Customer Churn Analysis using Python & Power BI

1. Project Overview

Customer churn is a major challenge for businesses, especially in the banking and financial sector, where acquiring new customers is more expensive than retaining existing ones. Understanding why customers leave and identifying high-risk segments is critical for improving customer retention.

This project focuses on analyzing customer churn data to identify key churn drivers and high-value at-risk customers using **Python** for data analysis and **Power BI** for interactive dashboard visualization.

---

2. Problem Statement

The objective of this project is to analyze customer data and answer the following business questions:

- What is the overall customer churn rate?
- Which customer segments are more likely to churn?
- How do factors such as tenure, activity status, product usage, balance, and credit score affect churn?
- Which high-value customers are at the highest risk of leaving?

The insights from this analysis can help businesses design targeted retention strategies and reduce customer churn.

---

3. Dataset Description

The dataset used in this project is a publicly available **Customer Churn Modelling dataset** obtained from Kaggle.

The dataset contains customer-level information including:

- **Demographic data:** Age, Gender, Geography
- **Account information:** Tenure, Balance, Credit Score, Estimated Salary
- **Product usage:** Number of Products, Has Credit Card
- **Engagement indicators:** Is Active Member
- **Target variable:** Exited (1 = Customer churned, 0 = Customer retained)

---

4. Tools & Technologies Used

- **Python:** Pandas, Matplotlib (for data cleaning and exploratory analysis)
- **Power BI Desktop:** Dashboard development and visualization

- **DAX:** KPI and churn metric calculations
- **Microsoft Excel:** Initial data inspection
- **Kaggle:** Dataset source

---

5. Data Cleaning & Preparation (Python)

Data preparation was performed using Python to ensure accuracy and consistency before visualization.

Key steps included:

- Loading the dataset using Pandas
- Checking and handling missing values
- Removing duplicate records
- Verifying and correcting data types
- Creating derived features such as:
    - Age groups
    - Balance buckets

A cleaned dataset was exported and used as the input source for Power BI.

---

6. Dashboard Design & Structure

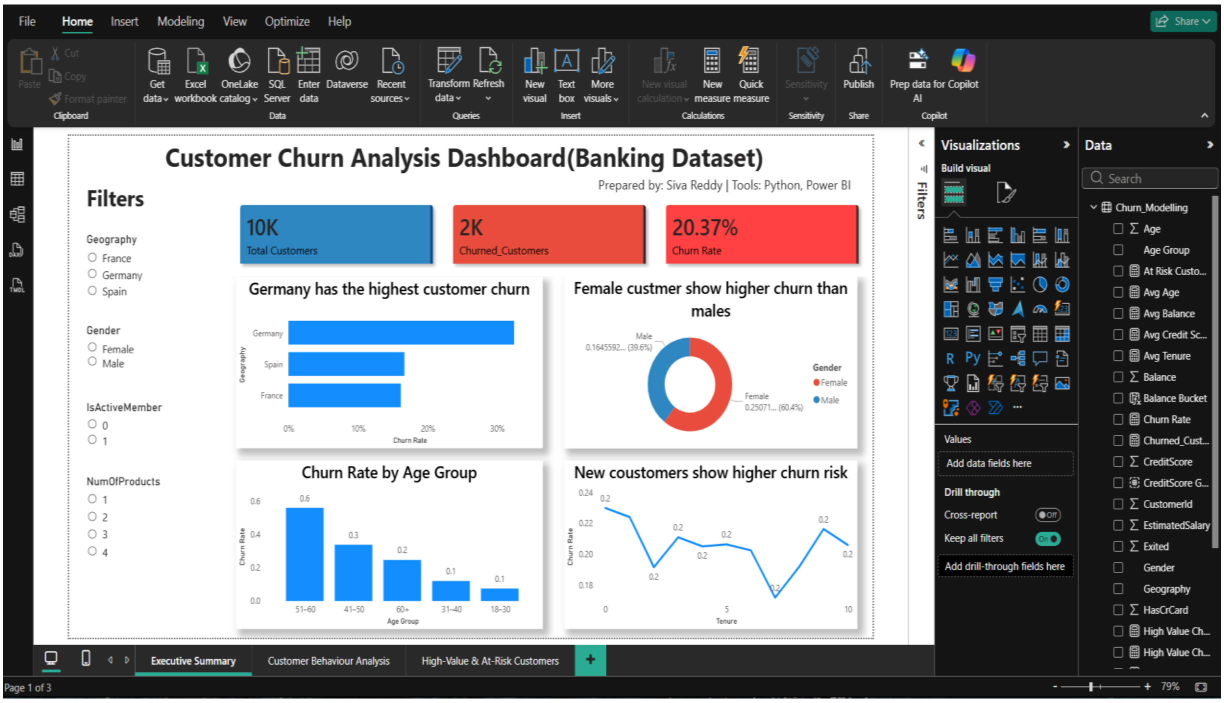
The Power BI dashboard is structured into three logical pages to guide stakeholders from high-level insights to actionable customer-level analysis.

Page 1: Executive Summary

Purpose: Provide a high-level overview of customer churn.

Key elements:

- Total Customers
- Overall Churn Rate
- Churned Customers
- Churn distribution by Geography, Gender, and Age Group
- Interactive slicers for filtering analysis



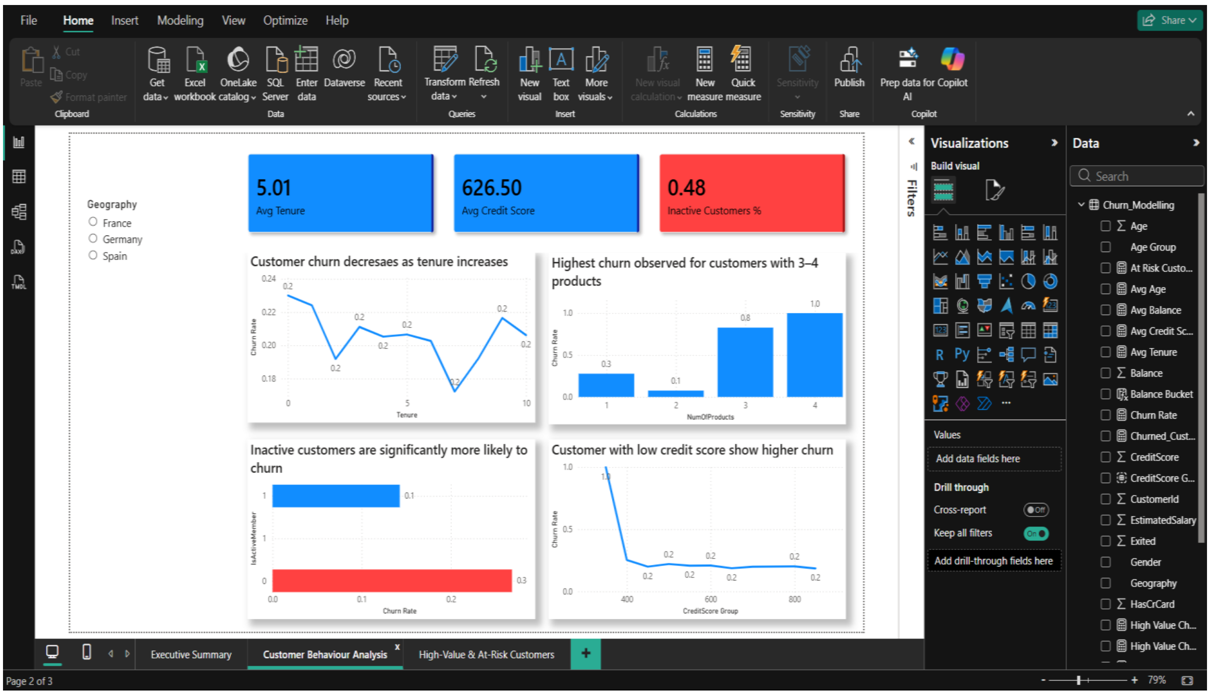
Page 2: Customer Behaviour Analysis

Purpose: Identify key factors influencing customer churn.

Key analyses:

- Churn rate by Tenure
- Churn rate by Number of Products
- Churn comparison between Active and Inactive customers
- Churn trends across Credit Score groups

This page focuses on understanding **why customers churn**.



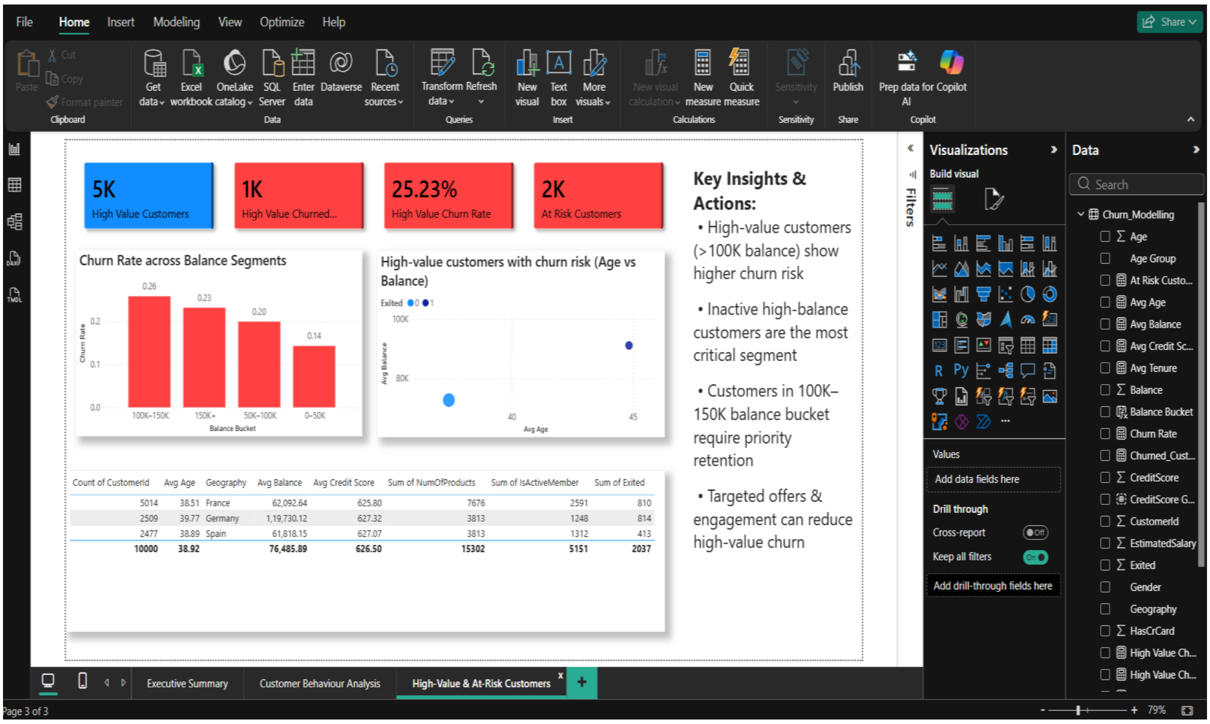
Page 3: High-Value & At-Risk Customers

Purpose: Identify customer segments that require immediate retention action.

Key analyses:

- High-value customers based on account balance
- High-value customer churn rate
- At-risk customers (high balance and inactive)
- Churn rate across balance segments
- Scatter analysis of Age vs Balance to visualize churn risk
- Detailed customer table for targeted decision-making

This page answers **who the business should focus on**.



---

7. Key Insights

The analysis revealed several important insights:

- Customer churn decreases as tenure increases, indicating higher risk among new customers.
- Inactive customers are significantly more likely to churn compared to active customers.
- Customers with 3 or more products show the highest churn rates.
- Low credit score customers have a higher probability of churning.
- High-value customers (balance > 100K) exhibit a higher churn risk, especially when inactive.

---

8. Business Recommendations

Based on the insights, the following recommendations are suggested:

- Implement onboarding and engagement programs for new customers.
- Design targeted retention campaigns for inactive customers.
- Introduce loyalty benefits for high-value customers.
- Offer personalized product recommendations for customers with low credit scores.
- Monitor high-risk segments proactively to reduce future churn.

---

9. Conclusion

This project demonstrates an end-to-end data analysis workflow using Python and Power BI, from data cleaning and exploration to dashboard creation and business insight generation.

The interactive dashboards enable stakeholders to explore churn patterns dynamically and identify high-risk customer segments. The findings can support data-driven decision-making and help organizations reduce customer churn through targeted retention strategies.

SQL Analysis for Customer Churn

SQL queries were written to validate churn metrics and customer segmentation used in the Power BI dashboard. These queries helped analyze churn behavior across geography, tenure, activity status, and high-value customers.

```
# To get total table...

select * from churn_modelling;


# Query to find the total churn percentage...

select (select count(*) from churn_modelling

where exited = 1)* 100 / (select count(*) from churn_modelling) as churn_percentage;


# Query to find country wise churn rate which is having high churn percentage...

SELECT Geography, ROUND(100 * SUM(Exited)/COUNT(*), 2) AS churn_rate_pct

FROM (

  SELECT * FROM Churn_Modelling

) AS t

WHERE Geography IN ('France','Spain','Germany')

GROUP BY Geography

ORDER BY churn_rate_pct DESC;


# Query to check whether churn rate is higher among older customers compared to younger customers

SELECT age_group, ROUND(100 * SUM(Exited)/COUNT(*), 2) AS churn_rate_pct, COUNT(*) as customers

FROM (

  SELECT *,

    CASE

      WHEN Age BETWEEN 18 AND 30 THEN '18-30'

      WHEN Age BETWEEN 31 AND 40 THEN '31-40'

      WHEN Age BETWEEN 41 AND 50 THEN '41-50'

      WHEN Age BETWEEN 51 AND 60 THEN '51-60'

      ELSE '60+' END AS age_group

  FROM Churn_Modelling
```

```
) AS t
GROUP BY age_group
ORDER BY age_group ;
```

Query to find is churn higher among new customers (0–2 years) or long-term customers...

```
SELECT tenure_group, ROUND(100 * SUM(Exited)/COUNT(*), 2) AS churn_rate_pct, COUNT(*) AS customers
FROM (
 SELECT *,
  CASE
   WHEN Tenure BETWEEN 0 AND 2 THEN '0-2'
   WHEN Tenure BETWEEN 3 AND 5 THEN '3-5'
   WHEN Tenure BETWEEN 6 AND 9 THEN '6-9'
   ELSE '10+' END AS tenure_group
 FROM Churn_Modelling
) AS t
GROUP BY tenure_group
ORDER BY FIELD(tenure_group, '0-2','3-5','6-9','10+');
```

Query to find Do customers with high account balances churn more or less...

```
SELECT balance_bucket, ROUND(100 * SUM(Exited)/COUNT(*), 2) AS churn_rate_pct, COUNT(*) AS customers
FROM (
 SELECT *,
  CASE
   WHEN Balance = 0 THEN '0'
   WHEN Balance BETWEEN 1 AND 50000 THEN '1-50k'
   WHEN Balance BETWEEN 50001 AND 100000 THEN '50k-100k'
   ELSE '100k+' END AS balance_bucket
 FROM Churn_Modelling
) AS t
GROUP BY balance_bucket
ORDER BY FIELD(balance_bucket, '0','1-50k','50k-100k','100k+');
```

```sql
# Query to find Is low credit score linked to high churn...

SELECT credit_bucket, ROUND(100 * SUM(Exited)/COUNT(*), 2) AS churn_rate_pct, COUNT(*) AS customers

FROM (

 SELECT *,

  CASE

   WHEN CreditScore < 550 THEN 'Low (<550)'

   WHEN CreditScore BETWEEN 550 AND 699 THEN 'Medium (550-699)'

   ELSE 'High (700+)' END AS credit_bucket

 FROM Churn_Modelling

) AS t

GROUP BY credit_bucket

ORDER BY FIELD(credit_bucket, 'Low (<550)','Medium (550-699)','High (700+)');
```

```sql
# Query to find Are inactive customers more likely to leave...

SELECT IsActiveMember,

       ROUND(100 * SUM(Exited)/COUNT(*), 2) AS churn_rate_pct,

       COUNT(*) AS customers

FROM (

 SELECT * FROM Churn_Modelling

) AS t

GROUP BY IsActiveMember

ORDER BY IsActiveMember DESC;
```

```sql
# Query to find Does number of products influence churn...

SELECT NumOfProducts, ROUND(100 * SUM(Exited)/COUNT(*), 2) AS churn_rate_pct, COUNT(*) AS customers

FROM (

 SELECT * FROM Churn_Modelling

) AS t

GROUP BY NumOfProducts

ORDER BY NumOfProducts;
```

Query to find Among customers with balance > 100,000, what is the churn rate...

```
SELECT
  COUNT(*) AS customers,
  SUM(Exited) AS churned,
  ROUND(100 * SUM(Exited)/COUNT(*), 2) AS churn_rate_pct
FROM (
  SELECT * FROM Churn_Modelling
) AS t
WHERE Balance > 100000;
```

Query to find Do male or female customers churn more...

```
SELECT Gender, ROUND(100 * SUM(Exited)/COUNT(*), 2) AS churn_rate_pct, COUNT(*) AS customers
FROM (
  SELECT * FROM Churn_Modelling
) AS t
GROUP BY Gender
ORDER BY churn_rate_pct DESC;
```

10. Author

Project by: Peram Venkata Siva Reddy
Role Targeted: Data Analyst
Tools Used: Python, Power BI, DAX