

# An Analysis of Credit Utilization and Risk Across Customer Segments

Emily Hubbard

January 27, 2026

## Abstract

This report analyzes credit utilization patterns for 10,127 customers to assess portfolio financial health, identify high-risk utilization behavior using a defined threshold, and determine which customer segments may benefit from targeted financial guidance. Findings emphasize the skewed nature of utilization, risk concentration by demographic segment, and a moderate negative relationship between credit limits and utilization.

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Business Problem and Questions . . . . .	3
1.2	Key Definitions . . . . .	3
<b>2</b>	<b>Dataset</b>	<b>3</b>
2.1	Data Source and Scope . . . . .	3
2.2	Variables Used . . . . .	4
<b>3</b>	<b>Methods</b>	<b>4</b>
3.1	Data Preparation . . . . .	4
3.2	Tools and Techniques . . . . .	4
<b>4</b>	<b>Results</b>	<b>5</b>
4.1	What does the overall distribution of credit utilization look like across customers, and how prevalent is high-risk utilization behavior? . . . . .	5
4.2	Are higher credit limits associated with healthier utilization behavior? . . . . .	6
4.3	How does credit utilization vary across income categories, and what degree of variability exists within each income group? . . . . .	7
4.4	How does the prevalence of high-risk utilization differ across income and age segments? . . . . .	8
4.5	Which customer segments have both elevated credit utilization risk and enough customers that targeted financial guidance could make a meaningful impact? . . . . .	9
4.6	What does overall financial health look like when these patterns are considered collectively? . . . . .	10
<b>5</b>	<b>Executive Summary and Recommendations</b>	<b>10</b>
5.1	Executive Summary . . . . .	10
5.2	Recommendations . . . . .	10

<b>6</b>	<b>Limitations</b>	<b>11</b>
<b>A</b>	<b>Appendix</b>	<b>12</b>

# 1 Introduction

## 1.1 Business Problem and Questions

Credit utilization is a widely used indicator of financial health and credit risk, yet it is often evaluated based on broad assumptions about customer behavior. For lenders and financial advisors, understanding not just how prevalent high utilization is, but where it concentrates within a customer portfolio, is critical for effective risk management and targeted financial guidance. This analysis seeks to move beyond general averages by examining how credit utilization and high-risk behavior vary across customer demographics and credit characteristics, with the goal of identifying segments where proactive intervention may have the greatest impact. This report will answer the following questions:

1. What does the overall distribution of credit utilization look like across customers, and how prevalent is high-risk utilization behavior?
2. Are higher credit limits associated with healthier utilization behavior?
3. How does credit utilization vary across income categories, and what degree of variability exists within each income group?
4. How does the prevalence of high-risk utilization differ across income and age segments?
5. Which customer segments have both elevated credit utilization risk and enough customers that targeted financial guidance could make a meaningful impact?
6. What does overall financial health look like when these patterns are considered collectively?

## 1.2 Key Definitions

- **Credit Utilization:**  $\text{Card Balance} / \text{Credit Limit}$
- **High-risk Threshold:**  $\text{Utilization} \geq 80\%$

# 2 Dataset

## 2.1 Data Source and Scope

This analysis uses an anonymized credit card customer dataset obtained from Kaggle, containing records for 10,127 individuals. The dataset is cross-sectional in nature and includes demographic attributes such as age and income category, as well as credit-related features including credit limit and card category. Behavioral metrics, particularly average credit utilization, provide insight into how customers manage available credit. Together, these features support an in-depth exploration of utilization patterns, risk concentration, and financial health across customer segments.

## 2.2 Variables Used

Table 1: Key Variables Used in the Analysis

Variable Name	Description	Source Type
CLIENTNUM	Unique identifier for each customer holding a credit card account.	Provided
Age_Band	Categorical grouping of customers into age ranges used for demographic segmentation and comparative analysis.	Derived
Income_Category	Annual income category of the account holder (<\$40K, \$40K-\$60K, \$60K-\$80K, \$80K-\$120K, >\$120K, Unknown).	Provided
Credit_Limit	Total credit limit assigned to the customer's credit card.	Provided
Avg_Utilization_Ratio	Average ratio of credit used relative to the total credit limit.	Provided
Risk_Tier	Categorical classification of credit utilization into Low, Moderate, or High risk based on pre-defined utilization thresholds.	Derived
High_Risk_Flag	Binary indicator identifying customers classified as high utilization risk based on the Risk_Tier definition.	Derived
Utilization_Band	Categorical grouping of average utilization ratios into percentage ranges for distribution analysis and visualization.	Derived

## 3 Methods

### 3.1 Data Preparation

Before conducting the analysis, the following preparation steps were performed prior to exploratory and segmented analysis:

- Created a comprehensive data dictionary documenting both original and derived variables; the full dictionary is provided in the Appendix.
- Converted the dataset into an Excel table to enable efficient filtering, sorting, and aggregation.
- Verified and corrected data types to ensure all quantitative fields were stored as numeric values rather than text.
- Engineered derived variables, including age bands, utilization bands, risk tiers, and a binary high-risk flag based on predefined thresholds.

### 3.2 Tools and Techniques

This analysis was conducted in Microsoft Excel using the following tools and techniques:

- Pivot Tables

- Advanced Formulas: COUNTIFS, IF, VLOOKUP, CORREL
- Statistical Analysis ToolPak
- Data Visualization
- Conditional Formatting
- Correlation Analysis

## 4 Results

### 4.1 What does the overall distribution of credit utilization look like across customers, and how prevalent is high-risk utilization behavior?

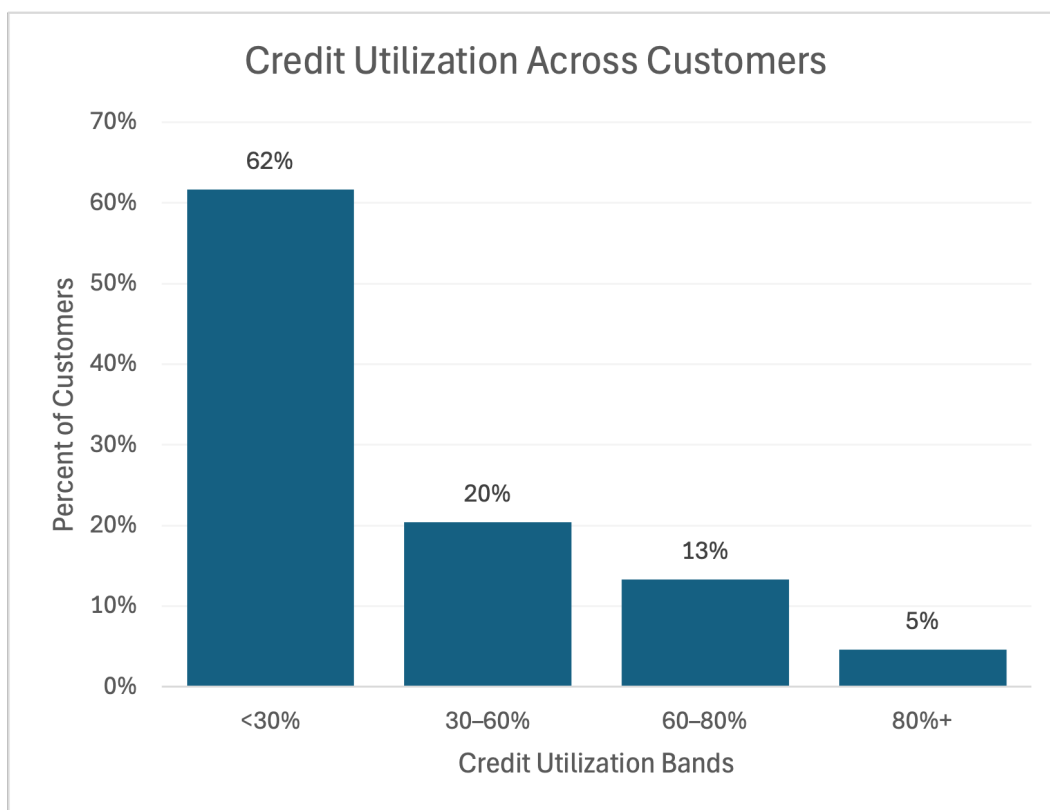


Figure 1: Distribution of credit utilization across customers.

Figure 1 illustrates the distribution of credit card utilization across customers. Credit utilization exceeding 80% is commonly considered high risk in the financial industry. As shown, the majority of customers (approximately 62%) maintain low utilization levels below 30%, while only a small minority (around 5%) meet or exceed the high-risk threshold. This skewed distribution indicates that elevated credit risk is not widespread across the portfolio, but rather concentrated within a relatively small subset of customers, motivating further segmentation analysis in subsequent sections.

## 4.2 Are higher credit limits associated with healthier utilization behavior?

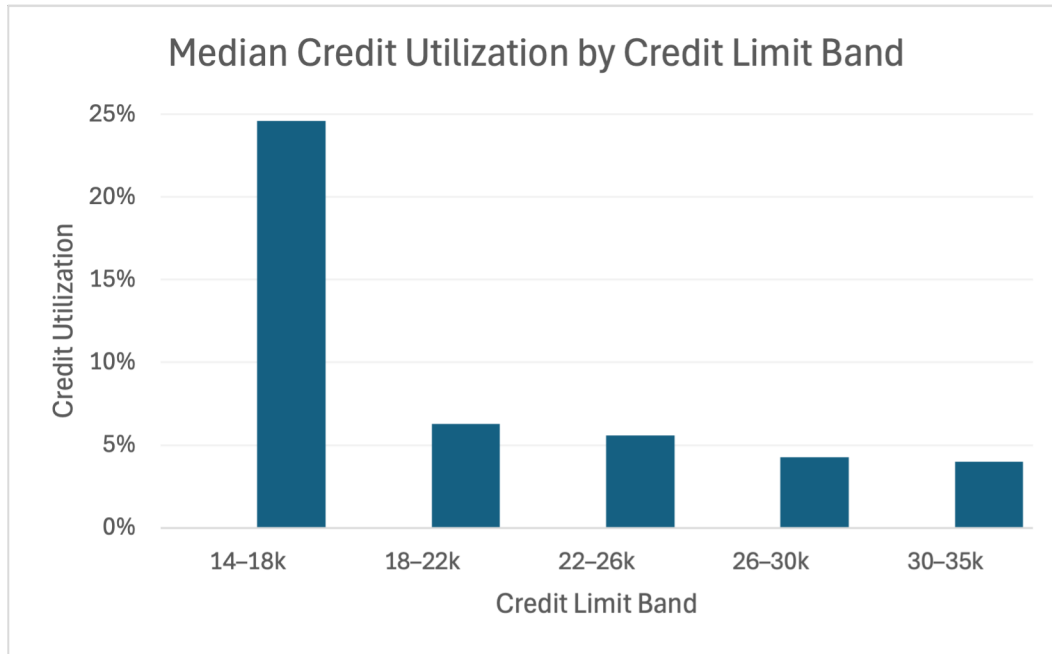


Figure 2: Median utilization by credit limit band.

Figure 2 shows median credit utilization across credit limit bands. A moderate negative correlation is observed between credit limit and credit utilization ( $r = -0.48$ ), indicating that customers with higher credit limits tend to maintain lower utilization ratios. This pattern is consistent with standard lending practices, as financial institutions typically assign higher credit limits to customers with stronger credit profiles and lower perceived risk. Importantly, this relationship reflects correlation rather than causation.

### 4.3 How does credit utilization vary across income categories, and what degree of variability exists within each income group?

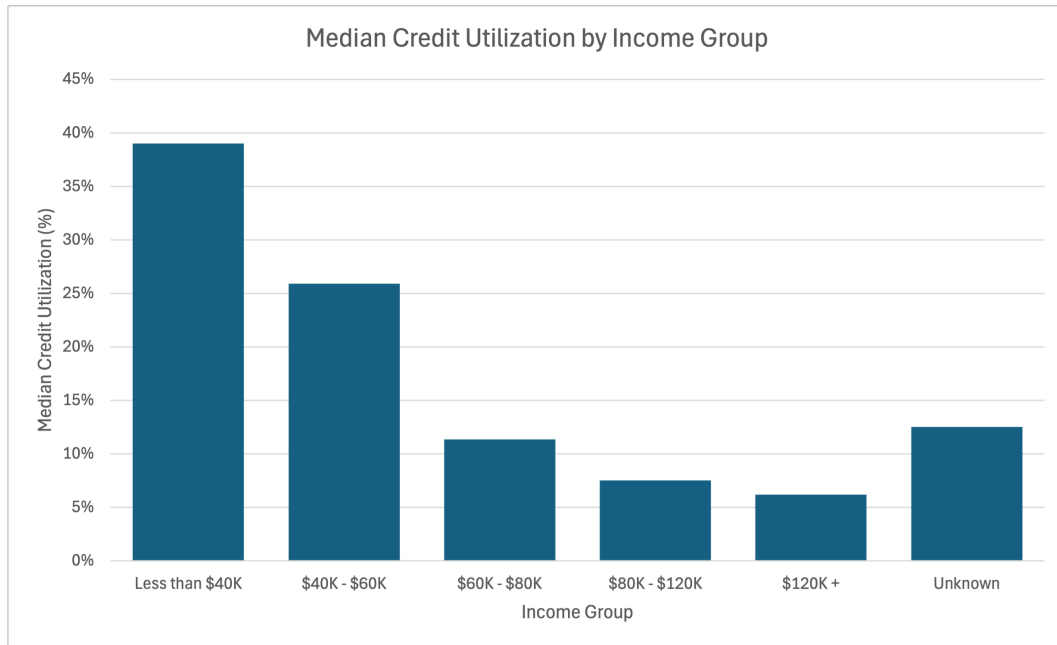


Figure 3: Median credit utilization by income category.

Figure 3 shows median credit utilization by income group. Median utilization declines steadily as income increases, with customers earning less than \$40K exhibiting substantially higher utilization than all other groups. This pattern is consistent with the expectation that higher-income customers typically have greater access to cash and liquid assets, reducing their reliance on revolving credit.

The “Unknown” income category displays relatively elevated utilization; however, this segment cannot be meaningfully interpreted due to missing income information and is treated as a limitation of the dataset rather than a behavioral finding.

#### 4.4 How does the prevalence of high-risk utilization differ across income and age segments?

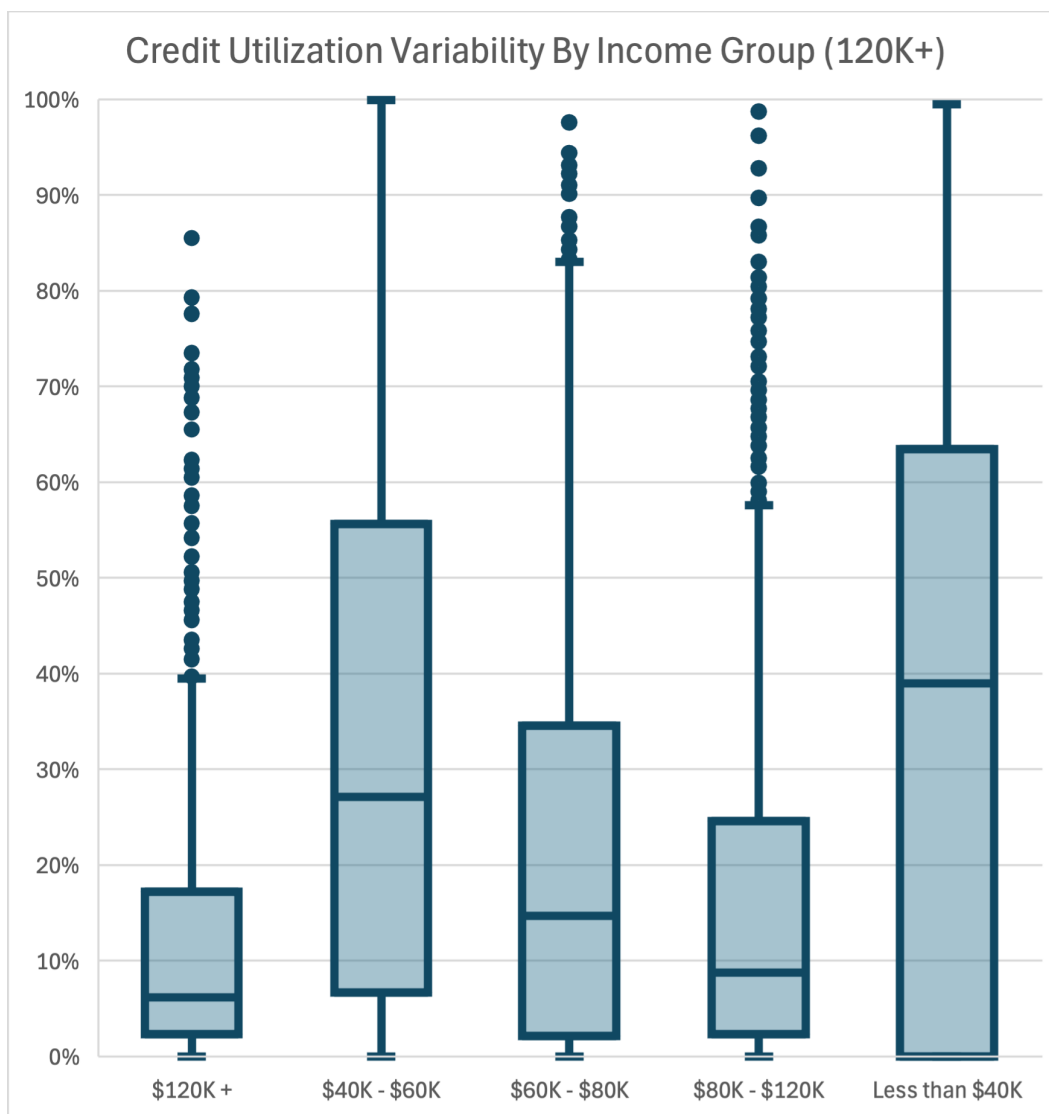


Figure 4: Credit utilization distributions within income categories.

Figure 4 illustrates the distribution and variability of credit utilization within income groups using box-and-whisker plots. While median utilization generally declines as income increases, substantial variability exists within every income category, including among higher-income customers. Notably, group that make greater than \$60K contain a non-trivial amount of high-utilization outliers, indicating that elevated credit risk is not exclusive to lower-income customers. This highlights the importance of examining within-group dispersion rather than relying solely on averages or medians when assessing financial risk.

#### 4.5 Which customer segments have both elevated credit utilization risk and enough customers that targeted financial guidance could make a meaningful impact?

Income Group/Age	30	30-39	40-49	50-59	66 +	Grand Total
Less than \$40K	13.48%	6.80%	7.40%	8.45%	7.96%	7.78%
\$40K - \$60K	2.33%	5.97%	4.74%	5.39%	9.45%	5.42%
\$60K - \$80K	0.00%	5.73%	1.70%	2.26%	3.08%	2.64%
\$80K - \$120K	0.00%	0.38%	1.93%	1.95%	3.57%	1.69%
\$120K +	0.00%	1.01%	0.00%	0.29%	0.00%	0.28%
Unknown	5.26%	2.45%	2.30%	2.89%	2.50%	2.61%
Grand Total	7.69%	4.78%	4.19%	4.64%	6.58%	4.62%

Figure 5: High-risk share by income and age segment.

Figure 5 presents a heatmap of high-risk credit utilization rates segmented by income and age group. Select high-risk segments are shown in the figure below to inform decisions about which groups should receive targeted financial guidance.

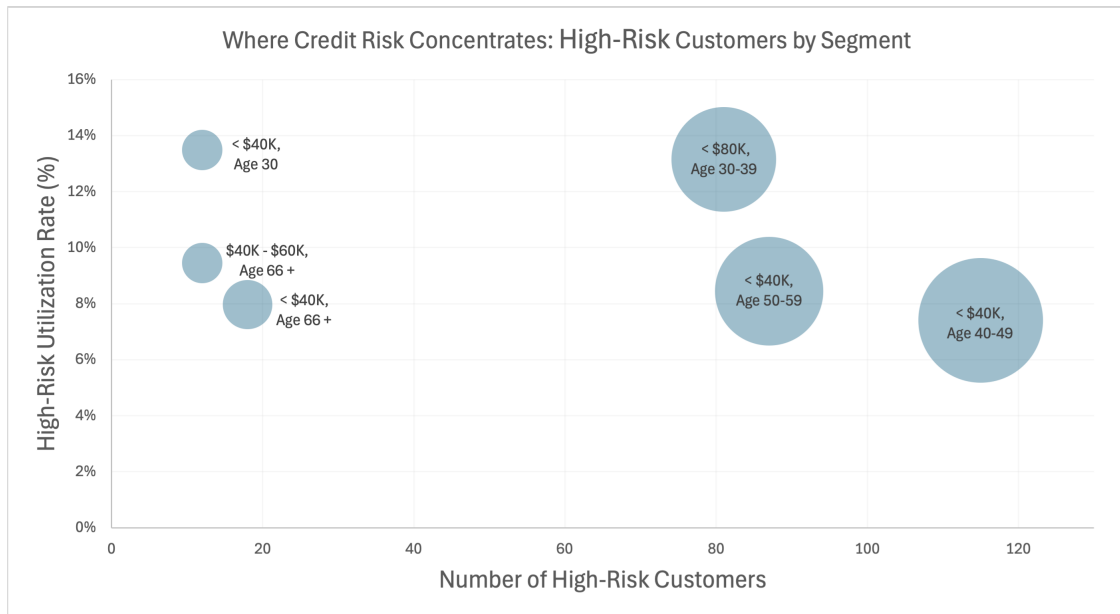


Figure 6: High-risk customer segments by risk rate and scale (bubble size reflects volume).

Figure 6 visualizes the concentration of high-risk customers by segment, with bubble size representing the amount of customers classified in each segment. This format enables simultaneous comparison of both risk prevalence and segment scale.

The analysis indicates that customers with incomes below \$40K consistently exhibit elevated credit utilization risk across multiple age groups, making this income segment a clear priority for proactive financial guidance. In particular, customers aged 30–39 with incomes below \$80K emerge as a high-impact segment, combining relatively high risk rates with a large population of high-risk customers, as reflected by their prominent position and bubble size in the visualization.

Additionally, customers aged 40–49 and 50–59 within the sub-\$40K income group represent substantial concentrations of high-risk customers, suggesting that financial strain persists beyond early career stages for lower-income individuals. While smaller in size, customers aged 66+ in the \$40K–\$60K income range display disproportionately high risk rates, indicating a potential need for more tailored, age-specific credit management support.

Overall, prioritizing lower-income segments, particularly those where elevated risk coincides with meaningful customer volume, offers the greatest potential impact for targeted financial guidance and risk mitigation efforts.

#### 4.6 What does overall financial health look like when these patterns are considered collectively?

To summarize the overall financial health of the portfolio and consolidate the primary findings from the analysis, a set of key performance indicators (KPIs) is presented below. These metrics provide a high-level view of utilization health, risk concentration, and structural relationships within the data.

Table 2: Key performance indicators (portfolio-level).

Category	KPI	Value
Portfolio Size	Total Customers	10,127
Utilization Health	% Low Utilization (< 30%)	61.68%
Utilization Health	% Moderate Utilization (30–80%)	33.70%
Utilization Health	% High Utilization ( $\geq$ 80%)	4.62%
Central Tendency	Median Utilization	11.35%
Concentration	Largest High-Risk Segment	< 40K, Age 40-49
Concentration	High-Risk Share (Top 3 Segments)	28.99%
Correlation	Credit Limit vs Utilization (r)	-0.48

## 5 Executive Summary and Recommendations

### 5.1 Executive Summary

At the portfolio level, financial health appears largely stable, with most customers maintaining low or moderate credit utilization. High-risk utilization is present but limited to a small share of customers and is concentrated within specific demographic segments rather than spread evenly across the portfolio. This pattern suggests that targeted, segment-driven interventions are likely to be more effective than broad risk management strategies. The following recommendations outline the most impactful interventions.

### 5.2 Recommendations

- **Prioritize lower-income customer segments (<\$40K) for proactive financial guidance.** These customers consistently exhibit higher credit utilization risk across multiple age groups and represent a substantial share of the high-risk population, making them the most impactful targets for early intervention.

- **Focus targeted outreach on customers aged 30–39 with incomes below \$80K.** This segment combines elevated utilization risk with a large number of high-risk customers, suggesting that timely guidance during early- to mid-career stages could improve long-term credit outcomes and reduce future risk exposure.
- **Provide tailored support for customers aged 66+ in the \$40K–\$60K income range.** Although smaller in size, this group exhibits disproportionately high utilization rates, indicating potential financial strain related to fixed or transitional incomes that may benefit from simplified credit management or planning resources.
- **Adopt segment-level monitoring rather than portfolio-wide risk strategies.** Because high-risk utilization is concentrated rather than widespread, ongoing monitoring by income and age segment can help identify emerging risk patterns early and enable targeted, preventative interventions.

## 6 Limitations

This analysis is subject to several limitations. The dataset is cross-sectional and represents a single snapshot in time, limiting the ability to assess changes in utilization behavior or risk dynamics over time and restricting the analysis to descriptive patterns rather than causal inference. Risk classifications are based on predefined utilization thresholds, which reflect common industry practices but represent policy-based definitions rather than predictive risk models. Additionally, certain demographic categories—most notably the “Unknown” income group—cannot be meaningfully interpreted and therefore limit segment-level conclusions.

# A Appendix

Original Variables	
CLIENTNUM	Client number. Unique identifier for the customer holding the account
Attrition_Flag	Internal event (customer activity) variable - if the account is closed then 1 else 0
Customer_Age	Demographic variable - Customer's Age in Years
Gender	Demographic variable - M=Male, F=Female
Dependent_count	Demographic variable - Number of dependents
Education_Level	Demographic variable - Educational Qualification of the account holder (example: high school, college graduate, etc.)
Marital_Status	Demographic variable - Married, Single, Divorced, Unknown
Income_Category	Demographic variable - Annual Income Category of the account holder (< \$40K, \$40K - 60K, \$60K - \$80K, \$80K-\$120K, > \$120K, Unknown)
Card_Category	Product Variable - Type of Card (Blue, Silver, Gold, Platinum)
Months_on_book	Period of relationship with bank
Total_Relationship_Count	Total no. of products held by the customer
Months_Inactive_12_mon	No. of months inactive in the last 12 months
Contacts_Count_12_mon	No. of Contacts in the last 12 months
Credit_Limit	Credit Limit on the Credit Card
Total_Revolving_Bal	Total Revolving Balance on the Credit Card
Avg_Open_To_Buy	Open to Buy Credit Line (Average of last 12 months)
Total_Amt_Chng_Q4_Q1	Change in Transaction Amount (Q4 over Q1)
Total_Trans_Amt	Total Transaction Amount (Last 12 months)
Total_Trans_Ct	Total Transaction Count (Last 12 months)
Total_Ct_Chng_Q4_Q1	Change in Transaction Count (Q4 over Q1)
Avg_Utilization_Ratio	Average Card Utilization Ratio

Figure A.1: Credit utilization distributions within income categories.

Derived Variables	
Income_Code	Ordinal numeric code representing income category order without implying continuous or equal income differences.
Age_Band	Categorical grouping of customers into age ranges for demographic segmentation and comparative analysis.
Limit_Band	Categorical grouping of customers by total credit limit to analyze behavior across credit capacity levels.
Risk_Tier	Categorical classification of credit utilization into Low, Moderate, or High based on predefined thresholds.
High_Risk_Flag	Binary indicator identifying customers classified as High utilization risk based on Risk_Tier.
Utilization_Band	Categorical grouping of average utilization ratios into percentage ranges for distribution analysis and visualization.

Figure A.2: Output of descriptive statistics for original dataset.

Column1	
Mean	0.2748936
Standard Error	0.0027396
Median	0.176
Mode	0
Standard Deviation	0.2756915
Sample Variance	0.0760058
Kurtosis	-0.794972
Skewness	0.718008
Range	0.999
Minimum	0
Maximum	0.999
Sum	2783.847
Count	10127
Confidence Level(95.0%)	0.0053701

Figure A.3: Definitions for variables derived from the original dataset.

Income_Category	Income_Code
\$120K +	4
\$80K - \$120K	3
\$60K - \$80K	2
\$40K - \$60K	1
Less than \$40K	0
Unknown	-1

Figure A.4: Reference table for derived Income\_Code variable.