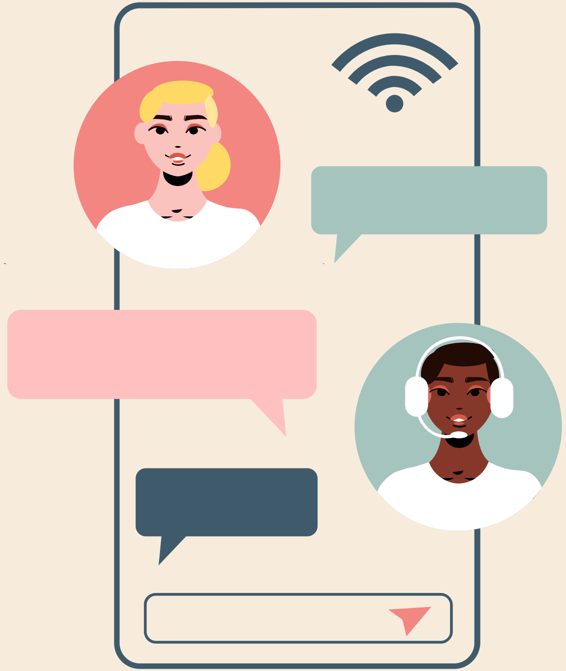


Customer Churn Prediction in the Telecom Industry:

A Data-Driven Approach to Retention Strategies

Business Problem & Objective

Business Question:

"How can we predict customer churn in the telecom industry and enable proactive customer retention?"

Why is it important?

Reduce Revenue Loss
Retaining customers is cheaper than acquiring new ones

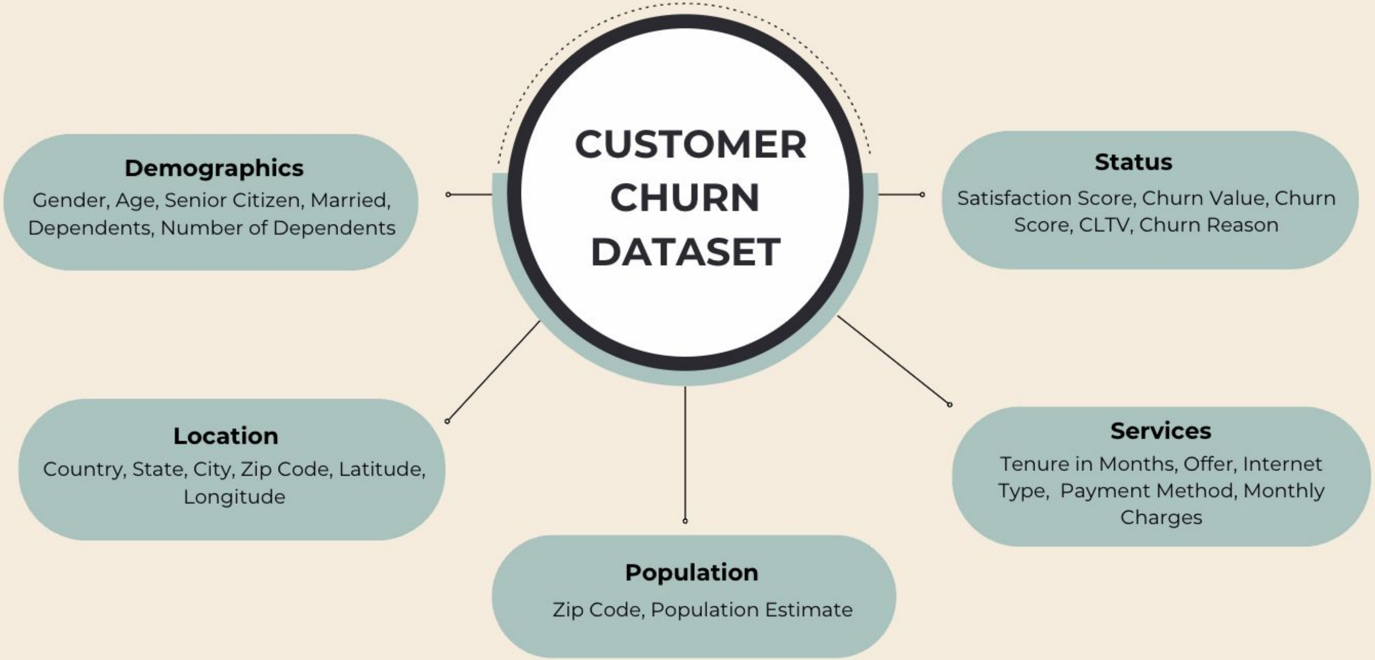
Improve Customer Experience
Identifies pain points (pricing, service issues) to enhance satisfaction.

Optimize Marketing & Retention
Helps target at-risk customers with personalized offers

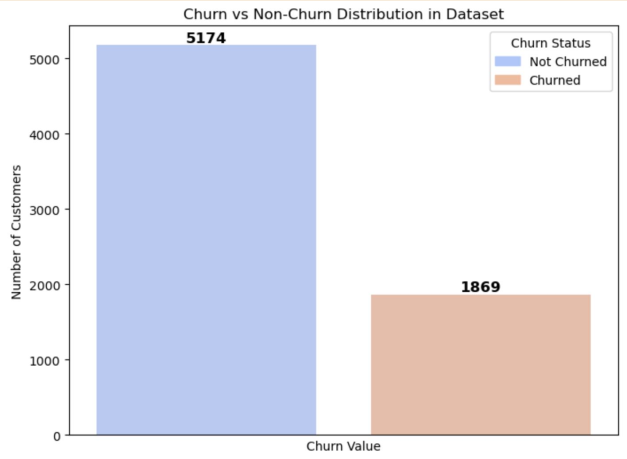
Gain Competitive Advantage
Enables proactive actions to retain customers before they switch to competitors.

Data Overview & Key Features

Source: IBM Telco Customer Churn Dataset *(7043 records, 62 initial features)*

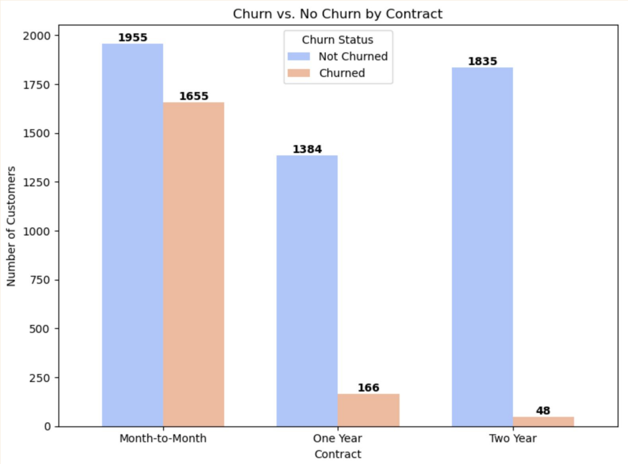


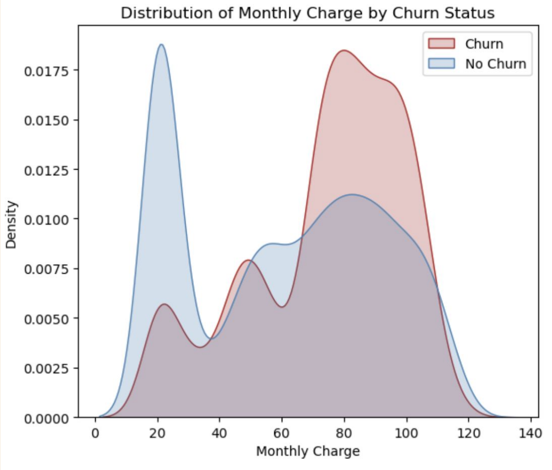
Exploratory Data Analysis



Churn label Distribution Imbalance

Trend in Contract type
Short tenure-> high churn





High monthly prices-> Higher churn

Data Preprocessing

FEATURE SELECTION & REDUCTION

- Started with 62 columns, reduced to 37 columns
- Removed irrelevant features (e.g., Customer IDs, Country, State)
- Dropped redundant features (e.g., ‘Under 30’ covered by ‘Age’)
- Excluded non-predictive features (e.g., Churn Category, Churn Reason)

HANDLING MISSING DATA & OUTLIERS

- Numerical: Mean imputation
- Categorical:Replaced with “Missing” category
- Log Transformation: Reduced skewness in highly skewed features (e.g., Monthly Charges, Tenure)
- Missing flags added for relevant data

DATA TRANSFORMATION

- StandardScaler: Applied to normalize numerical data
- One-Hot Encoding: For nominal categorical variables
- Ordinal Encoding: For Satisfaction Score

Candidate Models

1 **Logistic Regression**

- A simple baseline model to predict if a customer churns (1) or stays (0).
- Helps explain how features like monthly charges or contract type influence churn probability.

2 **Random Forest**

- Combines decision trees to improve accuracy and handle mixed data types.
- Identifies key drivers of churn, like satisfaction score and internet service.

3 **Support Vector Machine (SVM)**

- Effective for separating churners and non-churners in complex data patterns.
- Handles imbalanced datasets well, such as the higher number of non-churners.

4 **Gradient Boosting**

- Focuses on hard-to-predict cases by correcting errors in sequential models.
- Captures subtle patterns, such as churn trends in medium-tenure customers.

5 **LightGBM**

- Optimized for large datasets with fast training and high accuracy.
- Handles imbalanced data effectively, perfect for our churn prediction task.

Model Selection and Evaluation

01 Choosing Candidate Models based on their suitability for classification tasks

02 Hyperparameter Tuning using GridSearchCV to find the best combination of parameters

03 Performing Cross-Validation (10-Fold) to evaluate model Generalization

04 Training and Evaluating Models on Test Data to measure real-world performance

05 Final Model Selection Based on multiple evaluation metrics.

Model Performance Comparison

MODEL	TEST ACCURACY	TEST F1-SCORE	TEST AUC-ROC
RANDOM FOREST	0.954578	0.908832	0.978460
GRADIENT BOOSTING	0.955287	0.911888	0.990790
LIGHTGBM	0.956707	0.914446	0.990944
SVM	0.948190	0.894964	0.985874
LOGISTIC REGRESSION	0.953868	0.908322	0.989865

Feature Importance

Satisfaction Score – The most predictive factor; higher scores correlate with lower churn.

01.

Monthly Charge – Higher charges increase churn, especially with add-ons.

02.

Total Charges – Long-term users churn less; high initial costs raise churn.

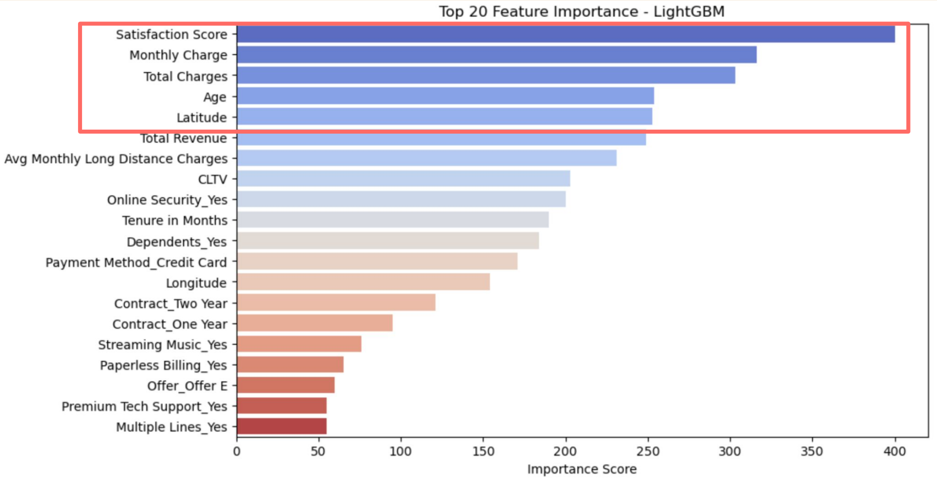
03.

Age – Older customers, often linked to long-term contracts, are less likely to churn compared to younger users.

04.

Latitude – Regional differences affect churn due to service and competition.

05.



Business Insights & Recommendations

Retention Strategies
Target At-Risk Customers, Enhance Customer Support , Loyalty Programs, Proactive Customer Engagement

Service & Experience Improvement
Service Quality Enhancement, Better Customer Segmentation

Market & Pricing Strategies
Diversify Telco Services by Age, Pricing Optimization, Regional Expansion & Customization

Limitations & Future Enhancements

Missing Data – Key features like Offer and Internet Type had missing values, requiring imputation.

Class Imbalance – Churn data was imbalanced, requiring resampling techniques for better model performance.

Potential Overfitting – Some models showed high performance on training data, requiring cross-validation & tuning.

External Factors Excluded – Factors like competitor pricing, customer sentiment, were not available in the dataset.

Using **time-series** churn analysis to track customer behavior trends.

Incorporate customer reviews & feedback for better **sentiment-based** churn prediction.

Leverage Churn Reason data to build a **multi-class classification model**, identifying why customers churn.

Thank You

Team 13