# MORTGAGE CLASSIFICATION

Ryan Sims

# PROBLEM STATEMENT

- **Purchase vs. Refinance**

- **Multitude of columns per mortgage**

Given all the data on mortgages, its quite difficult to easily navigate through and pick out important pieces.

Can we classify between purchase and refinance

Classifying purchase vs. refinance enables ability to look at important features contributing to classification
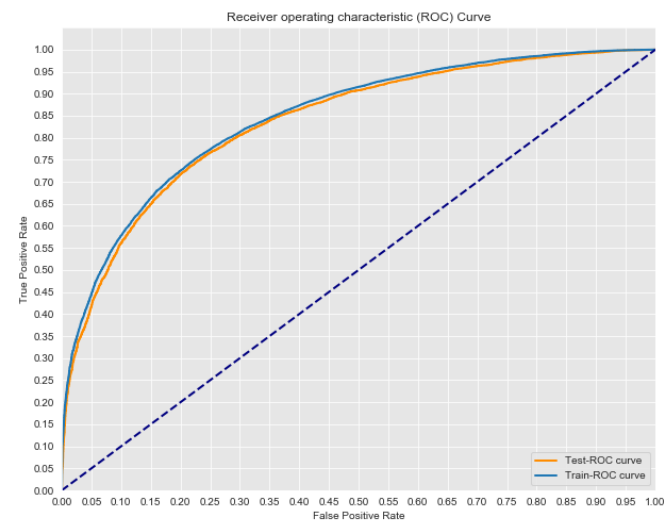A well generalizing model will reproduce results with new data

# METHODOLOGY

- **Supervised Learning**

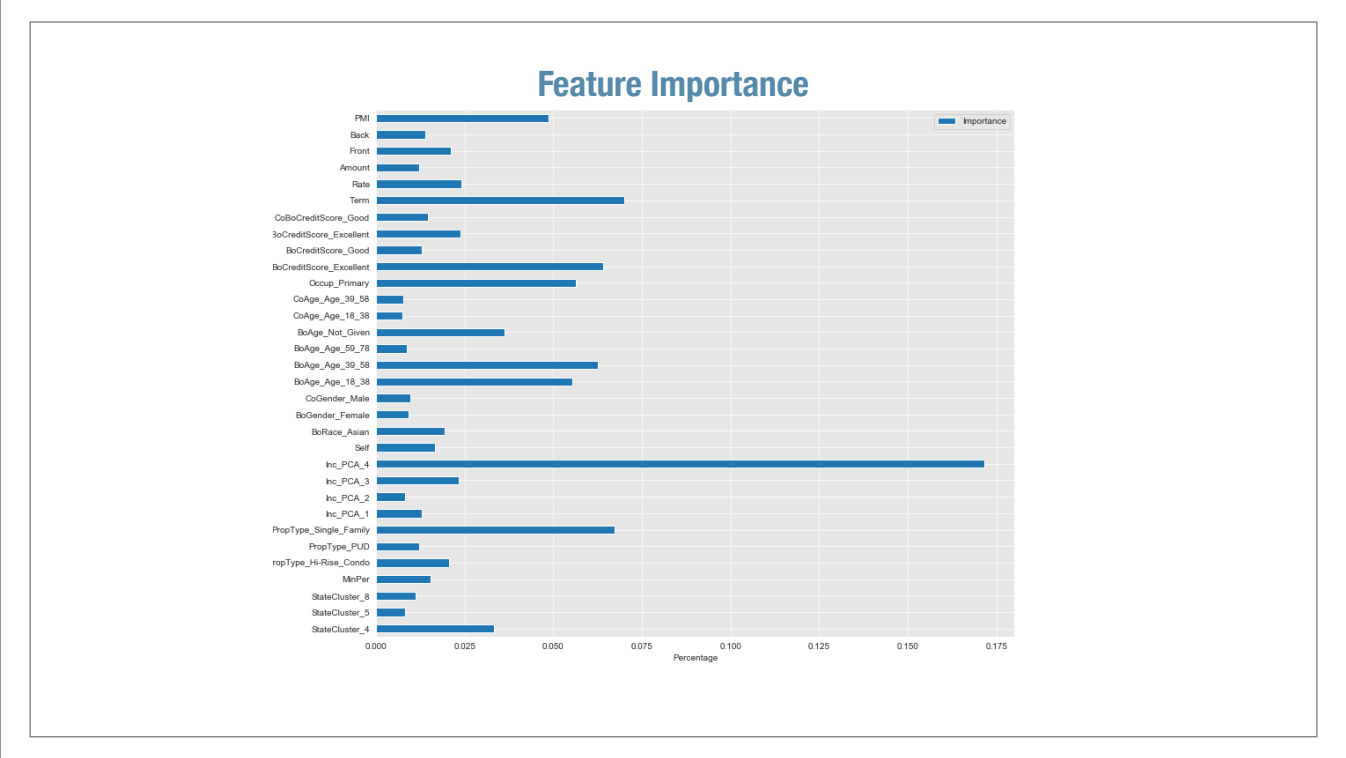- **Unsupervised Learning:**

    - **KMeans**

    - **PCA**

Supervised learning for classification
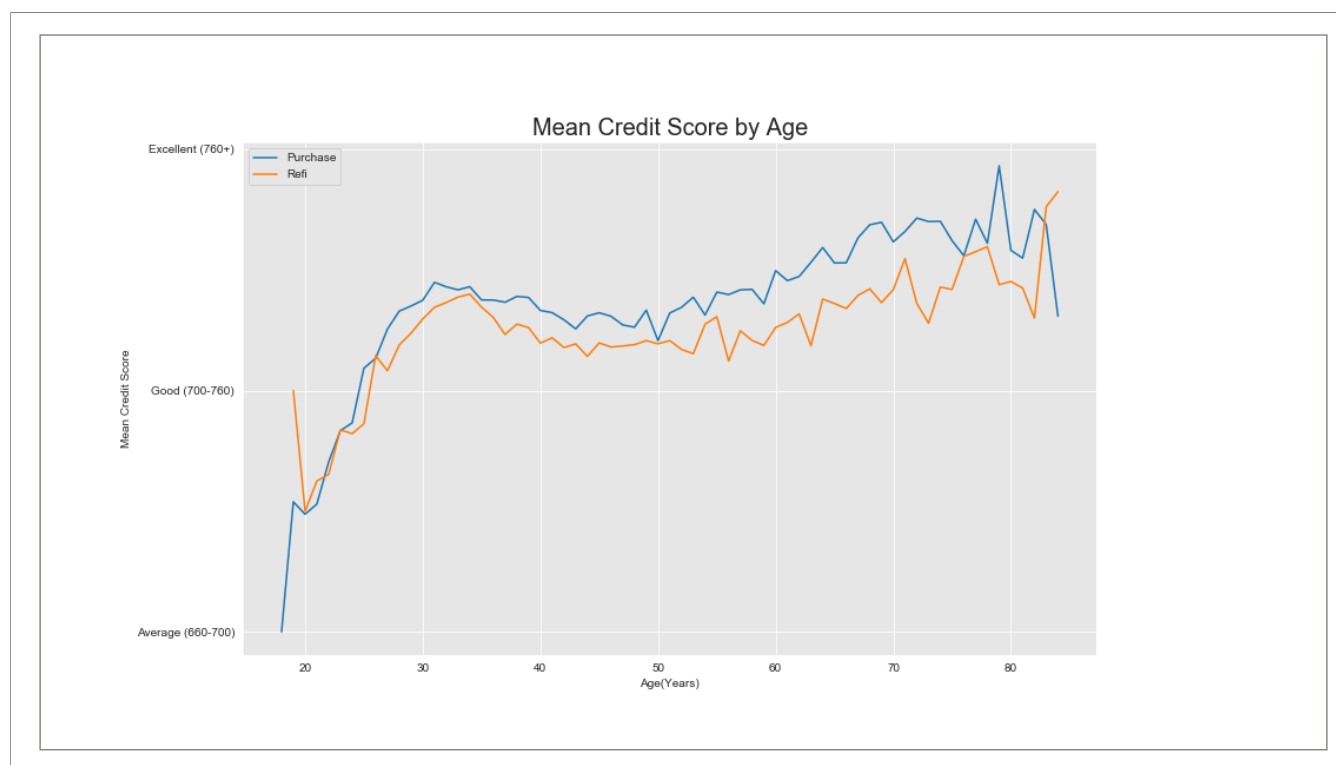Unsupervised for clustering and dimensionality reduction
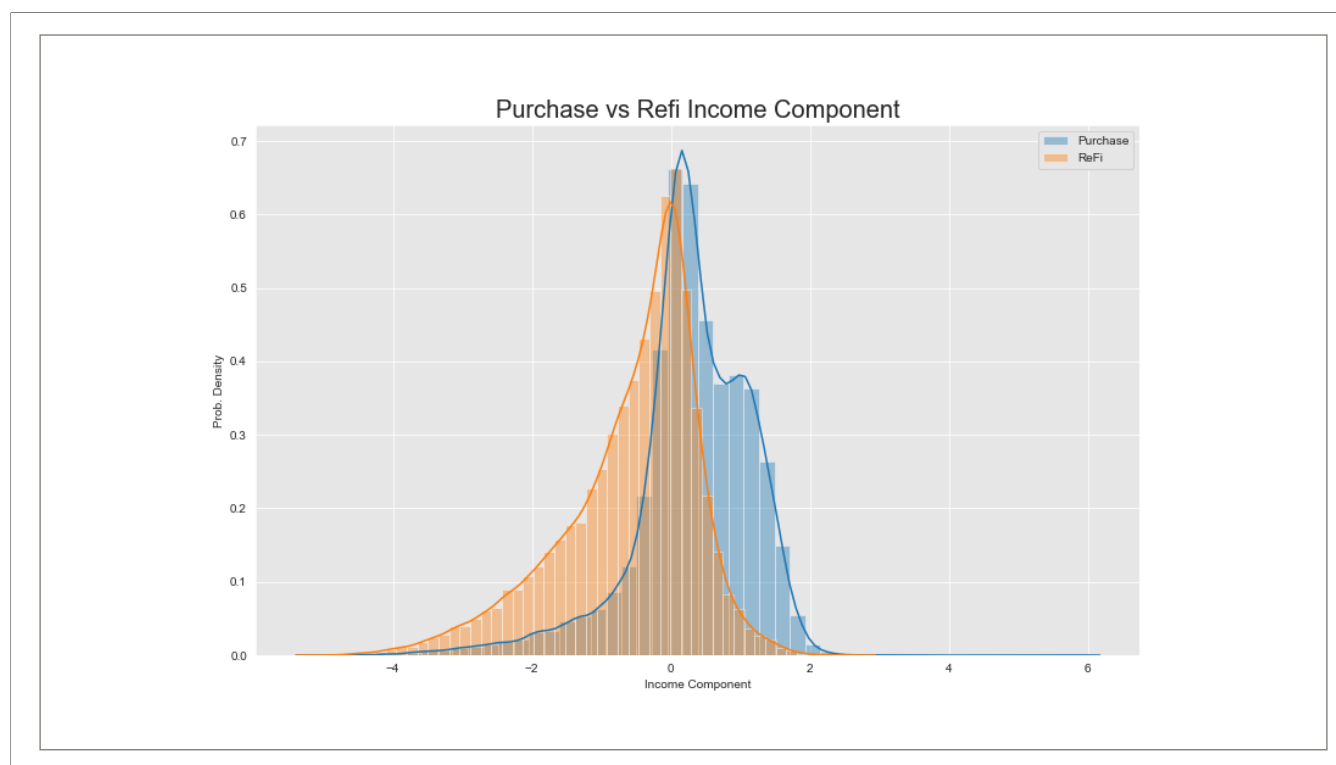
# FINDINGS



Receiver operating characteristic (ROC) Curve

Accurate model with good generalization

Feature Importance

Features that were useful in deciding between purchase or refi

Mean credit score by age

Purchase vs Refi Income Component

Notice difference between purchase and refi

# SUGGESTIONS

- Pre-screen borrowers by age

- Credit score importance

Ages younger than 40 more likely to purchase vs over 40 more likely to refi

# FUTURE WORK

- Classify on First time buyer

- Incorporate more geographical data

Look at importances that signify a first time buyer, since it was important in classifying purchase vs refi
Get geographical data to be more useful

# Thank You