Predicting Customer Churn

Business Problem

Customer churn, the departure of customers, impacts revenue and profitability. The project aims to identify reasons for churn and develop strategies for its reduction in the telecom industry.

Project objective:

Identify and prevent customer churn within a telecommunications company to enhance business performance.

Data Overview

Data is sourced from Kaggle.com

Data Features:

- Total day, evening, night and international minutes used as well as their charges.
- Number of customer service calls.
- State and Area codes.
- International plan
- voicemail plan
- Churn numbers

Data Preprocessing

Checked for:

- Missing data
- Duplicates

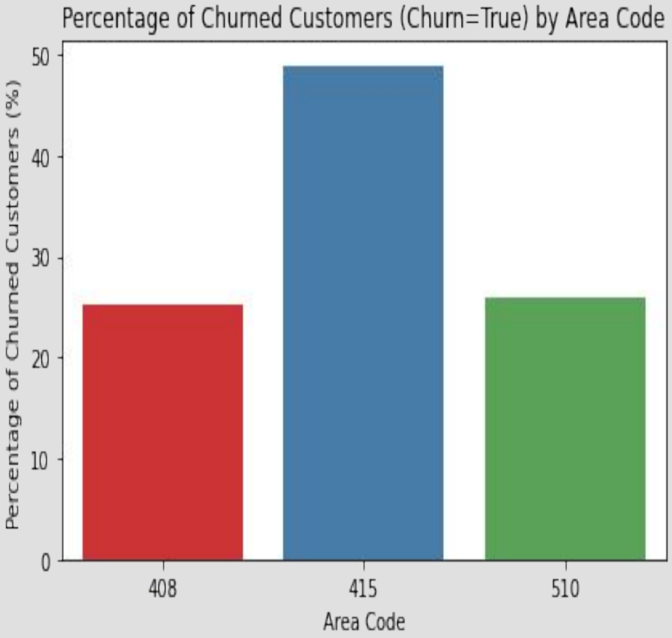
Target Variable:

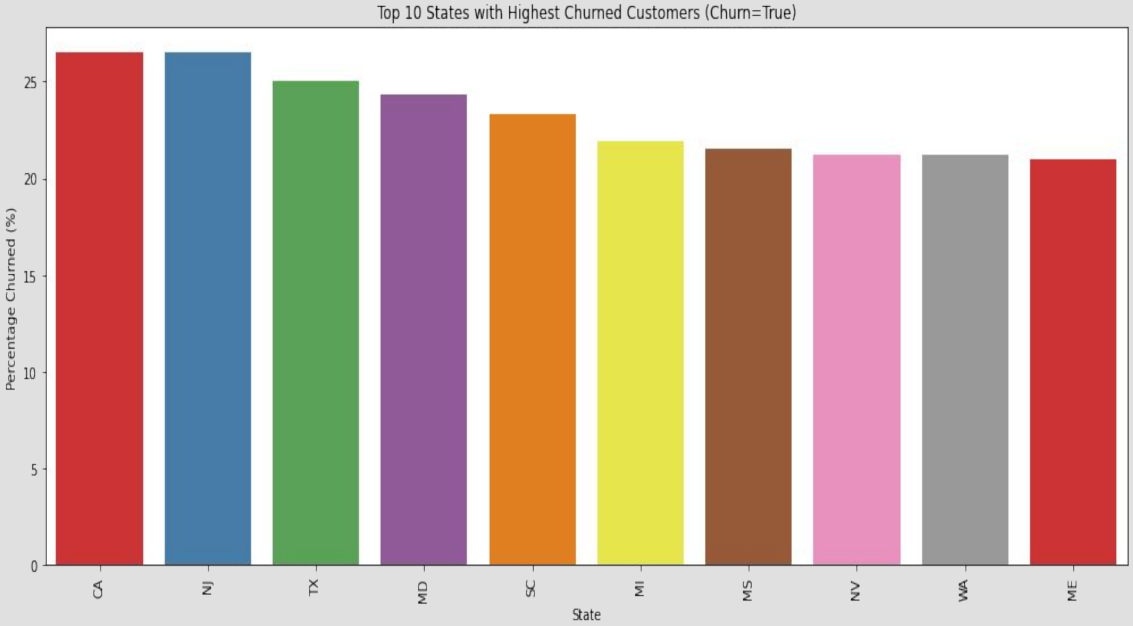
- Churn

Dropped Features that were not significant such as :

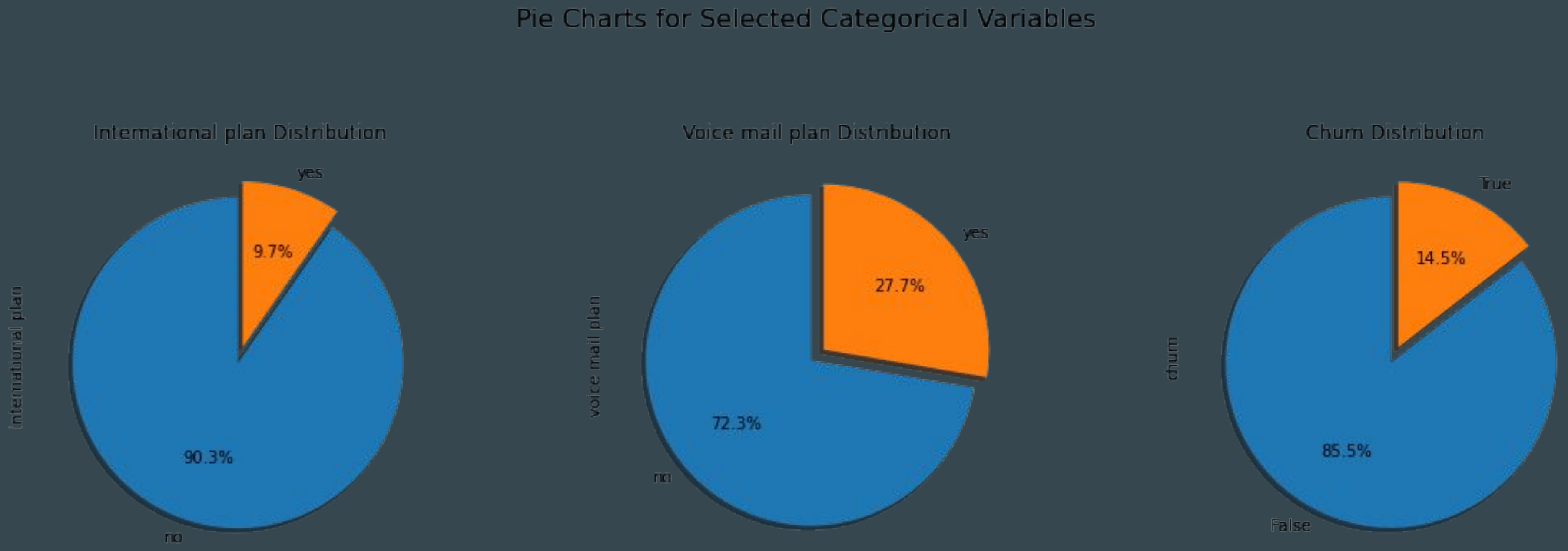
- Phone numbers

Exploratory Data Analysis

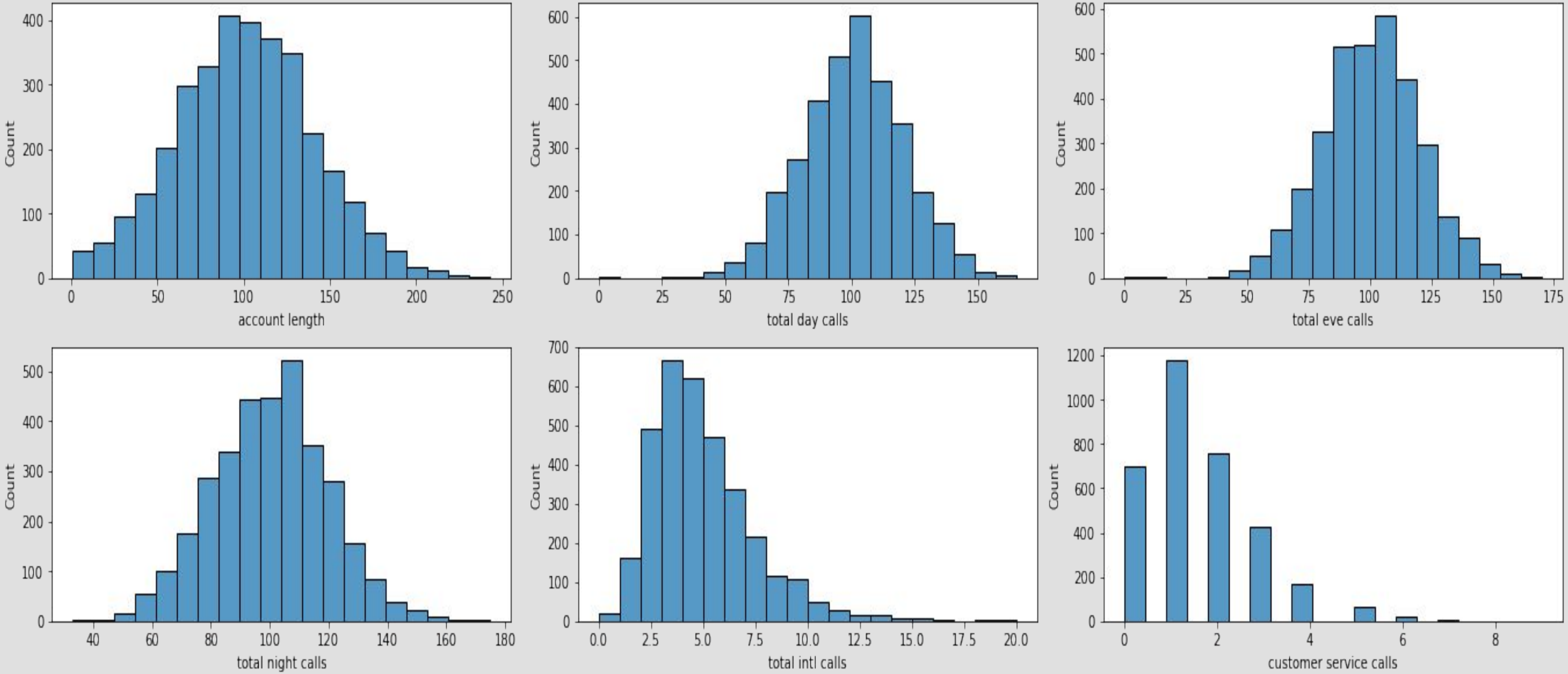




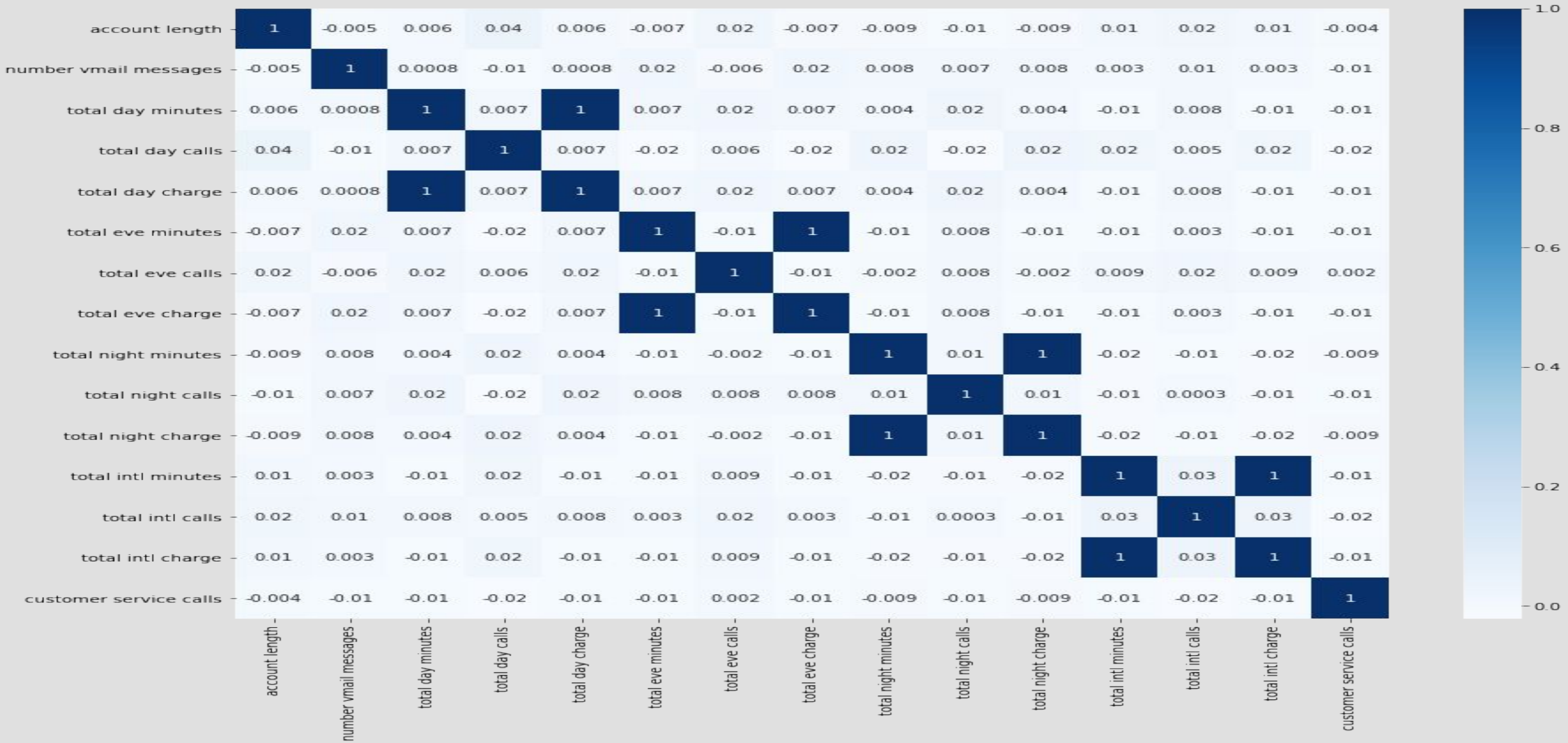
Exploratory Data Analysis...cont



Distributions of features



Correlation Matrix Numeric Features



Modeling

- Our analysis unveils limitations in feature relevance, model selection, potential external influences, and the risk of overfitting.

  Models used:

- Logistic regression
- Decision Tree
- Random Forest

Measures of Model prediction

Precision	• How accurate are the model's positive guesses
Accuracy	• Overall, how many guesses are right
F1 Score	• A balance of accuracy and completeness.
Recall	• How well does the model catch all the positive cases

Results of Model prediction

Decision Tree Classifier

* Precision: 79.8%
* Accuracy: 94.8%
* F1 Score: 80.9%
* Recall: 82%

Random Forest Model

* Precision: 67.9%
* Accuracy: 92.7%
* F1 Score: 75.7%
* Recall: 85.6%

Logistic Regression

* Precision: 55.5%
* Accuracy: 88.7%
* F1 Score: 64.7%
* Recall: 77.5%

Model Summary

- Though Decision Tree Model outperforms the other two models, we will pick the Random Forest Model as it has the highest recall and aligns better with our business requirements as it minimizes the risk of losing customers who were incorrectly classified.

Limitations

Feature Relevance:

- Not all dataset features were impactful for predicting churn, leading to potential model inefficiencies.
- lack of detailed information on states and area codes, prevented further in-depth exploration

Modeling:

- Model Selection: Despite using Logistic Regression, the focus was on tree-based models. Exploring models like gradient boosting or neural networks might enhance performance.
- Further research and analysis, considering additional data sources and alternative modeling approaches, would enhance our understanding and strategy for reducing customer churn effectively.

External Factors:

- Temporal Changes: The data might exhibit unnoticed trends over time, affecting the analysis.
- Bias Concerns: Potential biases, such as geographic or demographic, in the dataset could skew results.

Generalization:

- Overfitting Risk: Models, especially complex ones like Random Forest, might overfit, necessitating robust validation and regularization.

Recommendations

- Focus on improving customer service as it's a major determinant of customer satisfaction.
- Engage with customers who have high daily usage as they are more likely to churn.
- Consider special plans or offers for long-term customers and regions with high churn rates to incentivize loyalty.

Conclusion

- In summary, we have derive actionable recommendations, emphasizing improvements in customer service, engagement with high-usage customers, and the implementation of loyalty incentives to mitigate churn.
- Additionally, we underscore the importance of addressing states or area codes with high churn rates, as these regions may require tailored strategies for effective churn reduction.
- Alternative modeling approaches, remain essential for refining churn reduction strategies.