

# **Banking Churn Rate – Mitigation Strategies**

Thomas Poozhikala

Springboard Data Science Career Track

<https://github.com/tpoozhikala>

---

Slide template provided by Sacramento State University

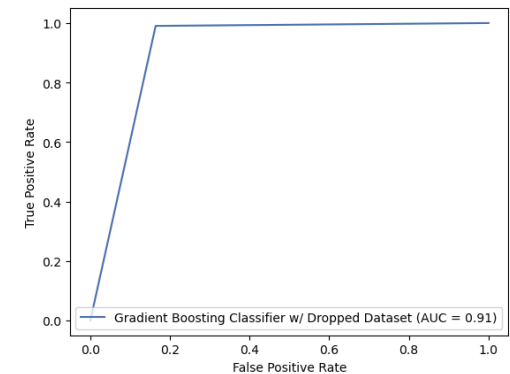
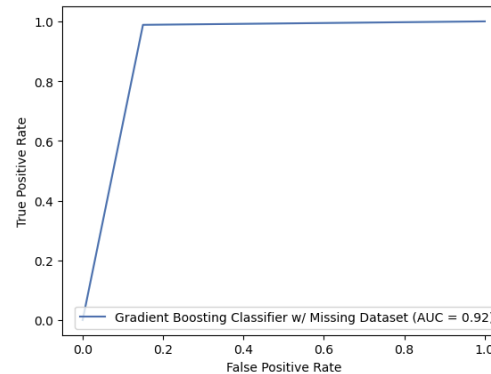
# Introduction of Problem & Objective

---

- Problem
  - Determining or predicting churn rate is key in increasing business effectiveness and allowing for targeted marketing and resources to increase customer retention. This is especially the case for a bank who released their user data on Kaggle for assistance on predicting churn in order to reduce it.
- Objective: The goal of this project is to provide a churn prediction model and mitigation strategies to increase overall bank revenue by 20% for the next financial quarter.

# Recommendation & Key Findings

| 'Dropped' Correlation List 'Attrition_Flag' |           | 'Missing' Correlation List 'Attrition_Flag' |           |
|---|-----------|---|-----------|
| Attrition_Flag                              | 1.000000  | Attrition_Flag                              | 1.000000  |
| Total_Trans_Ct                              | 0.364835  | Total_Trans_Ct                              | 0.380618  |
| Total_Ct_Chng_Q4_Q1                         | 0.281174  | Total_Ct_Chng_Q4_Q1                         | 0.288937  |
| Total_Revolving_Bal                         | 0.268979  | Total_Revolving_Bal                         | 0.263335  |
| Avg_Utilization_Ratio                       | 0.185690  | Avg_Utilization_Ratio                       | 0.179838  |
| Total_Trans_Amt                             | 0.170178  | Total_Trans_Amt                             | 0.178078  |
| Total_Relationship_Count                    | 0.149125  | Total_Relationship_Count                    | 0.150889  |
| Total_Amt_Chng_Q4_Q1                        | 0.133285  | Total_Amt_Chng_Q4_Q1                        | 0.128559  |
| Gender_M                                    | 0.045186  | Gender_M                                    | 0.047382  |
| Income_Category_\$60K - \$80K               | 0.028017  | Income_Category_\$60K - \$80K               | 0.026338  |
| Income_Category_\$40K - \$60K               | 0.023828  | Credit_Limit                                | 0.024437  |
| Education_Level_High_School                 | 0.018403  | Education_Level_College                     | 0.018937  |
| Credit_Limit                                | 0.017750  | Marital_Status_Married                      | 0.017872  |
| Marital_Status_Married                      | 0.015278  | Income_Category_\$40K - \$60K               | 0.015050  |
| Education_Level_College                     | 0.012472  | Card_Category_Silver                        | 0.014854  |
| Education_Level_Uneducated                  | 0.007257  | Education_Level_High_School                 | 0.009930  |
| Card_Category_Gold                          | 0.003949  | Education_Level_Uneducated                  | 0.006614  |
| Card_Category_Silver                        | 0.003173  | Income_Category_\$80K - \$120K              | 0.005148  |
| Income_Category_\$80K - \$120K              | 0.001899  | Education_Level_Graduate                    | 0.004847  |
| Card_Category_Blue                          | -0.001150 | Avg_Open_To_Buy                             | 0.000803  |
| Education_Level_Graduate                    | -0.001751 | Income_Category_\$120K +                    | -0.000831 |
| Avg_Open_To_Buy                             | -0.006245 | Marital_Status_missing                      | -0.002723 |
| Marital_Status_Divorced                     | -0.006688 | Card_Category_Gold                          | -0.003397 |
| Dependent_count                             | -0.008122 | Marital_Status_Divorced                     | -0.006305 |
| Marital_Status_Single                       | -0.011789 | Card_Category_Blue                          | -0.008463 |
| Income_Category_\$120K +                    | -0.014601 | Education_Level_missing                     | -0.009100 |
| Education_Level_Post-Graduate               | -0.018346 | Education_Level_Post-Graduate               | -0.013375 |
| Months_on_book                              | -0.018599 | Marital_Status_Single                       | -0.013437 |
| Card_Category_Platinum                      | -0.022579 | Income_Category_missing                     | -0.017580 |
| Customer_Age                                | -0.024498 | Card_Category_Platinum                      | -0.020284 |
| Income_Category_Less than \$40K             | -0.033687 | Dependent_count                             | -0.020550 |
| Education_Level_Doctorate                   | -0.043378 | Months_on_book                              | -0.022270 |
| Gender_F                                    | -0.045186 | Income_Category_Less than \$40K             | -0.023134 |
| Months_Inactive_12_mon                      | -0.160701 | Customer_Age                                | -0.035203 |
| Contacts_Count_12_mon                       | -0.203447 | Education_Level_Doctorate                   | -0.039483 |
|   |           | Gender_F                                    | -0.047382 |
|   |           | Months_Inactive_12_mon                      | -0.156552 |
|   |           | Contacts_Count_12_mon                       | -0.211327 |

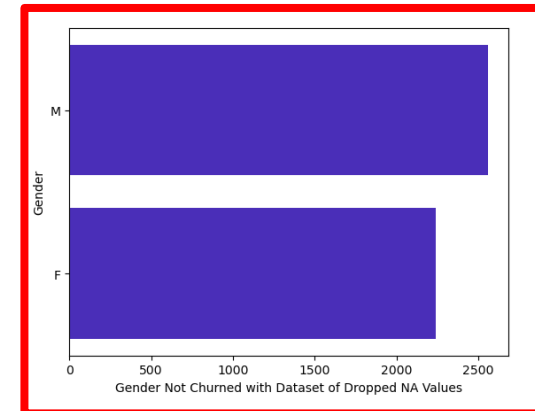
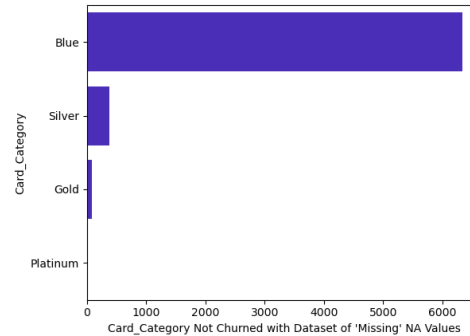
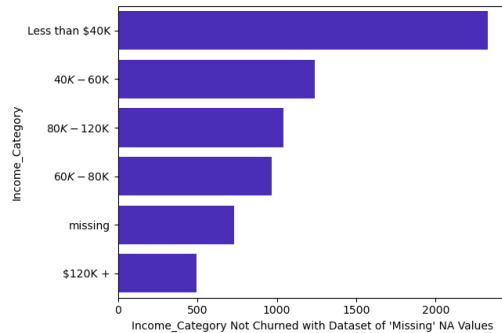
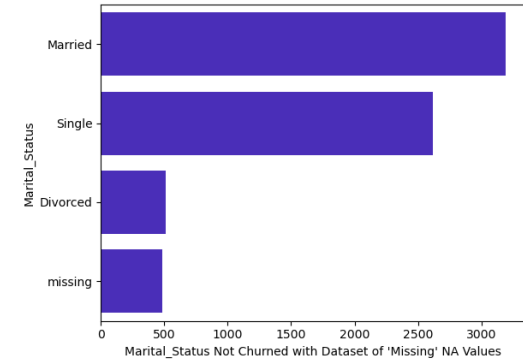
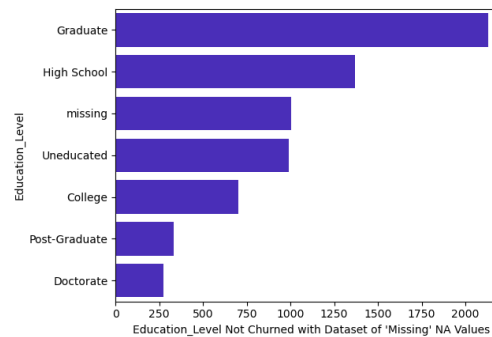
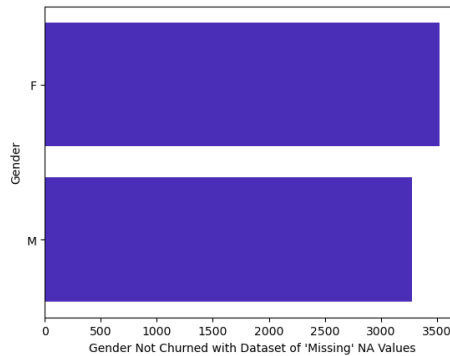


\*ROC\_AUC Scores from best performer and easily reproducible Gradient Boosting Model

\*Correlation List showcasing correlations with Target Attrition\_Flag column

**Recommended Churn Mitigation Strategies:** The bank along with deploying a hyperparameterized Gradient Boosting model can focus marketing deals towards the higher to likely attrite audience of females or older demographic or maximize marketing to the not likely to churn population of younger married males with children or dependents.

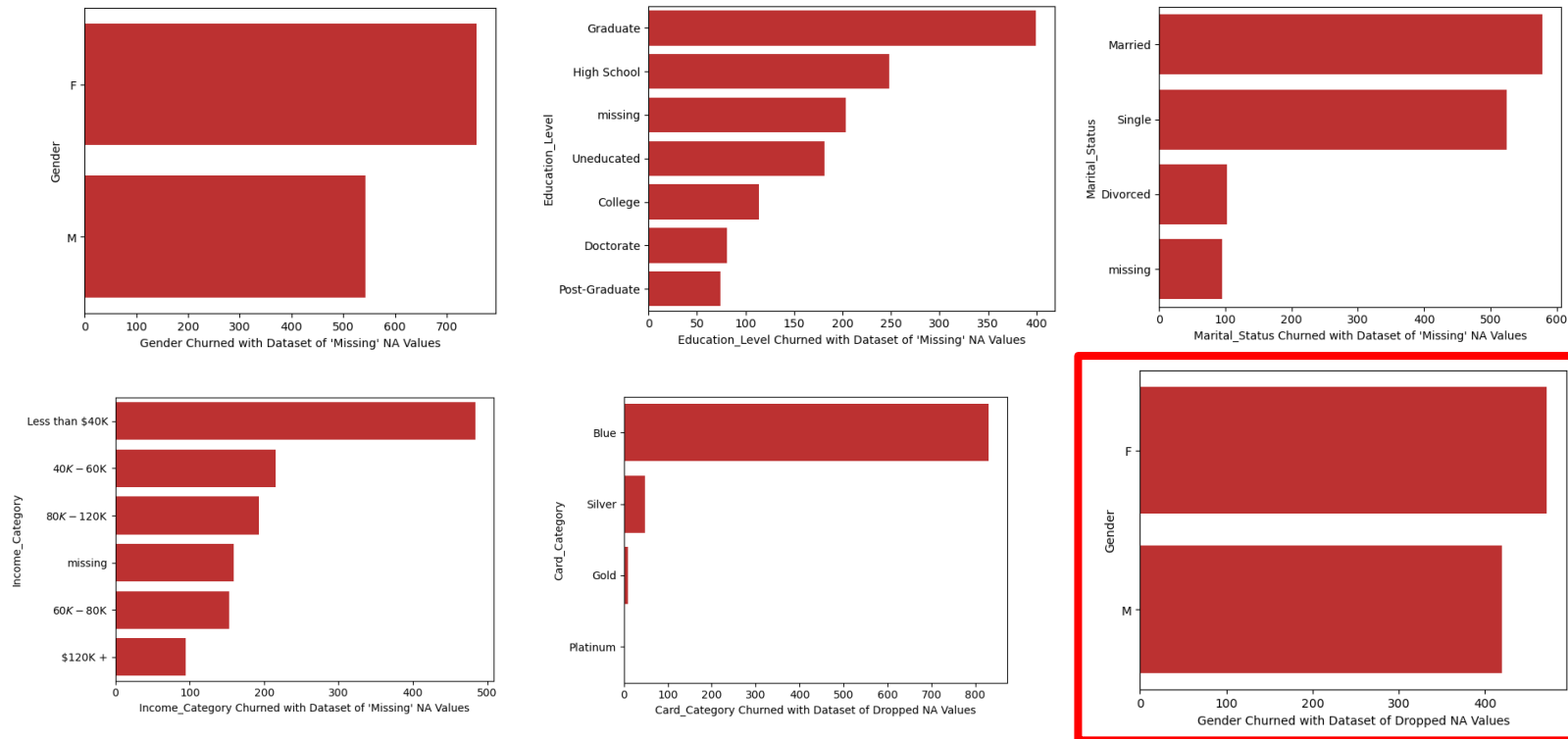
# Assessing Not Churn User Base



\*Graphs displaying just assessment from 'Missing' Dataset values

\*Gender differed within 'Dropped' Dataset between Churn and Not Churned

# Assessing Churn User Base



\*Graphs displaying just assessment from 'Missing' Dataset values

\*Gender differed within 'Dropped' Dataset between Churn and Not Churned

**User Base Assessment:** Most common user base are females, graduate degree holders, married individuals, those making less than 40k, and majority Blue card holders. Females less likely to disclose personal information.

# Establishing a Predictive Model

---

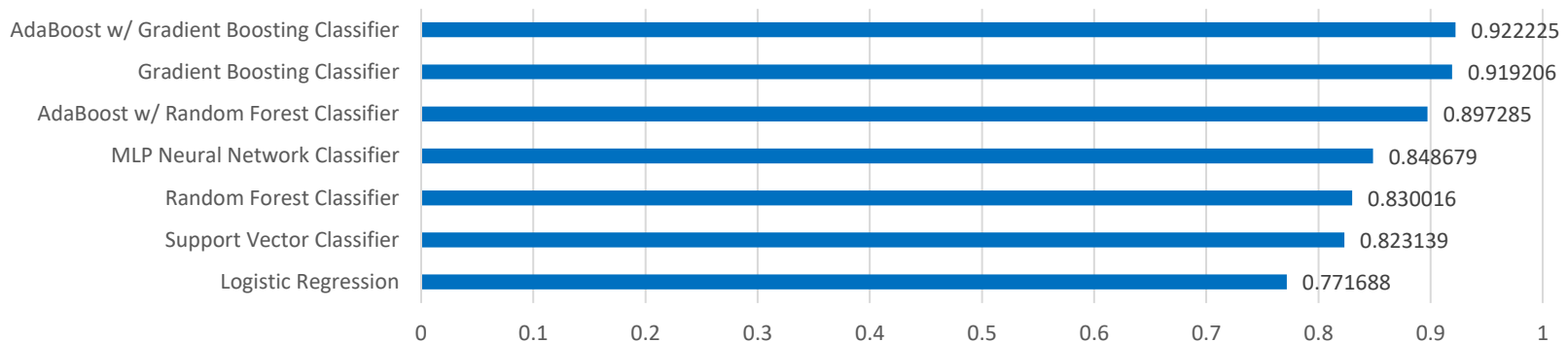
- Two Datasets (Missing and Dropped) constructed to deal with 'Missingness' of values.
- Each categorical feature given associated k-1 dummy variable
- Predictive models trained with 70/30 split and 5-fold Cross Validated for optimal hyperparameters
- Model performance assessed based on Accuracy, Precision, Recall, f1 score, **ROC\_AUC score**, and Training Time (Sec)

| Model Name                                      | Accuracy_score | Precision_Score | Recall_score | f1_Score | ROC_AUC_score | Training Time (Sec) |
|---|----------------|-----------------|--------------|----------|---------------|---------------------|
| Gradient Boosting Classifier w/ Missing Dataset | 0.967914       | 0.927273        | 0.850000     | 0.886957 | 0.919206      | 846.87              |
| Gradient Boosting Classifier w/ Dropped Dataset | 0.968366       | 0.935780        | 0.836066     | 0.883117 | 0.913248      | 603.65              |

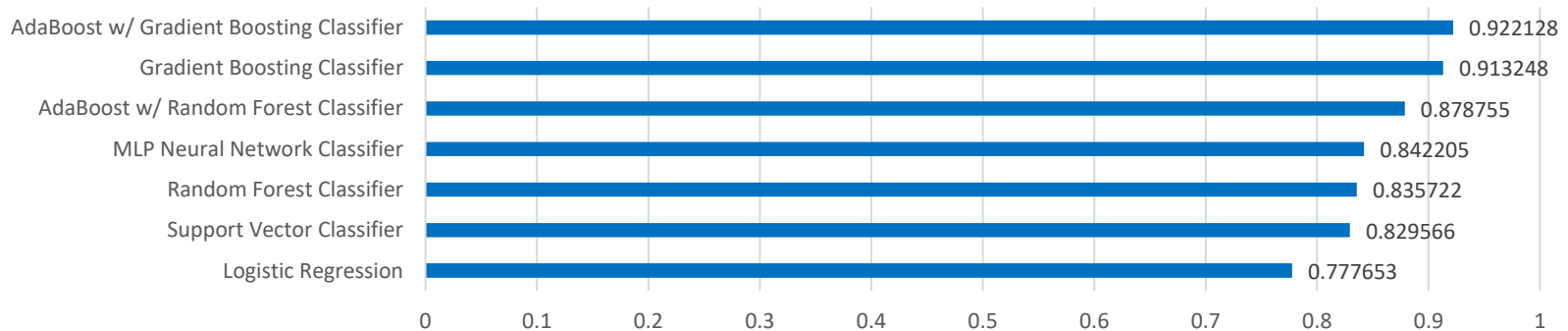
\*Table Displaying Best and Easily Reproducible Gradient Boosting Model Metrics

## 'Missing' & 'Dropped' Dataset Model Comparison

ROC\_AUC\_score Missing Dataset



ROC\_AUC\_score Dropped Dataset



\*Bar graphs displaying Prediction model ROC\_AUC scores for Missing and Dropped Datasets

## ‘Missing’ vs. ‘Dropped’ & Modeling Insights

---

- **Logistic Regression** is the **worst performer** for both the missing and dropped datasets per predicting churn.
- Gradient Boosting Classifier is the best performer overall when applied with an AdaBoost Classifier model.
- However, the **individual Gradient Boosting Classifier** seems **sufficient** with a limitation of training time for reproducibility and scale.
- Imputing the original ‘Unknown’ values with ‘missing’ slightly improved the stronger performing Gradient Boosting and AdaBoost Classifier models.
- Future work to reduce Gradient Boosting Classifier training time and uncover feature importances and discover methods to improve MLP Neural Network Classifier.



# Summary

---

- Gradient Boosting Model is the best performer with an intuitive sense of reproducibility to deploy to production.
- Suggested Mitigation Strategies include:
  - Focus marketing deals towards the higher to likely attrite audience of females and older demographic.
  - Or maximize marketing to the not likely to churn population of younger married males with children or dependents.

# Thank You! Questions?

<https://github.com/tpoozhikala/Bank-Churnrate/tree/main>

