DUKE FUQUA SCHOOL OF BUSINESS

Applied Probability and Statistics

# Multidimensional Assessment of Home Mortgage Approval Rate in Washington State

MQM A9

# Agenda

- Introduction and Data Wrangling Process
- Data Exploration and Visualization
- Correlation Analysis
- Regression Analysis - Customer-driven Model
- Regression Analysis – Intuitive and Stepwise Selection Models
- Prediction & Empirical Results
- Summary

DUKE
FUQUA
SCHOOL OF BUSINESS

# Introduction and Data Wrangling Process

Some of the key observations from the exploration were as follows:

Aim to predict the likeliness of a home loan application getting approved

Base on the data from Kaggle for the state Washington in **2016**

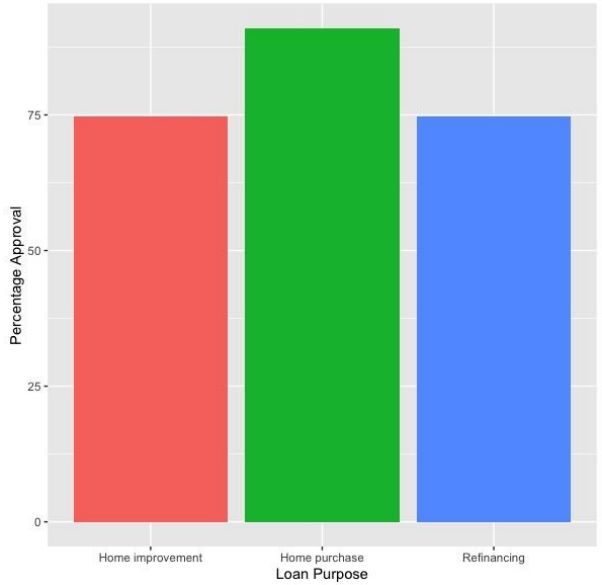
Include **19** variables divided into **4** major categories

DUKE
FUQUA
SCHOOL OF BUSINESS

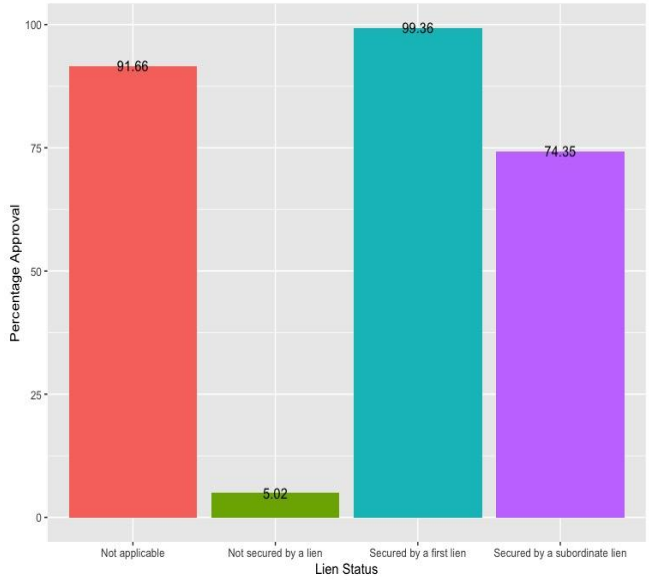
# Data Exploration and Visualization

Some of the key observations from the exploration were as follows:

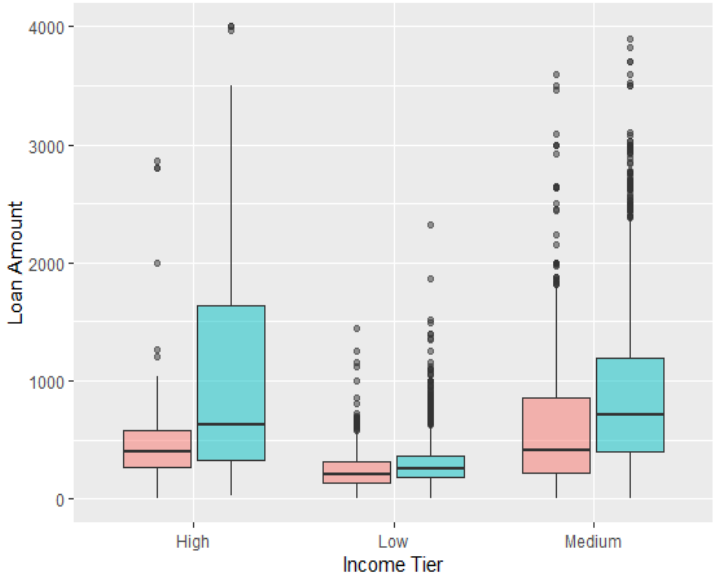
**Approval vs. Loan purpose**




**Approval vs. Lien status**




**Approval vs. Income tier**

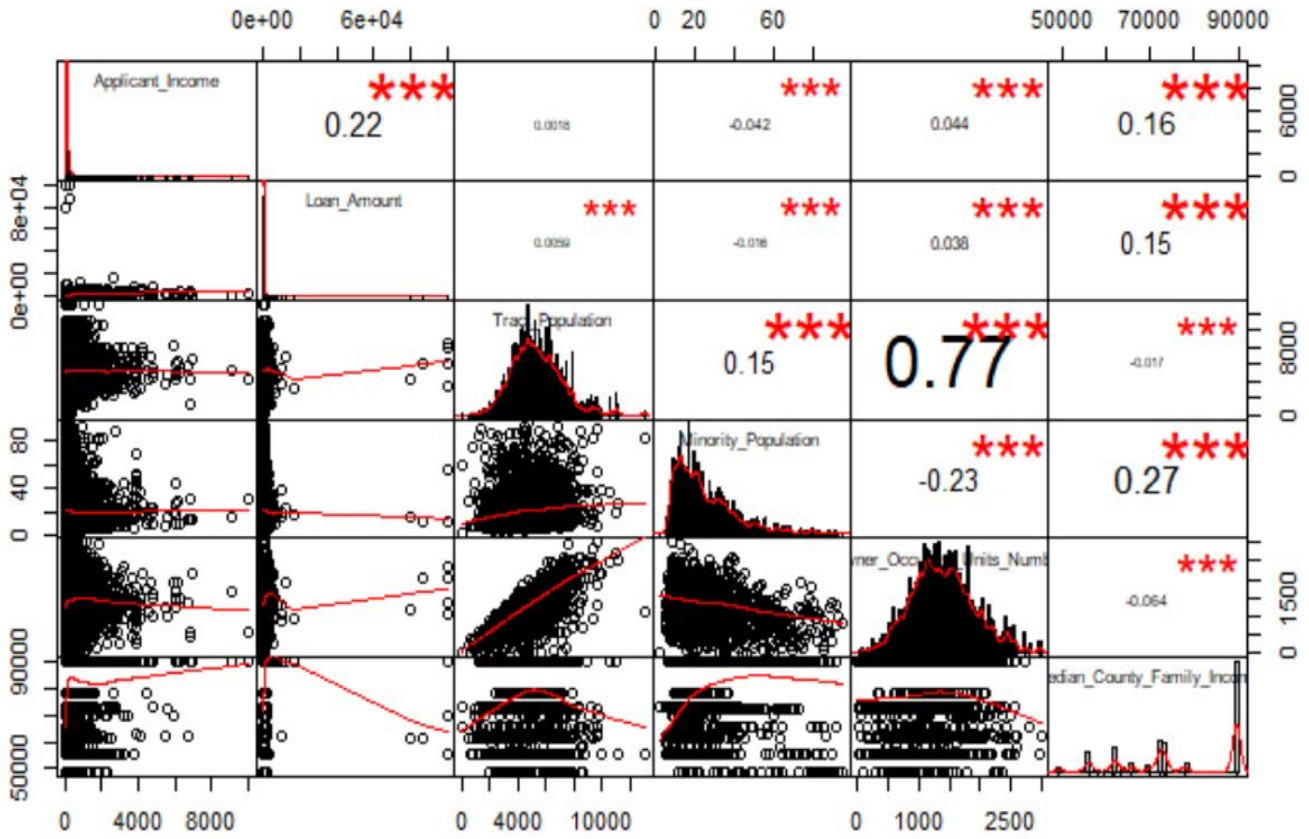



- The effect of purpose on likelihood of approval
- The vast difference in approval rates, when a loan is secured vs unsecured
- Applicant income is also a key factor based on the loan amount for the chances of approval

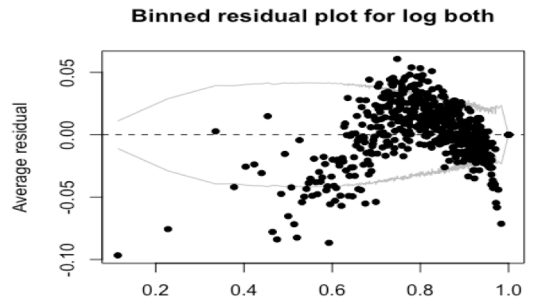
DUKE FUQUA SCHOOL OF BUSINESS

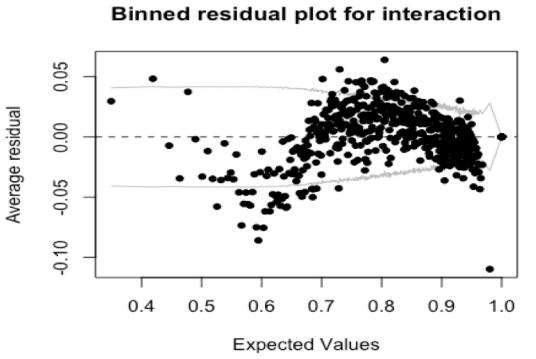
# Correlation Analysis

## Correlation Between Numerical Variables







# Regression Analysis – Customer-driven Model

Customer-driven model includes **6** variables in **4** categories, with an AIC of **305522** and McFadden's $R^2$ of **0.04**

$$Approval = Applicant\_Gender + Applicant\_Race + Applicant\_Income + County\_Name + Loan\_Amount + Property\_Type$$

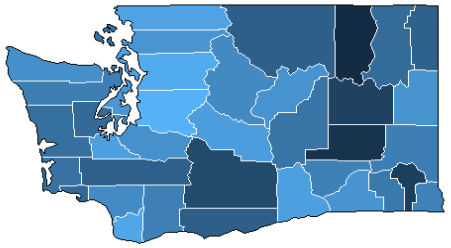
- For customers' practical use: intuitive and easily accessible

DUKE FUQUA SCHOOL OF BUSINESS

# Regression Analysis – Intuitive & Stepwise Selection Models

Intuitive Model includes **9** variables across **4** categories, with an AIC of **284914** and McFadden‘s $R^2$ of **0.0950;** Final AIC Stepwise Selection Model includes **15** variables, with an AIC of **274828** and McFadden's $R^2$ of **0.1273**

## Geographic



- Vary in different County
- Low minority population

## Demographic

- Female
- White
- High income

## Loan


- Home purchasing
- Conventional loan

## Property

- One-to-four family dwelling

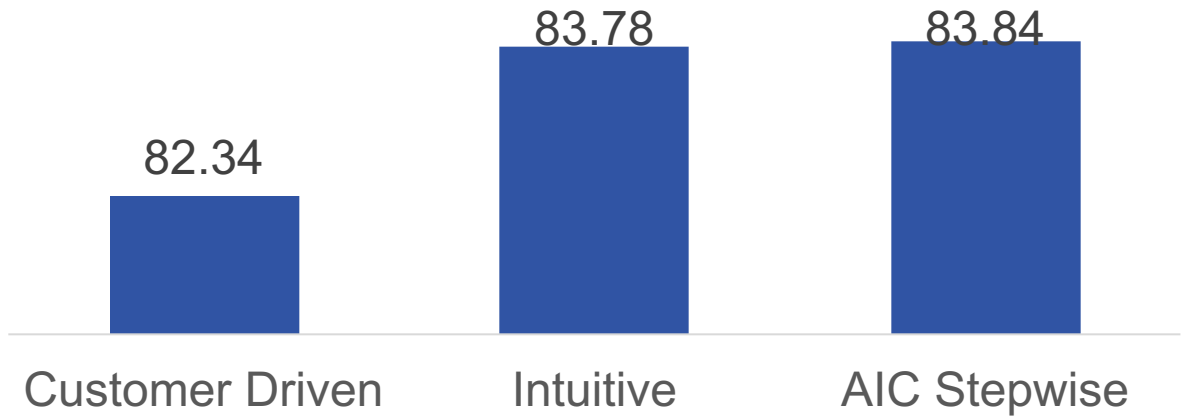
DUKE FUQUA SCHOOL OF BUSINESS

# Prediction & Empirical Results

| | Approval (A) | Deny (D) |
|---|---|---|
| **True Prediction (T)** | T-A | T-D |
| **False Prediction (F)** | F-A | F-D |

## Accuracy Rate:

- E(Approval)/SUM(Approval)
- Evaluate the accuracy of model's prediction ability



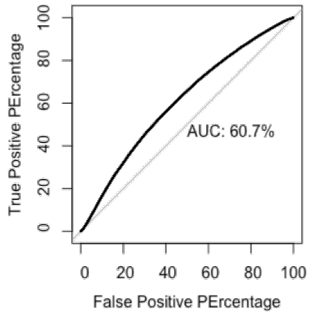

DUKE FUQUA SCHOOL OF BUSINESS

# Prediction & Empirical Results

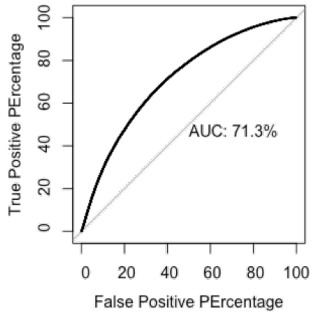
| | Approval (A) | Deny (D) |
|---|---|---|
| **True Prediction (T)** | T-A | T-D |
| **False Prediction (F)** | F-A | F-D |

## ROC & AUC

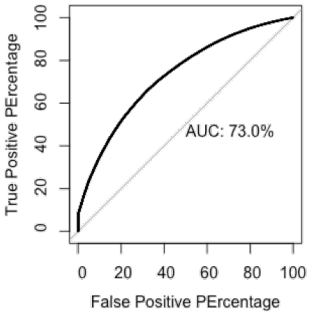
- ROC: Receiver operating characteristics
- AUC: Area under the curve
- Check classification model's performance




60.7%




71.3%




73.0%

DUKE FUQUA SCHOOL OF BUSINESS

# Summary

- Data visualization suggests that loan approval links with demographic, geographic, property-related and loan-related information
- We have built 3 models using logistic regressions: Customer-driven Model for borrowers, Intuitive Model based on tract information, and a model using AIC Stepwise Selection for loan issuers

DUKE
FUQUA
SCHOOL OF BUSINESS