Predicting and preventing customer churn

SAROJ RAUT & SOUKHNA WADE

BELLEVUE UNIVERSITY

Introduction and Background

Customer attrition means that a customer stops his or her business relationship with a company. It is the loss of customers by a business. Businesses classify a customer as churned(inactive) after a certain period when the customer is not active or make any transaction from the company. By losing customers, the business loses both revenue and the marketing costs involved in replacing them with new customers.

Problem Statement

The purpose of this project is to develop an AI model for looking into customer activity and pattern to provide customers that are at risk of likely to leave soon. The model provides customers that are at risk or will likely leave soon by looking into customer activity and pattern. The model also provides LOB an ability to utilize model output data for identifying merchants at risk, finding microsegments of merchants at risk, and develop corrective action to lower attrition.

Scope

2018 Churn modeling data from Kaggle site will be used for the prediction of customers at a Bank.

Perform exploratory data analysis.

Data will be cleansed before model selection.

Identify and visualize which factors contribute to customer churn.

Predict the probability of a customer who is likely to attrite using machine learning techniques

Classify if a customer is going to churn or not.

Exploratory Data Analysis

- Goal: get a general sense of the data
  - shape, distribution, data type, summary.
- Data Discovery and Analysis using Python, R, and Visualization.
- Bottom line: it is always well worth looking at your data.

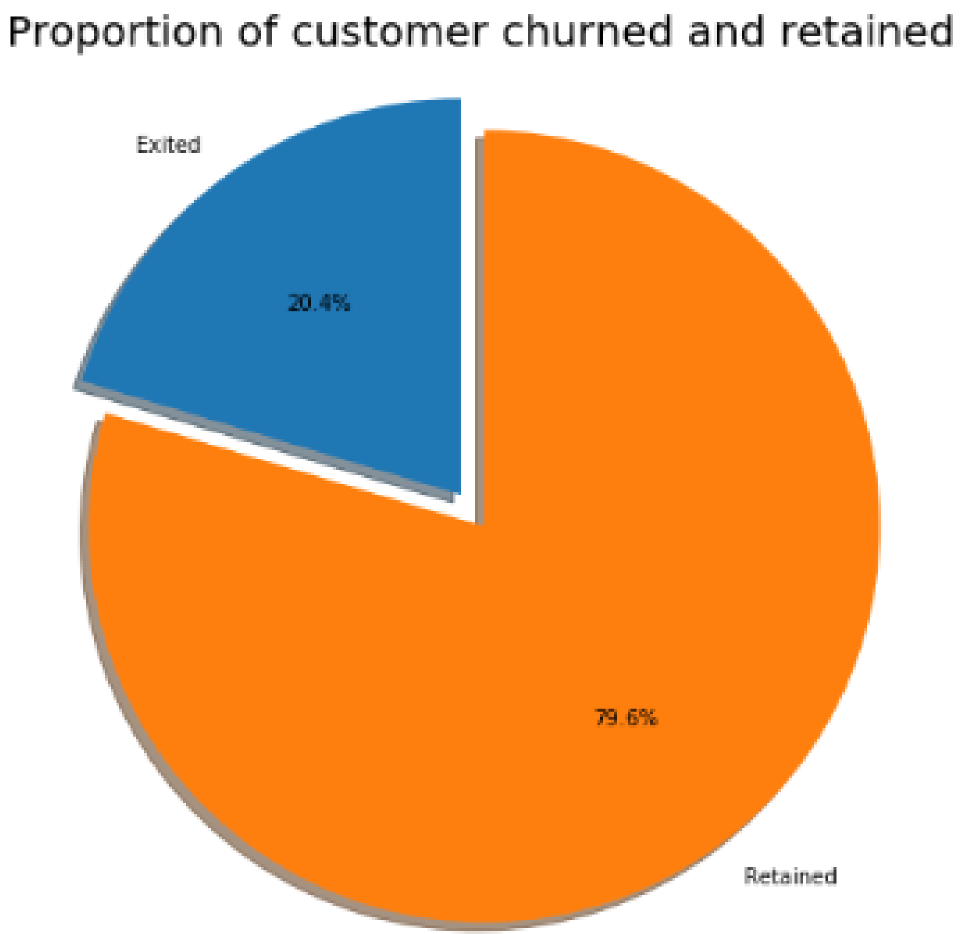
Methods

- The data set offered information on variables such as:
  - 'Status' relation with categorical variables.
  - Understand how the given attributes relate to the 'Exit' status.
  - Relations based on the continuous data attributes.
  - Arrange columns by data type for easier manipulation.
  - Data prep pipeline for test data.

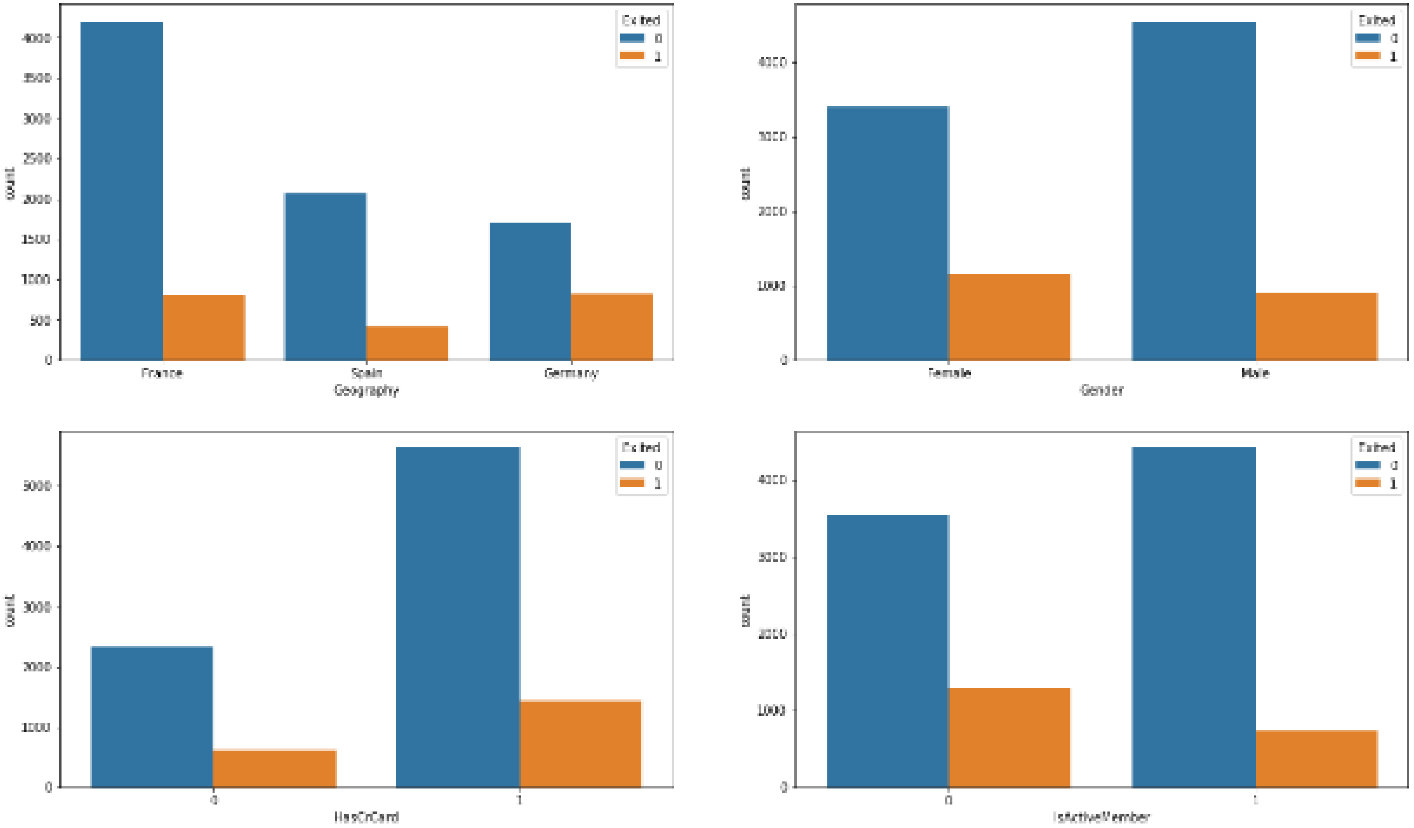
Results of data understanding

- About 20% of the customer have churned.
- Majority of the data is from France.
- There is more female customer churn than that of male customers.
- Most of the customer that churned are those with credit cards.
- Inactive member have a greater churn than the customer that are active.
- Neither the product nor the salary has a significant effect on the likelihood to churn.

Graph Analysis



Graph Analysis



Question from the Analysis

- The data appears to be a snapshot as some point in time e.g. the balance is for a given date which leaves a lot of questions:
  - What date is it and of what relevance is this date
  - Would it be possible to obtain balances over a period of time as opposed to a single date.
  - There are customers who have exited but still have a balance in their account! What would this mean? Could they have exited from a product and not the bank?
  - What does being an active member mean and are there difference degrees to it? Could it be better to provide transaction count both in terms of credits and debits to the account instead?
  - A break down to the products bought into by a customer could provide more information topping listing of product count

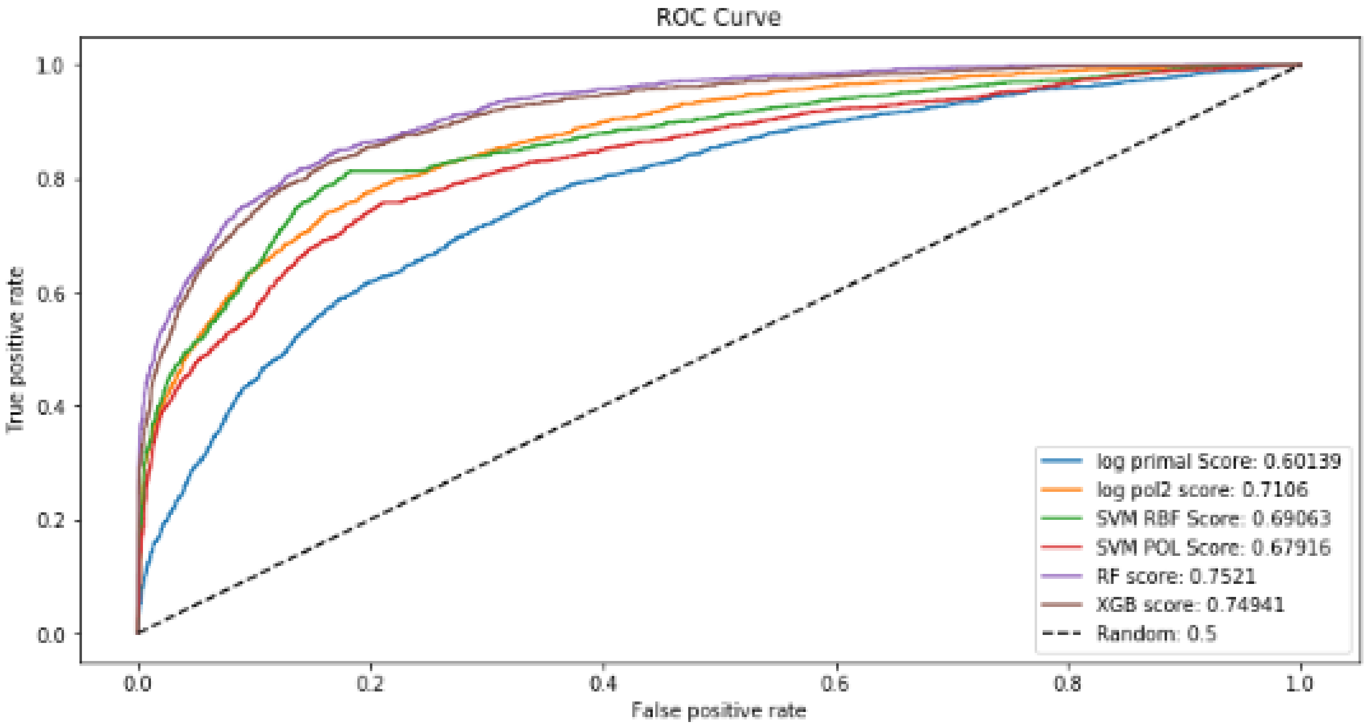
Feature Engineering

- Add features that are likely to have an impact on the probability of churning.
- Split the train and test sets
- Data prep for model fitting
  - ➢Categorical variable
  - ➢Continuous variable

Model fitting and selection

- Logistic regression
- SVM in the primal and with different Kernels
- Ensemble models
- Review best model fit accuracy
- The performance in predicting 1's (Customers who churn)

Performance measurement



Model deployment

- From the review of the fitted models above, the best model that gives a decent balance of the recall and precision is the random forest where according to the fit on the training set, with a precision score on 1's of 0.88, out of all customers that the model thinks will churn, 88% do actually churn and with the recall score of 0.53 on the 1's, the model is able to highlight 53% of all those who churned.

Conclusion

In conclusion it can be said that the Business need to update and upgrade their business as per the changing need of the market. To gain a competitive edge in the market, it is necessary to focus on areas that are identified as risk and take necessary corrective actions for it.

References

- https://blog.hubspot.com/service/what-is-customer-churn
- https://www.optimove.com/resources/learning-center/customer-churn-prediction-and-prevention
- https://www.datacamp.com/community/tutorials/predicting-employee-churn-python