CAPSTONE PROJECT

CUSTOMER CHURN PREDICTION

Contents

- Business Problem Overview & Solution Approach
- Data Overview
- Features Extracted
- Models Built
- Best Model Performance
- Important Features
- Churn vs Important Features
- Value Calculation
- Scope for Further Analysis

Business Problem Overview & Solution Approach

- A Direct to Home (DTH) service provider is facing challenges to retain existing customers
- One account can have multiple customers tagged to it
- Account churn is the major problem that the company is facing

- Need to build a prediction model to identify potential customers
- Provide insights to the company to minimize the churn
- Provide business recommendations to the company on segmented offers

Data Overview

- The given data includes 11260 rows and 19 features
- All unique values in Account ID
- 16.8% of churners and 83.2% of non-churners were present in the data
- Data entry error in Login_device treated as 'Other' category
- Missing values in all features except Account ID and Churn
- Outliers present in numerical features
- Data was divided into Train, Validation and Test sets for model building

Features Extracted

- From the exploratory analysis, it was seen that Male customers churned more
  - Eleven new interaction variables were extracted from Marital_Status, Payment and Login_device with respect to the customer's Gender
- Churn pattern varied for Account_user_count with respect to Service_Score
  - One ratio variable was created
- Total of 29 features

Models Built

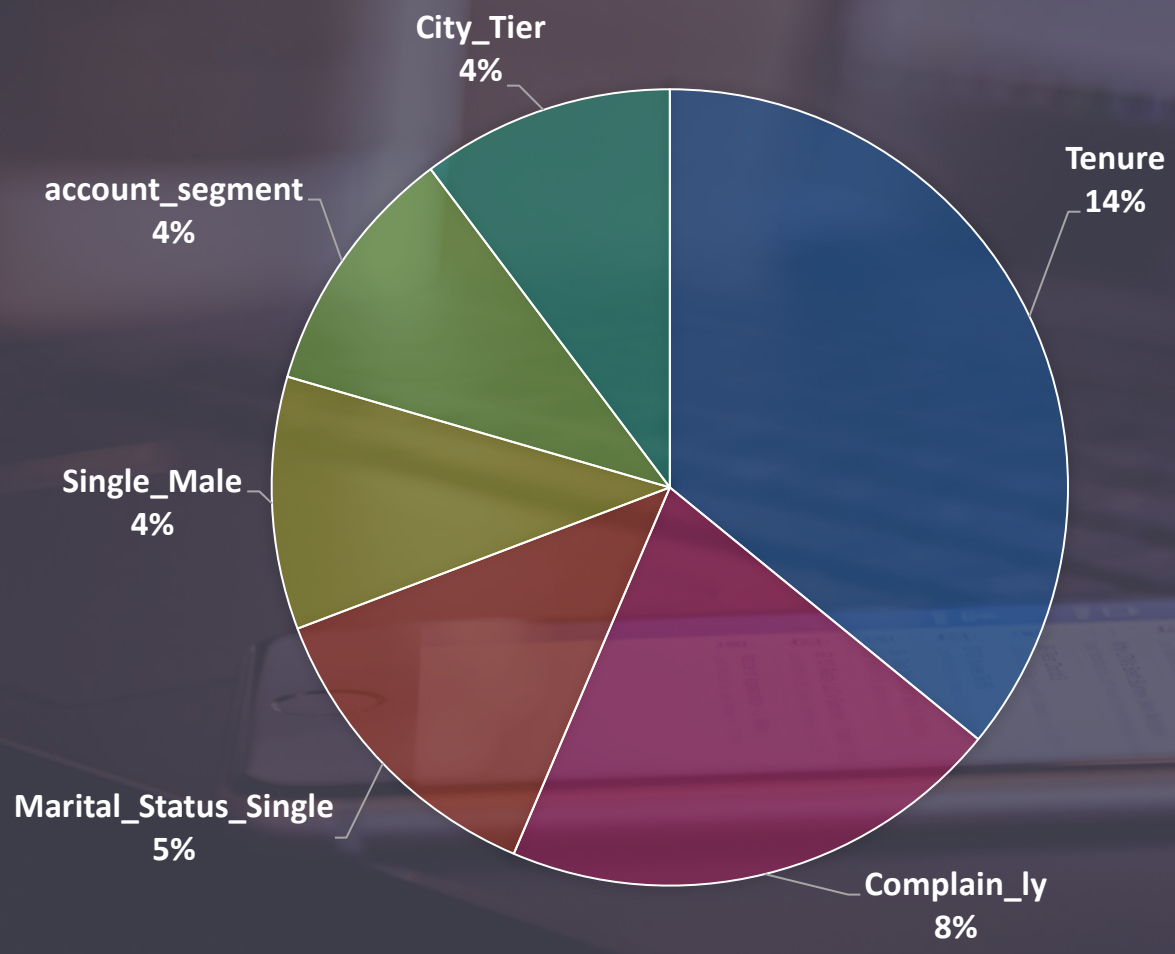
- Models were identified with a cross-validation technique
- 10 individual models were built including 5 tuned models
- Top models are:
  - Bagging model with base Tuned Decision Tree
  - Tuned Bagging model with base Tuned Decision Tree
    - 84% of people were correctly predicted to churn
    - Only 61% of people were correctly predicted from total predicted churners
- 4 stacking models were built with combinations of previously created models
  - Stacking4

Best Model Performance

Data / Metric	Accuracy	Recall	Precision	F1-Score
Validation	0.97	0.95	0.90	0.92
Test	0.98	0.96	0.91	0.93

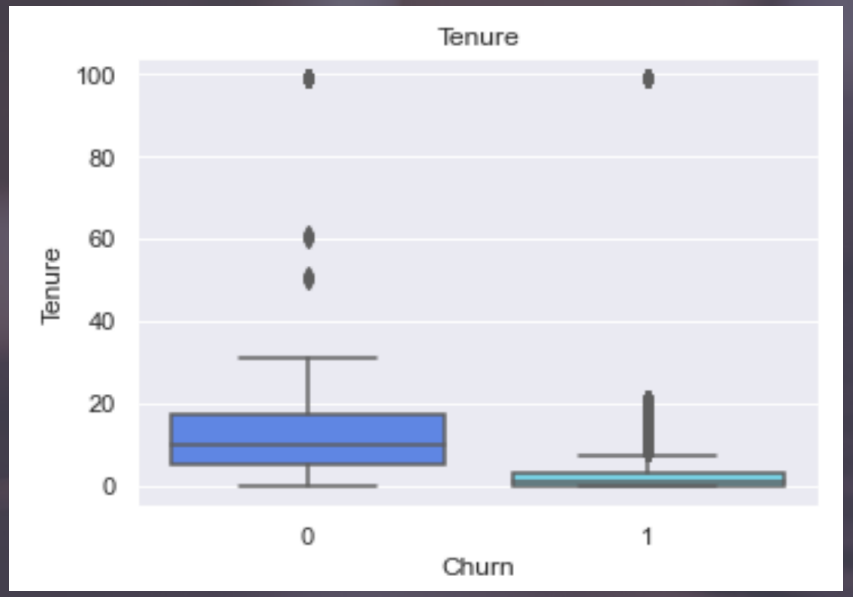
- Stacking4 is the best model
- The model is giving a generalized performance, with even better results in the test data

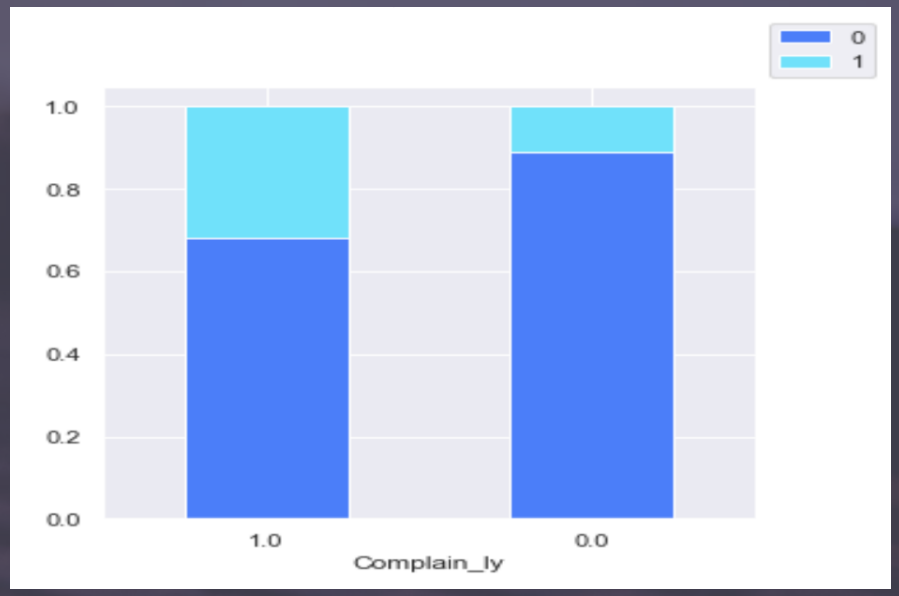
Important Features



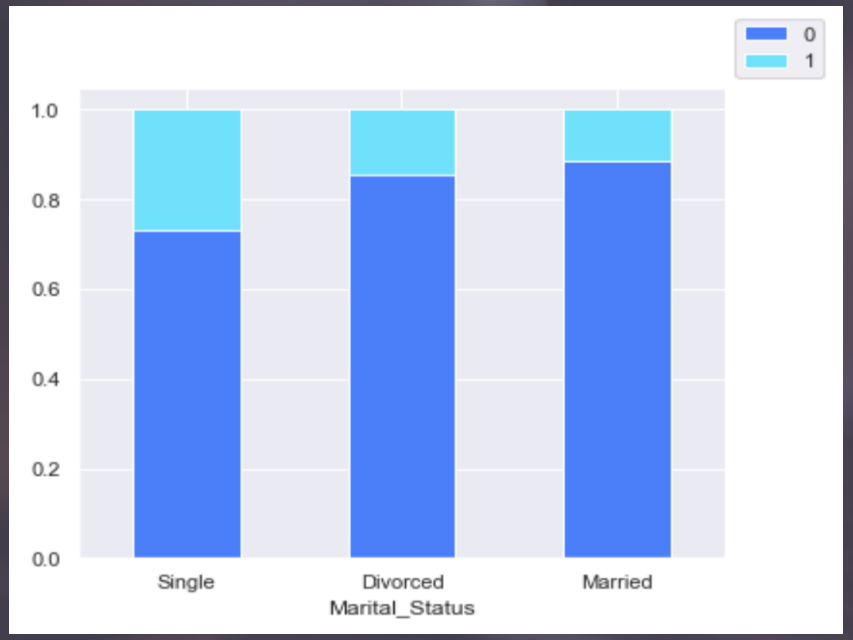
- All features were given some significant importance in model building
- Top 6 features are shown

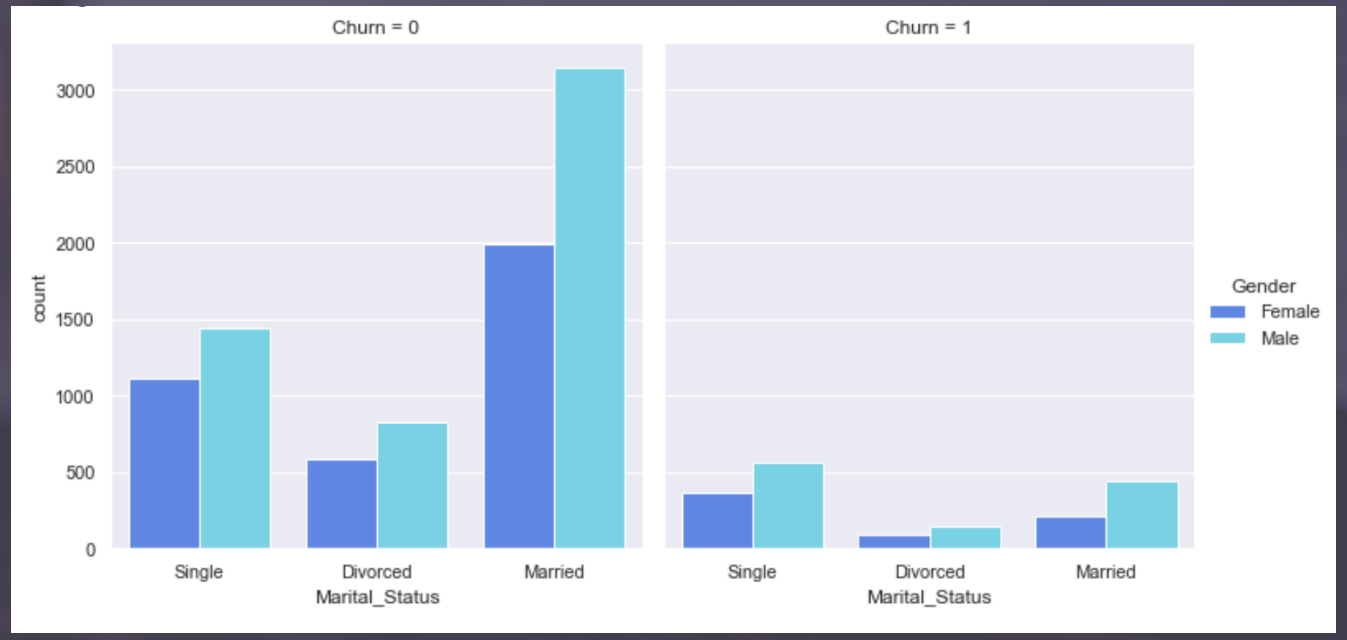
Churn vs Important Features



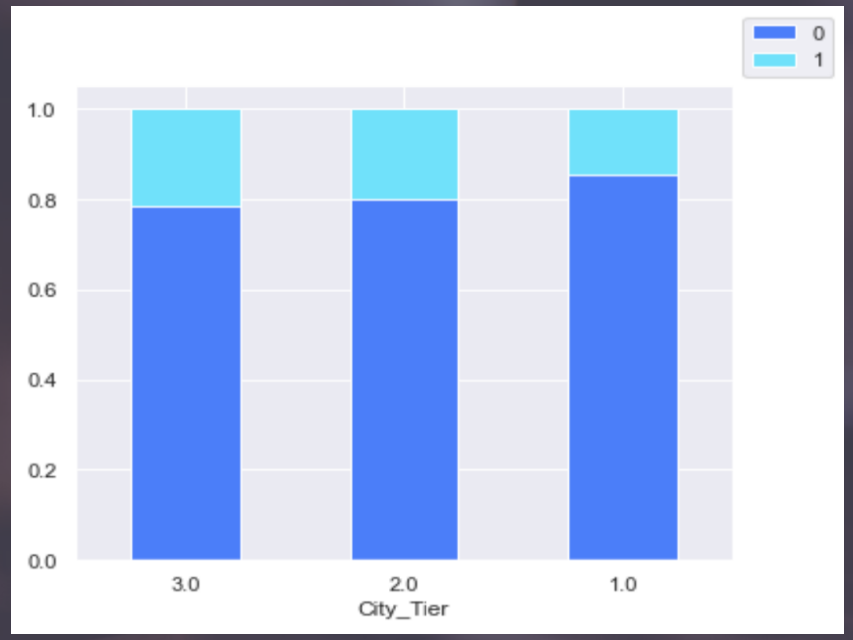


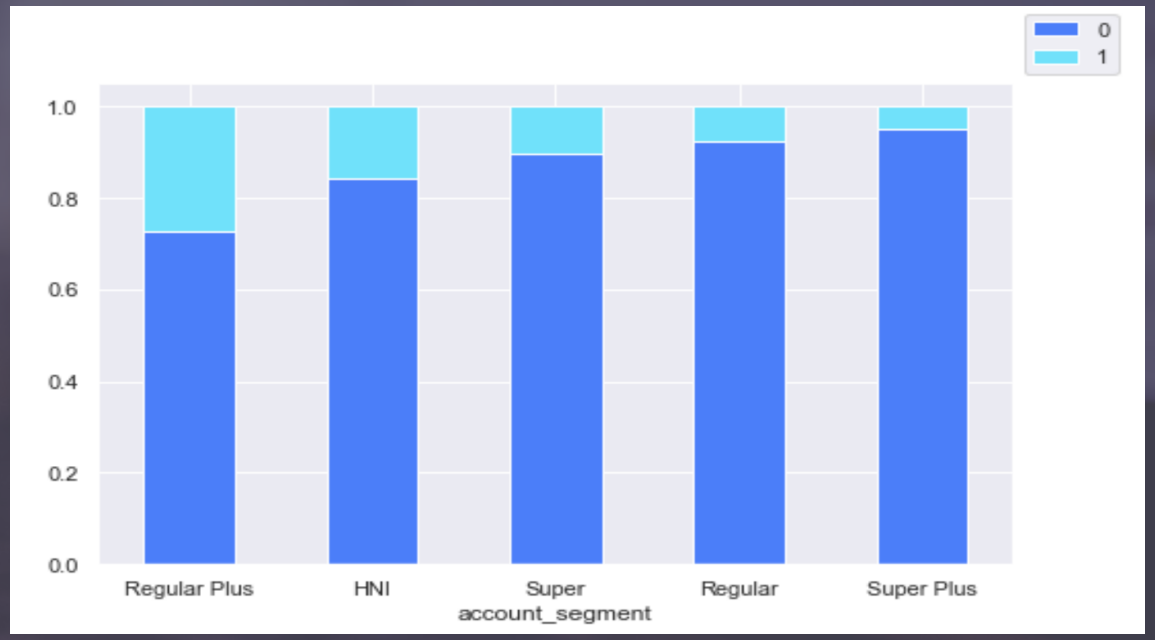
- Customers with lower Tenure are the ones who churned more
- Higher the number of complaints raised by a customer, higher the chances for the customer to churn





- Single customers show higher churn rate
- Male customers churn more regardless of their marital status





- Customers from City_Tier3 have churned more
- Customers from Regular Plus account segment have churned the most

Value Calculation

- Best model predictions on 2252 observations in the test data include:
  - 364 customers correctly predicted as churners
  - 15 customers wrongly predicted as non-churners
  - 38 customers wrongly predicted as churners
- Assuming a profit/loss of $50 and an operational cost of $5, the overall profit made will be $17,260 with a profit/customer of $7.66
- If this model were to make predictions on 20K customers in future, the Net Profit made by the company would be an estimate of $153,200

Scope for Further Analysis

- Data such as Age and average monthly income of customer would provide better understanding of customer needs
- Data on average usage of the DTH services and preference of genres may help improve the marketing strategy
- Models can be built separately for each category of account segments to make efficient predictions


THANK YOU