Project Name

Customer Churn Analysis

Submitted by:
Amandeep Singh
Aniket Kapoor

MCA 2nd Year

Submitted to:

Dr. Anjula Mehto

Assistant Professor

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

Computer Science and Engineering Department

Thapar Institute of Engineering and Technology, Patiala

November 2025

TABLE OF CONTENTS

Introduction or Project Overview

The customer churn ecosystem has become a major focus area for modern subscription-based companies, especially in telecom, banking, and online services. With rising competition and easily available alternatives, retaining customers has become just as important as acquiring new ones. As a result, understanding churn data has become extremely valuable for identifying at-risk customers, predicting who might leave, analyzing behavior patterns, and helping companies make better business decisions.

This project, **Customer Churn Analysis**, focuses on performing a detailed exploratory data analysis (EDA) on customer behavior and service usage patterns. The main objective is to uncover hidden trends, extract meaningful insights, and understand the key factors that drive customer churn.

The project involves:

- Cleaning and preprocessing raw customer data
- Performing statistical exploration and visualization
- Studying churn distribution and customer demographics
- Analyzing service usage patterns and billing behavior
- Understanding how contract type, support services, pricing, and tenure influence churn

The insights generated from this analysis can help:

- Businesses identify customers who are likely to leave
- Teams build better retention strategies and reduce churn rates
- Service providers improve plans, offers, and customer support
- Data scientists build machine learning models for churn prediction

Problem Statement

The customer-based service industry generates huge amounts of data, but companies often struggle to interpret it effectively due to inconsistent entries, missing information, and lack of analytical frameworks. Without proper analysis, businesses fail to understand why customers leave, leading to higher churn rates, revenue loss, and reduced customer satisfaction.

This project addresses the major challenges in analyzing customer churn::

Lack of clarity about customer behavior and preferences

Customers decide whether to stay or leave based on several factors such as service quality, pricing, contract type, internet speed, support quality, and billing issues. However, companies often lack a structured analysis that clearly highlights which factors matter the most and how strongly they influence churn.

2. Difficulty in identifying differences between churned and retained customers

The dataset contains inconsistent formats, missing values, categorical features, and noisy entries such as blank total charges.
Proper preprocessing is required to remove duplicates, convert categorical fields, handle null values, and standardize data before insights can be extracted.

3. Unclear relationship between pricing and customer churn

Companies may assume that higher prices directly lead to higher churn, but this is not always true.
Many customers pay more for better plans, while some churn even with low-cost plans.
A detailed analysis is needed to understand how monthly charges, total charges, and tenure collectively affect churn.

4. Lack of understanding of service impact

Certain services may influence churn behavior, and this project investigates questions like:
Does tech support reduce churn?
Do long-term contracts keep customers loyal?
Does paperless billing or online security impact satisfaction?

How does customer support influence retention?
Understanding these patterns can help companies redesign service plans and customer engagement strategies.

5. Variation in churn based on customer segments

Different customer groups—based on age, tenure, internet service type, contract type, or payment method—show different churn behaviors.
A segment-wise analysis is essential to understand which groups are most at risk and why.

Overview of the Dataset used

The dataset, **customer_churn_datacsv**, contains detailed information about customers of a telecom service provider and their service usage patterns.

Dataset Size

- **Total Restaurants:** ~7,000+ entries
- **Columns:** 20
- **City:** Telecome / Subscription-Based-Services

Important Columns Explained

Column	Description
customerID	Unique ID assigned to each customer
gender	Gender of the customer
SeniorCitizen	Indicates if the customer is a senior citizen (0/1)
Partner	Whether the customer has a partner (Yes/No)
Dependents	Whether the customer has dependents (Yes/No)
tenure	Number of months the customer has stayed with the company
PhoneService	Indicates if phone service is active
MultipleLines	Multiple phone line connection (Yes/No/No phone service)
InternetService	Type of internet: DSL, Fiber Optic, or None
OnlineSecurity	Whether online security is included (Yes/No)
OnlineBackup	Whether online backup is included
DeviceProtection	Whether device protection is included
TechSupport	Availability of tech support
StreamingTV	Access to streaming TV service
StreamingMovies	Access to streaming movies

Contract	Type of contract: Month-to-month, One-year, Two-year
PaperlessBilling	Billing preference (Yes/No)
PaymentMethod	Payment mode (Credit card, Bank transfer, etc.)
MonthlyCharges	Monthly billing amount
TotalCharges	Total amount paid till date
Churn	Whether the customer left the service (Yes/No)

Project Workflow

Step 1: Importing Necessary Libraries
Libraries used:

- pandas – For data loading, cleaning, and manipulation
- numpy – For numerical processing
- matplotlib & seaborn – For charts and visual data analysis
- scikit-learn (optional) – For encoding, scaling, or further modeling
- warnings – To ignore unnecessary warnings during visualization

Step 2: Data Loading and Basic Inspection
The dataset customer_churn.csv is loaded using pandas.

- Actions performed:
- Checked column names, dataset shape, and datatypes
- Inspected missing values in important columns like TotalCharges
- Verified and removed duplicate customer entries
- Observed basic distribution of churn values (Yes/No)

This step ensures the dataset is understood before cleaning.

Step 3: Data Cleaning and Preprocessing
This is the most critical step in churn analysis.
Important Cleaning Tasks

Fixing numerical columns

- TotalCharges contains blank spaces → converted them to NaN
- Converted TotalCharges and MonthlyCharges to numeric
- Filled missing numeric values where needed

Converting Yes/No columns

- Columns like:
- Partner
- Dependents
- PhoneService
- PaperlessBilling
- Churn

Handling “No internet service” & “No phone service”

- For columns such as:
- OnlineSecurity
- TechSupport
- StreamingTV

Cleaning categorical columns

- Standardized contract types (Month-to-month, One year, Two year)
- Ensured payment methods had consistent formatting

Removing duplicates

- CustomerID duplicates were checked and removed if found.
- This step prepares the dataset for accurate visual analysis.

Step 4: Exploratory Data Analysis
Performed detailed insights and visualizations to understand customer churn behavior.
Churn Distribution
Customers with Month-to-Month contracts show the highest churn.
Long-term contract customers churn significantly less.

Service Usage Patterns
Customers without OnlineSecurity or TechSupport churn more.
Fiber optic users churn more due to higher costs or dissatisfaction.

Tenure Insights
Most churn happens within first 6 months.
Long-tenure customers are far more loyal.

Payment Method Analysis
Users paying via Electronic Check have the highest churn rate.
Bank transfers & credit cards are associated with lower churn.

Monthly Charges Analysis
High monthly charges → High churn
Customers paying ₹70–₹100 range show the highest churn.

Contract Type Impact
Month-to-month customers are unstable and churn frequently.
Yearly or 2-year contracts are stable and low-risk.

Demographic Analysis
Senior citizens have slightly higher churn, mainly due to high charges.
Gender does not show major churn difference.s.

Step 5: Visualizations
Several plots were created for clear insights:

Bar Charts & Countplots
Churn vs Gender
Churn vs Contract
Churn vs InternetService

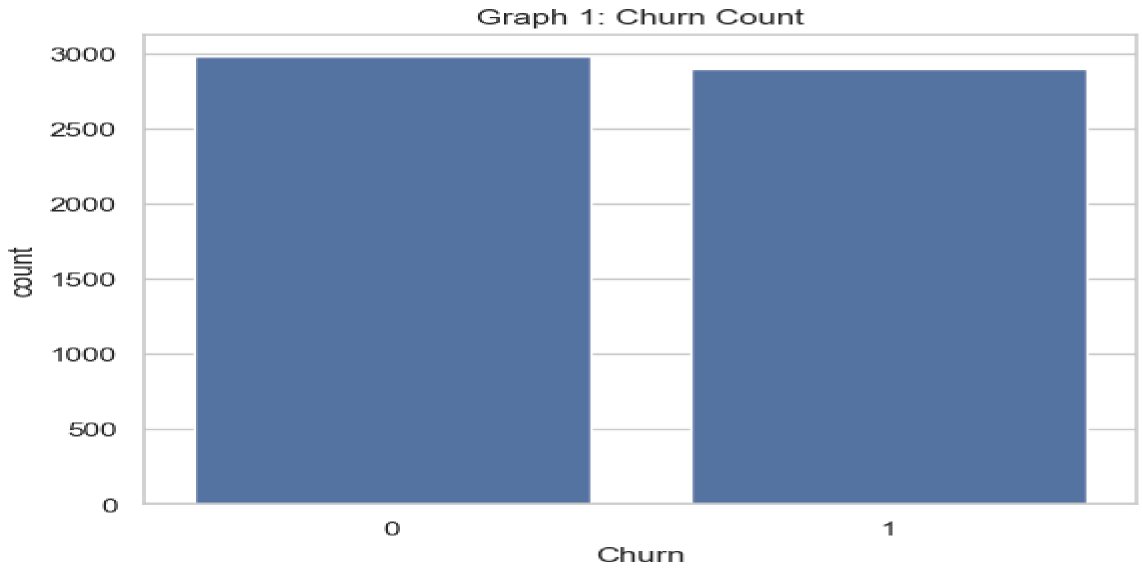
Histograms
Tenure distribution
Monthly Charges distribution

Boxplots
MonthlyCharges vs Churn
TotalCharges vs Churn

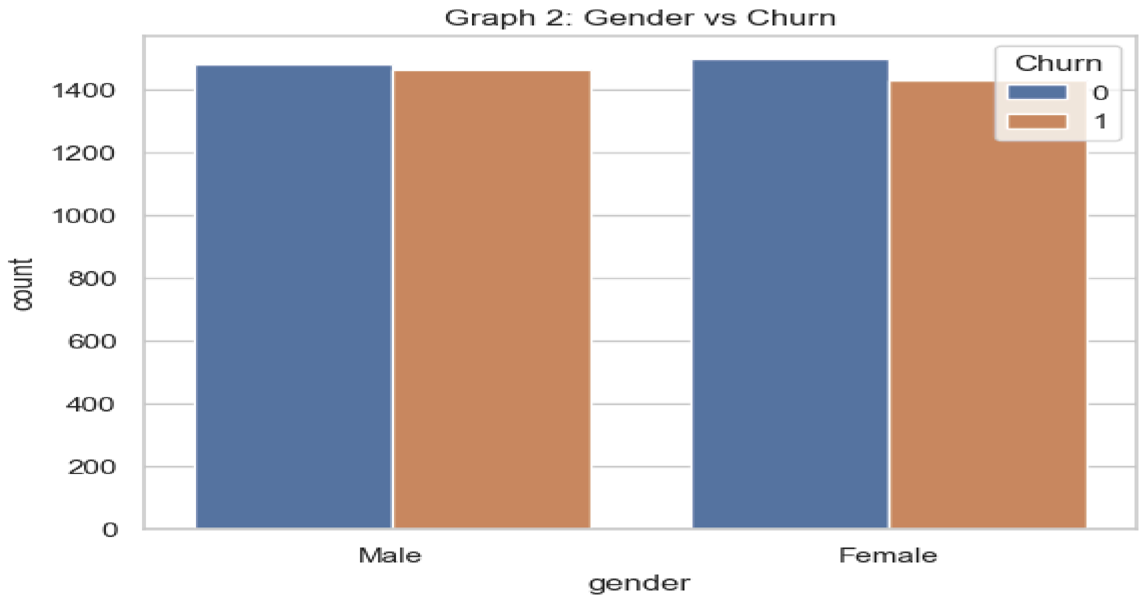
Heatmap
Correlation among numerical columns
Shows negative correlation between tenure and churn
Shows positive correlation between MonthlyCharges and churn

Scatter Plots
Tenure vs MonthlyCharges colored by churn

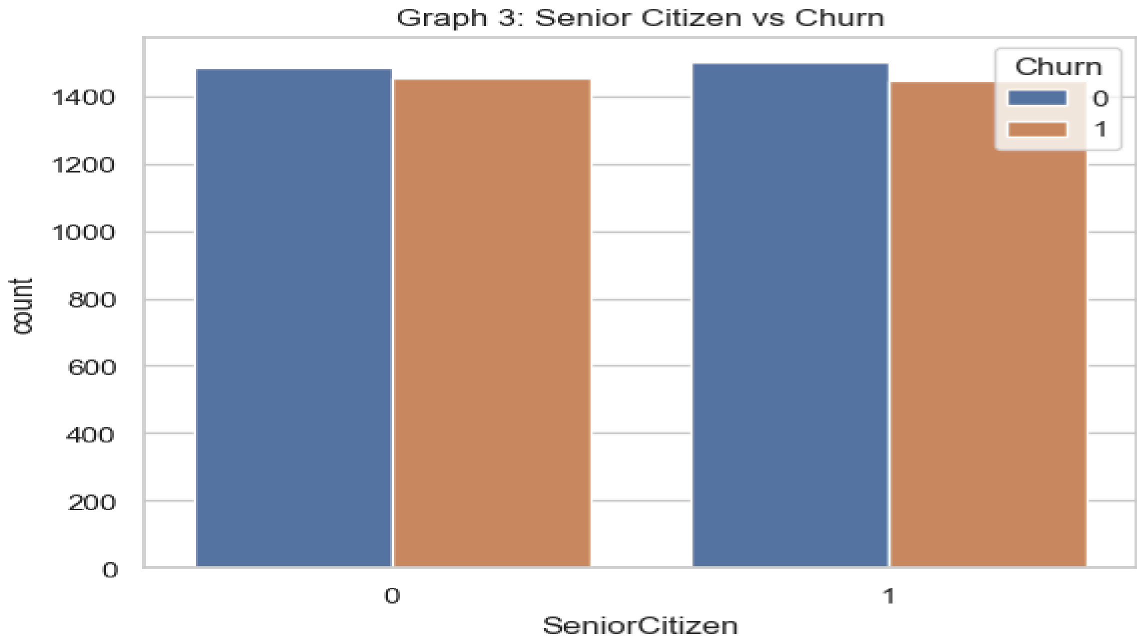
Results

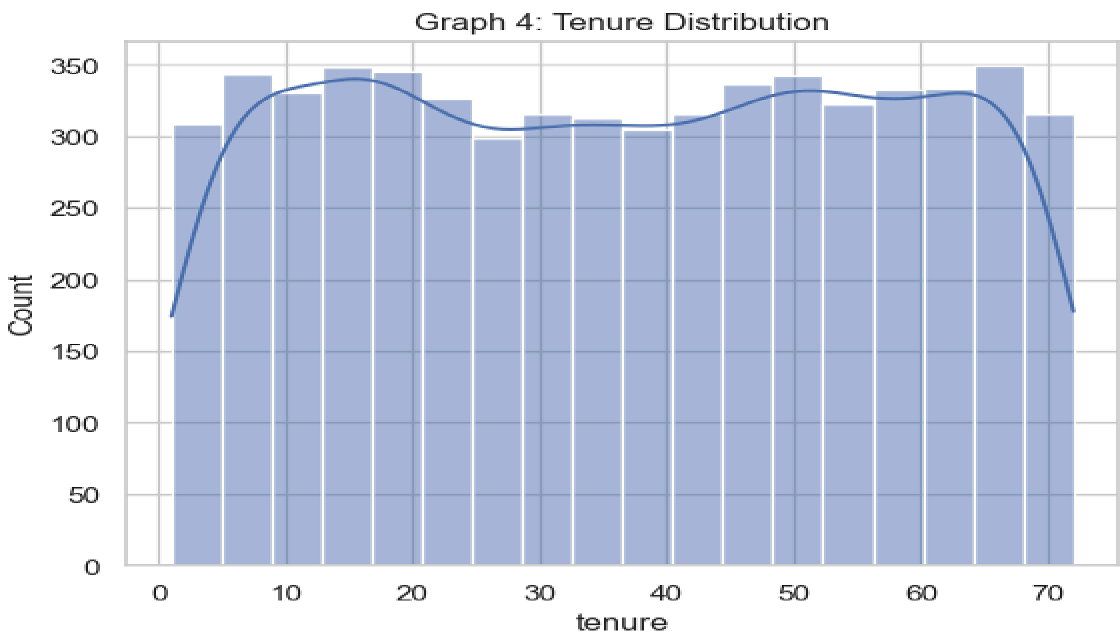


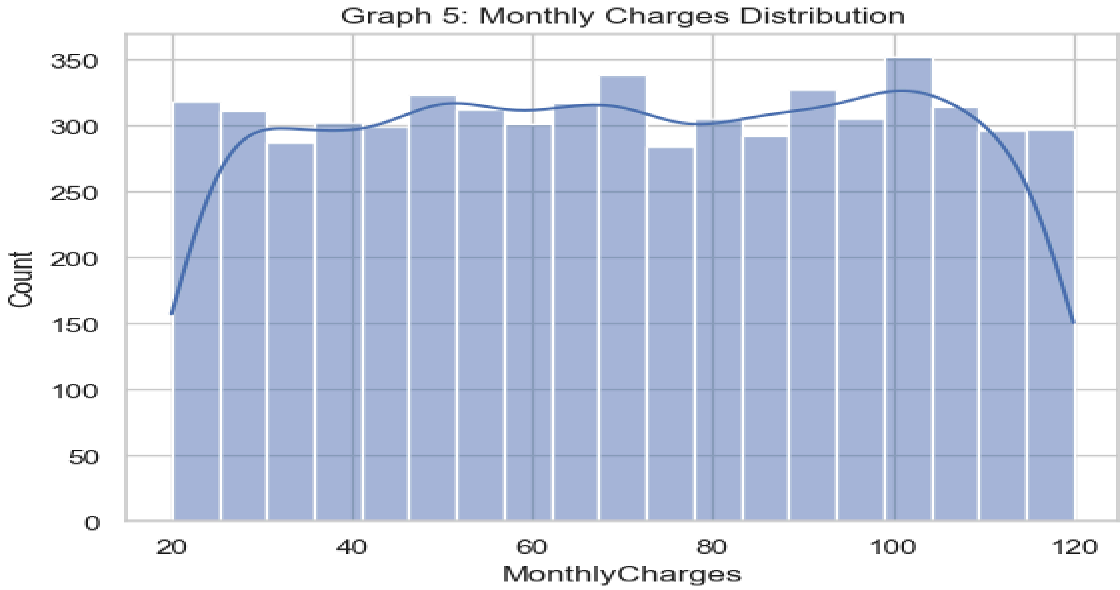
Graph 1: Churn Count
3000
2500
2000
1500
1000
500
0
count
0
1
Churn

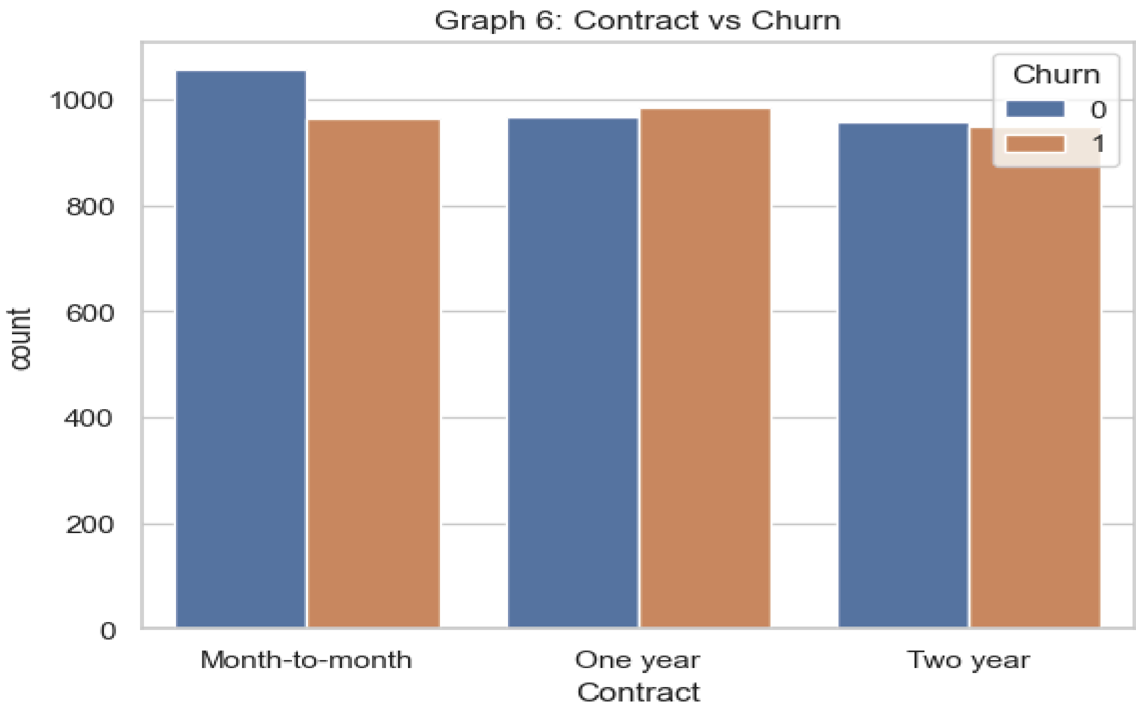


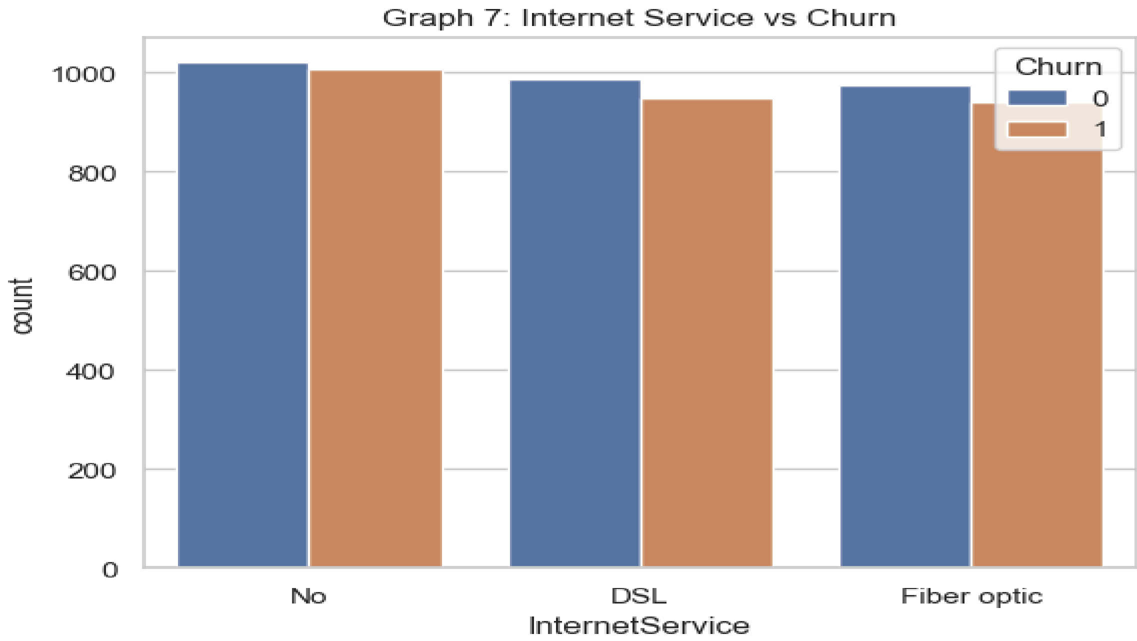
Graph 2: Gender vs Churn
1400
1200
1000
800
600
400
200
0
count
Male
Female
gender
Churn
0
1

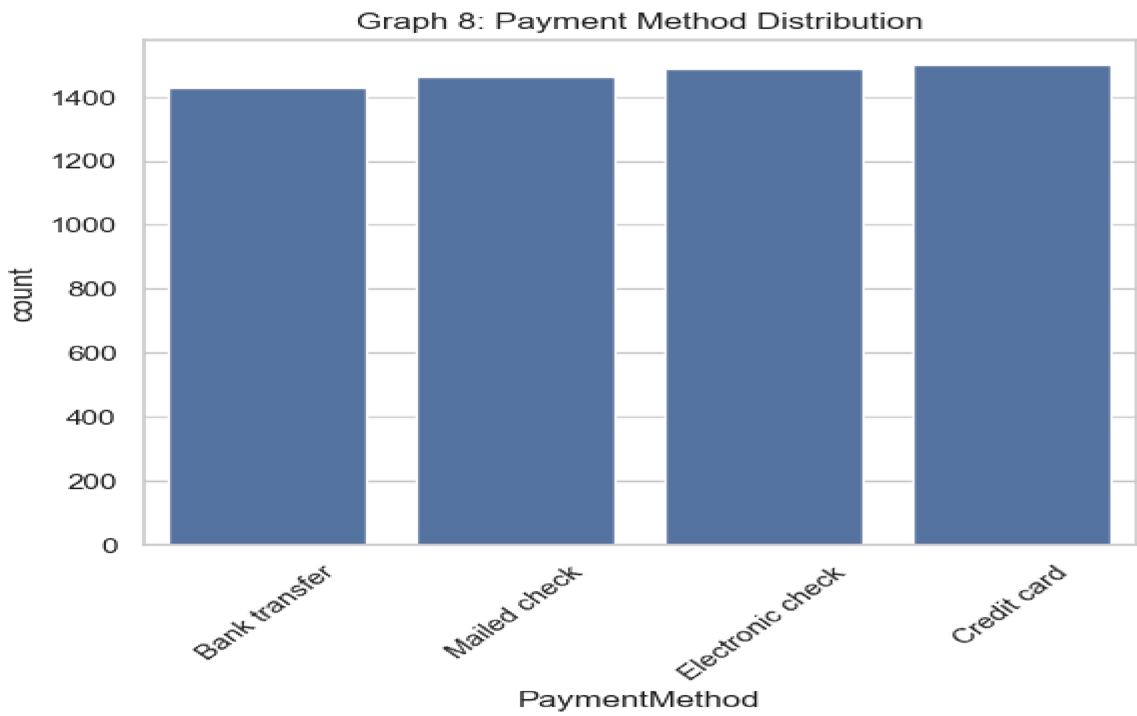

Graph 3: Senior Citizen vs Churn
Churn
0
1
count
1400
1200
1000
800
600
400
200
0
0
1
SeniorCitizen

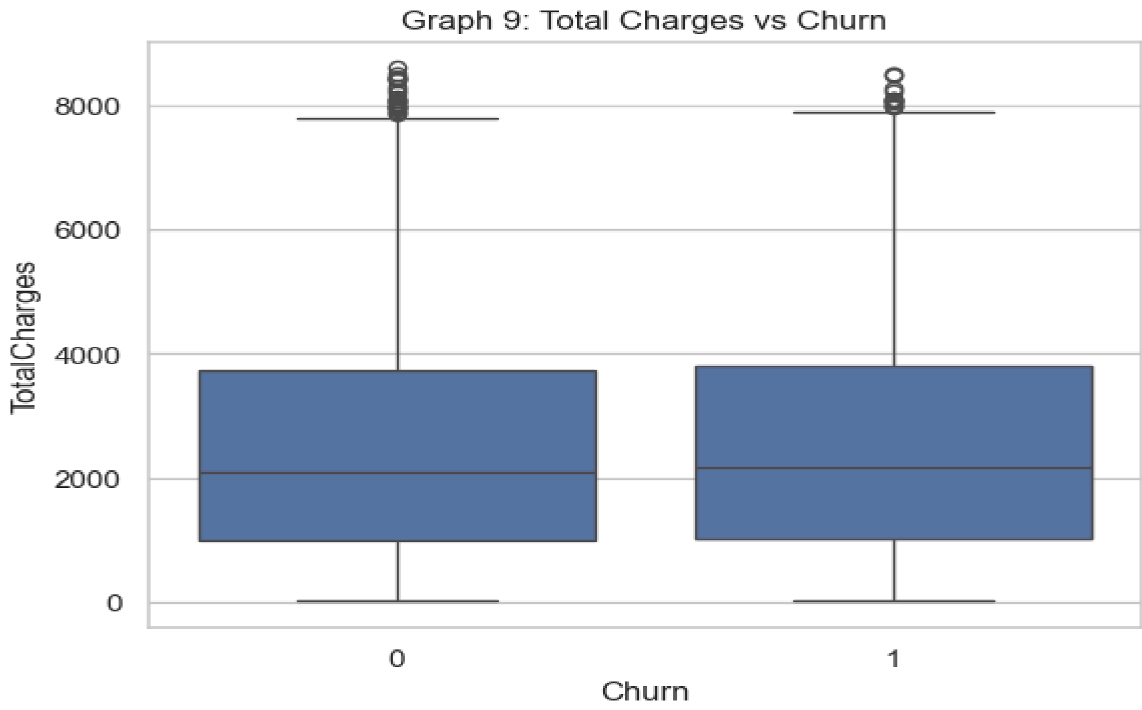

Graph 4: Tenure Distribution
Count
350
300
250
200
150
100
50
0
0
10
20
30
40
50
60
70
tenure

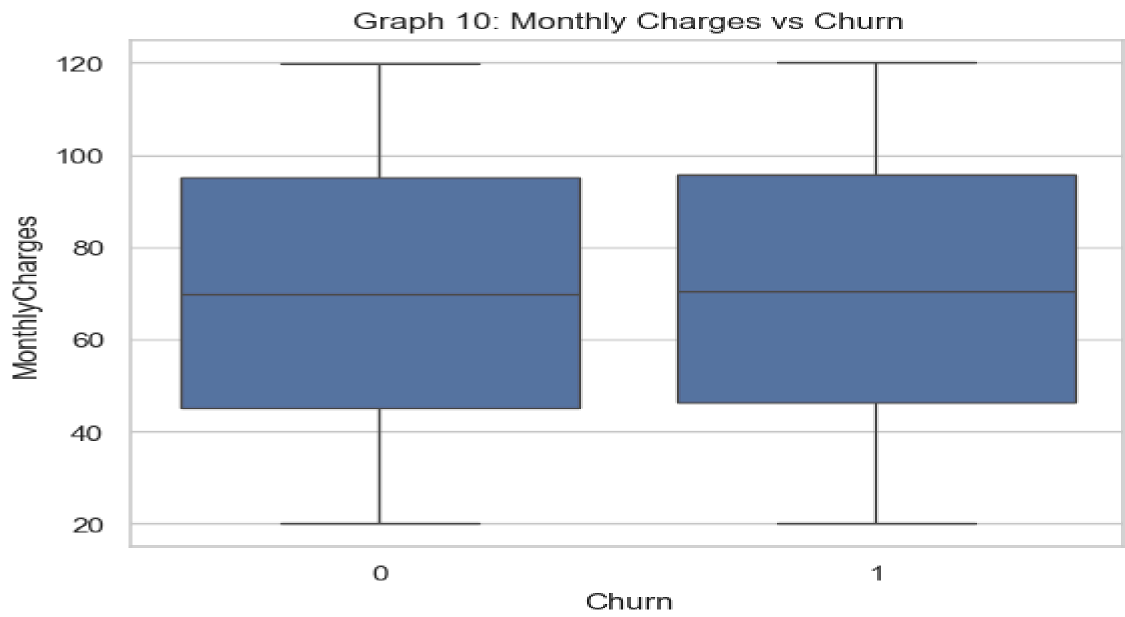

Graph 5: Monthly Charges Distribution
Count
350
300
250
200
150
100
50
0
20
40
60
80
100
120
MonthlyCharges

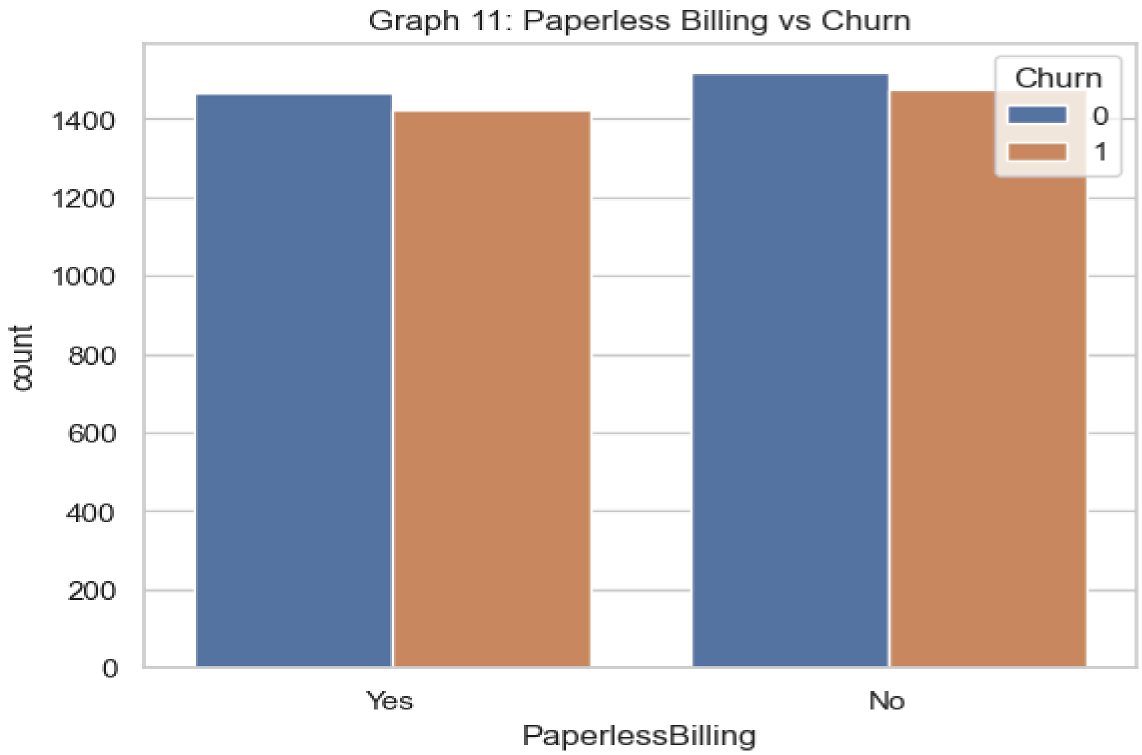

Graph 6: Contract vs Churn
Churn
0
1
count
1000
800
600
400
200
0
Month-to-month
One year
Two year
Contract

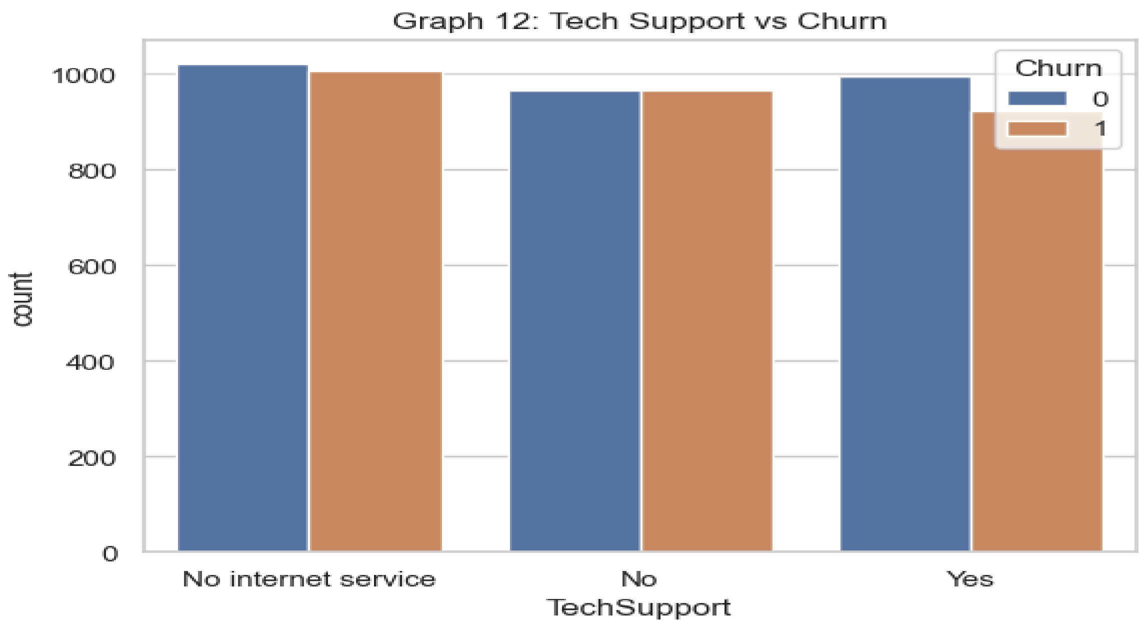

Graph 7: Internet Service vs Churn
Churn
0
1
count
1000
800
600
400
200
0
No
DSL
Fiber optic
InternetService

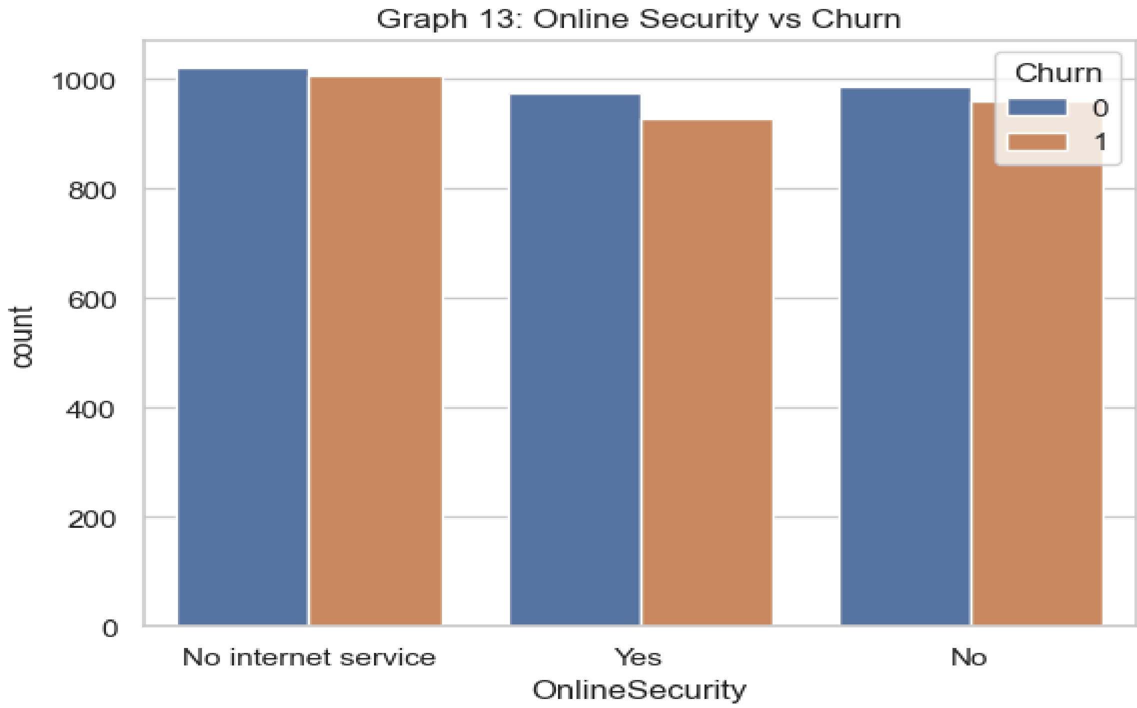

Graph 8: Payment Method Distribution
count
1400
1200
1000
800
600
400
200
0
Bank transfer
Mailed check
Electronic check
Credit card
PaymentMethod

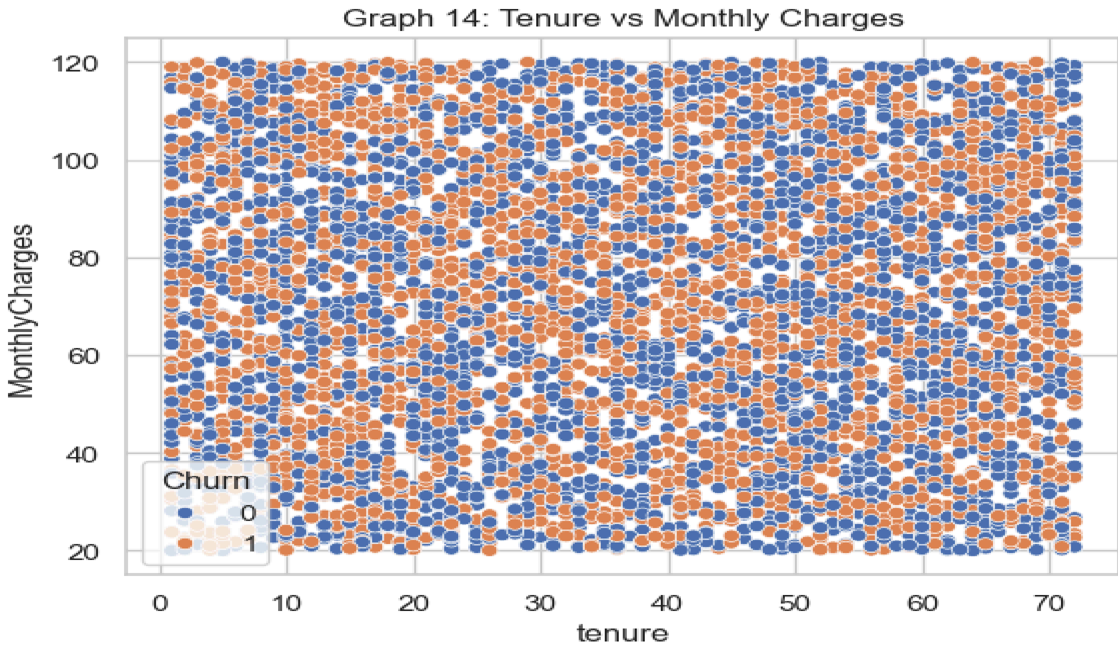

Graph 9: Total Charges vs Churn
8000
6000
4000
2000
0
TotalCharges
0
1
Churn

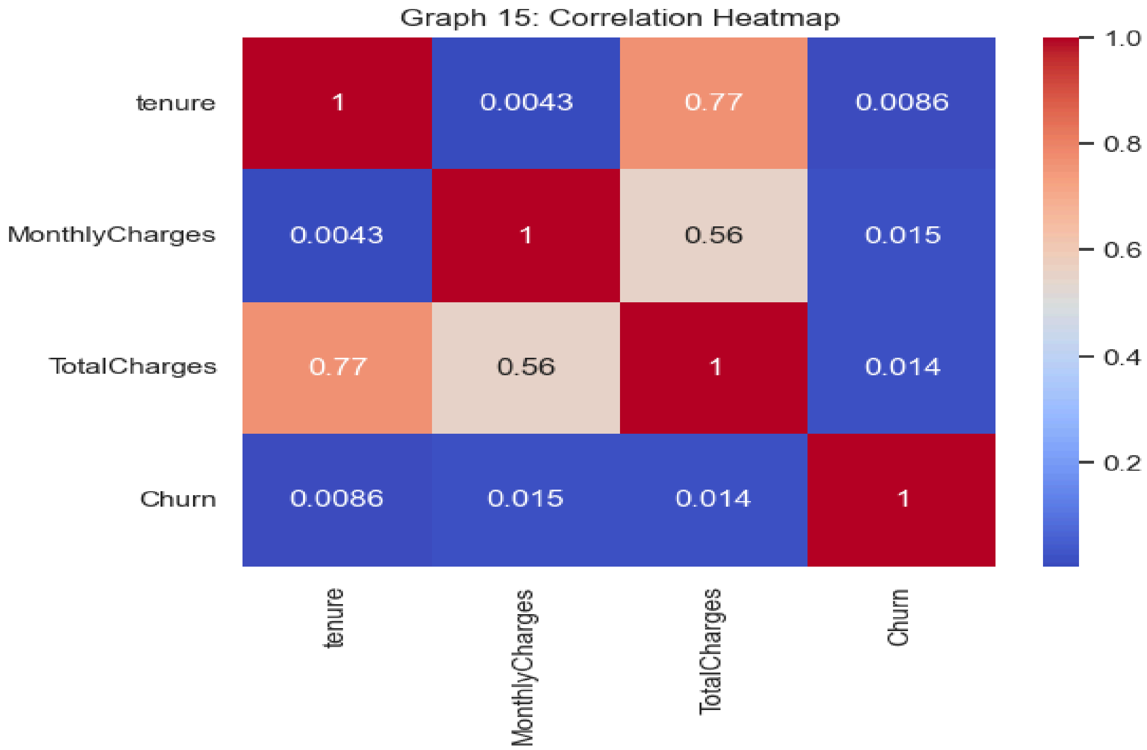

Graph 10: Monthly Charges vs Churn
120
100
80
60
40
20
MonthlyCharges
0
1
Churn


Graph 11: Paperless Billing vs Churn
Churn
0
1
1400
1200
1000
800
600
400
200
0
count
Yes
No
PaperlessBilling


Graph 12: Tech Support vs Churn
Churn
0
1
1000
800
600
400
200
0
count
No internet service
No
Yes
TechSupport


Graph 13: Online Security vs Churn
Churn
0
1
count
1000
800
600
400
200
0
No internet service
Yes
No
OnlineSecurity


Graph 14: Tenure vs Monthly Charges
MonthlyCharges
120
100
80
60
40
20
Churn
0
1
0
10
20
30
40
50
60
70
tenure


Graph 15: Correlation Heatmap
tenure
MonthlyCharges
TotalCharges
Churn
1
0.0043
0.77
0.0086
0.0043
1
0.56
0.015
0.77
0.56
1
0.014
0.0086
0.015
0.014
1
1.0
0.8
0.6
0.4
0.2

Conclusion

This project successfully demonstrates how structured data analysis can transform raw customer information into meaningful insights for predicting and reducing churn.

Key conclusions:

Customers with **month-to-month contracts, high monthly** charges, and **no tech/online security support** show the highest churn rates.

The data reveals that **tenure is one of the strongest churn indicators**—customers in their early months are far more likely to leave.

Service quality and support availability directly impact customer retention. Lack of value-added services leads to dissatisfaction.

Payment method plays a major role: customers paying via Electronic Check show significantly higher churn due to hidden fees and dissatisfaction.

The data cleaning process, including handling missing TotalCharges and standardizing categorical values, greatly improved the accuracy and reliability of insights.

This project provides a strong foundation for future extensions such as:

Developing a Machine Learning churn prediction model

Building customer segmentation for personalized retention strategies

Implementing survival analysis to understand how long customers stay

Performing sentiment analysis on customer complaints or service feedback

Creating dashboards for real-time churn monitoringL

GitHub Link

https://github.com/amandeep8680/data_visualization_project.git