Case 4 Customer Retention

# Executive Summary

*Brief introduction of problem. Summarizes key findings. Summarizes insights behind key findings.*

# Problem

*Clear description of the problem, from an application and theoretical point of view. Outlines the report.*

# Lit. Review

*Discusses and cites existing works in the theoretical and application realm.*

# Methods

*Discusses types of variables, sample size, and sampling techniques (if any). Discusses the model(s) and its assumptions and limitations.*

## Logistic

## SVM

## 

## Decision Trees

## 

## Random Forest

# Data

*Discusses how data was handled, i.e. cleaned and preprocessed. Discusses distributions, correlations, etc.*

Our dataset included 15 variables and 500 observations, with no missing values or NA's. While we observed a number of zeroes, these were correlated with the acquisition column, which indicates whether a prospect was acquired. As part of the data cleaning process, we looked into the correlations (graph below) among these variables and removed those with the highest correlation to reduce multicollinearity. Also, we removed one of the customer ID fields. The variables removed from the dataset include *customer, duration, profit, ret\_exp, ret\_exp\_sq, freq, freq\_sq, crossbuy, and sow*. No additional data cleaning was necessary.

## Correlation

Talk about the variables that we removed for this analysis

A blue and white squares with red text

Description automatically generated

# Results

*Presents and discusses the results from model(s). Discusses relationships between covariates and response, if possible, and provides deep insights behind relationships in the context of the application.*

# Conclusions

*Concludes with a summary of the aim and results. Discusses alternative methods that can be used.*