

A thick black L-shaped frame surrounds the text. It starts at the top left, goes right, then down, then right again at the bottom right.

# ALCOHOL CONSUMPTION: PREDICTING USAGE AND RISK FACTORS

Module 3 Project

Sara Robinson

# Introduction

- Purpose
- Implications
- My Background

# The Data Science Process: OSEMN

# Obtain

- UC Irvine Machine Learning Repository
- Drug Consumption

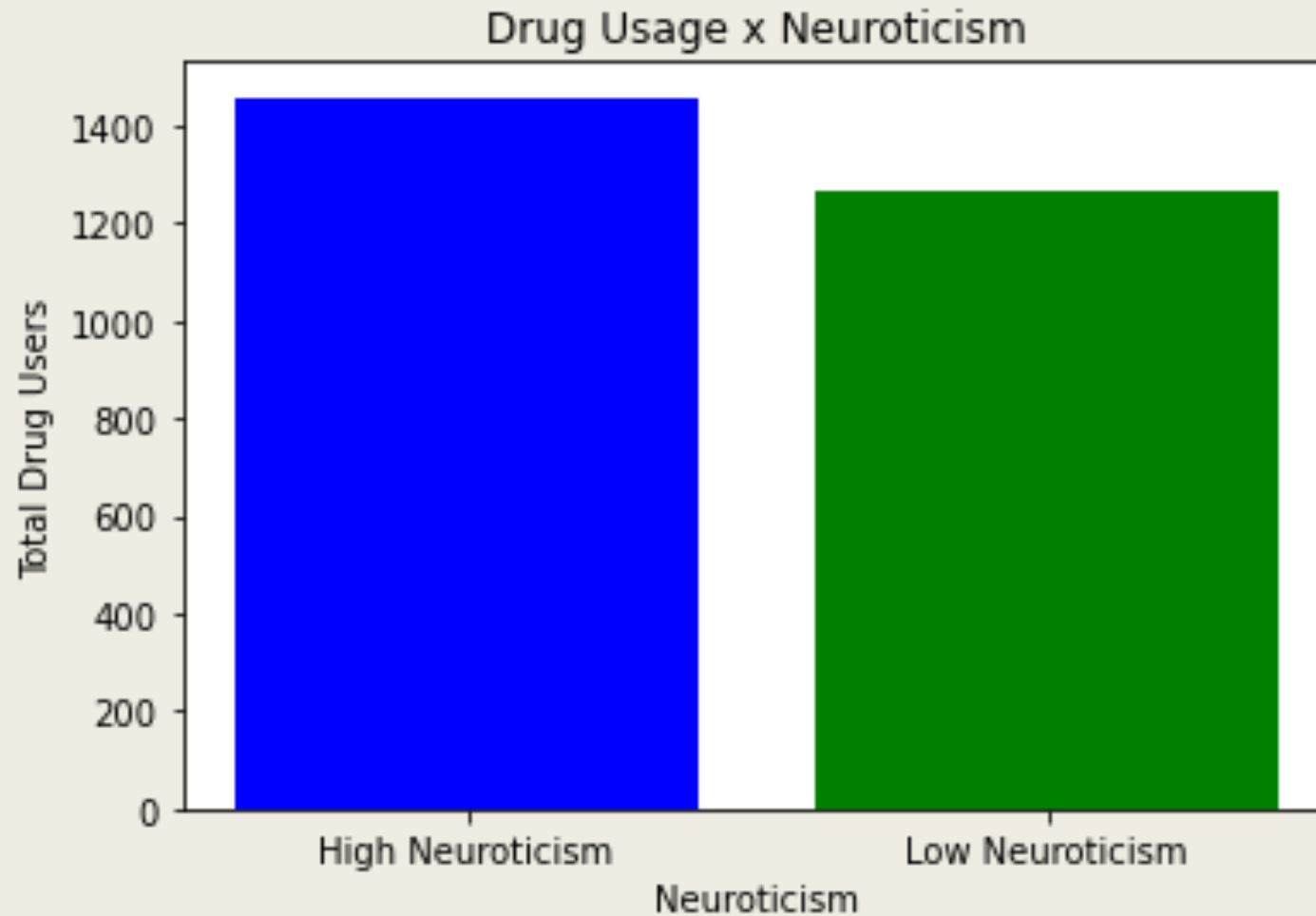
# Scrub

- Placeholder Values
- Types of Users
  - *Binary Classification*

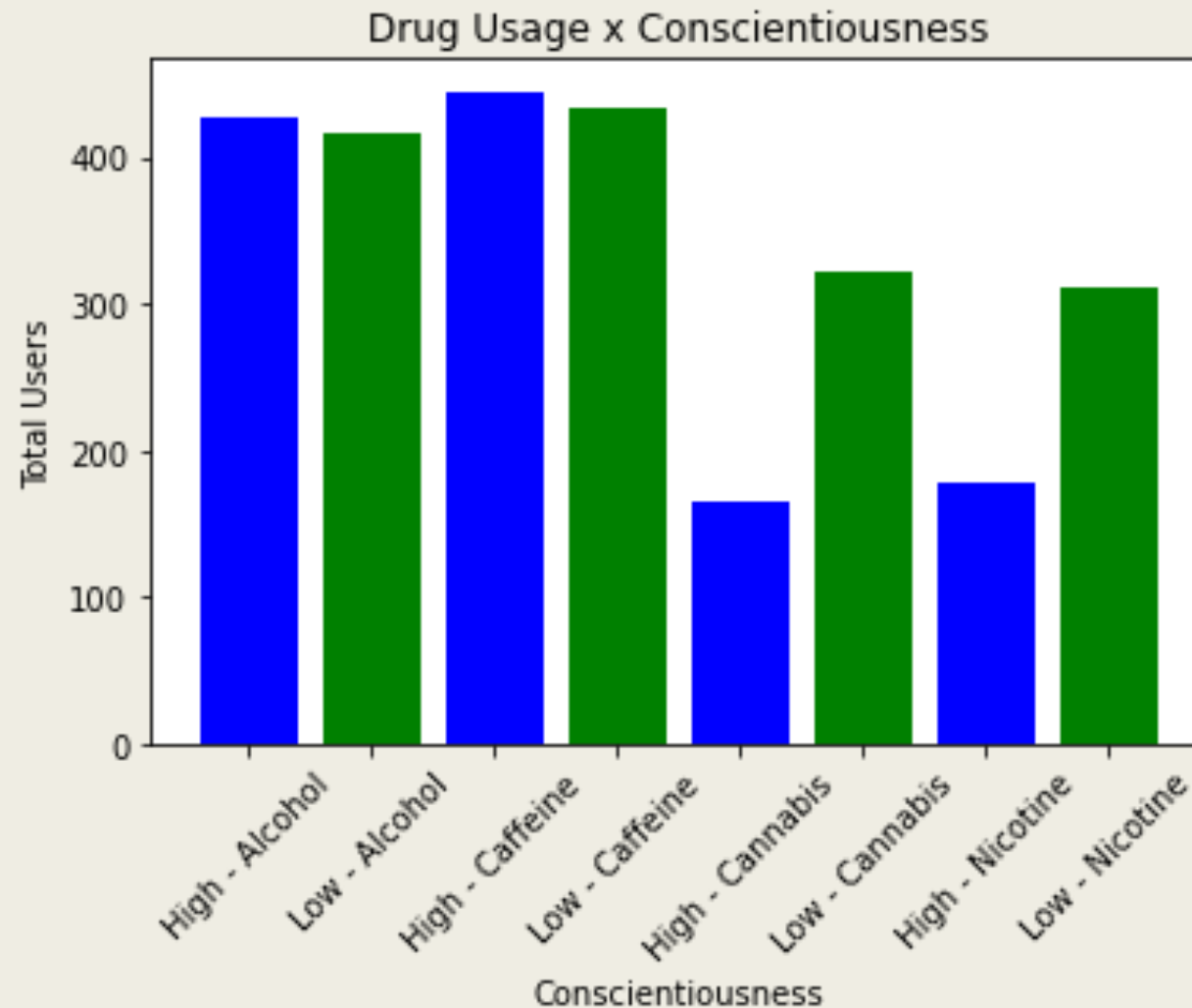
# Explore

- Personality Score Measures
- Level of Education
- Exploratory Data Analysis (EDA)

Are those who score higher on Neuroticism more likely to be drug users than those who score lower?

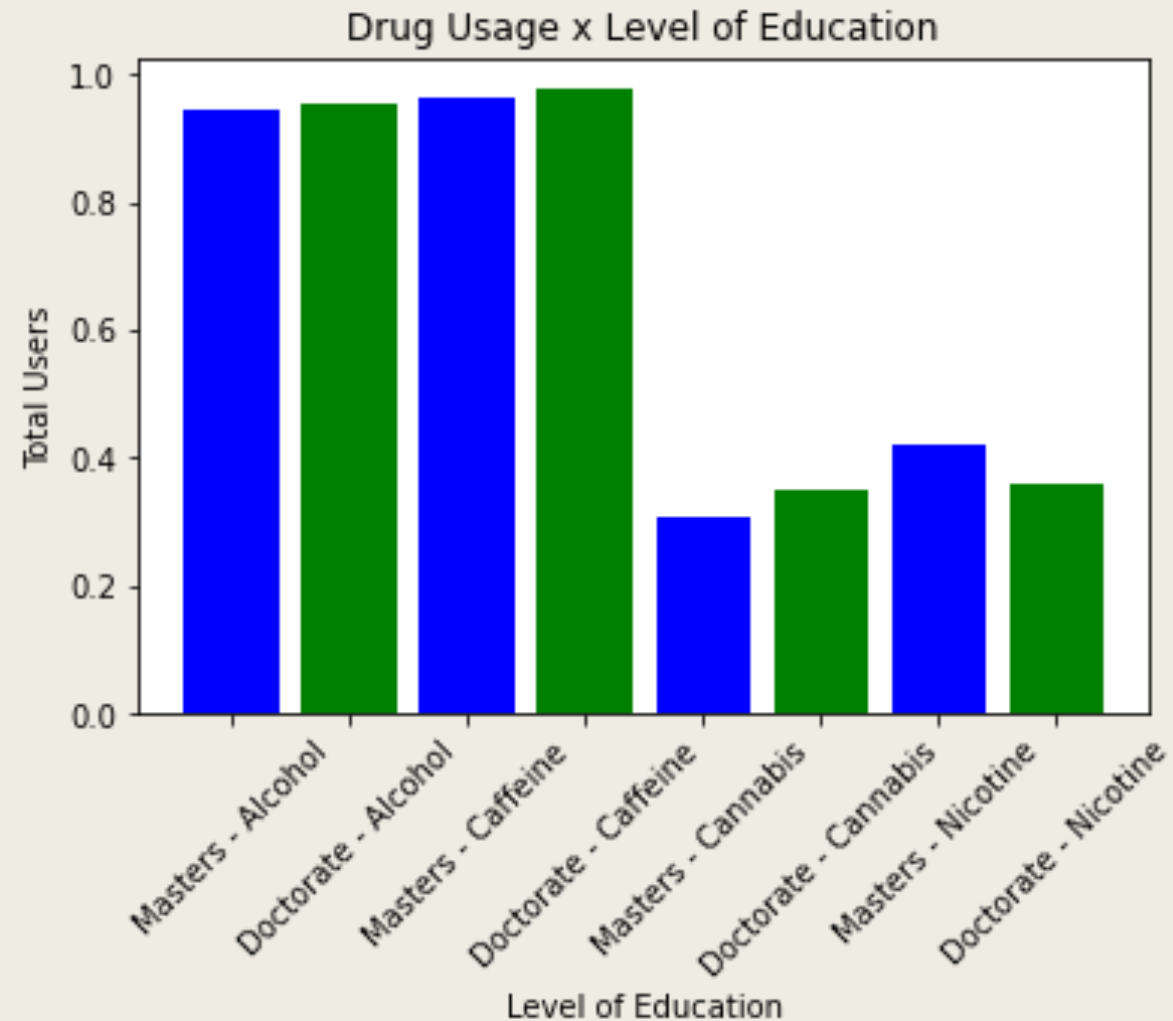


Are those who score lower on Conscientiousness more likely to use drugs than those who score higher?





Which makes you more likely to be a drug user, having a Doctorate or a Masters degree?



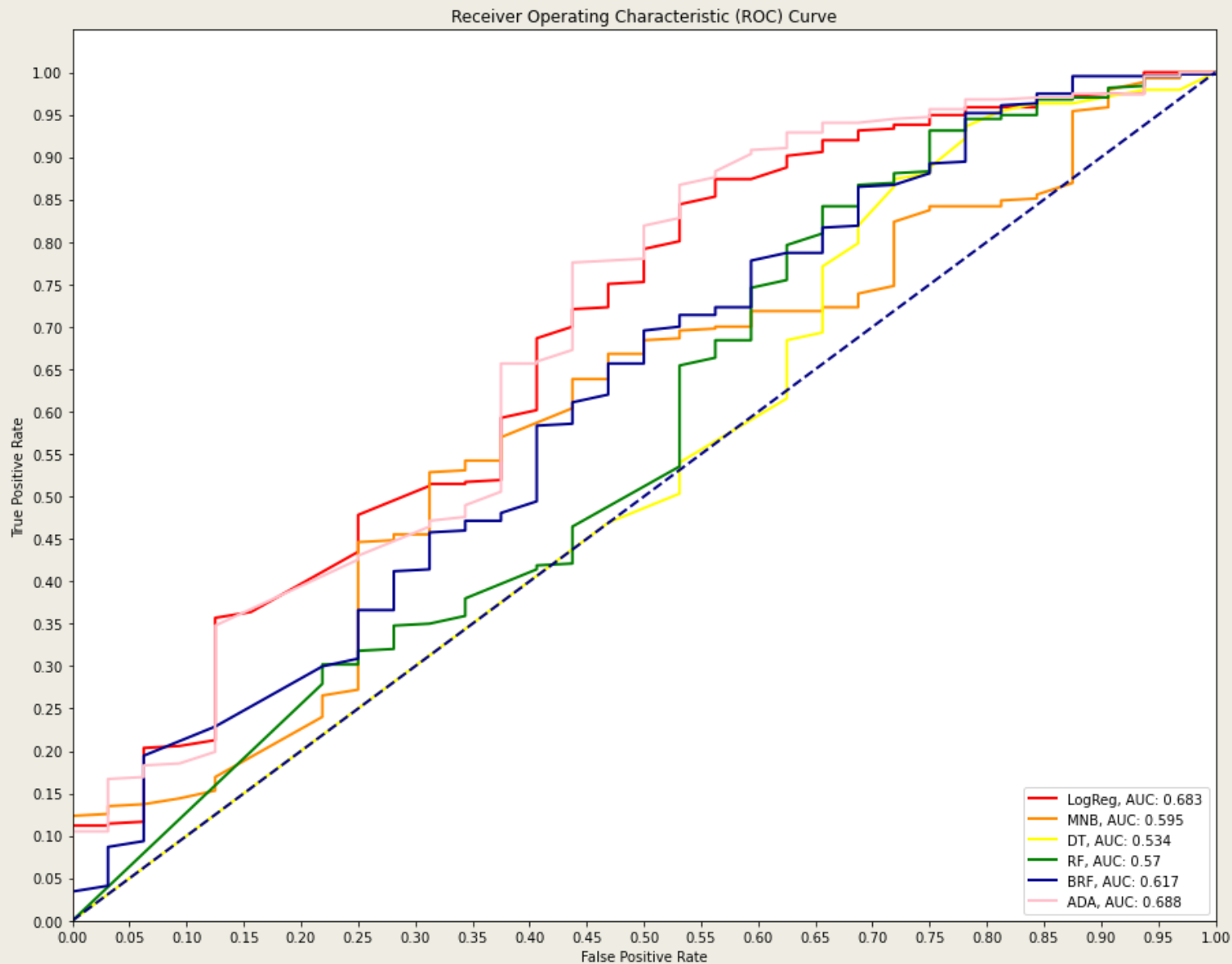
# Explore

- Personality Score Measures
- Level of Education
- Exploratory Data Analysis (EDA)

# Model

- 6 Models

# ROC Graph Comparing Original Models



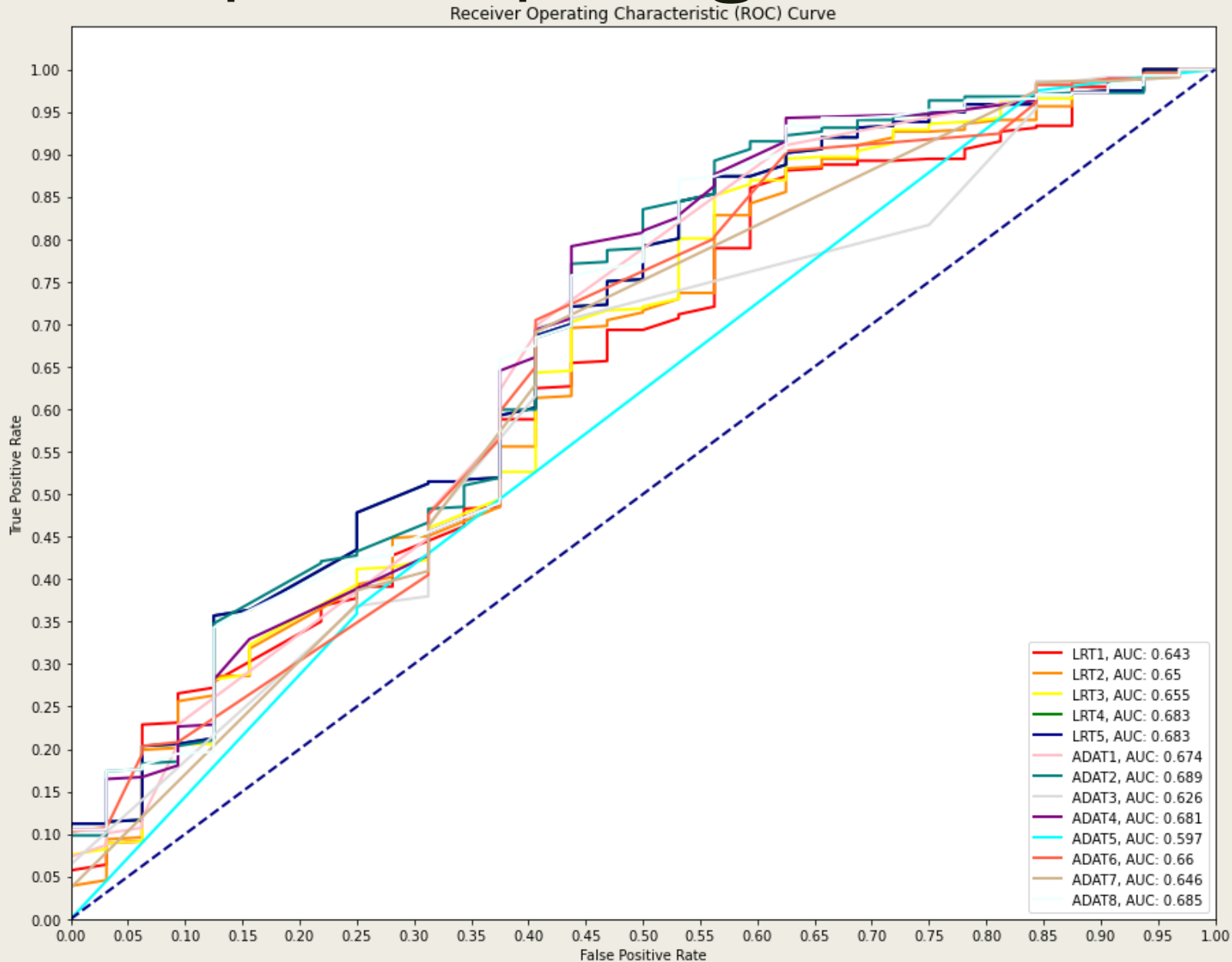
# Comparing Original Model Scores

Model	Training Accuracy	Testing Accuracy	Training F1 Score	Testing F1 Score
LogReg	0.925	0.934	0.961	0.966
MNB	0.925	0.932	0.961	0.965
DT	0.942	0.906	0.969	0.950
RF	0.942	0.921	0.969	0.959
BRF	0.609	0.586	0.735	0.726
ADA	0.925	0.930	0.961	0.963

# Model

- Tuned Top Performing Models

# ROC Graph Comparing Tuned Models



# Comparing Tuned Model Scores

Model	Training Accuracy	Testing Accuracy	Training F1 Score	Testing F1 Score
LRT1	0.927	0.932	0.962	0.965
LRT2	0.927	0.932	0.962	0.965
LRT3	0.926	0.932	0.962	0.965
LRT4	0.925	0.934	0.961	0.966
LRT5	0.925	0.934	0.961	0.966
ADAT1	0.926	0.934	0.962	0.966
ADAT2	0.924	0.930	0.960	0.963
ADAT3	0.927	0.932	0.962	0.965
ADAT4	0.923	0.930	0.960	0.963
ADAT5	0.927	0.932	0.962	0.965
ADAT6	0.926	0.932	0.962	0.965
ADAT7	0.927	0.932	0.962	0.965
ADAT8	0.924	0.930	0.960	0.963



# iNterpret

- Final Model
- Important Features
  - *Age*
  - *Drug comorbidity*

# Further Exploration

- Identifying features of users within drug classes
- Multi-class users instead of binary
- Personality score measurements

THANK YOU

