Customer Churn Prediction Project

Author: Aarohi Bhatt

Date: [22/12/2024]

1. Introduction

What is Customer Churn?

Customer churn means customers leaving a service (like a telecom company). Predicting churn helps companies keep customers by offering them discounts or better plans. Customer churn is defined as when customers or subscribers discontinue doing business with a firm or service.

Customers in the telecom industry can choose from a variety of service providers and actively switch from one to the next. The telecommunications business has an annual churn rate of 15-25 percent in this highly competitive market.

Individualized customer retention is tough because most firms have a large number of customers and can't afford to devote much time to each of them. The costs would be too great, outweighing the additional revenue. However, if a corporation could forecast which customers are likely to leave ahead of time, it could focus customer retention efforts only on these "high risk" clients. The ultimate goal is to expand its coverage area and retrieve more customers loyalty. The core to succeed in this market lies in the customer itself.

Customer churn is a critical metric because it is much less expensive to retain existing customers than it is to acquire new customers.

To reduce customer churn, telecom companies need to predict which customers are at high risk of churn.

To detect early signs of potential churn, one must first develop a holistic view of the customers and their interactions across numerous channels, including store/branch visits, product purchase histories, customer service calls, Web-based transactions, and social media interactions, to mention a few.

As a result, by addressing churn, these businesses may not only preserve their market position, but also grow and thrive. More customers they have in their network, the lower the cost of initiation and the larger the profit. As a result, the company's key focus for success is reducing client attrition and implementing effective retention strategy.

Objectives

I will explore the data and try to answer some questions like:

- What's the % of Churn Customers and customers that keep in with the active services?
- Is there any patterns in Churn Customers based on the gender?
- Is there any patterns/preference in Churn Customers based on the type of service provided?
- What's the most profitable service types?
- Which features and services are most profitable?
- Many more questions that will arise during the analysis

2. Dataset Used

-Source: [Kaggle Telecom Dataset]
(https://www.kaggle.com/datasets/blastchar/telcocustomer-churn)

-Sample Data:

CustomerID	Gender	Contract	Tenure	MonthlyCharges	Churn
001	Male	Monthly	5	70	Yes
002	Female	Yearly	24	50	No

(This is a small example. Real data has 7,000+ rows.)

Features:

- **Demographics** (gender, age, partner status).
- **Account info** (tenure, contract type, payment method).
- **Services used** (Internet, phone, streaming).
- **Target Variable**: Churn (Yes/No).

3. Steps Followed

Step 1: Clean the Data

- Removed empty rows.

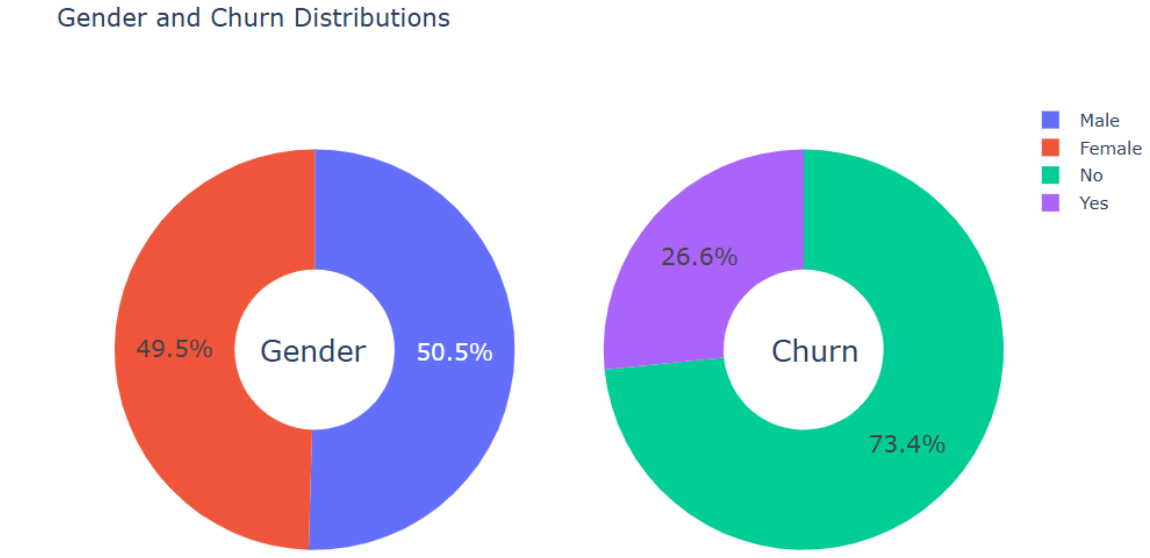
- Converted words (like "Yes"/"No") to numbers (1/0) for the computer to understand.

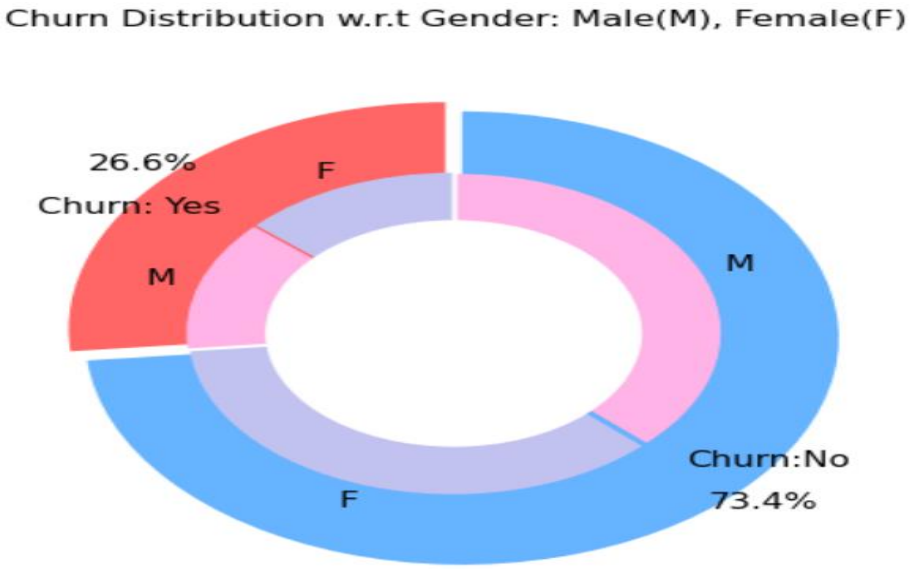
Step 2: Analyze the Data

- Found interesting patterns:

  - “Monthly contract” customers churn more than yearly ones.

  - Customers paying “high monthly charges” are more likely to leave.

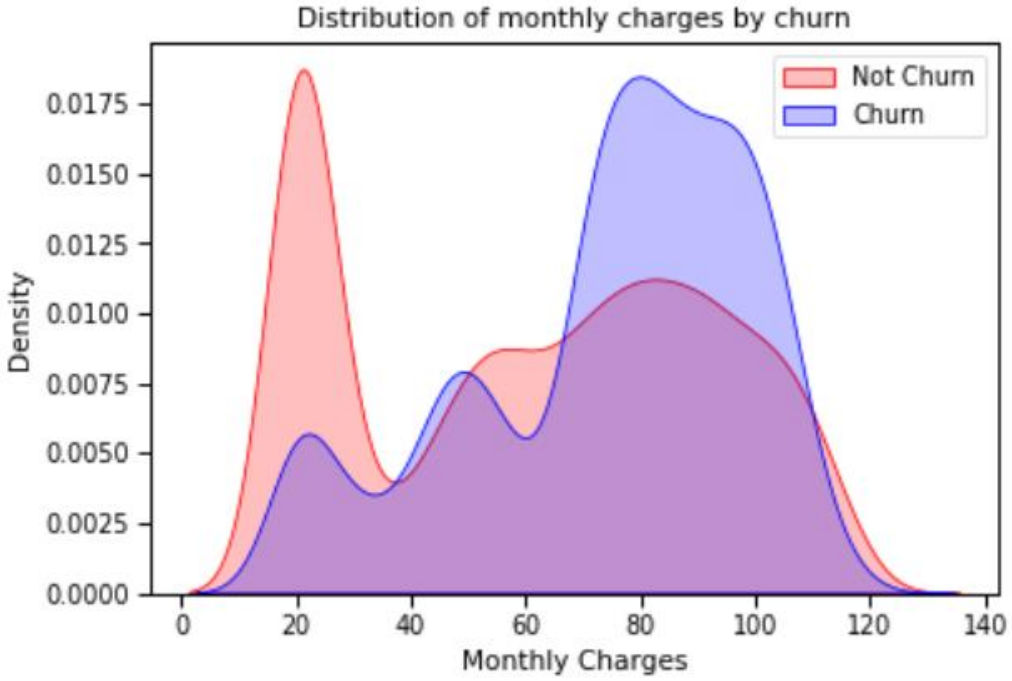


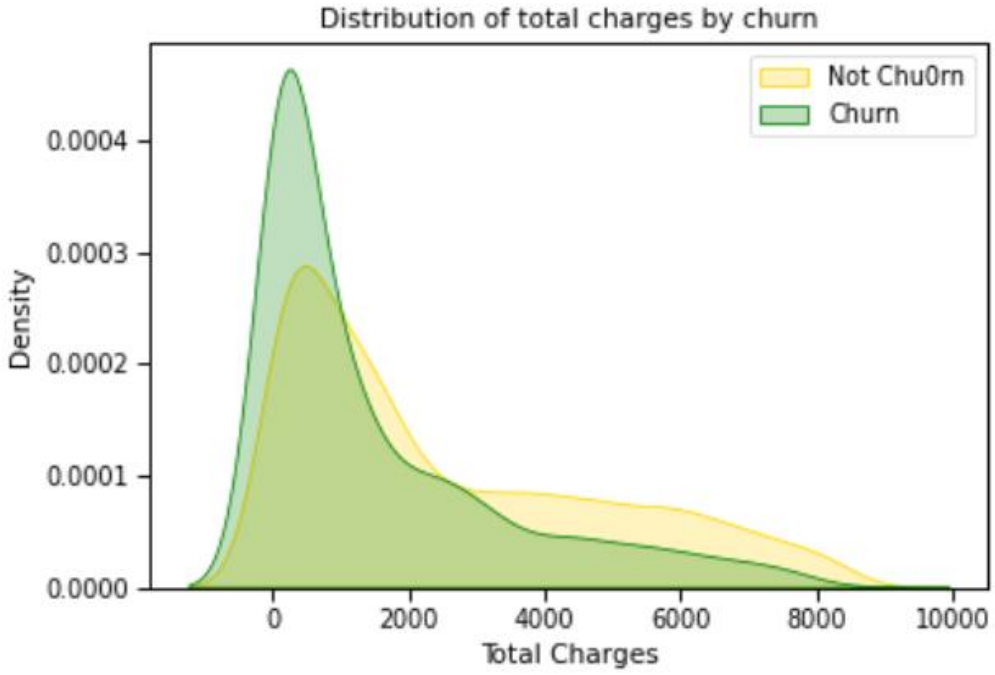


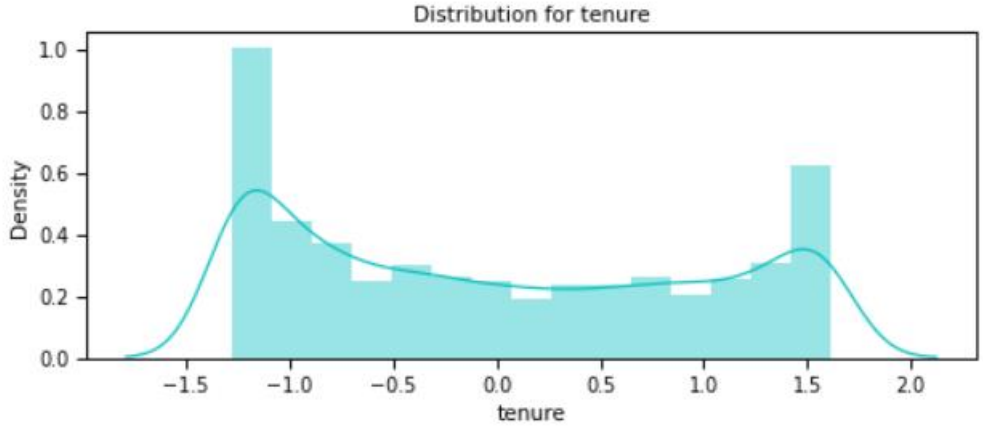
Step 3: Build a Prediction Model

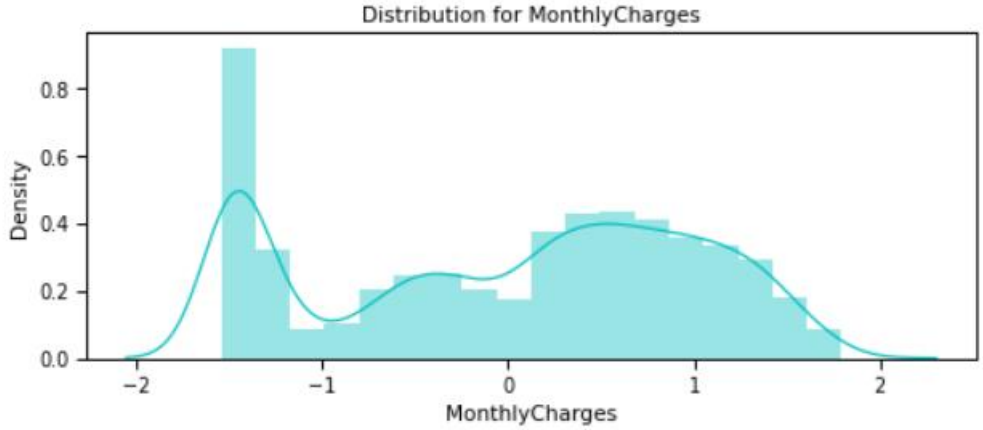
- Used Random Forest (a simple AI tool) to predict churn.

- Accuracy: 80% (i.e., the model is correct 8 out of 10 times).









Step 4: Key Findings

Feature Importance

- **Top 5 Features** Influencing Churn:
    1. **Tenure** (Longer tenure = less churn).
    2. **Contract Type** (Month-to-month = high churn).
    3. **Monthly Charges** (Higher charges = more churn).
    4. **Online Security** (No service = high churn).
    5. **Tech Support** (No service = high churn).

4. Business Suggestions

1. Offer discounts to monthly users to switch to yearly plans.

2. Reduce prices for high-paying unhappy customers.

3. Call at-risk customers (e.g., those with 1-month contracts) to offer help.

5. How You Can Try This

Tools Needed

- Excel/Google Sheets (for data cleaning).

- Python : use `sklearn` library.

6. Conclusion

- Why this matters: If a telecom company uses this, they can save customers and earn more money.

- What I learned: How to find patterns in data and make simple predictions.

7.Appendix

- **Dataset Sample**:

CustomerID	Gender	Contract	Tenure	MonthlyCharges	Churn
001	Male	Monthly	5	70	Yes
002	Female	Yearly	24	50	No

- **Full Code**:
  Python Code Snippet:

```
from sklearn.ensemble import RandomForestClassifier
model = RandomForestClassifier()
model.fit(X_train, y_train)
predictions = model.predict(X_test)
```