C-3DevO / **Churn_in_Telecom_phase_3_project**

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

Churn_in_Telecom_phase_3_project

Public

1 Branch 0 Tags Go to file t Go to file Add file About Code

C-3DevO Final Commit 9948ee9 · 1 minute ago

.ipynb_checkpoints	Finished	1 hour ago
cleaned_data	Finished	1 hour ago
images	Readme done	3 hours ago
original_data	structure and data under...	last week
Power Point Brian ...	Final Commit	1 minute ago
README.md	Finished	1 hour ago
notebook.pdf	Final Commit	1 minute ago
notebook_001.ipynb	Finished	1 hour ago
presentation.pdf	Final Commit	1 minute ago

No description, website, or topics provided.

Readme
Activity
0 stars
0 watching
0 forks

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Languages

README

Jupyter Notebook 100.0%

SyriaTel Customer Churn: A Data-Driven Analysis for Strategic Intervention

Project Overview

This project analyzes customer churn in the telecommunications sector, a critical challenge that directly impacts profitability due to the higher cost of acquiring new customers compared to retaining existing ones. The case study focuses on **SyriaTel**, a telecom provider experiencing ongoing customer losses, with the objective of identifying which customers are most likely to churn and the key factors driving this behavior.

The primary goal is to **predict churn early**, allowing SyriaTel to implement targeted retention strategies such as personalized offers, improved customer support, and customized plan adjustments to reduce customer churn and enhance long-term profitability.

Business Objectives

1. **Predict Customer Churn** – Build machine learning models using account details, usage patterns, and service interactions.
2. **Identify Key Churn Drivers** – Highlight factors like international plans, high usage, and frequent support calls.
3. **Measure Model Effectiveness** – Ensure high predictive accuracy using metrics like **F1 score ≥ 89%** and **precision ≥ 97%**.

4. **Recommend Retention Strategies** – Translate insights into actionable interventions to retain at risk customers.

Data Source

- **Dataset**: SyriaTel Customer Churn Dataset
- **Records**: 3,333 records and 21 features
- **Key Info**: Binary Target Variable(Churn) and customer account details, usage behavior (calls, minutes, charges), and interactions with customer service.

Data Cleaning & Preparation

- Dropped non-predictive columns such as `phone number` (unique per customer, no modeling value)
- Standardized categorical fields (`intl_plan` , `vmail_plan` , `state`) for consistency
- Verified dataset completeness: **No missing values and No duplicates found
- Renamed columns for clarity and modeling i.e. `account length` → `acc_length` , `total day minutes` → `ttl_day_mins`)
- Retained outliers, as removing them would discard important customer usage behavior of heavy callers or frequent customer care contacts.
- Stored `area_code` as a string instead of numeric, since it acts as a categorical label, not a continuous value
- Identified class imbalance in the target variable `churn` (14.5% churn vs. 85.5% retained), requiring special handling during modeling.

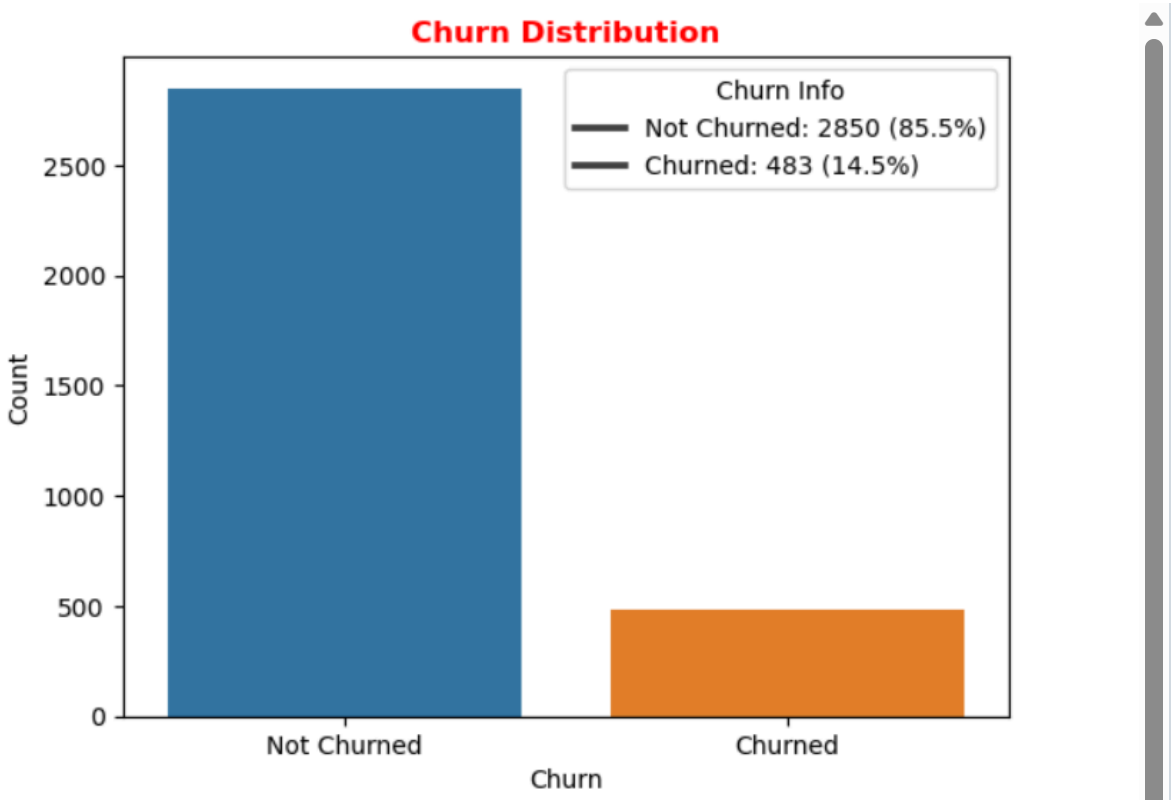
Feature Engineering

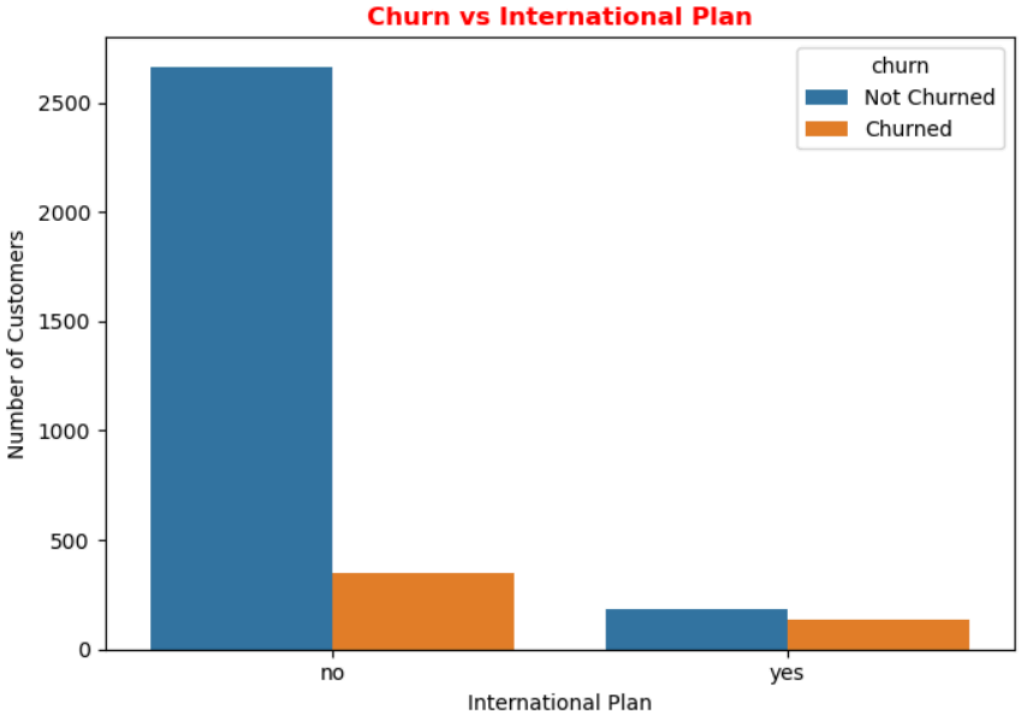
- Created aggregate usage features (total calls, minutes, and charges)
- Derived ratios such as share of international usage, proportion of day calls These engineered features help capture customer behavior patterns and improve churn prediction accuracy.

Key Visualizations

Objective 1: Churn Profiles

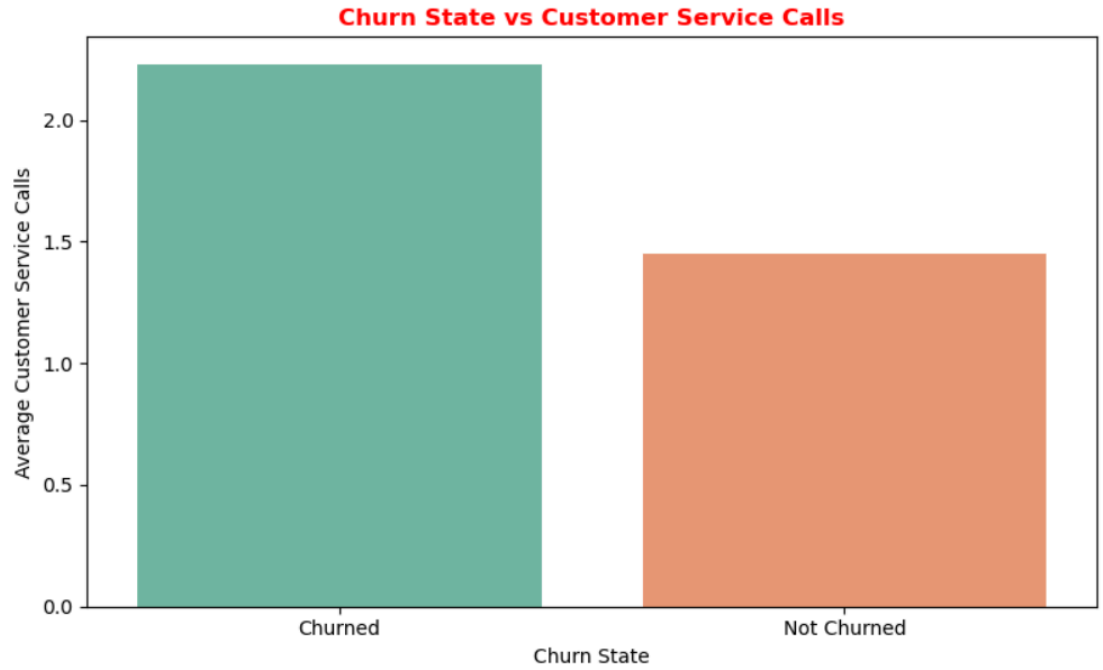
- Overall churn rate: **14.5%** of customers churned, indicating class imbalance (85.5% did not churn).
- Customers with **International Plans** exhibited higher churn rates (~42%), compared to ~11% for those without.

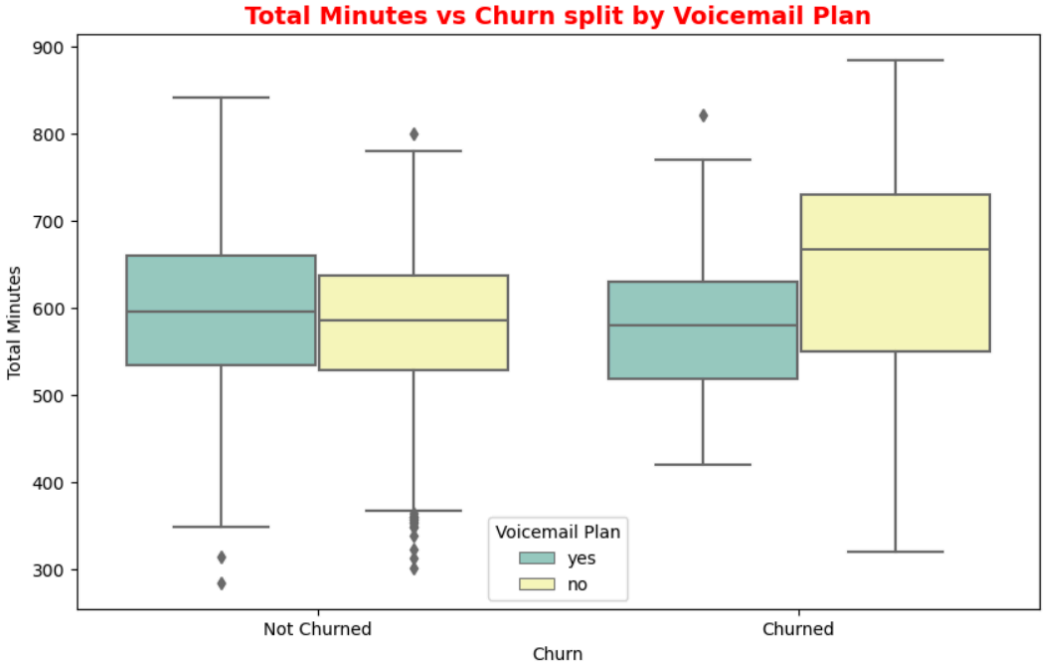

Churn Distribution
Churn Info
Not Churned: 2850 (85.5%)
Churned: 483 (14.5%)
Count
2500
2000
1500
1000
500
0
Not Churned
Churned
Churn



Objective 2: Contributing Factors

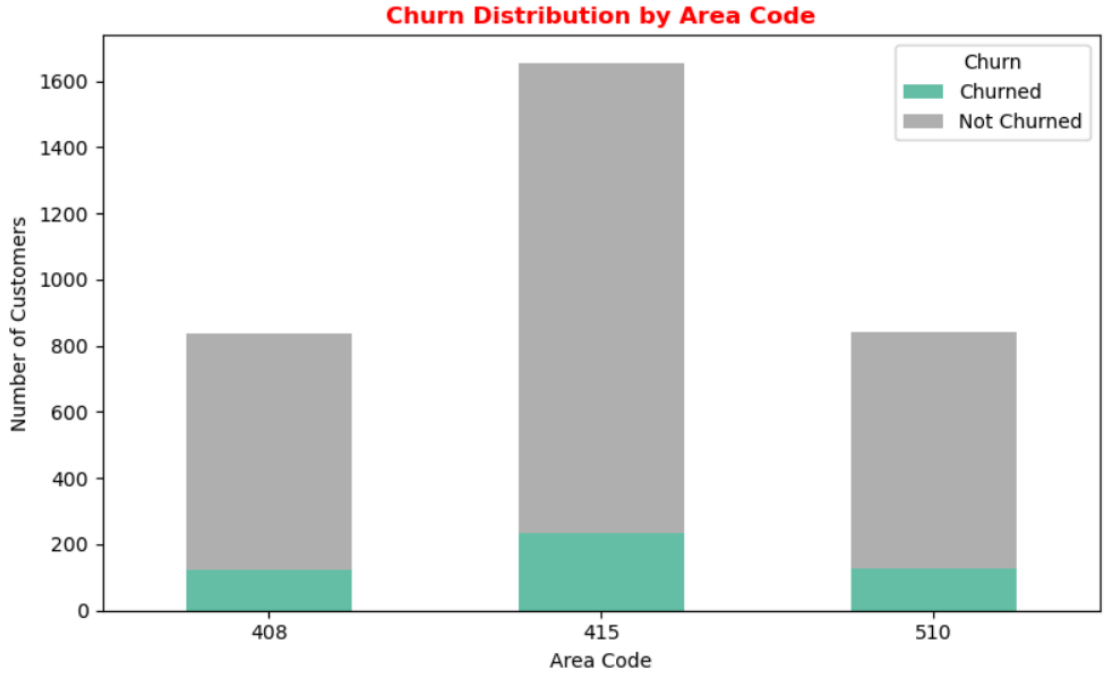
- **Customer service calls** and **high total charges** were strongly associated with churn.
- **Voicemail plan adoption** is associated with lower churn variability and thus customers with voicemail plans tend to be more stable.


Churn State vs Customer Service Calls
Average Customer Service Calls
2.0
1.5
1.0
0.5
0.0
Churned
Not Churned
Churn State


Total Minutes vs Churn split by Voicemail Plan
Total Minutes
900
800
700
600
500
400
300
Voicemail Plan
yes
no
Not Churned
Churned
Churn

Objective 3: Geographic Factors

- Area code 415 has the highest number of customers and elevated churn, while 408 and 510 show lower churn despite similar customer volumes.
- **High-churn states** include New Jersey, Texas, Maryland, and Minnesota, suggesting targeted regional interventions.
- Churn is influenced more by **plan type and service experience** than geography alone.



Tools & Technologies

- **Python 3.11**
- **Pandas, NumPy** – Data processing

- **Seaborn, Matplotlib** – Visualization
- **Jupyter Notebook** – Interactive Analysis
- **Scikit-learn, Statsmodels** – Hypothesis Testing and Modeling
- **Markdown, PDF, PowerPoint** – Reporting and Presentation

Strategic Recommendations

1. Focus Retention on At-Risk Segments

- Identify customers with **International Plans** and those making **frequent customer support calls**.
- Offer proactive outreach, loyalty rewards, or usage based incentives to these high risk groups.

2. Improve Quality of Service and Billing

- Optimize customer service to reduce repeat or unresolved support calls.