Capstone Project: Increasing user retention by Predicting Churn

Telecom companies face a critical challenge: customer churn. This presentation outlines our strategy to predict and proactively prevent customer attrition, ensuring sustained revenue and customer loyalty.

By :

Indrapal Singh (Roll Number :30)

Anuj Soni (Roll Number :29)

Khushi Dasra (Roll Number :31)

Devdatt Mishra (Roll Number :28)

K. Yukti (Roll Number :32)

Under the Guidance of :

Mr. Ishan Marwah

Conducted by :

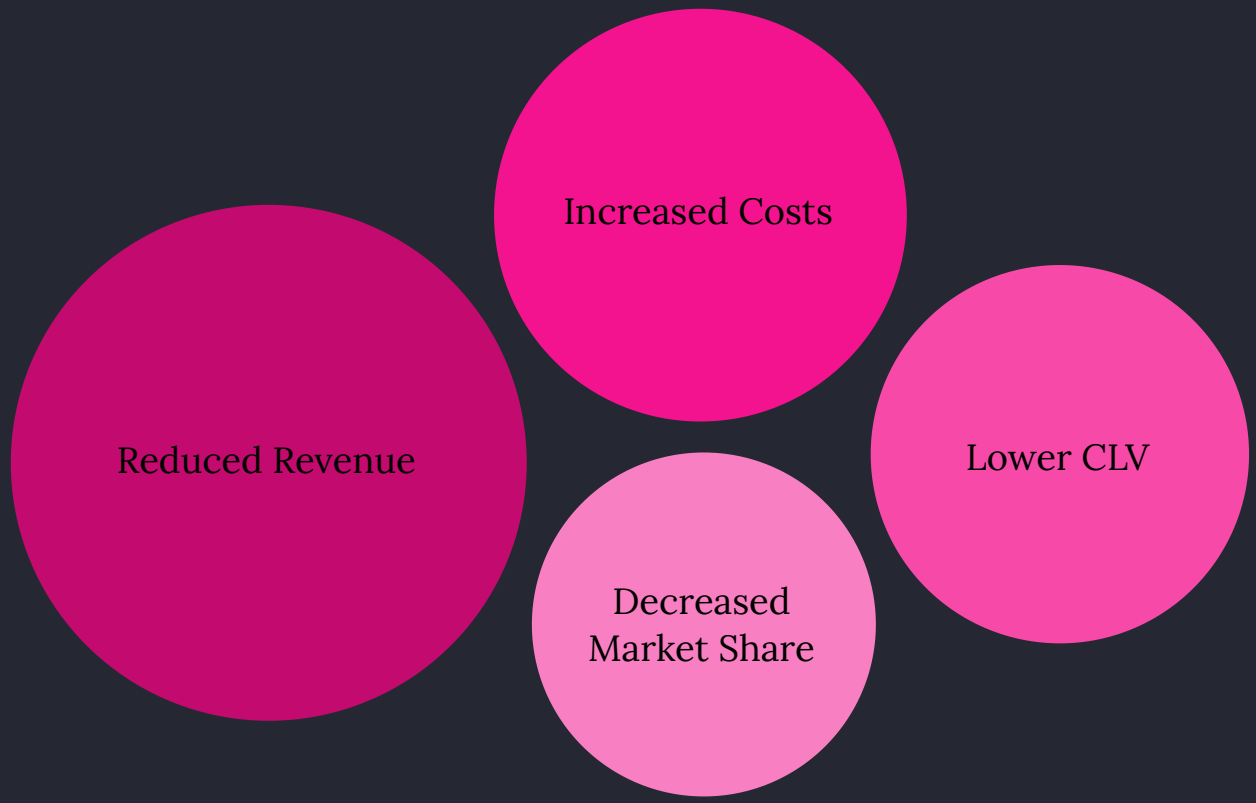
DSSC

PROBLEM STATEMENT

Understanding the Impact of Customer Churn

Customer churn significantly erodes revenue and market share for telecom providers. Identifying at-risk customers is paramount for implementing targeted retention strategies.

- Direct revenue loss from departing customers.
- Increased acquisition costs to replace lost subscribers.
- Negative impact on brand perception and customer lifetime value.



Our objective is to develop a robust machine learning model capable of accurately classifying potential churners versus non-churners, enabling timely intervention.

DATA & APPROACH

Building the Foundation: Data Integration and Feature Engineering

Comprehensive Data Merge

We integrated four distinct datasets: customer demographics, usage patterns, complaint records, and billing history to form a holistic view of each subscriber.

Master Dataset Creation

A master dataset was constructed comprising 8 numerical and 2 categorical features, carefully selected to capture critical indicators of churn.

Strategic Feature Engineering

Key features were engineered, including complaint metrics (frequency, type), detailed usage patterns (data, call minutes), and billing history (average spend, payment delays).

Robust Train-Test Split

The master dataset was split into an 80-20 ratio for training and testing, ensuring the model's ability to generalize to unseen data.

MODEL DEVELOPMENT

Selecting the Optimal Predictive Model

We evaluated three prominent machine learning algorithms to identify the most effective for churn prediction:

- Logistic Regression: A baseline for binary classification.
- Decision Tree: Interpretable, captures non-linear relationships.
- Random Forest: Ensemble method for enhanced accuracy and robustness.

Hyperparameter tuning with GridSearchCV was crucial for optimizing model performance.

The Random Forest model emerged as the top performer, achieving a **72% cross-validation accuracy** and an impressive **80% accuracy** on the prediction dataset with optimized parameters (n_estimators=200, max_depth=5).

TECH STACK & DEPLOYMENT

From Model to Interactive Application

Python Ecosystem

Leveraged **pandas** for data manipulation, **numpy** for numerical operations, and **scikit-learn** for machine learning.

Machine Learning Core

Implemented **RandomForestClassifier** with **GridSearchCV** for robust model training and optimization.

Streamlit Web Application

Deployed an intuitive and interactive web application using **Streamlit** for real-time churn predictions.

Model Serialization

Utilized **joblib** for efficient serialization and deserialization of the trained machine learning model.

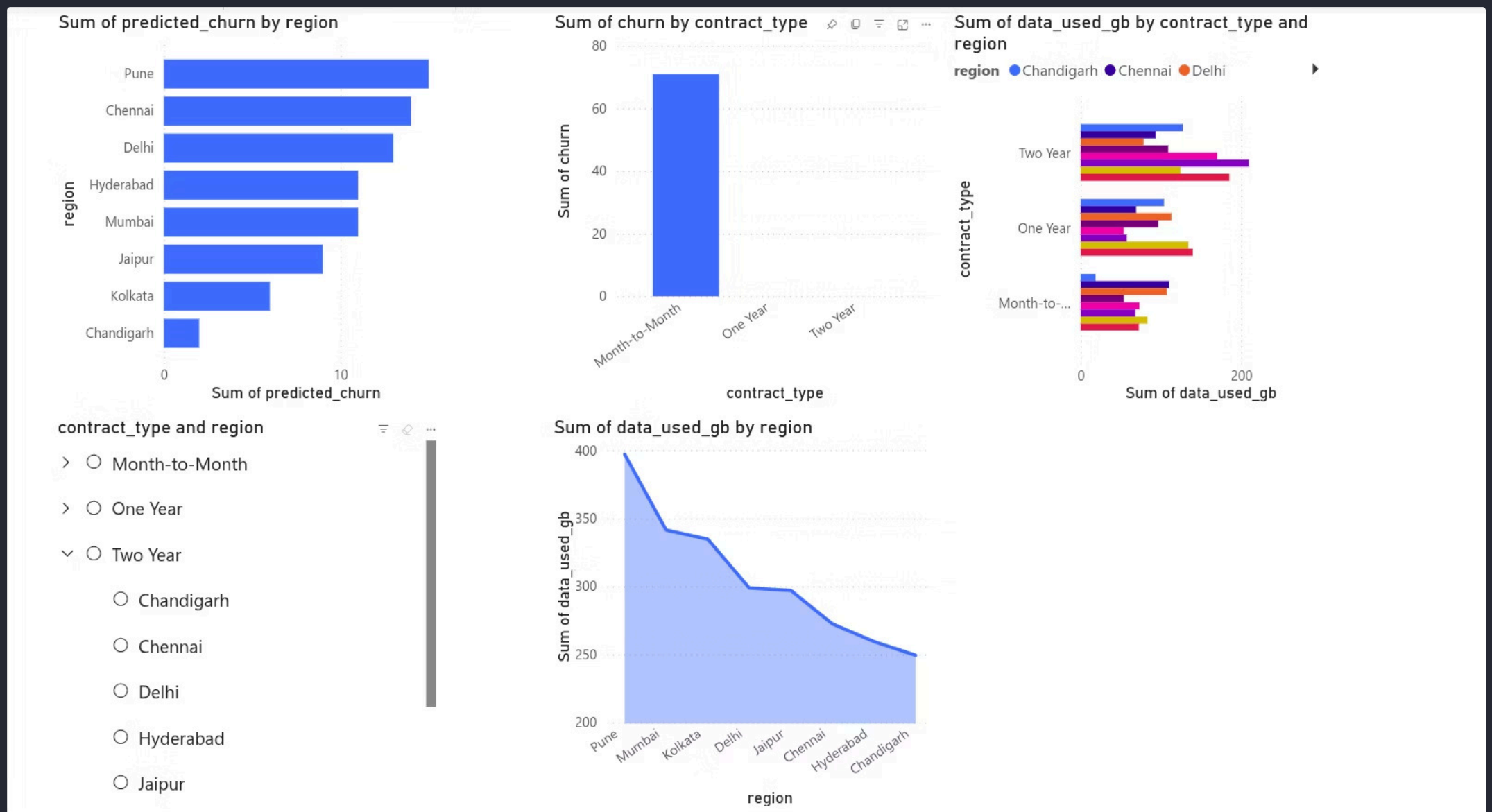
The interactive Streamlit dashboard enables stakeholders to input customer data and receive instant churn probability scores, enhancing decision-making.

INTERACTIVE DASHBOARD

Visualizing Churn: Our Power BI Dashboard

Our interactive Power BI dashboard provides a dynamic view of churn prediction. Users can explore key metrics, filter by customer segments, and understand the driving factors behind potential churn.

This dashboard empowers business users to proactively identify at-risk customers, allowing for timely and targeted interventions to improve retention rates and safeguard revenue.



Thank You!