	00:00:02s/3s/9
Random Forest()	91.36%	0.5593	0.5546	0.622	0.129	00:00:04s/4s/13
Balanced Random Forest()	60.38%	0.6239	0.6267	0.628	0.134	00:00:24s/24s/1:14
Bagging Classifier()	91.12%	0.5520	0.5551	0.623	0.132	00:00:20s/19s/58
Balanced Bagging Classifier()	76.42%	0.5785	0.5790	0.628	0.134	00:00:11s/11s/34
ADABoost()	91.47%	0.6181	0.6181	0.624	0.134	00:00:15s/15s/47
Gradient Boost()	91.46%	0.6282	0.6321	0.633	0.138	00:00:55s/55s/2:49
XGBoost()	91.47%	0.6283	0.6320	0.634	0.140	00:01:38s/1:37/4:51

ALGORITHM TUNING ROC AUC CURVE



ROC AUC Scores
Logistic regression: 0.626
KNearestNeighbors: 0.51
Randomforest: 0.621
Bagged Tree Classifier: 0.628
Balanced Bagged Tree : 0.628
ADABoost: 0.623
Gradient Boosting: 0.633
XGBoost: 0.634
Tuned XGBoost: 0.49998

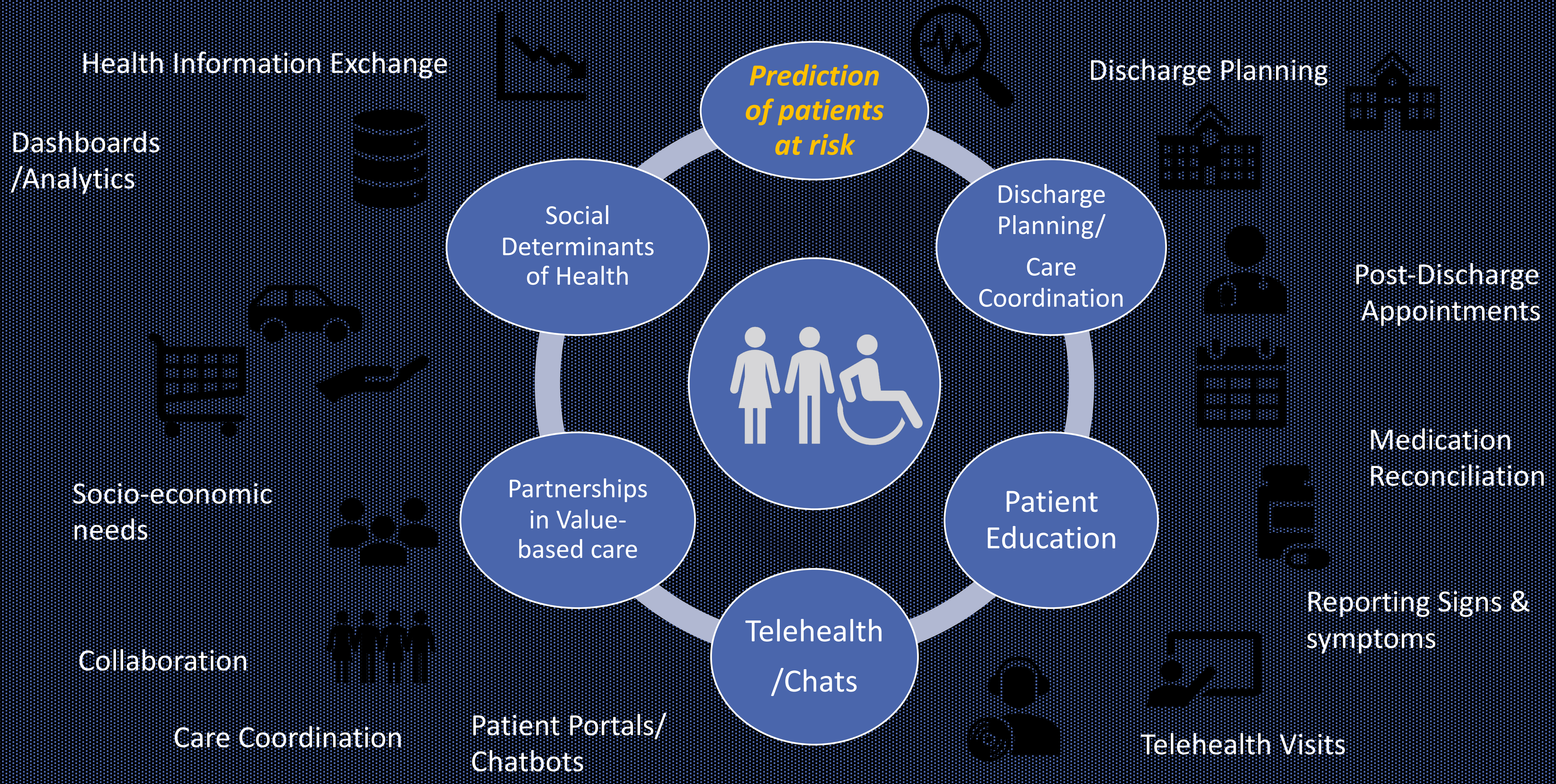
INTERPRETATION

- *True Positive Rate (TPR)* is the proportion of actual readmissions that the test correctly predicted would be readmitted.
- *False Positive Rate (FPR)* is the proportion of patients whom the model predicted would be readmitted, but were not.
- ROC is the graphical representation of the balance between TPR and FPR at *every* possible decision boundary
- AUC is a single number that can evaluate a model's performance, regardless of the chosen decision boundary
- If we were to choose a boundary of .8, readmittance probability above .8 is a readmission, everyone below is not.

INTERPRETATION

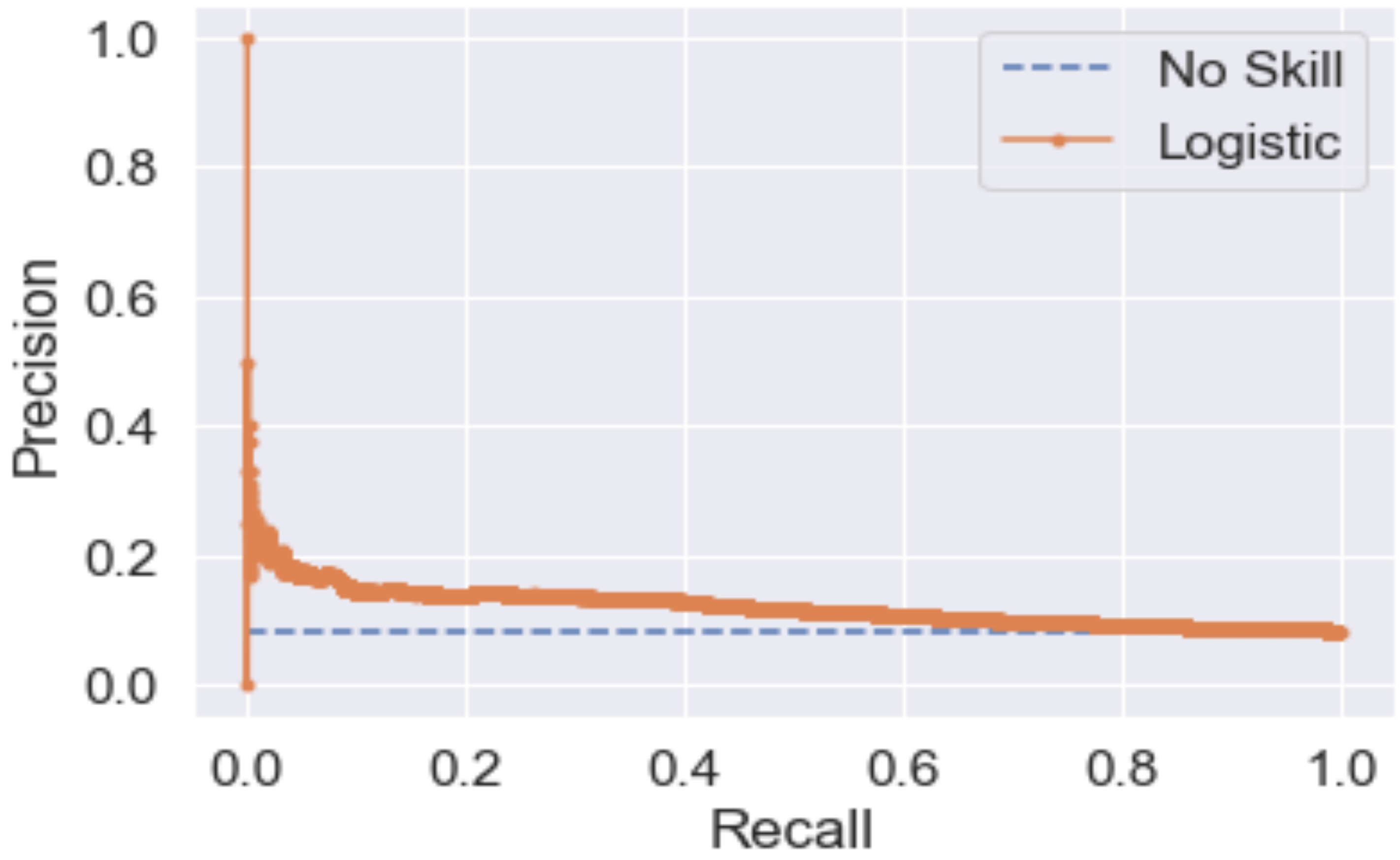
- Out of 46 input variables explored, the final model used=24.
- The final model XGBoost Classifier was selected as the most accurate model to predict readmissions.
- Compared to random predictions, results from our predictive model (AUC=.64: Accuracy=91.47%) is encouraging and a good baseline for further improving our model
- Fine tuning of the model will help the adoption of the model as a clinical decision system for evaluating readmission

CARE TRANSITIONS



PRECISION RECALL CURVE

```
AUC:  0.5
Logistic: f1=0.000 auc=0.122
```



FUTURE WORK

- Implement dashboard of Inpatient encounters and their predicted readmission risk scores and rankings and to measure and monitor reduction in readmission rates
- Explore further model improvements/tuning resampling techniques to improve the accuracy of the predictive model
- Experiment on further feature selection and elimination
- Take best models and ensemble
- Use alternative metrics for evaluation & thresholding
- Add interventions as additional predictors for readmissions.

THANK YOU!

Keep in touch!