CISD 43 Final Project

Group Member: YongSheng Lin

Loan Data Analysis

Goal of the Project:

I chose to use the loan dataset for my project because my current job is related to mortgage loan collection. I know the data might not be most current, but it would be interesting to see model that apply, and relationship between different factors. I will be demonstrating the focus around interest rate and fico score. Hopefully can add other fields to test out.

Conclusion:

The project involves analyzing a loan dataset with a focus on interest rates and FICO scores. I've performed K-means clustering and found that the clusters indicate different risk profiles for loans. The project includes MongoDB integration, preprocessing, and using RapidMiner.

The result was consistent show clusters are relatively well-separated in the space defined by Interest Rate and FICO Score. Lower FICO scores and higher interest rates are grouped differently compared to higher FICO scores and lower interest rates. The clustering indicates different risk profiles for loans. For example, lower FICO scores and higher interest rates could be higher-risk loans.