Pharmaquant Hackathon

Ayan Saha
Ayush Ghatak
Siddhartha Sen
Soumyadeep Roy

Data Description

Clinical Features (Medical and health-related attributes)

1. Number of lab procedures (num_lab_procedures)
2. Number of procedures (num_procedures)
3. Number of medications (num_medications)
4. Number of outpatient visits (number_outpatient)
5. Number of emergency visits (number_emergency)
6. Number of inpatient visits (number_inpatient)
7. Number of diagnoses (number_diagnoses)

Treatment-Related Features (Hospitalization and medication-related attributes)

1. Admission type (admission_type_id)
2. Discharge disposition (discharge_disposition_id)
3. Admission source (admission_source_id)
4. Time in hospital (time_in_hospital)
5. Change of medications (change)
6. Diabetes medications (diabetesMed)
7. 24 medication-related features (metformin to metformin.pioglitazone)

Demographic Features :

1. Race
2. Gender (gender)
3. Age (age)
4. Weight (weight)

Objective of our Analysis

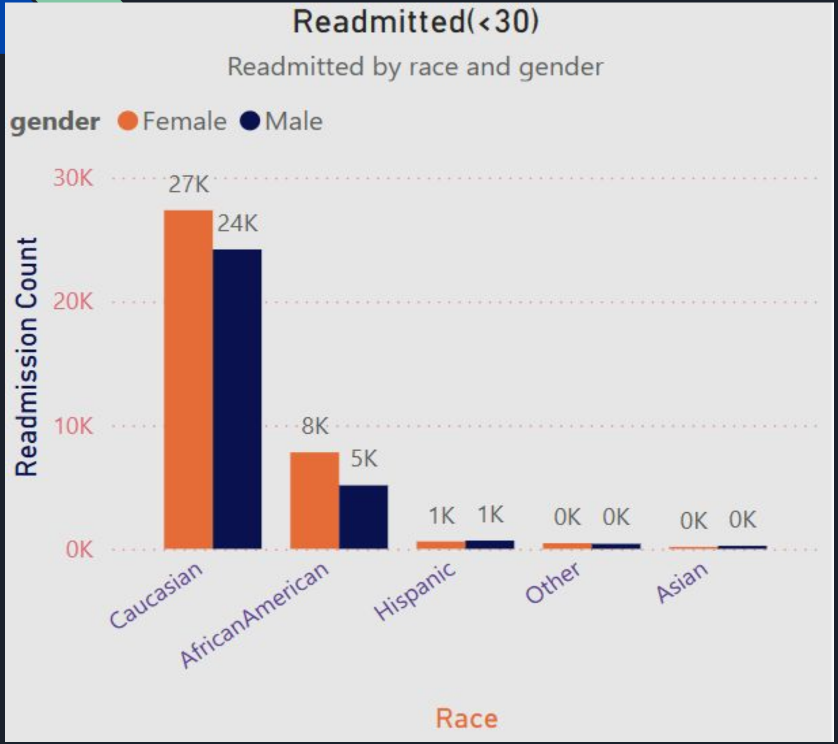
We are trying to analyse readmission rates based on different clinical features, demographic features and treatment related features.

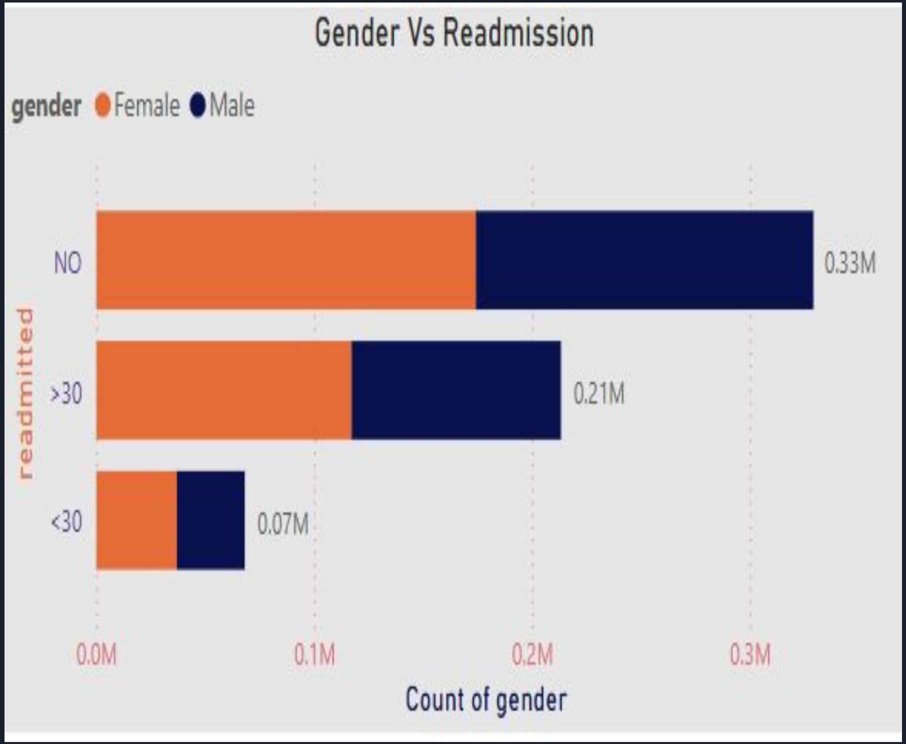
Reason behind the chosen objective :

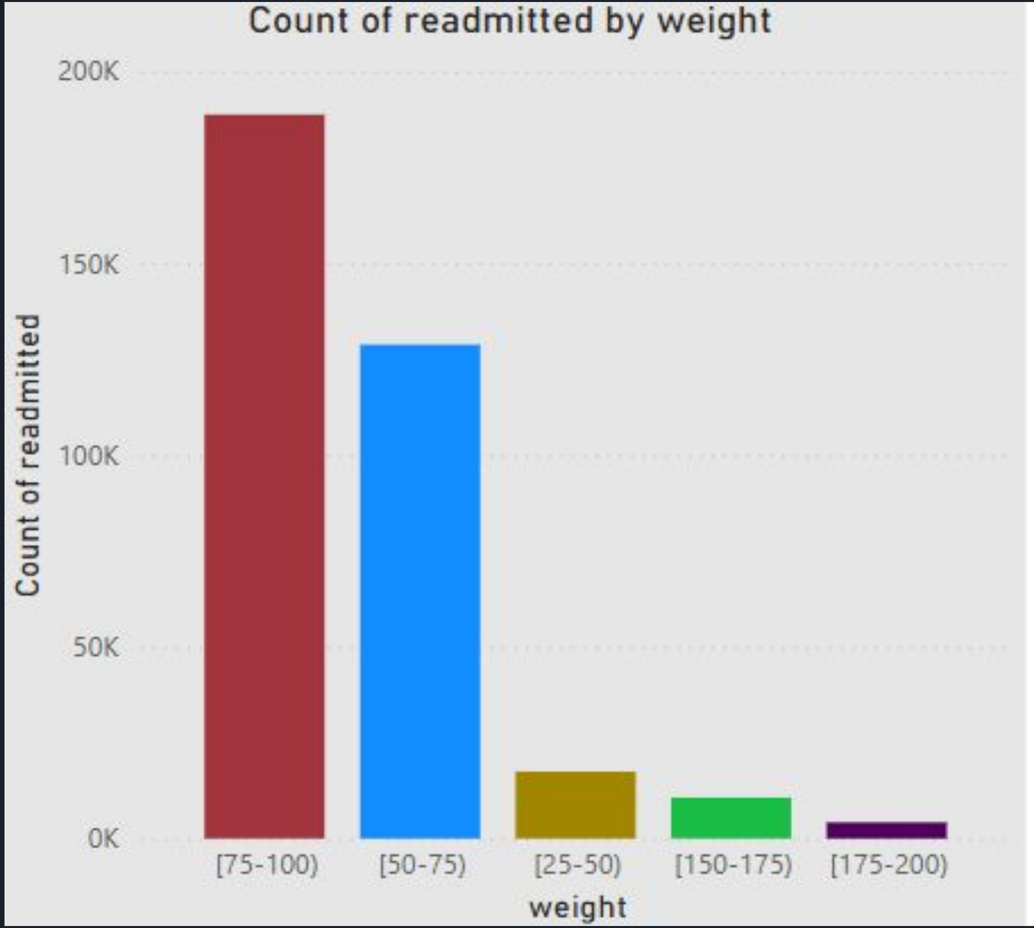
There is a certain threshold value of readmission rates above which hospitals get penalised in US. HRRP is a Medicare value-based program that encourages hospitals to improve communication and coordination with patients to reduce avoidable readmissions. This is in relation to the fact that 83% of the hospitals in US face penalty because of violating this HRRP act. We are conducting our analysis to address this issue from the hospital's point of view.

https://www.cms.gov/medicare/quality/value-based-programs/hospital-readmissions

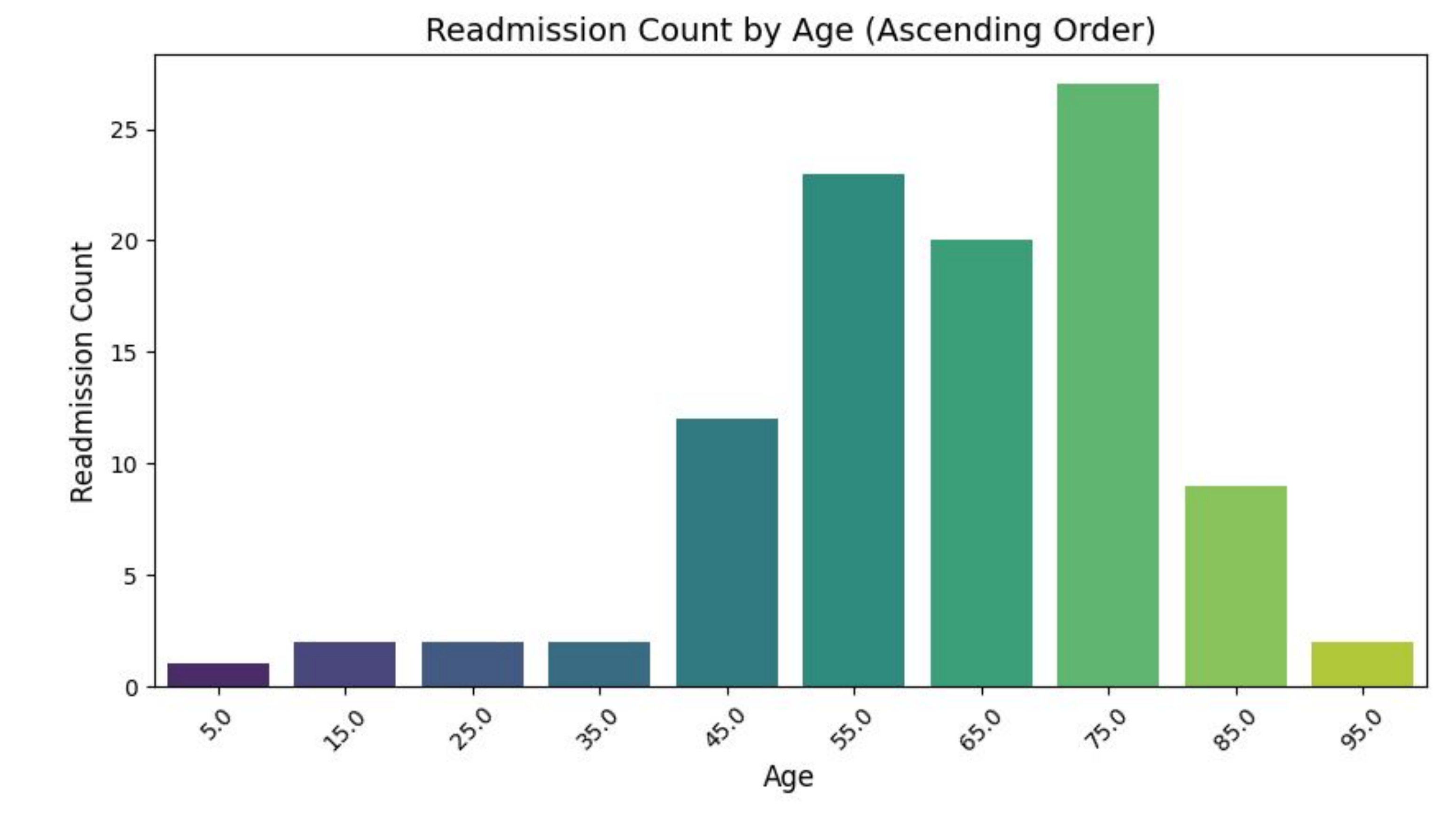
Re-admission counts based on Demographic features (Patient background and personal attributes) :



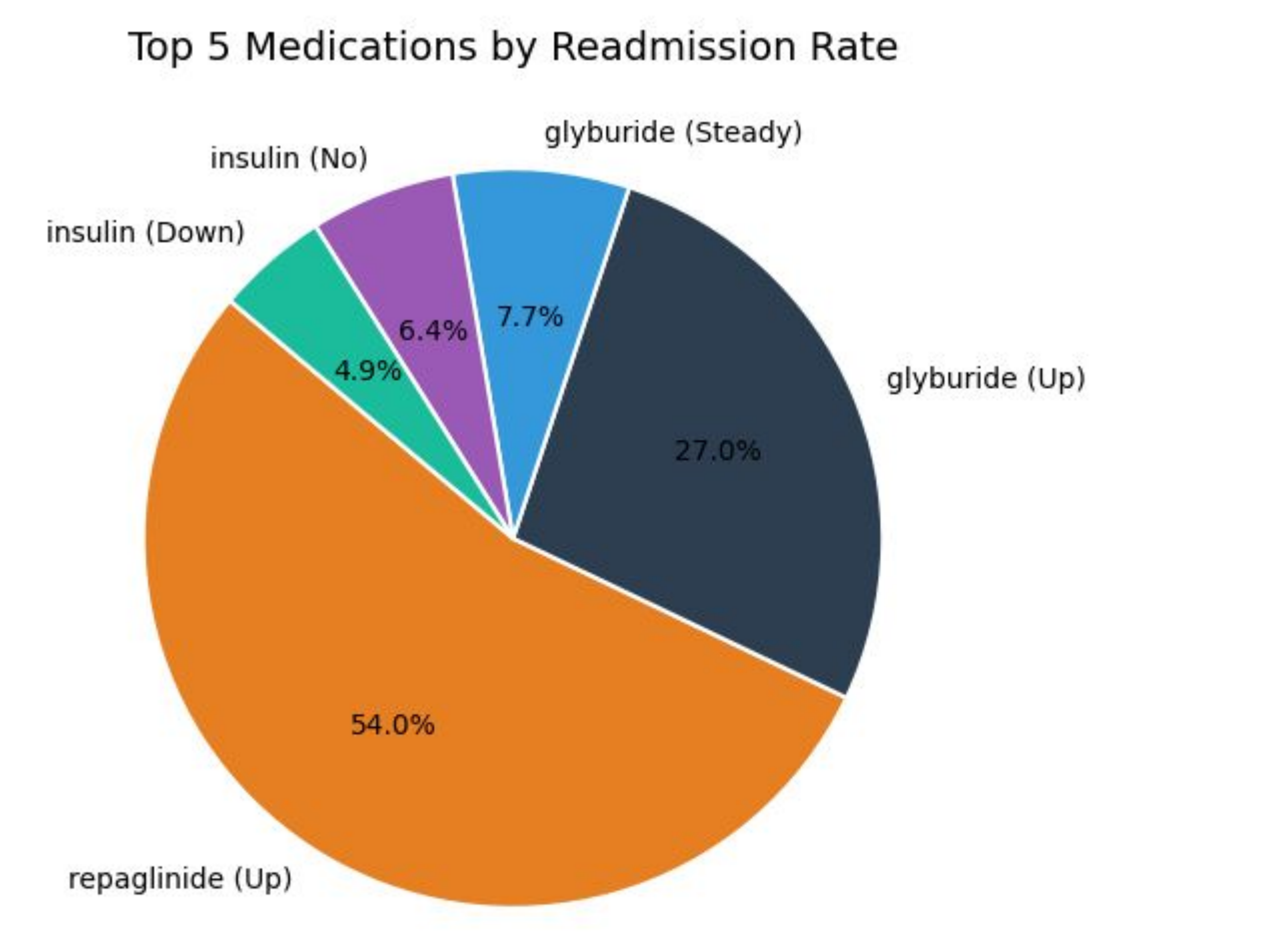




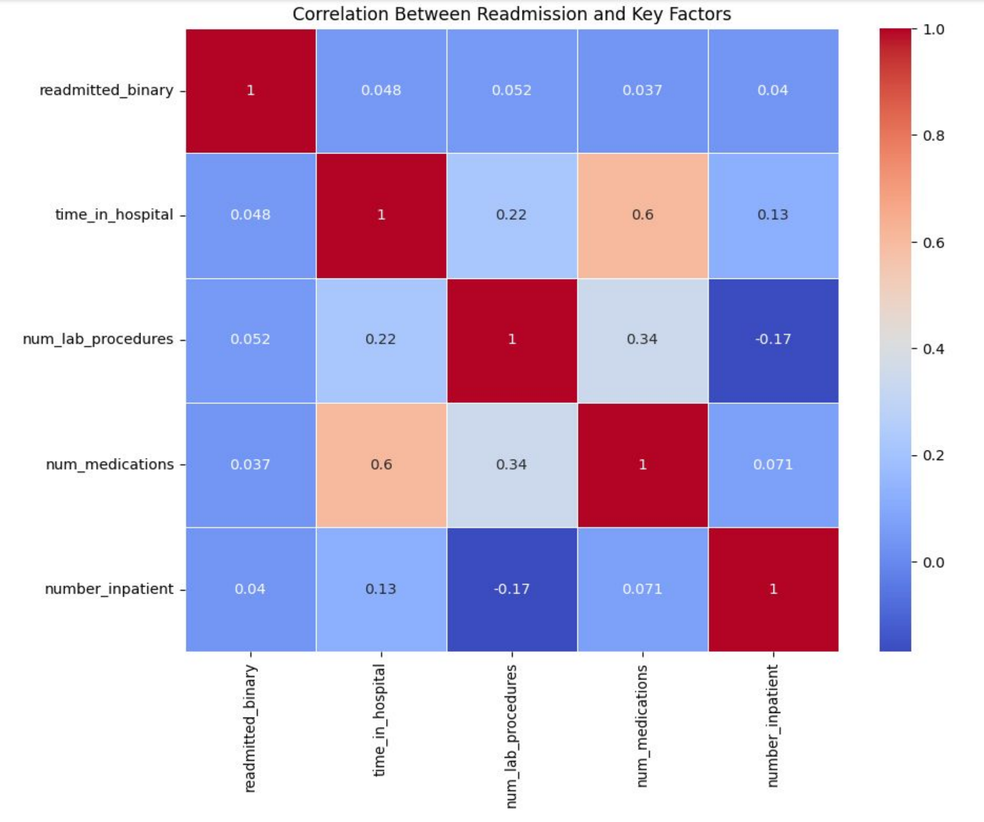
Within 30 days readmission(Weight in pounds)



Readmission Count by Age (Ascending Order)
Readmission Count
0
5
10
15
20
25
5.0
15.0
25.0
35.0
45.0
55.0
65.0
75.0
85.0
95.0
Age


Top 5 Medications by Readmission Rate
glyburide (Steady)
insulin (No)
insulin (Down)
glyburide (Up)
repaglinide (Up)
7.7%
6.4%
4.9%
27.0%
54.0%

Correlation Heatmap



Strongest relationship: Time in hospital & medications (0.6).

Weak correlations:

Time in hospital & lab procedures (0.22).

Lab procedures & medications (0.34).

Very weak correlations:

Time in hospital & inpatient visits (0.13).

Medications & inpatient visits (0.071).

Weak negative correlation: Lab procedures & inpatient visits (-0.17).

Conclusion: These factors alone are not strong predictors of readmission.

However, correlation doesn’t imply causation. Therefore, we need to proceed with further tests to actually find the significant predictors.

Data Cleaning and Preprocessing

We removed the following data columns from our analysis : Unnamed, encounter_id, Patient_nbr, payer_code, medical_specialty, max_glu_serum, A1C result. All these columns are irrelevant to our analysis and some of them occur with huge no of null values.

The ‘weight’ column has a considerable no of null values but at the same time its significant in analyzing the readmission rates. We could have used KNN to impute the null values but since we have high no of predictors here its not suggested to use KNN.

So we are imputing the null weight values by replacing them with the median value of weight of those patients who have been clustered based on similar demographic features.

Analysis

Which features best predict readmission within 30 days ?

We have treated this as a binary classification problem for which we are using logistic regression , random forest and XGBoost.

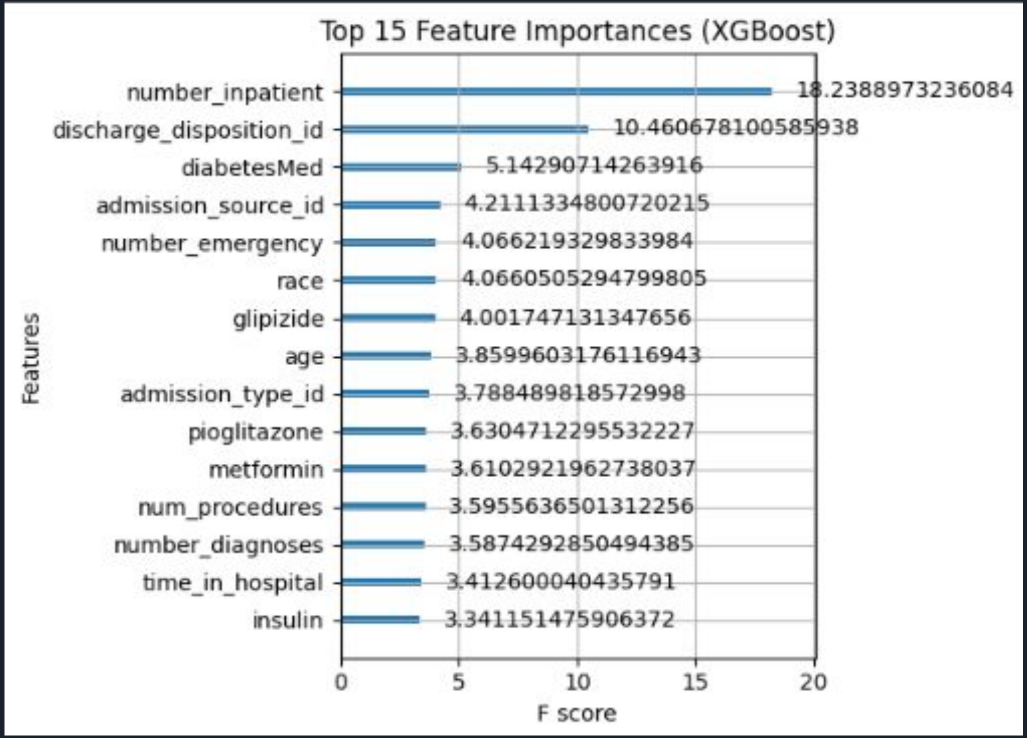
Results

```
=== Logistic Regression ===
              precision    recall  f1-score   support

           0       0.89      1.00      0.94     17609
           1       0.43      0.01      0.03      2227

    accuracy                           0.89     19836
   macro avg       0.66      0.51      0.48     19836
weighted avg       0.84      0.89      0.84     19836

ROC-AUC Score: 0.6340

=== Random Forest ===
              precision    recall  f1-score   support

           0       0.89      1.00      0.94     17609
           1       0.56      0.01      0.02      2227

    accuracy                           0.89     19836
   macro avg       0.72      0.50      0.48     19836
weighted avg       0.85      0.89      0.84     19836

ROC-AUC Score: 0.6310

=== XGBoost ===
              precision    recall  f1-score   support

           0       0.89      1.00      0.94     17609
           1       0.45      0.02      0.04      2227

    accuracy                           0.89     19836
   macro avg       0.67      0.51      0.49     19836
weighted avg       0.84      0.89      0.84     19836

ROC-AUC Score: 0.6565
```

On the basis of accuracy, all our 3 models perform equally well in predicting whether a patient would be readmitted within 30 days.

However, on the basis of ROC-AUC score, XGBoost performs slightly better than the other 2 because of its capability to handle non linear relationships (common in healthcare data and the fact that it uses both l1 and l2 regularization to prevent overfitting.

We find that the following predictors are most crucial in predicting the readmission of patients within 30 days. Also, the – model fits the data best as is clear from the result table. This can be used as an estimate of a patient getting readmitted within 30 days.



Actionable Insights

- Implement post-discharge follow-up programs for high-risk patients.
- Improve care coordination to reduce unnecessary readmissions.
- Analyze which discharge dispositions lead to frequent readmissions.
- Optimize discharge planning and provide better transition care.
- Track medication adherence and adjust prescriptions as needed.
- Provide patient education programs on diabetes self-management.
- Strengthen outpatient and telehealth services to reduce ER visits.
- Improve triage systems to ensure better resource utilization.

Identification of different patient clusters

We have segmented diabetic patients into distinct groups based on demographics, treatment history, hospital usage, and health indicators to identify patterns in readmission, severity, or healthcare needs.

We choose relevant features like:

age,gender,race,time_in_hospital,no_lab_procedures,
no_procedures,no_of_medications,no_inpatients,no_outpatients,no_e
mergency,change_diabetesMed

Summary of the clusters:

cluster	age	gender	race	time_in_hospital	num_lab_procedures	num_procedures	num_medications	number_inpatient	number_outpatient	number_emergency	change	diabetesMed
0	62.002133	0.447741	0.222174	3.757320	40.769149	1.269052	12.492341	0.437561	0.249758	0.097634	0.000000	0.000000
1	62.897242	0.463914	0.206551	7.912683	56.270978	2.719106	25.490849	0.535050	0.238962	0.113364	0.687633	0.953135
2	55.984577	0.482512	3.288901	3.903608	41.737538	1.166070	13.698155	0.492977	0.248141	0.163040	0.513082	0.821537
3	60.701889	0.468896	0.205778	3.245257	38.488433	0.866431	13.608021	0.423982	0.261643	0.114476	0.547853	1.000000
4	55.402341	0.423555	0.255304	4.555596	45.214338	0.907644	17.712326	3.902158	2.432151	1.730432	0.547366	0.858266

Cluster	Patient Type	Key Characteristics	Actionable Insights
0	Older, Low-Risk, Non-Medicated	- Age ~62 - Shortest hospital stay (3.75 days) - Low inpatient/emergency visits - **No medication changes or diabetes meds prescribed**	Likely pre-diabetic or stable diabetics. Prevent progression via lifestyle guidance
1	Elderly, High Resource, Medicated	- Highest hospital stay (7.91 days) - Most lab tests (56+), meds (25.5), and procedures - High medication use and changes	High-cost, high-risk group. Needs coordinated care, close follow-up post discharge
2	Mid-age, Poor Control	- Younger (~56), high race diversity - Poor glycemic control (high emergency visits) - High diabetes med use (82%)	Focus on improving med adherence, A1C control, and follow-ups
3	Stable, Low Procedure Use	- Balanced profile - Fewest procedures & low resource use - Fully medicated (100%)	Good management group. Continue same care strategy
4	Mid-age, Emergency Heavy	- Younger (~55), **high emergency visits (1.73)** & outpatient use - High `number_inpatient` (3.9) - Medicated but few changes	Possibly unstable or uncontrolled cases. Consider emergency care diversion efforts

cluster	readmitted	proportion
0	NO	0.606263
	>30	0.304247
	<30	0.089490
1	NO	0.520645
	>30	0.355385
	<30	0.123970
2	NO	0.602864
	>30	0.298540
	<30	0.098595
3	NO	0.541570
	>30	0.356157
	<30	0.102273
4	>30	0.514813
	<30	0.254206
	NO	0.230980

Thank You !!