

### Business Case:

No-Churn Telecom is an established Telecom operator in Europe with more than a decade in Business. Due to new players in the market, telecom industry has become very competitive and retaining customers becoming a challenge. In spite of No-Churn initiatives of reducing tariffs and promoting more offers, the churn rate (percentage of customers migrating to competitors) is well above 10%. No-Churn wants to explore possibility of Machine Learning to help with following use cases to retain competitive edge in the industry.

### Challenges Faced

**Handling Categorical Data:** Converting categorical variables like state, area code, international plan, and voicemail plan into meaningful numerical representations.

**Feature Selection:** Identifying the most impactful features and eliminating redundant or less relevant ones.

**Feature Transformation:** Applying techniques like standardization, normalization, and log transformation to improve model performance.

**Creating New Features:** Deriving new features, such as interaction terms or customer service interaction frequency.

**Detecting Outliers:** Using techniques like Z-score, IQR method, or clustering algorithms to identify unusual customer behavior.

**Handling Outliers:** Deciding whether to remove, cap, or transform extreme values to reduce their impact on model performance.

**Imbalanced Data:** Addressing class imbalance using techniques like oversampling, undersampling, or SMOTE.

**Evaluation Metrics:** Focusing beyond accuracy, using precision, recall, and F1-score to measure performance effectively.

**Ensemble Learning vs. Simplicity:** Choosing between single models and ensemble techniques like bagging, boosting, or stacking.

### Model Comparison Report:

#### Before Hyperparameter Turning

- The **Decision Tree** and **Random Forest** models achieved **100% accuracy**, indicating **overfitting**, as the model might have memorized the training data.
- **Bagging**, **Extreme Gradient Boosting (XGBoost)**, and **LightGBM** models also performed well on the training set with accuracy scores of **99.85%**, **99.94%**, and **98.78%**, respectively.
- The **Support Vector Machine (SVM)** had the lowest training performance, with an accuracy of **68.02%** and an F1 score of **18.42%**, suggesting it struggled to learn from the training data.

#### **Test Dataset Insights:**

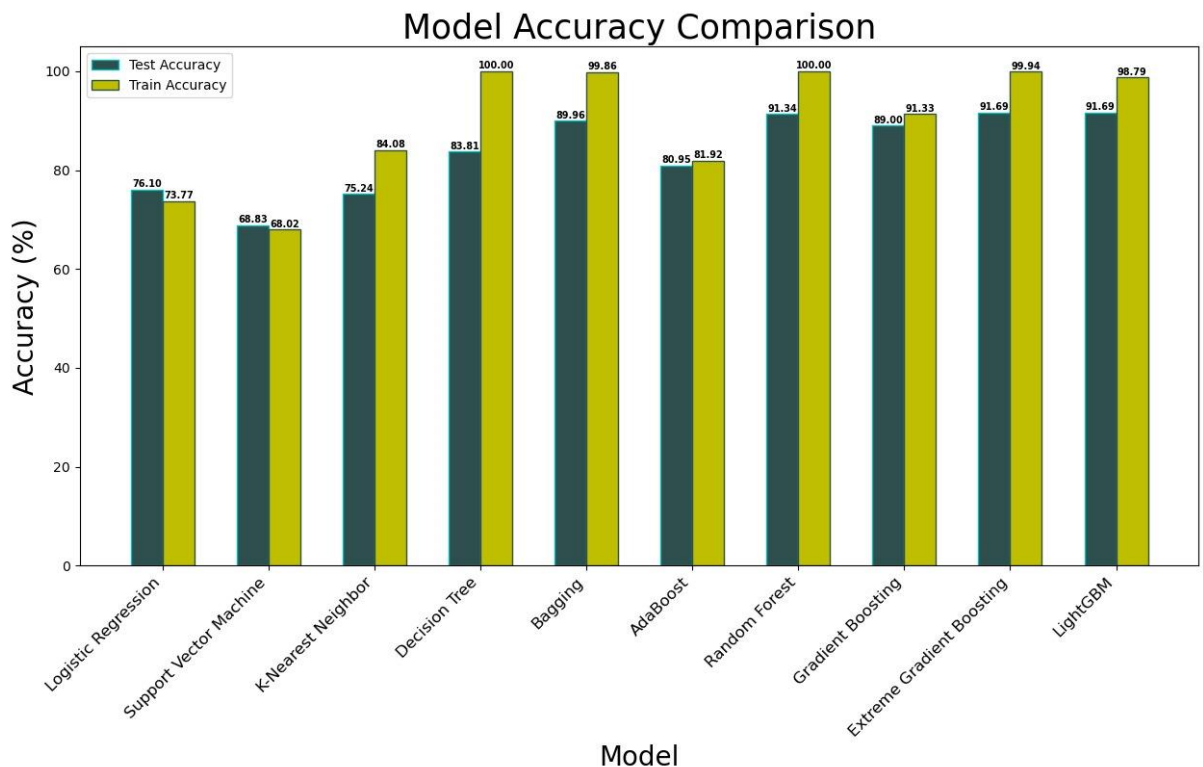
- On the test set, **Random Forest**, **Gradient Boosting**, **Extreme Gradient Boosting (XGBoost)**, and **LightGBM** models achieved high accuracy scores of **91.34%**, **89.00%**, **91.68%**, and **91.68%**, respectively.
- The **Bagging** model showed a good balance with an accuracy of **89.95%** and the highest F1 score of **67.23%** among all models, suggesting robust performance on unseen data.

## Customer Churn Rate Report

- The **Support Vector Machine (SVM)** and **K-Nearest Neighbor (KNN)** models performed poorly on the test set, with low F1 scores of **21.05%** and **43.25%**, respectively.

### Recommendation:

- The **Bagging** model is recommended for deployment as it offers the best balance between accuracy and F1 score, indicating both high precision and recall.
- Additionally, **Gradient Boosting**, **XGBoost**, and **LightGBM** are also strong contenders due to their high accuracy and balanced performance.



### After Hyperparameter Turning

#### Overfitting Indicators:

- Training Set:** Many models, especially **AdaBoost**, **Random Forest**, **Gradient Boosting**, **Extreme Gradient Boosting**, and **LightGBM**, show **100% accuracy**, indicating **overfitting**.
- Test Set:** These same models have lower accuracy, with **LightGBM** being the highest at **94.11%**, showing that the model did not generalize well to unseen data.

#### 2. Model Performance:

- Best Performing Model on Test Data:** **LightGBM** achieved the highest **accuracy (94.11%)**, **precision (84.14%)**, **recall (73.05%)**, and **F1 score (78.21%)** on the test data, making it the **best choice**.
- Gradient Boosting** also performed well with **93.51% accuracy** and an **F1 score of 75.25%**, slightly behind **LightGBM**.

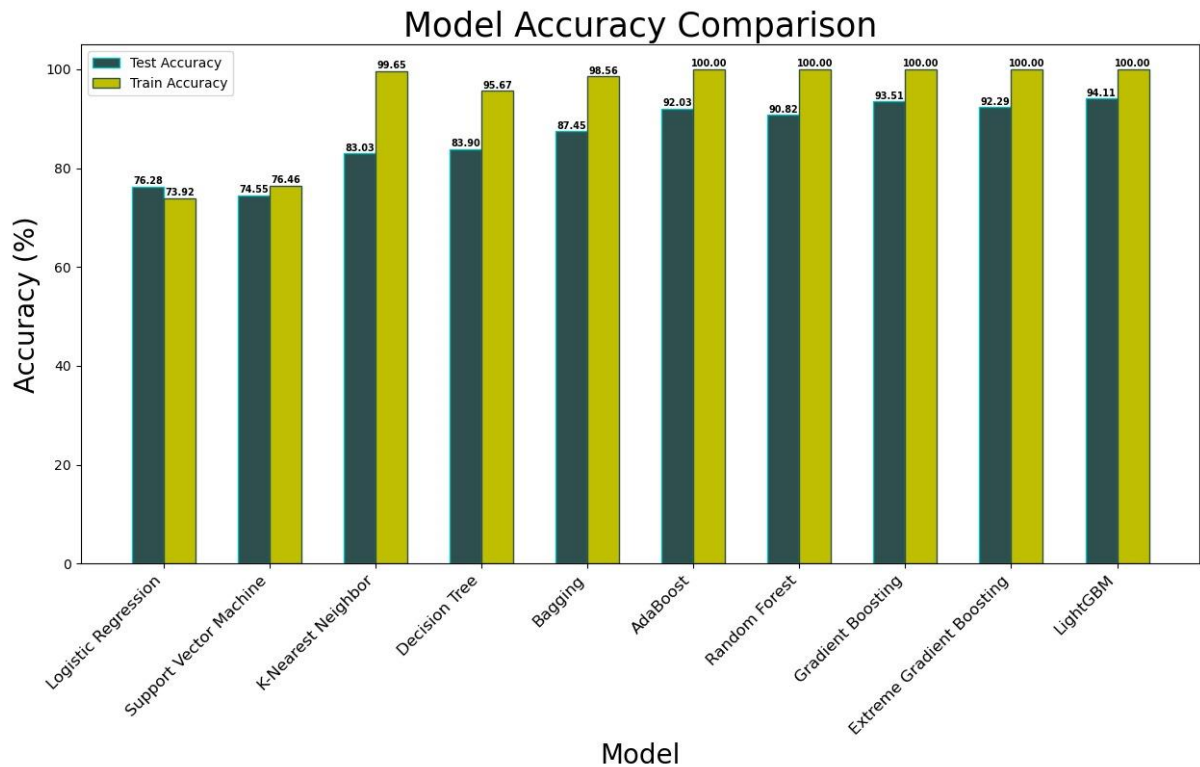
#### 3. Underperforming Models:

## Customer Churn Rate Report

- **Logistic Regression** and **Support Vector Machine (SVM)** have the **lowest precision** and **F1 scores**, indicating they might not be well-suited for this dataset.
- **K-Nearest Neighbor (KNN)** shows a **large performance drop** from **99.65% (train)** to **83.03% (test)** accuracy, highlighting **overfitting**.

### 4. Balanced Performance:

- **Bagging** model shows relatively **balanced metrics** with **87.45% accuracy** and **60.49% F1 score**, suggesting it is more **generalized** than some **ensemble methods**.



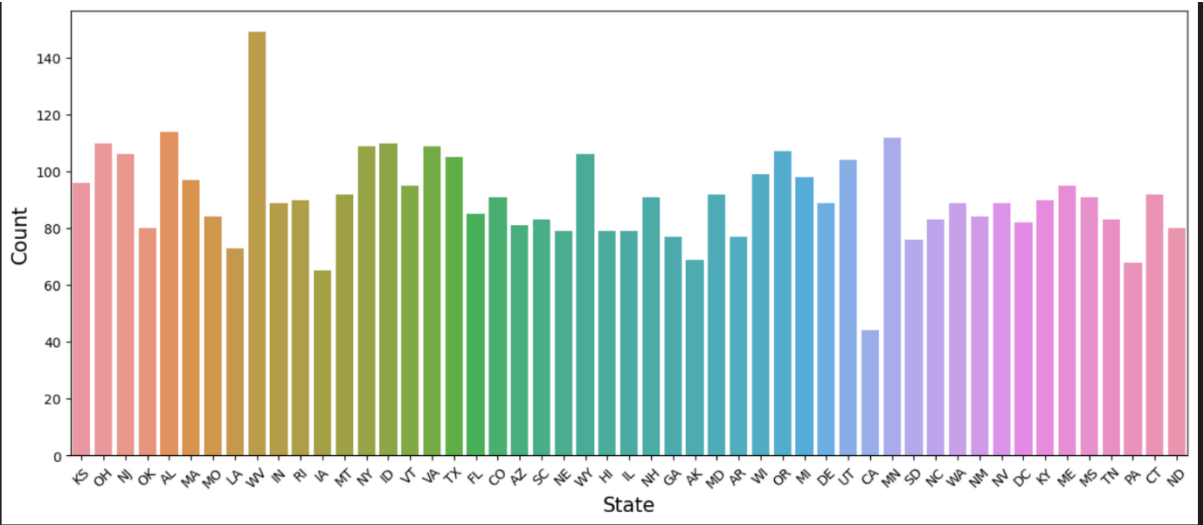
### Recommendation:

The **LightGBM** model is the **best choice** after **hyperparameter tuning** due to:

- **Highest accuracy** on the test set (**94.11%**).
- **Strong precision** and **F1 score**, indicating **balanced performance**.
- **Good generalization**, with a **manageable drop** between **train and test metrics**, avoiding **extreme overfitting**.

Customer Churn Rate Report

Exploratory Data Analysis Report

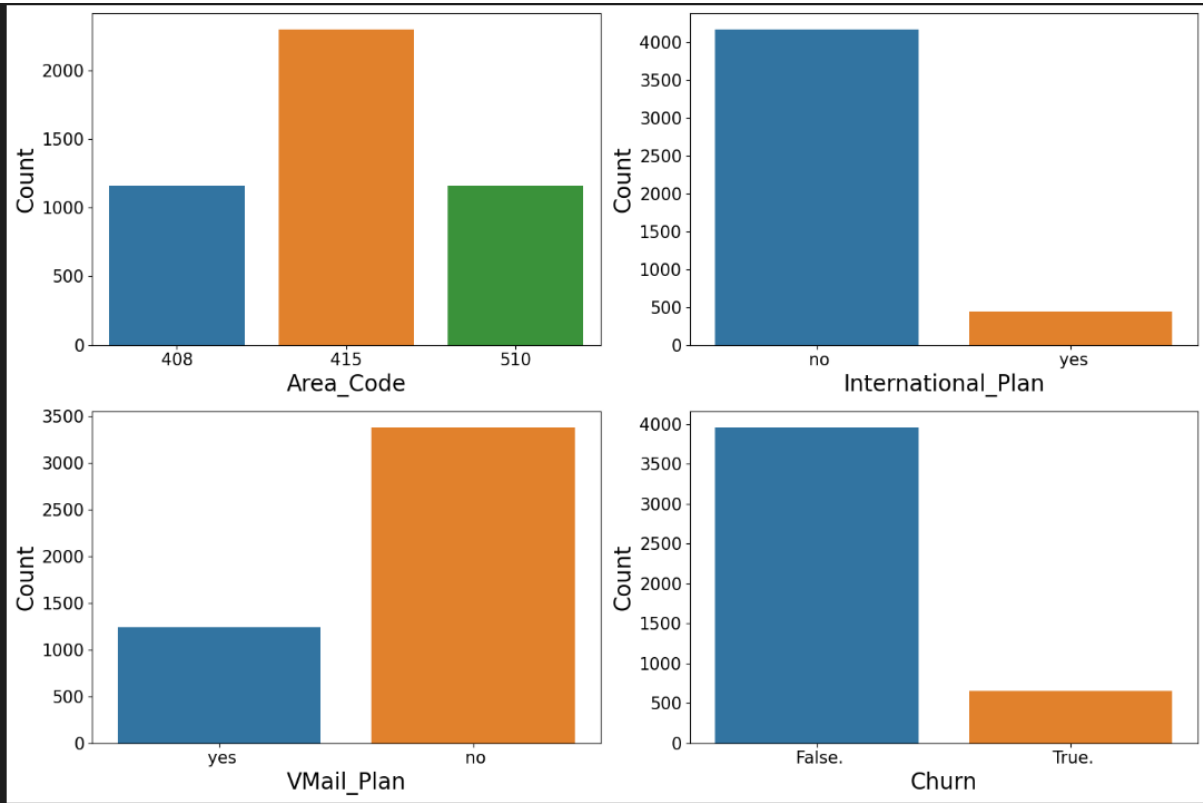


The customer distribution across states shows significant variance.

**WV (West Virginia)** has the highest count of customers.

States like **MO (Missouri)**, **AL (Alabama)**, **OK (Oklahoma)**, **NJ (New Jersey)**, and **OH (Ohio)** also show higher customer counts.

Some states, such as **MT (Montana)** and **DE (Delaware)**, have relatively lower customer representation.



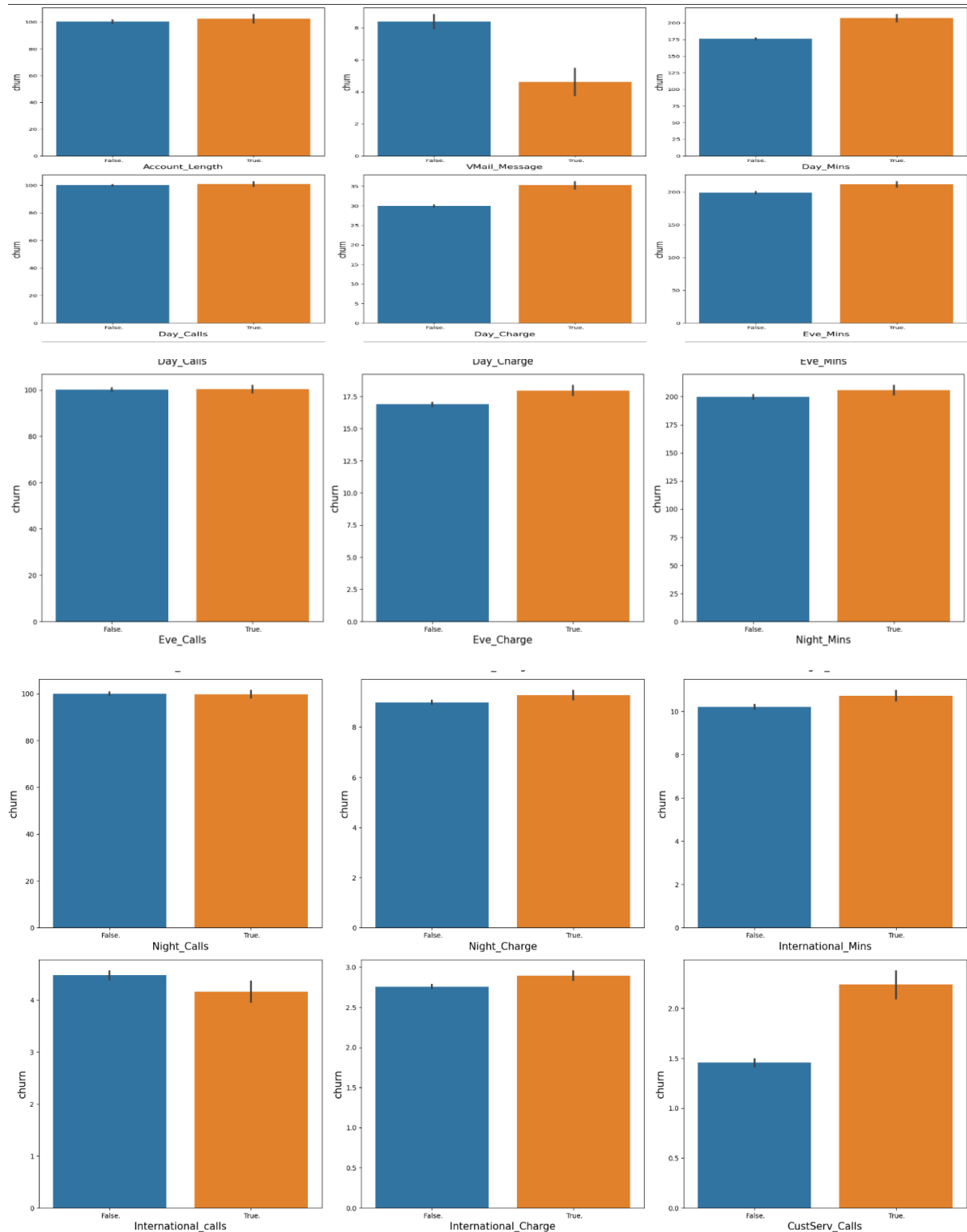
**Area Code Distribution:** Area code 415 has the highest count, while 408 and 510 are lower and almost similar.

## Customer Churn Rate Report

**International Plan:** A large majority of customers do not have an international plan, indicating it might not be widely adopted.

**Voicemail Plan:** Most customers do not have a voicemail plan, suggesting a possible trend or lack of demand for this feature.

**Churn Analysis:** The churn rate is quite low, as most customers have not churned (False), which is a positive indicator for customer retention.



## Customer Churn Rate Report

1. **Account\_Length:** There is no significant difference in churn rate between shorter and longer account lengths, indicating that the duration of a customer's association with the service provider may not be a strong predictor of churn.
2. **VMail\_Message:** Customers with voicemail messages enabled have a lower churn rate. Offering or promoting voicemail services could be a strategy to reduce churn.
3. **Day\_Mins:** Higher day minutes are associated with a higher churn rate. Customers who use the service more during the day might be more likely to leave, possibly due to dissatisfaction with call quality or cost.
4. **Day\_Calls:** Similar to account length, the number of calls made during the day does not significantly affect churn, suggesting that call volume is not a strong churn predictor.
5. **Day\_Charge:** Higher day charges lead to a higher churn rate, aligning with the observation for day minutes. High costs may drive customers away.
6. **Eve\_Mins:** Evening minutes show a similar trend as day minutes, where higher usage corresponds to higher churn. This might indicate that heavy users are more likely to switch to a different provider.
7. **Eve\_Calls vs. Churn**

Similar to day calls, the number of evening calls does not appear to have a strong correlation with churn.

Both categories (True/False) have almost identical churn rates.
8. **Eve\_Charge vs. Churn**

There is a small increase in churn for customers with a higher evening charge (True).

This suggests that high evening charges may contribute to customer dissatisfaction and churn.
9. **Eve\_Mins vs. Churn**

Customers with more evening minutes (True) show a slightly higher churn rate.

This might indicate that heavy evening usage contributes to churn, potentially due to cost concerns.
10. **Night\_Mins vs. Churn**

The churn rate appears slightly higher for customers with more night minutes (True).

This could mean that users with high night-time usage are more likely to churn, possibly due to pricing or service dissatisfaction.
11. **Night\_Calls and Night\_Charge:** There is no clear impact of night calls and charges on churn. Nighttime usage patterns do not seem to influence customer retention.
12. **International\_Mins:** Higher international minutes slightly increase churn, which could suggest that international callers may be looking for better international calling plans elsewhere.
13. **International\_Calls:** Fewer international calls seem to correlate with higher churn, which might indicate that customers with fewer international needs are not finding enough value in the service.

## Customer Churn Rate Report

14. **International\_Charge:** Similar to international minutes, higher charges do not significantly impact churn, showing that pricing for international calls might not be a major factor in customer decisions.
15. **CustServ\_Calls:** A higher number of customer service calls is strongly associated with a higher churn rate. This is a critical insight, as it indicates customer dissatisfaction. Improving customer service quality and issue resolution could help reduce churn.

### Overall Recommendations:

- Focus on reducing charges or offering better plans for heavy day and evening users.
- Improve customer service experience to prevent churn from dissatisfied customers.
- Promote services like voicemail to enhance customer retention.
- Evaluate pricing strategies, especially for heavy users and international callers, to remain competitive.

### Power BI Dashboard



The purpose of this report is to analyze customer churn using the insights provided in the 'Customer Churn Rate Dashboard' generated through Power BI. The analysis aims to identify key factors contributing to churn and provide actionable insights to reduce churn rates.

### Key Metrics

- **Total Customers:** 4,617
- **Total Churned Customers:** 656

## Customer Churn Rate Report

- **Customer Churn Rate:** 14%
- **Average Calls Per Day:** 300
- **Average Charges for International Calls:** 2.78

### Analysis

#### **Churn by Voice Mail Plan**

- Customers without a Voice Mail Plan account for 85% of churn (559 customers), while those with a Voice Mail Plan account for 15% (97 customers).

#### **Churn by International Plan**

- 72% of churned customers do not have an International Plan (470 customers), while 28% do (186 customers).

#### **Churn by State**

- The highest churn is observed in states such as NJ, TX, WA, WV, and MD.
- States with higher call volumes per day include WV, NJ, TX, MD, and WA.

### Conclusion

- The absence of a Voice Mail Plan significantly correlates with higher churn rates.
- Customers without an International Plan are also more likely to churn.
- Specific states demonstrate higher churn, suggesting potential regional issues or opportunities for targeted retention strategies.

### Recommendations

- Introduce promotional offers for Voice Mail Plans and International Plans to reduce churn.
- Develop targeted retention campaigns in high-churn states.
- Enhance customer service and engagement strategies, particularly for regions and plans with higher churn.