Customer Churn Analysis

Presented By : Subeena k.k

Introduction to Churn Analysis

In today’s highly competitive market landscape, customer retention has become a critical factor for ensuring long-term business sustainability and growth. Churn analysis is a strategic data-driven approach aimed at understanding and minimizing customer attrition. This process involves analyzing historical customer data to uncover patterns, behaviors, and potential triggers that lead to customers discontinuing their association with a company.

By leveraging advanced analytics and machine learning techniques, businesses can not only predict which customers are most likely to churn but also identify the underlying factors influencing their decisions. These insights enable organizations to implement targeted retention strategies, enhance customer satisfaction, and ultimately foster stronger brand loyalty. Effective churn analysis allows companies to be proactive rather than reactive, transforming customer data into actionable intelligence for smarter decision-making.

Problem Statement

Companies are losing valuable customers but lack insights into why churn is happening or how to prevent it.

Challenge:

- High churn impacts revenue and growth.
- Data is available, but insights are hidden.
- No clear method to predict or respond to churn effectively.

Project Solution:

✅ Build a complete ETL process to collect, clean, and structure customer data.
✅ Develop a Power BI dashboard for visual exploration.
✅ Perform multi-level analysis:

- 👤 Demographic
- 🌍 Geographic
- 💳 Payment & Account Info
- 🛠️ Service Usage

✅ Analyze churner profiles to identify:

- 📣 Marketing gaps
- 🔁 Retention opportunities

✅ Design a method to:

- 🔍 Predict future churners
- 📊 Track metrics like Total Customers, Churn Rate, and New Joiners

Data Overview

This dataset contains telecom customer details used for churn analysis.

- Total Focus Areas:
  - Customer Info: ID, Gender, Age, State, Referrals
  - Services Used: Phone, Internet, Streaming, Security
  - Billing: Monthly Charge, Total Revenue, Payment Method
  - Churn Info: Status (Stayed/Churned), Category, Reason
- Key Field: Customer_ID (unique key)
- Target Field: Customer_Status (to analyze churn)

Step 1: Data Extraction, Cleaning, and Preparation Using SQL Server

Key Steps Involved:

1. **Create Database:**
    - A new database named db_churn is created to manage the churn analysis project.
2. **Data Import (Extract):**
    - The raw customer data is provided in a CSV file format.
    - Using SQL Server Management Studio (SSMS), we utilized the Import Wizard to load this data into staging tables.

3. Data Exploration (Transform):

- Performed initial data checks:
  - Explored distinct values for key columns.
  - Calculated percentage distribution of categorical values (e.g., Gender, Internet Service).
  - Checked for null values and data inconsistencies.

4. Data Cleaning & Transformation:

- Cleaned the null or invalid entries.
- Inserted clean data into production tables for reliable analysis.

5. View Creation (Load):

- Created SQL views to simplify querying for Power BI and machine learning models.
- These views aggregate and join key data fields such as revenue, tenure, service usage, and churn status.

Step 2: Data Transformation in Power BI

In this step, I used Power BI to transform data for analysis and build an interactive dashboard for churn insights. The data was imported from the SQL Server database and enhanced using DAX expressions and data modeling techniques.

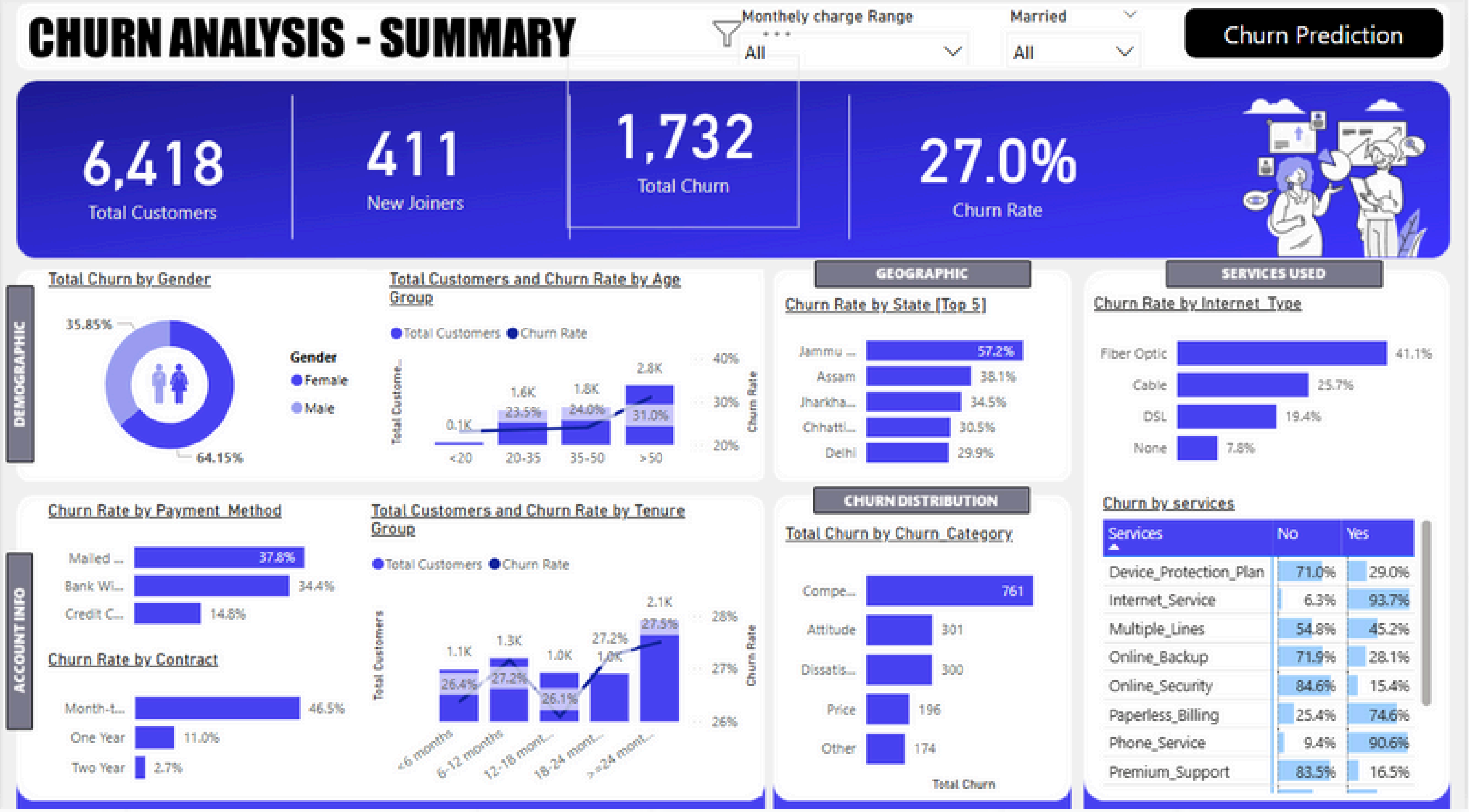
Transformations Done:

- Churn Status Column
- → 1 for Churned, 0 for Stayed
- Monthly Charge Bins
- → < 20, 20–50, 50–100, > 100
- Age Grouping Table
- → < 20, 20–35, 36–50, > 50 + sort key
- Tenure Grouping Table
- → <6, 6–12, 12–18, 18–24, ≥24 Months + sort key
- Services Table (Unpivoted)
- → Columns converted to Service and Status fields

📊 Dashboard Metrics & Visuals

Key KPIs displayed:

- ✅ Total Customers
- 🚪 Total Churn
- 📈 Churn Rate
- 🆕 New Joiners



Visuals include:

- Churn Breakdown by:
  - Gender
  - Age group
  - State
  - Internet type
  - Contract type
  - Tenure group
  - Churn category & reason
- All visuals are interactive and filterable using slicers.

Key Insights from Dashboard

📊 **Overall Metrics**

- Total Customers: 6,418
- Total Churned: 1,732
- New Joiners: 411
- Churn Rate: 27.0% (Significantly high – needs action)

👥 **Demographic Insights**

- Churn by Gender:
  - Females churn more (64.15%) than males (35.85%)
- Churn by Age Group:
  - Customers aged >50 have the highest churn rate (31%)
  - Age groups 20–35 and 35–50 are also at risk (24–25%)

🌍 **Geographic Insights**

- Highest churn rates observed in:
  - Jammu: 57.2%
  - Assam: 38.1%
  - Jharkhand: 34.5%
  - → These regions may need focused retention campaigns.

💳 Account & Billing Insights

- Churn by Contract:
  - Month-to-month contracts have the highest churn (46.5%)
  - Annual contracts reduce churn drastically (as low as 2.7%)
- Payment Method:
  - Mailed checks show the highest churn (37.8%)
  - Credit card users are more stable (14.8%)

🛠️ Service Usage Insights

- Internet Type:
  - Fiber Optic users churn more (41.1%) than DSL or Cable
- Churn by Services (when not used):
  - Device Protection Plan: 71% churned
  - Online Backup: 71.9% churned
  - Premium Support: 83.5% churned
  - → 🚨 Customers not using value-added services are more likely to churn.

📦 Churn Distribution

- Churn Categories:
  - Competitor-related churn dominates (761 customers)
  - Followed by attitude issues (301) and dissatisfaction (300)
  - → Retention strategy should address competitor offers and service satisfaction.

Step 3: Predicting Customer Churn Using Machine Learning

In this step, I developed a churn prediction model using the Random Forest Classifier in Python via Jupyter Notebook.

✅ **Tools & Technologies Used**

- Jupyter Notebook (Python via Anaconda)
- Libraries: pandas, numpy, matplotlib, seaborn, scikit-learn, joblib
- Data Source: SQL Server views (vw_ChurnData, vw_JoinData) imported to Excel

🔧 **Data Preparation**

- Imported and preprocessed customer churn data:
    - Dropped non-predictive columns (Customer_ID, Churn_Category, Churn_Reason)
    - Label-encoded all categorical features
    - Converted Customer_Status to binary (Stayed = 0, Churned = 1)
    - Split into training and testing sets (80/20)

🤖 Model Building

- Applied Random Forest Classifier with 100 estimators
- Trained model on the prepared dataset

📈 Model Evaluation

- Evaluated using confusion matrix and classification report
- Plotted feature importance chart to understand key drivers of churn

🔮 Making Predictions

- Loaded new customer data (vw_JoinData)
- Encoded using previously fitted label encoders
- Predicted churn on new data
- Filtered and saved predicted churners to a CSV file

Step 4: Visualization of Predicted Churners in Power BI

In the final step, we created a dedicated Churn Prediction dashboard in Power BI to visualize customers who were predicted to churn using the trained ML model.

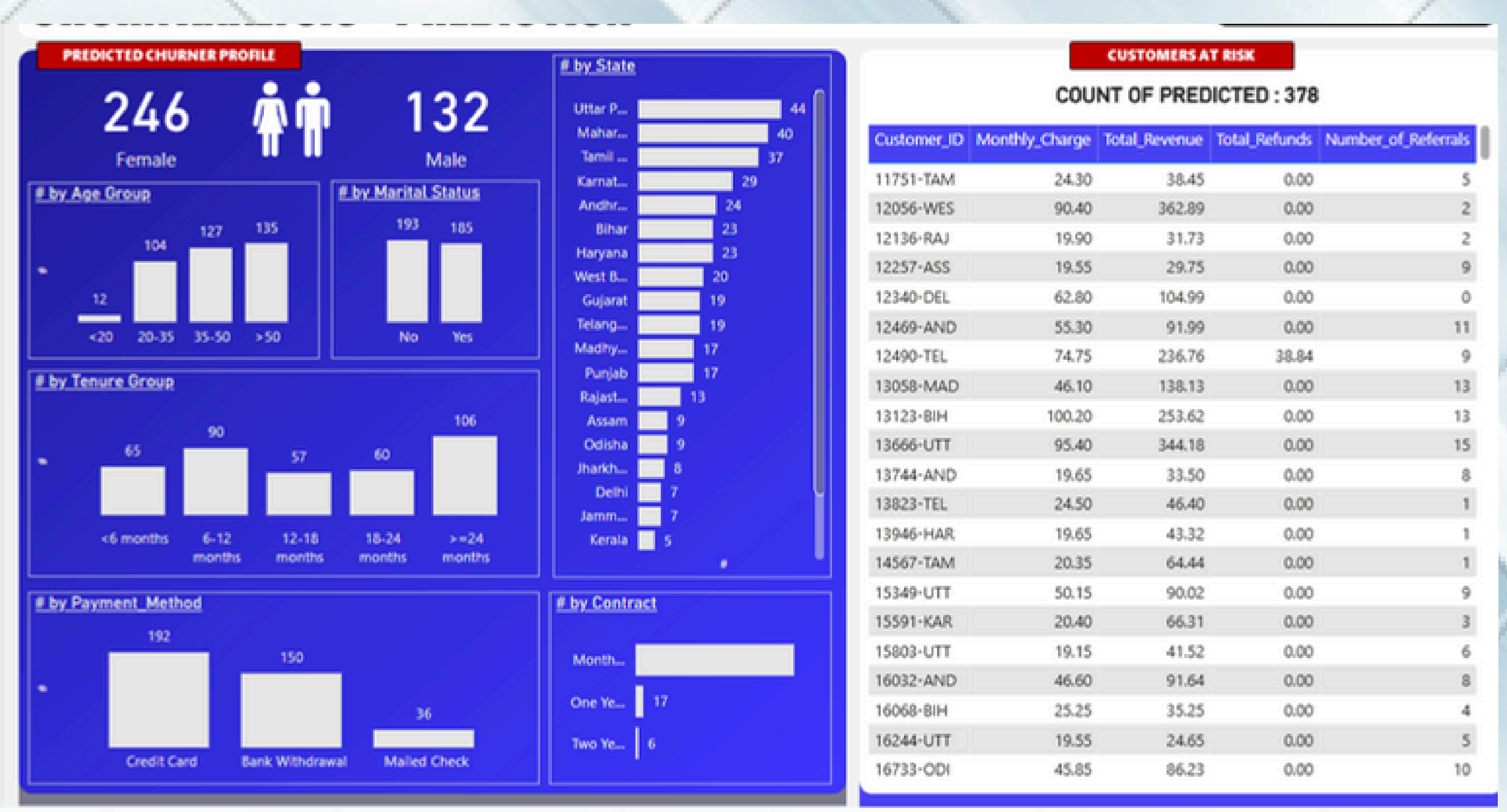
✅ **Data Integration**

- Loaded the prediction results (Predictions.csv) either via:
  - Direct CSV import, or
  - SQL Server connection (if saved back to DB)

📊 **Measures Created**

- Count Predicted Churner: Total predicted churners
- Title Predicted Churners: Custom label showing count dynamically

Insights from Churn Prediction Dashboard



Customer_ID	Monthly_Charge	Total_Revenue	Total_Refunds	Number_of_Referrals
11751-TAM	24.30	38.45	0.00	5
12056-WES	90.40	362.89	0.00	2
12136-RAJ	19.90	31.73	0.00	2
12257-ASS	19.55	29.75	0.00	9
12340-DEL	62.80	104.99	0.00	0
12469-AND	55.30	91.99	0.00	11
12490-TEL	74.75	236.76	38.84	9
13058-MAD	46.10	138.13	0.00	13
13123-BIH	100.20	253.62	0.00	13
13666-UTT	95.40	344.18	0.00	15
13744-AND	19.65	33.50	0.00	8
13823-TEL	24.50	46.40	0.00	1
13946-HAR	19.65	43.32	0.00	1
14567-TAM	20.35	64.44	0.00	1
15349-UTT	50.15	90.02	0.00	9
15591-KAR	20.40	66.31	0.00	3
15803-UTT	19.15	41.52	0.00	6
16032-AND	46.60	91.64	0.00	8
16068-BIH	25.25	35.25	0.00	4
16244-UTT	19.55	24.65	0.00	5
16733-ODI	45.85	86.23	0.00	10

- Total Predicted Churners: 378
- More females (246) than males (132) are at risk
- Age >50 has the highest churn (135 customers)
- Tenure >24 months shows most churn (106), indicating long-term user dissatisfaction
- Month-to-month contracts dominate churn cases
- Credit Card is the most common payment method among churners
- Top churn-prone states: Uttar Pradesh, Maharashtra, Tamil Nadu

Conclusion

This customer churn analysis project successfully covered the end-to-end data journey – from data extraction and transformation to churn prediction and dashboarding. Using a combination of SQL Server, Power BI, and Python (Random Forest ML model), we built a robust system to:

- Analyze historical churn behavior
- Predict potential future churners
- Visualize insights for better decision-making

With a total churn rate of 27% identified, and 378 customers predicted to churn, the project provides clear direction for targeted retention strategies.

Recommendations

1. **Target At-Risk Groups:**
    - Focus on customers >50 years and long-tenure users with declining satisfaction.
    - Retain month-to-month contract users through loyalty programs or upsell to annual plans.
2. **Geo-Specific Campaigns:**
    - Prioritize churn prevention in Uttar Pradesh, Maharashtra, and Tamil Nadu.
3. **Service Improvement:**
    - Boost adoption of value-added services like Online Backup, Premium Support, and Protection Plans.
    - Offer incentives or education around service benefits.
4. **Contract Optimization:**
    - Encourage shift from flexible contracts to long-term ones via discounts or bonus features.
5. **Ongoing Prediction:**
    - Automate the churn prediction pipeline using SSIS + Jupyter + Power BI refresh to keep predictions real-time.

THANK YOU