PREDICTING HOSPITAL READMISSIONS AND PATIENT OUTCOMES USING MIMIC-III AND DIABETES DATA

Predicting Patient Outcomes: Uncovering Insights from Hospital Readmissions

CAPSTONE PROJECT

Project Overview :

The goal of this project is to analyze hospital readmissions and patient outcomes by combining two datasets:

- Diabetes 130-US Hospitals Dataset (Readmissions Data).
- MIMIC-III Clinical Database (ICU Data).

This will allow us to investigate which clinical factors affect readmissions, explore the relationships between various patient-level data, and build models to predict patient outcomes based on combined datasets.

Data Preview

The datasets given for the analysis includes the data of 10 years (1999-2008) covering more than 1,00,000 patients from 130 hospitals in United States.

- Diabetes 130-US hospitals dataset (readmissions data)

  Total Rows: 101766

  Total Columns : 50

- MIMIC-III Clinical Database (ICU data)

  Total Rows: 100000

  Total Columns : 18

Diabetes 130-US Hospitals Dataset (Readmissions Data)

- Patient identifiers: - encounter _id, patient_nbr
- Patient demographics:- race, gender, age, 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_medicationsClinical
- Results :- max_glu_serum, Alcresult
- Medication Details :- diabetesMed, change, 23 features for medications
- Readmission Indicator: - readmitted

MIMIC-III Clinical Database (ICU data)

- Patient identifiers: - Patient_ID, ICU_Admission_ID
- Other Derails: - ICU_Length_of_Stay , Diagnoses, Number_of_Lab_Tests
- Vital Signs: - Blood_Glucose, Hemoglobin, WBC, Heart_Rate, Blood_Pressure_Systolic, Blood_Pressure_Diastolic, SpO2, Respiratory_Rate, Temperature
- Medication Details :- Medications, Number_of_Medications
- Readmission Indicator: - Readmission Flag

Data Preprocessing -

- Identified null values in the data sets.
- Filled the '?' and other null values with respective column and group of columns using data engineering.
- Dropped the irrelevant columns and columns with high missing values as weight, payer code etc.
- Used Data Featuring techniques for the exploratory data analysis.

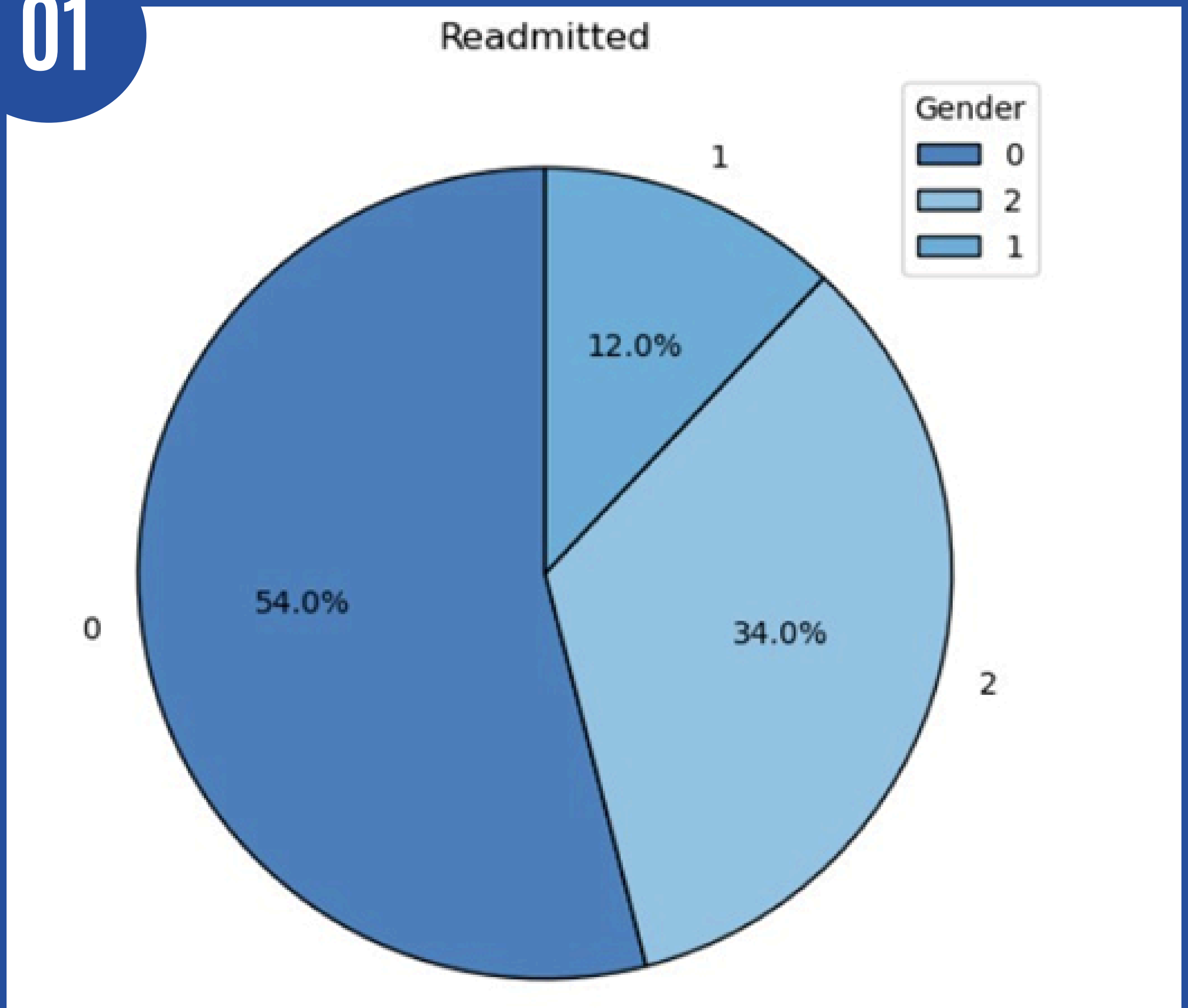
Challenges Encountered -

- Major challenge was to acquire domain knowledge of the medical world.
- Uncovered critical factors driving readmissions through data integration.
- Imbalanced data
- Different imputation methods
- Reductant Features

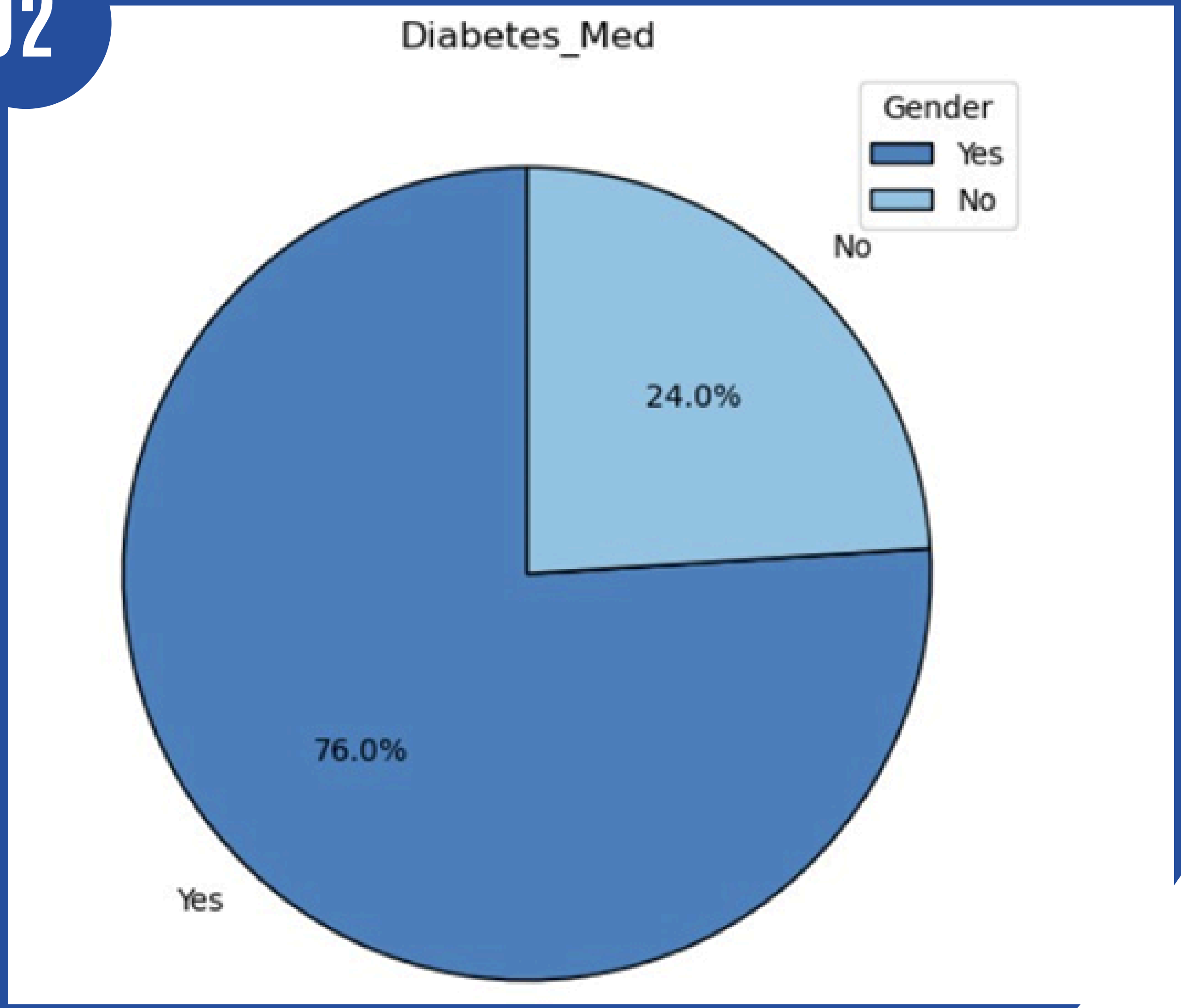
EXPLORATORY DATA ANALYSIS

UNIVARIATE ANALYSIS

01



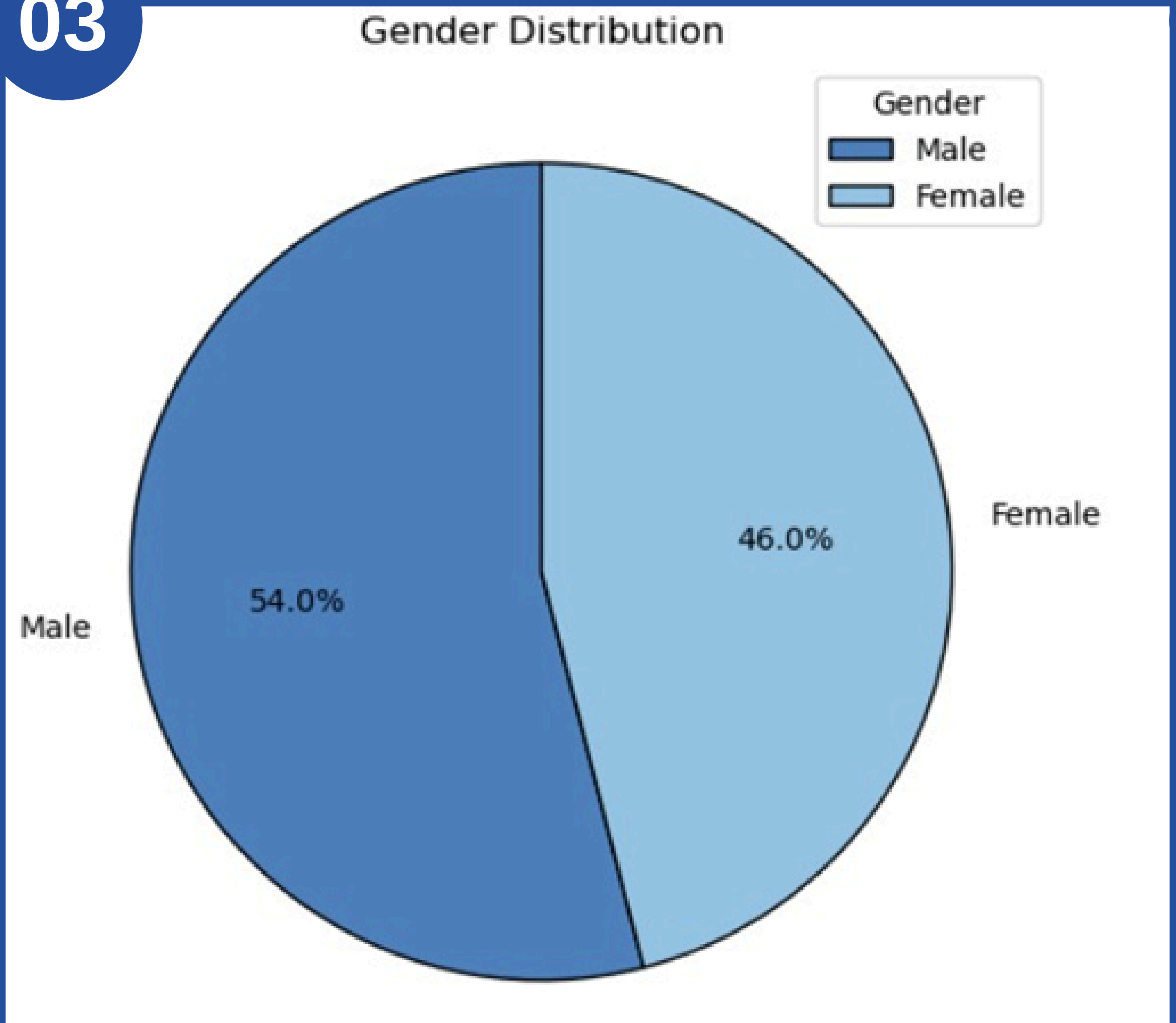
02



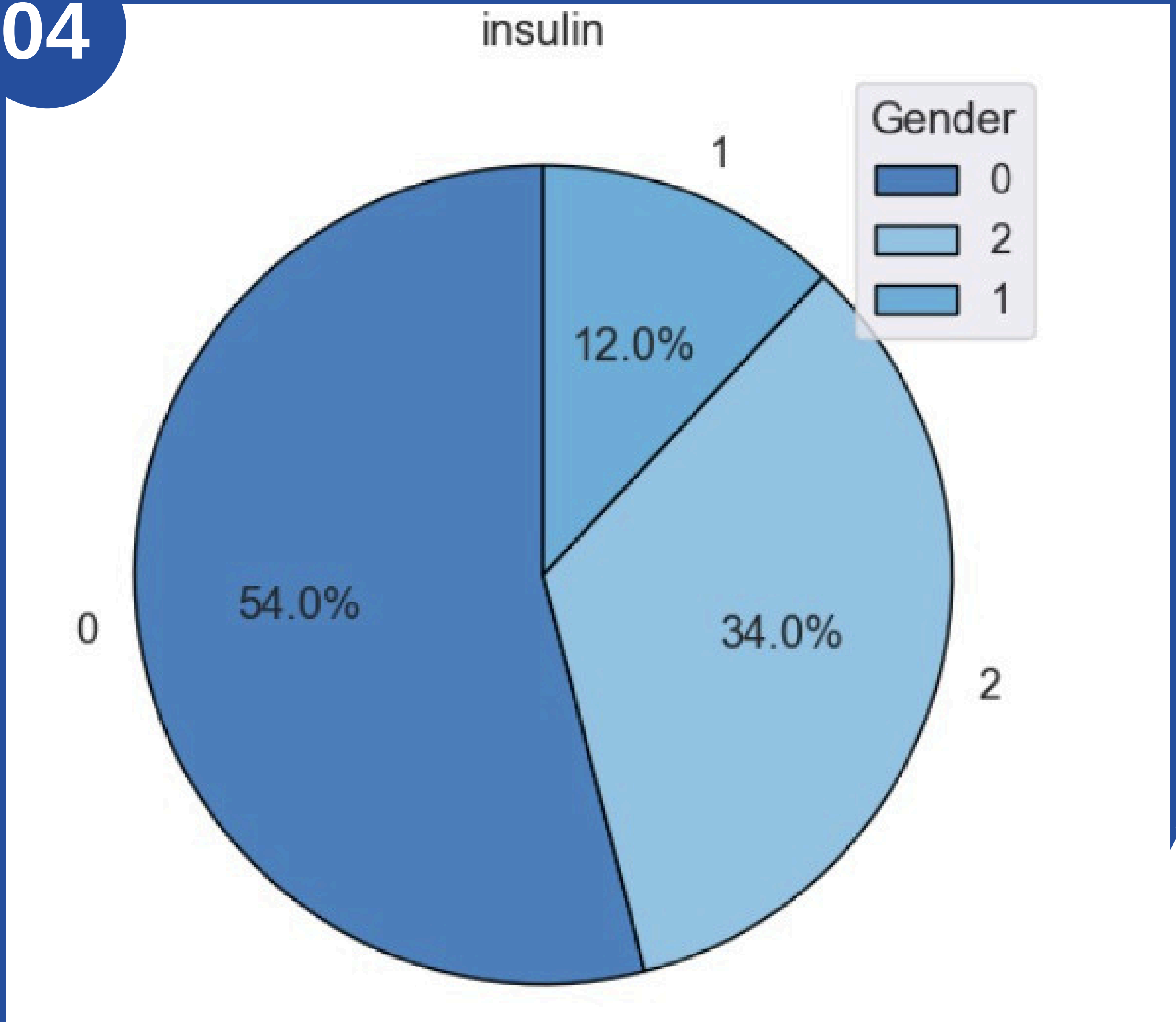
2 = >30 readmissions 48594
1 = <30 readmissions 15930
0 = no readmissions 76801

UNIVARIATE ANALYSIS

03

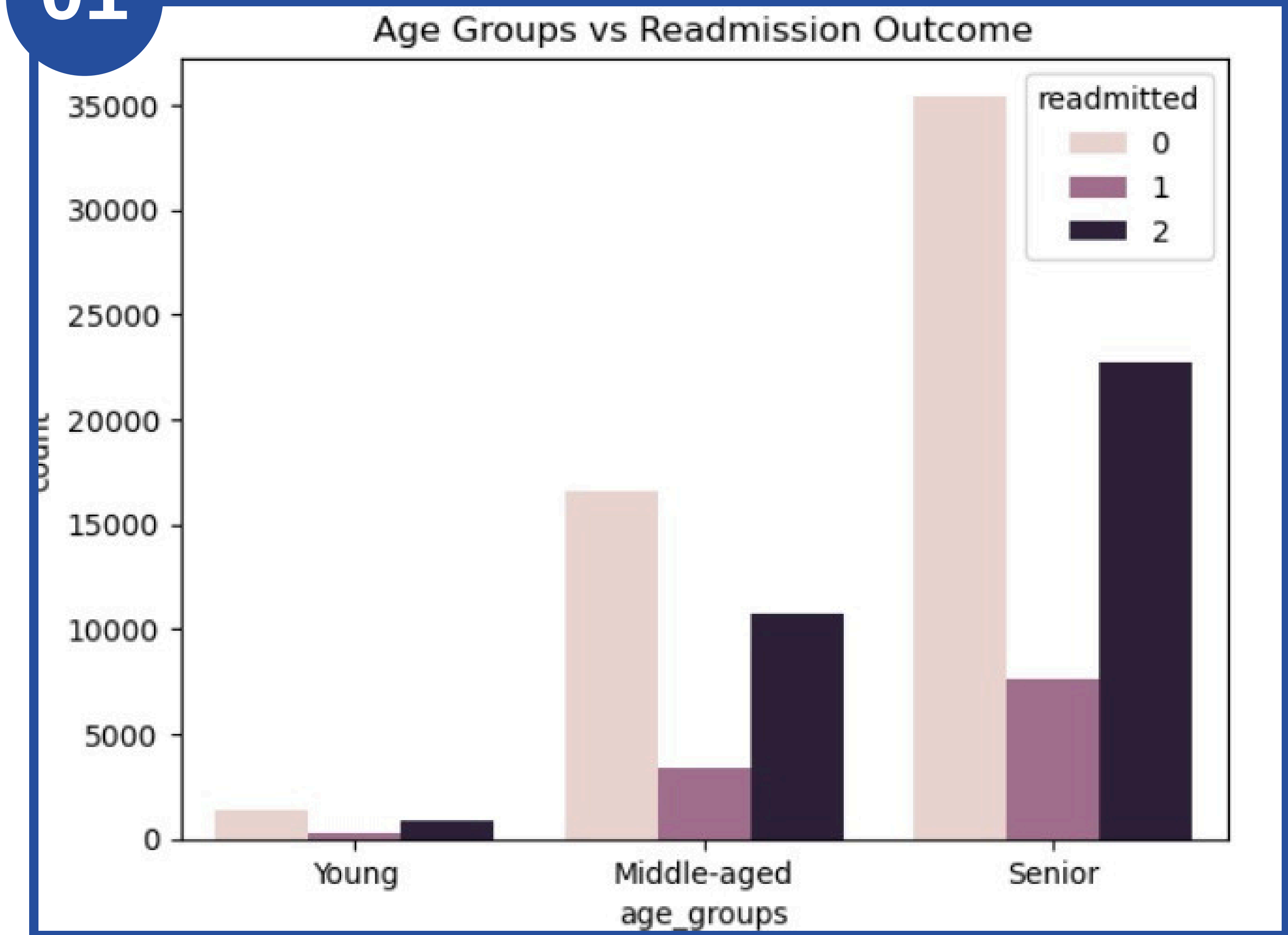


04

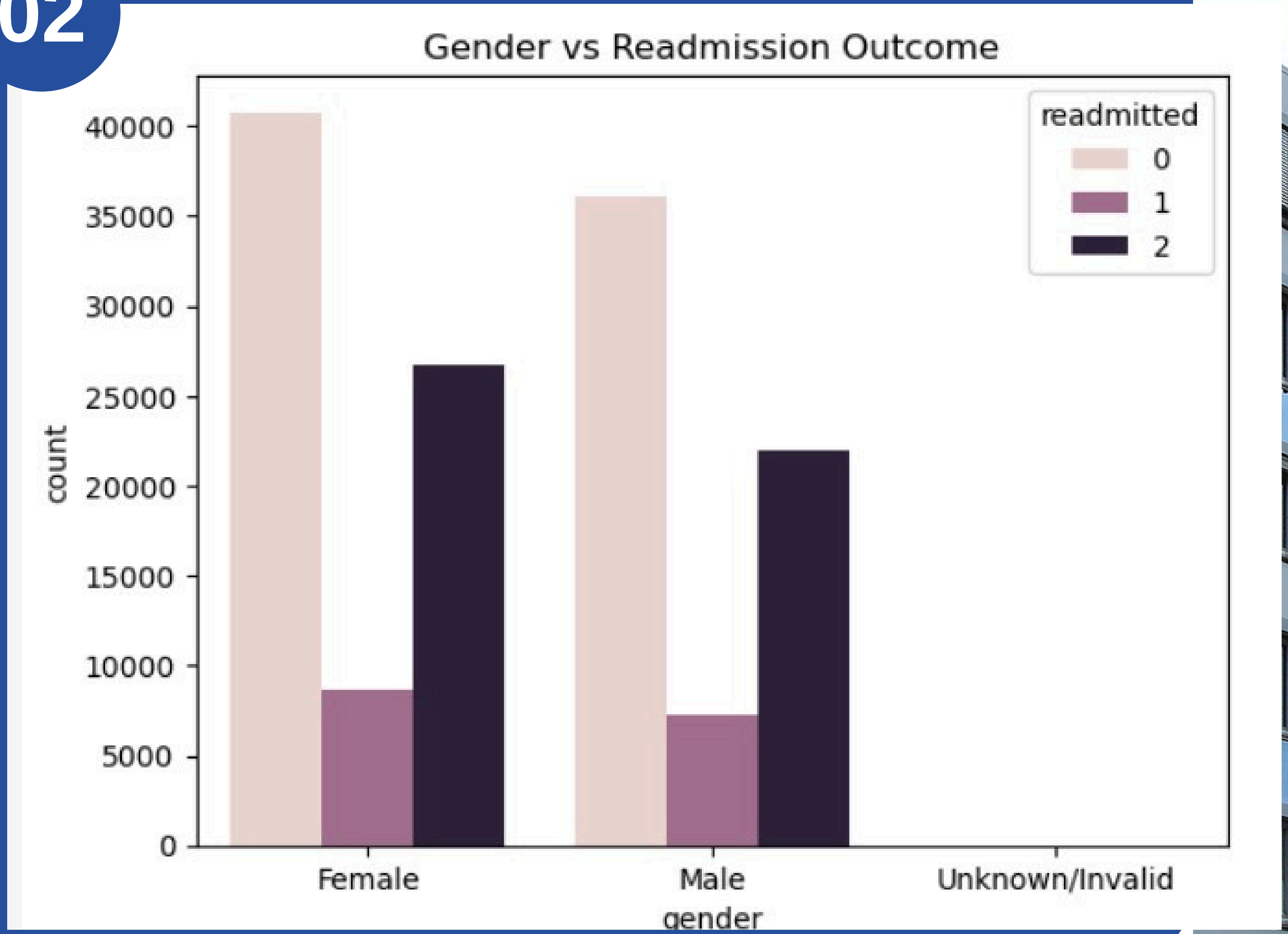


BIVARIATE ANALYSIS

01



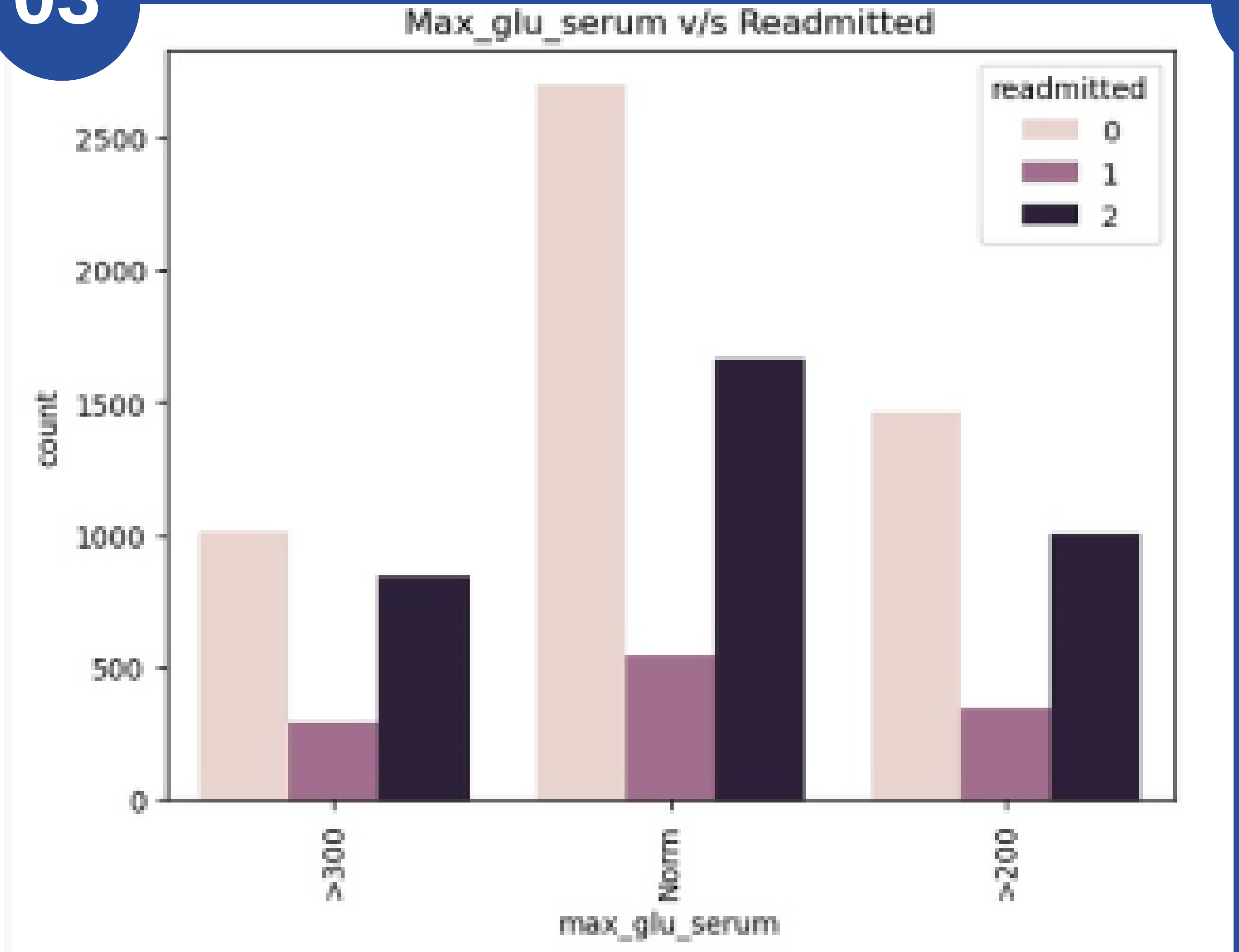
02



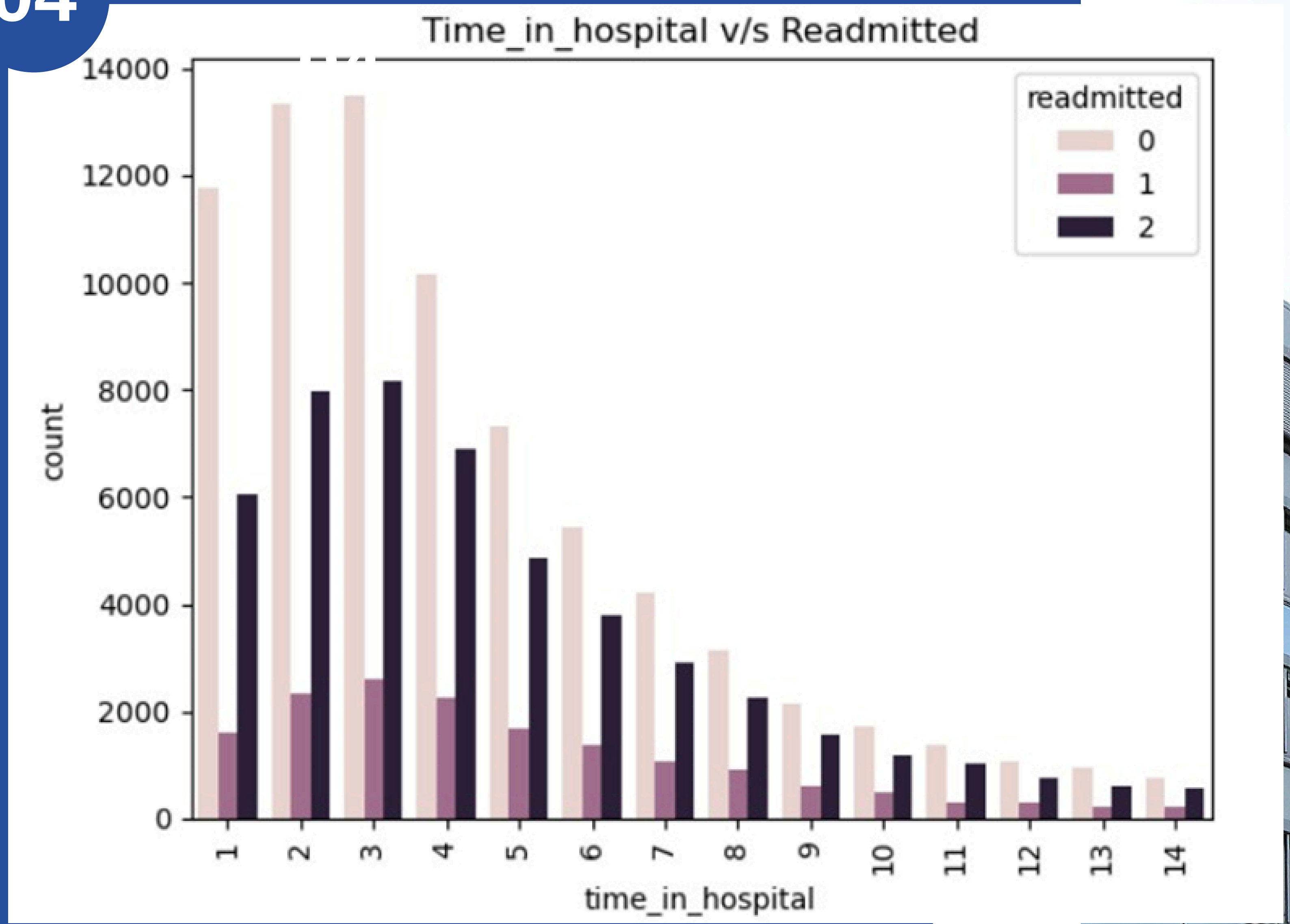
2 = >30 readmissions 48594
1 = <30 readmissions 15930
0 = no readmissions 76801

BIVARIATE ANALYSIS

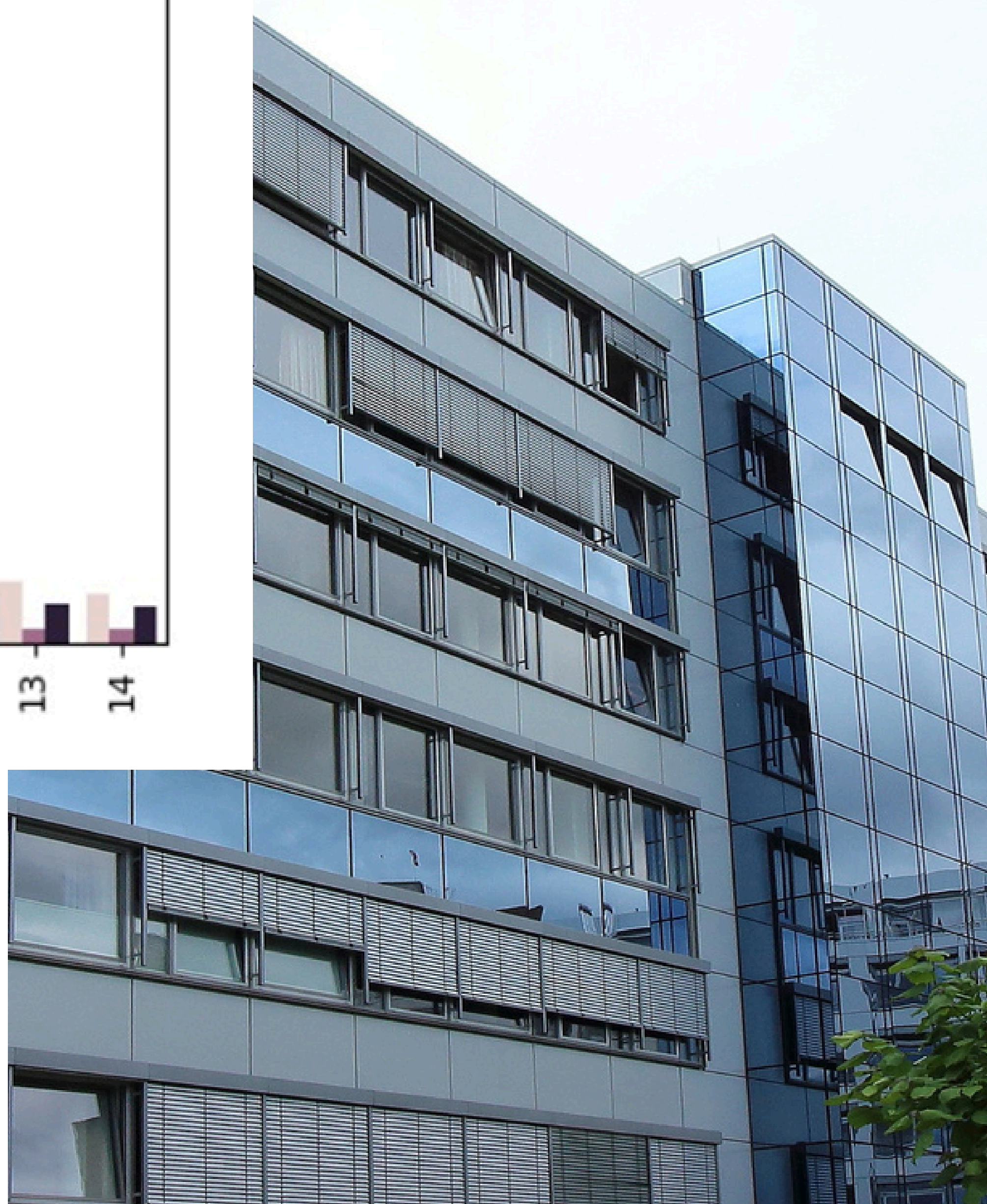
03



04

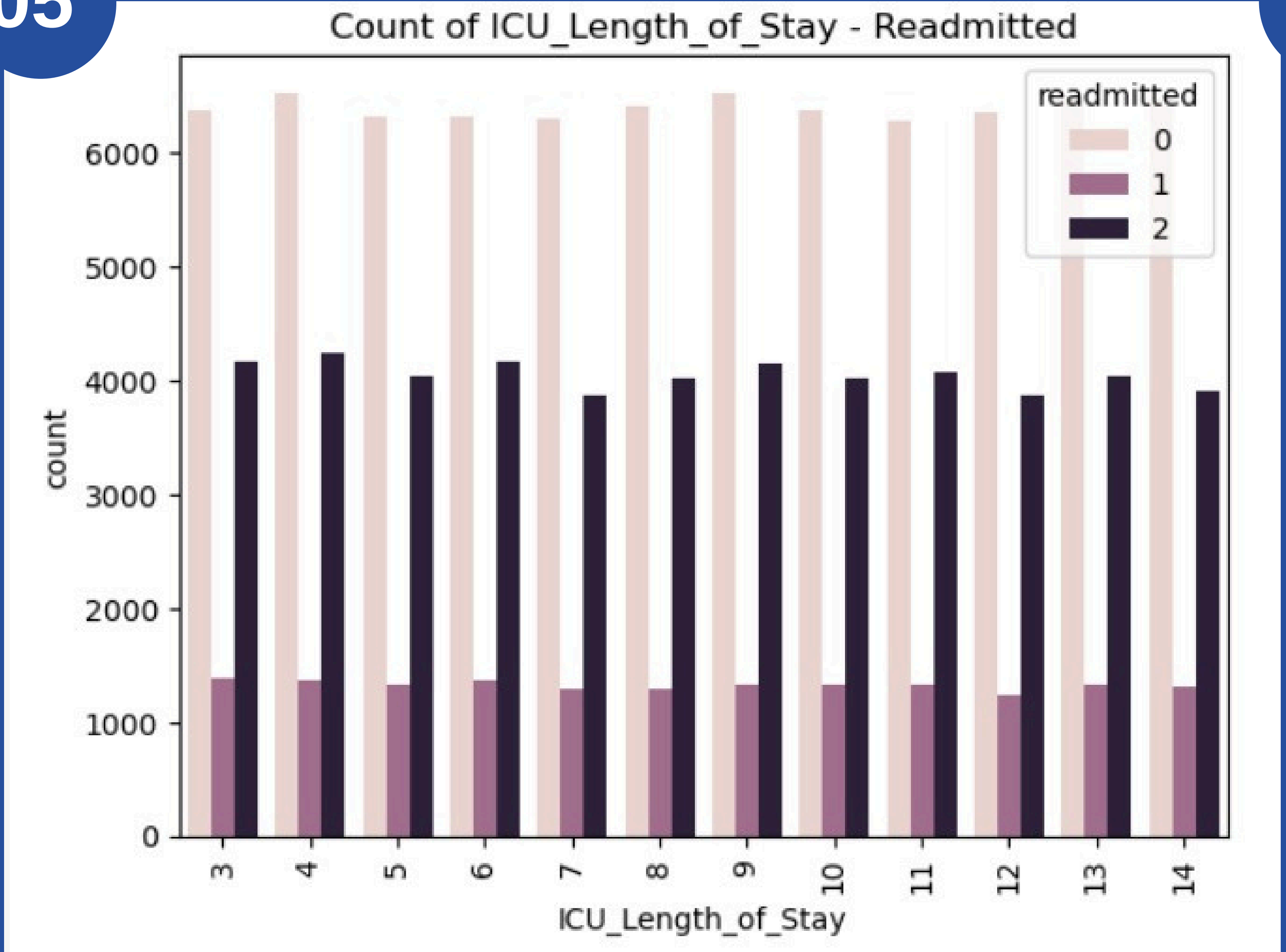


2 = >30 readmissions 48594
1 = <30 readmissions 15930
0 = no readmissions 76801

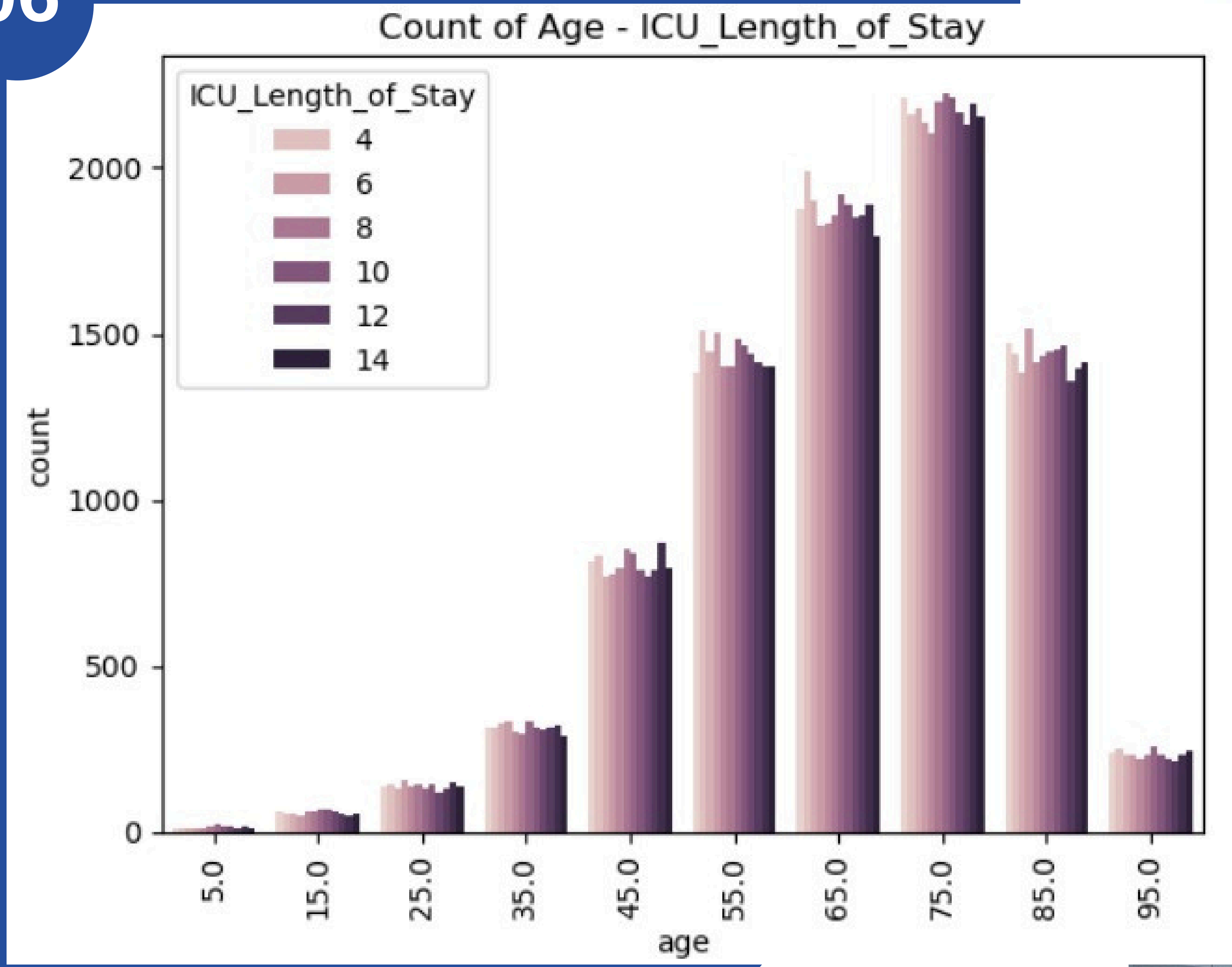


BIVARIATE ANALYSIS

05

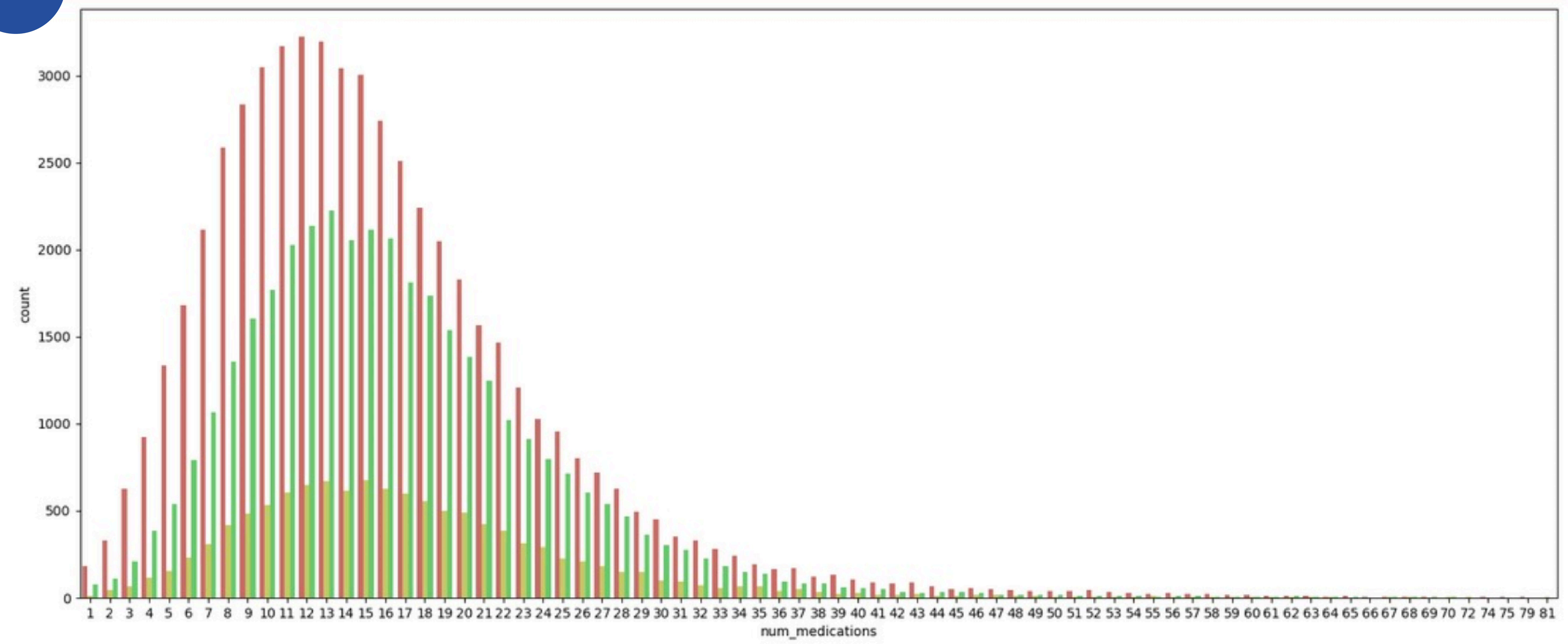


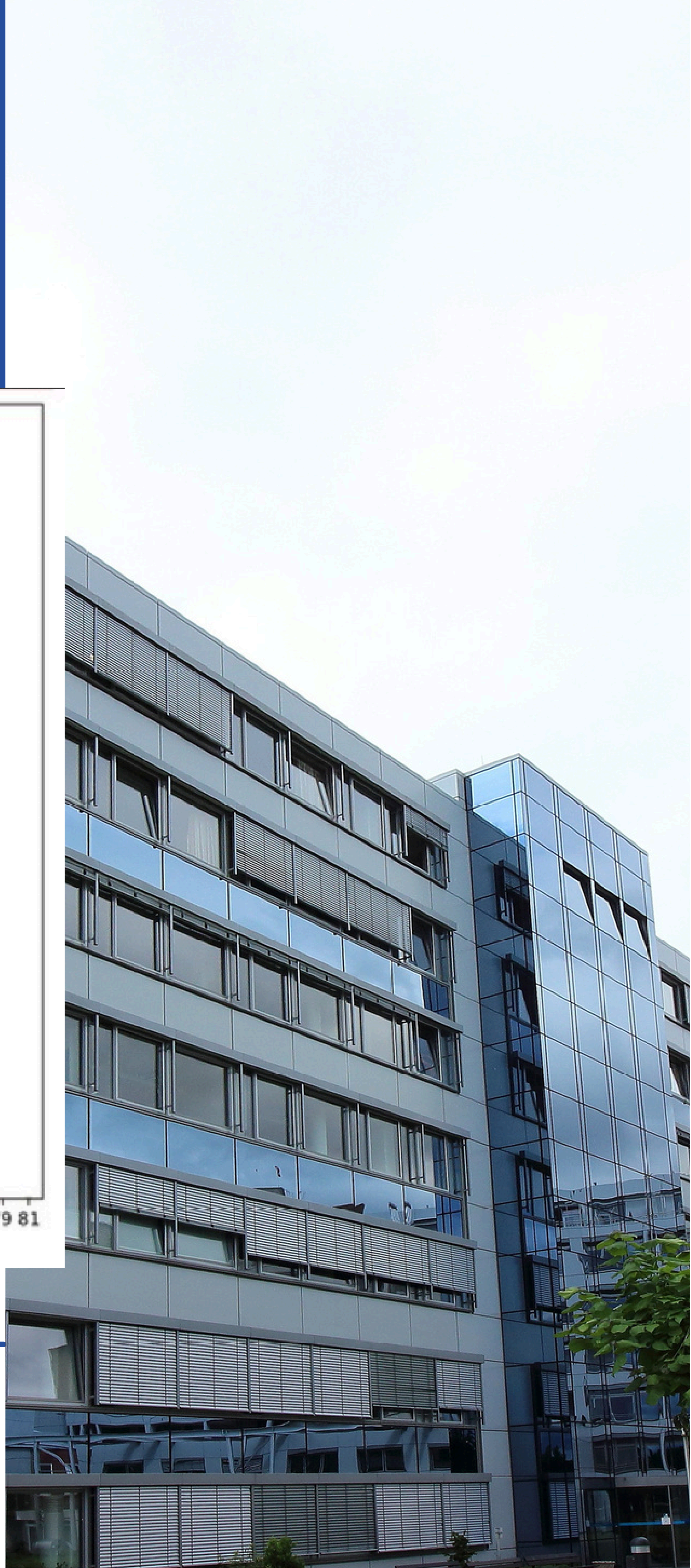
06



BIVARIATE ANALYSIS

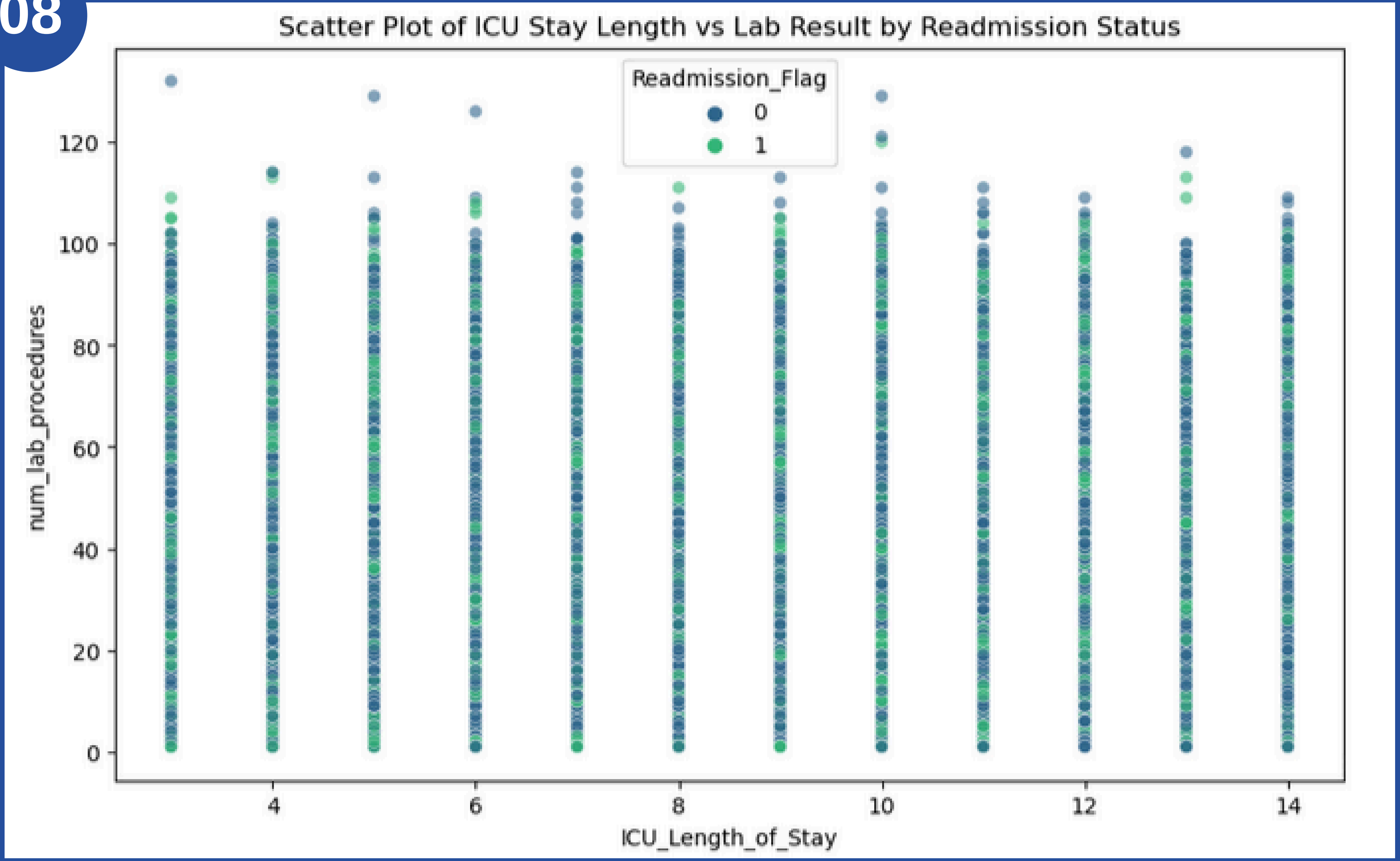
07



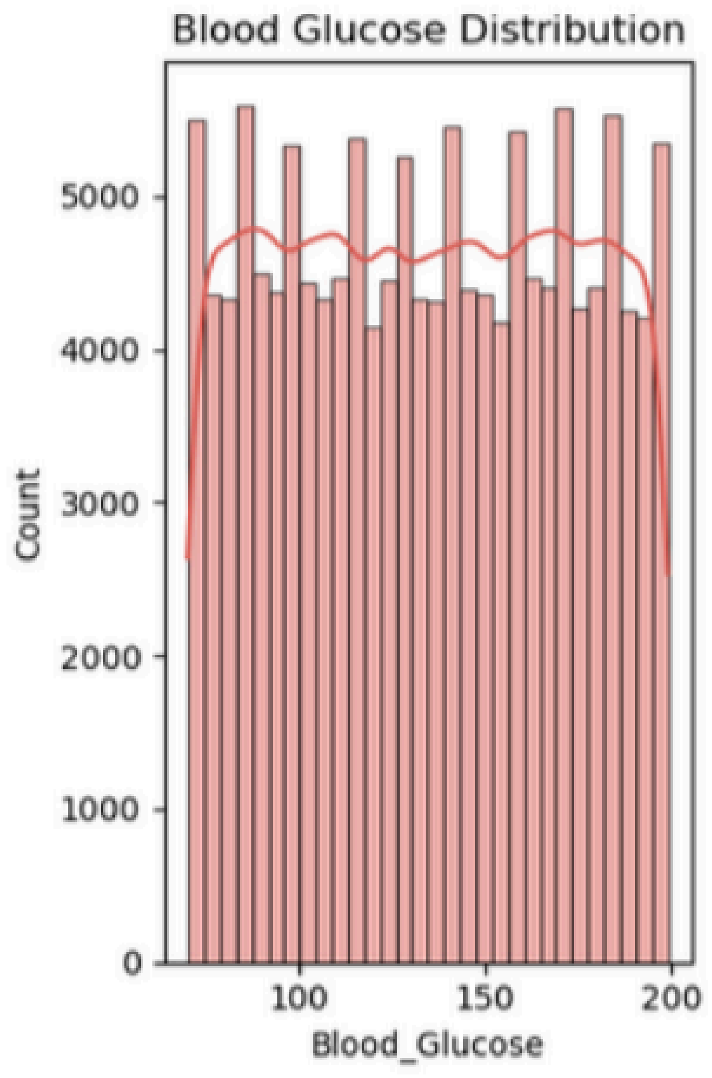


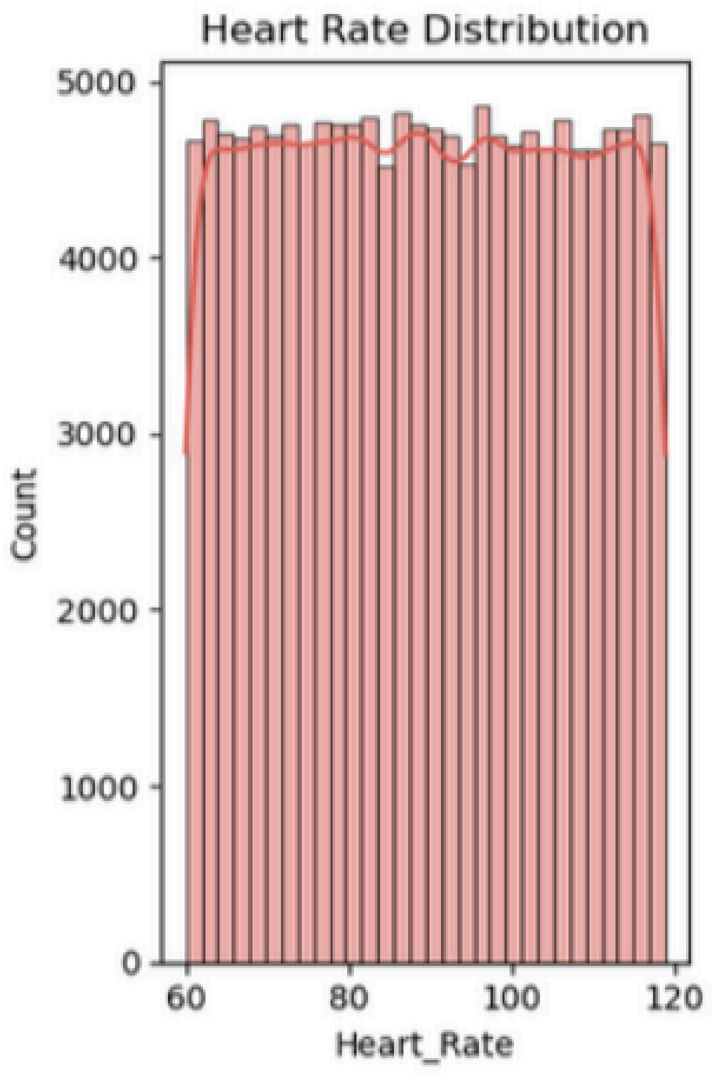
BIVARIATE ANALYSIS

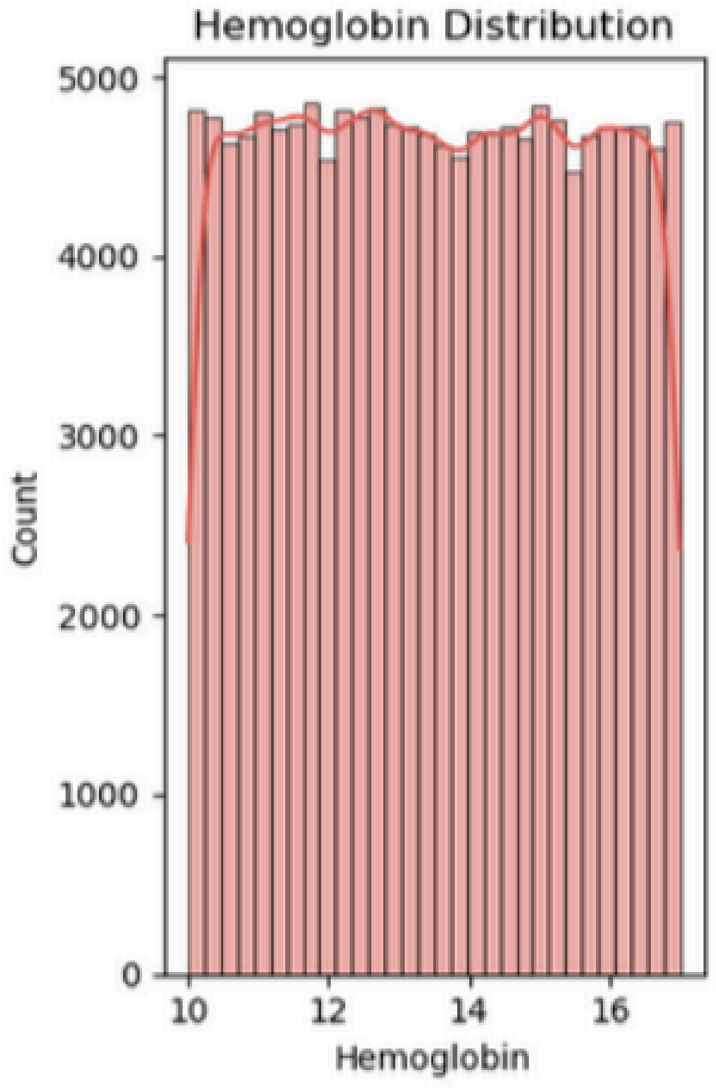
08

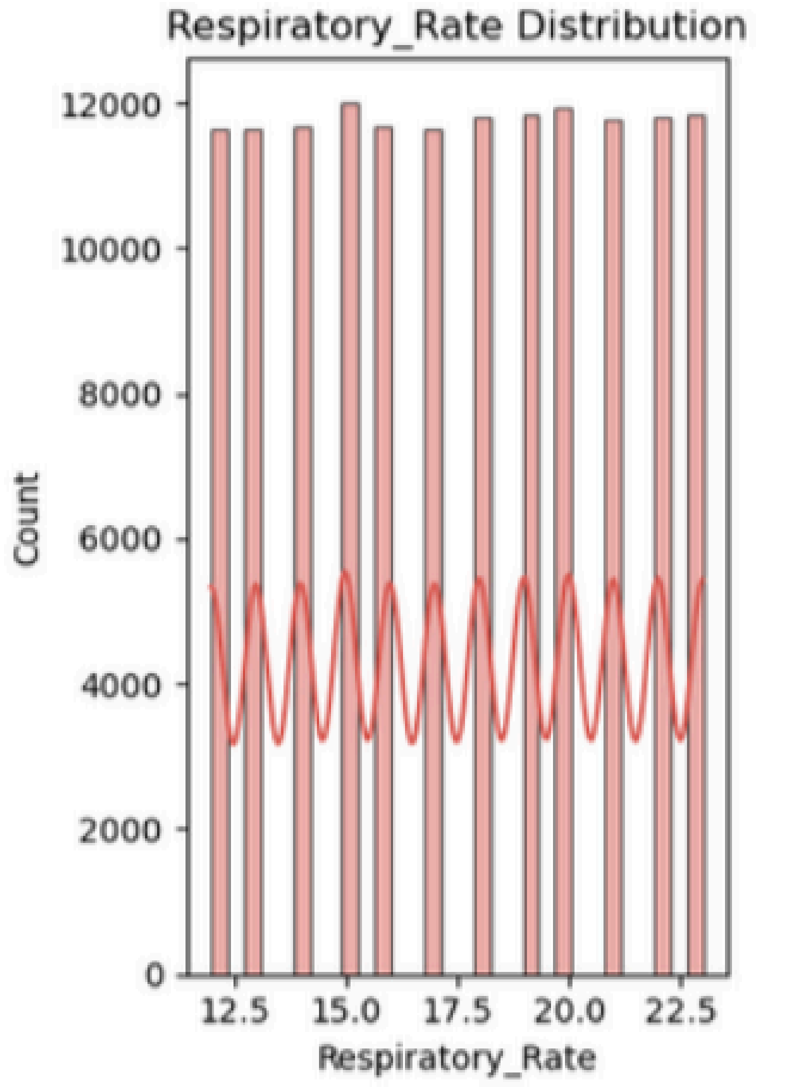


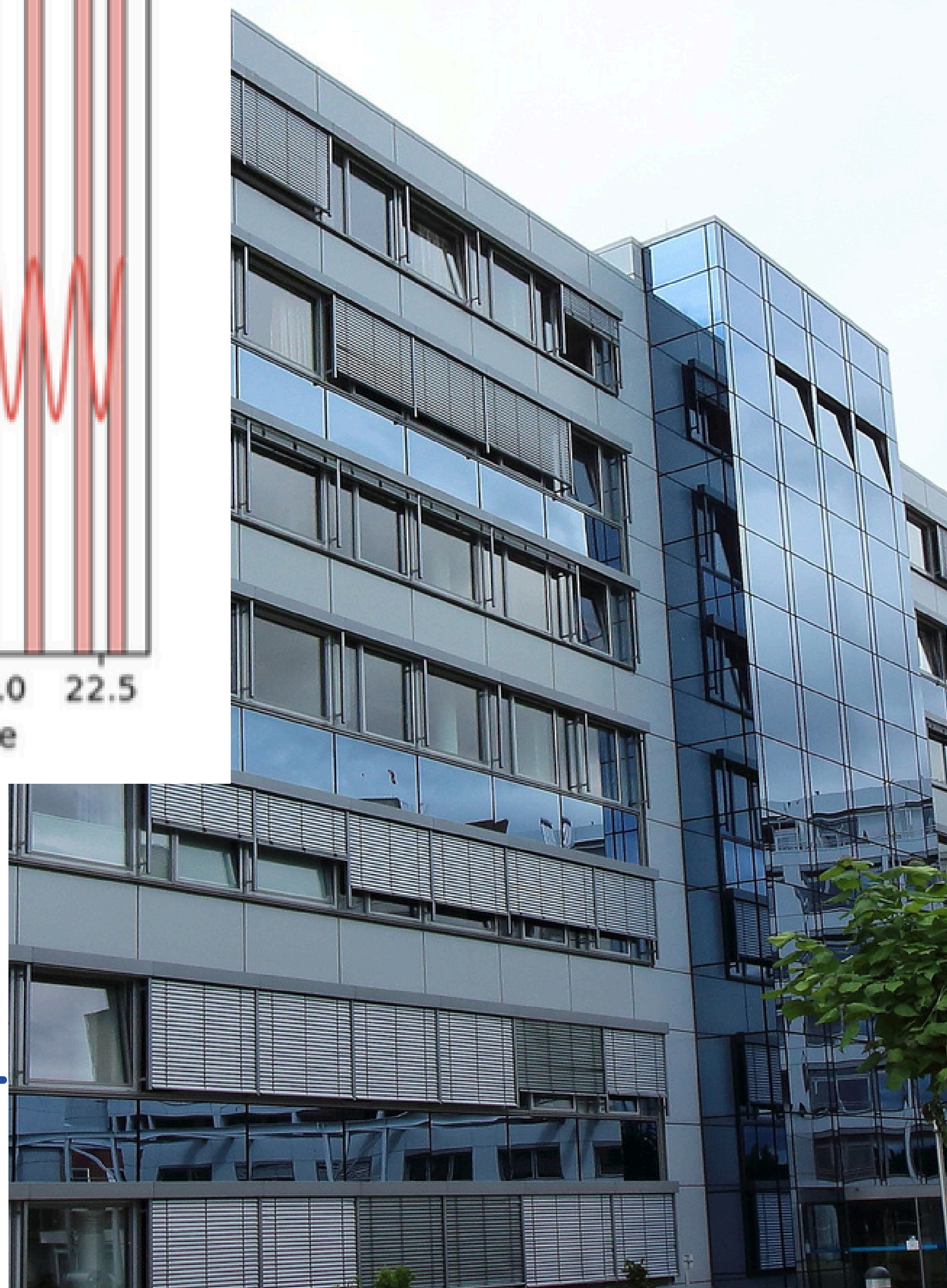
DISTRIBUTION OF VITAL SIGNS



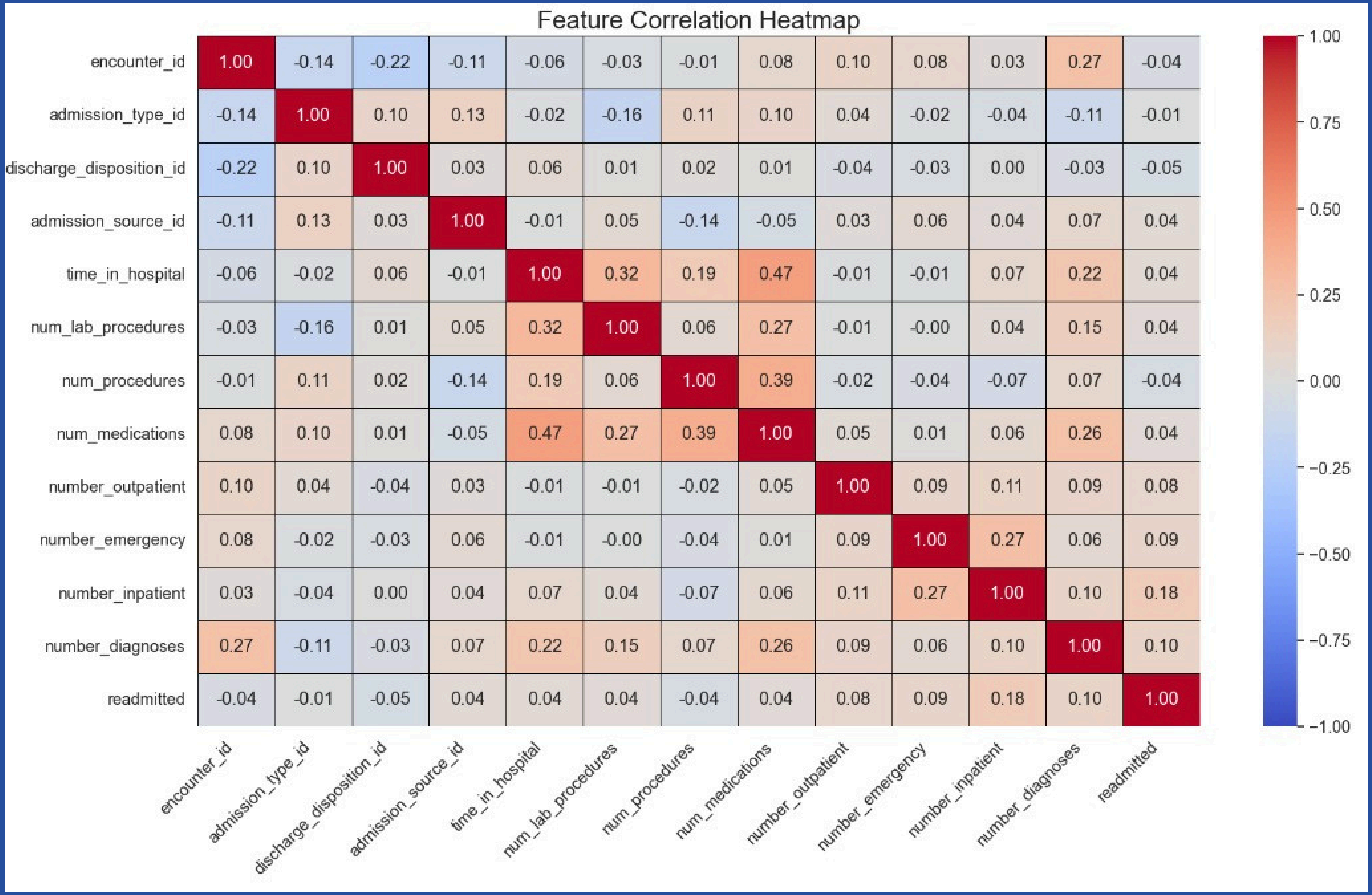








CORRELATION



1. Strong positive correlations:

- num_medications with time_in_hospital (0.47)
- num_medications with num_lab_procedures (0.27)
- num_lab_procedures with time_in_hospital (0.32)
- number_inpatient with number_diagnoses (0.27)

2. Target variable (readmitted) analysis:

- Weak positive correlations with number_inpatient (0.18) and number_diagnoses (0.10)
- Minimal to no correlation with other features

MODEL PERFORMANCE COMPARISON

Models	Accuracy	Balanced Accuracy	ROC_AUC	F1 Score
RandomForestClassifier	1.00	1.00	None	1.00
LogisticRegression	1.00	1.00	None	1.00
DecisionTreeClassifier	1.00	1.00	None	1.00
LGBMClassifier	1.00	1.00	None	1.00
KNeighborsClassifier	0.98	0.95	None	0.98

Insights and Findings from Exploratory Data Analysis

- num_medications, time_in_hospital, and num_lab_procedures are key indicators of extended ICU stays or hospital time
- number_inpatient and number_diagnoses play minor roles in predicting readmission
- The majority of patients were not readmitted (indicative of successful treatments or lower risk patients)
- Longer stays in ICU (evidenced by correlations) are associated with older demographics and multiple diagnoses

Recommendations

- Implement remote monitoring programs for high-risk patients' post-discharge, especially those with frequent inpatient visits or multiple diagnoses.
- Increase focus on preventive care for the elderly to reduce ICU admissions.
- Introduce personalized treatment plans to reduce unnecessary medication use and prevent adverse drug interactions.
- Promote early diagnostic tests and annual health check-ups for high-risk patients to detect complications early.
- Create age-specific care plans for this demographic, emphasizing proactive health screenings and managing comorbidities.

THANK YOU