Churn Modelling Analysis

# Executive Summary

This project involves analyzing customer churn to identify patterns, trends, and actionable insights that can help reduce customer attrition. We used a combination of Python (Jupyter Notebook) and Power BI for data analysis and visualization.

# Problem Statement

Customer churn refers to when a customer stops doing business with a company. The objective of this analysis is to understand the key drivers of churn using data insights and machine learning, and provide recommendations to improve customer retention.

# Data Overview

The dataset used in this analysis is `Churn\_Modelling.csv`, which was loaded and stored into an SQLite database. It contains demographic and account-related attributes for 10,000 customers, including:  
- Customer ID  
- Gender  
- Age  
- Geography  
- Tenure  
- Balance  
- Credit Score  
- Estimated Salary  
- Exited (churn indicator)

# Exploratory Data Analysis (SQL-based)

Key findings from SQL queries performed on the dataset include:

* - Churn Rate by Gender: Churn is slightly higher among females than males.
* - Age & Churn: Customers over age 50 showed significant churn rates.
* - Tenure vs Churn: Churn patterns were observed across different tenure durations.
* - Top 5 Ages with Highest Churn: Specific age groups churned more frequently.
* - Top 5 High Salary Customers Who Churned: Some high earners still exited, indicating financial strength alone is not predictive.
* - Average Customer Age: 42.37
* - Average Estimated Salary: $100193.91
* - Average Credit Score: 651.85
* - Geography with Highest Retention: Germany (example based on output)

# Power BI Dashboard And Jupyter Notebook (Python-SQL) Insights

# 10 Important Analyses from the Jupyter Notebook

1. **Gender-Based Churn Analysis**
   * **Male customers** have a **16.47% churn rate**, with **899 out of 5,458** males churning.
   * **Insight:** While males churn less frequently than females (based on the query structure), targeted retention strategies for high-risk genders are needed.
2. **Age and Churn Correlation**
   * **Top 5 ages with highest churn:**
     + **46 years (91 churns)**, 40 years (89), 43 years (88), 45 years (87), 48 years (80).
   * **Insight:** Customers aged **40–50** are most likely to churn, suggesting mid-life financial or service dissatisfaction.
3. **High-Value Churned Customers**
   * **Top 5 churned customers by salary** earned **€199K–€199.8K**. Most were from **Germany (3/5)**.
   * **Insight:** High earners are leaving, especially in Germany. Investigate premium service gaps or competitor offers.
4. **Geography Impact on Retention**
   * **Germany** has the **highest retention** (likely due to higher salaries or better service).
   * **Spain** has the **lowest retention** (highest churn linked to salary or engagement issues).
   * **Insight:** Regional strategies are critical—Spain may need improved customer engagement.
5. **Average Customer Metrics**
   * **Average Age:** **38.92 years**
   * **Average Salary:** **€100,083**
   * **Average Credit Score:** **650.56**
   * **Insight:** Typical customer profile: late-30s, moderate salary, fair credit score.
6. **Tenure-Based Churn (Partial Data)**
   * The notebook includes a query for tenure-based churn, but results are incomplete.
   * **Insight:** Tenure analysis could reveal if long-term or short-term customers are more likely to leave.
7. **Active vs. Inactive Members**
   * Queries hint at **IsActiveMember** impact but lack explicit results.
   * **Insight:** Inactive members likely churn more. Re-engagement campaigns could help.
8. **Credit Score Trends**
   * Average credit score is **650.56**, indicating moderate risk.
   * **Insight:** Customers with lower scores may need financial incentives to stay.
9. **Data Quality Issues**
   * **Null values** in HasCrCard and Geography (e.g., Row 7 has Geography = None).
   * **Insight:** Cleanse data to avoid skewed analysis, especially for geography-based insights.
10. **Duplicate Records**
    * Rows **9999–10001** show duplicate entries (e.g., Customer ID **15628319** appears twice).
    * **Insight:** Deduplicate data to ensure accurate churn rate calculations.

# 10 Important Points from Page 1 with Detailed Analysis:

1. **Total Customer Churn (Exited):**
   * Total churn is **1,538** out of **6,380** customers (~24.1% churn rate).
   * **Germany** has the highest churn (815), followed by **France (473)** and **Spain (250)**. This suggests Germany may have service or competitive issues.
2. **Customer Base Size:**
   * **6,380 customers** in total, providing a substantial dataset for analysis.
3. **Credit Score Insights:**
   * **Maximum CreditScore is 850**, indicating some customers have excellent credit. This could be segmented for retention strategies.
4. **Salary Distribution:**
   * **Highest EstimatedSalary is €199.97K**, useful for understanding high-value customers at risk of churn.
5. **Average Tenure:**
   * **4.98 years** average tenure, implying many customers are mid-term. Low tenure customers may churn faster.
6. **Total Balance:**
   * **€764.66M** total balance, showing significant deposits held. High-balance customers may need targeted retention.
7. **Salary by Geography (Sum):**
   * **France (€261.11M)**, **Germany (€253.79M)**, and **Spain (€250M)** have comparable salary sums, but Germany’s higher churn suggests non-salary factors drive attrition.
8. **Agent Metrics (Avg/Max/Min):**
   * Agent-related data (e.g., average, max, min) is listed but unclear without values. Potentially tied to customer service performance.
9. **Age by Geography:**
   * **Germany** has the highest age-related bars (79), possibly indicating older customers churn more. **France (39)** and **Spain (18)** show younger demographics.
10. **Visual Trends:**
    * Graphs suggest **Germany** has the highest metrics in multiple categories (churn, age, etc.), warranting deeper investigation into local market dynamics.

# 10 Important Points from Page 2 with Detailed Analysis:

1. **Balance by Gender:**
   * **Males** have slightly higher average balance (**€120.51K**) than **females (€118.98K)**. Gender-based retention strategies may differ.
2. **Active Members by Geography:**
   * **Sum of IsActiveMember** is visualized but lacks exact numbers. Geographic trends (e.g., Europe vs. Asia) could reveal engagement disparities.
3. **Churn by Gender & Exit Status:**
   * **3.24K (50.69%) exited males** vs. **3.07K (48.14%) exited females**. Near-equal churn, but males are marginally higher.
4. **Active/Inactive Segmentation:**
   * **High Balance - Active (1,521)** and **Low Balance - Active (1,599)** dominate, while inactive segments are smaller. Targeting inactive customers could reduce churn.
5. **Global Distribution:**
   * Mentions of **South America, Europe, Asia**, and **Atlantic Ocean** hint at multinational operations, requiring region-specific strategies.
6. **Customer Segments:**
   * Four segments: **High/Low Balance** × **Active/Inactive**. High-balance inactive customers are prime targets for re-engagement.
7. **Churn Rate Calculation:**
   * **0.63% (0.04K)** of a subgroup (possibly new customers) churned, but context is unclear. Further segmentation needed.
8. **Visual Data on Activity:**
   * Bar charts show **active members** are concentrated in certain regions/categories, but exact numbers are missing.
9. **Gender-Based Insights:**
   * **Female (0)** and **Male (1)** labels in a graph may indicate binary gender segmentation, with males slightly more prone to churn.
10. **Unlabelled Metrics:**
    * Some graphs (e.g., "6.59") lack context, suggesting incomplete data presentation. Clarification is needed for actionable insights.

# Summary of Key Findings:

* **Germany** has the highest churn, possibly linked to age or service issues.
* **High-balance inactive customers** are critical for retention efforts.
* **Gender differences** in balance and churn are minor but notable.
* **Geographic and segment-based strategies** are essential due to varying trends.
* Data presentation gaps (e.g., unlabelled graphs) limit some insights.

# Recommendations:

1. Investigate Germany’s high churn with surveys or competitor analysis.
2. Launch re-engagement campaigns for inactive high-balance customers.
3. Segment retention strategies by geography, age, and activity status.
4. Clarify ambiguous metrics (e.g., agent performance) for deeper analysis.