**Problem Statement**

**Credit card churn** is a critical issue for banks and financial institutions. Retaining existing customers is often more cost-effective than acquiring new ones. The goal of this project is to analyse customer behaviour using a real-world bank dataset and uncover actionable insights to reduce customer churn and improve service personalization.

**Objective**

* Analyse a bank’s **credit card customer dataset** to understand key features contributing to churn.
* Perform **Exploratory Data Analysis (EDA)** to uncover trends and patterns.
* Segment customers into distinct groups using **Customer Segmentation** techniques.
* Draw meaningful conclusions to aid strategic business decisions.

**Dataset**

* Source: [Kaggle – Bank Customer Churn Prediction](https://www.kaggle.com/datasets/sakshigoyal7/credit-card-customers)
* Size: ~10,000 records
* Features include: Credit Limit, Total Transactions, Months Inactive, Avg. Open To Buy, etc.

About(Kaggle)-

A manager at the bank is disturbed with more and more customers leaving their credit card services. They would really appreciate if one could predict for them who is going to get churned so they can proactively go to the customer to provide them better services and turn customers' decisions in the opposite direction

I got this dataset from a website with the URL as https://leaps.analyttica.com/home.

Now, this dataset consists of 10,000 customers mentioning their age, salary, marital\_status, credit card limit, credit card category, etc. There are nearly 18 features.

We have only 16.07% of customers who have churned. Thus, it's a bit difficult to train our model to predict churning customers.

**Data Dictionary-**



**Work Done**

**Exploratory Data Analysis (EDA)**

**Data Quality & Structure**

* Missing values (encoded as 'Unknown') were observed in Education\_Level and Income\_Category.
* Gender was uniformly distributed, although the count of females was higher.
* Age, Dependents Count, and Marital Status were normally distributed, with no major skew between churned and existing customers.

**Descriptive Statistics**

Basic statistics were computed to understand feature distributions:

* Features like Credit Limit and Avg\_Open\_To\_Buy were heavily right-skewed, driven by a small subset of customers with extremely high limits.
* Measures like mean, median, standard deviation, and interquartile range (IQR) helped identify variability and outliers.

**Inferential Statistics & Hypothesis Testing**

1. Age vs Credit Limit

* Correlation: 0.002, p-value: 0.803  
  → No significant relationship; age does not influence credit limits.

2. Credit Limit: Attrited vs Existing

* T-test, p-value: 0.016  
  → Statistically significant: Churned customers tend to have lower credit limits.

3. Education Level vs Card Category

* Chi-square test, p-value: 0.480  
  → No significant association; education doesn't affect card choice.

4. Transaction Amount across Income Categories

* ANOVA, p-value: 0.546  
  → No income-based variation in transaction amount.

5. Utilization Ratio vs Churn

* Correlation: -0.178, p-value: 0.000  
  → Customers with lower utilization are more likely to churn.

6. Months Inactive vs Churn

* Mann-Whitney U test, p-value: 0.000  
  → Churned customers are significantly more inactive.

7. Relationship Count vs Transaction Amount

* Correlation: -0.347, p-value: 0.000  
  → More bank products ≠ more usage. In fact, engagement may dilute with product count.

8. Marital Status vs Churn

* Chi-square, p-value: 0.109  
  → No significant difference; marital status is not predictive of churn.

9. Gender vs Credit Limit

* Correlation: 0.421, p-value: 0.000  
  → Moderate correlation; males generally have higher credit limits.

10. Card Category vs Transaction Amount

* Kruskal-Wallis H-test, p-value: 0.000  
  → Premium cardholders transact significantly more.

**Key Behavioral Insights**

* Churned customers:
  + Are less active, with lower transaction count and amounts.
  + Have lower revolving balances and lower card utilization ratios.
  + Tend to own fewer bank products.
* Credit limit plays a critical role:
  + Low credit limit females are more likely to churn.
  + Platinum cardholders aged 51–60 had the highest churn rate (38%).
  + Across all age groups, customers with the highest limits (₹31,200–₹34,516) had 0% attrition.
* Demographic variables like education, marital status, and income showed little to no predictive value.

**Actionable Business Insights**

1. Two distinct customer segments:
   * Majority: Lower-limit, standard cardholders.
   * Minority: High-limit, premium cardholders with disproportionate influence on average financials.
2. Churn risk is strongly tied to:
   * Low credit limits
   * Inactivity (especially >3 months)
   * Low utilization ratio
3. Customer engagement strategies should:
   * Shift from demographics to behavior-based segmentation.
   * Focus on usage patterns, product depth, and transaction frequency.
4. Promotions should encourage upgrades:
   * Card category significantly influences engagement, offer tier-based benefits to retain mid-tier users and upsell.
5. Female platinum cardholders churn more:
   * Possibly due to unmet expectations, banks should investigate experience gaps and personalize high-tier offerings.

**Final Takeaways**

* Activity level and perceived credit empowerment are far stronger churn predictors than demographics.
* Customer segmentation, not generalization, is key to proactive churn reduction.
* Monitoring low engagement customers and introducing early intervention strategies is essential for improving retention.

**Feature Engineering & Preprocessing**

Data Cleaning & Transformation:

* Unknowns Dropped: Records with 'Unknown' in Education\_Level and Income\_Category were dropped instead of imputed, business context was lacking to justify replacements.
* Marital Status Re-mapped: Converted into binary values, "Married" as Yes, all others (example- Divorced, Single) grouped as No.
* Card Categories Grouped: Consolidated Silver, Gold, Platinum into a single Premium category for meaningful clustering.
* Feature Redundancy: Dropped Avg\_Open\_To\_Buy due to high correlation with Credit\_Limit.

Column-wise Transformations:

Used ColumnTransformer to apply different preprocessing steps:

| **Transformation** | **Applied To** |
| --- | --- |
| OneHotEncoder | Categorical vars like Gender, Education\_Level, etc. |
| StandardScaler | Continuous features like Customer\_Age, Months\_on\_book |
| PowerTransformer | Skewed vars like Credit\_Limit, Utilization\_Ratio |
| MinMaxScaler | Transaction metrics like Total\_Trans\_Amt, Trans\_Ct |

**Customer Segmentation**

Dimensionality Reduction:

* Applied PCA to reduce features to 2 principal components.
* Helped visualize and understand customer distribution in lower-dimensional space.

Clustering:

* Used KMeans Clustering with 3 segments (k=3), chosen via elbow and silhouette analysis.
* Fitted the full pipeline using:  
  Preprocessing → PCA → KMeans.

**Cluster Insights**

**Cluster 0: Young Credit Builders**

* Age: Younger customers
* Bank Tenure: Shorter relationship (Months\_on\_book)
* Gender: Balanced distribution
* Credit Limit: Lower
* Utilization Ratio: Higher
* Revolving Balance: Higher
* Income: Mostly in <$40k group

**Cluster 1: Wealthy Low-Risk Users**

* Age: Intermediate
* Bank Tenure: Long-standing customers
* Gender: Predominantly Male
* Credit Limit: Highest
* Utilization Ratio: Lowest
* Revolving Balance: Lowest
* Card Category: Majority Silver+ (premium)
* Income: Highest among clusters

**Cluster 2: Older High-Debt Customers**

* Age: Oldest segment
* Bank Tenure: Longest relationships
* Gender: Predominantly Female
* Credit Limit: Low
* Utilization Ratio: High
* Revolving Balance: High
* Income: Concentrated in <$60k range

**General Observations:**

* Transaction Amount: Similar across all 3 clusters.
* Marital Status: Uniformly distributed; not a major differentiator.
* Behaviour proved more meaningful than demographics in segmentation.

**Conclusion**

The analysis reveals that churn is primarily driven by behavioural patterns, particularly low engagement, high inactivity, and low credit utilization, rather than demographic factors. Customer segmentation further uncovered three distinct personas, enabling targeted retention strategies. These insights equip decision-makers with actionable levers to reduce attrition and optimize customer value.