Subscriber Retention: Halting the Churn

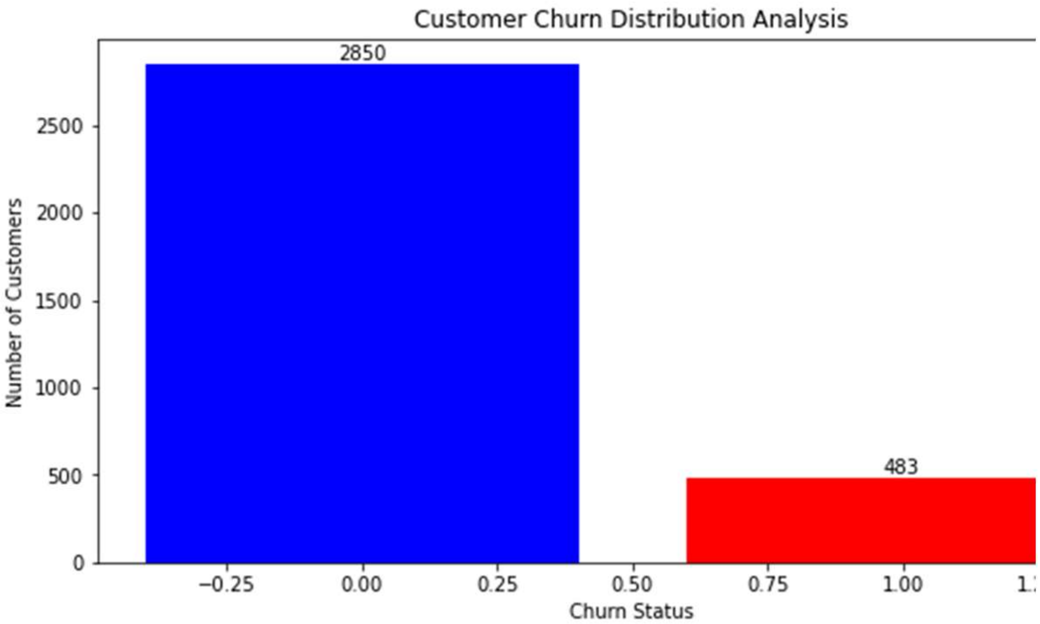
Deon Durrant

Business and Data Understanding

Problem Statement

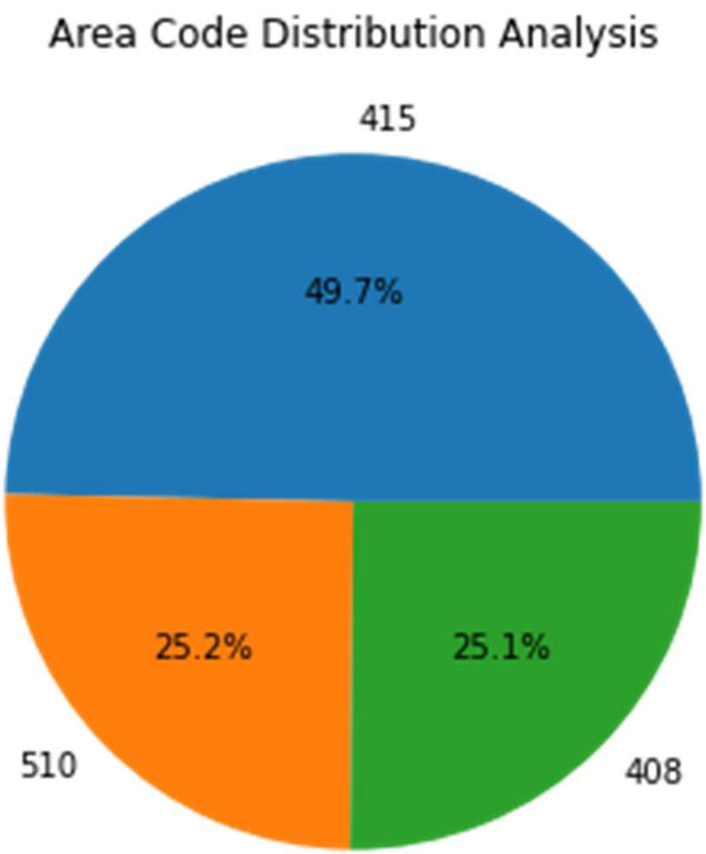
SyriaTel, a major telecommunications provider, is interested in minimizing resources expended on customers who are likely to terminate their services, a phenomenon known as churn. Customers' churn intentions maybe predictive by identifying and isolating patterns hidden in the data.

Business and Data Understanding



Churn Distribution Analysis

- 2850 instances “not-churned “
- 483 instances “churned”
- **SyriaTel** has a churn rate of **14.49**%

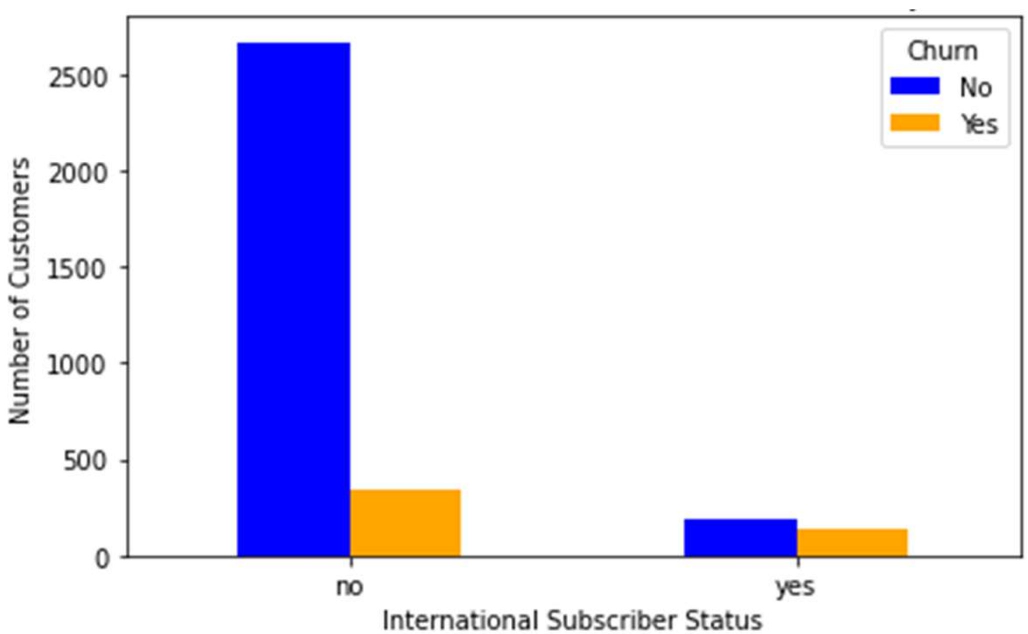


Business and Data Understanding

Area Code Distribution Analysis

- 49.7% of the customers are from 415 area code
- 25.2 % in area code 510
- 25.1% in area code 408

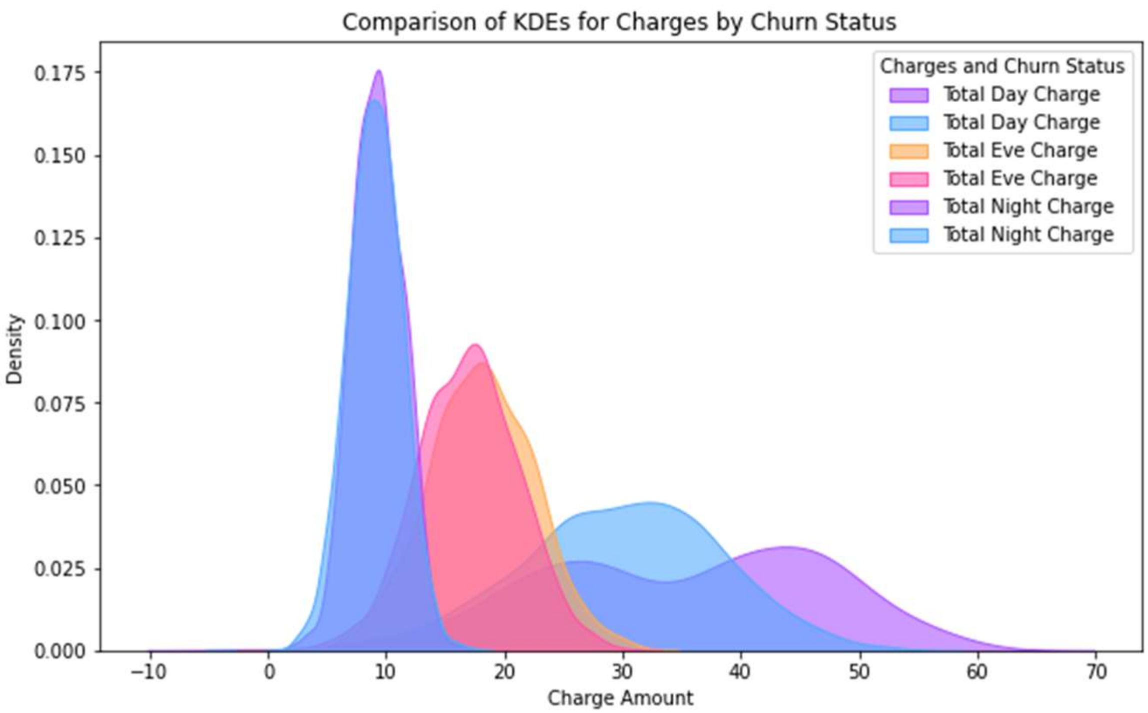
Business and Data Understanding



International Subscribers Churn Distribution

- **10.38%** subscribers have an international plan
- **39.56%** of international plans subscribers terminated the service

Business and Data Understanding



Churn Rates and Telephone Charges by Time of Day

Day Usage Churn

- Churned customers have two distinctive peaks indicating different customer behavior within the subset

Evening Usage Churn

- Non-churned extends tail to the right, higher total evening charge for some non-churned customers.

Night Usage Churn

- High degree of overlapping, feature is not distinguishing churn and not-churn

Modeling Evaluation: Analysis of models’ performances

Model Comparison

Model	Precision	Recall	F1 Score	Accuracy	AUC Score
Logistic Regression	0.56	0.17	0.26	0.86	**0.7453**
KNN	0.70	**0.41**	**0.51**	0.89	0.6882
Decision Tree	0.75	0.39	**0.51**	0.858	0.6847
Random Forest	**0.88**	0.31	0.46	**0.90**	0.6538

Precision: Random Forest (0.88), lowest rate of false churn(positives).

Recall: KNN (0.41) highest rate of true positives/churn

F1 Score: KNN and Decision Tree (0.51 overall performance.

Accuracy: Random Forest (0.90), overall correctness of the predictions.

AUC Score: Logistic Regression (0.7453), ability to distinguish between churn and not-churn

Modeling Evaluation: Analysis of models’ performances

Confusion Matrices Model Performance

Model	True Negatives (TN)	False Positives (FP)	False Negatives (FN)	True Positives (TP)
Logistic Regressi on	838	19	119	24
KNN	832	25	**85**	**58**
Decision Tree	838	19	87	56
Random Forest	**851**	**6**	98	45

KNN

- highest **TP** accurately predict churn
- lowest **FN** where the model incorrectly predicts 'No Churn', but customer churns.

•Random Forest

- highest **TN** predicts 'No Churn
- lowest **FP**, predicts 'Churn' but customer does not churn

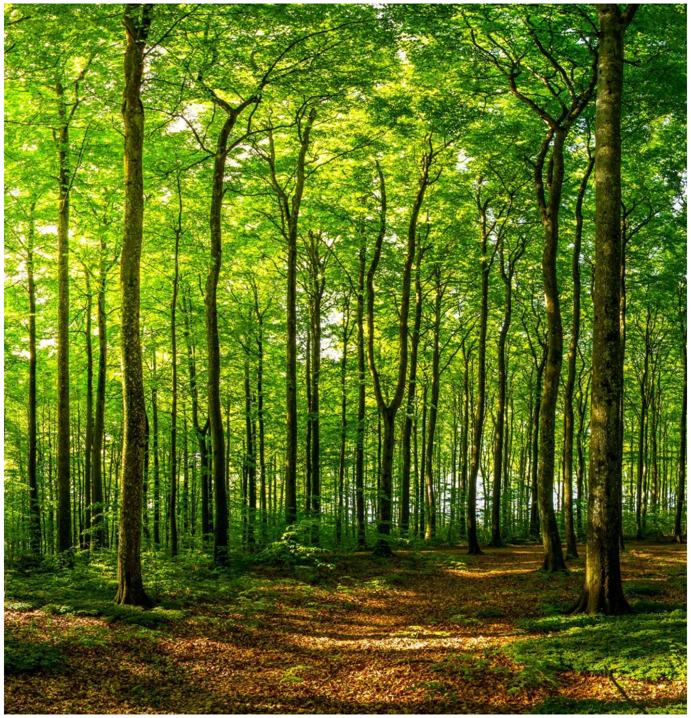
Modeling
Model Selection and Tuning

Random Forest selected because

- It achieved a high recall, the highest precision, F1 score, and accuracy scores among the models evaluated.
- The ability to handle highly-correlated variables
- Easy to evaluate variable importance or contribution to model
- Reduce risk of overfitting compared to other models

Hyperparameter Tuning

- GridSearchCV: to obtain the best parameters for the model



Modeling Evaluation: Model Improvement Analysis

Metric	rf	**model_rf**	Difference %
Precision	**0.88**	0.80	-9.09%
Recall	0.31	**0.49**	+58.06%
F1-score	0.46	**0.61**	+32.61%
Accuracy	90%	**91%**	+1.11%
AUC Score	0.65	**0.73**	+12.31%

Precision.

- model_rf shows a decrease in precision by **9.09%**
- higher tendency to classify not-churn(negatives) cases as churn(positives)

Recall

- model_rf is **58.06%** higher than rf
- better at identifying actual positive cases overall.

F1-Score

- **32.61%** improvement
- more robust in balancing between missing churns and maintaining accuracy in those predictions.

Accuracy

- **1.11%** improvement
- makes correct predictions on a slightly higher percentage of the total dataset

AUC Score

- **12.31%,** improvement
- enhanced predictive performance.

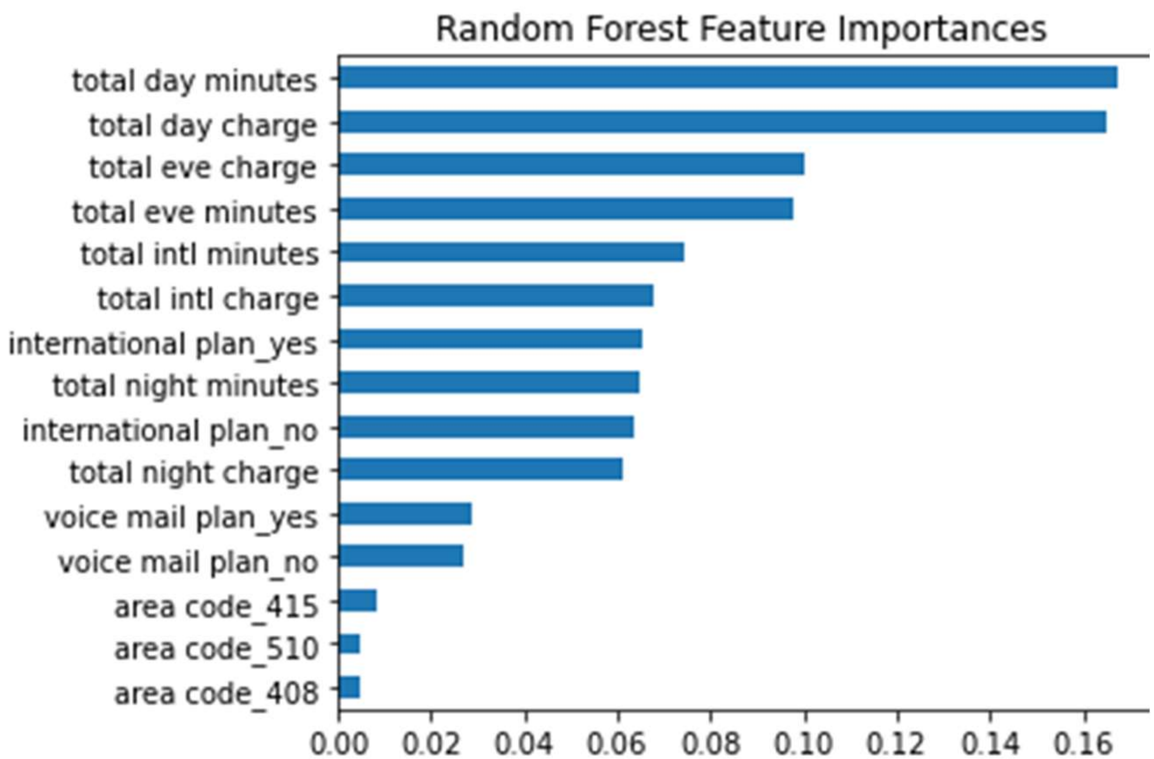
Modeling

Evaluation: Model Improvement Analysis

Model	True Negatives (TN)	False Positives (FP)	False Negatives (FN)	True Positives (TP)
rf	851	6	98	45
model_rf	839	18	**73**	**70**

- TN are slightly lower at 839, drop in correctly identifying not-churn.
- FP increase to 18, model is more aggressive in predicting churn but with associated errors
- FN are reduced to 73, indicating better performance in catching positive cases compared to rf.
- Significantly higher TP at 70, suggesting better effectiveness at identifying churn.

Analysis of the feature importances



Daytime usage

- Total Day Minutes(**0.167243),** Total Day Charge(**0.164814)**
- Daytime usage is highest predictor of churn
- Day rate = 0.17

Evening time usage

- Total Evening Charge(**0.100384),** Total Evening Minutes (**0.097512**)
- Eve rate= 0.085

International usage and plan

- Total International Minutes (**0.074507),** Total International Charge **(0.067571**)
- International Plan: Yes **(0.065110),** No (**0.063774)**
- International rate= 0.27

Predictive Recommendations

Concentrated Usage Analysis

- Accurate and meticulous data collection for usage analysis
- Employ detailed analysis of usage patterns, including rate structure for different times of day.
- Allocate resources to track and analyze high-risk customers and employ mitigation strategies to stem churn

Customer Segmentation

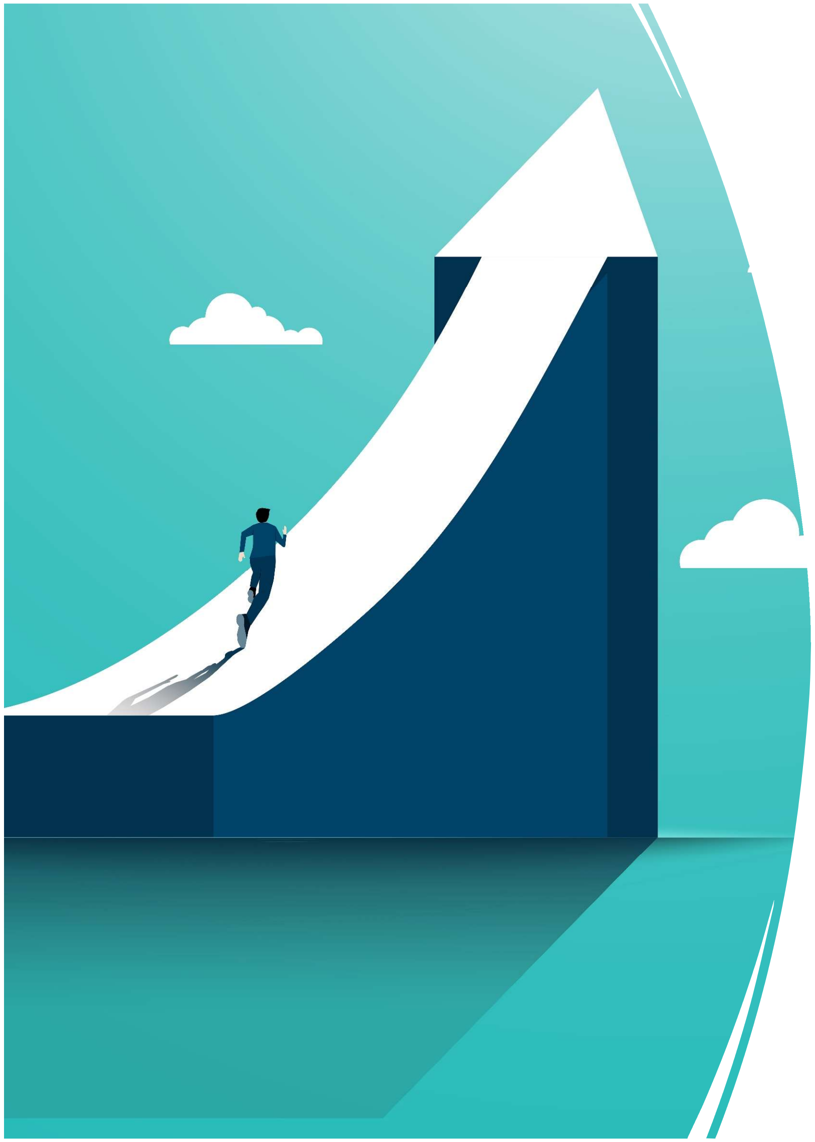
- Utilize insights on plan types and usage patterns to segment customers
- Allotment rate structure for international subscribers to curb churn

Invest in Subscription Analytics

- Acquire dashboards to monitor high-importance features in real-time, allowing for proactive measures to prevent churn

Survey customers at the point of cancel

- Utilize churn survey to gather data to understand why customers cancel their accounts.



Next Steps

Improve model predictive performance and accuracy

- Utilize other ensemble techniques such as Adaboost
- Conduct feature engineering: reveal hidden data or patterns in the dataset that are not immediately apparent
    - Create feature interactions
    - Binning of continuous variables

Questions

Thank You!

Email:DurrantDeon@gmail.com
LinkedIn: linkedin.com/in/dr-d-durrant