LINE OF CREDIT ON HOUSEHOLD'S USED CAR

CUSTOMER PROPENSITY MODEL REPORT

By: Mei Hwa Wong

CONTENT

PROBLEM DEFINITION

- National City Bank has a line of credit product for customers.

- The bank aim to offer this new product to prospective customers that are likely to take up the product.

- The marketing department needs to identify the top 100 customers the bank should approach to promote the product.

THE DATA

Training data

Current customers dataset

- containing 4000 observations

Customers credit dataset

- containing 5000 observations

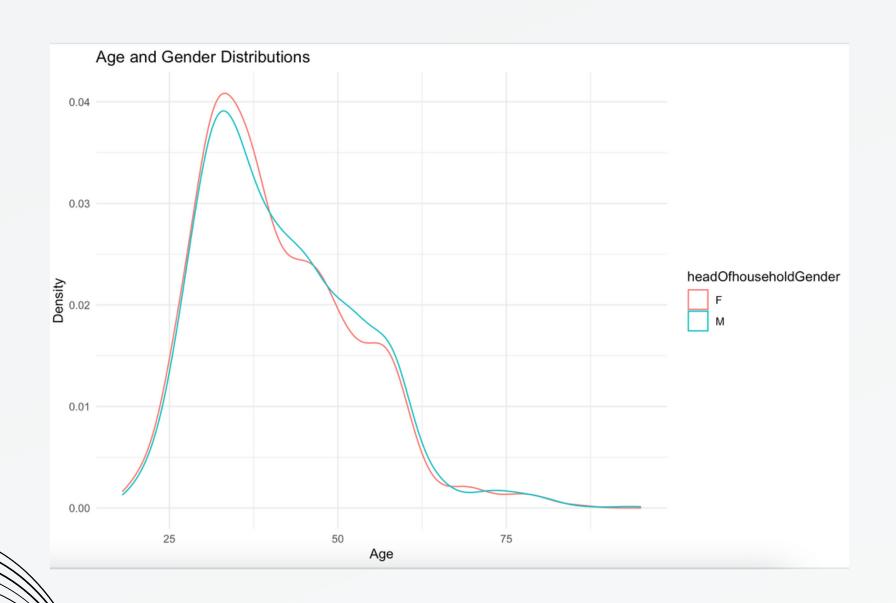Axiom describing customers' lifestyle, gender, race, etc. dataset
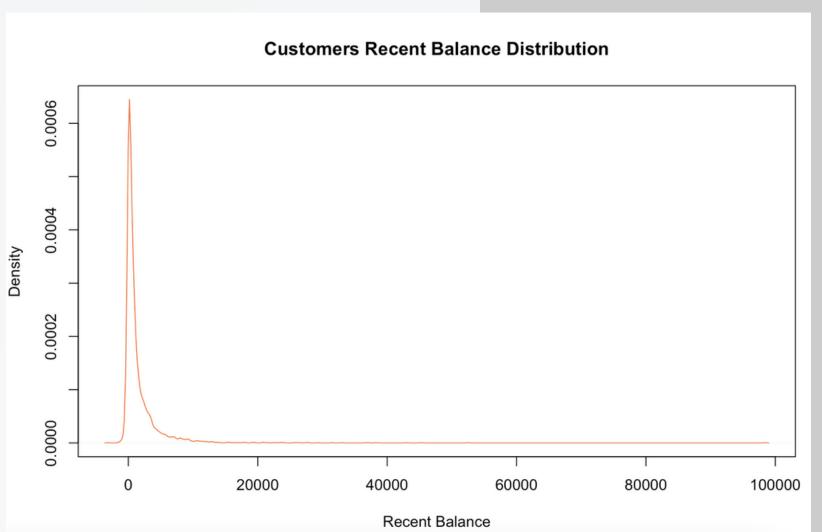
- containing 5000 observations

Prediction data

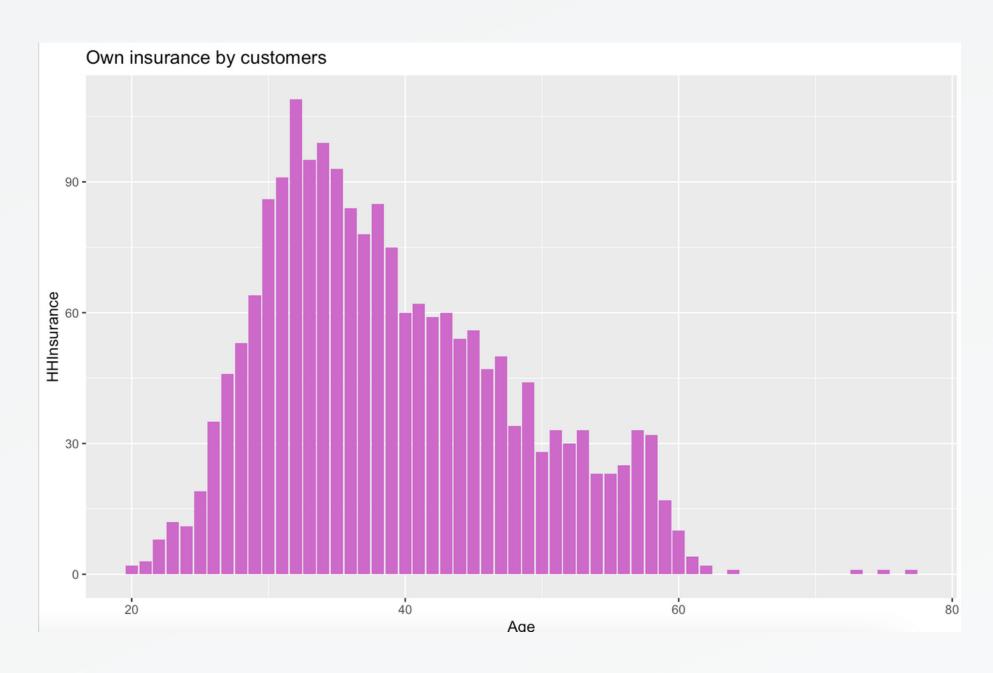Prospective customers dataset

- containing 1000 observations

OVERVIEW

Age and Gender Distributions

Customers Recent Balance Distribution

Customers are within 30-45 age group

Majority customers credit within recent balance below $20,000

OVERVIEW

Own insurance by customers

Customers in the younger age group owns HH insurance

METHODOLOGY

○ **Exploratory Data Analysis**

identify outliers, missing or incomplete records, unstructured data

○ **Data Preparation - Data Cleaning**

fill null values, create group or bins to categorize variables, remove outliers

○ **VTREAT package**

utilized VTREAT package to handle null values, create booleans for categorical data

○ **Select predictor variables**

○ **Apply modeling methods**

DATA PREPARATION

1. Remove outliers in Recent Balance column in training set.

2. Group 'Age' and 'Recent Balance' to bins to convert continuous variable to discrete variable.

3. Columns with " " have been replaced with NA to ensure VTREAT package can be applied properly.

4. Split data to training, validation and test at 75%, 15% and 10% respectively.

5. To prepare data for modeling, columns not included: -
   dataID, HHuniqueID, Call Start, Call End

MODEL METHODOLOGY

- 3 models were utilized in this machine learning exercise
- Supervised learning approach

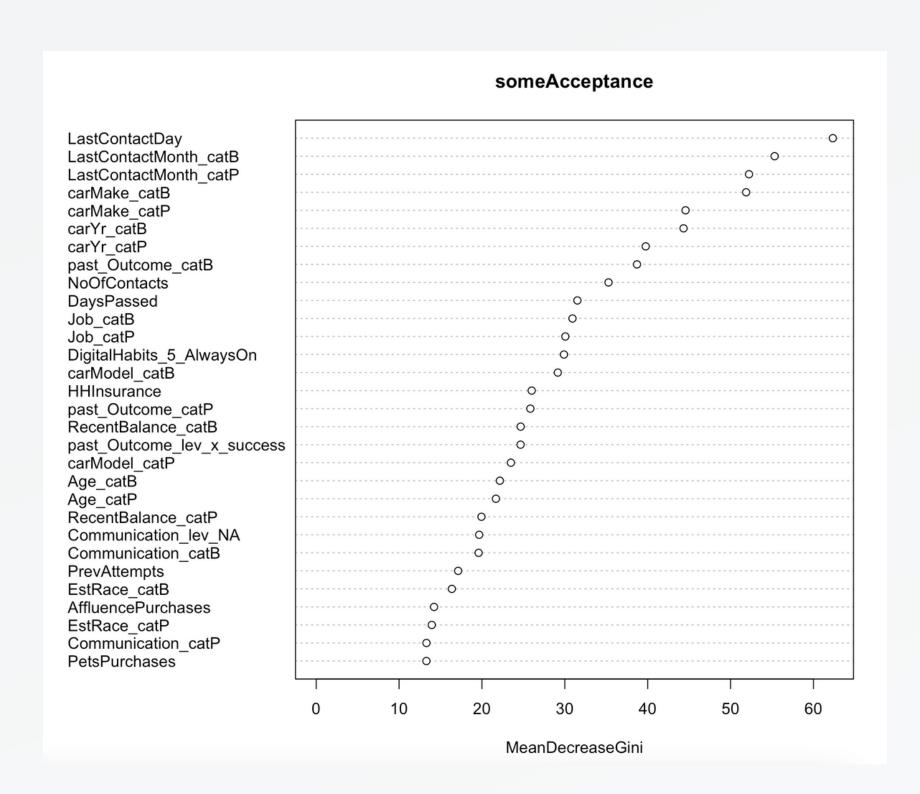|		
- To model probability of binary responses - Handles both categorical and numerical predictor - To find the best fitting model that describes the relationship	- Allows nonlinear relationship - Create tree that can accurately predict target variable for new input data - Handles outliers and missing values better	- Ensemble model to improve accuracy and reduce overfitting - Algorithm splits the data into subset and build decision trees - ntrees = 250 gave the best result
LOGISTIC REGRESSION	DECISION TREE	RANDOM FOREST

MODEL RESULTS

The model is evaluated based on accuracy and F1 - score.
The model with a higher accuracy is chosen.

Model	Accuracy	F1-Score
Logistic Regression	0.39	0.31
Decision Tree	0.72	0.55
Random Forest (n=250)	0.73	0.57

VARIABLE IMPORTANCE

someAcceptance

- Top important predictors in the random forest with 250 trees model

- Guide on features that have the biggest impact on the outcome

HIGHER TENDENCY FOR AGE GROUP IN 45+

💡 more receptive on this product option because this age group have the highest debt

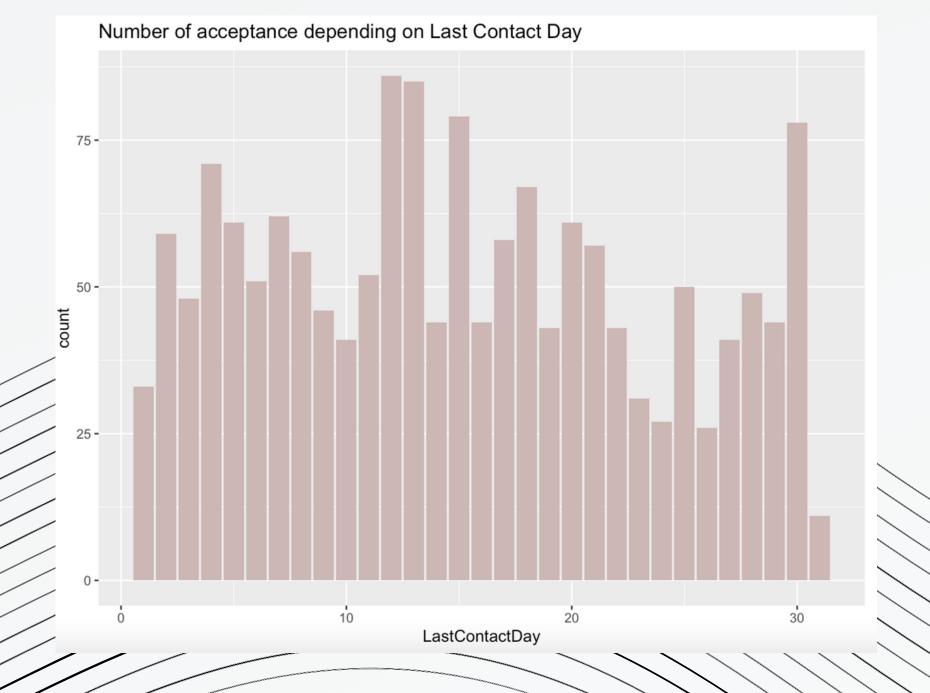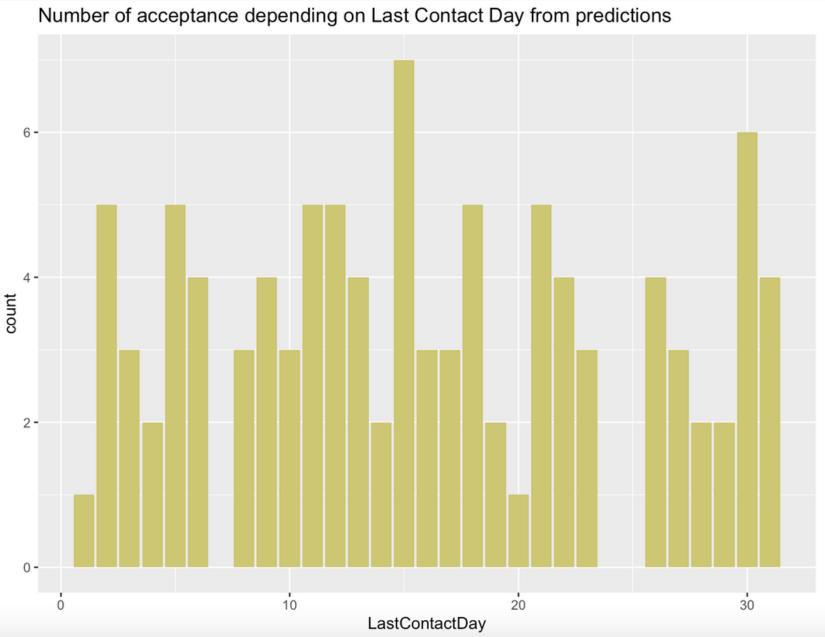💡 auto loan equity have a higher rate of approval hence is more attractive to customers in the age group

CAR YEAR LESS THAN 10 YEARS HAVE HIGHER ACCEPTANCE
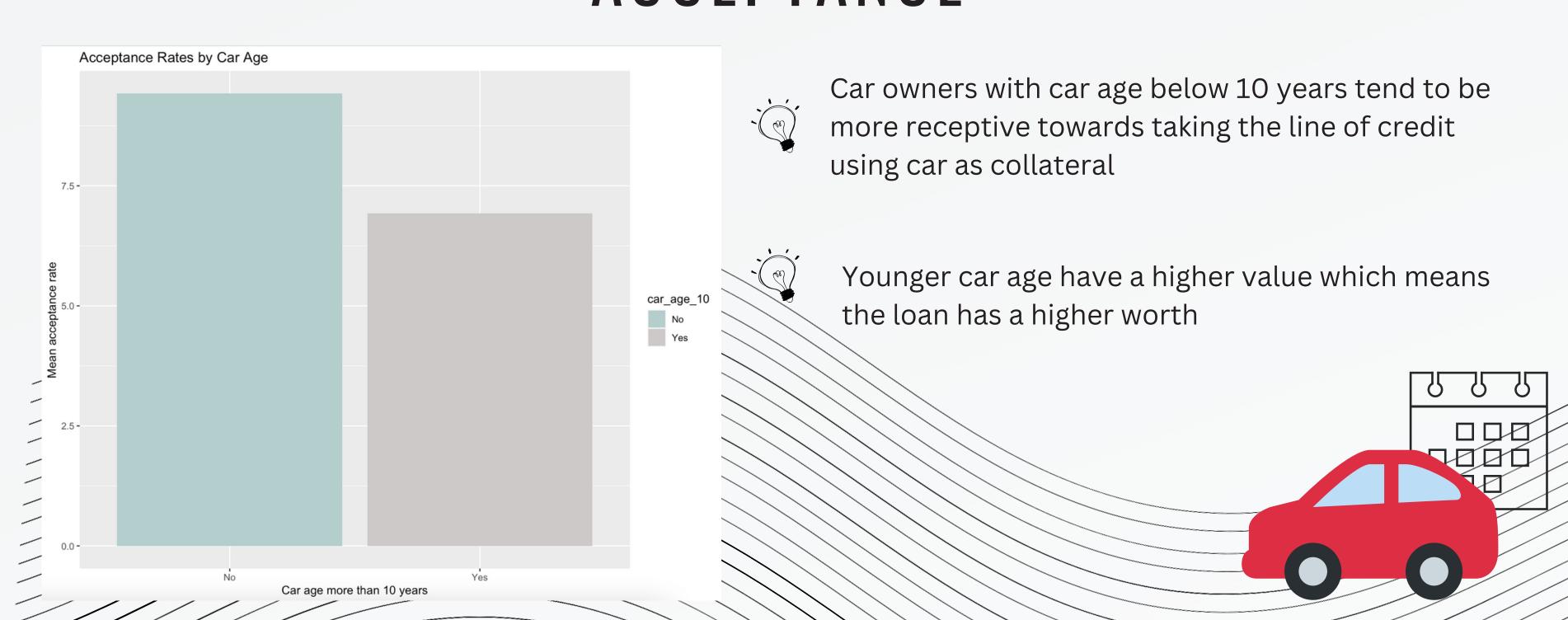

Acceptance Rates by Car Age

Car owners with car age below 10 years tend to be more receptive towards taking the line of credit using car as collateral

Younger car age have a higher value which means the loan has a higher worth
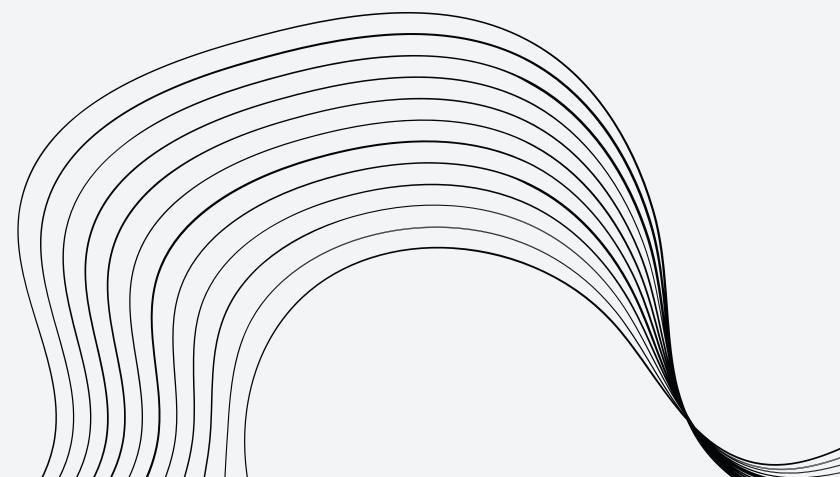
SUMMARY

- Customers in higher age group are more inclined to accept the line of credit against household used car.

- Closing of sales occur mostly in the middle and end of the month.

- Car owners with lower car age tend to accept the line of credit product.

Thank You