CUSTOMER CHURN PREDICTION

Presentation by :

Hemanth Nayak

INTRODUCTION

- **Customer churn is the term used to describe the phenomena where customers discontinue utilizing a business's goods or services after a certain amount of time. Customer retention is essential for long-term success for online services, telecoms, and subscription-based enterprises. By anticipating customer attrition, businesses may proactively identify at-risk clients and implement focused retention strategies.**

- **Utilizing machine learning and data analytics, customer churn prediction examines past customer behavior, use trends, grievances, and other pertinent variables. With this data, businesses may train predictive models to anticipate the probability of a client leaving in the near future.**

- **Businesses can apply prompt interventions, like tailored offers, better service, or increased assistance, by identifying possible churners early on. This eventually lowers revenue loss and raises customer satisfaction.**

PROBLEM STATEMENT

- **Goal:** Identify customers likely to leave a subscription service.
  Steps:
- Load customer data (usage, support tickets, demographics).
- Feature engineering (e.g., average monthly usage, complaints).
- Encode categorical variables.
- Train models (Logistic Regression, XGBoost).
- Evaluate using F1-score and confusion matrix.
- Recommend retention strategies.

PREPARE

PROJECT PROCESS

DATA SET

To determine whether customers are about to quit their subscription, we utilized a structured dataset with columns such as customer ID, gender, region, tenure months, plan type, total usage, and churn rate.

DATA CLEANING AND PREPROCESSING

- We have used python for data cleaning, handling missing values and duplicates.
- Used python for performing Exploratory Data Analysis (EDA), encode categorical variables, train models and evaluate scores and generate matrices.

DATA VISUALIZATION AND DASHBOARDS

We have used POWER BI and python to build visualizations and create dashboards.

Logistic Regression

By training the models using Logistic Regression and XGBoost , following results were generated.

- F1 Score : 0.886
- XGBOOST F1 : 0.8798
- Confusion Matrix : [[16 596]
  [44 2344]]

Detecting Churn :

- True Positives (TP) = 2344 Out of 2388 actual churners, 2344 were correctly predicted.
- Recall = 2344 / (2344 + 44) ≈ 98.16%
- We can say that the predicted churn is very likely true.
- False Positives (FP) = 596. These are 596 users predicted to churn but actually stayed.
- Precision = 2344 / (2344 + 596) ≈ 79.76%
- Insight: About 20% of predicted churners are false alarms
- Running expensive retention campaigns (e.g., discounts, calls) based on our prediction may waste resources on customers who would have stayed anyway.
- Actual churners: 2388 (about 80% of the dataset)
- Actual non-churners: 612 (about 20%)

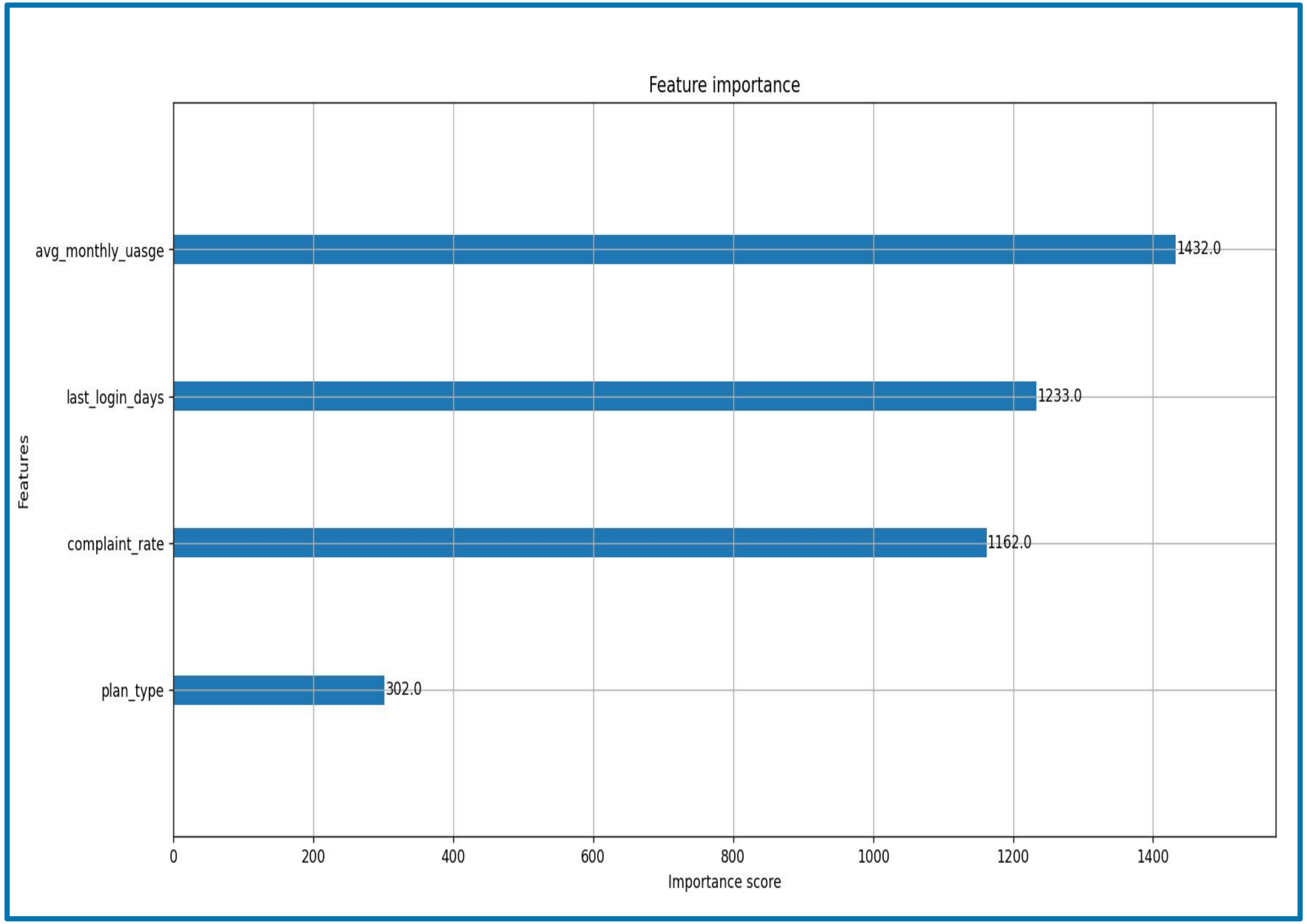
Feature Importance on churn

1. Average monthly usage – Most important
Importance score : 1432
- This is the most important feature. It is likely that the user's monthly usage frequency has the biggest impact on the model's forecasts (e.g., for churn, satisfaction, or engagement).

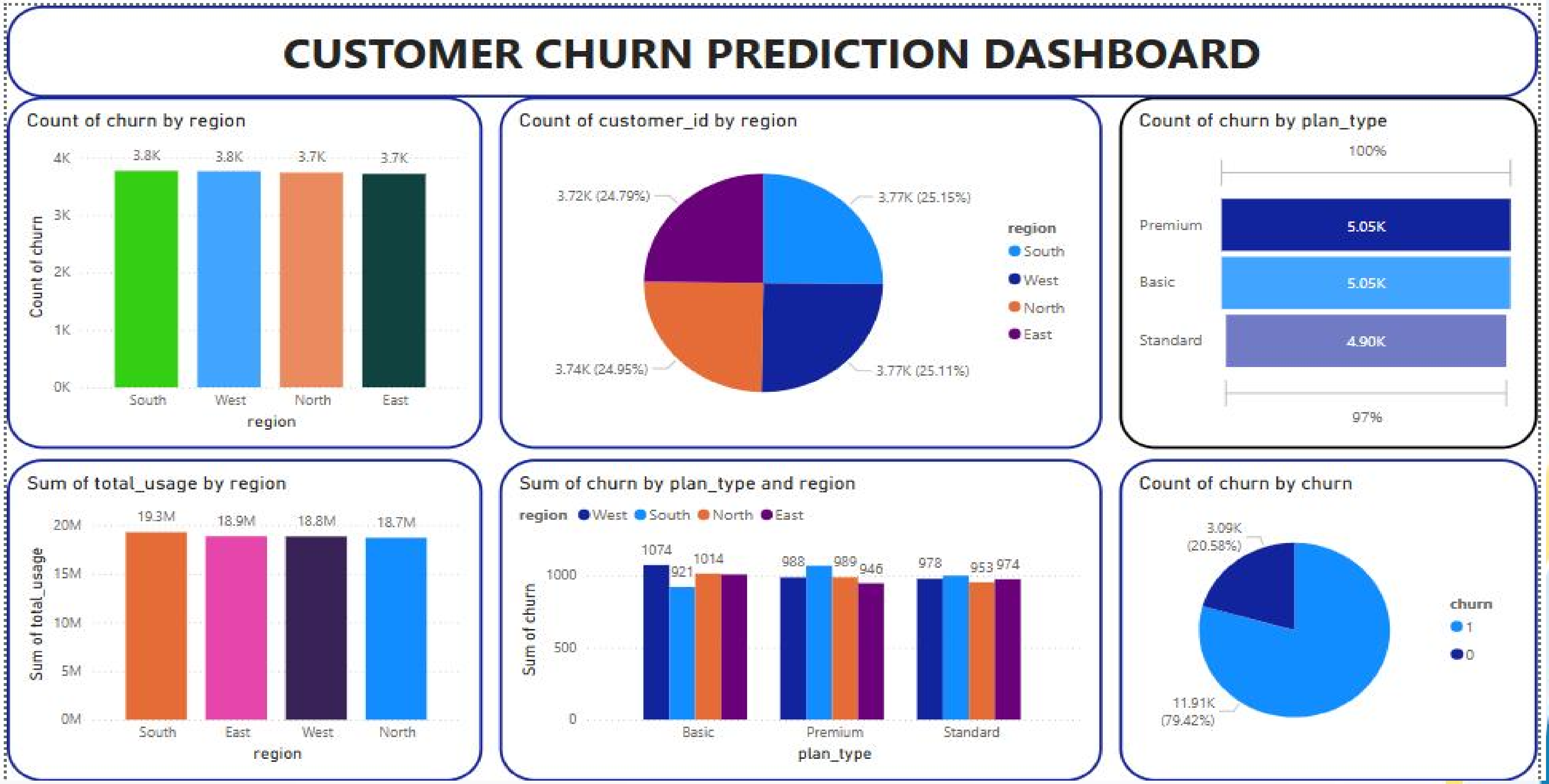
2. Last login days – Second most important
Importance score : 1233
- Indicates that recent user activity (e.g., how many days since last login) is a strong predictor.
- May signal user disengagement or satisfaction level.

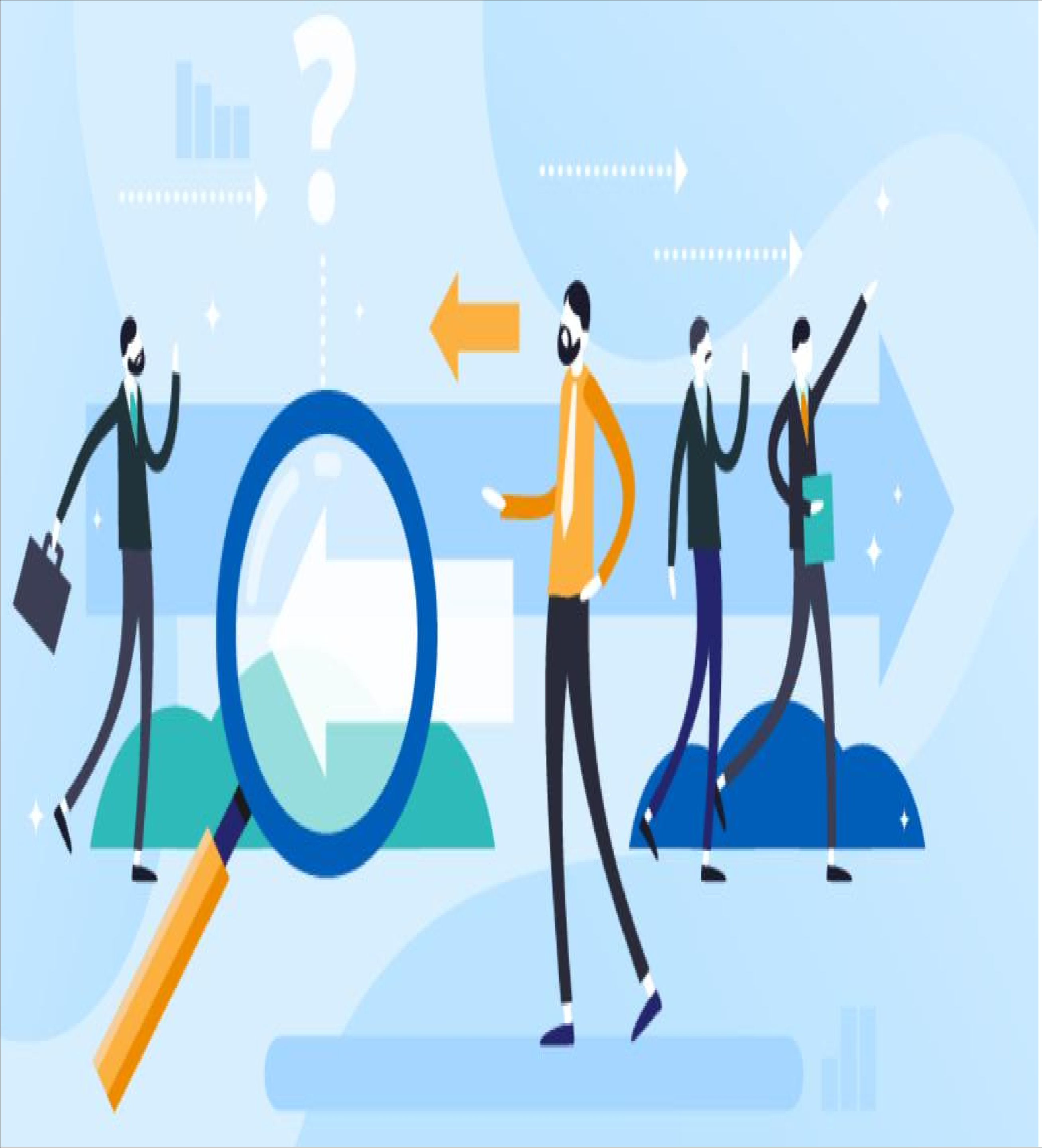
3. Complaint rate
Importance score : 1162
- Suggests user complaints are also highly relevant—potentially a strong signal of dissatisfaction or risk of churn.

4. Plan type – Least important
Importance score : 302
- Indicates that the type of plan a user is on contributes the least to the model's decisions.
- It may still matter, but not as much as behavior-based features like usage and login history.



Data Visualization and Dashboard (Power BI)





Key Insights

- User behavior metrics (usage frequency and recent activity) are more predictive than demographic or static features like plan type.
- High complaint rates are strongly associated with the target outcome, implying the importance of monitoring customer feedback.
- Plan type might not need as much attention for predictive modeling, but could still be relevant for strategic segmentation.

Final Thoughts

- The analysis of churn prediction reveals important details about consumer behavior. When predicting loss of customers, key metrics like average monthly consumption, last login activity, and complaint rate have the biggest impact. This emphasizes how crucial user experience, customer involvement, and service dependability are to keeping telecom customers.

- Understanding the factors that contribute to customer churn is crucial for preserving market share and profitability in a sector with fierce competition and low switching costs. By using predictive analytics, telecom companies may take action before clients depart, which lowers attrition and raises customer lifetime value.

Retention Strategies

- Target Users with Low Usage : Provide discounts, bundle offers, or use advice.
- Engage Inactive Users Again : Send offers or communications while there is no activity on the login.
- Respond to Users with High Complaints : Give assistance top priority and extend kindness.
- Efforts to Maintain Segments : To save money on false positives, use confidence scores.
- Programs for Behavior-Based Loyalty : Reward engagement and consistent use rather than just plan type.
- Automate Alerts for Churn : Send high-risk consumers CRM actions (emails, calls).
- Update the model frequently : Refine using updated data, and monitor the success of campaigns.

THANK YOU