CUSTOMER SEGMENTATION ANALYSIS REPORT

Introduction

This analysis seeks to uncover distinct customer segments within a comprehensive loan dataset, comprising key features such as financial loan applications, borrower characteristics, and loan performance metrics. The dataset, sourced from Kaggle, offers a unique opportunity to explore the complex relationships between loan attributes, repayment capacity, and financial profiles. By applying segmentation techniques, I aim to identify and characterize customer groups with similar financial behaviors and loan characteristics, enabling targeted loan product development and marketing strategies. This research addresses critical questions surrounding customer segment risk profiles, default probabilities, and optimal loan term adjustments. Ultimately, this analysis will provide actionable insights for lenders to refine their customer engagement and risk management strategies, enhancing overall loan portfolio performance.

Aim of Customer Segmentation Analysis

The aim of this customer segmentation analysis is to identify distinct customer groups based on their loan characteristics, repayment capacity, and financial profile, to better tailor loan products and marketing strategies.

Research Questions

1. What are the key characteristics of customer segments based on their financial behavior and loan attributes?

2. Which customer segment has the highest risk of default or inability to repay loans?

3. How can loan terms or interest rates be adjusted to attract or retain specific customer segments?

Methodology

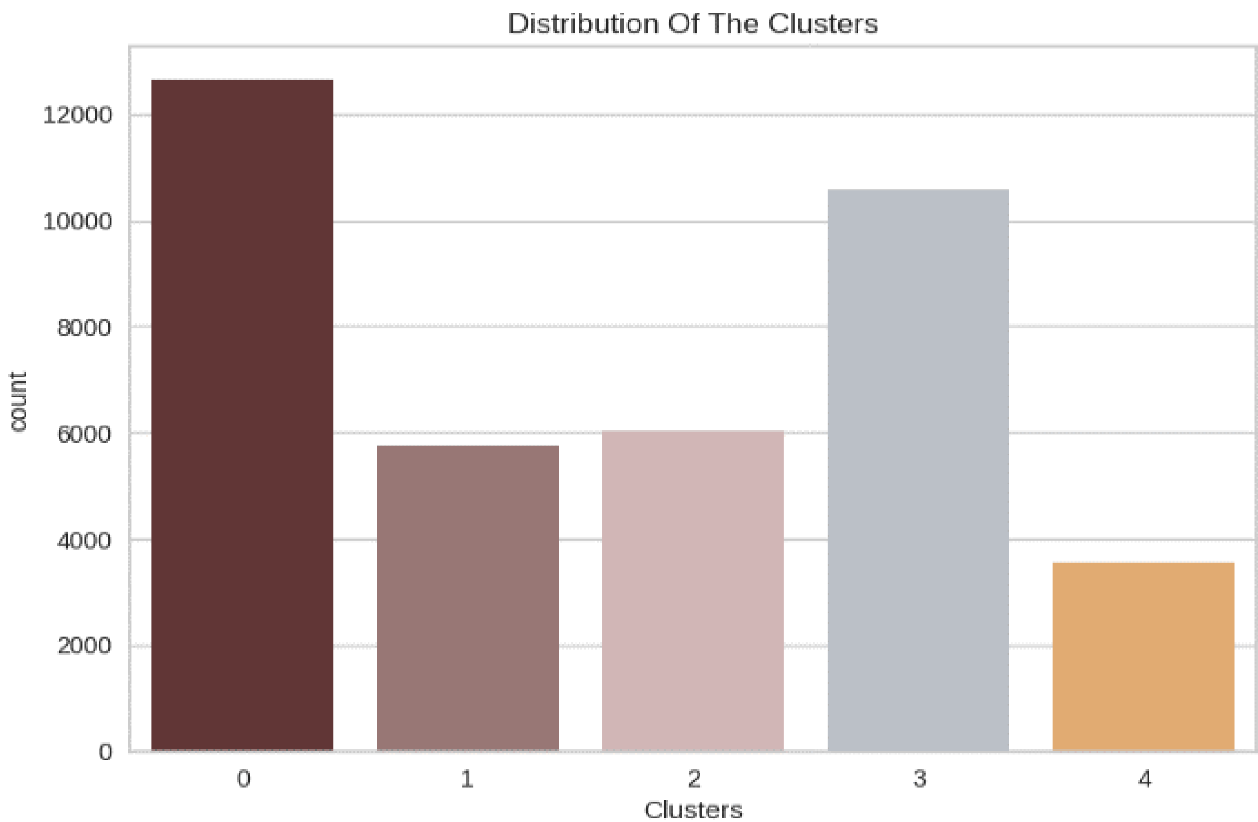
1. Three different datasets were downloaded from Kaggle. The datasets comprise loan facts table, loan dimension table and borrower dimension table.
2. These datasets were loaded in google collab notebook, an integrative development of python, where the entire project was executed.
3. Each dataset was thoroughly processed by converting columns to their appropriate data type, removing columns with missing values and dropping columns that were not relevant for the analysis.
4. After pre-processing, the three datasets were joined together to form a one complete table.
5. The merged data was further preprocessed for the clustering process. The preprocessing involved:
   a. Encoding categorical columns for easy analysis
   b. Scaling the data to reduce the effect of outliers on clustering outcome.

c. Reducing the dimension of the data using PCA dimensionality reduction method to reduce the complexity of the data.

6. An optimal cluster number of five was determined using the Elbow Method.
7. The Agglomerative clustering method was then fit to the preprocessed data to predict five (5) customer segment/clusters.
8. An exploratory analysis was performed on these clusters to gain insights into their distribution.
9. Visualizations were then created to answer each research question.

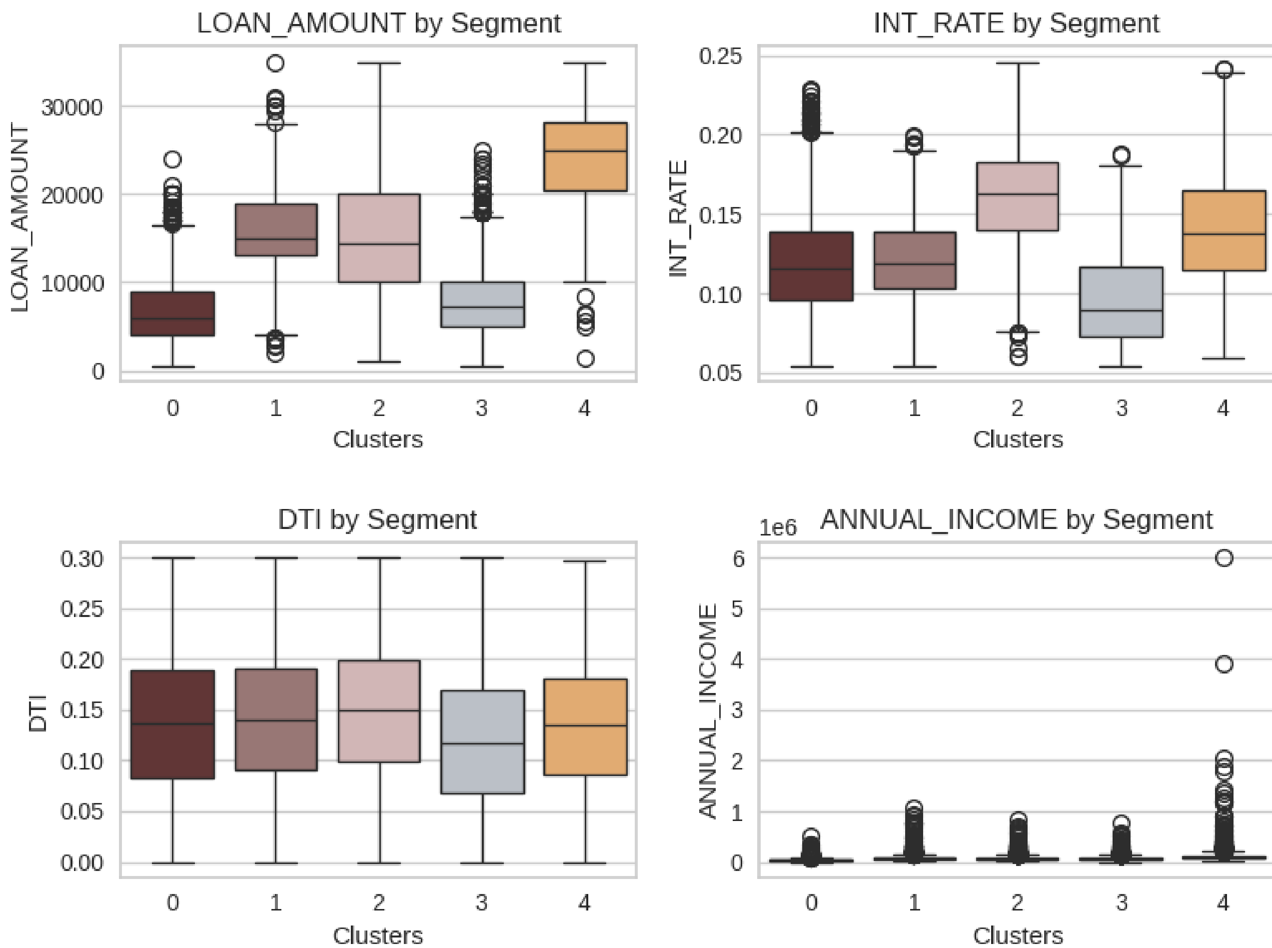
Results and Discussions

Distribution of Clusters



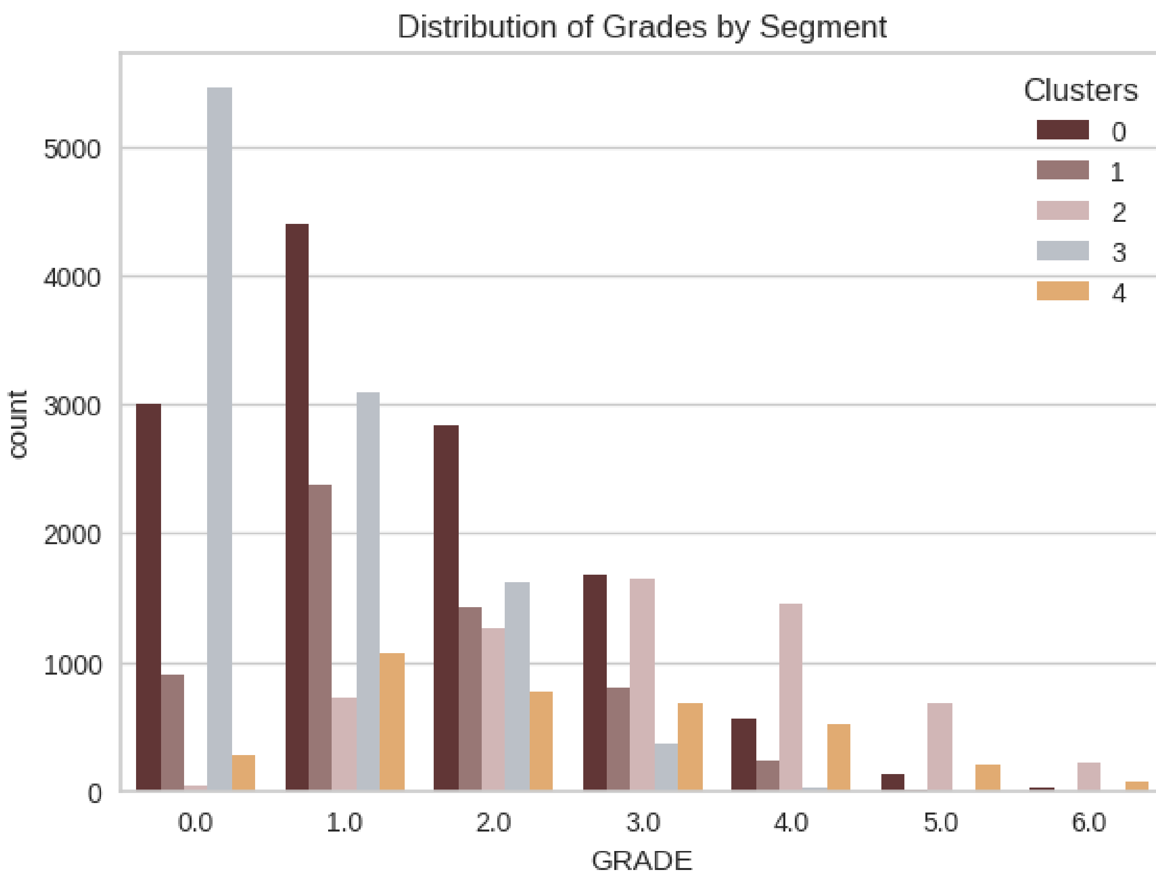
The clusters are not evenly distributed. There are significant differences in the number of observations in each cluster. Cluster 0 has the highest number of observations, while Cluster 4 has the lowest.

Research Question 1: What are the key characteristics of customer segments based on their financial behavior and loan attributes?



1. **Loan Amount**:
   - Segment 4 stands out with the highest loan amounts, suggesting that customers in this group are approved for or apply for larger loans.
   - Segment 3 shows a lower median loan amount compared to others, indicating more modest borrowing behavior.
2. **Interest Rate (INT_RATE)**:
   - Segments 4 and 3 have higher median interest rates, likely indicating higher-risk borrowers or loans with less favorable terms.
   - Segments 0 and 2 have lower median interest rates, suggesting better creditworthiness or favorable loan conditions.
3. **Debt-to-Income Ratio (DTI)**:
   - DTI levels are relatively consistent across clusters, but Segment 4 shows slightly higher debt-to-income ratios, which might indicate customers with heavier financial obligations.

4. **Annual Income**:
   - Segment 4 shows the highest variation in annual income, suggesting it includes both high-income and low-income borrowers, while other segments demonstrate more consistent income levels.

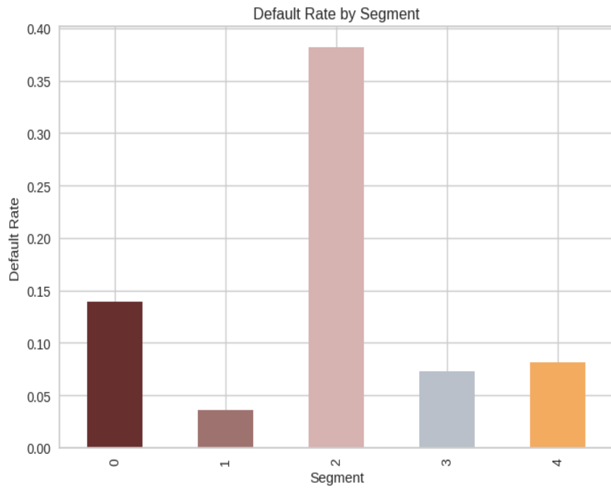


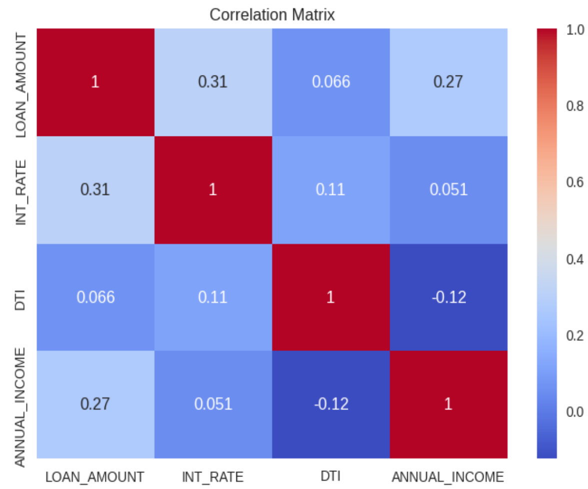
Grade Distribution:

- Segment 3 has the highest concentration of lower loan grades (0.0), indicating a preference or eligibility for higher-risk loans.
- Segments 0 and 1 also include customers with diverse grade distributions, whereas Segment 4 has relatively fewer borrowers across all grades.

These insights highlight how different segments vary in terms of financial risk, loan terms, and income levels, offering guidance for tailored financial products or risk management strategies.

<u>Research Question 2. Which customer segment has the highest risk of default or inability to repay loans?</u>





Segment 2 Has the Highest Default Rate: The bar chart clearly shows that Segment 2 has the highest default rate (close to 40%), making it the riskiest group in terms of repayment ability. The other clusters have lower default rate.

Debt-to-Income (DTI) Ratio Correlation: The heatmap reveals that DTI has a weak correlation with loan-related variables, such as loan amount (0.066) and interest rate (0.11), indicating that DTI alone may not be a strong predictor of default.

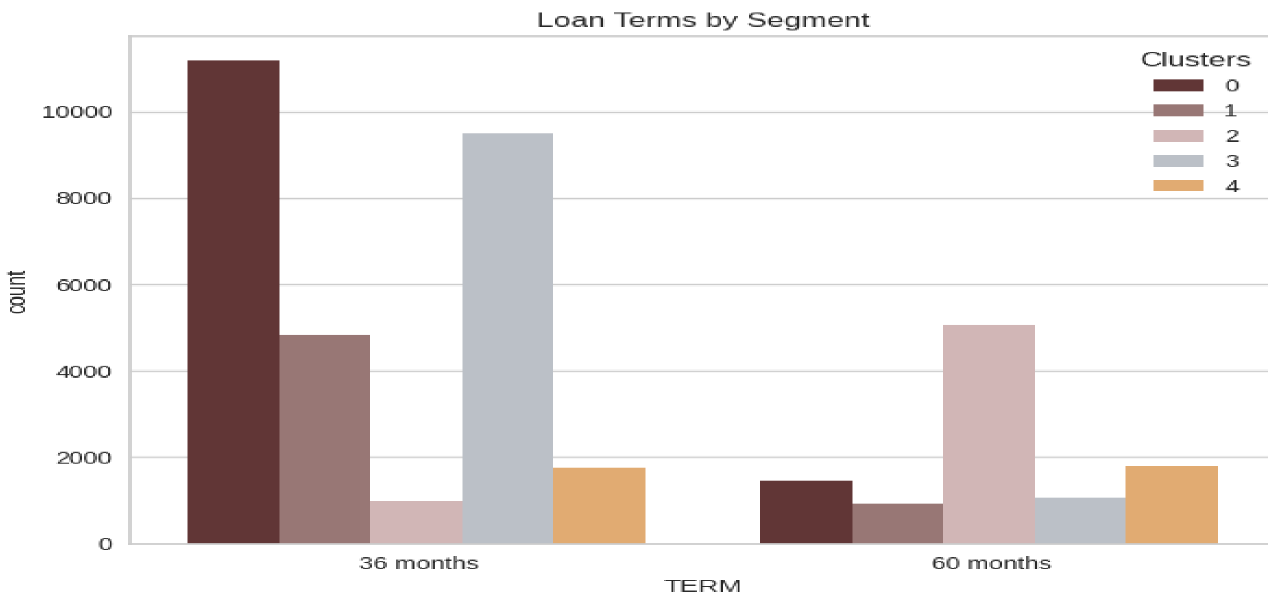
Interest Rate (INT_RATE) as a Risk Factor: The heatmap shows a moderate correlation (0.31) between interest rate and loan amount. Higher interest rates might be associated with higher default risk, particularly in Segment 2.

Loan Amount and Default Risk: Loan amount has a positive correlation with both interest rate (0.31) and annual income (0.27). Customers in Segment 2 might be those with higher loan amounts and insufficient financial capacity to repay.

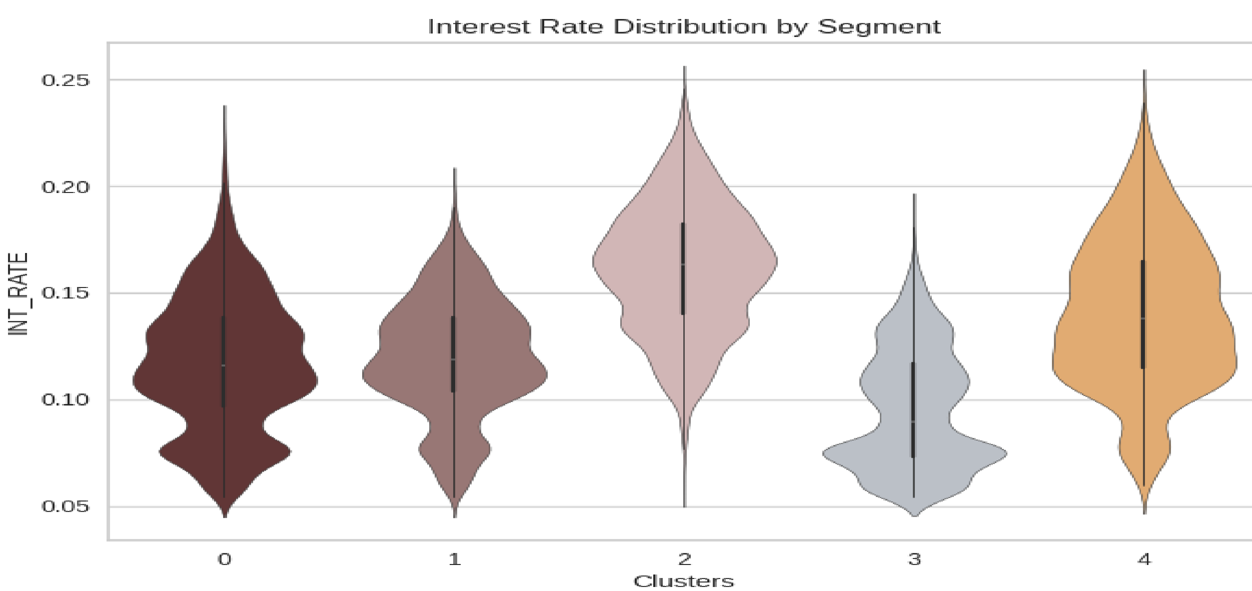
Weak Relationship with Annual Income: Annual income has a negative correlation with DTI (-0.12) and only a slight correlation with loan amount (0.27), suggesting that income alone is not a decisive factor in defaulting, especially for Segment 2.

These insights highlight Segment 2 as the highest-risk group, driven by high default rates and possibly influenced by higher interest rates and loan amounts.

Research Question 3. How can loan terms or interest rates be adjusted to attract or retain specific customer segments



1. **Short-Term Preference**: Segments 0 and 3 prefer 36-month loans; offer lower rates for short terms.
2. **Long-Term Focus**: Segment 2 prefers 60-month loans; target them with flexible payment options.
3. **High Volume Segments**: Segments 0 and 3 dominate; prioritize them with special offers.
4. **Balanced Preference**: Segment 4 chooses both terms equally; use competitive rates for both.
5. **Low Engagement**: Segment 1 has low volume; attract them with discounts and perks.



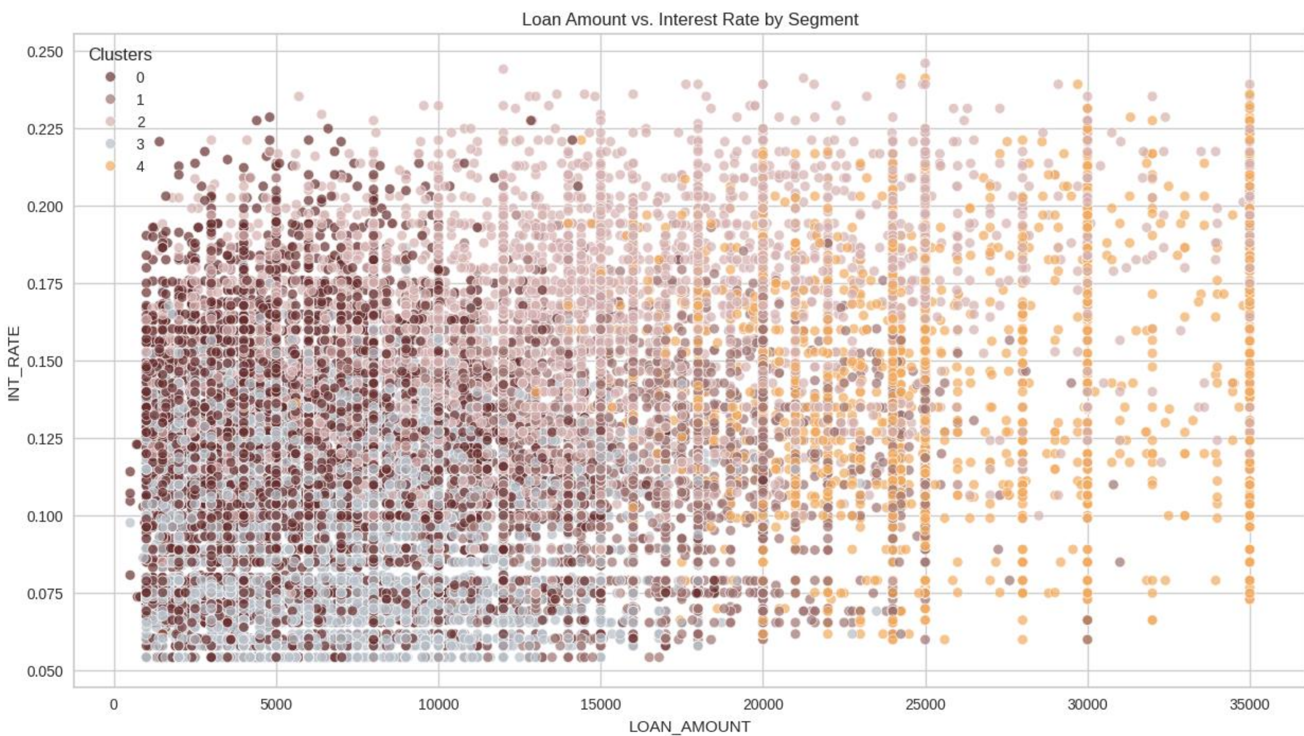
1. Segmented Interest Rate Preferences: Violin plots show distinct interest rate distributions across segments, indicating varying sensitivities.

2. Targeted Strategies: Tailor interest rates to attract/retain specific segments (e.g., lower rates for Segment 0, slightly higher for Segment 4).

3. Cross-Selling Opportunities: Identify segments (like Segment 2) that may be interested in additional financial products.

4. Risk Management: Analyze interest rate preferences to assess credit risk, applying stricter checks for segments preferring lower rates.

5. Competitive Advantage: Understand segment preferences to position loan offerings competitively and adjust interest rate strategies.



1. **Cluster 0**: Risk-Averse Borrowers Profile: These customers prioritize low-interest rates, regardless of loan amount. Strategy: Offer fixed-rate loans and explore cross-selling opportunities.
2. **Cluster 1:** Diverse Borrowers Profile: This diverse group has varying needs in terms of interest rates and loan amounts. Strategy: Provide a range of loan products with flexible terms and personalized marketing.
3. **Cluster 2:** High-Value Borrowers Profile: These customers are willing to pay higher interest rates for larger loans. Strategy: Offer high-limit loans with competitive rates and stringent credit checks.
4. **Cluster 3:** Small-Loan Seekers Profile: These customers need smaller loans with low interest rates. Strategy: Provide quick and easy online loan applications with competitive rates and low limits.
5. **Cluster 4:** Short-Term Borrowers Profile: These customers seek smaller loans with higher interest rates and shorter terms. Strategy: Offer short-term loans with flexible repayment options.

Profiling of Clusters/Segments

Cluster 0

- Tends to have lower loan amounts compared to other clusters.
- Has the lowest median interest rate compared to other clusters
- Generally lower DTI, indicating lower levels of debt.
- Wide range of annual income, with some outliers at the higher end.
- Diverse Creditworthiness
- Moderate default rate
- Prefers lower interest rates
- Prefers lower interest rates, with a wide range of loan amounts.

Cluster 1

- Moderate loan amounts, slightly higher than Cluster 0.
- Slightly higher interest rate compared to Cluster 0.
- Moderate DTI, indicating a balance between debt and income.
- Moderate to high annual income range.
- Moderately Creditworthy
- Very low default rate
- Moderate interest rate preference
- Moderate interest rate preference, with a wide range of loan amounts.

Cluster 2

- Moderate to high loan amounts.
- Moderate interest rate, higher than Clusters 0 and 1.
- Moderate to high DTI, suggesting higher debt levels.
- Moderate to high annual income range.
- Lower to Moderate Creditworthiness
- Highest default rate
- Moderate interest rate preference, slightly higher than Cluster 1
- Moderate interest rate preference, with a wide range of loan amounts.

Cluster 3

- Moderate loan amounts, similar to Cluster 1.
- Moderate interest rate, similar to Cluster 2.
- Moderate DTI, similar to Cluster 1.
- Moderate to high annual income range.
- Moderately Creditworthy
- Very low default rate
- Prefers lower interest rates
- Prefers lower interest rates, with a moderate range of loan amounts.

Cluster 4

- Highest loan amounts among all clusters.
- Highest median interest rate, indicating a willingness to pay higher rates.
- Moderate to high DTI, suggesting higher debt levels.
- Wide range of annual income, with outliers at both the lower and higher ends.
- Lower Creditworthiness
- Very low default rate
- Prefers higher interest rates
- Prefers higher interest rates, with a wide range of loan amounts.