Fall Term 2

DATA SCIENCE FOR BUSINESS

Team 30

ifood Início Restaurantes Mercados Bebidas Farmácias Pets Shopping

Busque por item ou loja

R. Fradique Coutinho, 1340

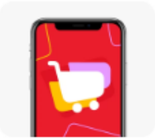
R$ 0,00
0 itens


Restaurantes


Mercado


Farmácia


Pet


Super


Bebidas


Shopping


Gourmet

Mercados próximo a você

Ver mais

super

iFood Super | Economize Aqui - Pompéia
4.5 • Mercado • 1.89 km
126-136 min • R$ 7,49

Dia

Dia Supermercado - Purpurina
4.4 • Mercado • 0.33 km
120-130 min • R$ 8,99

minuto Pão de Açúcar

Minuto Pão de Açúcar Sumarezinho
3.6 • Mercado • 0.48 km
120-130 min • R$ 7,99

Carrefour express

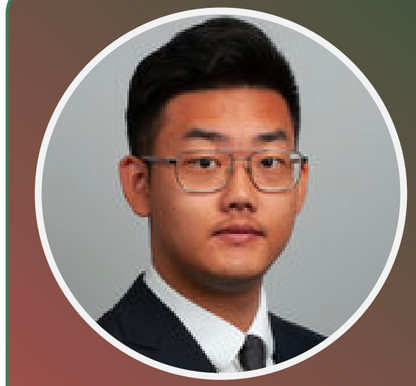
Our Team

Mahima Kriti

Director

Mandy Zhou

Manager

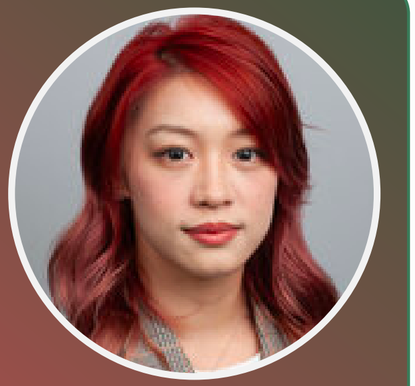


Haoming Bai

Business Head

Serena Wang

Specialist



Ella Hu

Marketing

Content

Overview

ifood

01

Industry:
Food and Grocery Delivery

02

Base:
Brazil

03

Business Performance:
Solid and healthy in the past 3 years

04

Problem:
The profit growth for the next 3 years is not promising

Business Understanding

Goal

Increase the profit of the next marketing campaign

Our strategy

- Identify the customers most likely to purchase the new gadget.
- Exclude non-respondents from the campaign to minimize wasted resources.
- Understand the characteristics of customer segments more likely to accept the ad campaign.

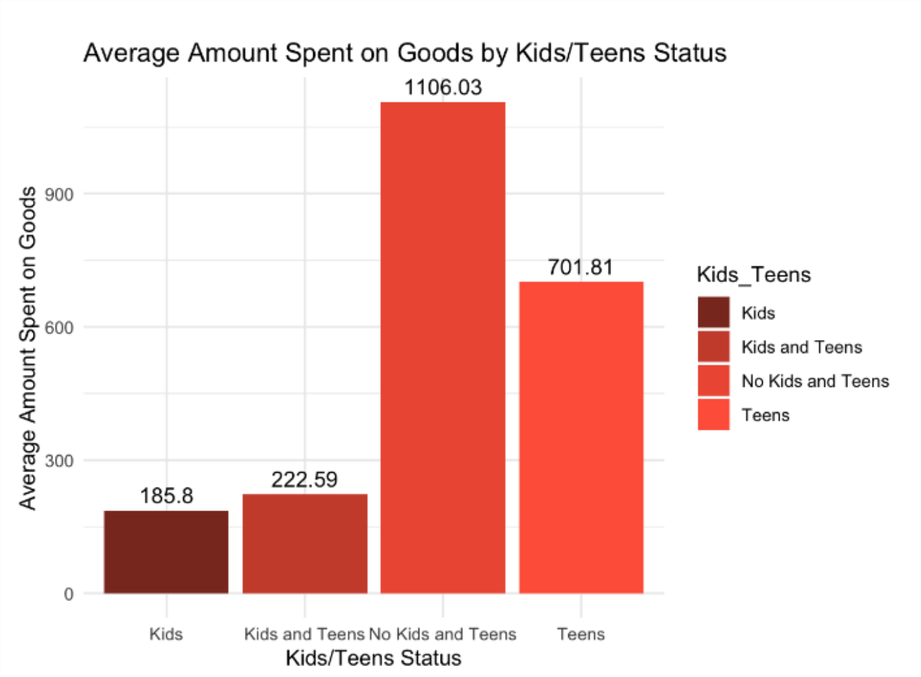
Data Understanding

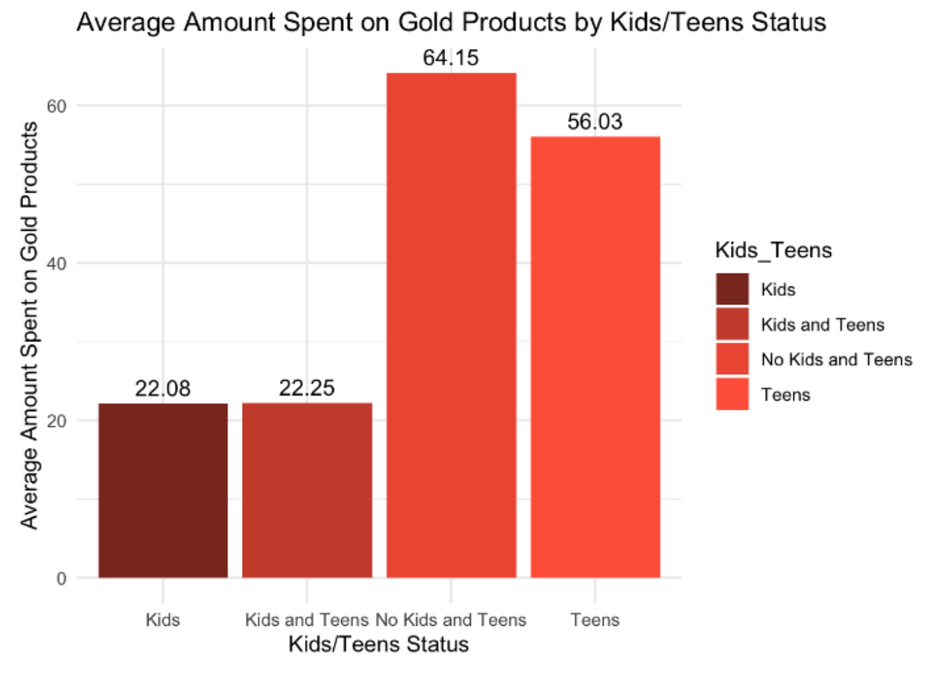
25 Original Features
2240 Observations

Types of data:

- **Demographics -** DOB, Marital Status, Education, Income
- **Activity -** Days of association, last purchase date
- **Purchases -** Purchase of meats/wines/fruits etc
- **Channels -** InStore/Catalog/Online
- **Campaign response -** Response to previous 5 campaigns (0/1)

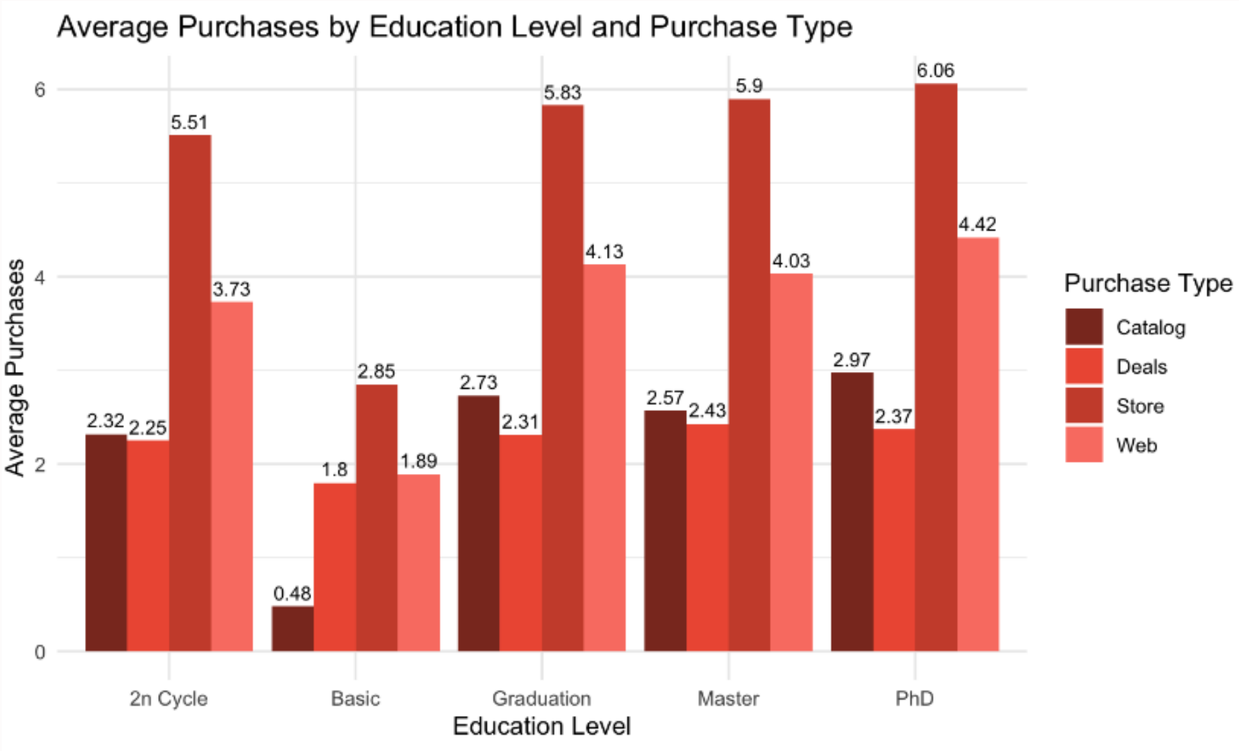
Data understanding





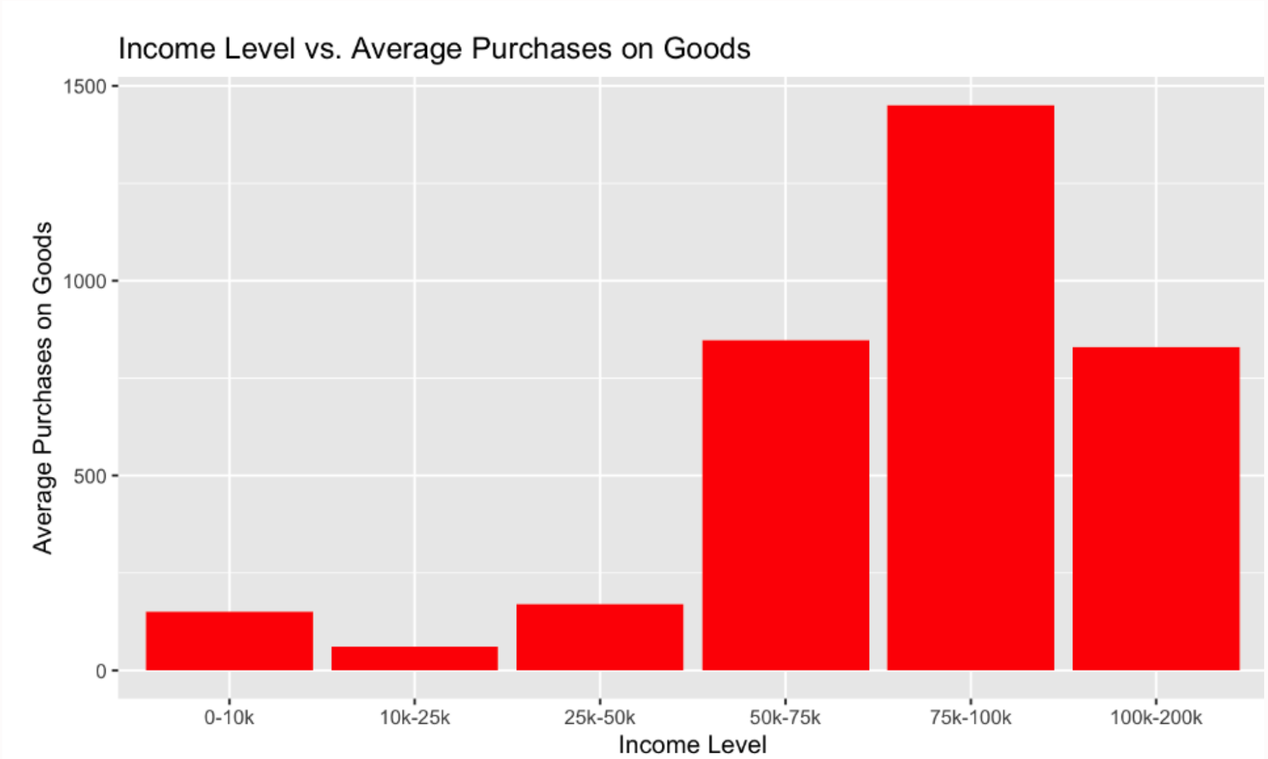
Customers who don’t have kids or teens are likely to spend more on both regular products and gold products.

Data understanding



Customers with a higher education background spent more on all purchase types.

Data understanding



The income group of 75-100k and above spent the most on goods.

Customer Segmentation

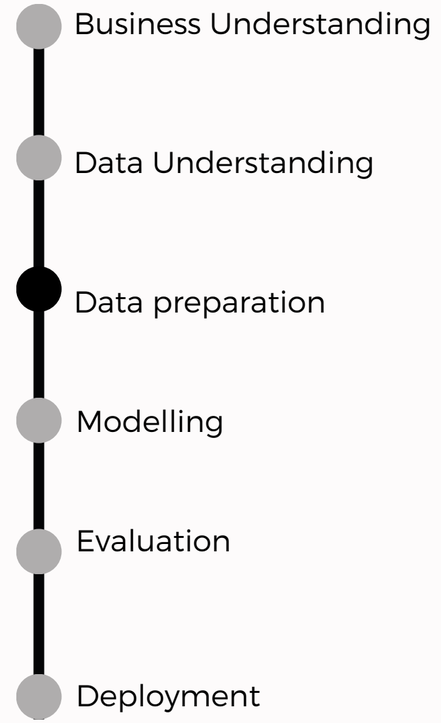
Target Audience #1	*Target Audience #2*	*Target Audience #3*
High income group (>75k) • *Great engagement with all campaigns* • *Dont use deals* • *Buy from all 3 sales channels*	Hoseholds with no kids/teens • *highest spending on Regular and Gold products*	Higly educated people • *Most purchases in store*

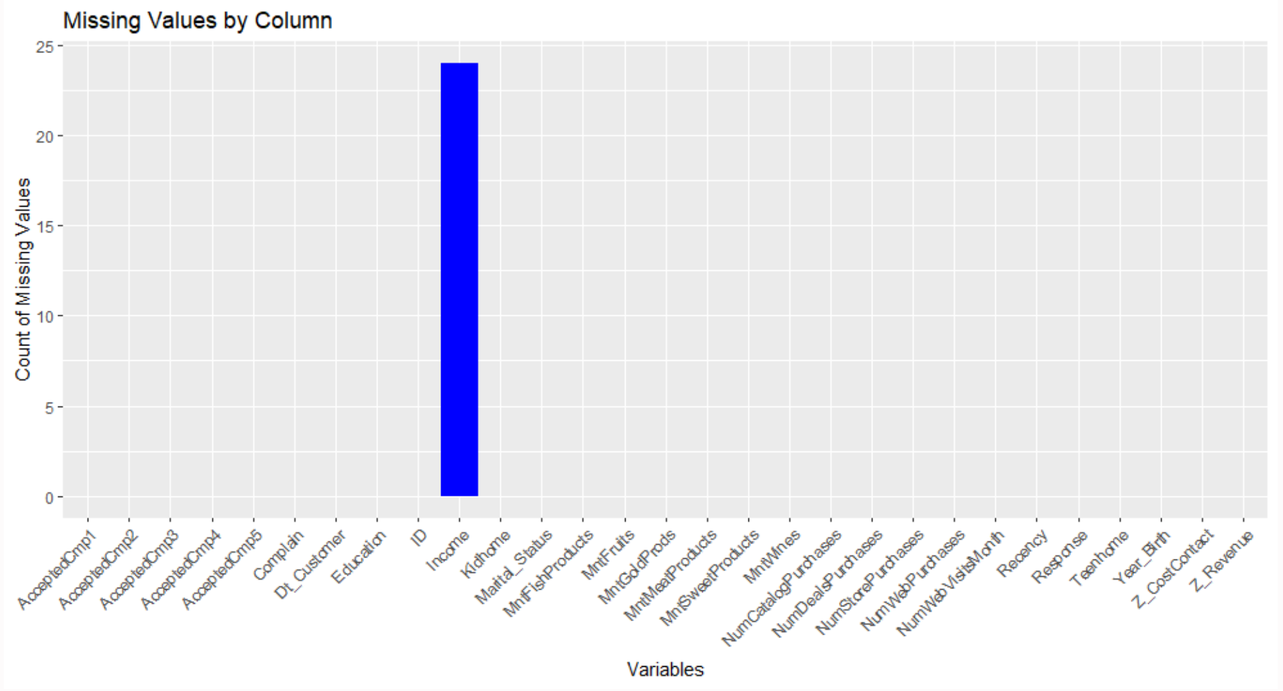
Data Preparation

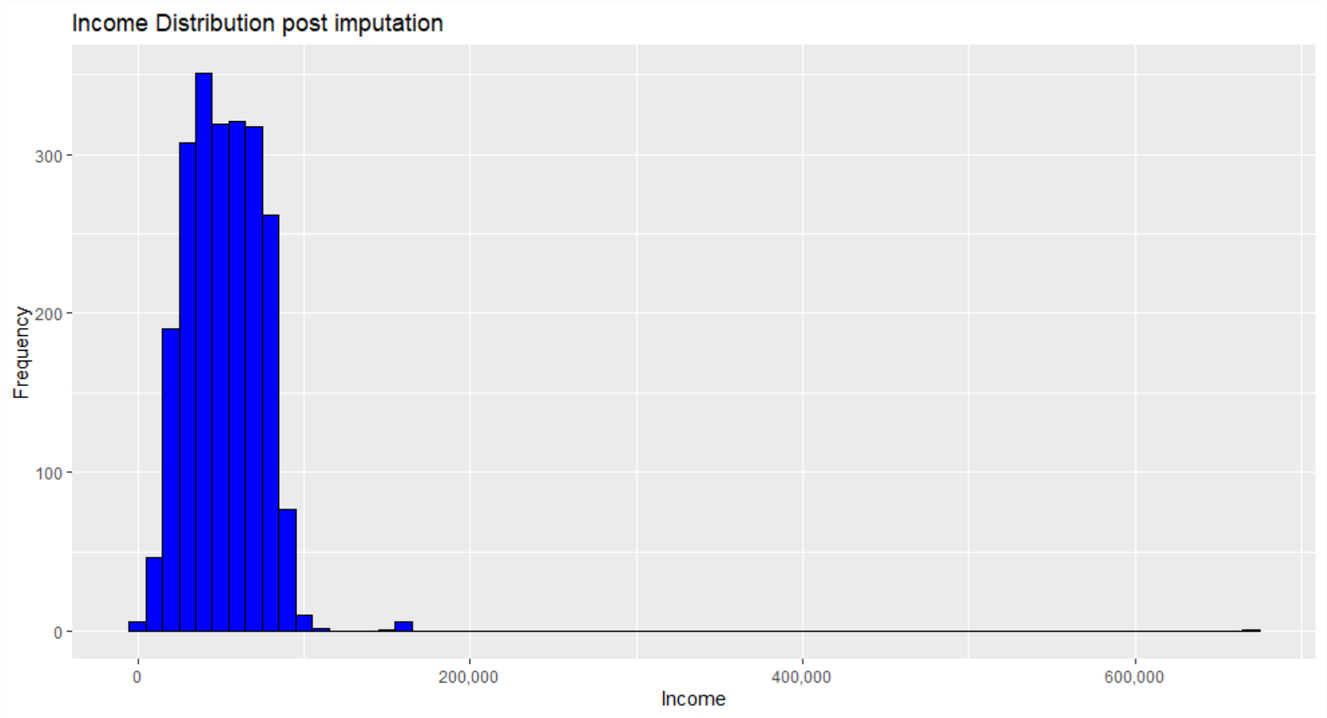
- **categorical columns**

1. Education- 6 different education categories, we converted into dummy variables
2. Marital status – The original dataset had 6 different categories of marital status we converted them to three categories – Single, Married and Together

- New fields created – Age, Days_until
- Missing values – Income had 24 NA values, they were imputed by simple linear regression
- Dropping columns – Z_Revenue(column of 3s), Z_Cost, Dt_Customer, ID





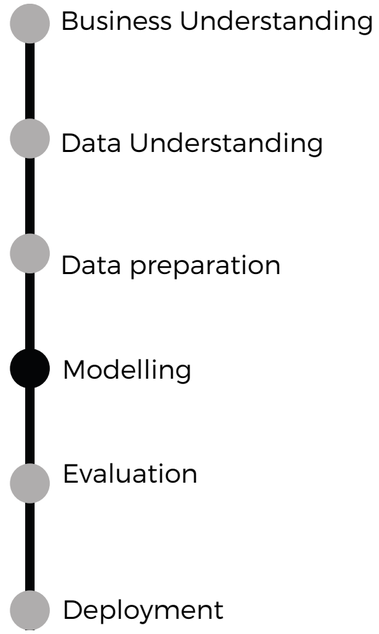


- Business Understanding
- Data Understanding
- **Data preparation**
- Modelling
- Evaluation
- Deployment

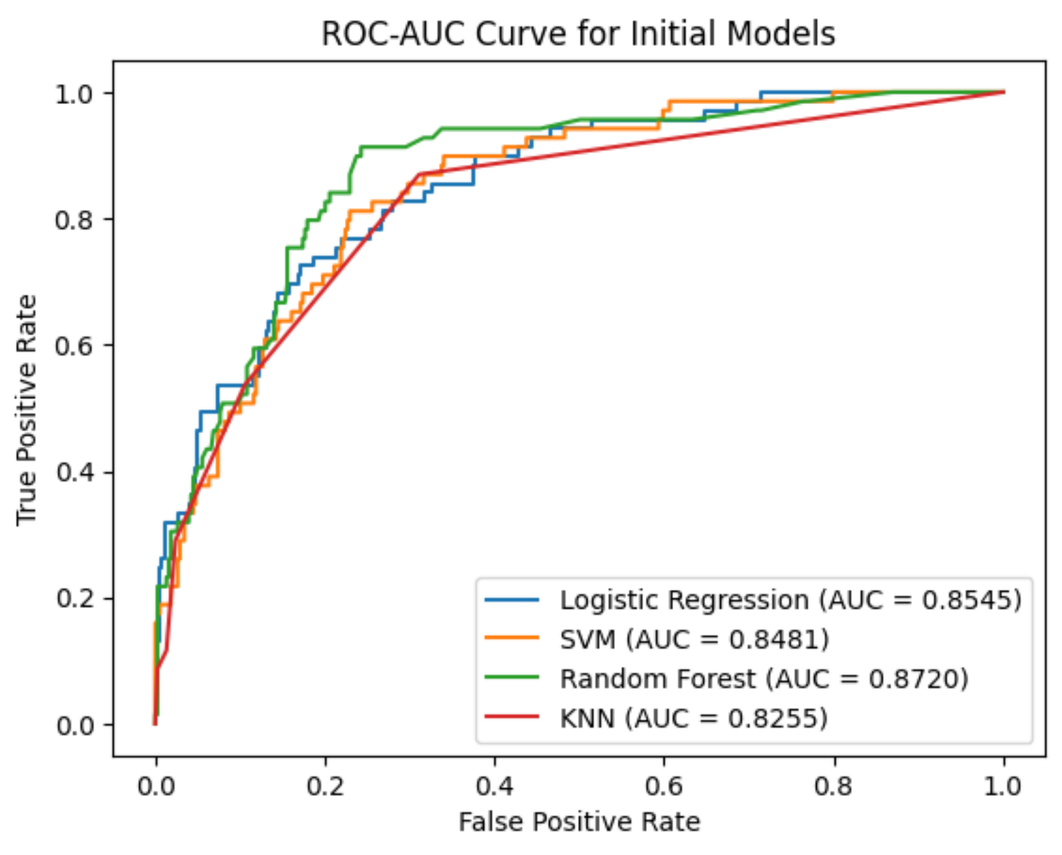
Modelling

4 Different classification models

1) Logistic Regression
2) Random Forest
3) Support Vector Machine
4) k-NN

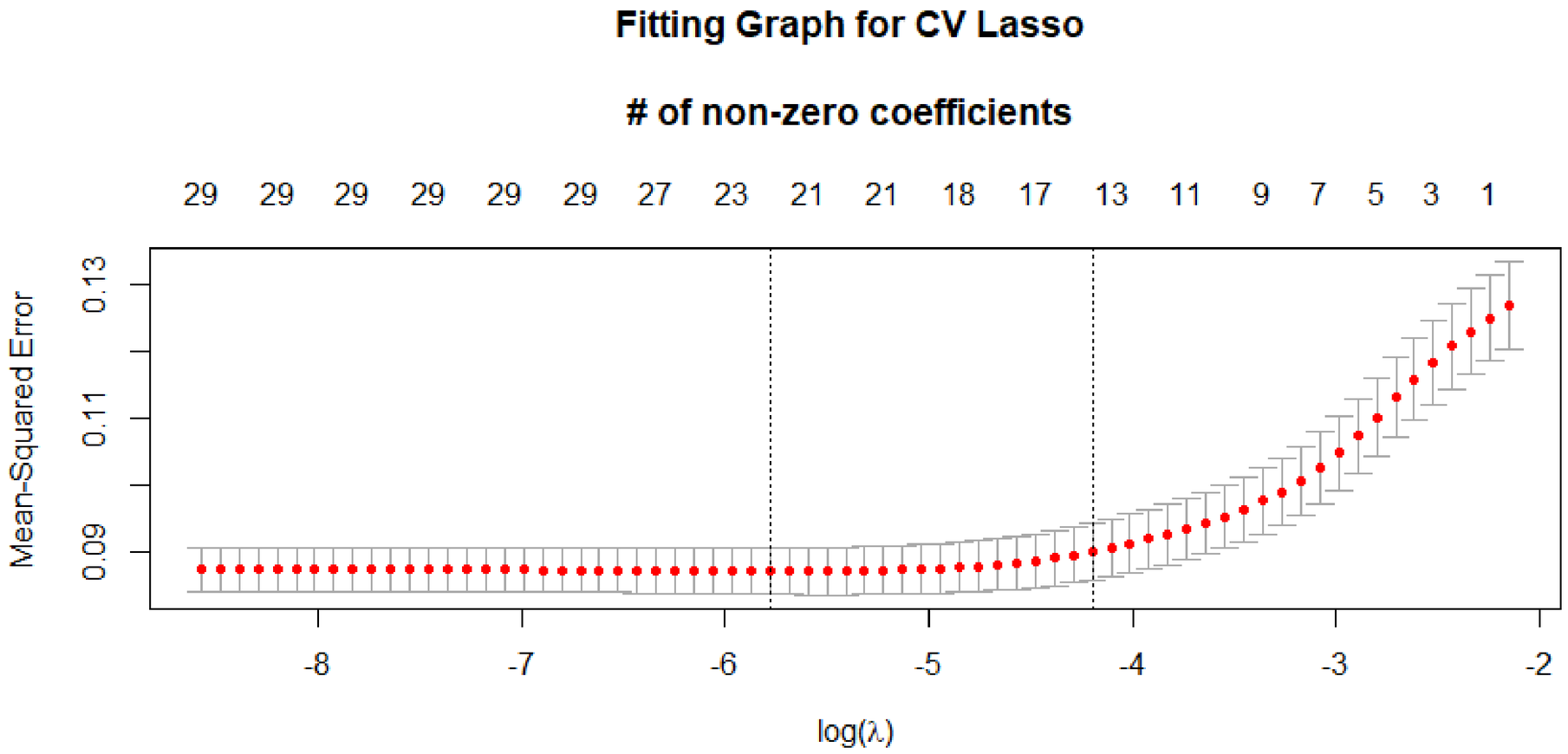


ROC-AUC curve for different models



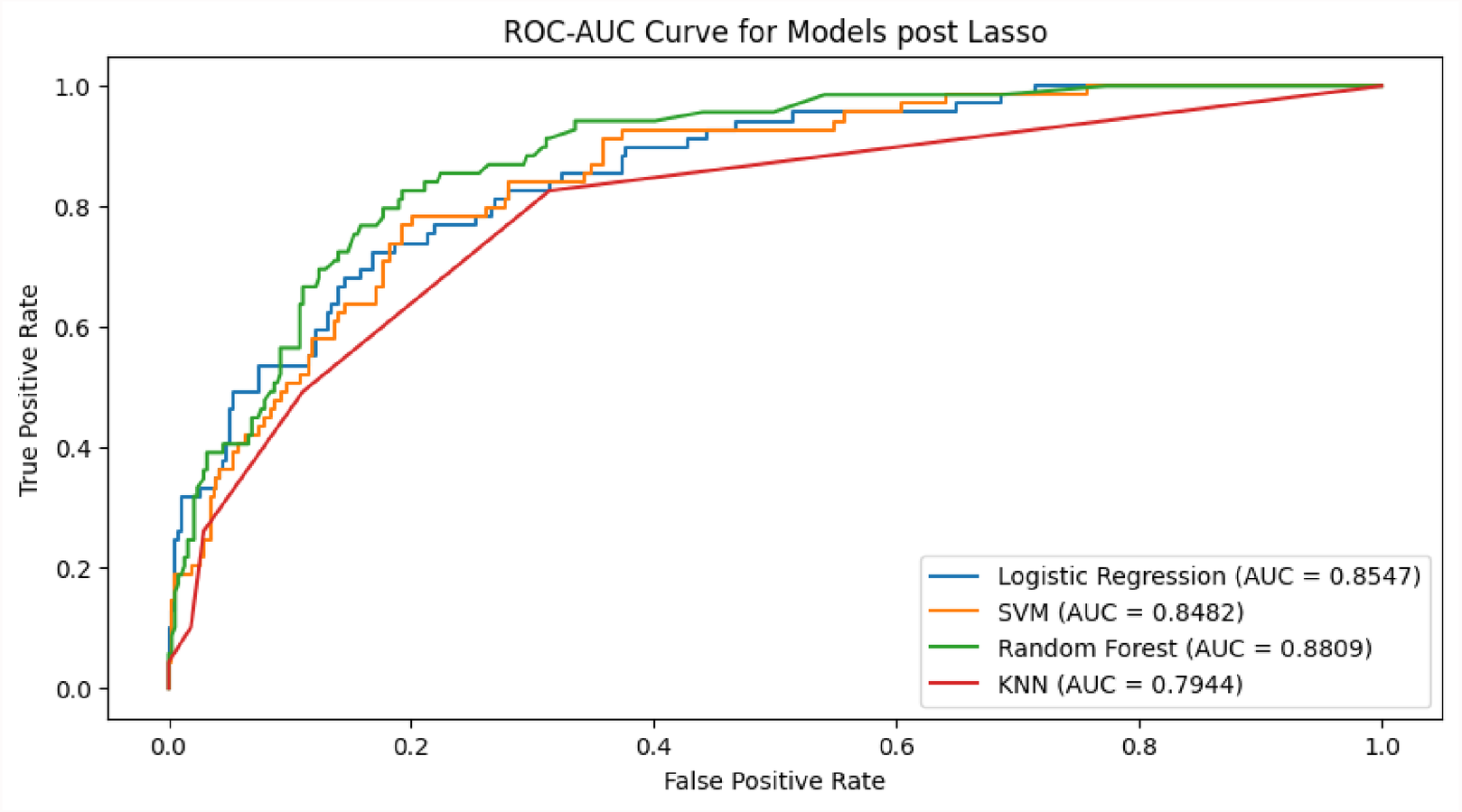
Random Forest performs the best among all 4 models

Modelling with feature selection using Lasso



- Business Understanding
- Data Understanding
- Data preparation
- **Modelling**
- Evaluation
- Deployment

ROC-AUC curve post Lasso for different models

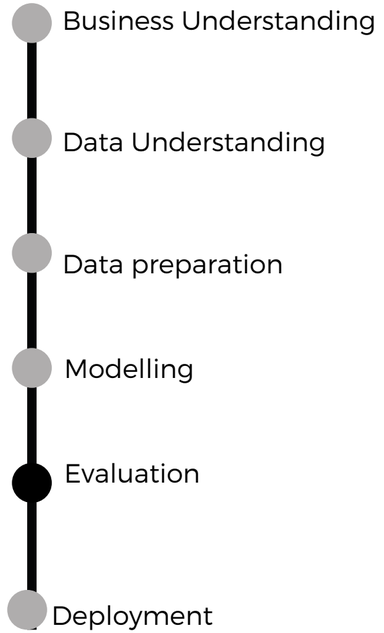


Random Forest still performs the best among all 4 models with 0.18 as the optimum threshold.

Evaluation

Evaluation metric - AUC

1) classification problem
2) assess the trade-off between True Positive Rate and False Positive Rate at different threshold
3)helps pick one threshold for your classification



Deployment

- Anticipate the positive response likelihood of prospective customers to a marketing campaign.
- Target specific customers whose predicted response is negative

Risks

- Class imbalance for the 6th marketing campaign
- Customer behavior changes with time
- Dataset for prediction may not be representative of population

THANK YOU