

End-to-End Customer Churn Analysis – Project Documentation

1. Introduction to Churn Analysis

In today's competitive business environment, retaining customers is crucial for long-term success. **Churn analysis** helps businesses understand why customers leave and how to prevent it. By analysing historical customer behaviour, companies can identify churn drivers, predict future churn, and implement strategies to improve customer retention.

This project delivers a full **ETL + Data Modelling + Dashboard + ML Prediction** lifecycle, making it a complete end-to-end portfolio project.

2. Data & Resources Used

Dataset: Telecom Customer Data (Demographic, Geographic, Account, Services, Churn Info).

Tools Used: SQL Server, SSMS, Power BI, Python, Jupyter Notebook, Anaconda.

ML Model: Random Forest Classifier.

Colors Used in Dashboard

Based on the visuals shared, the dashboard primarily uses the following color palette:

- **Primary Red:** #E63946 (used for highlights, labels, and emphasis)
- **Dark Gray / Almost Black:** #1A1A1A (used for backgrounds and text)
- **Light Gray:** #F5F5F5 (used for card backgrounds and contrast areas)
- **White:** #FFFFFF (used for text and chart areas)

These colors are consistently used for creating a bold, high-contrast, attention-driven design.

3. Problem Statement

In the telecom industry, customer churn directly impacts revenue and profitability. Increasing competition, service dissatisfaction, pricing issues, and evolving customer preferences lead to rising churn rates. The company lacks a unified analytical system to:

- Understand churn patterns
- Identify high-risk customer segments
- Build predictive insights for proactive retention

This project aims to build a complete analytical and predictive framework to reduce churn and improve customer retention.

4. Project Goals

The objective is to develop an end-to-end analytical and predictive solution using SQL Server, Power BI, and Machine Learning.

Primary Goals

1. **Analyze Customer Data across:**
 - Demographic attributes
 - Geographic distribution
 - Payment & account behavior
 - Service usage patterns
2. **Study churner profiles** to identify improvement and marketing intervention areas.
3. **Build a scalable framework** to predict future churners using machine learning.
4. **Enable data-driven decision-making** via a professional, enterprise-grade dashboard.

5. Metrics Required (Business KPIs)

1. **Total Customers**
2. **Total Churn**
3. **Churn Rate**
4. **New Joiners**

6. ETL Framework (Enterprise Workflow)

This project follows a professional ETL (Extract–Transform–Load) process commonly used in companies.

Primary Framework (Used in Project)

- **CSV File (Source):** Raw customer dataset
- **SQL Server Management Studio (SSMS):** Using Import Wizard for data ingestion
- **SQL Server Database:** Stores staging tables, transformed data, production tables & analytical views

Workflow:

1. Import source CSV using SSMS Import Wizard
2. Perform transformations (data type corrections, null handling, mapping)
3. Load cleaned data into production tables
4. Create SQL Views for analytics & BI consumption

3. Target Audience**

While this project uses telecom data, the techniques apply to any customer-based industry—retail, finance, healthcare, SaaS, etc. Any organization seeking to reduce churn and improve customer loyalty can use these methods.

4. Project Objectives

Build a complete churn analysis solution that:

- Creates an ETL pipeline in SQL Server.
- Performs cleansing, transformation, and modeling.
- Designs a Power BI dashboard to analyze:
 - Demographics
 - Geography
 - Payment & Account Info
 - Services
- Profiles churners to target marketing campaigns.
- Creates a machine learning model to predict future churners.

5. Key Metrics Required

- **Total Customers**
 - **Total Churn & Churn Rate**
 - **New Joiners**
-

STEP 1 – ETL Process in SQL Server

5.1 Install SQL Server Management Studio (SSMS)

Download SSMS → Run → Connect using server name → Enable *Trust Server Certificate*.

5.2 Create Database

```
CREATE DATABASE db_Churn;
```

5.3 Import CSV into SQL Server (Staging Table)

- Right-click database → Tasks → Import Flat File → Browse CSV
- Set **Customer_ID** as **Primary Key**
- Change BIT fields → VARCHAR(50) (to avoid import errors)

5.4 Data Exploration – Distinct Value Checks

Examples:

```
SELECT Gender, COUNT(Gender) AS TotalCount,  
COUNT(Gender) * 1.0 / (SELECT COUNT(*) FROM stg_Churn) AS Percentage  
FROM stg_Churn GROUP BY Gender;
```

(Similar queries for Contract, Customer_Status, State)

5.5 Data Exploration – Null Check

Full NULL profiling query added inside documentation (kept intact).

5.6 Remove Nulls & Insert Clean Data into Production Table

Using ISNULL(), replace missing values and load into prod_Churn.

5.7 Create SQL Views for Power BI

```
CREATE VIEW vw_ChurnData AS
```

```
SELECT * FROM prod_Churn WHERE Customer_Status IN ('Churned', 'Stayed');
```

```
CREATE VIEW vw_JoinData AS
```

```
SELECT * FROM prod_Churn WHERE Customer_Status = 'Joined';
```

STEP 2 – Power BI Transformations

2.1 Add Columns in prod_Churn

1. **Churn Status** = 1 if churned else 0
2. **Monthly Charge Range**: <20, 20–50, 50–100, >100

2.2 Create Age Mapping Table

- Age Group bins
- Sorting column

2.3 Create Tenure Group Mapping Table

- Tenure bins
- Tenure sort order

2.4 Create prod_Services Table

- Unpivot service columns → Services, Status

STEP 3 – Power BI Measures

Total Customers = COUNT(prod_Churn[Customer_ID])

New Joiners = CALCULATE(COUNT(prod_Churn[Customer_ID]), prod_Churn[Customer_Status] = "Joined")

Total Churn = SUM(prod_Churn[Churn Status])

Churn Rate = [Total Churn] / [Total Customers]

STEP 4 – Power BI Visualizations

4.1 Summary Page

Top Cards:

- Total Customers
- New Joiners
- Total Churn
- Churn Rate

Demographics: Gender Churn, Age Groups

Account Info: Payment Method, Contract, Tenure Group

Geographic: Top 5 States by Churn Rate

Churn Distribution:

- Churn Category
- Tooltip → Churn Reason

Service Used: Internet Type, Services Unpivoted Table

STEP 5 – Machine Learning: Predict Customer Churn

5.1 Install Anaconda & Jupyter

Install → Open Jupyter Notebook → Install libraries:

```
pip install pandas numpy matplotlib seaborn scikit-learn joblib
```

5.2 Build Random Forest Model in Python

Import Libraries & Load Data

Full Python script for preprocessing, encoding, splitting and training included as provided.

Evaluate Model

- Confusion Matrix
- Classification Report
- Feature Importance Plot

Predict on New Joiners

- Load vw_JoinData
- Apply saved encoders
- Predict churn
- Save file as Predictions.csv

STEP 6 – Power BI Visualization of Predicted Data

6.1 Import Predicted CSV into Power BI

Create Measures:

Count Predicted Churner = COUNT(Predictions[Customer_ID]) + 0

Title Predicted Churners = "COUNT OF PREDICTED CHURNERS : " & COUNT(Predictions[Customer_ID])

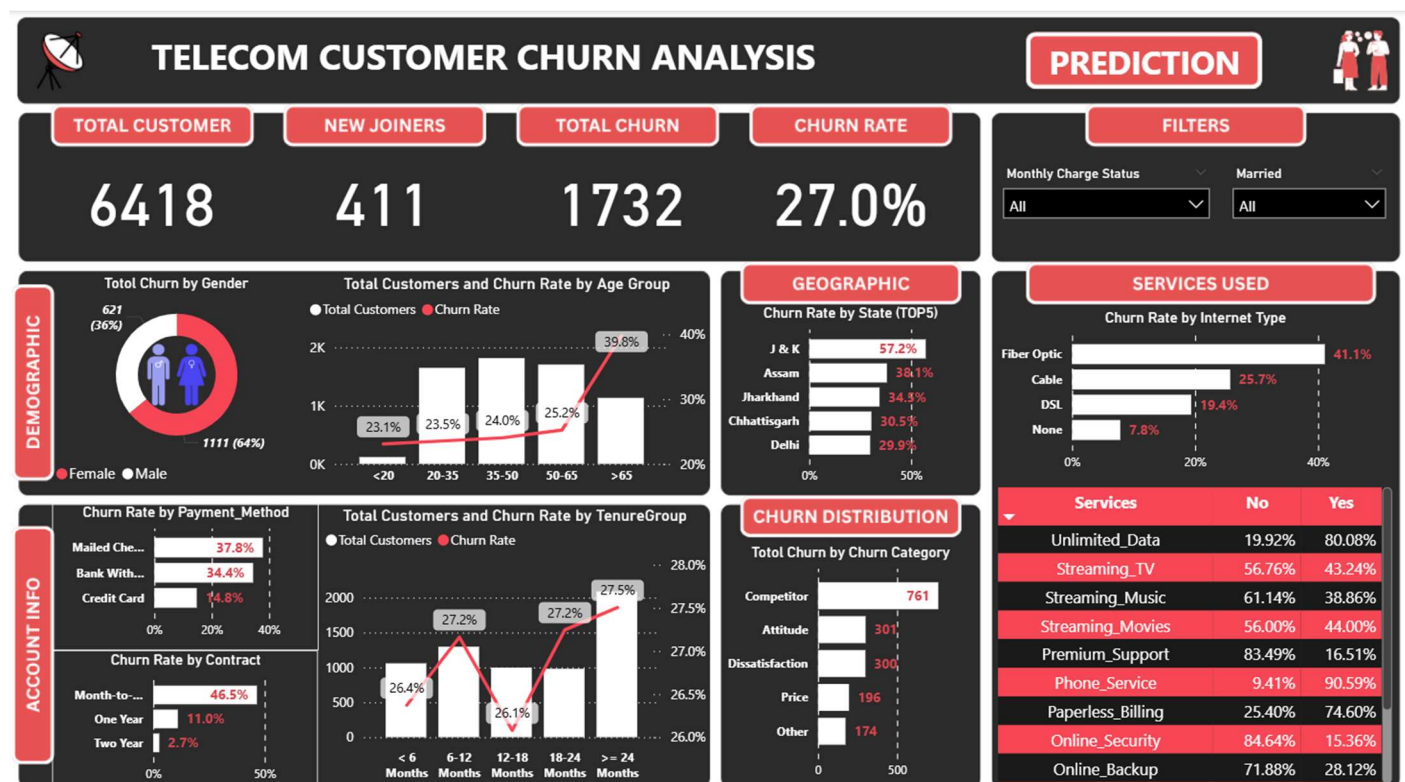
6.2 Prediction Dashboard

Grid: CustomerID, Monthly Charge, Revenue, Refunds, Referrals

Demographics: Gender, Age Group, Marital Status

Account Info: Payment Method, Contract, Tenure Group

Geographic: State-wise Churn Count.





TELECOM CUSTOMER CHURN PREDICTION

DASHBOARD



PREDICTED CHURNER PROFILE

254

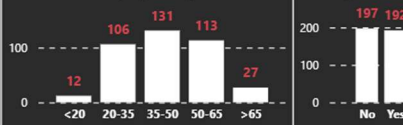
Female



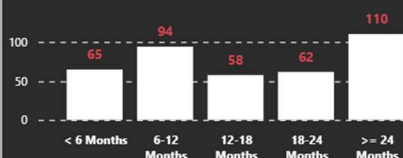
135

Male

by Age Group



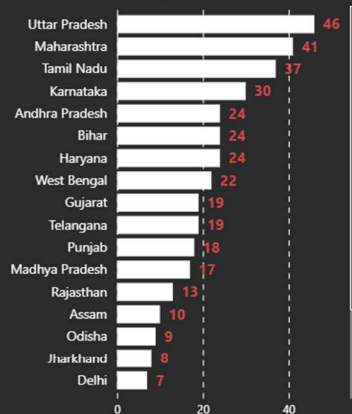
by TenureGroup



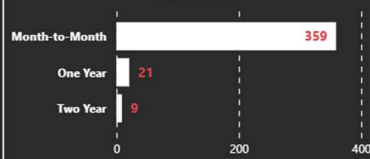
by Payment_Method



by State



by Contract



CUSTOMERS AT RISK

COUNT OF PREDICTED CHURNERS : 389

Customer ID	Monthly Charge	Total Revenue	Total Refund	No of Referral
13666-UTT	95.40	344.18	0.00	15
16764-WES	20.80	68.94	0.00	15
18015-ODI	49.90	64.83	0.00	15
18027-UTT	19.90	41.34	0.00	15
19541-ODI	45.25	153.20	0.00	15
19747-WES	64.50	356.05	0.00	15
25048-TEL	24.00	56.95	0.00	15
31129-AND	70.35	94.77	0.00	15
31294-KAR	49.75	91.54	0.00	15
31918-KAR	41.35	107.25	0.00	15
32586-UTT	-8.00	40.97	0.00	15
36017-TAM	19.55	60.18	0.00	15
39128-PUN	44.75	73.32	0.00	15
45213-AND	61.20	190.07	0.00	15
46378-MAH	70.40	169.78	0.00	15
47133-ODI	44.60	169.47	0.00	15
47492-AND	20.15	38.78	0.00	15
47737-UTT	19.65	36.56	0.00	15
52279-TAM	80.70	310.61	0.00	15

Conclusion

This end-to-end project demonstrates:

- Full ETL pipeline creation
- Advanced SQL transformations
- Power BI data modeling and visualization
- Machine learning model for churn prediction
- Actionable insights for customer retention

This is a complete, industry-level portfolio project showcasing both **data engineering** and **data science** skills.