Data Collection and Preprocessing Phase

Date	20 JUNE 2024
Team ID	team-739701
Project Title	HOSPITAL READMISSION PREDICTOR USING MACHINE LEARNING
Maximum Marks	6 Marks

Data Exploration and Preprocessing Report

Dataset variables will be statistically analyzed to identify patterns and outliers, with Python employed for preprocessing tasks like normalization and feature engineering. Data cleaning will address missing values and outliers, ensuring quality for subsequent analysis and modeling, and forming a strong foundation for insights and predictions.

Section	Description
Data Overview	Dimension: 5 rows × 29columns Descriptive statistics: _5 rows × 29 columns_

Data Overview table:

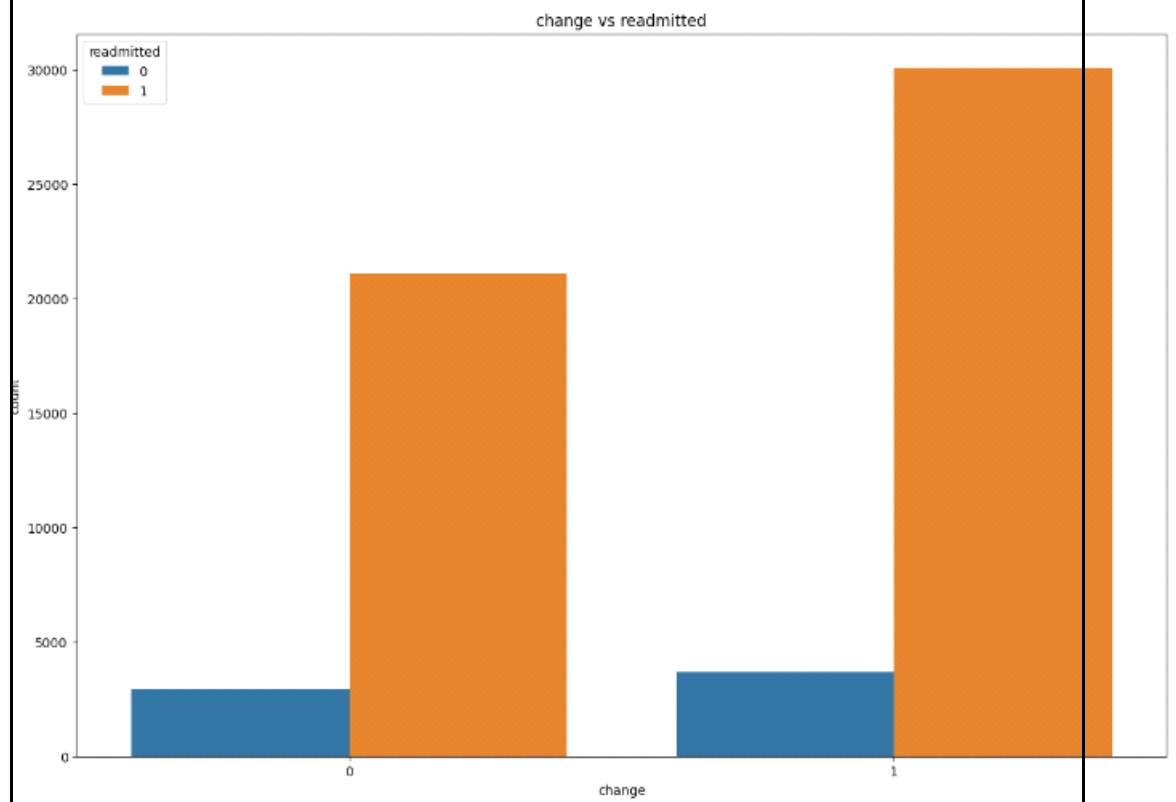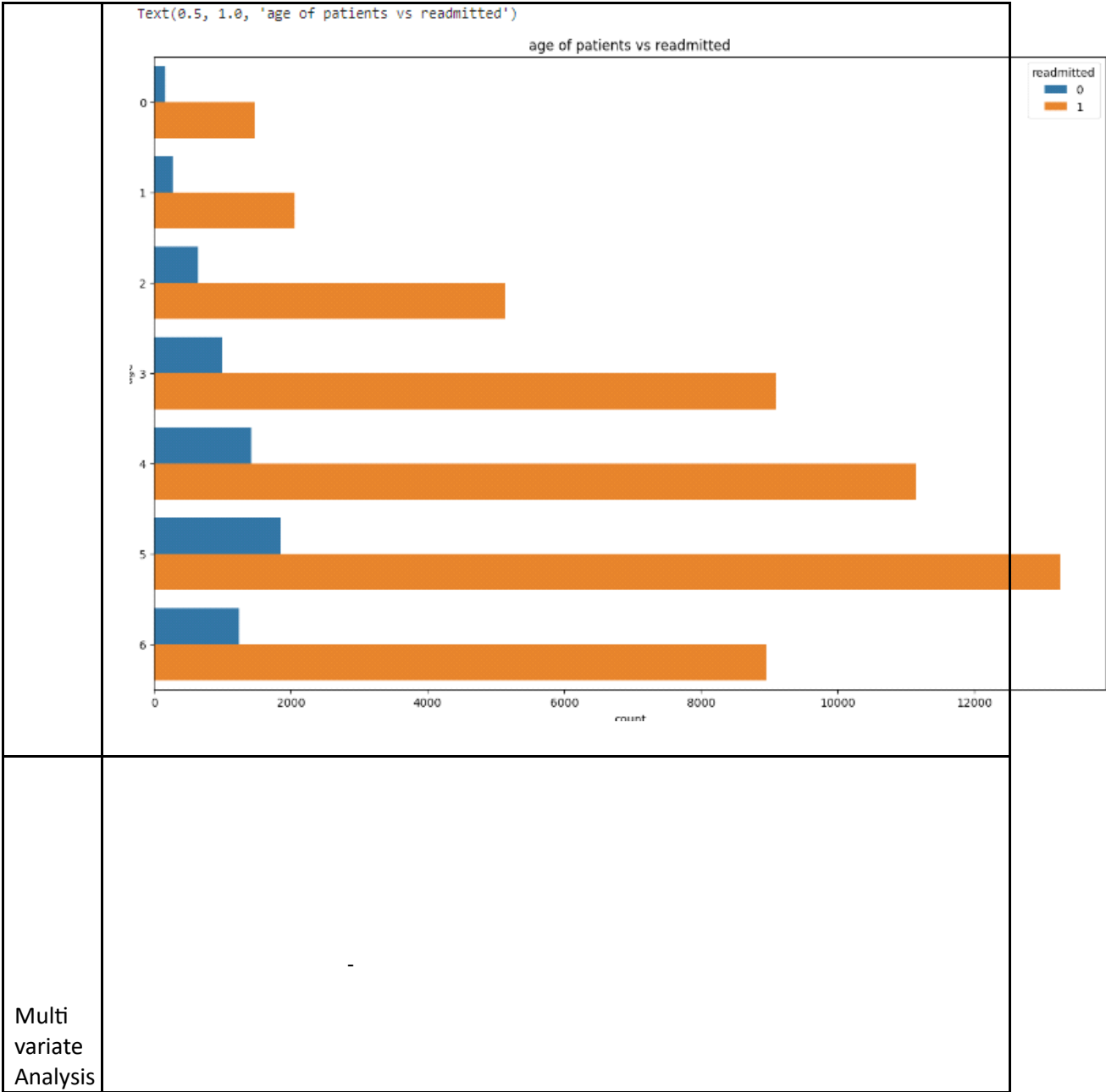|	race	gender	age	admission_type_id	discharge_disposition_id	admission_source_id	time_in_hospital	num_lab_procedure
0	Caucasian	Female	[0-30)	Others	Others	Referral	-1.632460	-0.10847
1	Caucasian	Female	[0-30)	Emergency	Discharged to home	Emergency Room	-0.271144	0.80852
2	AfricanAmerican	Female	[0-30)	Emergency	Discharged to home	Emergency Room	-0.825587	-1.62961
3	Caucasian	Male	[30-40)	Emergency	Discharged to home	Emergency Room	-0.825587	0.04419
4	Caucasian	Male	[40-50)	Emergency	Discharged to home	Emergency Room	-1.632460	0.40069

5 rows × 29 columns

| Univariate Analysis | |
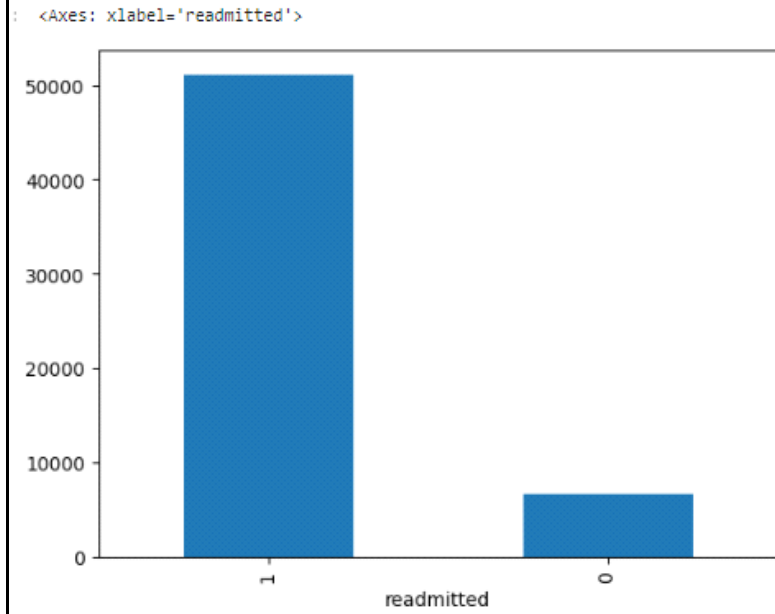
	-
Bivariate Analysis	

Text(0.5, 1.0, 'diabetesMed vs readmitted')

diabetesMed vs readmitted
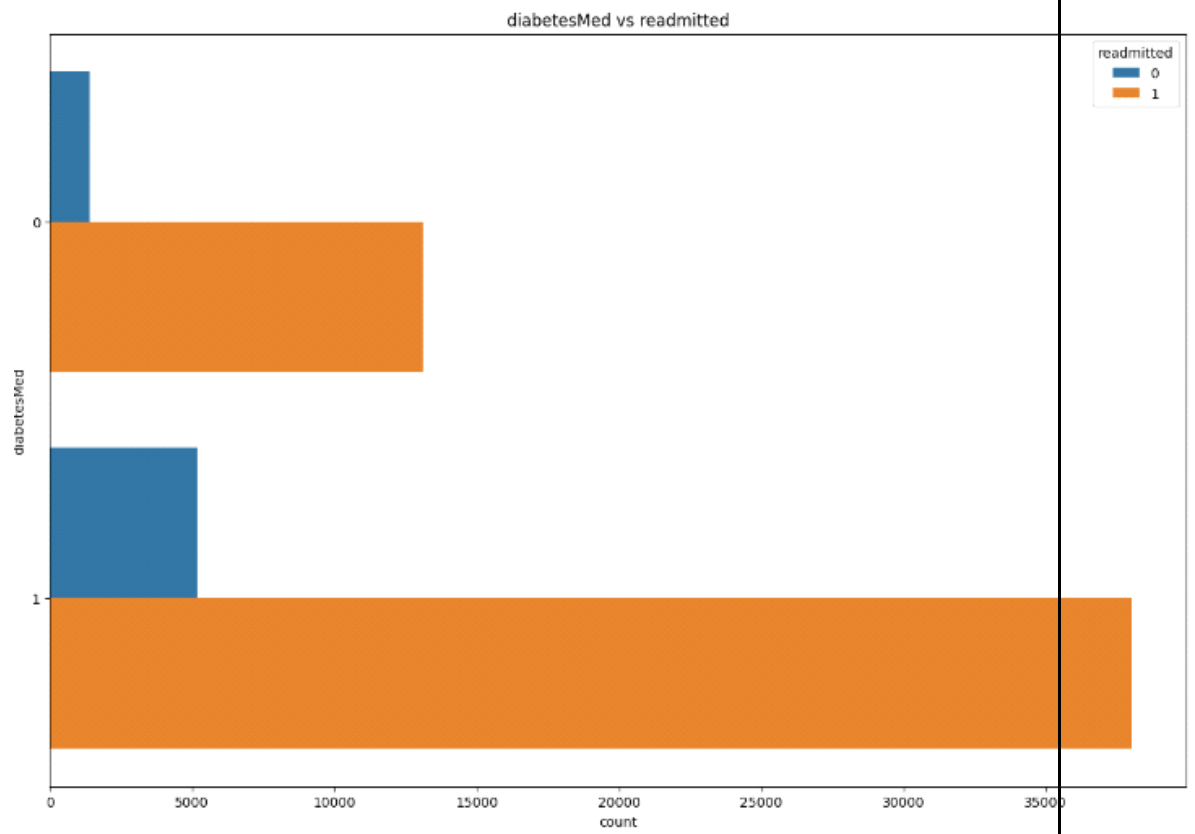
Text(0.5, 1.0, 'change vs readmitted')

change vs readmitted

	Text(0.5, 1.0, 'age of patients vs readmitted')
Multi variate Analysis	-

Outliers and Anomalies	-

Data Preprocessing Code Screenshots

Loading Data	

	race	gender	age	admission_type_id	discharge_disposition_id	admission_source_id	time_in_hospital	num_lab_procedu...
0	Caucasian	Female	[0-30)	Others	Others	Referral	-1.632460	-0.1084
1	Caucasian	Female	[0-30)	Emergency	Discharged to home	Emergency Room	-0.271144	0.8085
2	AfricanAmerican	Female	[0-30)	Emergency	Discharged to home	Emergency Room	-0.825587	-1.629
3	Caucasian	Male	[30-40)	Emergency	Discharged to home	Emergency Room	-0.825587	0.044
4	Caucasian	Male	[40-50)	Emergency	Discharged to home	Emergency Room	-1.632460	0.480

| Handling Missing Data | ```python
data['num_medications'].fillna(data['num_medications'].mean(),inplace=True)

data['number_outpatient'].fillna(data['number_outpatient'].mean(),inplace=True)

data['number_emergency'].fillna(data['number_emergency'].mean(),inplace=True)

data['number_diagnoses'].fillna(data['number_diagnoses'].mean(),inplace=True)

data['age_derived'].fillna(data['age_derived'].mean(),inplace=True)

data['count_Steady'].fillna(data['count_Steady'].mean(),inplace=True)

data['count_Down'].fillna(data['count_Down'].mean(),inplace=True)

data['count_Up'].fillna(data['count_Up'].mean(),inplace=True)

data['number_inpatient'].fillna(data['number_inpatient'].mean(),inplace=True)
``` |
| Data Transformation | ```python
le=LabelEncoder()

data['gender']=le.fit_transform(data['gender'])

data['age']=le.fit_transform(data['age'])

data['admission_type_id']=le.fit_transform(data['admission_type_id'])

data['discharge_disposition_id']=le.fit_transform(data['discharge_disposition_id'])

data['admission_source_id']=le.fit_transform(data['admission_source_id'])

data['diag_1']=le.fit_transform(data['diag_1'])

data['diag_2']=le.fit_transform(data['diag_2'])

data['diag_3']=le.fit_transform(data['diag_3'])

data['metformin']=le.fit_transform(data['metformin'])

data['repaglinide']=le.fit_transform(data['repaglinide'])

data['glipizide']=le.fit_transform(data['glipizide'])

data['insulin']=le.fit_transform(data['insulin'])

data['change']=le.fit_transform(data['change'])

data['diabetesMed']=le.fit_transform(data['diabetesMed'])

data['readmitted']=le.fit_transform(data['readmitted'])
``` |

Feature Engineering	Attached the codes in final submission