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
- 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-

Column	Data Type	count	unique values	min_value	max_value	null values	Description
CLIENTNUM	int64	10127	10127	708082083	828343083	0	primary key for identifying the customer
Attrition_Flag	object	10127	2			0	attrited customer-customer has churned , existing-customer is active
Customer_Age	int64	10127	45	26	73	0	age of the customer
Gender	object	10127	2			0	gender of the customer
Dependent_count	int64	10127	6	0	5	0	no.of people relying financially on customer
Education_Level	object	10127	7			0	education is categorised into-'High School', 'Graduate', 'Uneducated', 'Unknown', 'College','Post-Graduate', 'Doctorate' to group the customers based on their education level
Marital_Status	object	10127	4			0	indicates either the customer is married or unmarried
Income_Category	object	10127	6			0	denotes to which range of income does the customer belong
Card_Category	object	10127	4			0	customer cards are categorised into blue,gold,silver,platinum and has various benefits accordingly
Months_on_book	int64	10127	44	13	56	0	the time period for which the customer is with the bank
Total_Relationship_Count	int64	10127	6	1	6	0	credit cards, loans, deposits-no. of products from the bank that the customer has availed
Months_Inactive_12_mon	int64	10127	7	0	6	0	Number of months the customer was inactive in the last 12 months.(didn't avail services from the bank)
Contacts_Count_12_mon	int64	10127	7	0	6	0	Number of times the customer contacted customer service in the past 12 months
Credit_Limit	float64	10127	6205	1438.3	34516	0	maximum amount that a customer can spend on his/her card
Total_Revolving_Bal	int64	10127	1974	0	2517	0	balance not paid in full every mpnth that adds on to the next
Avg_Open_To_Buy	float64	10127	6813	3	34516	0	credit limit-total revolving balance: indicate the amount the customer can spend actually
Total_Amt_Chng_Q4_Q1	float64	10127	1158	0	3.397	0	q4 to q1 comparison: change in the spending behavior when compared between initial quarter and final quarter
Total_Trans_Amt	int64	10127	5033	510	18484	0	the total value of amount transferred through the card in 12 months
Total_Trans_Ct	int64	10127	126	10	139	0	total no. of transactions throughout the year
Total_Ct_Chng_Q4_Q1	float64	10127	830	0	3.714	0	q4 to q1 comparison: change in the spending behavior when compared between initial quarter and final quarter(when seen through frequency)
Avg_Utilization_Ratio	float64	10127	964	0	0.999	0	percentage of available credit that a customer uses depicted as utilization of the card 0-not using the credit and close to 1 implies higher credit usuage each month

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.