Portuguese Bank:
Customers' Subscription to a Financial Product

Developer: Claudia Xu

Highlights

- An overall understanding of the dataset including features
- and target variable through various visualizations and summary statistics.
  - Most features are categorical.
- Explore potentially important variables.
  - Most customers' previously campaign outcome is success.
- Three different machine learning algorithms (logistic regression, random forest, and xgboost) to predict whether a customer churned based on selected features. Based on evaluation metrics - AUC, xgboost model was the best.

Review Progress

During this sprint, I completed the following stories:

- Stories in Epic #1:
  - EDA:
    - Overall distribution of features.
    - Distribution of features across subscribed/not- subscribed customers.
  - Feature selection.
  - Modeling.
  - Hyper-parameter tuning.
- Stories in Epic #2:
  - Choose proper number of clusters
 - Compare clusters according to features used in segmentation

Visualization

▶ Though blue collar and retired men are roughly in the same age group, retired men appear to be in the largest group who subscribe to the financial products. Hence, job does not seem to be a big factor in predicting customers' decisions
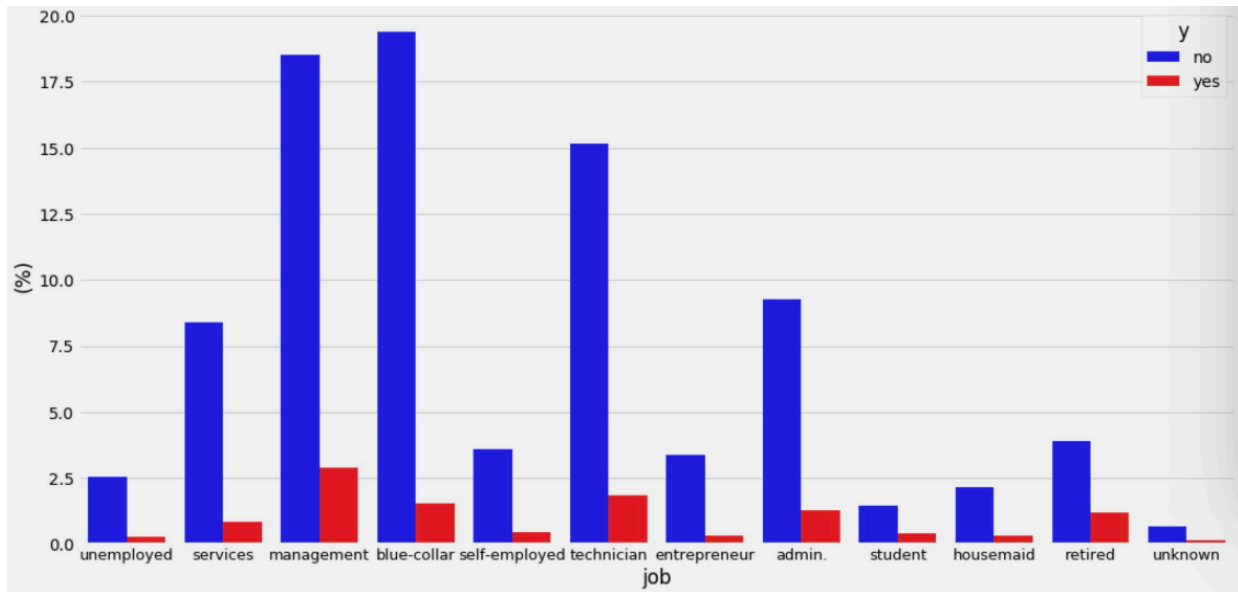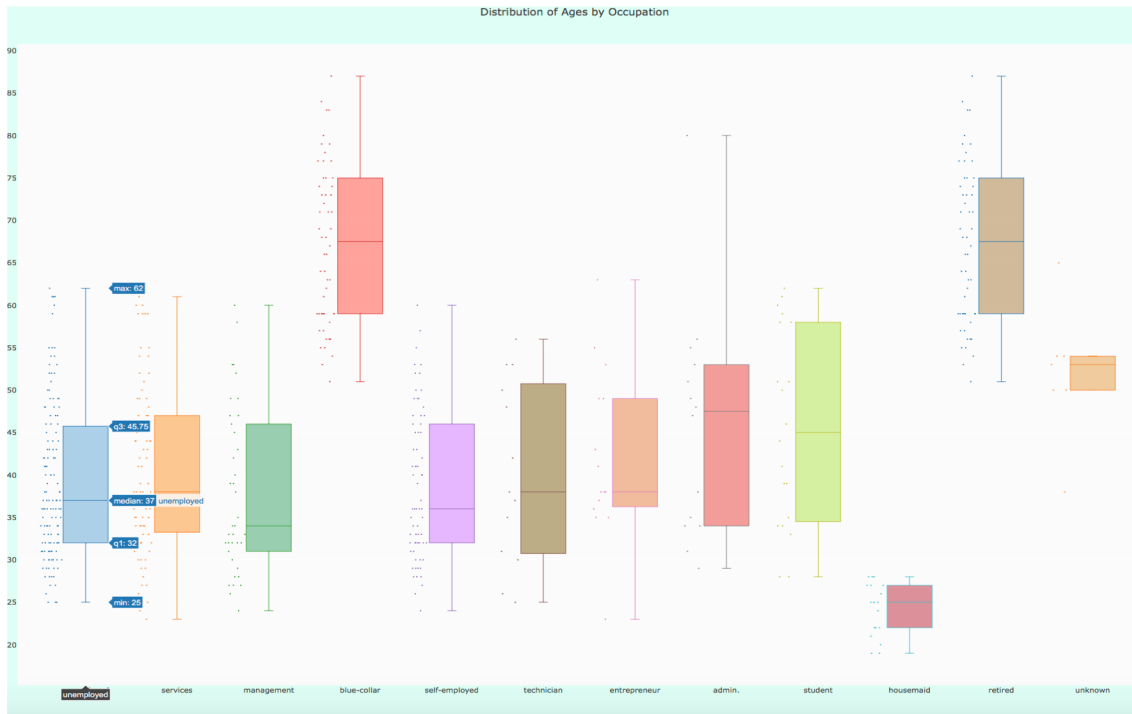
Analysis: Baseline and XGBoost Model

▶ Confusion metric indicates a relatively high number of false positive cases

▶ Accuracy score : 89.388357

▶ ROC score : 84.780000
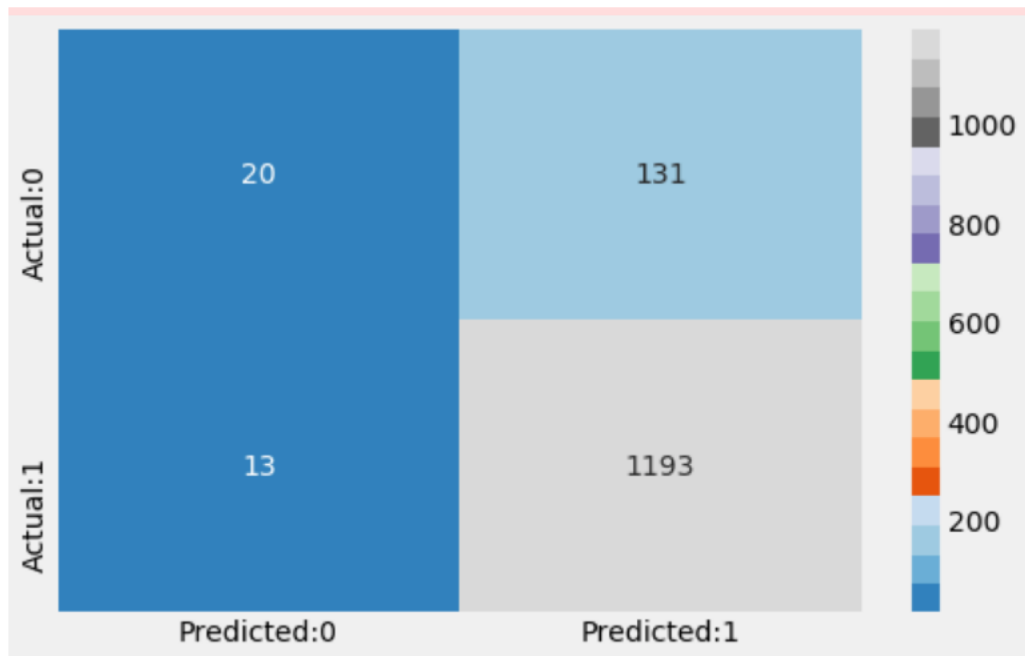
Lesson Learned

- Previous success of advertisement campaign and the amount of balance appear to be the biggest factors in predicting customers' decision whether to buy the product

- For customers contacted by the cellular phone, there is the largest proportion of people decide to subscribe to the bank's financial product.

- Based on the summary table of variables in logistic regression, the positive coefficient of previous campaign success is the largest in magnitude. This indicates that it has a positive effect on people saying yes. This makes sense because previous success might indicate that the customers might favor another bank product in the future.

- Therefore, if banks want to attract more people subscribe to their financial product, what they should do is to hire more people to work for them, improve the quality of conversation on the phone and run their campaigns. After all, current success might lead to the customers to become more interested in the bank service in the future. Consequently, it might promise a success of future advertisement, and enable the bank to make more money.

- NEVER put access key on github

Feature importance
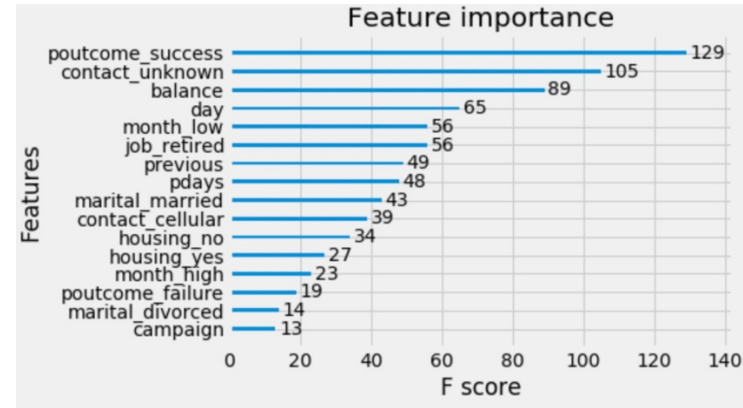
Variable	Coefficients
poutcome_success	1.467587
poutcome_failure	-0.896357
poutcome_unknown	-0.698403
job_retired	0.616200
contact_unknown	-0.610906
loan_yes	-0.404161
marital_married	-0.373824
job_unemployed	-0.368073
month_high	-0.320513
contact_cellular	0.294365

Recommendation

- The following stories should be finished in the next step
  - Enable user input options and predict whether the customer will buy the product
  - Construct a pipeline for this project, create database and load dataset on AWS.
  - Deploy the model onto AWS and develop the subscription prediction App. Keep track of customer subscription in the bank as more customers are added.