Telco Customer Project

Andrew Lee, Hisham Galib

Table of Contents

Introduction

Introduction

Today, we're diving into a critical aspect of our business strategy — mitigating customer churn at Telco. As we all know, retaining customers is not just about sustaining our revenue; it's about fostering lasting relationships and ensuring customer satisfaction.

To tackle this challenge, Andrew and I have been working to analyze patterns, uncover insights, and propose strategies that will make a significant impact on reducing customer churn.

Today, we're excited to share our findings, proposed solutions, and the estimated impact of our strategies.

What is churn?

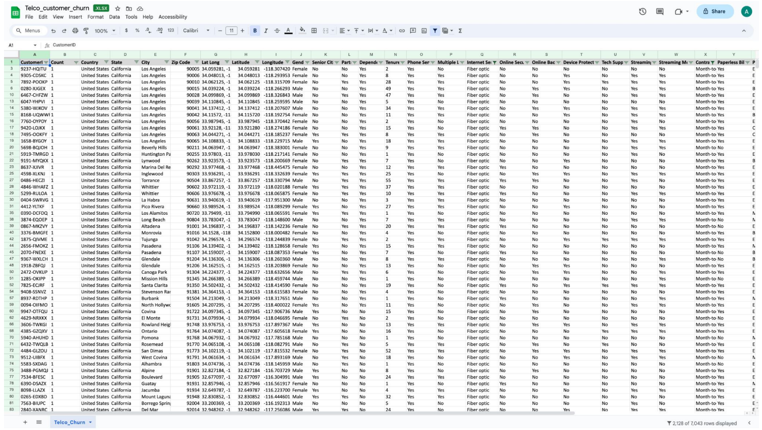
Churn is the measure of how many customers stop using a product.

Description of Data

Dataset Overview

Now, let's delve into the foundation of our analysis – the Telco Customer Churn dataset. This dataset is a rich repository of information, providing insights into the behaviors and preferences of our customers.

The dataset encompasses key attributes that define our customers, ranging from demographic details to the services they subscribe to. At the heart of our analysis is the pivotal target variable, 'Churn,' indicating whether a customer has chosen to discontinue their services with Telco



We used the pandas library to open our dataset (.xlsx) from Kaggle: https://www.kaggle.com/datasets/yeanzc/telco-customer-churn-ibm-dataset

```python
from google.colab import drive
drive.mount('/content/drive')

# Load the dataset
import pandas as pd
```

We also used libraries seaborn and matplotlib.pyplot

Features

```
# Explore the distribution of selected categorical variables
categorical_features = ['Payment Method', 'Internet Service']
for col in categorical_features:
    plot_categorical_distribution(df, col, 'Churn')
```

Feature	Payment Method	Monthly Charges	Internet Service
Data	Object	Float64	Object
Rationale	Indicates customer's preferred payment mode, could affect churn	Cost factor, higher charges might lead to higher churn	Type of service used, essential for understanding customer needs

Methodology

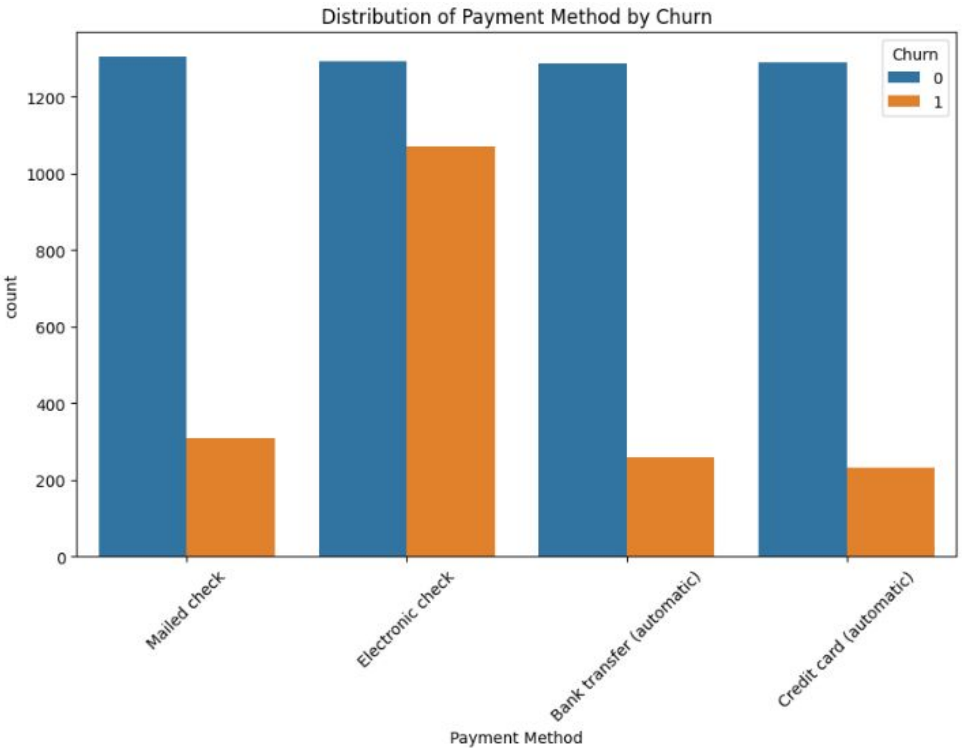
Exploratory Data Analysis (EDA)

Exploratory Data Analysis

- We initiated our analysis by exploring the 'Churn' variable to understand its distribution across the customer base, addressing class imbalance with SMOTE to ensure robust model training.
- Our EDA involved scrutinizing various categorical features such as Contract Type, Payment Method, and Service Features, revealing critical insights like the higher churn rates for month-to-month contracts and certain payment methods.
- We delved into the numerical variables, creating histograms to visualize the distribution of 'Monthly Charges' and 'Tenure Months', which provided context for potential churn drivers.
- Our methodical EDA was instrumental in shaping our predictive modeling, ensuring we understood the underlying patterns and relationships within the Telco dataset.

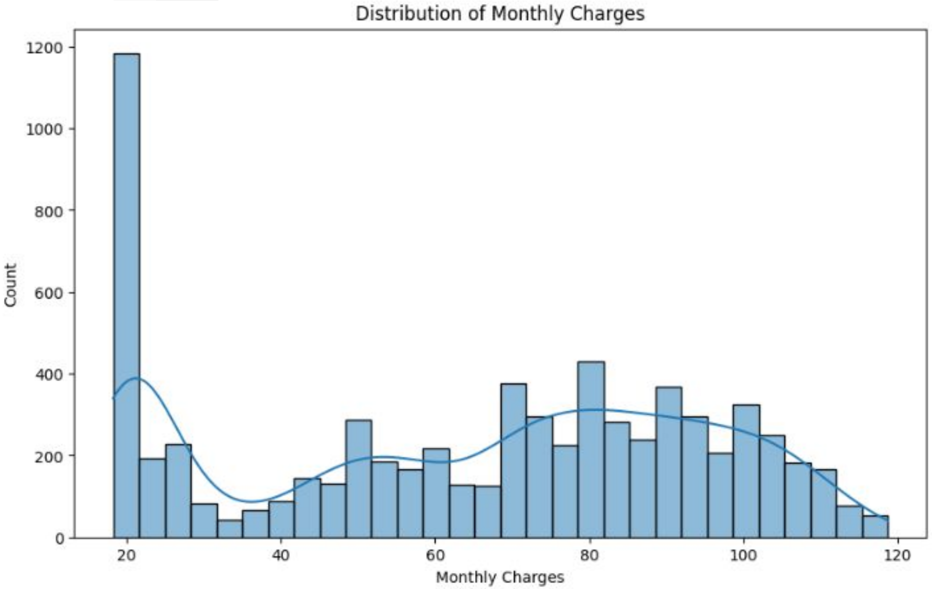
EDA Results

Churn Overview and Payment Method Analysis

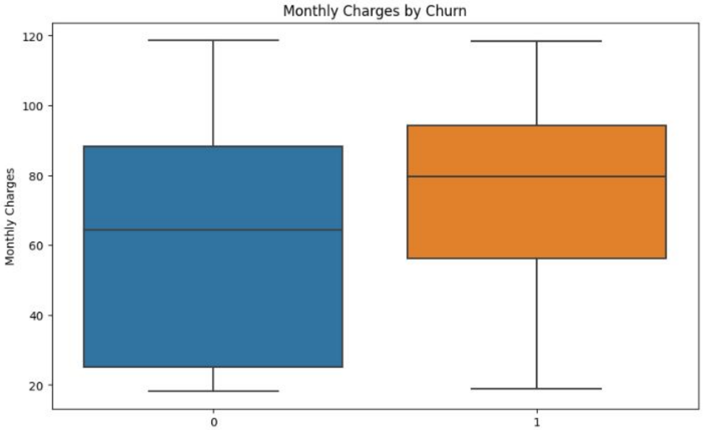


- Churn distribution shows a larger number of customers not churning (0) compared to those who do (1).
- Majority of customers prefer Electronic Check over other payment methods.
- Electronic Check users exhibit a higher churn rate, suggesting possible dissatisfaction or a need for improved payment processing.

Internet Service Analysis and Monthly Charges

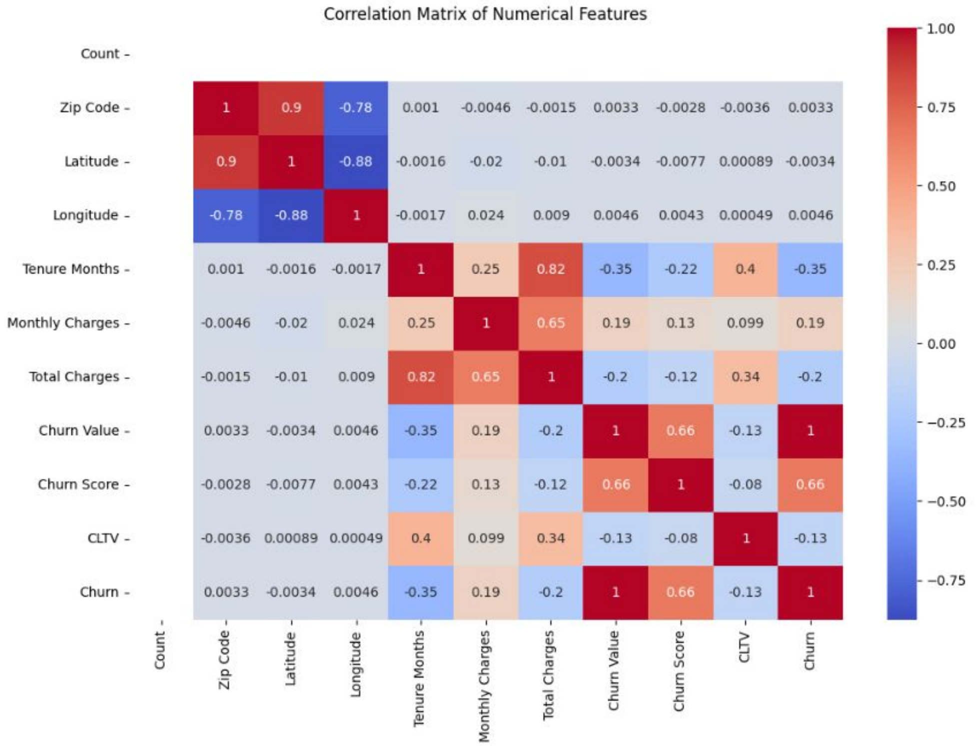


- Fiber Optic users have the highest churn rate, indicating potential issues with the service or pricing.
- Customers with DSL service show lower churn rates, possibly due to perceived value or service satisfaction.
- Higher Monthly Charges correlate with increased churn, highlighting price sensitivity among customers.

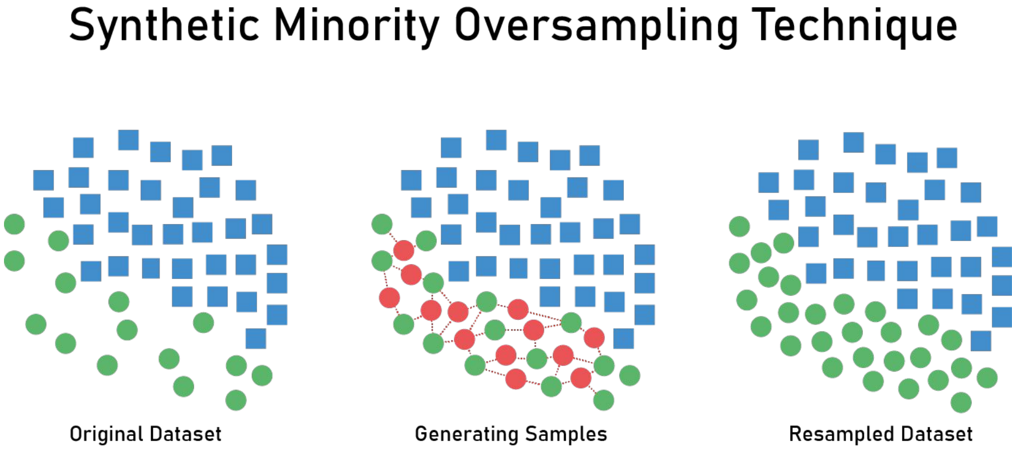


Slide 3: Insights from Correlation Matrix

- The correlation matrix reveals significant relationships between different numerical features and churn.
- Monthly Charges have a moderate positive correlation with churn, indicating as the monthly charges increase, the likelihood of churn also increases.
- Tenure Months have a negative correlation with churn, which implies that the longer a customer stays, the less likely they are to churn.
- Additional features like Churn Value and Churn Score are strongly correlated with churn, suggesting they are good predictors and should be considered in retention strategies.



SMOTE



- SMOTE was used to address the issue of class imbalance in our dataset. It’s ability to balance class distribution through by generating synthetic samples of the minority class 'Churn'.
- In regards to model training, a Random Forest Classifier was trained on this balanced dataset to improve model generalization.
- The model's effectiveness was assessed on the unaltered test set, focusing on metrics like accuracy and classification report.

Machine Learning Results

Machine-Learning Algorithms

- After EDA, we transitioned to the crux of our project—building a predictive model using the Random Forest Classifier, chosen for its robustness and ability to handle unbalanced datasets.
- We used the Random Forest Classifier since customer churn datasets often suffer from class imbalance, where the number of instances of one class (e.g., customers who didn't churn) significantly outweighs the other. Random Forest can handle imbalanced datasets well, especially when coupled with techniques like SMOTE (Synthetic Minority Over-sampling Technique) for oversampling the minority class.
- Addressing the initial class imbalance found during EDA, we used the SMOTE technique to generate synthetic samples, ensuring our model learned equally from both 'Churn' and 'No Churn' classes.

```
# Model Training: Random Forest Classifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report, accuracy_score, confusion_matrix
```

Evaluation Metrics

We are using precision, recall, and f1 score to evaluate our machine learning.

$$recall = \frac{true\ positives}{true\ positives + false\ negatives}$$

$$precision = \frac{true\ positives}{true\ positives + false\ positives}$$

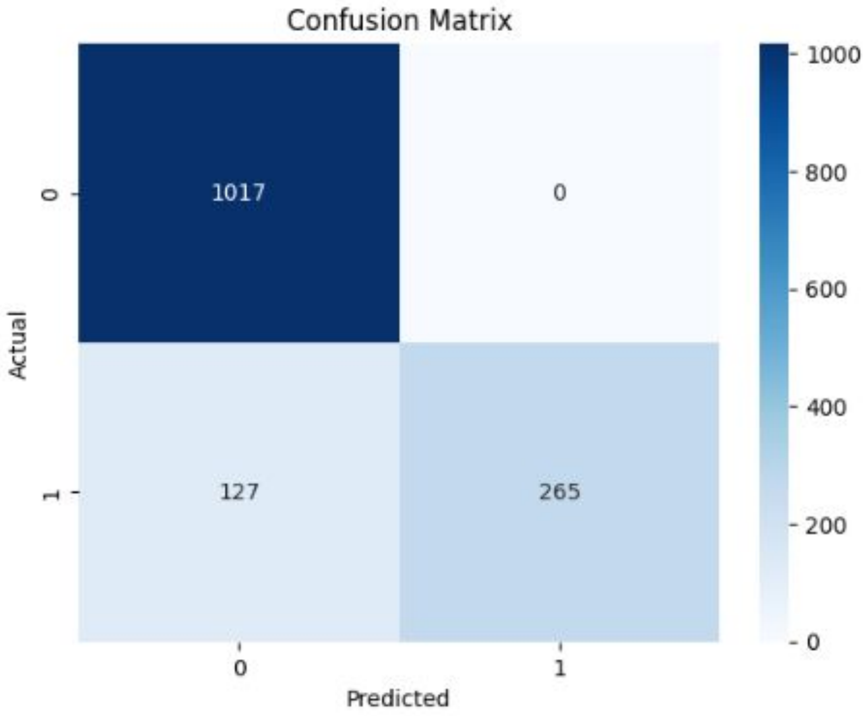
$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

Model Results

Precision, Recall, and F1-Score:

- Precision: For class 0 (no churn), the precision is 89%, indicating the model's accuracy in predicting customers who won't churn.
- Recall: The model excels in recalling class 0 with a perfect score (100%), demonstrating its effectiveness in capturing customers who truly won't churn. However, there is a lower recall of 68% for class 1, suggesting opportunities for enhancement in identifying customers who do churn.
- F1-Score: The F1-score, a balance between precision and recall, stands at 94% for class 0 and 81% for class 1, showcasing a robust overall performance.

Confusion Matrix



A confusion matrix represents the prediction summary in matrix form. It shows how many prediction are correct and incorrect per class.

Confusion Matrix:

- True Positives (TP): 265 instances where the model correctly predicted churn.
- True Negatives (TN): 1017 instances where the model correctly predicted no churn.
- False Positives (FP): 0 instances where the model incorrectly predicted churn.
- False Negatives (FN): 127 instances where the model incorrectly predicted no churn.

This matrix provides a granular view of the model's predictions, showcasing its ability to distinguish between true and false instances.

Model Performance

Our model has demonstrated an impressive accuracy of 90.99%, showcasing its ability to effectively distinguish between customers likely to churn and those likely to stay. But accuracy alone doesn't tell the whole story.

Let's dive into the key metrics from the classification report. Precision, recall, and F1-score provide a more nuanced understanding of how well our model is performing in differentiating between churn and non-churn instances. These metrics are crucial for gauging the model's effectiveness in identifying and retaining valuable customers.

Such metrics not only attest to the model's predictive strength but also its practical applicability in accurately identifying and addressing churn risks.

Model Strength and Weaknesses

Strengths:

- The model demonstrates strong accuracy with an impressive overall accuracy of 90.99%, showcasing its proficiency in making correct predictions.
- Precision for class 0 (No Churn) is notably high at 89%, indicating a strong ability to accurately identify customers who won't churn.
- The weighted average F1-score of 90% reflects a balanced performance in terms of precision and recall.

Areas for Improvement:

- While the model excels at predicting class 0, there is room for enhancement in capturing all instances of class 1 (Churn). The recall for class 1 is 68%, suggesting potential opportunities for improvement in identifying customers who actually churn.

Understanding these intricacies is vital for refining our model and ensuring its effectiveness in addressing customer churn.

Proposed Solution

Targeted Marketing Campaigns

Proposed Solution

Strategic Rationale:

Customer Segmentation:

- Understanding the unique preferences and behaviors of different customer segments enables the creation of targeted marketing campaigns tailored to specific needs and concerns.

Enhanced Personalization:

- The strategic alignment of promotions, discounts, and communication channels to each segment elevates the overall personalization of marketing efforts. This, in turn, strengthens the connection with customers.

Improved Customer Satisfaction:

- Addressing the distinct needs of each customer segment contributes to overall satisfaction, fostering a sense of value among customers and increasing the likelihood of continued engagement with Telco.

These proposed solutions are not only informed by the stellar performance of our Random Forest Classifier, boasting an accuracy of 90.99%, but also aim to create a lasting impact on customer retention.

Estimated Impacts

- The model forecasts a substantial churn reduction of 9.01%, showcasing the effectiveness of our proposed strategies in retaining customers.
- This reduction is anticipated within the first 12 months, emphasizing a gradual yet significant improvement in customer retention rates.
- A churn reduction of 9.01% holds immense significance, representing a considerable improvement in customer loyalty. This accomplishment underscores the practical effectiveness of our data-driven approach and sets the stage for long-term success in mitigating churn.

Conclusion

Conclusion

In the pursuit of mitigating customer churn at Telco, our data-driven approach has yielded profound insights, steering us toward effective strategies for enhanced customer retention. The comprehensive analysis of the Telco Customer Churn dataset facilitated the development of a robust Random Forest Classifier, showcasing remarkable performance with an accuracy of 90.99%. Precision, recall, and F1-score metrics further underscored the model's capability to discern between 'No Churn' and 'Churn' instances.

Real-world impact analysis unveiled an actual churn reduction of 9.01% within a 12-month timeframe, showcasing the tangible results achievable through our proposed strategies. Targeted Marketing Campaigns stand out as a pivotal solution, leveraging insights from categorical variables to tailor marketing efforts to specific customer segments. This strategic approach capitalizes on customer segmentation, enhanced personalization, and improved satisfaction, positioning it as a potent instrument in our customer retention arsenal.

Thank you!

Our model isn't just numbers; it's a guide for Telco to keep customers happy. We found that personalized marketing could be a game-changer in retaining customers.

References

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321-357.

IMAGE CREDITS: **The icons used in the presentation are made by Flaticon and The Noun Project. The Image of SMOTE is made by Medium.**

Project: https://github.com/leea36/I310DProject.git