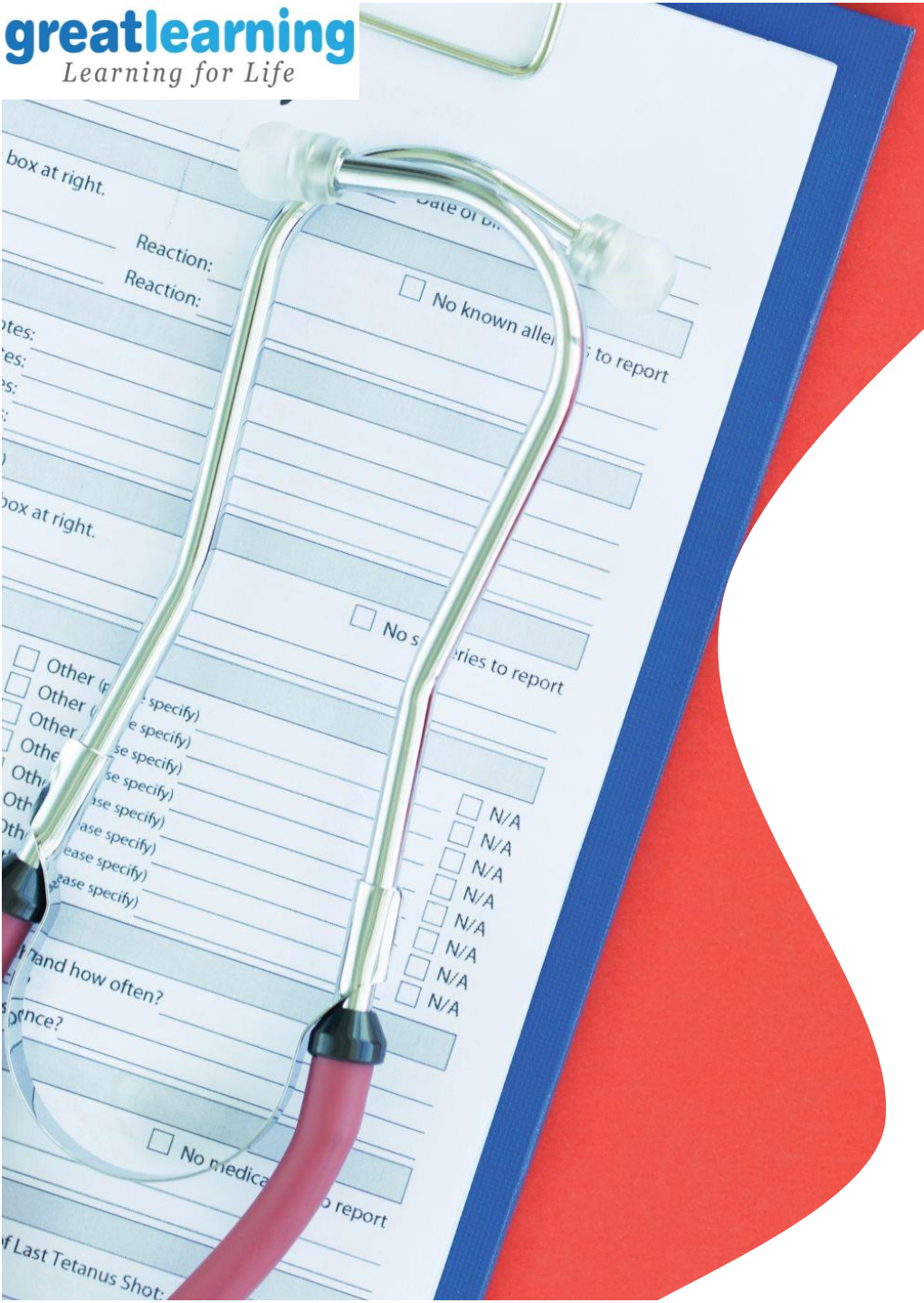
greatlearning
Learning for Life



{PGPDSE-GGN-JAN21-GROUP-4 } PRESENTS

DIABETIC PATIENT'S READMISSION PREDICTION

SUBMITTED BY:

MOHIT SAHU

RAHUL MAMGAI

RAVI RAJ CHAUBEY

SACHIN PATHANIA

SUSHIL

GUIDED BY:

SRAVAN MALLA

greatlearning
Learning for Life

BUSINESS PROBLEM

To identify the factors that lead to the high readmission rate of diabetic patients within 30 days post discharge and to correspondingly predict the high-risk diabetic patients who are most likely to get readmitted.

Also, to identify the medicines that are the most effective in the treatment.

Why it's needed?

1

Diabetes Mellitus (DM) is a chronic disease where the blood has high sugar level. It is a progressive disease that can lead to a significant number of health complications and profoundly reduce the quality of life.

2

Hospital readmission is a high-priority health care quality measure and target for cost reduction. The burden of diabetes among hospitalized patients, however, is substantial, growing, and costly, and readmissions contribute a significant portion of this burden.

3

Thirty-day readmission rates for hospitalized patients with DM are reported to be between
14.4 and 22.7% much higher than the rate for all hospitalized patients (8.5–13.5%).

greatlearning
Learning for Life

DATASET PREVIEW

The data subset used for analysis covers 10 years of diabetic patients encounters data (1999 – 2008) with over 100,000 diabetic patients and variables including length of stay, medicines, in-patient visits etc. from 130 hospitals in the United States.

VARIABLES

Patient identifiers :- encounter_id, patient_nbr

Patient demographics :- race, gender, age, weight ,payer_code

Admission and Discharge details :- admission_source_id ,admission_type_id, discharge_disposition_id

Patient medical history :- number_outpatient ,number_inpatient, number_emergency

Patient Admission details :- medical_speacialty ,diag_1, diag_2 and diag_3, time_in_hospital, number_diagnoses, num_lab_procedures, num_procedures, num_medications

Clinical Results :- max_glu_serum, A1cresult

Medication Details :- diabetesMed, change, 23 features for medications

Readmission Indicator :- readmitted

DATATYPES

Category:36, Numerical:13, Boolean:1

OTHER DETAILS

- Classification Problem
- Total Rows : 101766
- Total Columns : 50

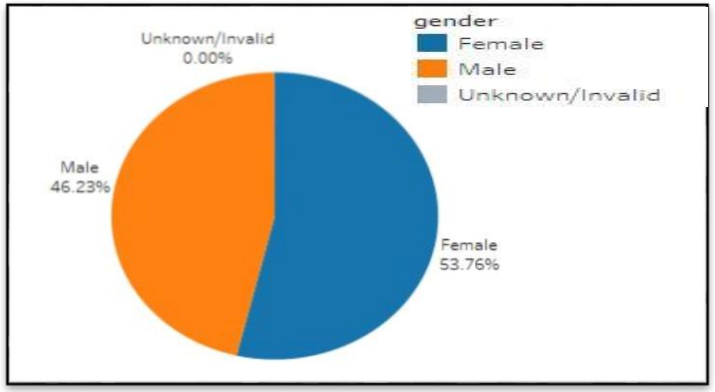
greatlearning
Learning for Life

EXPLORATORY DATA ANALYSIS

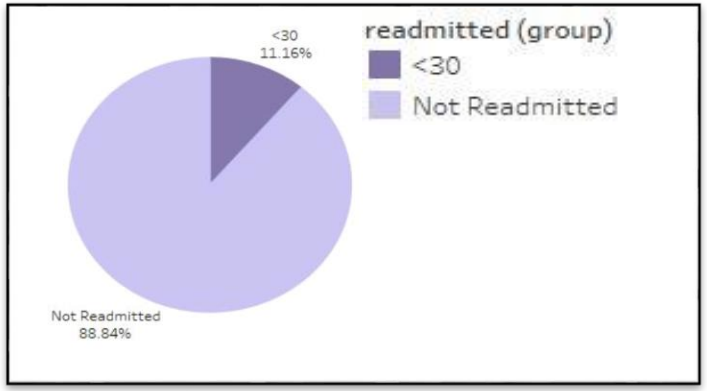
greatlearning
Learning for Life

UNIVARIATE ANALYSIS

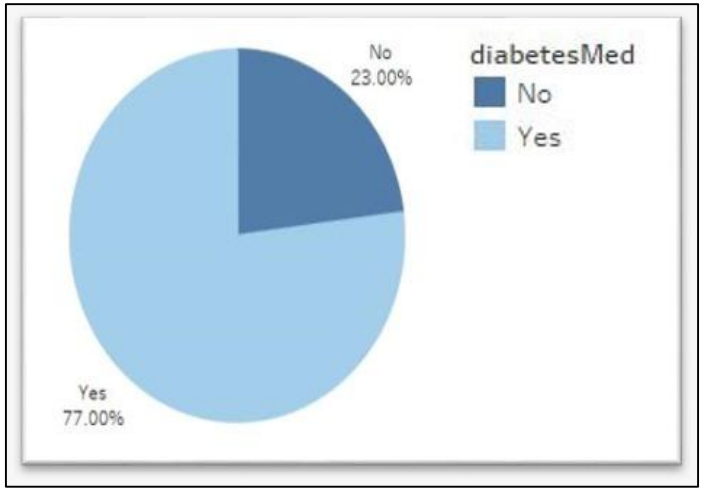
Gender

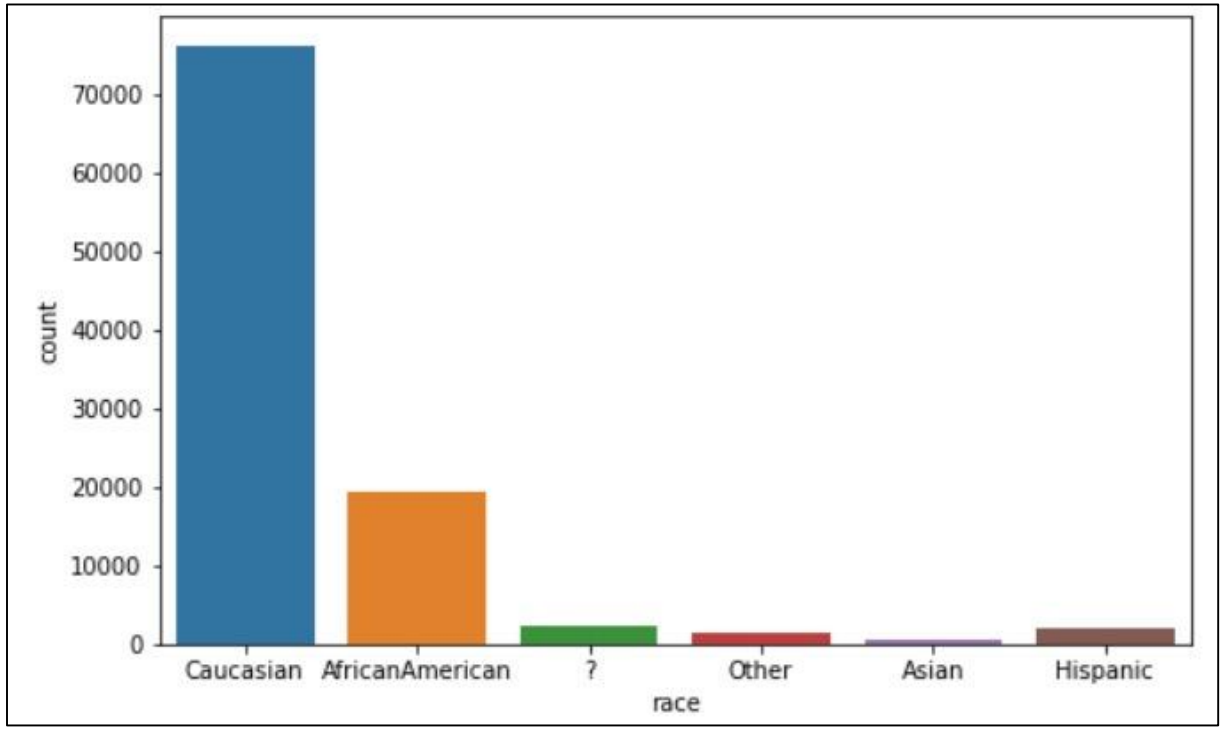


Readmitted



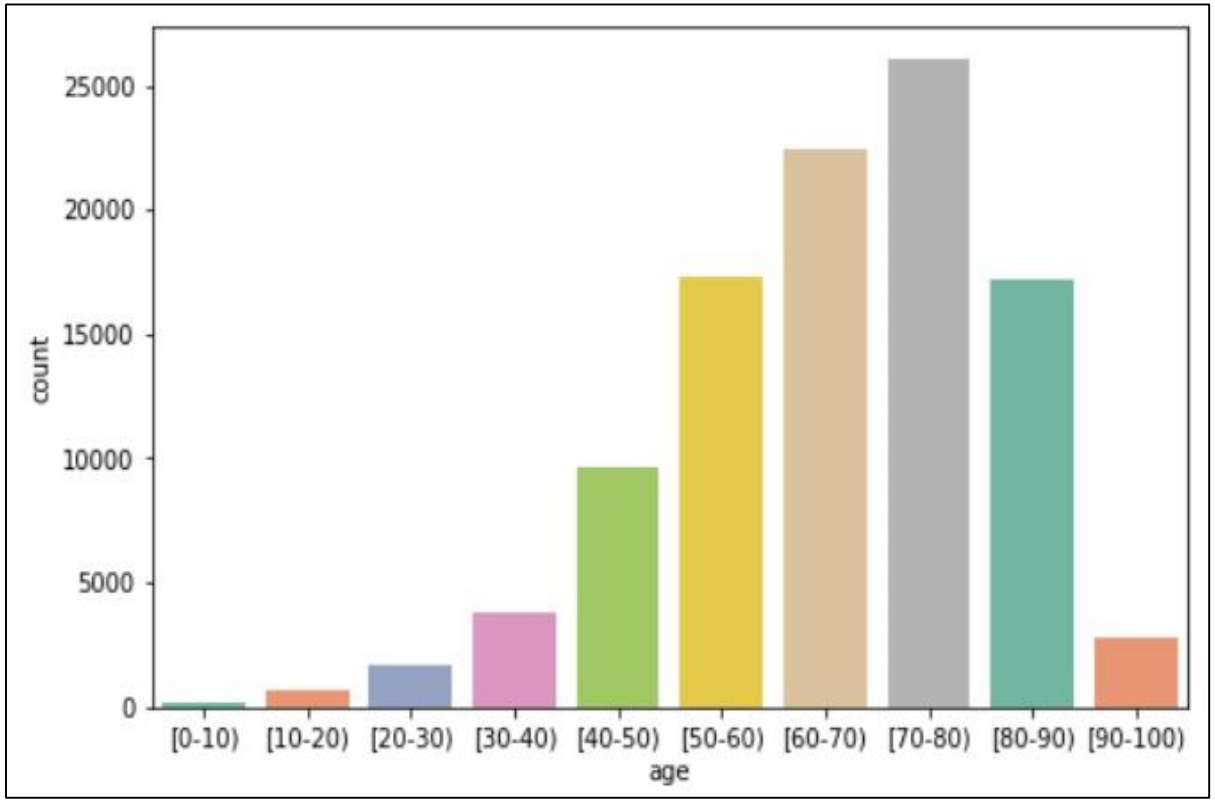
DiabetesMed




70000
60000
50000
40000
30000
20000
10000
0
count
Caucasian
AfricanAmerican
?
Other
Asian
Hispanic
race

Race

Age

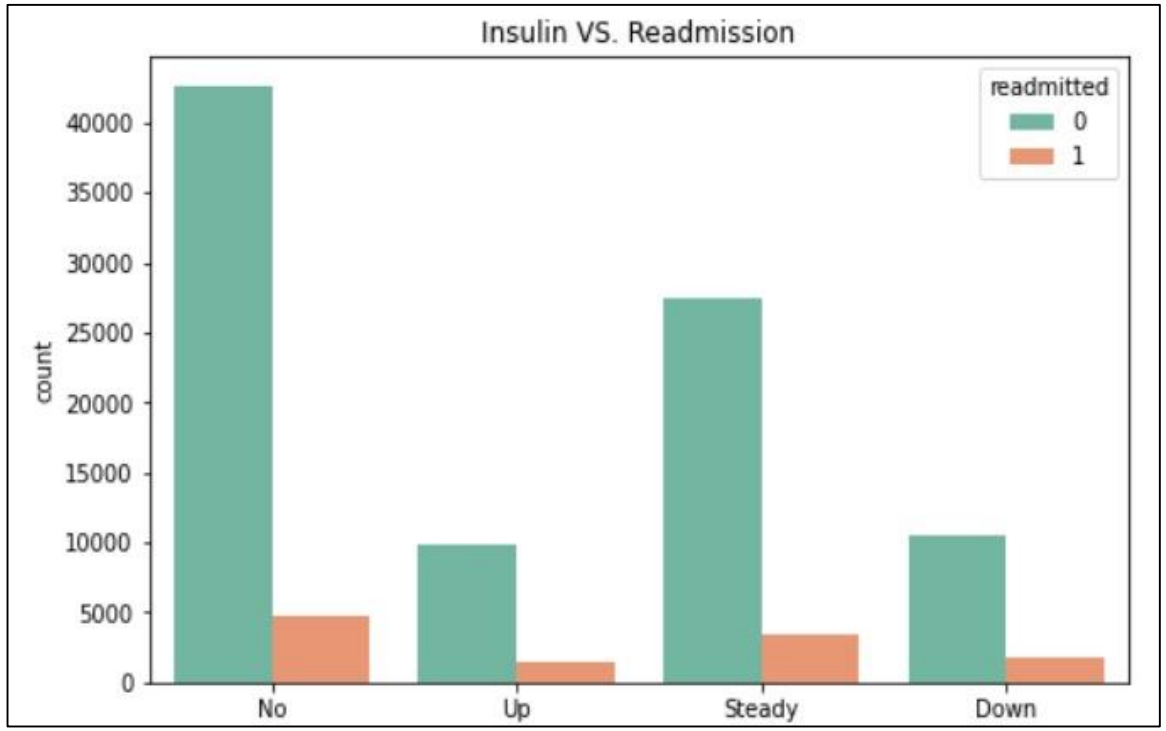

25000
20000
15000
10000
5000
0
count
[0-10)
[10-20)
[20-30)
[30-40)
[40-50)
[50-60)
[60-70)
[70-80)
[80-90)
[90-100)
age

greatlearning
Learning for Life

BIVARIATE ANALYSIS

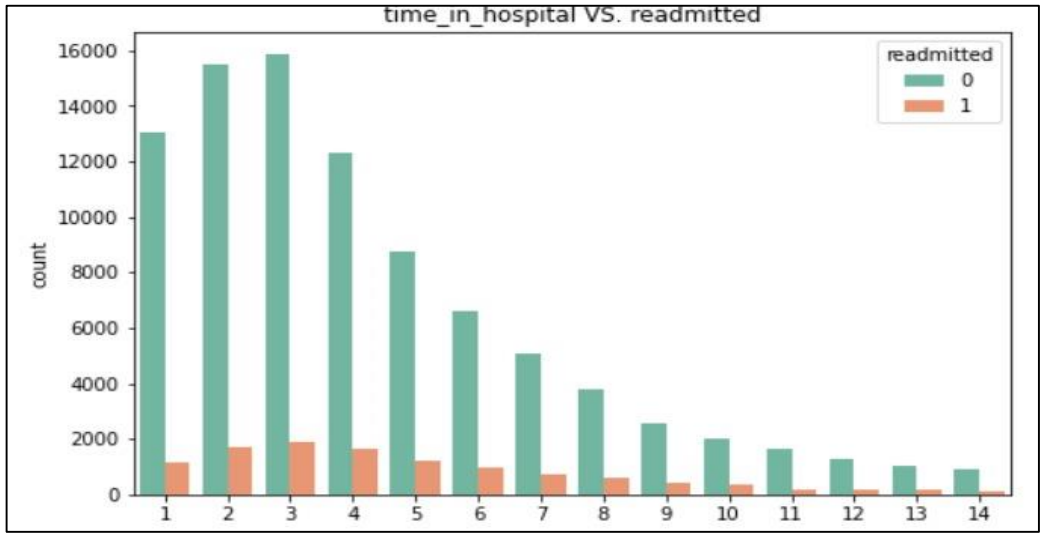
Insulin Vs Readmission

No. of Counts



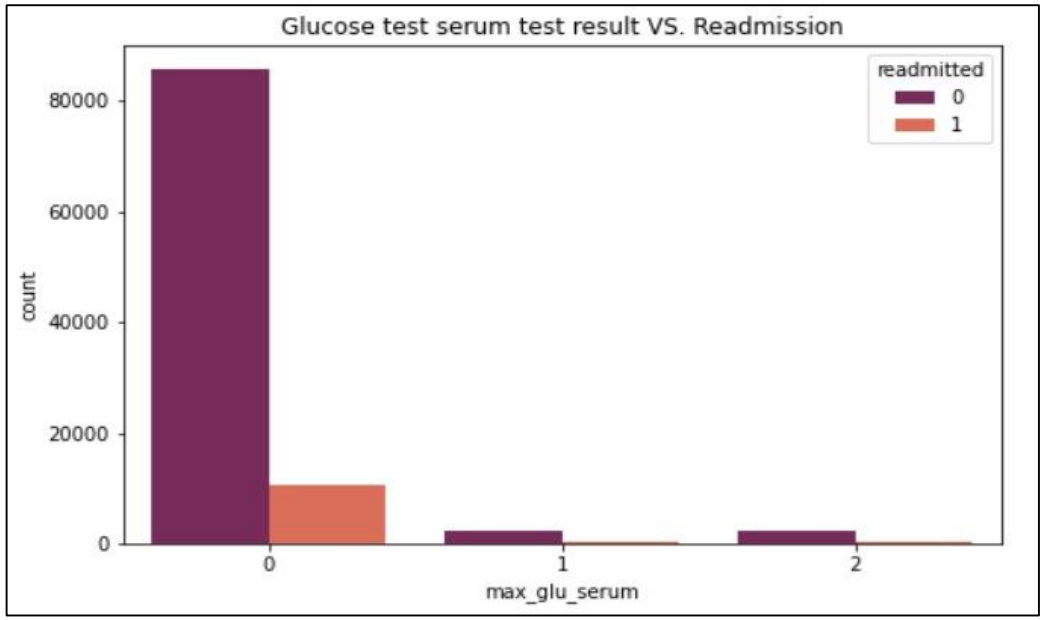
Time Vs Readmission

No. of Counts



Max_Glucose_serum Vs Readmission

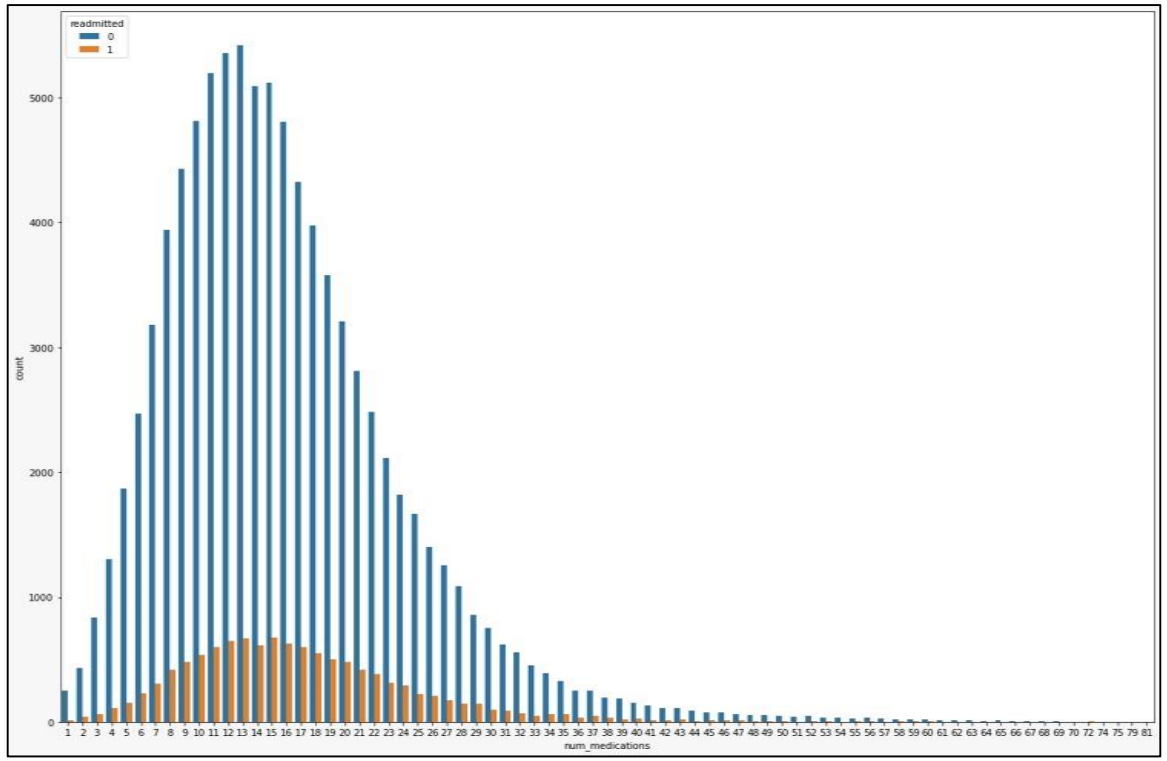
No. of Counts



greatlearning
Learning for Life

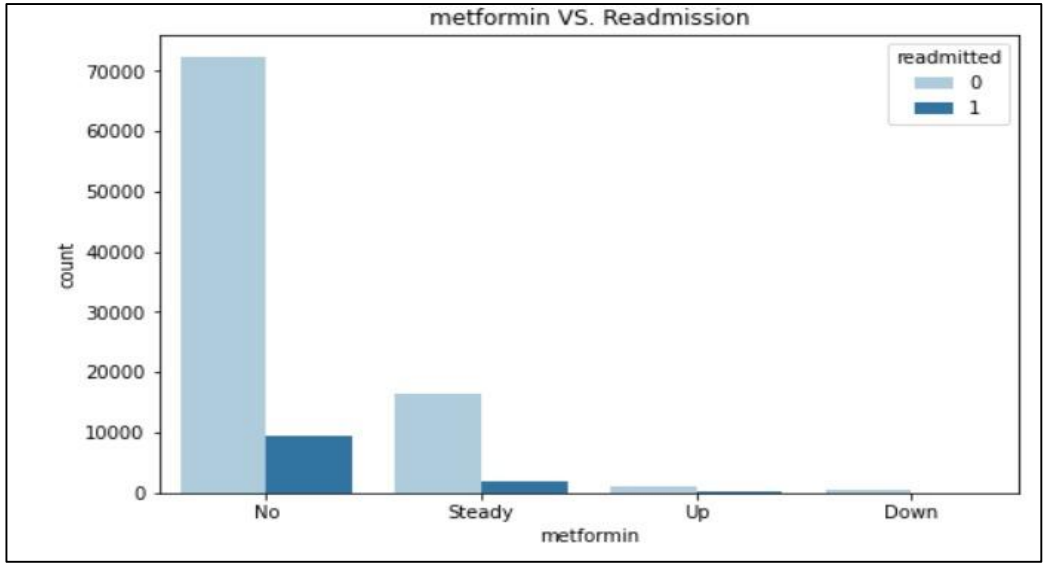
Num_Medications Vs Readmission

No. of Counts



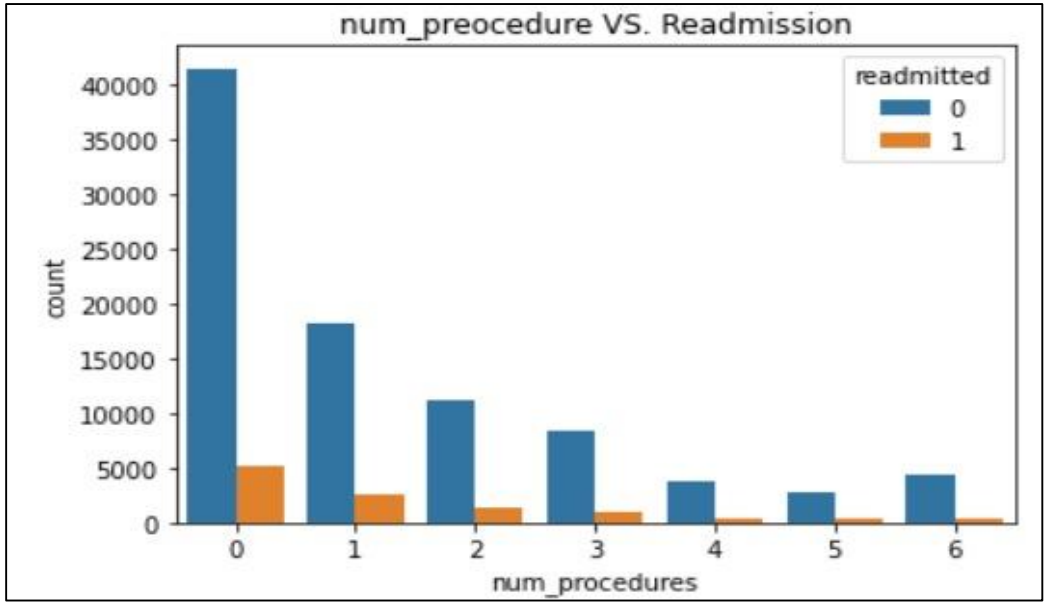
Metformin Vs Readmission

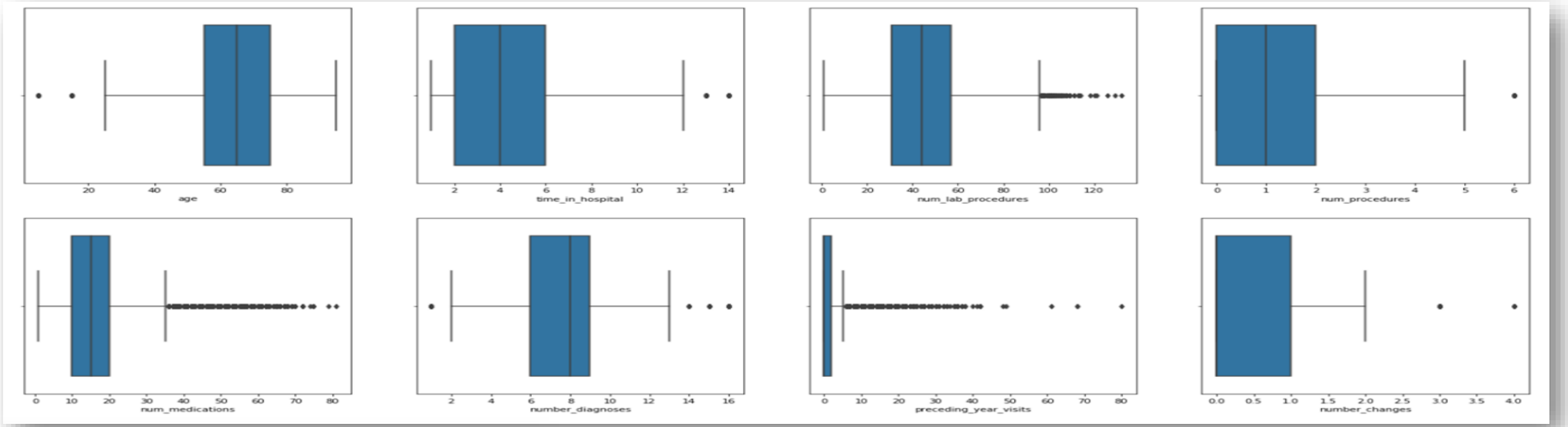
No. of Counts

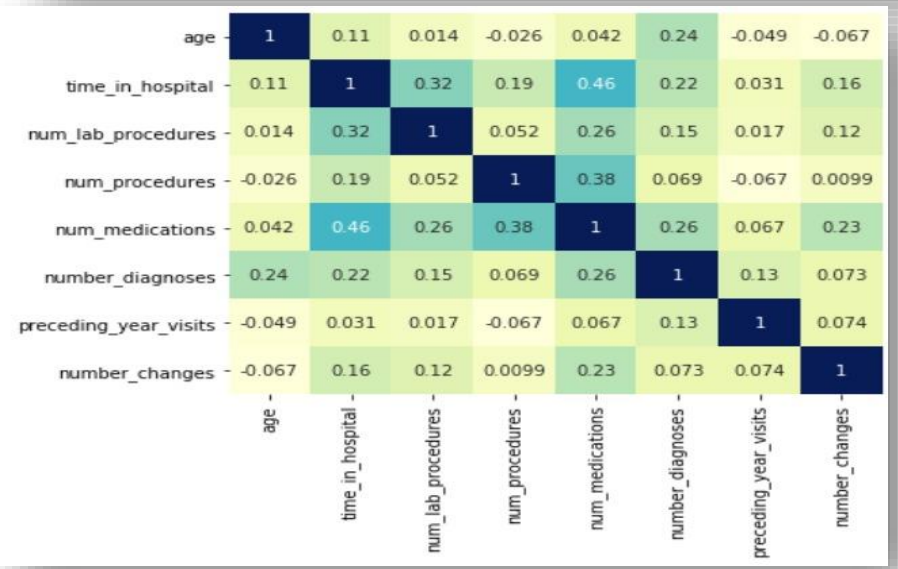


Num_Procedure Vs Readmission

No. of Counts







SUMMARY

- Age has Outliers, having low values indicating that , it's is left skewed.
- Time, Number of procedure, Number of Lab procedure and preceding year visits have Outliers with high value making the distribution right skewed .
- Independents Correlation has been shown through Heatmap.

greatlearning
Learning for Life

Missing Values

Weight
Payer_Code
Medical_specialty

They all have more than 40% missing value

Encoding

Gender	
Female	0
Male	1

Diabetes Med	
No	0
Yes	1

Age	
[0-10)	5
[10-20)	15
[20-30)	25
[30-40)	35
[40-50)	45
[50-60)	55
[60-70)	65
[70-80)	75
[80-90)	85
[90-100)	95

Feature Engineering

Number_Outpatient
Number _Inpatient
Number _Emergency

All combine to form
Service_utilization

greatlearning
Learning for Life

MODELS WITH BALANCED DATA WITHOUT TUNING

Logistic Regression (class_weight="balanced")		
Metrics	Train Set Result	Test Set Result
Accuracy_score	59.23%	59.90%
F-1 Score	25.50%	26.10%
Precision_Score	16.10%	16.60%
Recall_Score	61.10%	61.20%
AUC_ROC Score	60%	60.53%

Decision Tree Classifier (class_weight="balanced")		
Metrics	Train Set Result	Test Set Result
Accuracy_score	80.5%	80.47%
F-1 Score	99.9%	14.06%
Precision_Score	99.9%	40.29%
Recall_Score	100%	13.8%
AUC_ROC Score	99.9%	51.5%

Random Forest Classifier (class_weight="balanced")		
Metrics	Train Set Result	Test Set Result
Accuracy_score	88.57%	88.44%
F-1 Score	99.98%	0.29%
Precision_Score	99.98%	41.66%
Recall_Score	99.97%	0.14%
AUC_ROC Score	99.98%	50.06%

XG Boost Classifier(scale_pos_weight=7.73)		
Metrics	Train Set Result	Test Set Result
Accuracy_score	68.37%	67.80%
F-1 Score	40.50%	25.78%
Precision_Score	27.38%	17.56%
Recall_Score	77.74%	48.43%
AUC_ROC Score	75.58%	59.38%

LightGBM Classifier (class_weight="balanced")		
Metrics	Train Set Result	Test Set Result
Accuracy_score	63.64%	64.09%
F-1 Score	32.10%	27.06%
Precision_Score	20.84%	17.68%
Recall_Score	69.81%	57.70%
AUC_ROC Score	67.82%	65.51%

greatlearning
Learning for Life

MODELS WITH BALANCED DATA WITH TUNING

Decision Tree Classifier (class_weight="balanced")		
Metrics	Train Set Result	Test Set Result
Accuracy_score	60.61%	62.03%
F-1 Score	26.44%	26.34%
Precision_Score	16.95%	16.97%
Recall_Score	60.05%	58.79%
AUC_ROC Score	61.06%	60.62%

Random Forest Classifier (class_weight="balanced")		
Metrics	Train Set Result	Test Set Result
Accuracy_score	63.67%	67.19%
F-1 Score	25.43%	25.76%
Precision_Score	17.10%	17.43%
Recall_Score	49.56%	49.29%
AUC_ROC Score	59.30%	59.41%

XG Boost Classifier(scale_pos_weight=7.73)		
Metrics	Train Set Result	Test Set Result
Accuracy_score	62.42%	62.81%
F-1 Score	28.26%	26.80%
Precision_Score	18.17%	17.34%
Recall_Score	63.49%	58.96%
AUC_ROC Score	63.32%	61.14%

LightGBM Classifier (class_weight="balanced")		
Metrics	Train Set Result	Test Set Result
Accuracy_score	61.28%	61.79%
F-1 Score	27.04%	26.79%
Precision_Score	17.25%	17.20%
Recall_Score	62.25%	60.56%
AUC_ROC Score	61.93%	61.26%

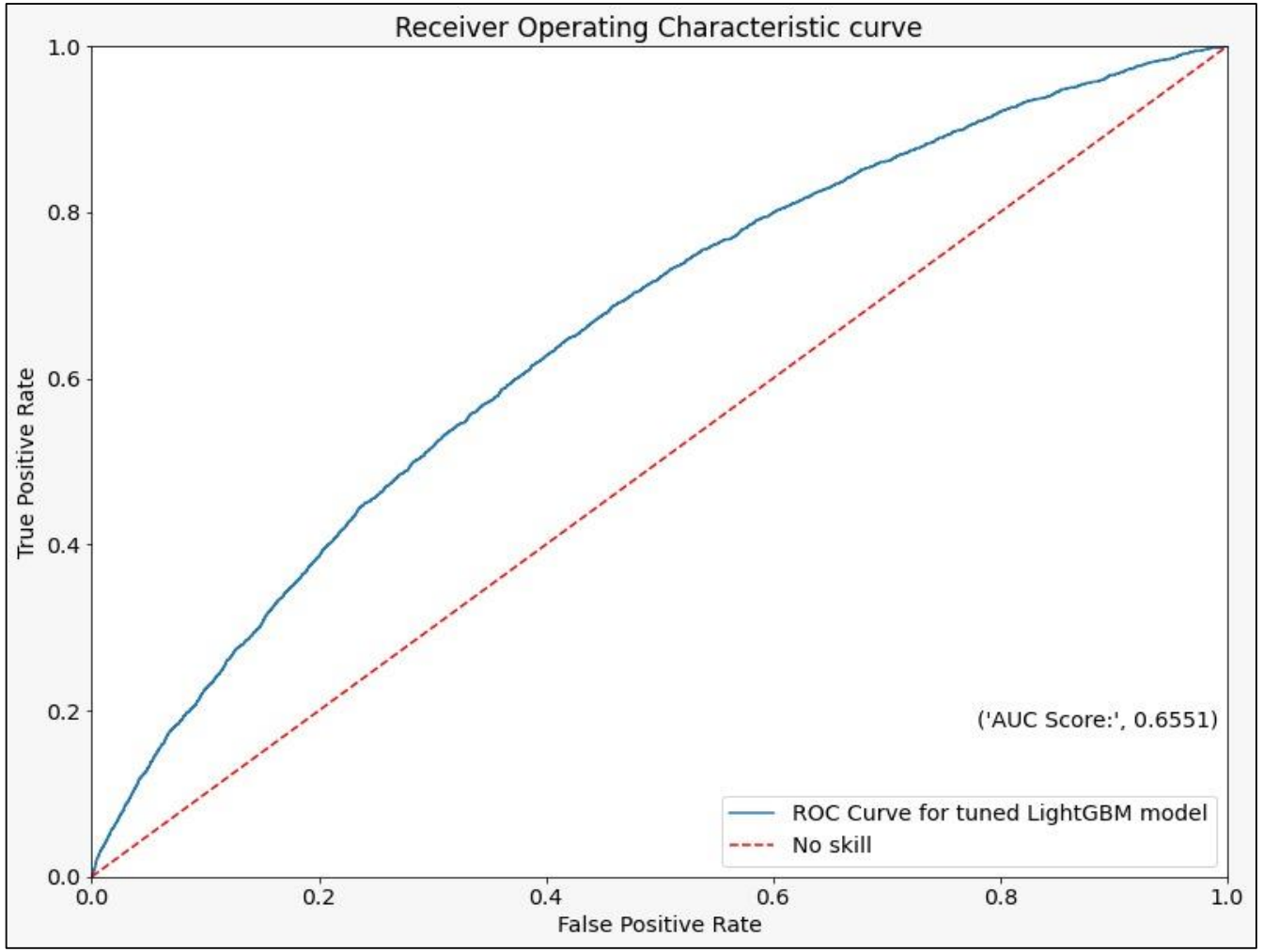
greatlearning
Learning for Life

SELECTED MODEL

LightGBM Classifier (class_weight="balanced")		
Metrics	**Train Set Result**	**Test Set Result**
Accuracy_score	61.28%	61.79%
F-1 Score	27.04%	26.79%
Precision_Score	17.25%	17.20%
Recall_Score	62.25%	60.56%
AUC_ROC Score	61.93%	65.51%
Final Model Chosen is LGBM !!!!!		

CONFUSION MATRIX

16080	9874
1336	2052



greatlearning
Learning for Life

FINDINGS

If the patient has the following characteristics he has a high probability of being readmitted :

- High preceding year visits.
- If the patient is discharged to another medical facility or discharged to home with health services.
- High number of diagnoses.
- If the patient is given diabetes medicines.
- If the primary diagnosed disease was of circulatory system.
- If Metformin and/or insulin is not being given or the dosage is low.
- If secondary diagnosis was coming to be Diabetes.
- If A1C test was not performed.

CHALLENGES ENCOUNTERED

- The major challenge was to acquire sufficient domain knowledge of the medical world.
- Huge amount of Data
- Different Imputation methods
- Imbalanced Data
- Reductants Features
- Hyper parameter tuning

greatlearning
Learning for Life

BUSINESS RECOMMENDATION

- Medical facilities can take precautionary measures with patients during their initial admission by making A1C and Maximum Glucose Serum test compulsory and providing the treatment accordingly.
- Providing extra attention and care to high-risk patients.
- A follow-up with the discharged patients should be done to keep a track of their health and to counsel them from time-to-time.
- High-risk patients' current medicines' regime should be re-evaluated and the most effective medicines should be considered.
- Most Effective Medications :- Metformin, Glipizide, Insulin.
- The annual plans, financials and infrastructure / inventory should be planned accordingly by taking into account the predicted readmissions.

greatlearning
Learning for Life

THANK YOU