Credit Card Approval Prediction Model

Data Science, Mandatory 2, Group 3

MAY 23, 2019

TALKING POINTS

Research Question

Dataset

Analysis

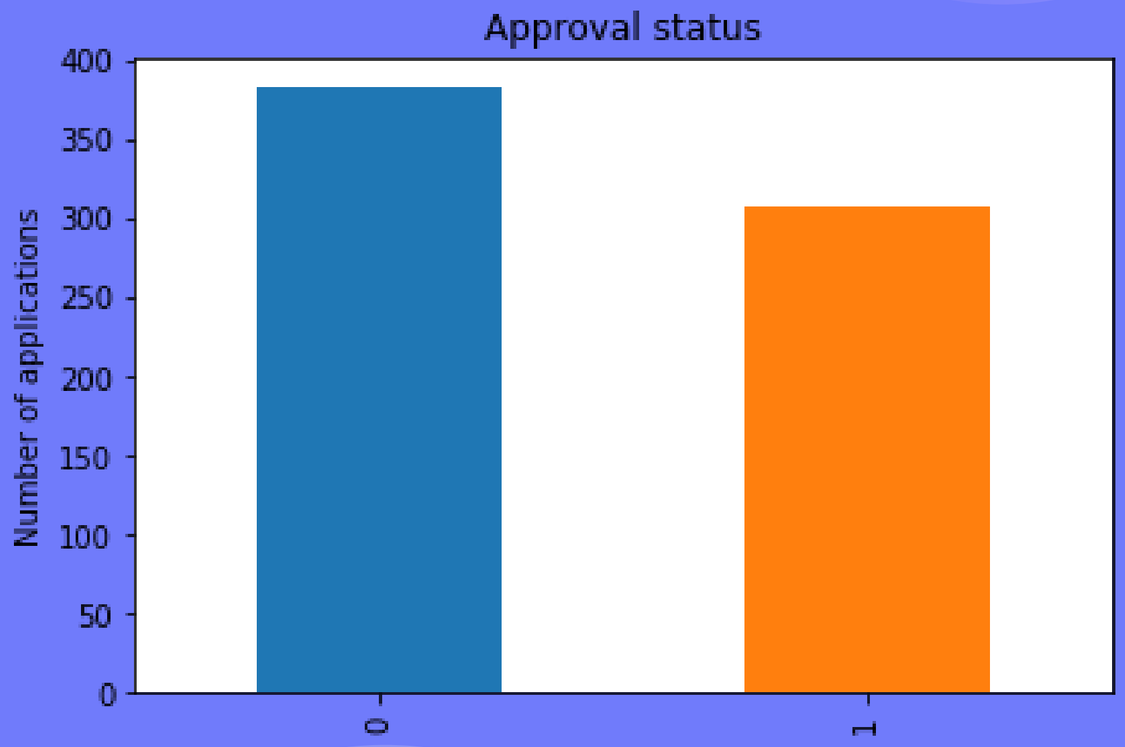
Model selection & tuning

Findings

Based on the UCI credit card approval dataset, what model has the highest predictive power to forecast credit card applications approval? Which attribute(s) have the highest influence?

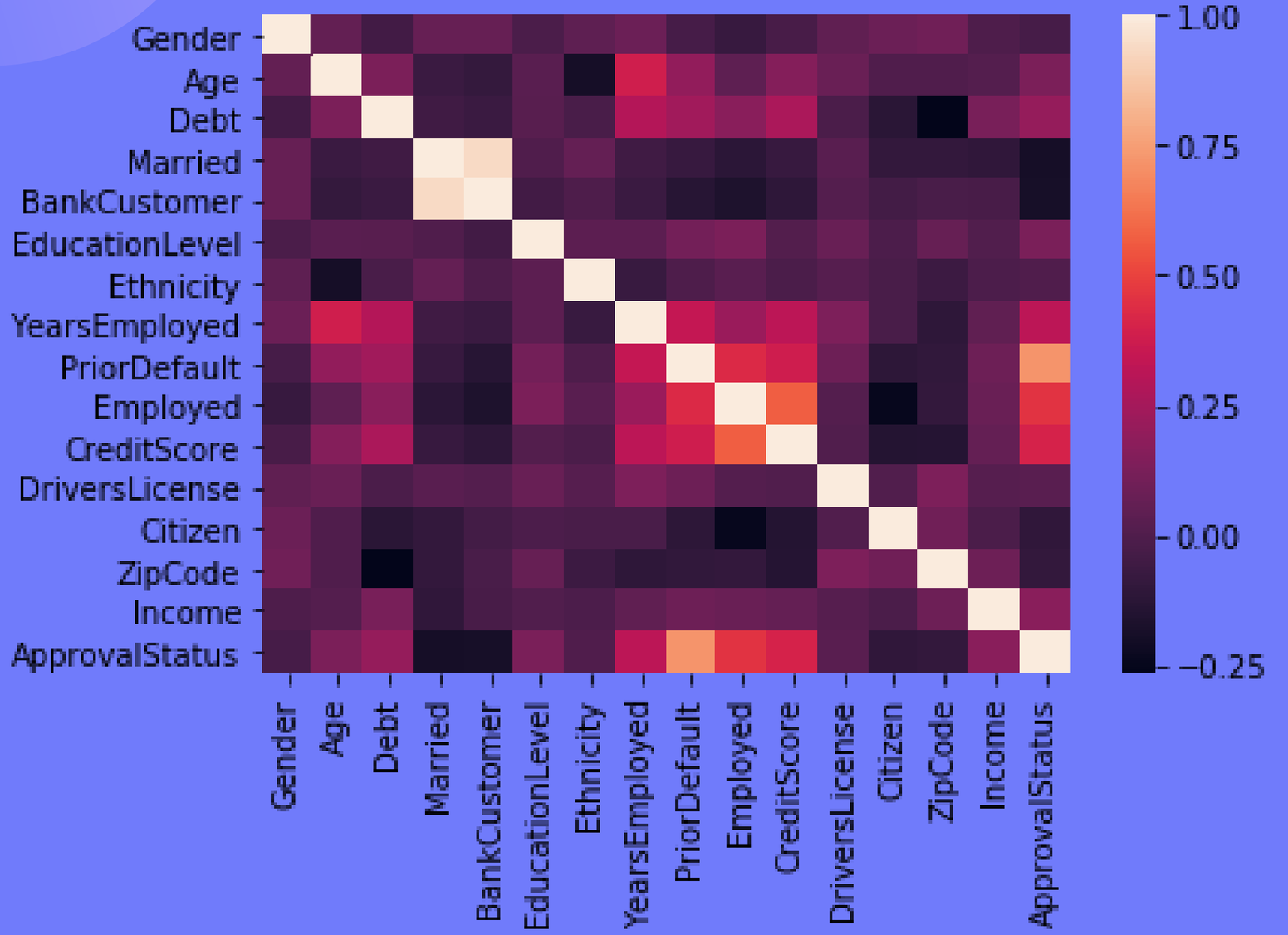
WHY WE CHOSE THIS DATASET?

- CREDIT CARD APPLICATIONS AND APPROVAL DECISIONS
- ANONYMIZED DATA
- 15 ATTRIBUTES
- 95% HOLDING 690 OBSERVATIONS



DEPENDANT VARIABLE

CORRELATION



DATA PREPROCESSING

MISSING VALUES

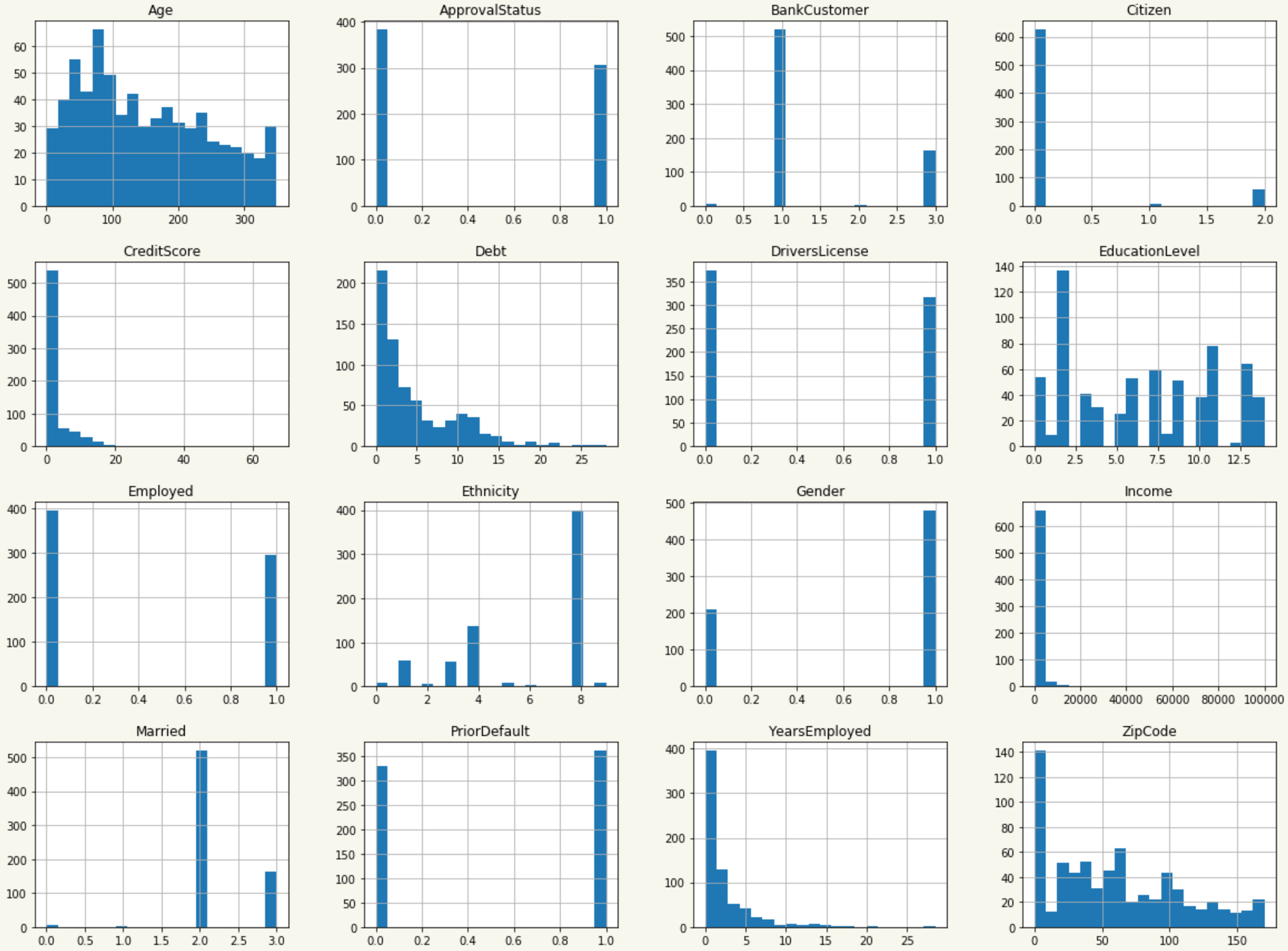
Mean imputation and most common feature value.

DATA TRANSFORMATION

Converted categorical values with label encoding.

DATA SPLITTING

Validation	25%
Testing	25%
Training	50%


Age
ApprovalStatus
BankCustomer
Citizen
CreditScore
Debt
DriversLicense
EducationLevel
Employed
Ethnicity
Gender
Income
Married
PriorDefault
YearsEmployed
ZipCode

Model selection

```
Random Forrest
AUC:  0.910370624494463 4
F1:  0.860959467398999 8
Precission:  0.884670905614301 8
Recall:  0.841490196078431 3
Accuracy:  0.846061545325396 5
*********
Naive Bayes
AUC of Naive Bayes is:  0.876713580453518 4
F1:  0.827552782173301 1
Precission:  0.752103311539960 9
Recall:  0.920784313725490 1
Accuracy:  0.780406642950230 7
*********
Logistic regression
AUC of Logistic Regression is:  0.918517502022145 9
F1:  0.873770502519346 2
Precission:  0.910217143506180 2
Recall:  0.845254901960784 3
Accuracy:  0.780406642950230 7
```

LOGISTIC REGRESSION

AUC	F1	PRECISION	RECALL	ACCURACY
0.943	0.903	0.934	0.879	0.630

RANDOM FOREST

AUC	F1	PRECISION	RECALL	ACCURACY
0.901	0.902	0.942	0.866	0.898

HIGHEST CORRELATION

The most correlated values that influence approval - **PriorDefault**, **Employed**, **CreditScore** and **YearsEmployed**

BEST PERFORMING MODEL

Highest predictive power to forecast credit card applications approval is **Random Forest**

THANK YOU!