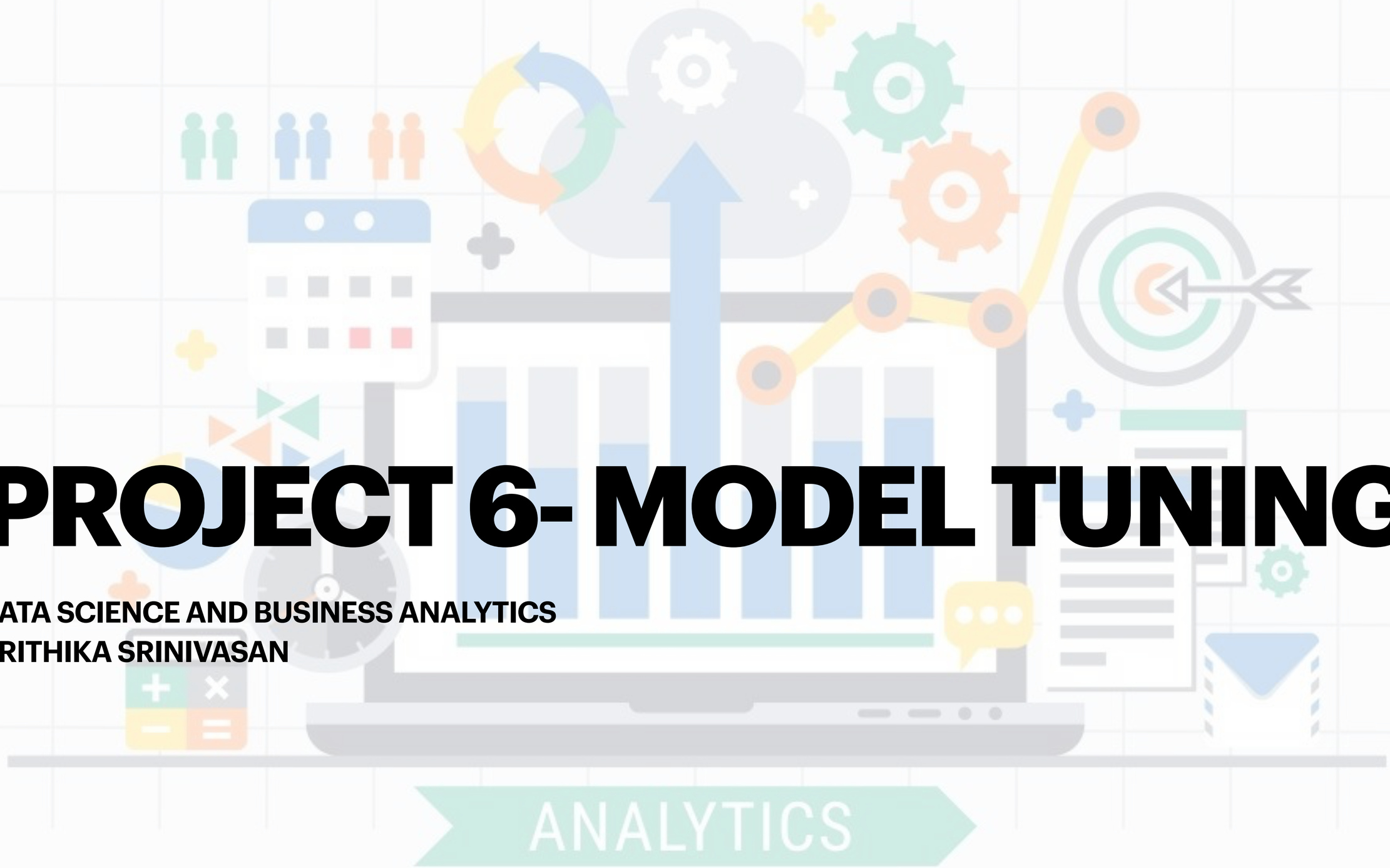


PROJECT 6- MODEL TUNING

DATA SCIENCE AND BUSINESS ANALYTICS
KRITHIKA SRINIVASAN

BUSINESS OVERVIEW

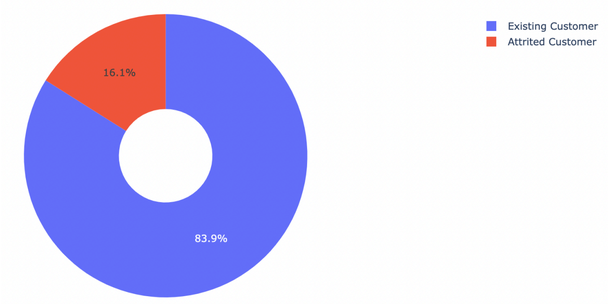
- Thera bank saw a decline in the number of users of their credit cards
- Customers leaving credit card services would lead the bank to loss
- The bank wants to identify customers who will leave their credit card services and why
- Objectives:
  - Explore and visualize the dataset
  - Build a classification model to predict if the customer is going to churn or not
  - Optimize the model using appropriate techniques
  - Generate a set of insights and recommendations that will help the bank

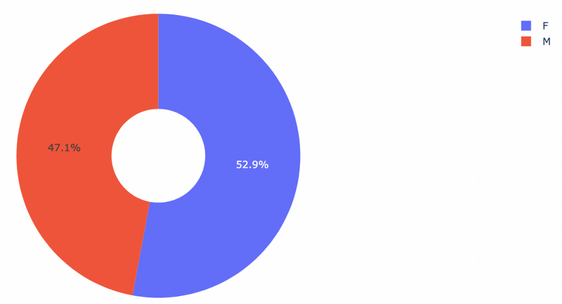
SOLUTION APPROACH

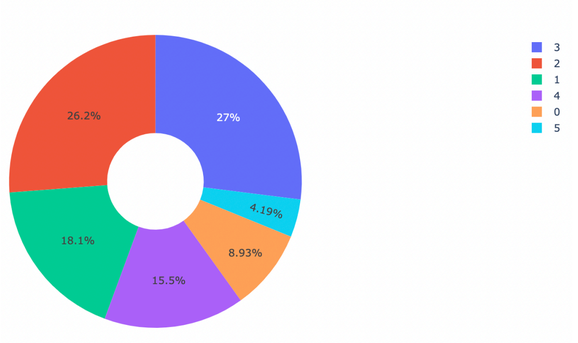
- This is a clear classification problem
- The approach I took was to solve the problem first with multiple classification algorithms, perform oversampling and under sampling, performing hyper tuning and then deciding on the best model for prediction
- The above methods of Supervised Learning seem apt for this problem as we need to predict if a customer will buy churn or not (1 or 0)
- Prior to implementing these algorithms, in-depth EDA was carried out to identify patterns in the data

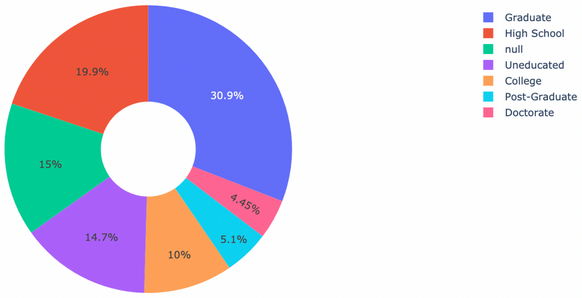
KEY FINDINGS AND INSIGHTS (EDA)

- 16% of the customers are churned
- 52% of the customer base is Female
- 27% of the customers have 3 dependents
- 31% of the customers are graduates



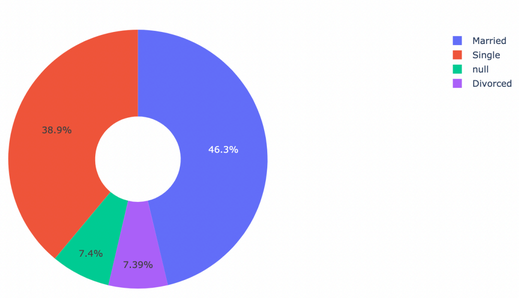


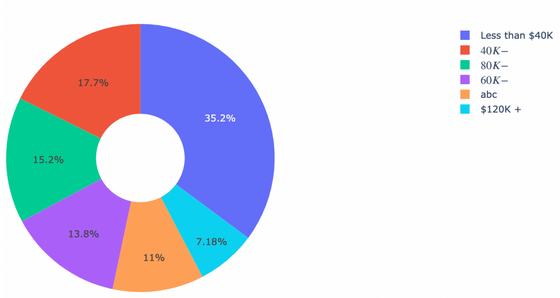


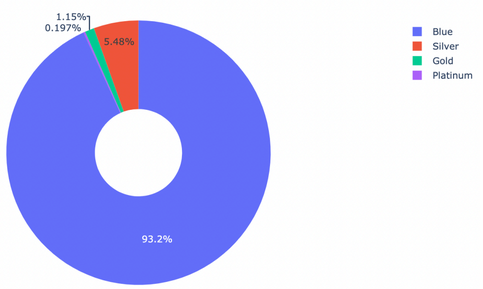


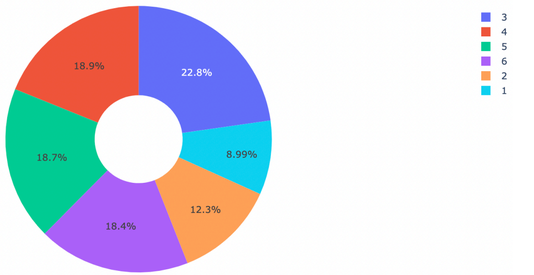
KEY FINDINGS AND INSIGHTS (EDA)

- 46.3% of customers are married
- 39% of customers are single
- 35.2% of customers earn less than 40K
- 93% of customers have the blue credit card
- 23% of customers hold 3 products



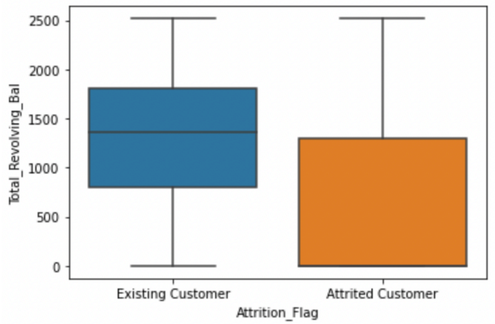


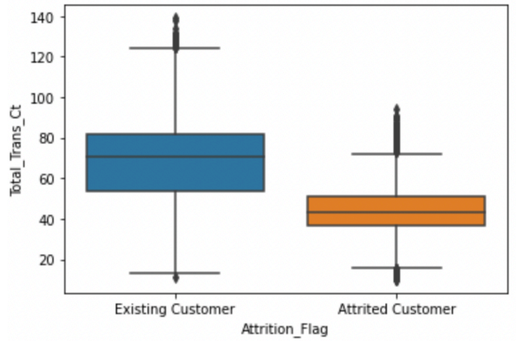


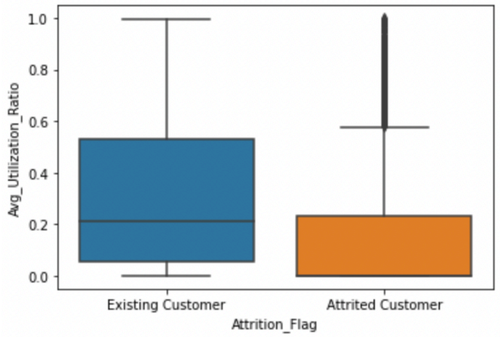


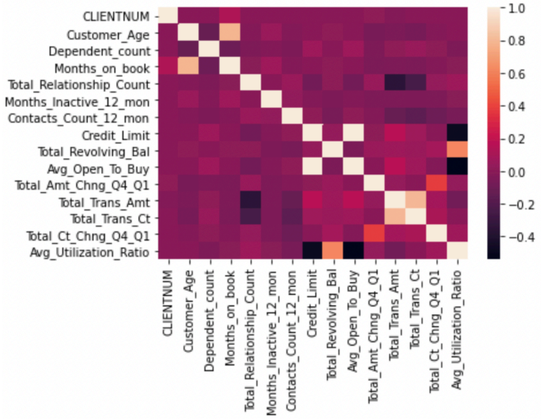
KEY FINDINGS AND INSIGHTS (EDA)

- Total Revolving Balance is lower for Attrited customers
- Total Transaction Count is lower for Attrited customers
- Utilization Ratio is lower for Attrited customers
- High correlation between the following columns:
  - Customer_Age and Months_on_book
  - Avg_Open_To_Buy and Credit_Limit
  - Total_Trans_Ct and Total_Trans_Amt
  - Total_Revolving_Bal and Avg_Utilization_Ratio









MODEL BUILDING AND EVALUATION

- Ran 6 classification models to see which has the best cross-validation score
- Performed oversampling and undersampling

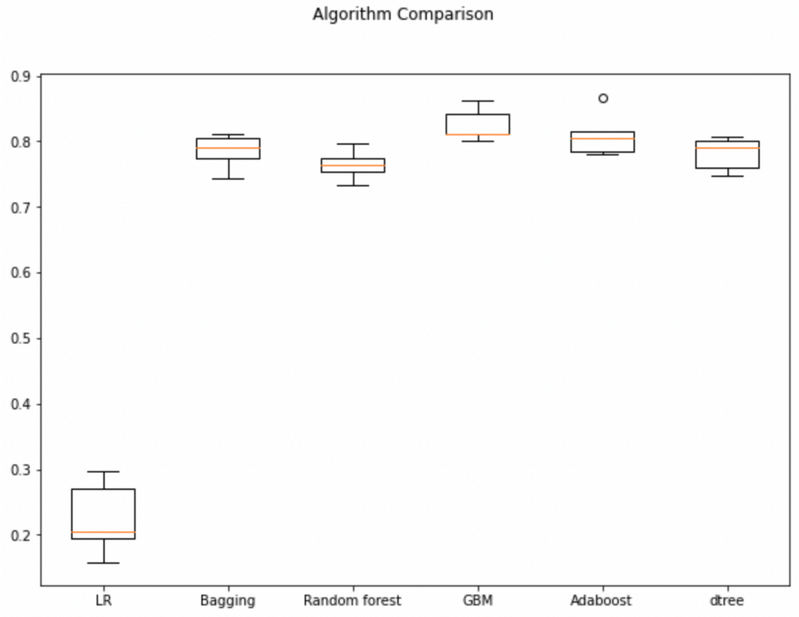
CROSS VAL SCORES- OVERSAMPLING

```
LR: 83.84017394985473
Bagging: 95.01862648886835
Random forest: 97.11718524504992
GBM: 96.07779637861033
Adaboost: 95.66597394600628
dtree: 93.64582732013315

Training Performance:
```

CROSS VAL SCORES- UNDERSAMPLING

```
LR: 80.74201988487702
Bagging: 90.6771323914181
Random forest: 93.75196232339091
GBM: 94.05651491365778
Adaboost: 92.72632129774988
dtree: 89.64992150706436
```



- Metric of importance: Recall

MODEL EVALUATION & PERFORMANCE

TRAINING SET PERFORMANCE COMPARISON

	Random Forest	Random Forest tuned with Random Search	Gradient Boost- Oversampled	Gradient Boost- Oversampled tuned with Random Search	Gradient Boost- Undersampled	Gradient Boost- Undersampled tuned with Random Search
Accuracy	1.0	1.0	0.970779	0.950971	0.972336	1.0
Recall	1.0	1.0	0.970779	0.961365	0.978484	1.0
Precision	1.0	1.0	0.970779	0.941787	0.966599	1.0
F1	1.0	1.0	0.970779	0.951475	0.972505	1.0

VALIDATION SET PERFORMANCE COMPARISON

	Random Forest	Random Forest tuned with Random Search	Gradient Boost- Oversampled	Gradient Boost- Oversampled tuned with Random Search	Gradient Boost- Undersampled	Gradient Boost- Undersampled tuned with Random Search
Accuracy	0.950148	0.950642	0.954590	0.926950	0.937808	0.943731
Recall	0.861963	0.877301	0.883436	0.895706	0.947853	0.960123
Precision	0.833828	0.826590	0.842105	0.719212	0.739234	0.756039
F1	0.847662	0.851190	0.862275	0.797814	0.830645	0.845946

TEST SET PERFORMANCE WITH SELECTED MODEL

	Accuracy	Recall	Precision	F1
0	0.944225	0.969231	0.753589	0.847914

BUSINESS RECOMMENDATIONS

- Based on the analysis performed, it looks like customers who do not use their credit cards as much are the ones that are churned
- The bank should look into promotions and deals that might give the customers cash back or points for all purchases using the credit card
- There could also be a rewards program on big purchases like air travel, furniture etc.
- This would provide incentive to use the card for all purchases and reduce the churn the bank is seeing