

## **Introduction**

The purpose of this project is to analyse loan application data and identify patterns that impact loan approval. The study aims to assess credit risk factors and recommend strategies to optimize loan approvals while minimizing defaults.

## **Data Analysis & Key Insights in Application Data Set**

### **1. Income and Loan Amount Relationship**

- Clients with academic degrees tend to have lower income compared to others.
- Defaulters generally have lower income than non-defaulters.

### **2. Loan Amount vs. Education & Marital Status**

- Married clients with academic degrees apply for higher loan amounts and have fewer outliers.
- Single clients with academic degrees show a slim boxplot with no outliers.
- Clients with higher education levels are more likely to take large credit loans.

### **3. Correlation Among Key Variables**

- **AMT\_CREDIT** and **AMT\_GOODS\_PRICE** are highly correlated, meaning as home prices increase, the loan amount increases.
- **AMT\_CREDIT** and **AMT\_ANNUIITY (EMI)** shows a high correlation, indicating that a higher loan amount leads to a higher EMI.
- **AMT\_CREDIT, AMT\_GOODS\_PRICE, and AMT\_ANNUIITY** is all interrelated, which might limit heatmap insights.

### **4. Loan Amount vs. Age and Family Size**

- **AMT\_CREDIT** is inversely proportional to **DAYS\_BIRTH**, meaning younger people take higher credit amounts.
- **AMT\_CREDIT** is inversely proportional to **CNT\_CHILDREN**, meaning clients with fewer children take higher loan amounts.
- **AMT\_INCOME\_TOTAL** is also inversely proportional to **CNT\_CHILDREN**, meaning clients with fewer children have higher income.
- Clients with fewer children tend to reside in densely populated areas where **both AMT\_CREDIT and AMT\_INCOME\_TOTAL** are higher.

### **5. Address Mismatch and Loan Default Risk**

- Clients whose **permanent address does not match their contact address** tend to have fewer children.
- Clients whose **permanent address does not match their work address** also tend to have fewer children.

## **Conclusions**

- Clients with higher education levels often take larger loans, but their risk varies based on marital status.
- Younger clients tend to take higher loan amounts, which could pose a risk if not assessed carefully.
- Clients with fewer children generally have higher income and take larger loans, but their address mismatch may indicate mobility, which could affect stability.

- A strong correlation exists between home price, loan amount, and EMI, meaning these factors should be assessed together.

## Recommendations

- **Loan Approval Strategy:** Clients with stable addresses and higher education should be prioritized for larger loans.
- **Risk-Based Pricing:** Loans should be priced based on education level, age, number of children, and address stability.
- **Targeted Loan Products:** Develop specialized loan products for younger applicants with structured EMI plans to reduce default risk.
- **Address Verification:** A stricter verification process should be in place for clients with address mismatches.

## Final Thoughts

This analysis provides a data-driven approach to improving loan approvals while minimizing risks. Implementing these recommendations can help financial institutions make informed lending decisions, optimizing profitability and customer satisfaction.

## Data Analysis & Key Insights in **Previous Application Dataset.**

### Univariate Analysis:

- The distribution of **AMT\_APPLICATION** is highly right-skewed, indicating that while most applicants request moderate amounts, a significant tail of high-value applications exists.
- A count plot of **PRODUCT\_COMBINATION** reveals which loan product types are most common, distinguishing standard offerings from niche products.

### Bivariate Analysis:

- The contract status distribution shows that Approved applications dominate the dataset, with fewer Refused, Cancelled, or Unused offers.
- Box plots of application amounts by contract status indicate that certain statuses (e.g., Refused) often involve higher requested amounts, suggesting distinct risk profiles.

### Outlier Quantification:

- Using the IQR method on **AMT\_APPLICATION**, the analysis identified that while no lower outliers exist, a substantial number of records exceed the upper bound, highlighting extreme loan requests that may influence credit risk.
- A Spearman correlation heatmap reveals strong positive interrelationships among key financial variables (e.g., **AMT\_APPLICATION**, **AMT\_CREDIT**, and **AMT\_GOODS\_PRICE**).

### Correlation Analysis:

- A Spearman correlation heatmap reveals strong positive interrelationships among key financial variables (e.g., **AMT\_APPLICATION**, **AMT\_CREDIT**, and **AMT\_GOODS\_PRICE**).
- These correlations suggest that larger loan requests and associated costs tend to move together, underscoring the need for integrated risk evaluation.

## Conclusions:

- The previous application data shows a predominantly standard range of loan applications with a notable tail of high amounts, indicating potential high-risk scenarios.
- Dominant product combinations and contract statuses provide insight into typical borrower behaviour, while outliers and strong interdependencies among financial metrics signal areas requiring closer scrutiny.

## Recommendations:

- **Risk-Based Underwriting:** Incorporate the observed extreme values and correlated financial metrics into risk assessment models to tailor loan approvals and pricing strategies.
- **Targeted Lending Products:** Develop specialized products for segments identified by unique product combinations or high-risk application amounts, ensuring enhanced monitoring and validation.

## Final Thoughts

This analysis of previous application data offers a comprehensive view of historical loan behaviours, forming a robust foundation for refining credit risk models and strategic lending decisions. Implementing these insights can help financial institutions improve profitability and customer satisfaction through more informed, data-driven lending practices.