

PREDICTING DIABETES TO INFORM CLINICAL DECISIONS ABOUT PATIENT CARE

LESLIE CARDONE
MAY 14, 2021

GOALS

Predict diabetes (all types) using data collected during patients first 24 hours in the ICU.

Minimize false negatives / maximize true positives

Target is binary

- 1 Has been diagnosed with diabetes
- 0 Has not been diagnosed



DATA
kaggleTM

130,157 SAMPLES
180 FEATURES

CATEGORICAL

CONTINUOUS

> 3 %
CORRELATION

> 3 %
CORRELATION

UNDER 60 % NAN

36 FEATURES

METHODS

LOGISTIC REGRESSION BASELINE #1

ONE FEATURE (HIGHEST GLUCOSE CONCENTRATION)

FILL ALL NAN WITH MEDIAN OF FEATURE

LOGISTIC REGRESSION BASELINE #2

15 FEATURES – ALL CONTINUOUS VARIABLES

SCALE AND REGULARIZE

FILL ALL NAN WITH MEDIAN OF FEATURE



METHODS

SCORING METRIC	MODEL #1	MODEL #2
ACCURACY		
PRECISION		
RECALL	0.21	0.26
F1	0.31	0.36

**LOW RECALL
HIGH FALSE NEGATIVES**

METHODS

LOGISTIC REGRESSION FINAL MODEL

ALL 36 FEATURES

FILL ALL NAN WITH KNN PREDICTION

SCALE AND REGULARIZE (L1 - LASSO)

TUNED WEIGHTS AND C VALUE (.001)

SCORE ON HOLD OUT SET

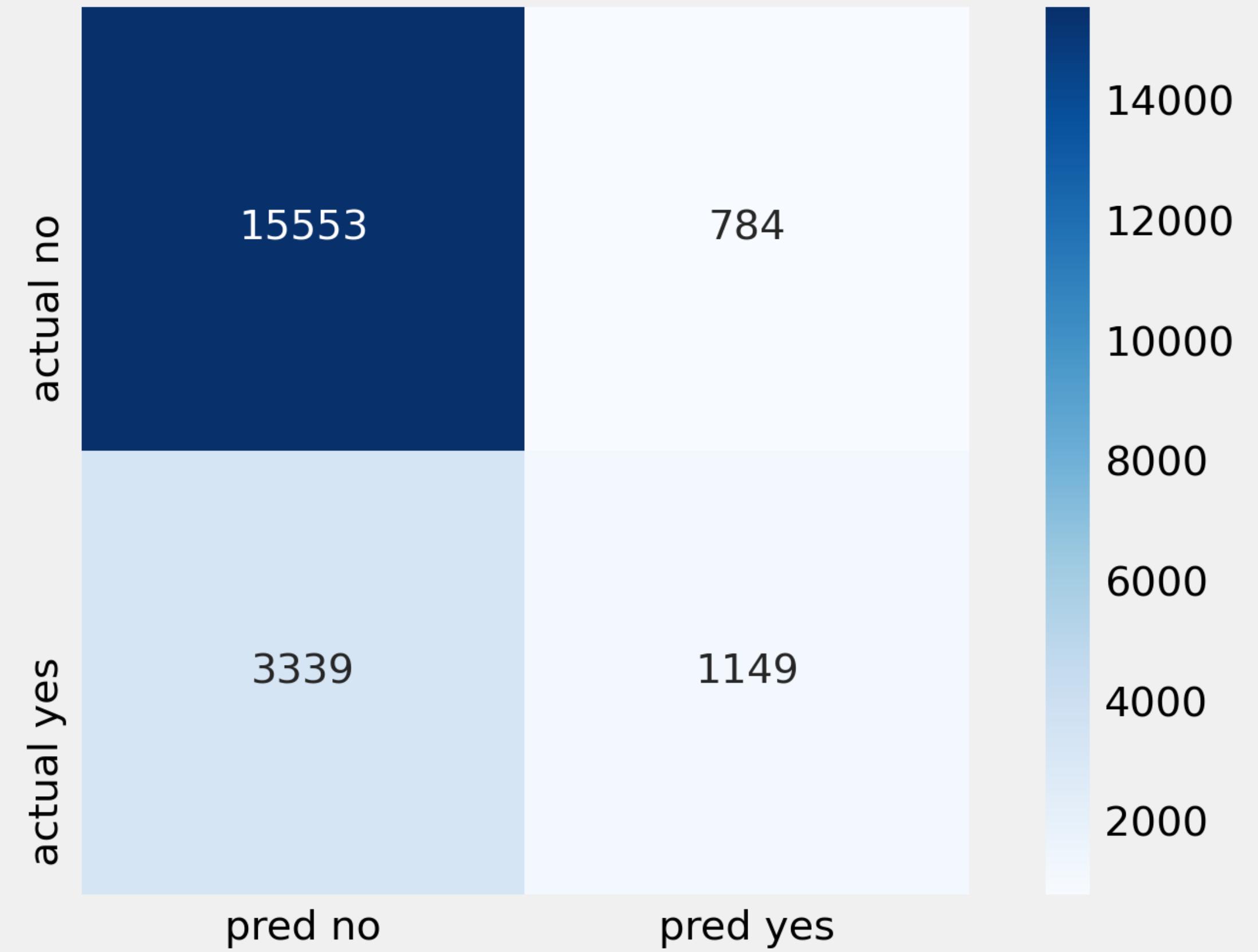


METHODS

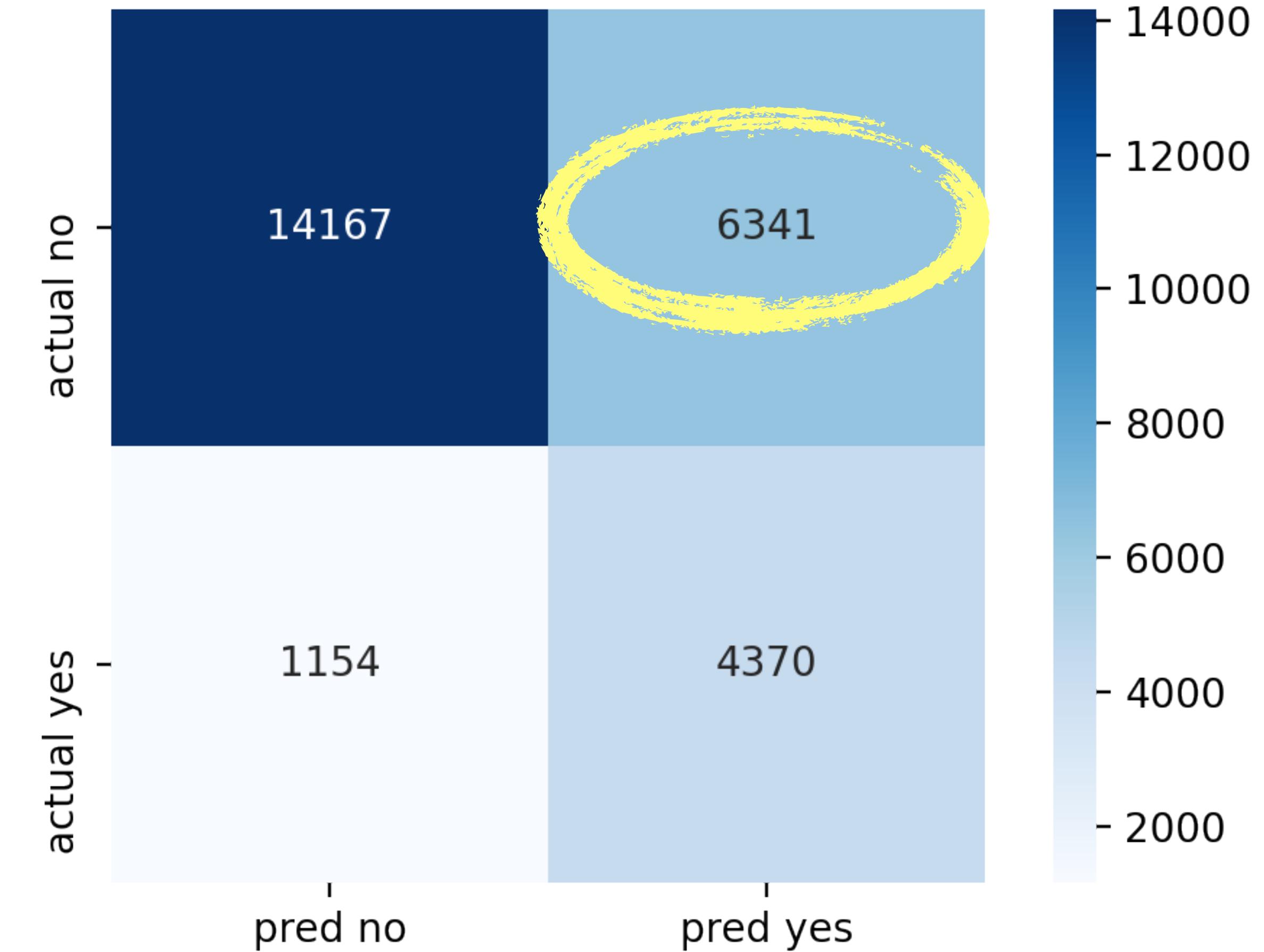
SCORING METRIC	MODEL #1	MODEL #2	FINAL
ACCURACY			
PRECISION			
RECALL			
F1			

FINDINGS

Confusion Matrix -- Limited Features

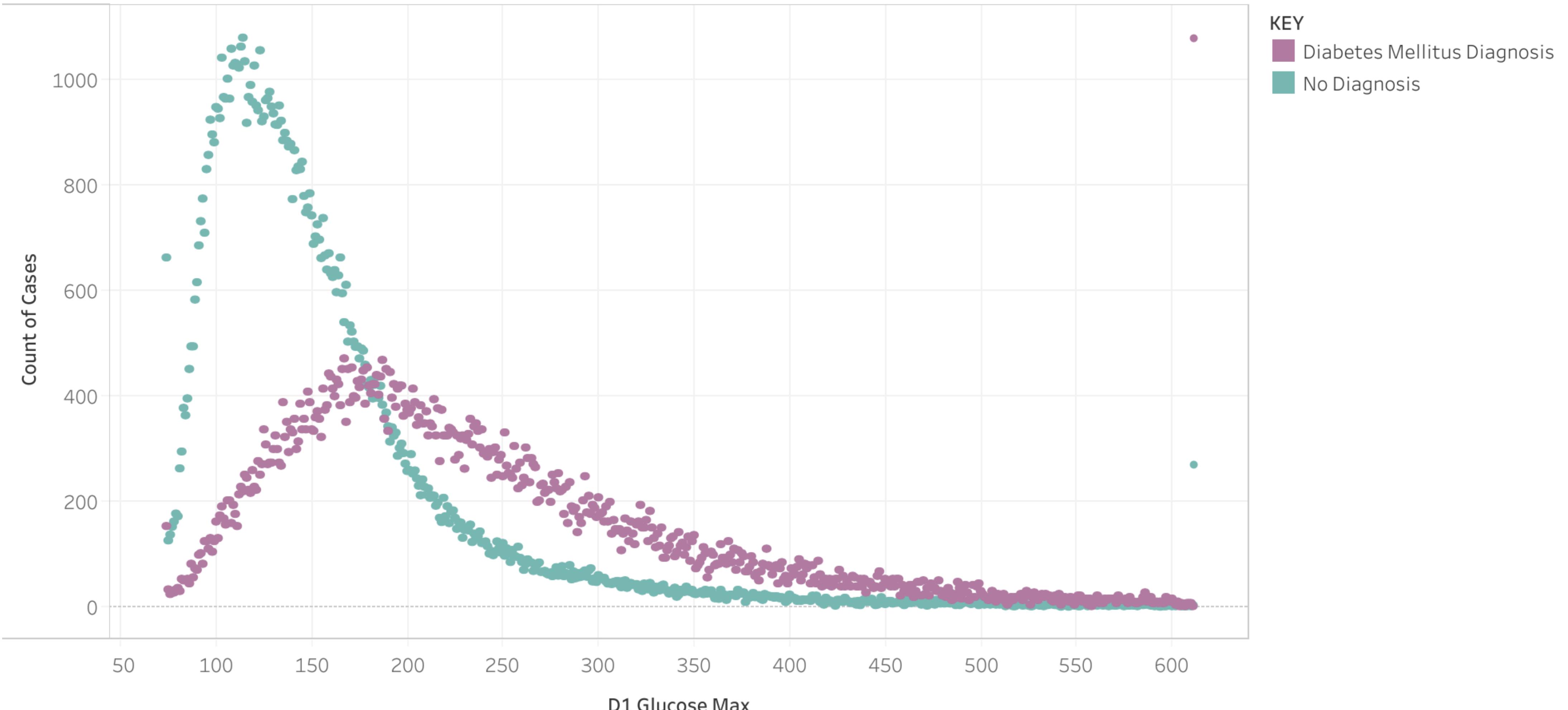


Confusion Matrix -- Final



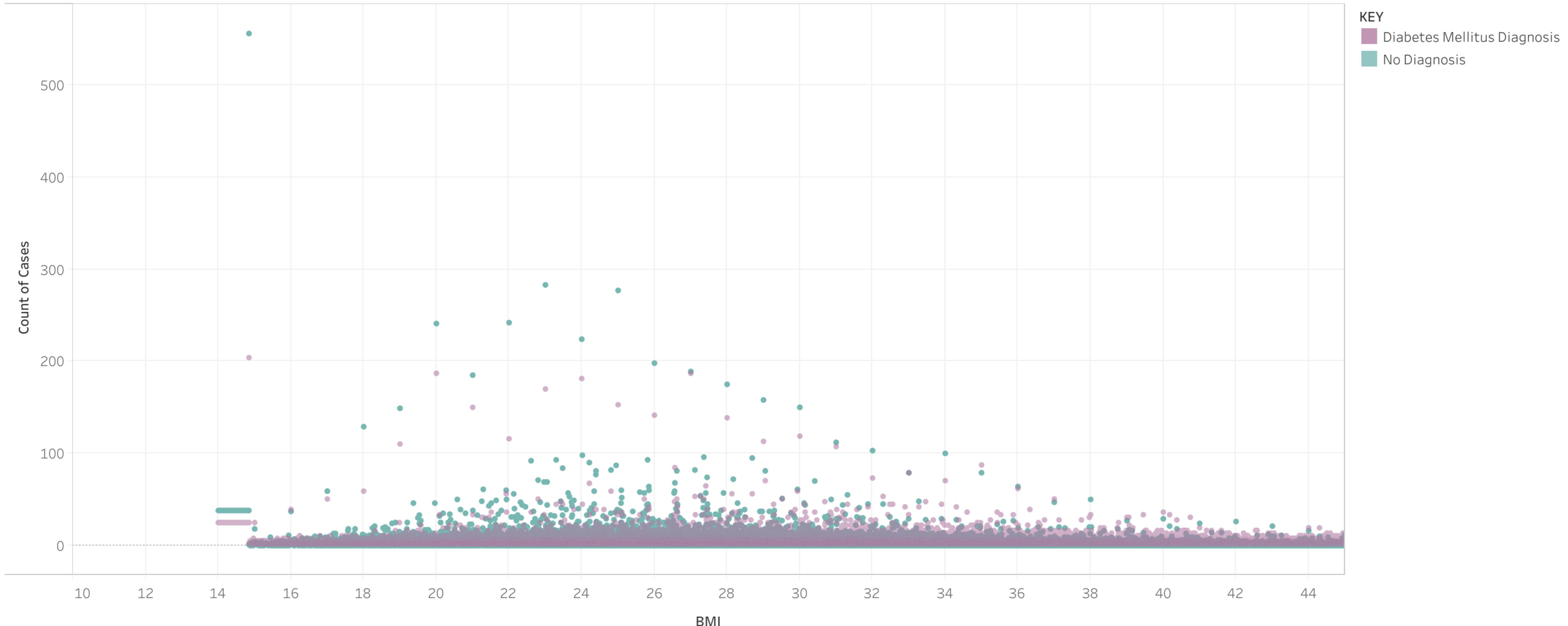
FINDINGS

Distribution of Diabetes Diagnosis



FINDINGS

Distribution of Diabetes Diagnosis



CONCLUSIONS

Class imbalance significantly affects the predictive capabilities to predict Diabetes (the minority class)

More features do not necessarily have an affect on the model

Complicated imputation did not improve any scoring metrics by much



FURTHER STUDIES

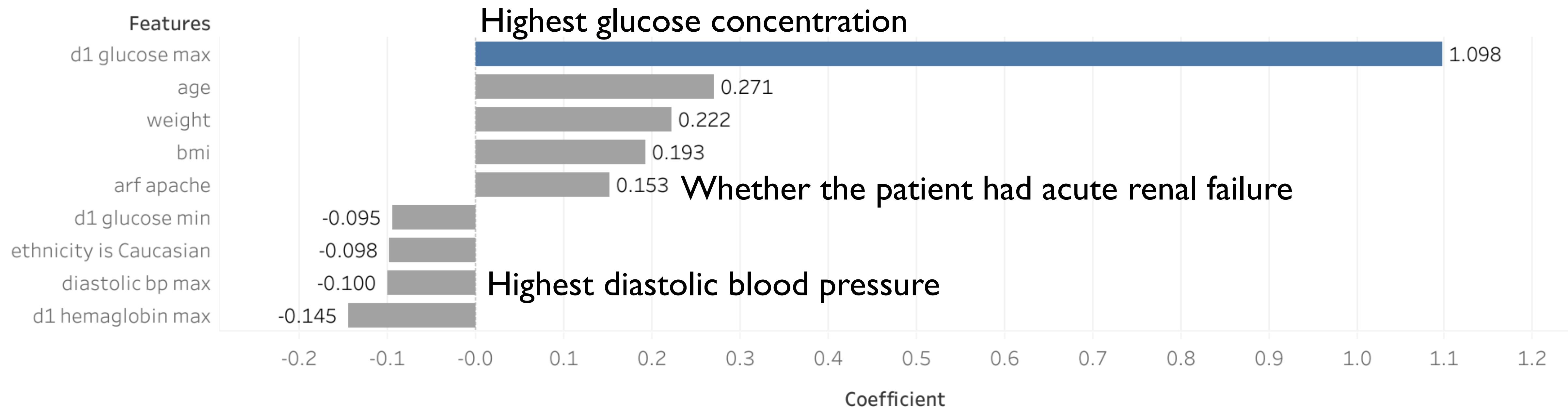
More feature engineering, create
feature interactions

Look at Gradient Boosting / XGBoost



APPENDIX

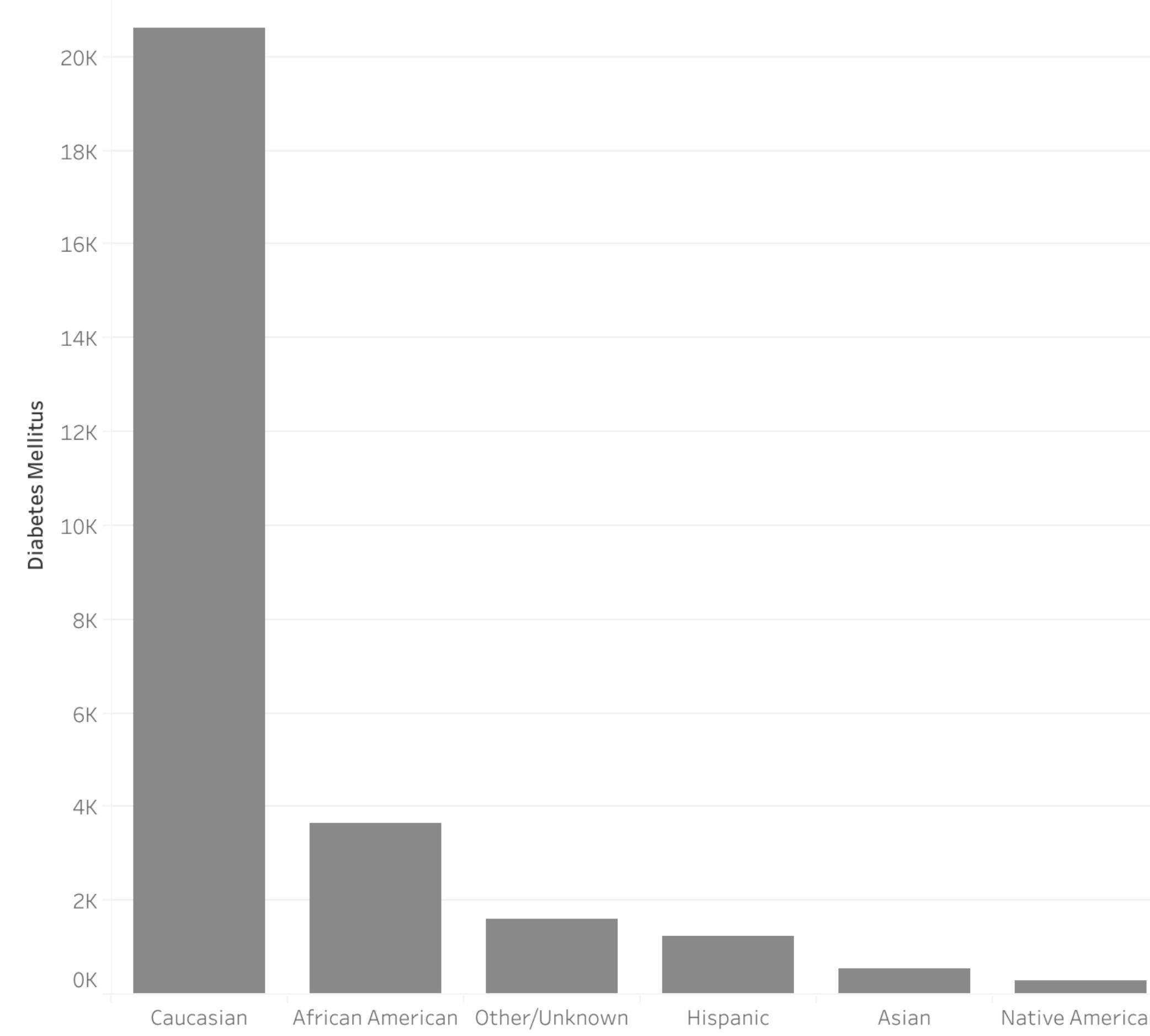
Feature "Importance"



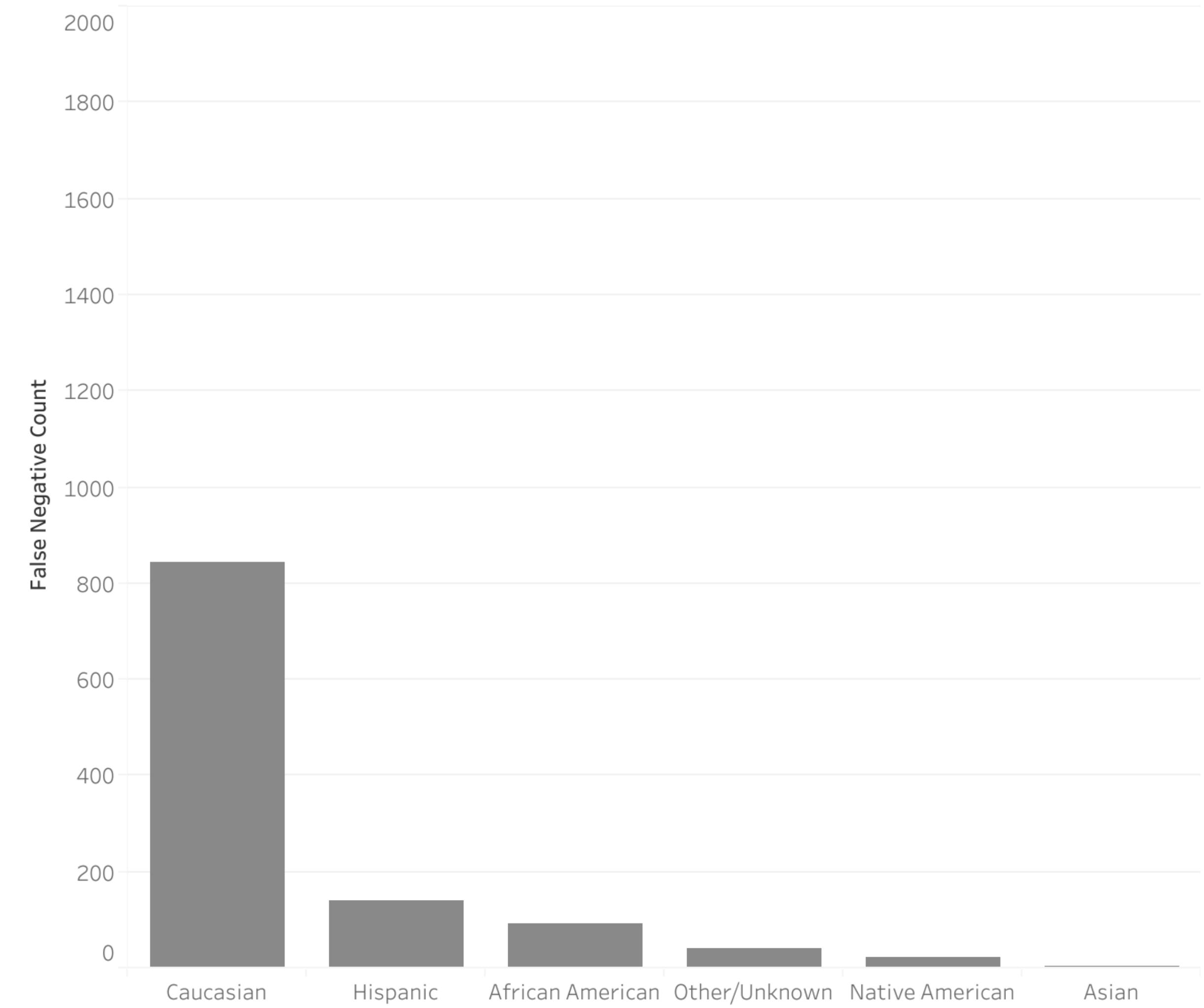
19 OF 36 FEATURES
AFTER REGULARIZATION

APPENDIX

Diabetes By Ethnicity

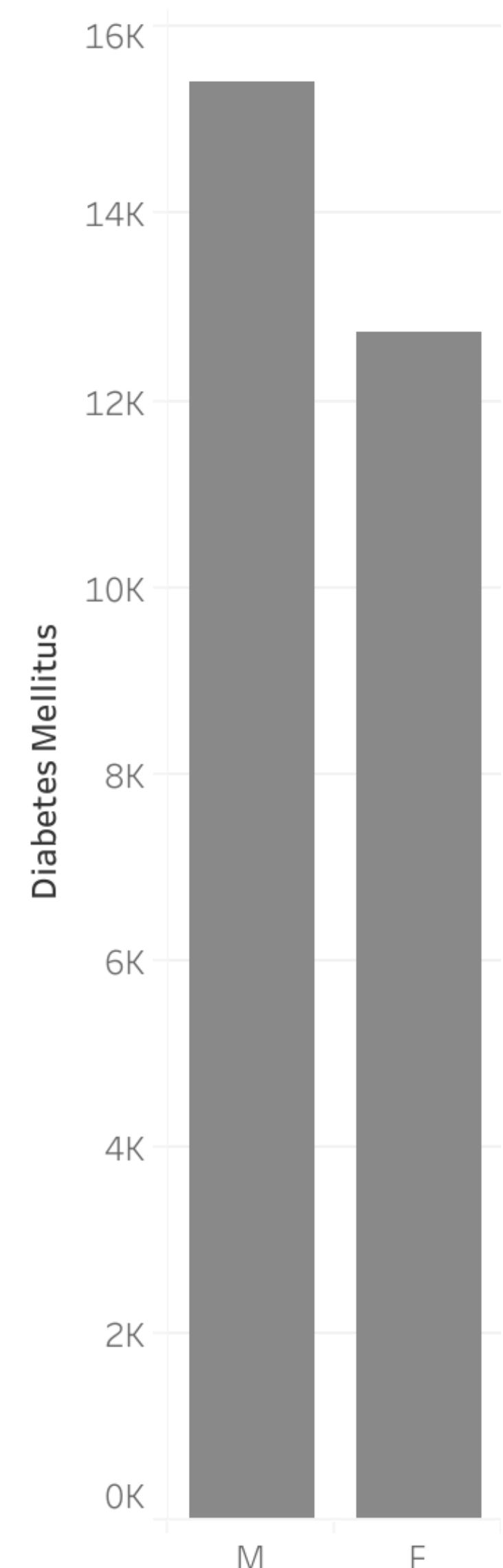


Diabetes Misclassification By Ethnicity

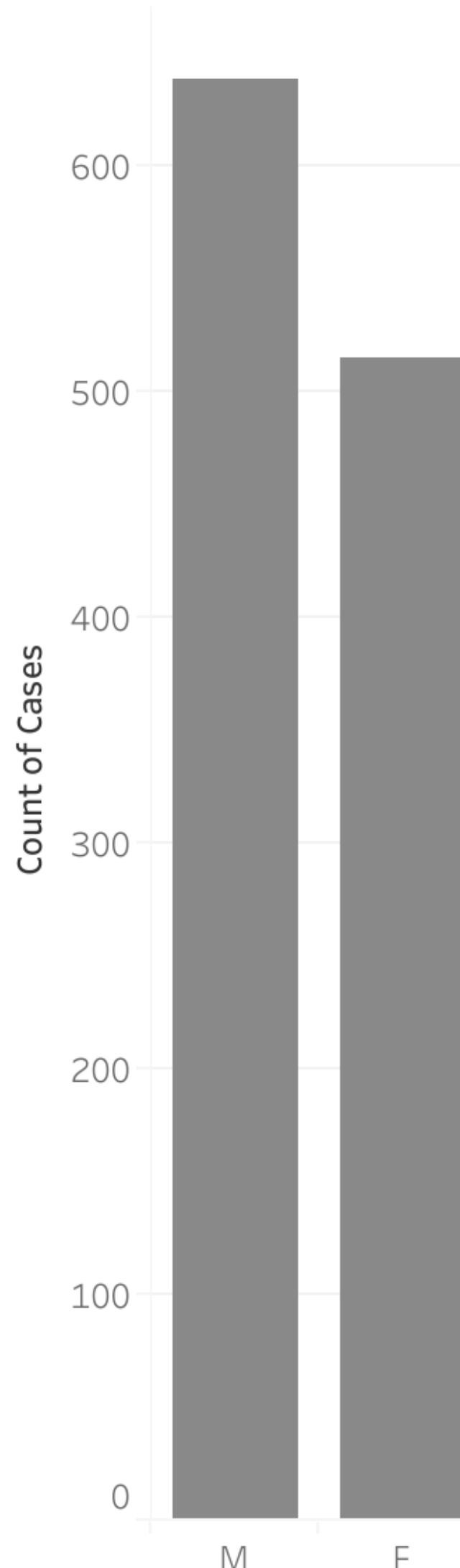


APPENDIX

Count of Diabetes Cases by Gender



Count of False Negatives by Gender



APPENDIX

Distribution of False Negatives by Age

