Bank Customer Churn Analysis and Risk Segmentation

❖ Abstract :

This project aims to predict customer churn in a bank to hep the bank to retain their customer. By analyzing customer demographics, account behavior, and financial attributes, we identify high-risk segments. Data was processed using SQL and Python, while Power BI dashboards provide actionable insights to support retention strategies and reduce customer attrition through targeted interventions.

❖ Introduction:

Customer retention is a critical factor for the long-term success of any banking institution. This project focuses on predicting customer churn .The likelihood that a customer will leave the bank by analyzing various customer attributes such as age, tenure, balance, product usage, and account activity. Using machine learning techniques, we aim to classify customers as potential churners or loyal clients. The data is extracted and explored using PostgreSQL, preprocessed and modeled in Python, and the results are visualized through interactive dashboards in Power BI. These dashboards highlight churn distribution, high-risk segments, and key drivers influencing churn. By understanding the patterns behind customer attrition, the bank can take proactive steps to retain valuable customers, improve service strategies, and optimize marketing efforts. This end-to-end project provides a data-driven approach to enhancing customer satisfaction and reducing churn rates in the competitive banking landscape.

❖ Data Overview:

The dataset used in this project contains information on 10,000 bank customers, with 14 features including demographics, account activity, and financial status. Key columns include Age, Gender, Geography, Tenure, Balance, Number of Products, Credit Card status, and whether the customer is an active member. The target variable Exited indicates if the customer has churned. This structured dataset enables detailed exploratory analysis and supports predictive modeling to identify patterns and risk factors associated with customer attrition.

❖ Tools:

PostgreSQL

- **Version** : PostgreSQL 16 , Windows 11
- **Description**: PostgreSQL is a powerful, open-source object-relational database system known for its reliability, feature robustness, and standards compliance.

Jupyter Notebook

- **Version** : JupyterLab 4.0 / Notebook 7.1 (Latest release in 2025)
- **Python Version**: Python 3.11
- **Description**: Jupyter Notebook provides an interactive environment for writing and running code, visualizing data, and documenting analysis.

Power BI Desktop

- **Version** : Power BI Desktop May 2025 Update (Version 2.128.x)
- **Description**: Power BI is a business analytics tool that enables data transformation, modeling, and rich interactive visualizations

❖ Data Preprocessing :

To use the data effectively, preprocessing was performed using Power Query in Power BI. This process followed the ETL (Extract, Transform, Load) approach. Data was extracted from PostgreSQL and connected to Power BI. In the transformation stage, data types were adjusted, queries were merged, tables were joined, and unnecessary values were handled by removing nulls, duplicates, and replacing values where needed to ensure clean and consistent data for analysis.

❖ Data Analysis :

The data analysis phase involved examining patterns and relationships between customer attributes and churn behavior. Using SQL, we explored churn rates across gender, geography, account balance, tenure, product usage, and salary levels. Key aggregations and groupings helped identify high-risk customer segments. In Python, further exploratory data analysis (EDA) was conducted to visualize feature distributions, correlations, and outliers using libraries like Pandas, Seaborn, and Matplotlib. We also checked for class imbalance and feature relationships with the target variable. These insights guided feature engineering and model development for accurate churn prediction and customer risk profiling.

❖ Predictive Model:

To predict churn probability, we first preprocessed the dataset by encoding categorical variables and scaling numerical features. We split the data into training and testing sets, then trained a classification model using algorithms like Random Forest. The model learned patterns from customer features such as age, balance, tenure, and activity status to predict the likelihood of churn. Predictive modeling is used to proactively identify high-risk customers, enabling the bank to take timely retention actions. It supports data-driven decision-making and helps reduce customer attrition effectively.

Classifiaction Model : In this project, a Random Forest Classifier was used as the classification model to predict customer churn probability. By training on key features like age, tenure, balance, income**,**

and credit score, the model accurately identified high-risk customers, enabling targeted retention strategies based on data-driven insights.

	precision	recall	f1-score	support
0	0.88	0.96	0.92	1607
1	0.76	0.48	0.59	393
accuracy			0.87	2000
macro avg	0.82	0.72	0.75	2000
weighted avg	0.86	0.87	0.85	2000

Precision: Measures how many predicted positive cases are actually positive.
Formula: TP / (TP + FP)
High precision = fewer false positives.

Recall: Measures how many actual positive cases were correctly predicted.
Formula: TP / (TP + FN)
High recall = fewer false negatives.

F1-Score: Harmonic mean of precision and recall. Balances both.
Formula: 2 × (Precision × Recall) / (Precision + Recall)

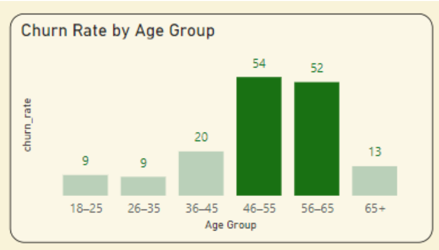
Support: The number of actual occurrences of each class in the dataset.
Indicates how many samples were evaluated per class.

❖ Insights From Analysis

A significant portion of the customer base 1 out of every 5 customers has churned, indicating the need for a stronger retention strategy.

Customers from Germany show the highest churn rate geographically, suggesting possible dissatisfaction or market-specific issues needing immediate attention.

Customers in the mid-age bracket exhibit a higher churn tendency, possibly due to changing financial goals or unmet service expectations.

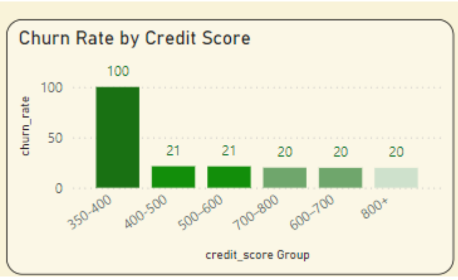


Newly joined customers (within 1 year) have the highest churn rate, signaling a need for better onboarding and early engagement strategies.

Surprisingly, even long-tenured customers are churning, emphasizing that loyalty alone doesn't guarantee retention—continuous value delivery is essential.

Both low balance (< ₹25,000) and very high balance (> ₹2,00,000) customers show elevated churn rates, suggesting dissatisfaction across different economic segments.

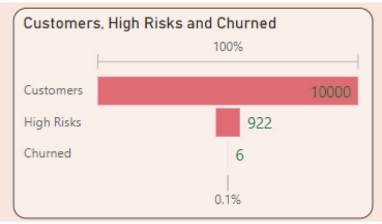
A sharp rise in churn is seen in customers with low credit scores (350–400)—likely due to financial instability or lack of tailored financial products.



Customers earning between ₹1–2 Lakhs are churning more compared to other income groups, indicating possible pricing or service-value mismatches.

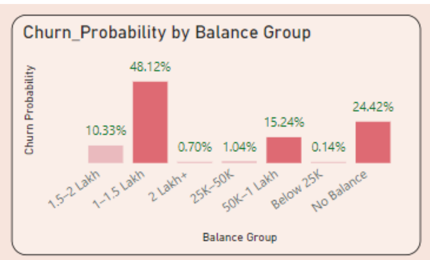
❖ Churn Risk Tracker

Your classification model flags 922 customers with high churn probability, offering a focused target for proactive retention efforts.



The scatter plot of risk level vs. age vs. churn probability reveals that middle-aged individuals, both male and female, are the most vulnerable segment.

Customers holding a balance between ₹1 to ₹1.5 lakhs show higher churn probability than those with zero balance—suggesting they may feel underserviced despite moderate savings.



Surprisingly, customers with 8–10 years of tenure are showing the highest churn probability, signaling long-term fatigue or declining satisfaction.

Individuals earning between ₹1 to ₹2 lakhs stand out as the most likely to churn, indicating a crucial need to revisit product fit and value offerings for this income group.

Customers with an average credit score of 642 are appearing frequently in the high-risk churn category, possibly due to limited credit access or financial stress.

❖ Data Storytelling With Dashboard

In this project, I utilized an interactive Power BI dashboard to visually narrate the customer churn story. By integrating key metrics such as churn rate, customer demographics, tenure, balance, credit score, and income, the dashboard provided a comprehensive view of churn behavior. Visual elements like KPI cards, funnel charts, scatter plots, and slicers allowed users to explore risk levels by geography, age group, and financial profile. The classification model outputs were embedded to highlight high-risk segments, helping to pinpoint vulnerable customer groups. This visual storytelling enabled data-driven insights, turning complex patterns into an intuitive, decision-friendly interface that empowered stakeholders to understand churn dynamics and take proactive retention actions effectively.

❖ Decision and Advice to Prevent Churn

Launch personalized retention campaigns (e.g., loyalty benefits, special offers) targeting the high-risk segment identified by the classification model.

Since 0–1 year tenure shows high churn, implement welcome journeys, onboarding support, and follow-up calls within the first 3–6 months.

Investigate region-specific factors behind Germany's high churn rate and tailor offerings or support services to that market.

Reassess pricing, service value, or benefits for mid-income customers to ensure affordability without sacrificing value.

Offer targeted financial advisory or perks to customers with this balance range who feel underserved despite holding assets.

Launch a "Loyalty Appreciation Program" or personalized feedback loop for long-tenured users at risk of churn.

Introduce credit improvement plans or lower-risk products for customers with weaker credit profiles to build trust and long-term relationships.

Since this age group shows high churn, use targeted lifecycle messaging and flexible offerings to retain their interest.

Continuously track risk levels using dashboards and enable data-driven intervention models with real-time alerts.

❖ Conclusion

In conclusion, this project provided valuable insights into customer churn patterns through data analysis and predictive modeling. By identifying key churn drivers such as tenure, age, balance, credit score, and income level, targeted strategies can now be implemented to reduce churn. The interactive dashboard served as a powerful storytelling tool, enabling clear visualization of risk segments. These findings empower stakeholders to make data-driven decisions, improve customer retention, and enhance overall business performance through focused and proactive engagement strategies.

❖ References

1. **Storytelling with Data**– by Cole Nussbaumer Knaflic

2. **Hands-On Data Analysis with Pandas** – by Stefanie Molin

3. **Krish Naik** YT Channel Machine learning classifiaction