Sprint 3 – Capstone

Lending Club: Loan Approval Prediction

- Vidya

Automating Loan Approvals

Objectives

- ❑ Objective 1: Predict Loan default
- ❑ Objective 2: Identify features influencing likely approval.

Solution:

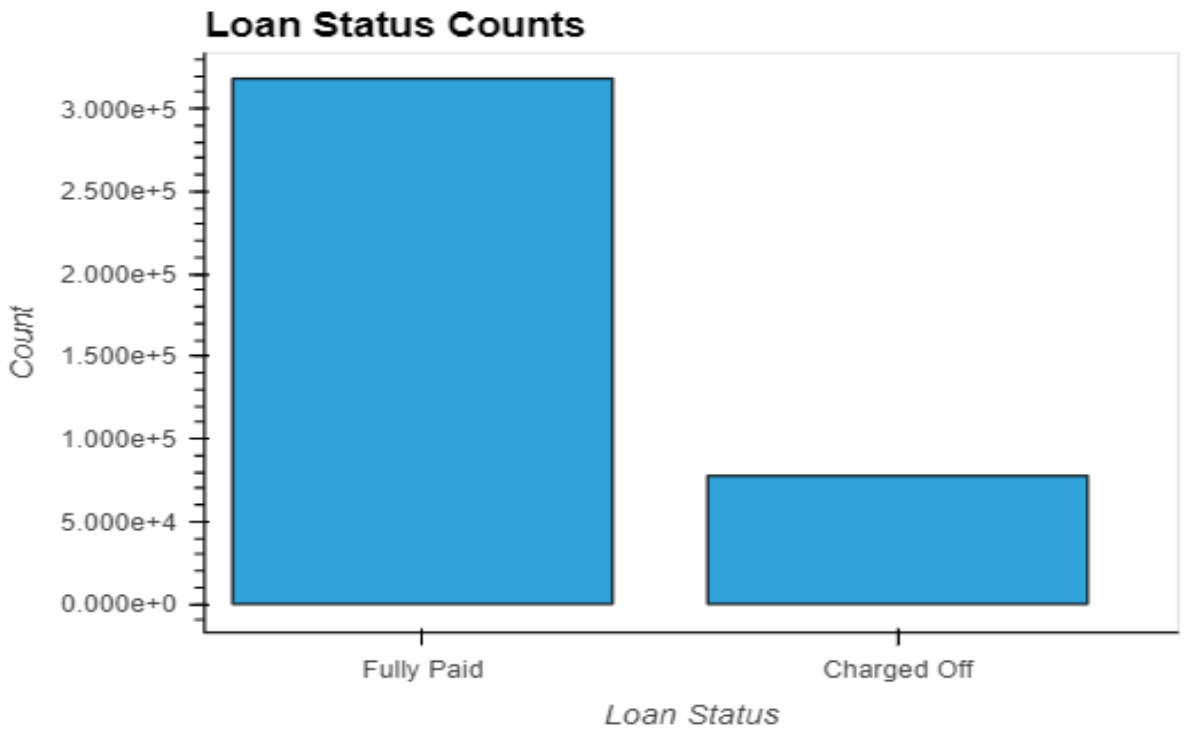
- **Problem**: Binary Classification Problem
- **Data**: Analyze historical loan data over 2008 to 2016
- **Modeling**: Use ML models for prediction of target class labels
- **Evaluate:** Determine best models

Impact:

- ✓ Reducing human intervention in loan approval process.
- ✓ Expedite decision-making and improve efficiency

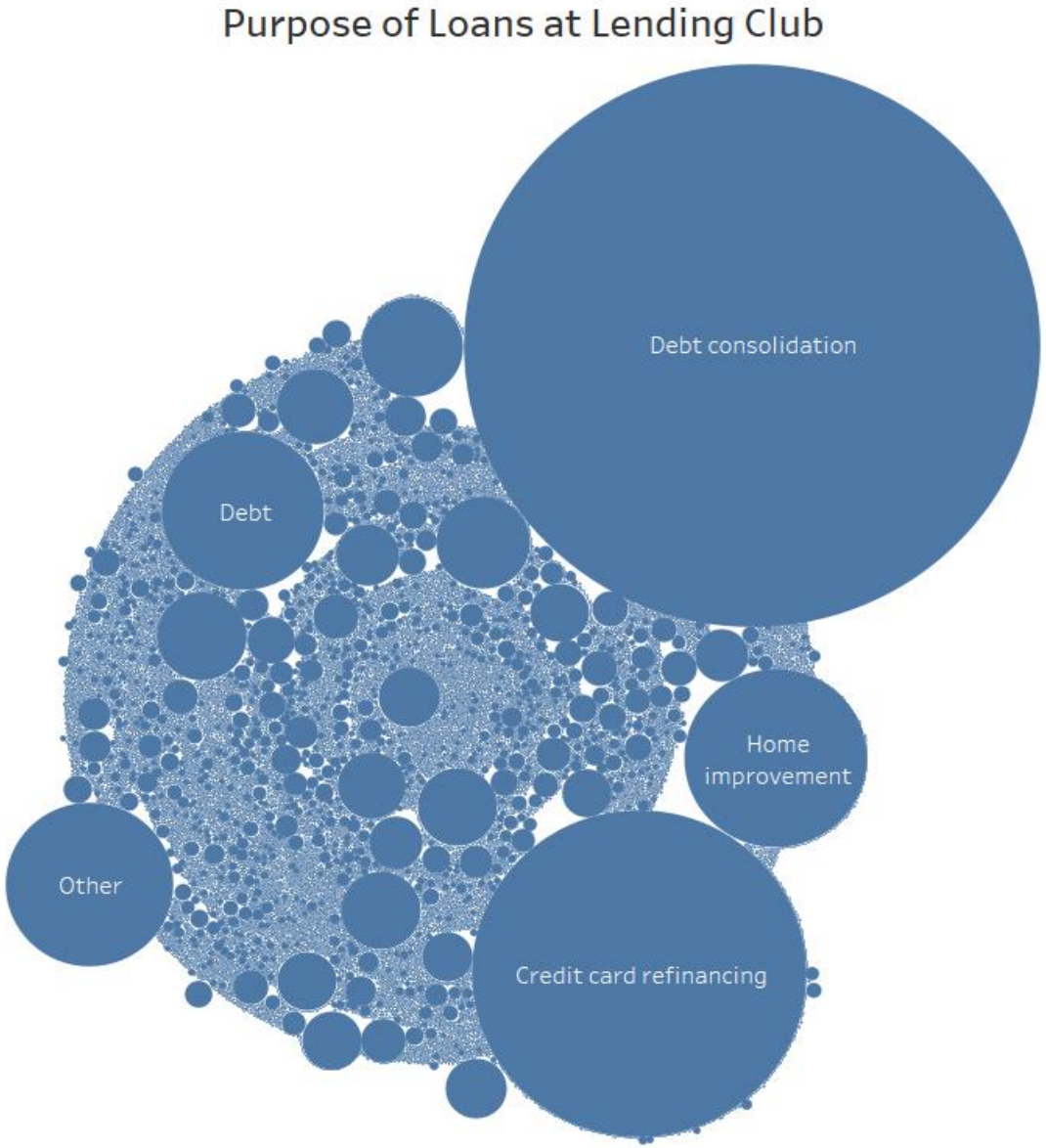
Data Preparation

- Dataset: (300k, 27)
- Issues:
  - Null values
  - Too many categories in some categorical columns
  - Imbalanced data (80:20)
  - Skewed distribution



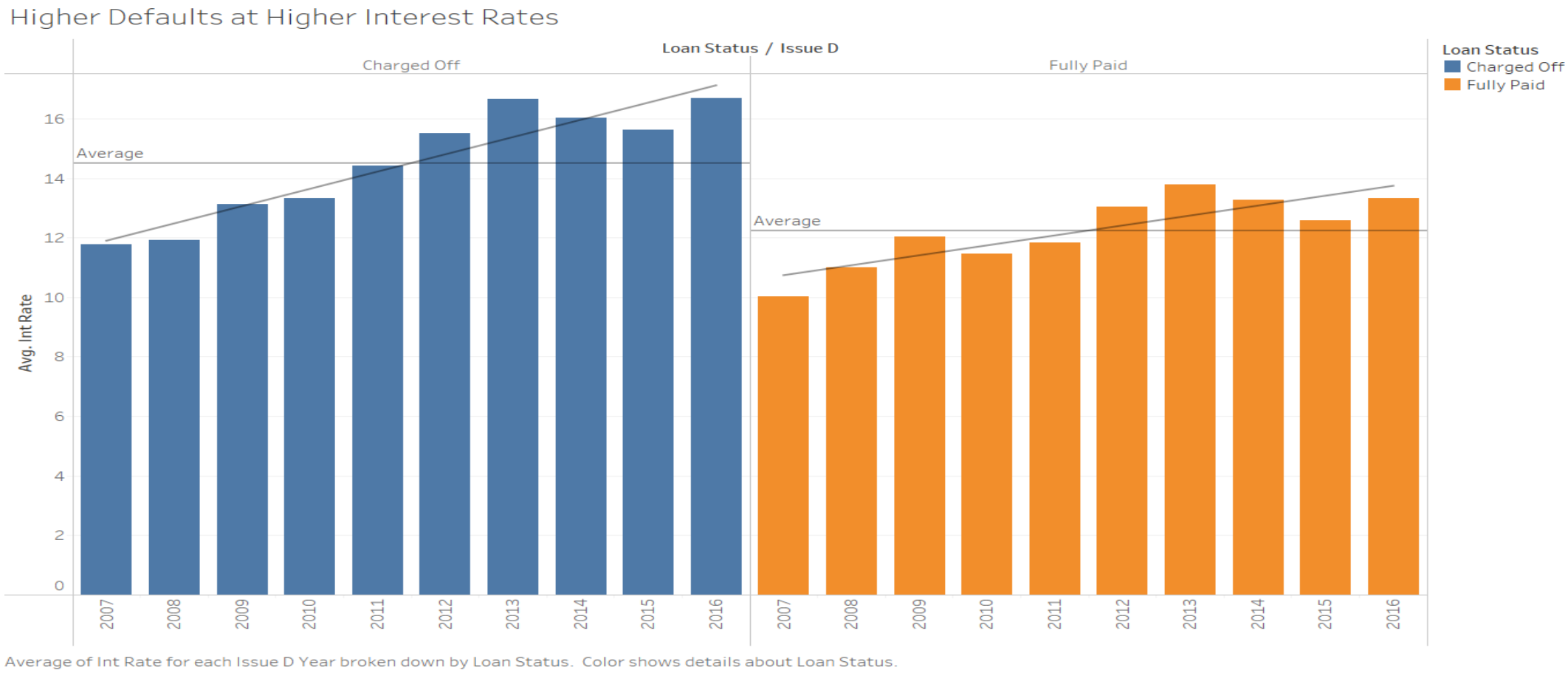
	mort_acc	emp_title	emp_length	title
% of Nulls	9.54	5.79	4.62	0.44

Types of Loans at Lending Club

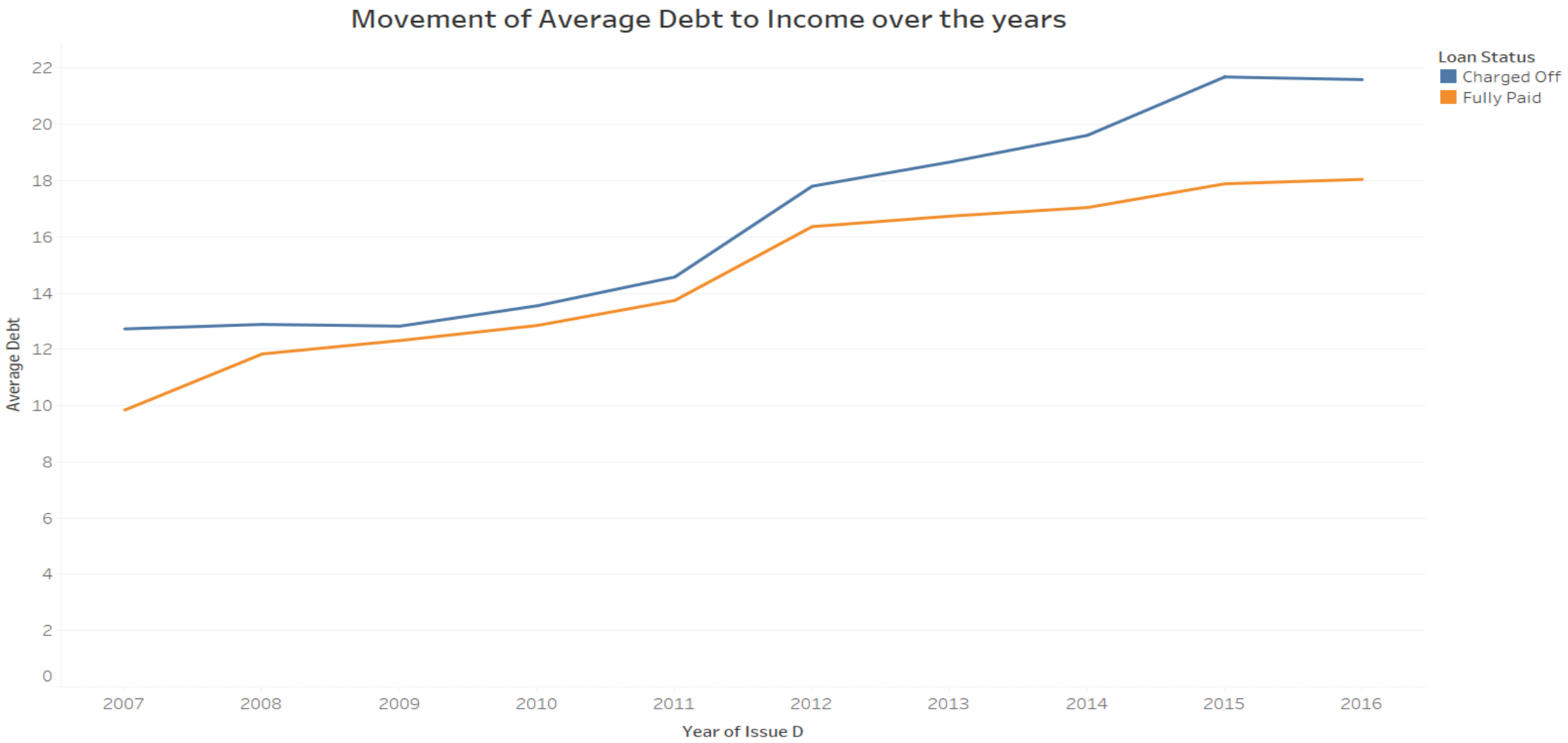


Title. Size shows count of Loan Status (copy). The marks are labeled by Title.

Story of Interest Rates



Is there a difference between Debt/Income?

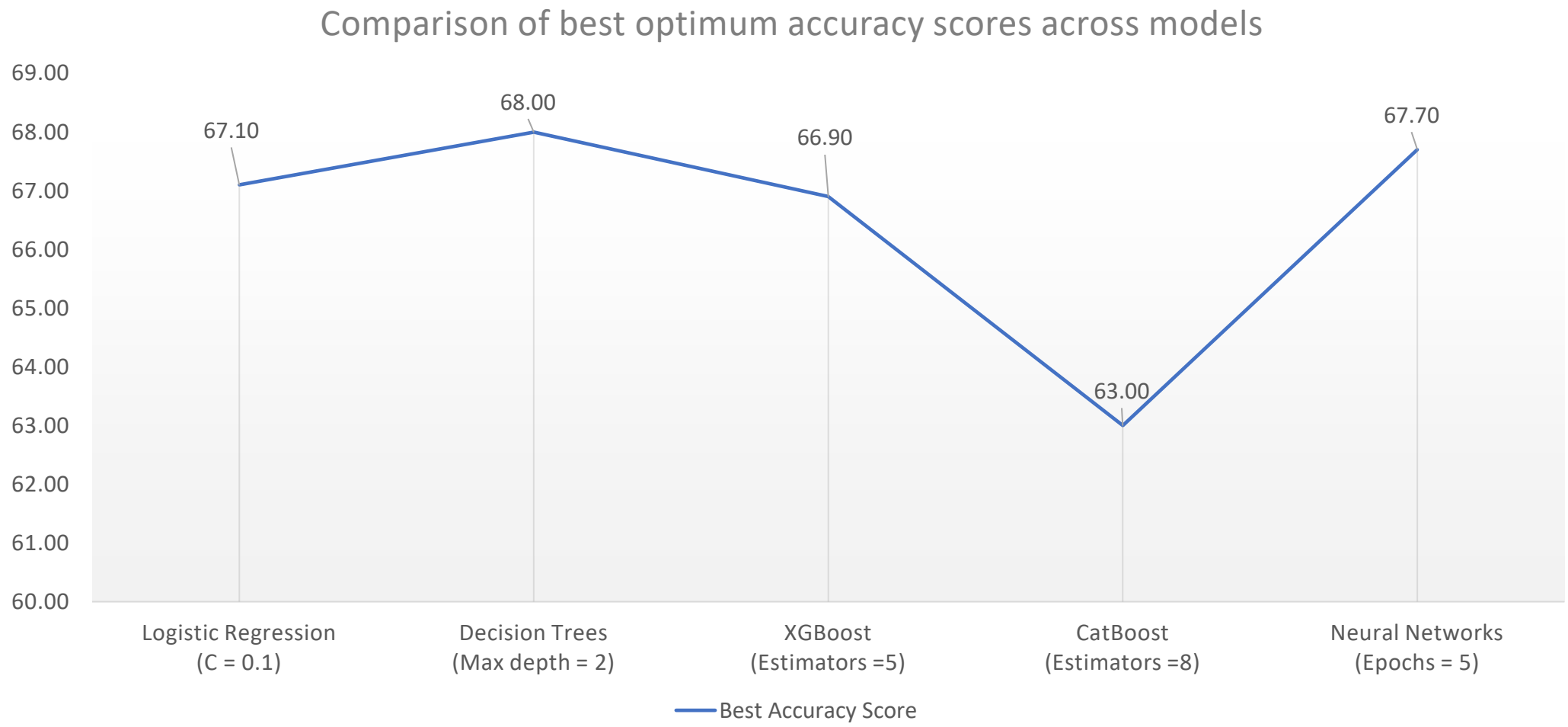


The trend of average of Dti for Issue D Year. Color shows details about Loan Status.

Machine Learning Models

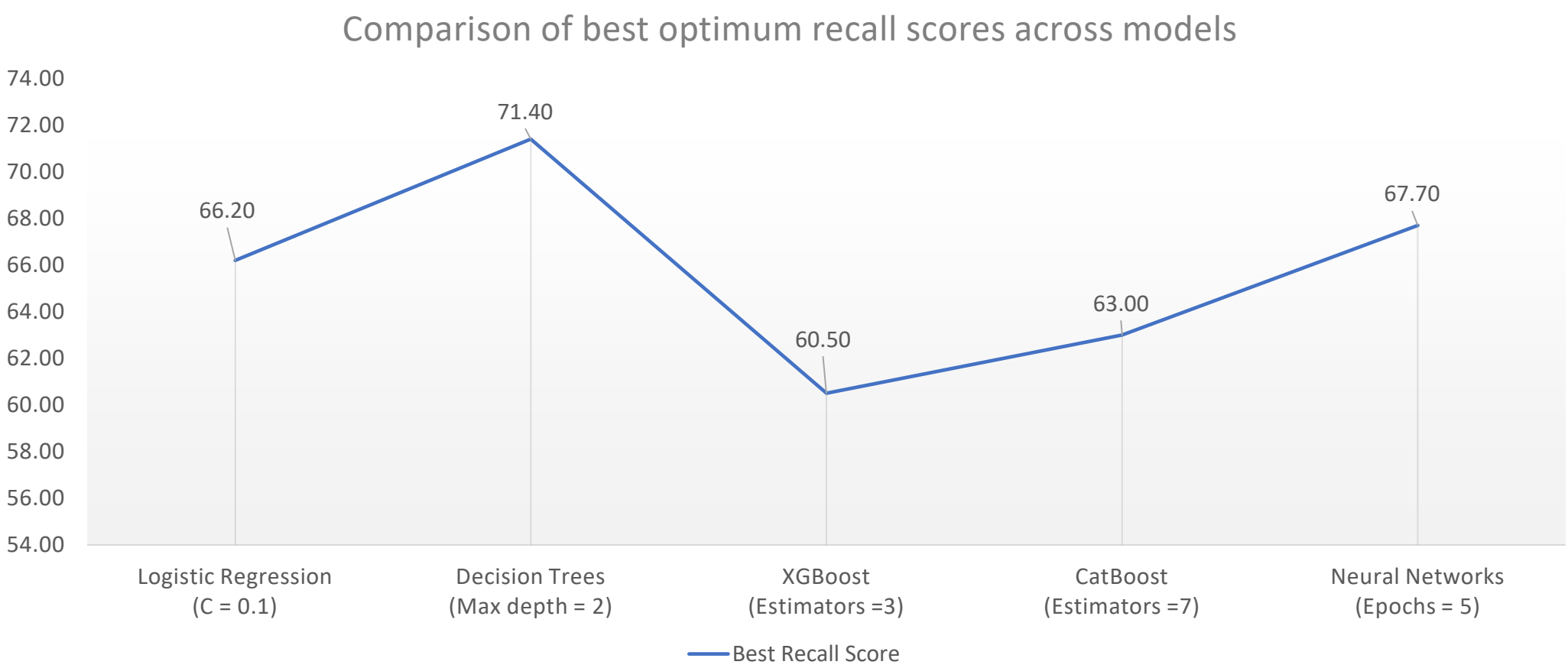
- Logistic Regression
- Decision Trees
- XGBoost
- Catboost
- Neural Network

How do different Models score in Accuracy?



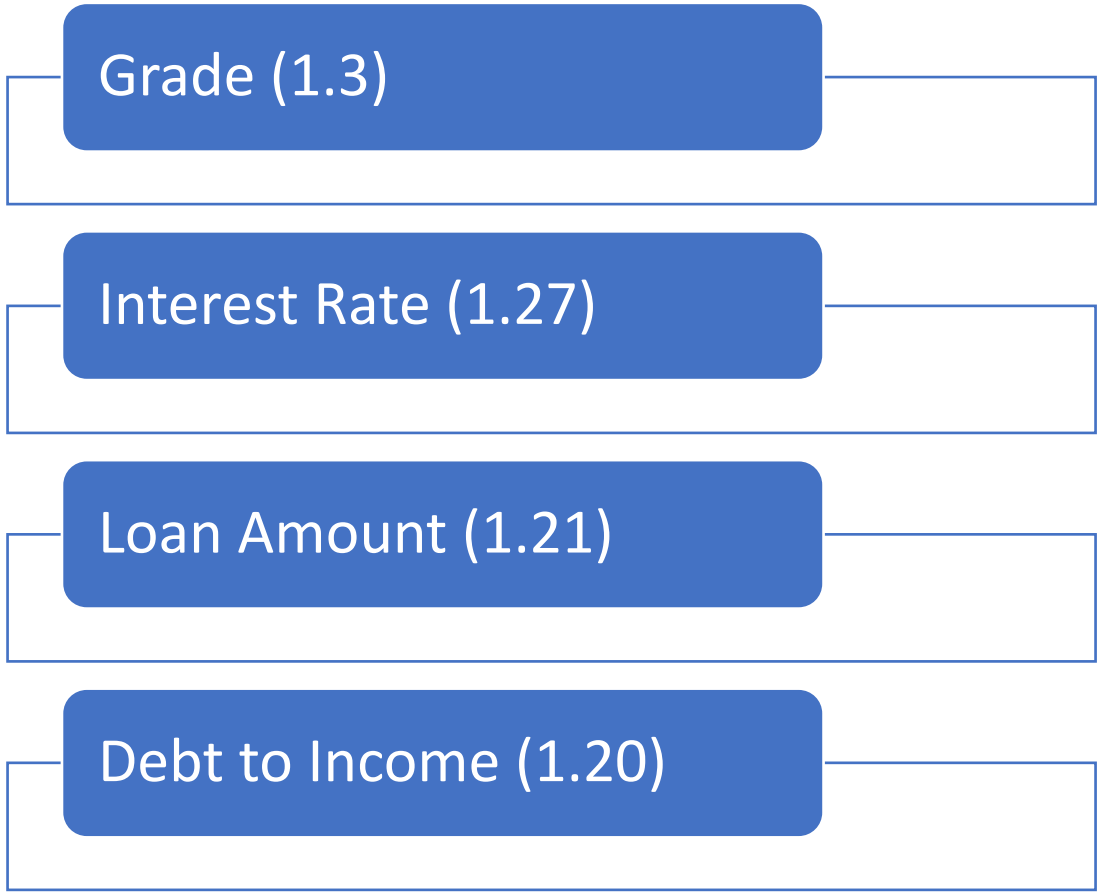
The best-performing model is Decision Trees, with 68% accuracy

How do the Models score in Recall?

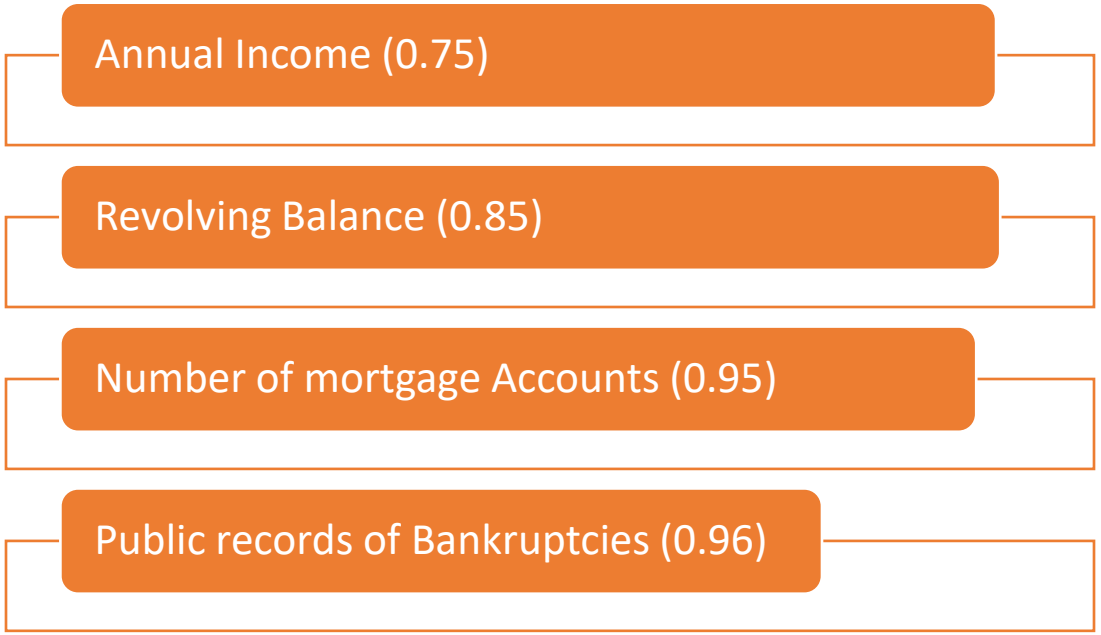


The best-performing model is Decision Trees, with 71.4% recall score

Features with highest odds ratio in prediction of Loan performance in Logistic Regression



Features that have the lowest Odds Ratio on Loan Performance prediction in Logistic regression



Further Tuning with more hyper parameters

Tuning thresholds for classification

Thank You

Vidya Sagar Botcha

Commercial Banker turned Data Scientist