

Customer Churn Prediction

Capstone Project - April 2023

Ryan Stamp - Data Scientist



Agenda

- Bio
- Industry Overview
- Why are our customers leaving?
- Key Stakeholders
- Can we predict the churn of future customers
- Current Data
- EDA
- Data Science Process
- Feature Engineering
- Modelling and Interpretation
- Summary and Next Steps

Bio

- Ryan Stamp
- Data Science Student
- Sales and Marketing Professional with more than 10 years experience
- Model Predictions and Machine Learning
- Lives in Auckland, New Zealand
- Half-marathon finisher and home-baker



Industry Overview

Financial Advisory Services industry valued at \$79.4billion (2020) and is predicted to be worth \$135.6billion by 2030.

North America sits as the largest regional market for this and is predicted to remain so during the forecasted period.

Why are our customers leaving?

Key Stakeholders

What value will this project bring?

- Lower customer acquisition costs (CAC)
- Reliable future forecasting

Can we predict the churn of future customers?

Data required

- customer transactions
- product details
- customer segments
- customer help desk records
- customer feedback and net promoter scores



Current Data

R Stamp April 2023

Kaggle Dataset

Four Dataframes

508932 records

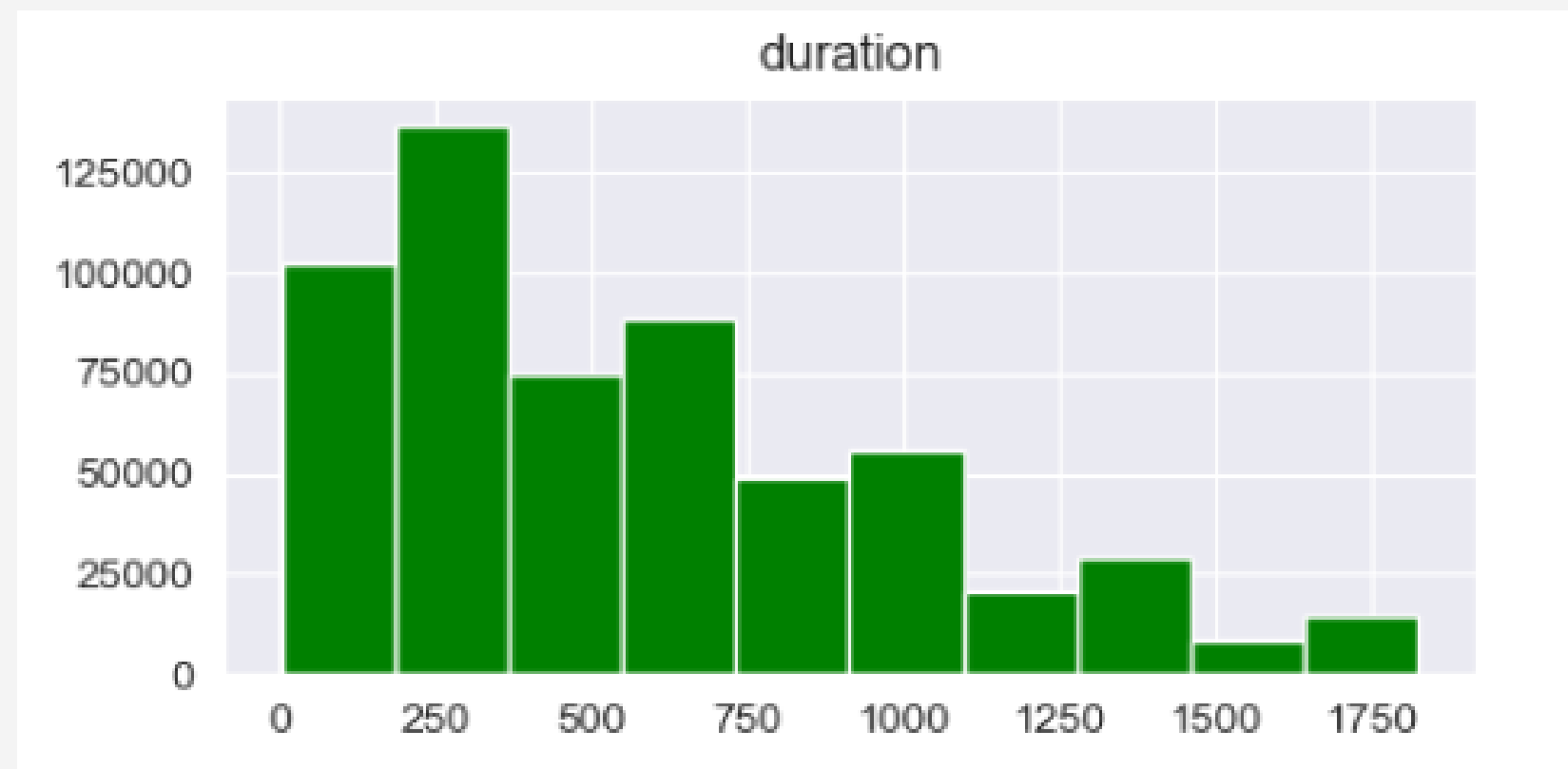
5 Years

Current Sources

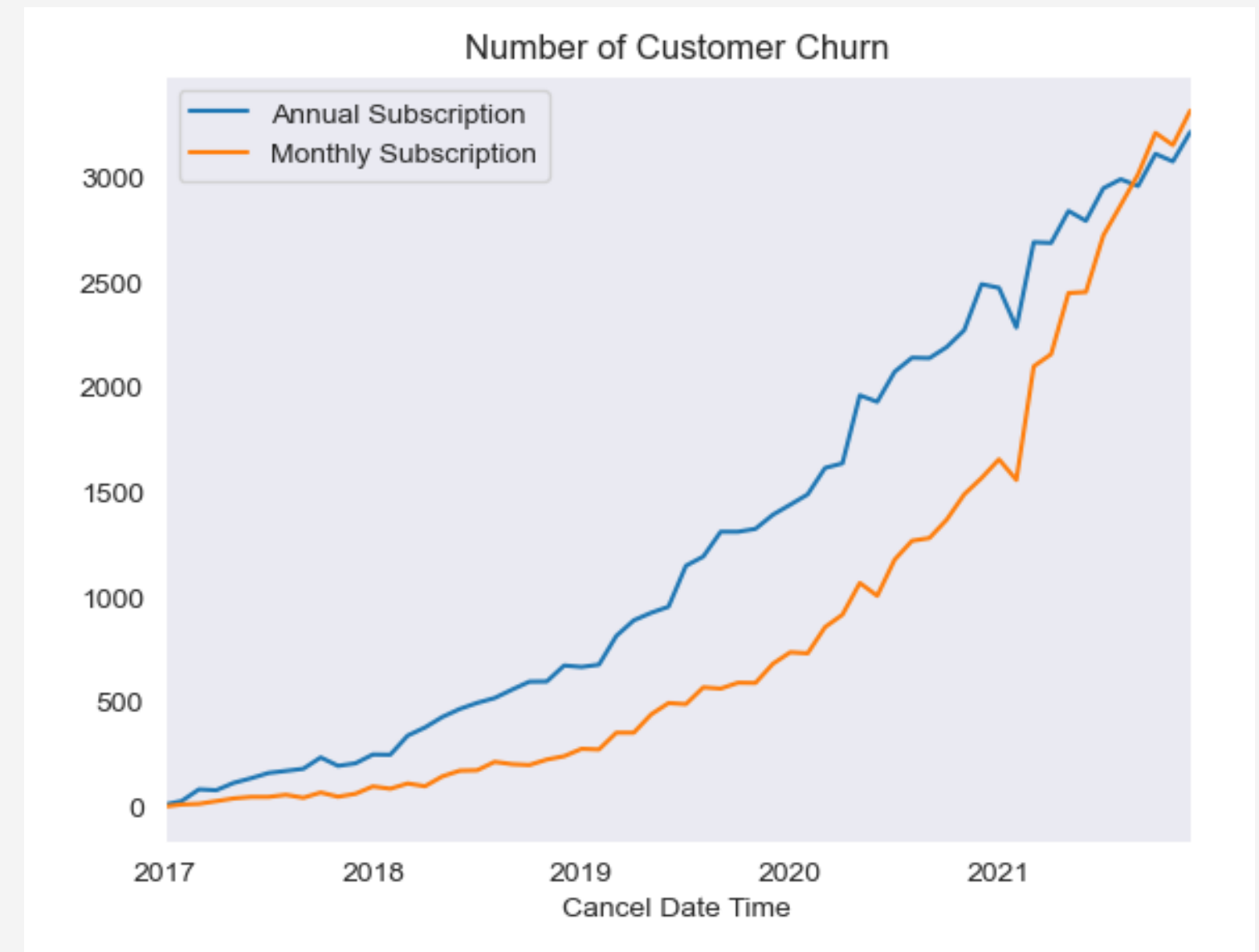
Future Sources

Exploratory Data Analysis

R Stamp April 2023



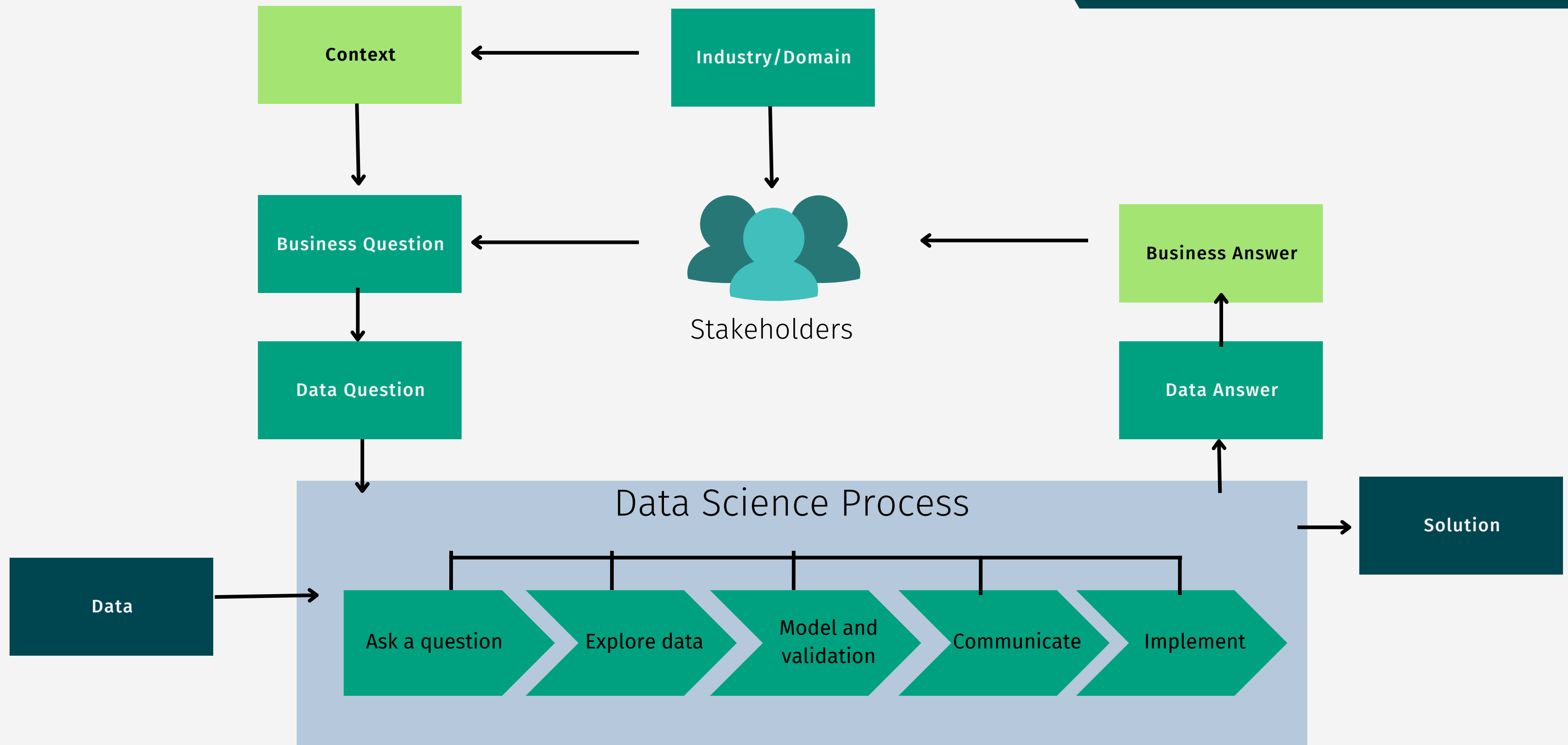
Distribution of subscription length (in days)



Rate of subscription cancellations

Data Science Process

R Stamp April 2023



Feature Engineering



The features we will use are:

-Churn (Target Variable)

- Duration
- Name
- Price
- Billing_cycle
- Agerange
- Gender
- Channel
- Reason

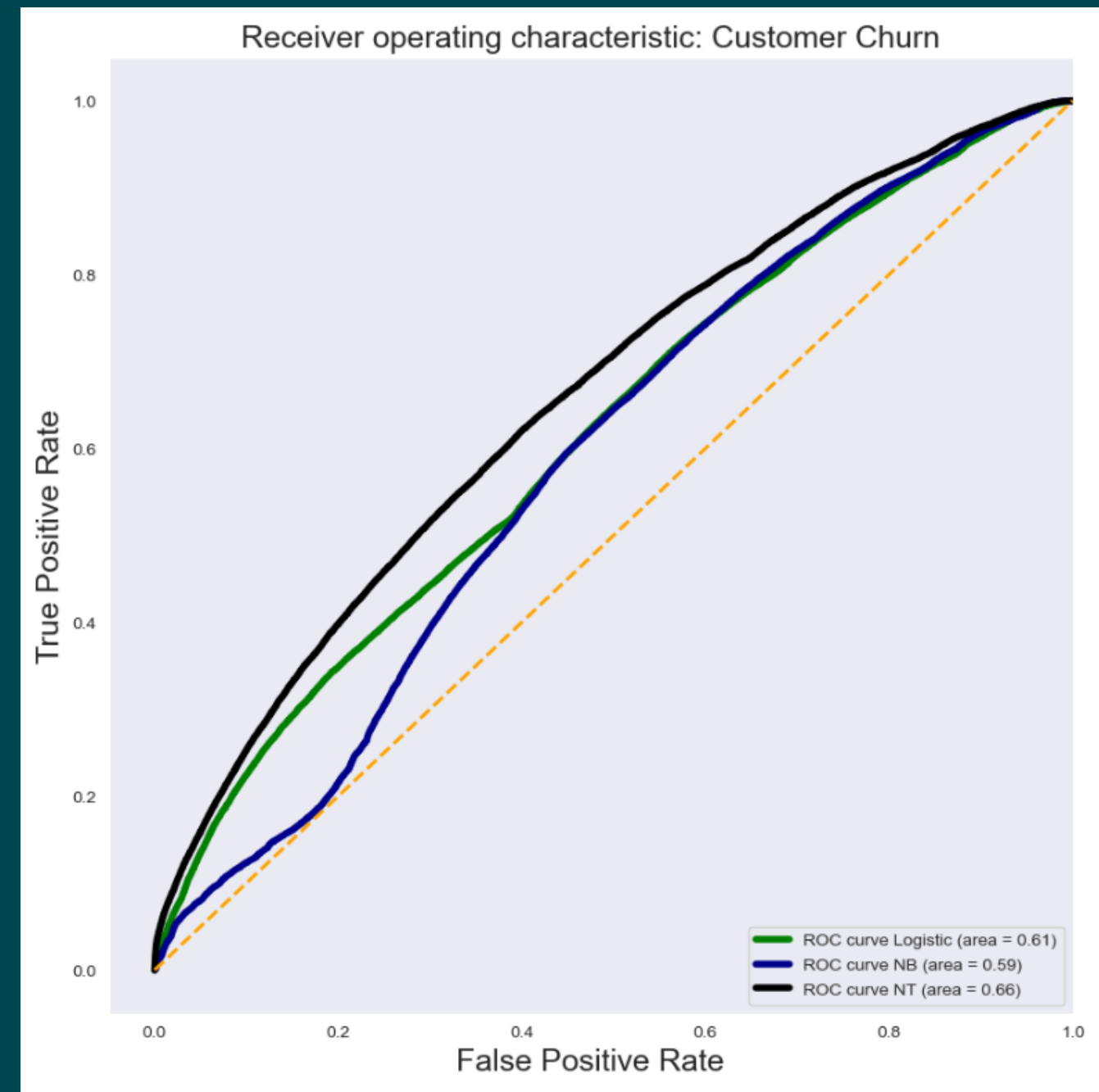
Modelling and Interpretation

Model	Precision	Recall	Accuracy
Multi-Layer Perceptron	0.063704	0.650615	0.116046
Logistic Regression	0.000000 (1.000000 for alternate target)	0.000000 (0.777870 for alternate target)	0.000000 (0.875059 for alternate target)
Naive Bayes	0.000000 (1.000000 for alternate target)	0.000000 (0.777870 for alternate target)	0.000000 (0.875059 for alternate target)

MLP model provides the strongest results yet not recommended to deploy (recall of .65).

Modelling

Receiver Operator Characteristic Curve - Customer Churn



Did we provide a solution?

- Value to stakeholders
- When and how can this project add value?

Summary and next steps

- Project Overview - data and modelling strengths and shortcomings

- What outcomes did we achieve?
Project is not ready to deploy

- Next steps?
Consult with stakeholders and data collection team

Thank you.

References

<https://www.alliedmarketresearch.com/financial-advisory-services-market-A06946>

<https://www.kaggle.com/datasets/gsagar12/dspp1>

<https://www.paddle.com/resources/customer-churn-analysis>

<https://developers.google.com/machine-learning/data-prep/construct/sampling-splitting/imbalanced-data#:~:text=A%20classification%20data%20set%20with,smaller%20proportion%20are%20minority%20classes.>