

# DIABETES MEDICATION PREDICTION

## Problem Statement

- Based on the patient food habits and hospitalization records, we need to understand whether a patient will require diabetes medication

## Task

- Based on a dataset of diabetes patients' history, attributes and hospital admission predict if the patient will require diabetes medication

# DIABETES MEDICATION PREDICTION

## About the dataset

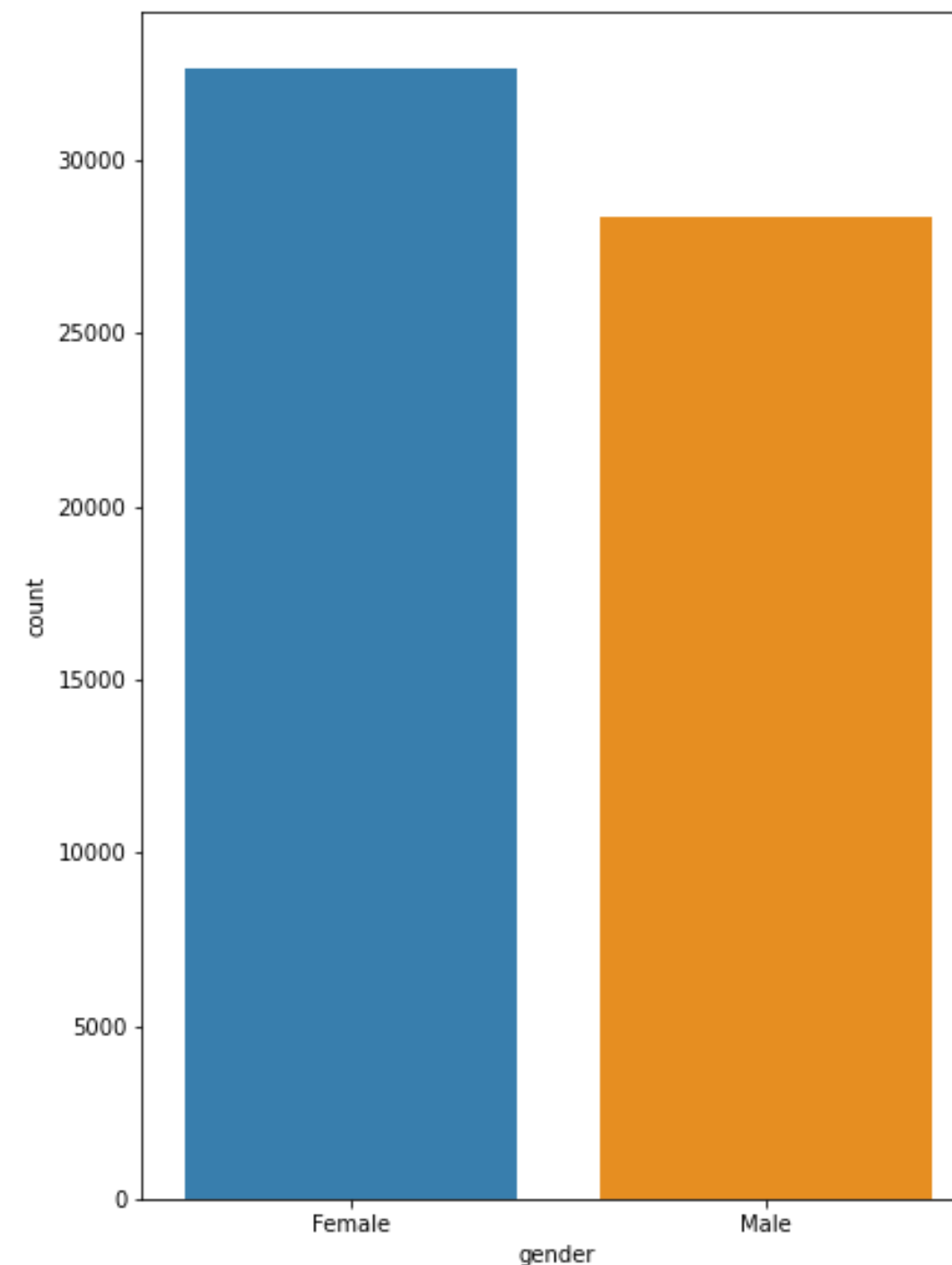
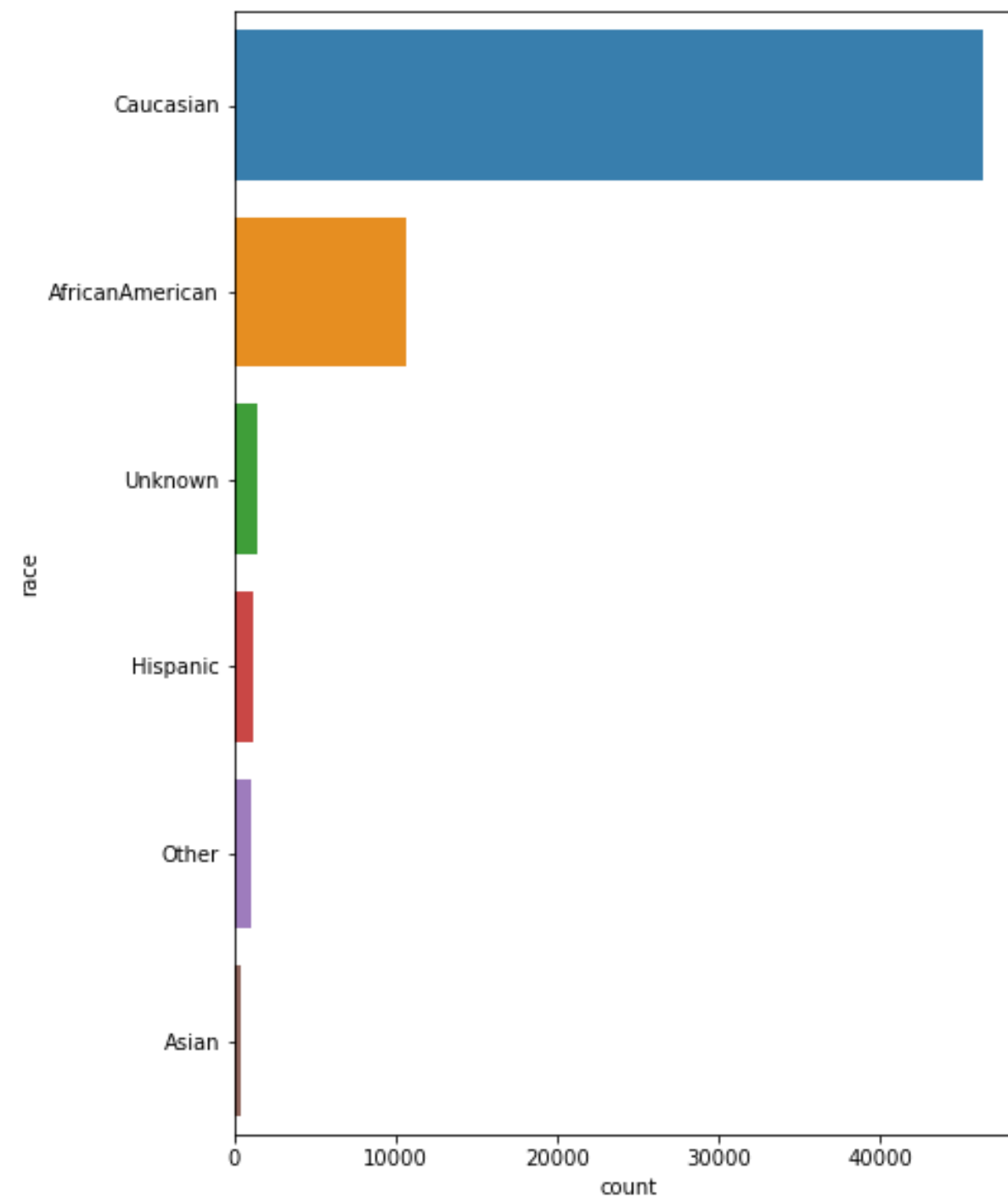
- encounter\_id : A calculated unique ID for each encounter with the patient
- Patient\_id: Unique ID for each patient
- race: Patient race
- gender: Patient gender
- age: Patient age
- weight: Patient weight
- Admission\_type\_id: The ID assigned while taking admission in the hospital
- Discharge\_disposition\_id: The ID assigned while discharging
- Admission\_source\_id; The ID of the physician for whom the patient got admitted
- Time\_in\_hospital : Time spent by the patient in the hospital
- tel\_1 - tel\_49 : Anonymous variables
- diabetesMed : Two unique values, Yes or NO, representing if the patient needs medicines for diabetes or not

## Shape of the dataset

- 61,060 instances and 50 features in the dataset

# DIABETES MEDICATION PREDICTION

## Exploratory Data Analysis

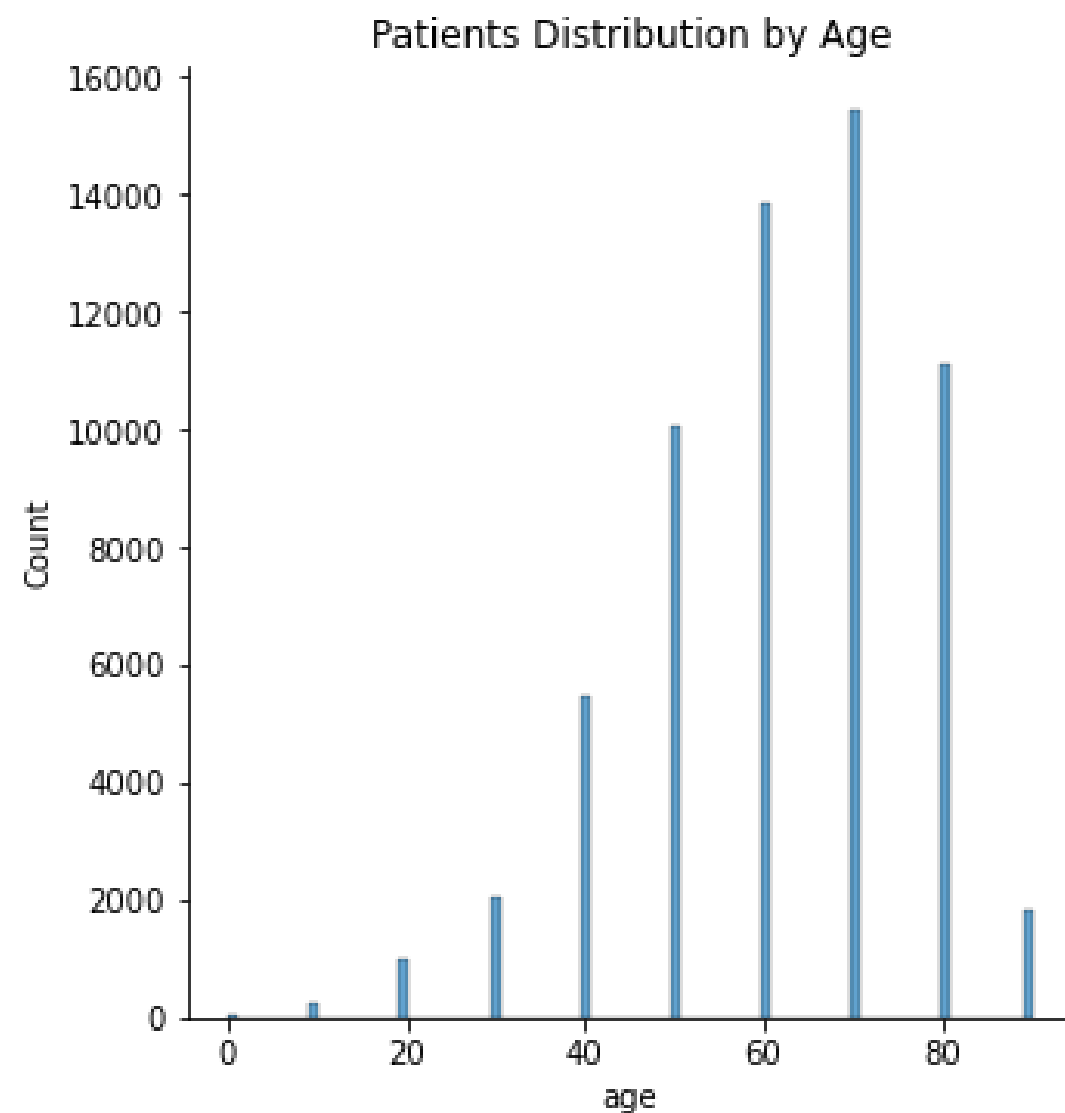


## Patient count by Race and Gender

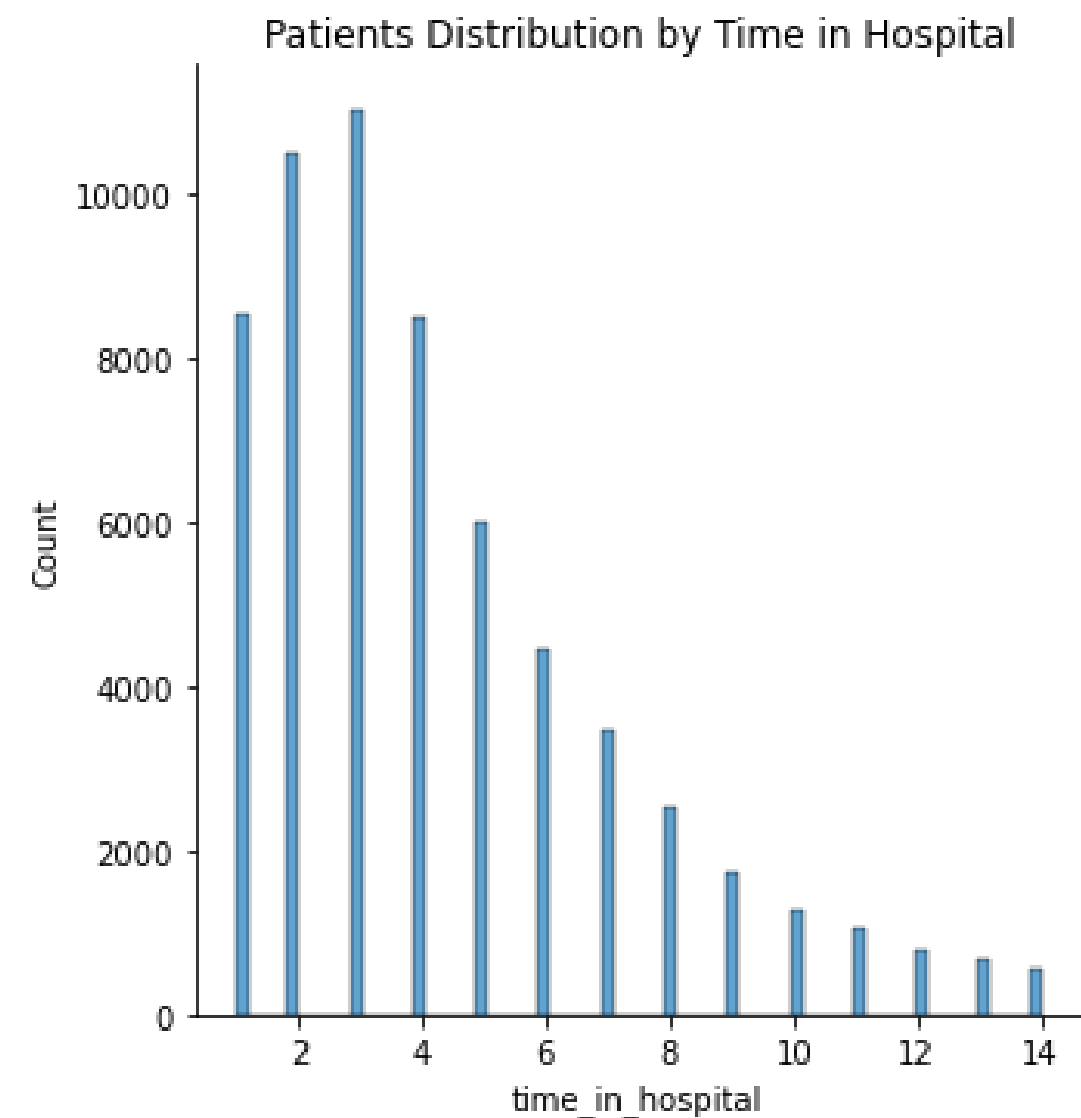
- It seems that people of Caucasian origin are prone to diabetes and require medication
- Females are slightly more prone to diabetes medication. However, this cannot be confirmed as this could be due to sample bias as well

# DIABETES MEDICATION PREDICTION

## Exploratory Data Analysis



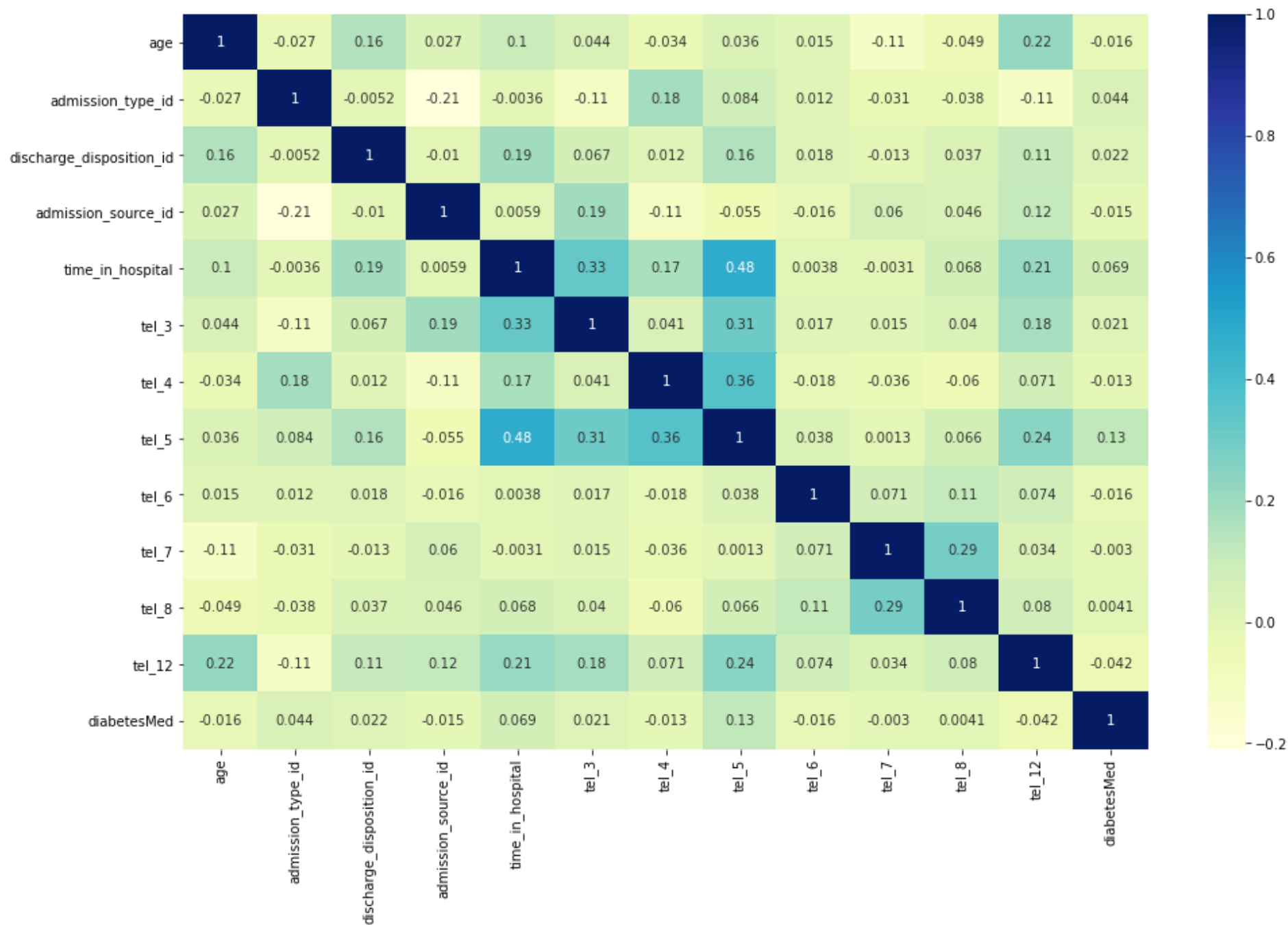
As expected, as people age they are more prone to diabetes and will require medication



Diabetes patients mostly require short term hospitalization

# DIABETES MEDICATION PREDICTION

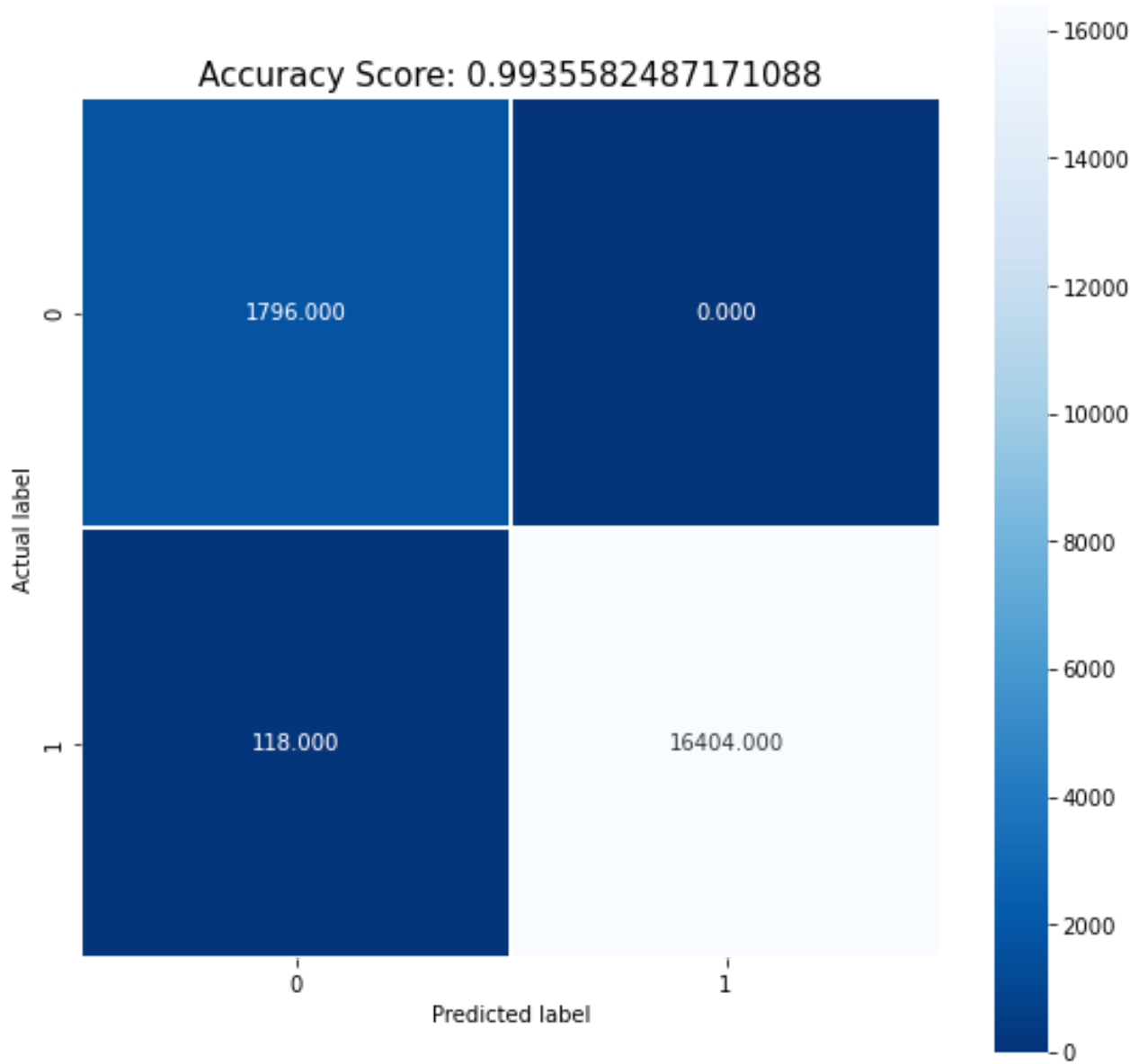
## Exploratory Data Analysis



No strong correlation observed in the dataset

# DIABETES MEDICATION PREDICTION

## Logistic Regression Model



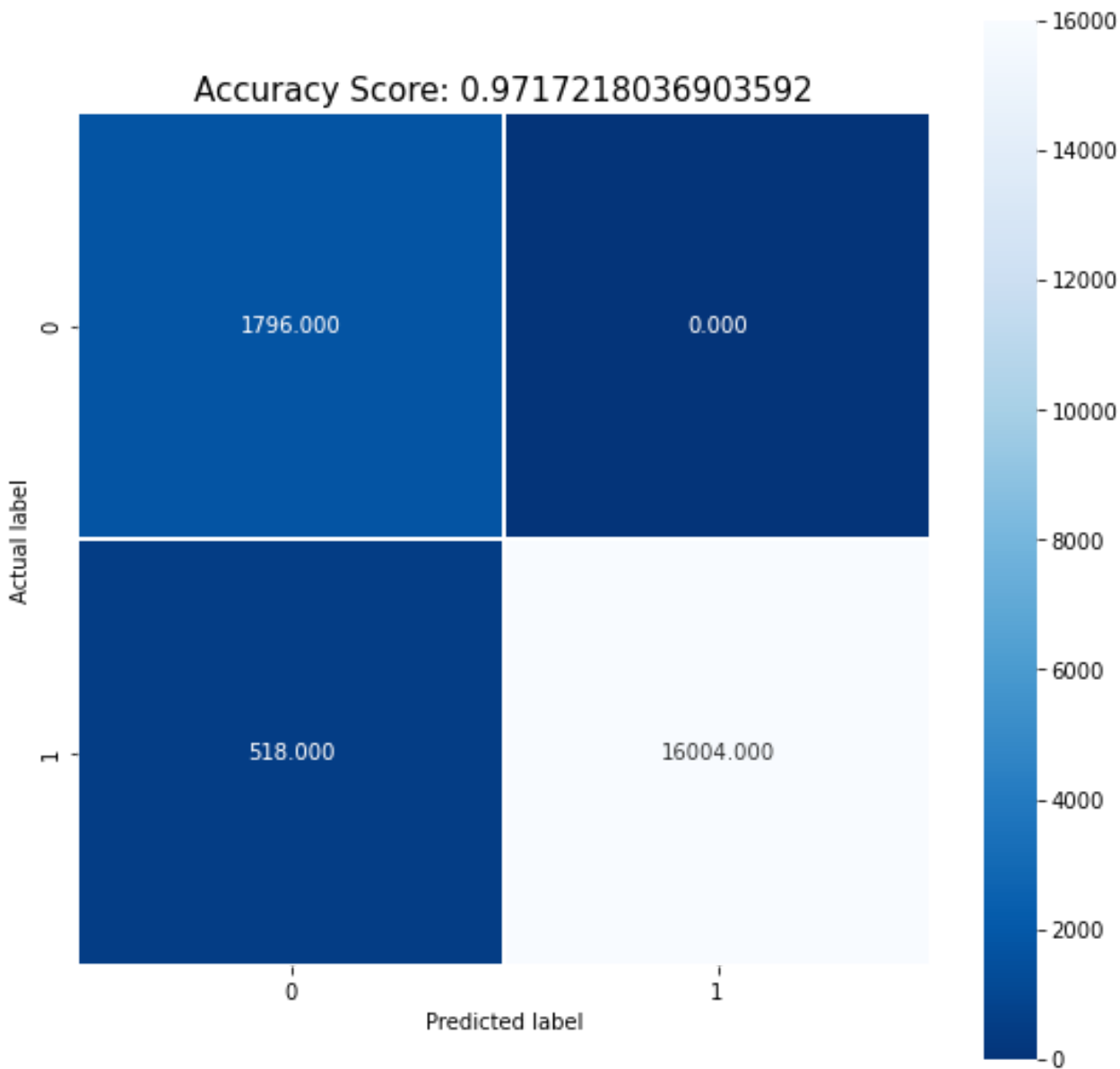
## Confusion Matrix

- We observe that the model is performing well with high Precision and Recall values

Classification report					
	precision	recall	f1-score	support	
0	0.94	1.00	0.97	1796	
1	1.00	0.99	1.00	16522	
accuracy			0.99	18318	
macro avg	0.97	1.00	0.98	18318	
weighted avg	0.99	0.99	0.99	18318	

# DIABETES MEDICATION PREDICTION

## Decision Tree Model



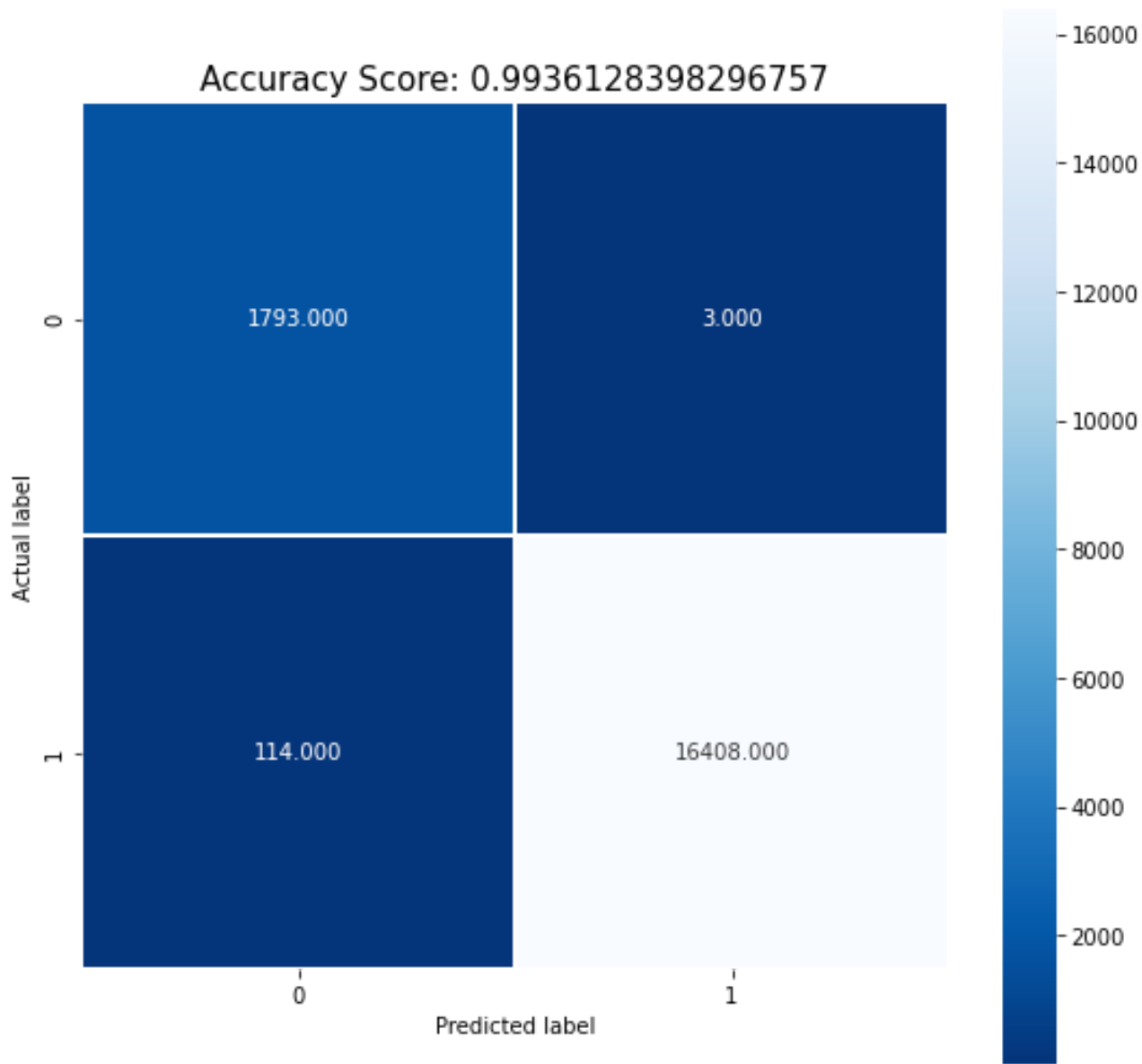
## Confusion Matrix

- We observe that the model is performing well with high Recall values

Classification report				
	precision	recall	f1-score	support
0	0.78	1.00	0.87	1796
1	1.00	0.97	0.98	16522
accuracy			0.97	18318
macro avg	0.89	0.98	0.93	18318
weighted avg	0.98	0.97	0.97	18318

# DIABETES MEDICATION PREDICTION

## Random Forest



## Confusion Matrix

- We observe that the model is performing well with high Recall values

Classification report					
	precision	recall	f1-score	support	
0	0.94	1.00	0.97	1796	
1	1.00	0.99	1.00	16522	
accuracy			0.99	18318	
macro avg	0.97	1.00	0.98	18318	
weighted avg	0.99	0.99	0.99	18318	



# IMPROVEMENT AREAS

**There is always scope for improvement!**

## **Formatting / Presentation**

- Better formatting of charts and Jupyter notebook code
- Some more detail in comments

## **Code**

- Use pipelines so that we don't have to repeat the code process for test case data
- Don't repeat code for similar functionality
- Detailed Exploratory Data Analysis

## **Model Tuning and Validation**

- Use GridSearchCV and other approaches to tune hyperparameters of the model
- Select the best performing model
- Analysis of errors

**THANK YOU!**