Project Report: End-to-End Churn Analysis

- Pranati Panjiar

1. Introduction

In today's competitive business environment, customer retention plays a critical role in long-term growth and profitability. Customer churn refers to the loss of customers over a specific period and directly impacts revenue and brand loyalty. Churn analysis helps organisations understand why customers leave, identify high-risk segments, and design proactive retention strategies.

This project presents an end-to-end churn analysis solution combining database management, business intelligence, and machine learning. By leveraging SQL Server, Power BI, and Python, the project analyses historical churn patterns and predicts future churners to support data-driven decision-making.

2. Target Audience

Although this project is based on a telecom customer dataset, the methodology and insights are applicable across multiple industries such as retail, finance, SaaS, and healthcare. This project is particularly relevant for data analysts, business intelligence professionals, and decision-makers focused on customer retention and growth strategies.

3. Project Objectives

The primary objectives of this project are:

- To build a complete ETL pipeline using SQL Server

- To analyze customer churn across demographic, geographic, account, and service dimensions

- To create an interactive Power BI dashboard for churn analysis

- To develop a machine learning model to predict future churners

- To provide actionable insights that support targeted retention strategies

4. Tools & Technologies Used

- **Database & ETL:** Microsoft SQL Server, SQL Server Management Studio (SSMS)

- **Visualization & BI:** Power BI, Power Query, DAX

- **Machine Learning:** Python (Jupyter Notebook), pandas, NumPy, scikit-learn

- **Model Used:** Random Forest Classifier

5. Data Preparation & ETL Process

Customer churn data was initially ingested into SQL Server using a staging table. Preliminary data exploration was performed to understand data distributions, detect missing values, and validate data consistency. Null values in service and churn-related columns were handled using appropriate default values to maintain data integrity.

After data cleaning, the transformed dataset was loaded into a production table. To support downstream analysis and prediction, two SQL views were created:

- A historical churn view containing churned and retained customers

- A new-joiner view containing recently onboarded customers

This structured ETL process ensured a reliable and scalable data foundation for analysis and modeling.

#	Customer_ID	Gender	Age	Married	State	Number_of_Referrals	Tenure_in_Months	Value_Deal	Phone_Service	Multiple_Lines	Internet_Service	Internet_Type	Online_Security	Online_Backup	Device_Protection_
1	11098-MAD	Female	30	Yes	Madhya Pradesh	0	31	Deal 1	Yes	No	Yes	Fiber Optic	Yes	Yes	No
2	11114-PUN	Male	51	No	Punjab	5	9	Deal 5	Yes	No	Yes	DSL	No	No	Yes
3	11167-WES	Female	43	Yes	West Bengal	3	28	Deal 1	Yes	Yes	Yes	Fiber Optic	Yes	Yes	Yes
4	11179-MAH	Male	35	No	Maharashtra	10	12	None	Yes	No	Yes	DSL	Yes	Yes	Yes
5	11180-TAM	Male	75	Yes	Tamil Nadu	12	27	Deal 2	Yes	No	Yes	DSL	Yes	No	No
6	11241-MAD	Female	41	Yes	Madhya Pradesh	4	11	None	Yes	No	Yes	Fiber Optic	No	Yes	Yes
7	11244-JAM	Female	20	No	Jammu & Kashmir	3	9	None	Yes	Yes	Yes	Cable	Yes	No	No
8	11251-UTT	Female	51	No	Uttarakhand	1	19	None	Yes	Yes	No	None	No	No	No
9	11262-HAR	Female	73	Yes	Haryana	5	32	None	Yes	Yes	Yes	Fiber Optic	No	No	No
10	11263-HAR	Female	41	No	Haryana	13	31	Deal 2	Yes	Yes	Yes	Fiber Optic	Yes	No	No
11	11264-MAH	Female	27	Yes	Maharashtra	14	17	Deal 5	No	No	Yes	DSL	No	No	No
12	11272-UTT	Female	65	No	Uttar Pradesh	0	19	Deal 5	Yes	Yes	Yes	Fiber Optic	No	No	No
13	11277-UTT	Male	66	Yes	Uttar Pradesh	10	23	None	Yes	No	Yes	DSL	No	Yes	Yes
14	11288-MAD	Male	52	No	Madhya Pradesh	6	24	None	Yes	No	No	None	No	No	No
15	11290-JAM	Female	70	Yes	Jammu & Kashmir	0	36	Deal 5	Yes	Yes	Yes	Fiber Optic	No	No	No
16	11301-WES	Female	31	No	West Bengal	7	5	Deal 3	Yes	Yes	Yes	Cable	No	No	No
17	11310-RAJ	Female	78	Yes	Rajasthan	0	15	Deal 2	Yes	Yes	Yes	Fiber Optic	No	No	Yes
18	11340-JAM	Female	21	No	Jammu & Kashmir	8	7	None	Yes	No	No	None	No	No	No
19	11348-MAH	Female	46	No	Maharashtra	11	19	Deal 3	Yes	No	Yes	Cable	Yes	Yes	No
20	11359-AND	Female	28	Yes	Andhra Pradesh	3	6	Deal 4	Yes	No	No	None	No	No	No
21	11370-TAM	Female	21	No	Tamil Nadu	15	10	Deal 4	Yes	No	Yes	Cable	No	No	No
22	11392-JAM	Female	39	Yes	Jammu & Kashmir	11	1	Deal 2	Yes	No	Yes	Fiber Optic	Yes	No	Yes
23	11392-KAR	Female	28	Yes	Karnataka	9	32	Deal 1	Yes	Yes	Yes	Fiber Optic	Yes	Yes	Yes
24	11410-AND	Male	80	No	Andhra Pradesh	1	4	None	Yes	Yes	Yes	Fiber Optic	Yes	No	No
25	11450-HAR	Male	60	Yes	Haryana	5	18	None	Yes	No	Yes	Cable	Yes	Yes	Yes
26	11464-WES	Male	25	No	West Bengal	12	5	None	Yes	Yes	Yes	Fiber Optic	No	Yes	No
27	11472-PUN	Female	33	No	Punjab	8	6	Deal 5	Yes	No	Yes	Cable	No	No	No
28	11474-TEL	Male	59	Yes	Telangana	0	18	None	Yes	No	Yes	Fiber Optic	Yes	No	No
29	11510-KER	Female	46	No	Kerala	10	4	None	Yes	No	Yes	Fiber Optic	No	No	No
30	11535-WES	Male	43	No	West Bengal	5	23	Deal 2	No	No	Yes	DSL	Yes	Yes	Yes
31	11540-DEL	Male	63	Yes	Delhi	15	15	Deal 2	Yes	Yes	Yes	Fiber Optic	No	Yes	Yes
32	11540-MAH	Female	50	Yes	Maharashtra	6	20	Deal 2	No	No	Yes	DSL	Yes	Yes	No
33	11543-WES	Female	35	Yes	West Bengal	0	7	None	Yes	No	Yes	DSL	Yes	No	No
34	11596-KAR	Female	40	No	Karnataka	15	11	Deal 4	Yes	No	Yes	Cable	Yes	No	No
35	11604-HAR	Female	38	Yes	Haryana	4	8	None	Yes	Yes	Yes	Fiber Optic	Yes	No	No

#	Premium_Support	Streaming_TV	Streaming_Movies	Streaming_Music	Unlimited_Data	Contract	Paperless_Billing	Payment_Method	Monthly_Charge	Total_Charges	Total_Refunds	Total_Extra_Data_Charges
1	Yes	No	Yes	Yes	Yes	Two Year	No	Bank Withdrawal	95.0999984741211	6683.39990234375	0	0
2	No	No	No	No	No	Month-to-Month	Yes	Bank Withdrawal	49.1500015258789	169.050003051758	0	10
3	Yes	Yes	Yes	Yes	No	Two Year	Yes	Bank Withdrawal	116.050003051758	8297.5	42.5699996948242	110
4	Yes	Yes	Yes	Yes	Yes	Two Year	Yes	Credit Card	84.4000015258789	5969.2998046875	0	0
5	Yes	Yes	Yes	No	No	Two Year	Yes	Credit Card	72.5999984741211	4084.35009765625	0	140
6	Yes	Yes	Yes	Yes	Yes	Month-to-Month	Yes	Bank Withdrawal	105.099998474121	3634.80004882813	0	0
7	No	Yes	Yes	Yes	No	One Year	No	Bank Withdrawal	76	1588.75	7.30000019073486	70
8	No	No	No	No	No	One Year	No	Bank Withdrawal	25.2000007629395	245.149993896484	0	0
9	No	Yes	Yes	No	Yes	Month-to-Month	Yes	Bank Withdrawal	95.0999984741211	95.0999984741211	0	0
10	No	Yes	Yes	Yes	Yes	One Year	Yes	Credit Card	99.6500015258789	6311.2001953125	0	0
11	No	No	No	Yes	Yes	Month-to-Month	Yes	Bank Withdrawal	-4	24.7999992370605	0	0
12	No	No	No	No	Yes	Month-to-Month	Yes	Bank Withdrawal	75.8499984741211	256.600006103516	0	0
13	No	No	No	No	Yes	One Year	Yes	Bank Withdrawal	54.7999923706605	731.299987792969	0	0
14	No	No	No	No	No	Two Year	Yes	Credit Card	20.2000007629395	917.450012207031	0	0
15	No	Yes	Yes	No	Yes	Month-to-Month	Yes	Bank Withdrawal	95.4499969482422	396.100006103516	0	0
16	No	No	No	No	Yes	One Year	Yes	Credit Card	51.5499992370605	1295.40002441406	0	0
17	No	Yes	Yes	No	Yes	One Year	Yes	Bank Withdrawal	102.150001525879	4735.35009765625	0	0
18	No	No	No	No	No	One Year	Yes	Bank Withdrawal	19.9500007629395	219.5	0	0
19	No	No	No	No	No	Month-to-Month	No	Bank Withdrawal	56.0499992370605	1522.65002441406	0	110
20	No	No	No	No	No	Two Year	No	Credit Card	20.5	290.549987792969	0	0
21	Yes	No	Yes	Yes	Yes	One Year	No	Credit Card	59.0999984741211	772.849975585938	0	0
22	Yes	No	No	No	Yes	One Year	No	Bank Withdrawal	84.4000015258789	4484.0498046875	0	0
23	Yes	Yes	Yes	Yes	Yes	Two Year	Yes	Bank Withdrawal	112.550003051758	7806.5	0	0
24	No	Yes	Yes	Yes	Yes	One Year	No	Credit Card	101.550003051758	5070.39990234375	0	0
25	Yes	Yes	No	No	Yes	Two Year	No	Credit Card	75.5500030517578	4707.85009765625	0	0
26	No	Yes	Yes	Yes	Yes	Month-to-Month	Yes	Credit Card	99.8499984741211	1992.55004882813	0	0
27	No	No	Yes	Yes	Yes	Month-to-Month	Yes	Bank Withdrawal	55.4000015258789	55.4000015258789	0	0
28	No	No	No	No	Yes	Month-to-Month	Yes	Bank Withdrawal	75.5999984741211	661.549987792969	0	0
29	Yes	Yes	Yes	Yes	Yes	Month-to-Month	Yes	Bank Withdrawal	95.9000015258789	827.450012207031	0	0
30	No	Yes	Yes	Yes	Yes	Two Year	Yes	Bank Withdrawal	58.75	3038.55004882813	0	0
31	No	Yes	Yes	Yes	Yes	Month-to-Month	Yes	Bank Withdrawal	104.099998474121	5135.14990234375	0	0
32	No	No	Yes	Yes	Yes	Two Year	No	Bank Withdrawal	46.2000007629395	2431.94995117188	0	0
33	No	No	No	No	Yes	Month-to-Month	No	Bank Withdrawal	49.4000015258789	874.799987792969	0	0
34	No	No	No	No	No	Month-to-Month	Yes	Bank Withdrawal	49.2000007629395	571.150024414063	0	80
35	Yes	Yes	Yes	Yes	Yes	One Year	Yes	Bank Withdrawal	104.400001525879	6721.60009765625	0	0

Total_Long_Distance_Charges	Total_Revenue	Customer_Status	Churn_Category	Churn_Reason
631.719970703125	7315.1201171875	Stayed	Others	Others
122.370002746582	301.420013427734	Churned	Competitor	Competitor had better devices
1872.97998046875	10237.91015625	Stayed	Others	Others
219.389999389648	6188.68994140625	Stayed	Others	Others
332.079986572266	4556.43017578125	Stayed	Others	Others
95.879997253418	3730.67993164063	Stayed	Others	Others
305.799987792969	1957.25	Stayed	Others	Others
20.5699996948242	265.720001220703	Stayed	Others	Others
4.8899998664856	99.9899978637695	Churned	Price	Price too high
90.0899963378906	6401.2900390625	Stayed	Others	Others
0	24.7999992370605	Churned	Dissatisfaction	Product dissatisfaction
133.020004272461	389.619995117188	Churned	Competitor	Competitor offered more data
136.630004882813	867.929992675781	Stayed	Others	Others
2112.32006835938	3029.77001953125	Stayed	Others	Others
126.440002441406	522.539978027344	Churned	Competitor	Competitor made better offer
641.419982910156	1936.81994628906	Stayed	Others	Others
1535.40002441406	6270.75	Stayed	Others	Others
78.5999984741211	298.100006103516	Stayed	Others	Others
1235.07995605469	2867.72998046875	Stayed	Others	Others
435.839996337891	726.390014648438	Stayed	Others	Others
476.140014648438	1248.98999023438	Stayed	Others	Others
1655.64001464844	6139.68994140625	Stayed	Others	Others
235.289993286133	8041.7900390625	Stayed	Others	Others
1217.16003417969	6287.56005859375	Stayed	Others	Others
398.790008544922	5106.64013671875	Stayed	Others	Others
522.059997558594	2514.61010742188	Stayed	Others	Others
13.0100002288818	68.4100036621094	Churned	Competitor	Competitor had better devices
189.990005493164	851.539978027344	Stayed	Others	Others
277.200012207031	1104.65002441406	Stayed	Others	Others
0	3038.55004882813	Stayed	Others	Others
2288.42993164063	7423.580078125	Stayed	Others	Others
0	2431.94995117188	Stayed	Others	Others
375.480010986328	1250.28002929688	Churned	Other	Deceased
114.959999084473	766.109985351563	Stayed	Others	Others
892.799987792969	7614.39990234375	Stayed	Others	Others

Figure 1: Cleaned customer churn data loaded into SQL Server

6. Power BI Data Transformation

Within Power BI, additional analytical features were created to enhance insights. These included:

- Churn status indicators

- Monthly charge ranges

- Age group and tenure group categorizations

- Service usage transformation through unpivoting

These transformations enabled flexible slicing of churn metrics across multiple business dimensions.

7. Key Metrics & Measures

The following business metrics were developed using DAX:

- Total Customers

- New Joiners

- Total Churn

- Churn Rate

These metrics formed the foundation of the interactive dashboard and allowed stakeholders to monitor churn trends effectively.

8. Churn Analysis & Visualisation

An interactive Power BI dashboard was developed to analyze historical churn patterns. The dashboard includes:

- An executive summary highlighting key churn KPIs

- Demographic analysis based on gender and age groups
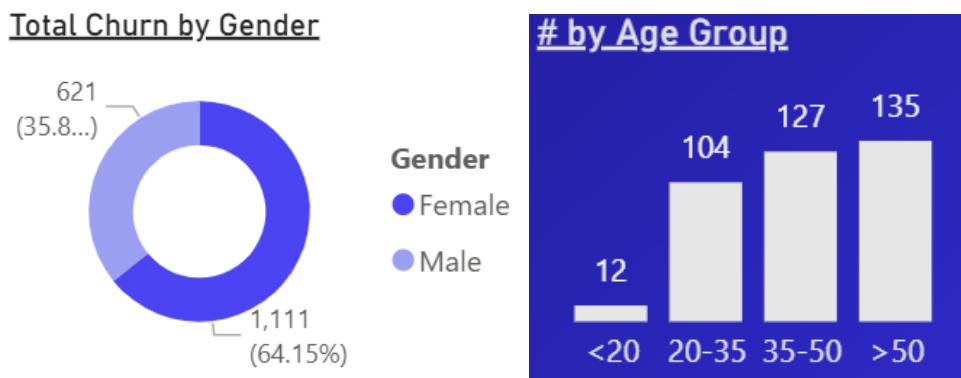
Figure 2: Churn distribution across demographic segments

- Account-level insights by payment method, contract type, and tenure
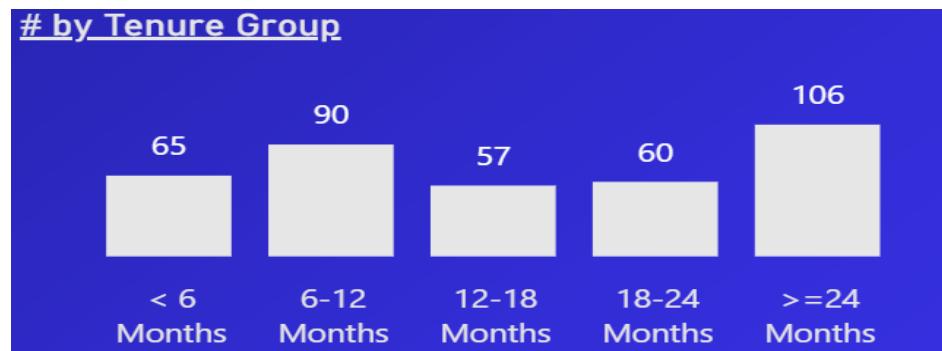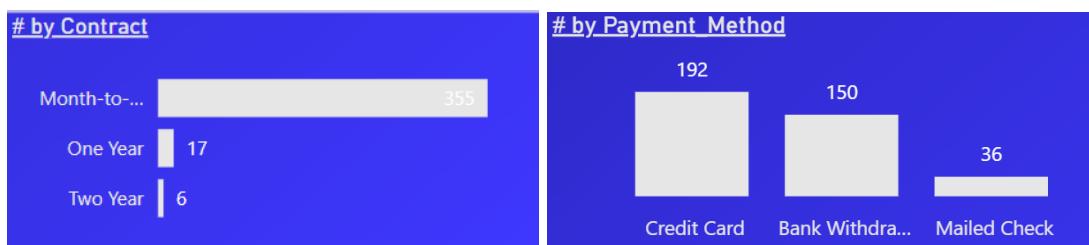
Figure 3: Impact of account and tenure characteristics on churn

- Geographic churn distribution across states
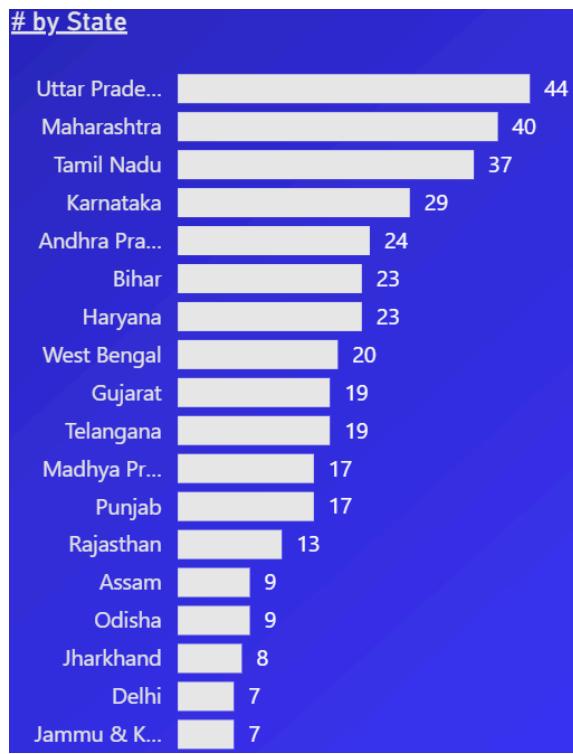
Figure 4: Geographic distribution of churn rates

● Service usage analysis identifying services associated with higher churn
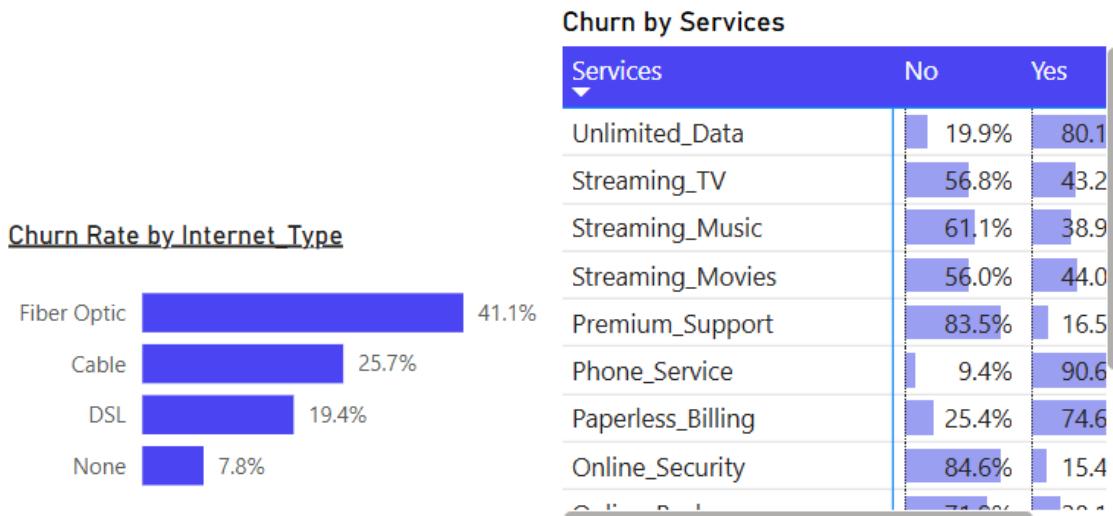

Figure 5: Service usage patterns associated with customer churn

● Churn category and churn reason distribution using tooltips

This visualization layer provides a clear understanding of *who is churning, where they are churning, and why*.

Figure 6: Executive summary of customer churn metrics

9. Predictive Modeling for Churn

To predict future churners, a Random Forest classification model was developed using Python. The model was chosen due to its robustness, ability to handle mixed data types, and strong performance in classification problems.

The modeling process involved:

- Encoding categorical variables

- Splitting data into training and testing sets

- Training the Random Forest classifier

- Evaluating performance using a confusion matrix and classification report

- Identifying key churn drivers using feature importance analysis

The model enables proactive identification of customers at high risk of churn.

```
Confusion Matrix:
[[783  64]
 [126 229]]
```

Figure 7: Random Forest churn prediction model performance

10. Churn Prediction Insights

The trained model was applied to newly joined customers to predict potential churn. Predicted churn results were exported and integrated into Power BI to create a dedicated churn prediction page.

This page highlights:

- High-risk customers

- Predicted churn distribution across demographics

- Account and contract-based churn patterns

- Geographic concentration of predicted churners

These insights allow businesses to prioritize intervention strategies and design targeted retention campaigns.
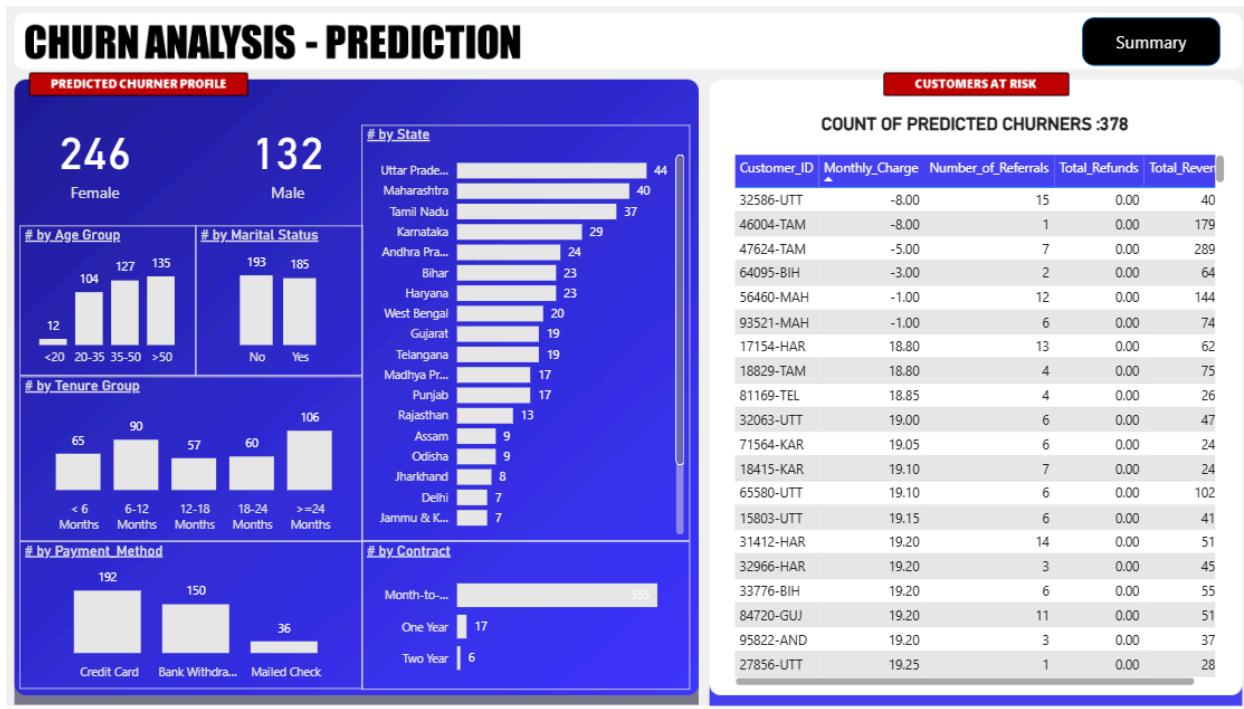
Figure 8: Predicted churners and risk segmentation

11. Business Impact

This project demonstrates how data analytics and machine learning can:

- Reduce customer attrition through early identification of churn risks

- Support data-driven marketing and retention strategies

- Improve decision-making through actionable insights

- Enable organizations to shift from reactive to proactive customer management

12. Conclusion

The End-to-End Churn Analysis project delivers a comprehensive analytics solution that combines ETL, visualization, and predictive modeling. By integrating historical analysis with future churn prediction, the project provides meaningful insights that support customer retention

and business growth. The approach is scalable and can be adapted across industries with minimal modification.

13. Appendix – Project Deliverables

- SQL scripts for ETL and data cleaning

- Power BI dashboard (.pbix) with interactive visualizations

- Python Jupyter Notebook for churn prediction modeling

- Feature importance analysis outputs

- Churn prediction results exported as CSV