HOME CREDIT DEFAULT RISK

Padma Shneha

Introduction

- Home Credit is a consumer finance provider.
- Little or no credit history
- Risk is higher when not provided with credit score
- They offer 3 types of loans:
  - ✓Point of Sale (POS) Loan
  - ✓Cash Loan
  - ✓Revolving Loan

Contents

- Problem Statements
- Data Analysis Pipeline
- Data Visualization
- Modelling
- Conclusion
- Reference

Problem Statement

To ensure that clients capable of repayment and are not rejected and that loans are given with a principal, maturity, and repayment calendar that will empower their clients to be successful.

Data Analysis Pipeline

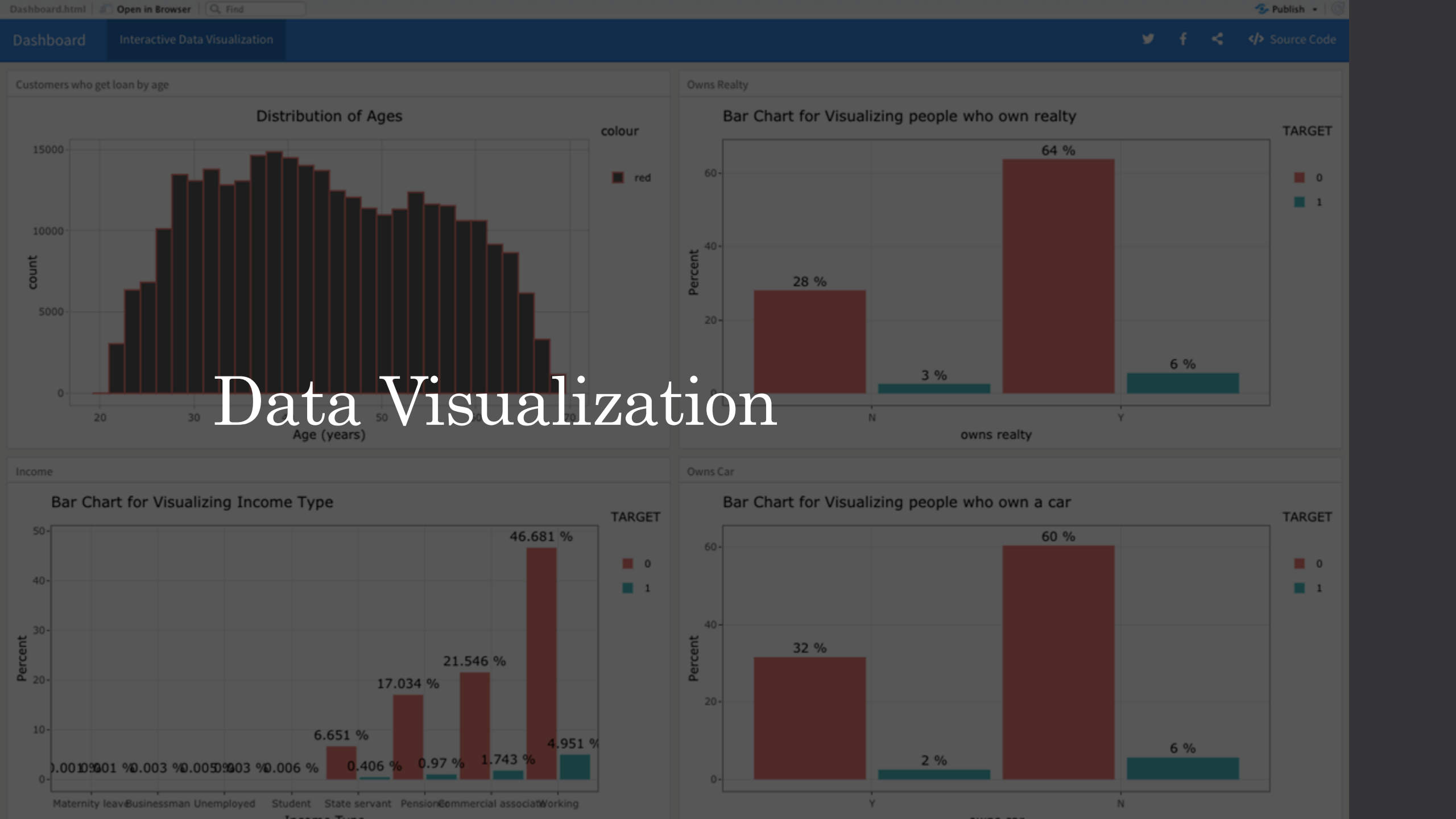
Data Preparation

- Missing Value Imputation
  - Mean, median
  - Features were removed if more than 85% of the data was missing
- Outlier Treatment
  - Visualized using Boxplot
  - Treated using log transformation
- Anomaly Detection
  - Days had negative values and they need to be converted to years and with positive values.
  - For eg: `DAYS_BIRTH : int -9461 -16765 -19046 -19005 -19932 -16941 -13778 -18850 -20099 -14469 ...`

Exploratory Data Analysis

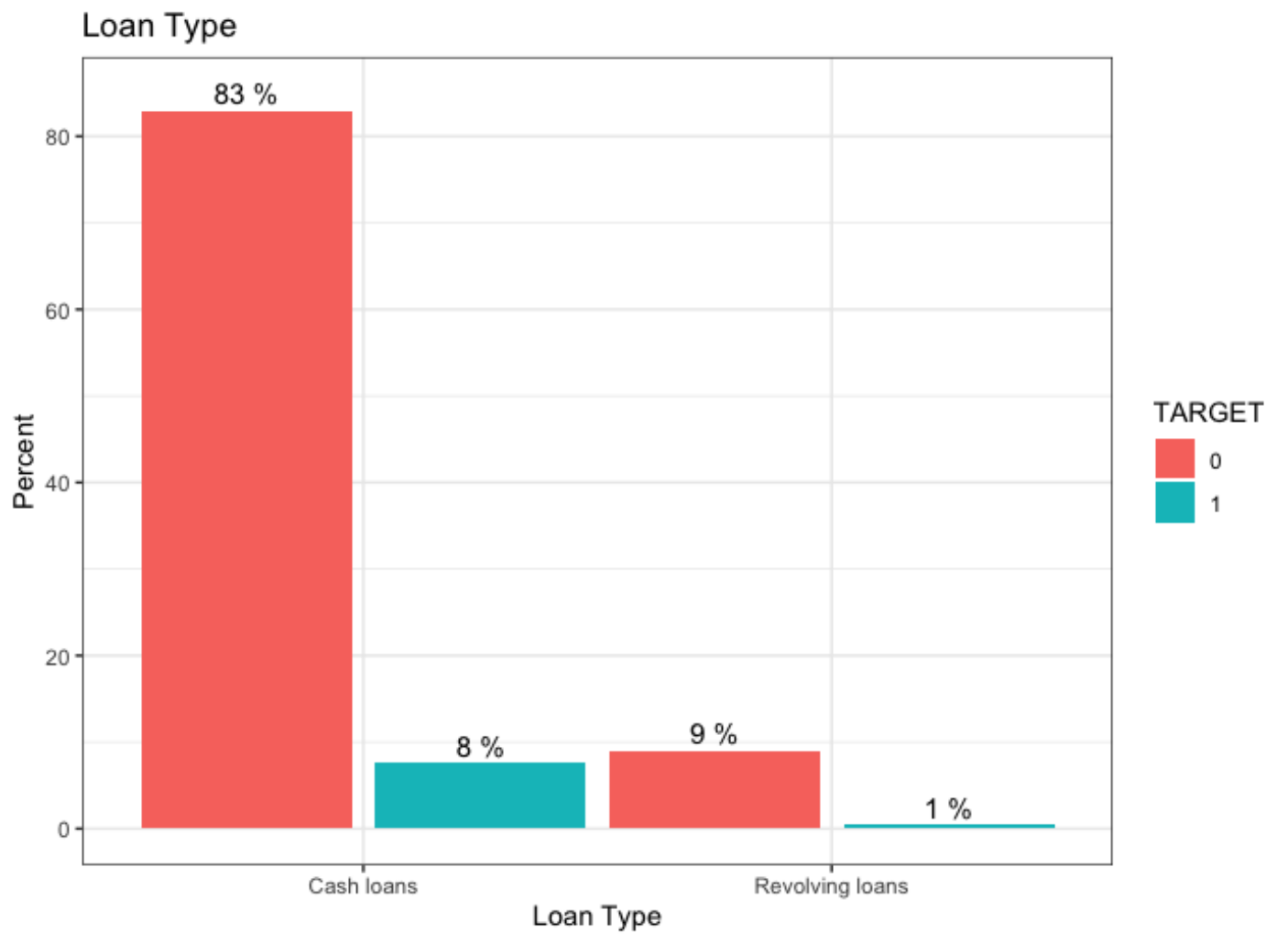
PCA

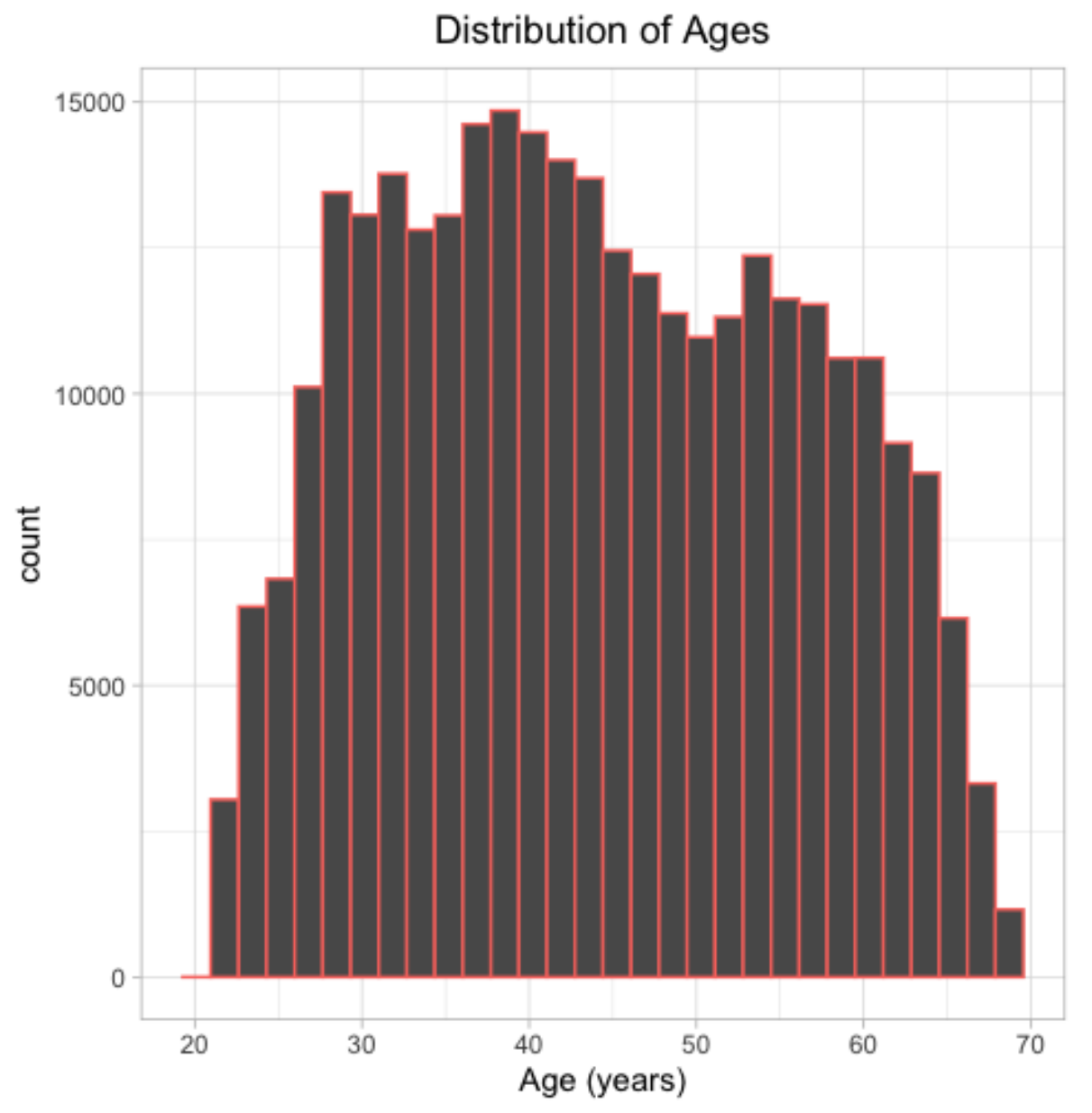
- Dimension Reduction
- To focus on most important and impactful data
- Avoid overfitting of data



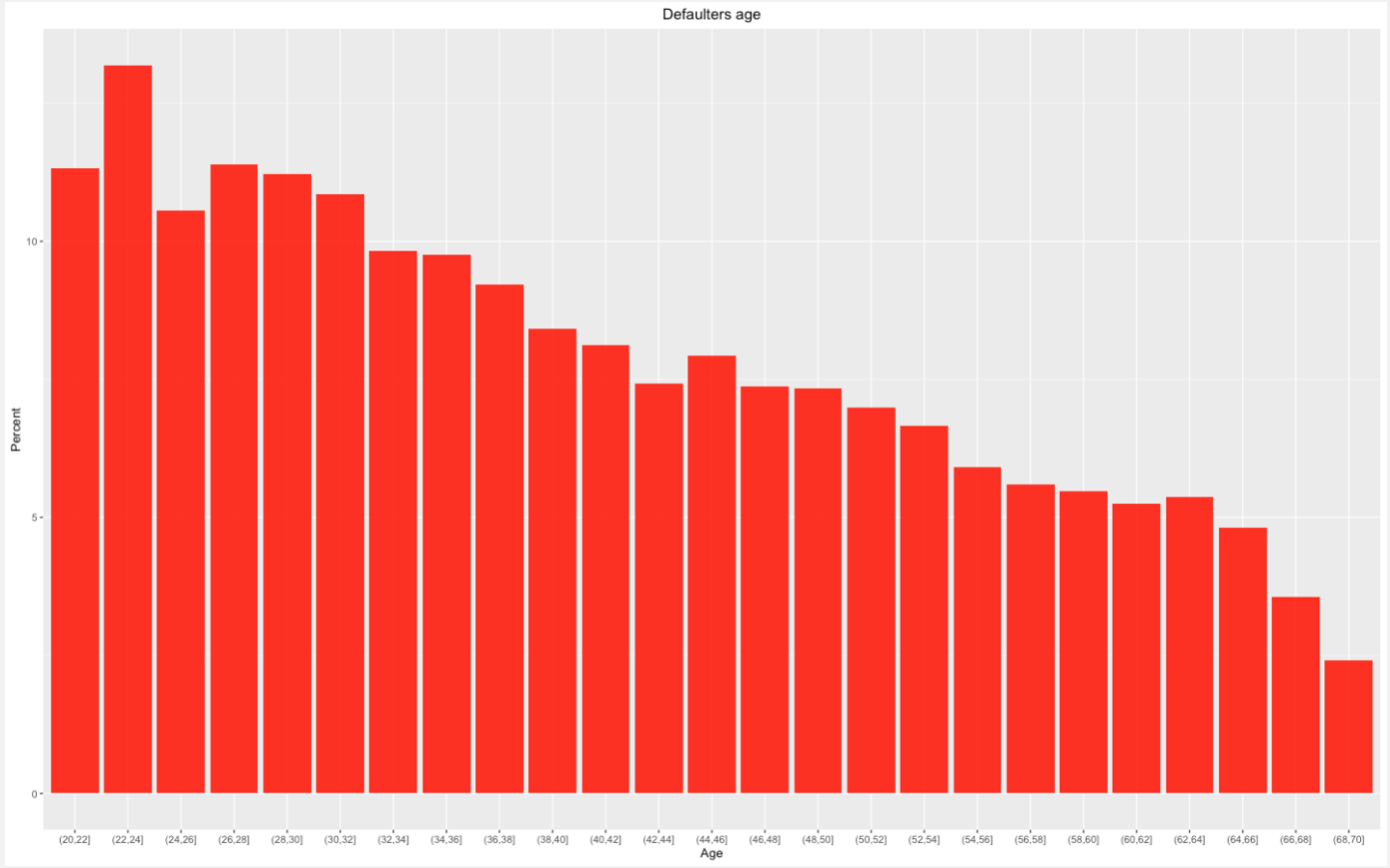
Data Visualization

Loan Type

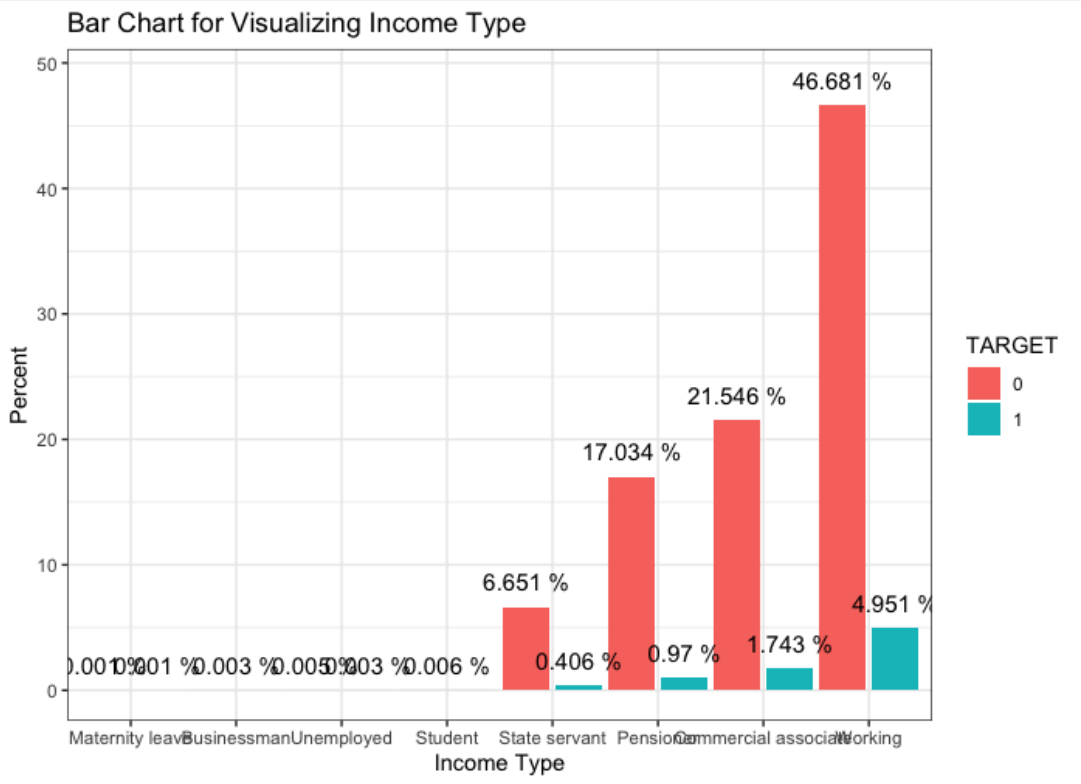

Loan Type
83 %
8 %
9 %
1 %
80
60
40
20
0
Percent
Cash loans
Revolving loans
Loan Type
TARGET
0
1



Distribution of clients age



Defaulters Age



Clients Occupation

Modeling Type

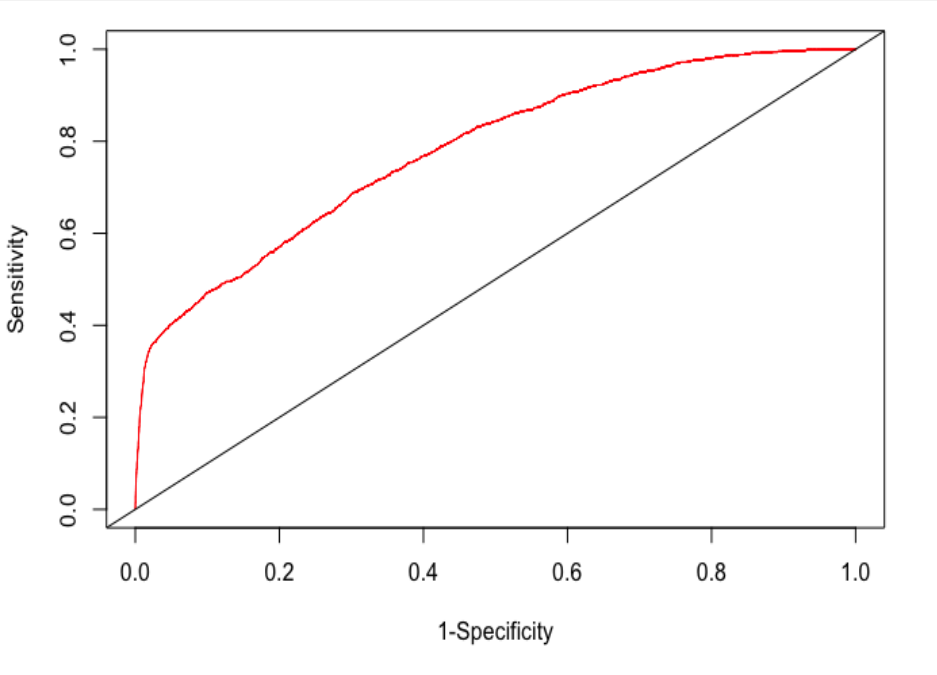
- Logistic Regression
- Support Vector Machine
- Random Forest

Classification Models

Model	TPR	FPR	Accuracy
Logistic Regression	75.393%	7.5%	83.27%
Support Vector Machine	62.37%	8.9%	73.76%
Random Forest	72.9%	6.4%	77.72%

Model Evaluation

ROC Curve



AUC Score

Model	Score
Logistic Regression	0.77
Support Vector Machine	0.69
Random Forest	0.71

Conclusion

- Logistic regression seems to provide more accuracy and TPR compared to other algorithms

- In the future, giving each feature a value/score depending on their importance and then considering that as a report for the future customers and avoid defaulters might help.

Reference

[1] https://www.homecredit.net/

[2] https://www.kaggle.com/c/home-credit-default-risk

[3] Gareth James, Daniel Witten, Trevor Hastie, Robert Tibshirani, An Introduction to Statistical Learning

[4] George Tauchen, Hao Zhou, Realized jumps on financial markets and predicting credit spreads (March 2010)

[5] Marco Lo Duca, Tuomas A. Peltonen, Assessing systemic risks and predicting systemic events (July 2012)

[6] S-M lin, J Ansell, G Andreeva, Predicting default of small business using different definitions of financial distress (March 2011)

[7] Vasant Dhar, Data Science and Prediction, (December 2013)

Thank You!!