LOAN RISK ANALYZER

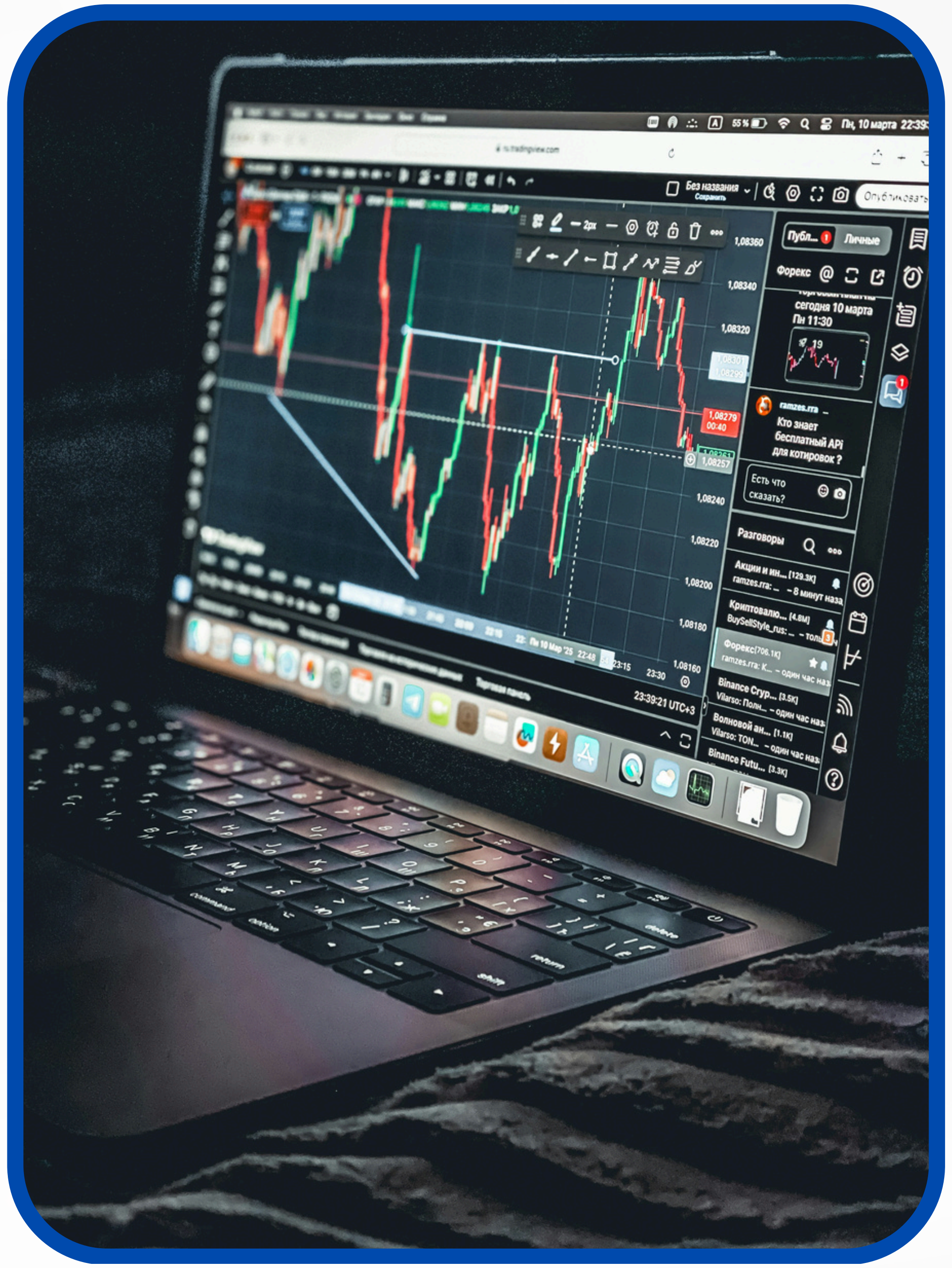
A DEEP DIVE INTO FINANCIAL RISK, FRAUD DETECTION, AND CUSTOMER BEHAVIOR

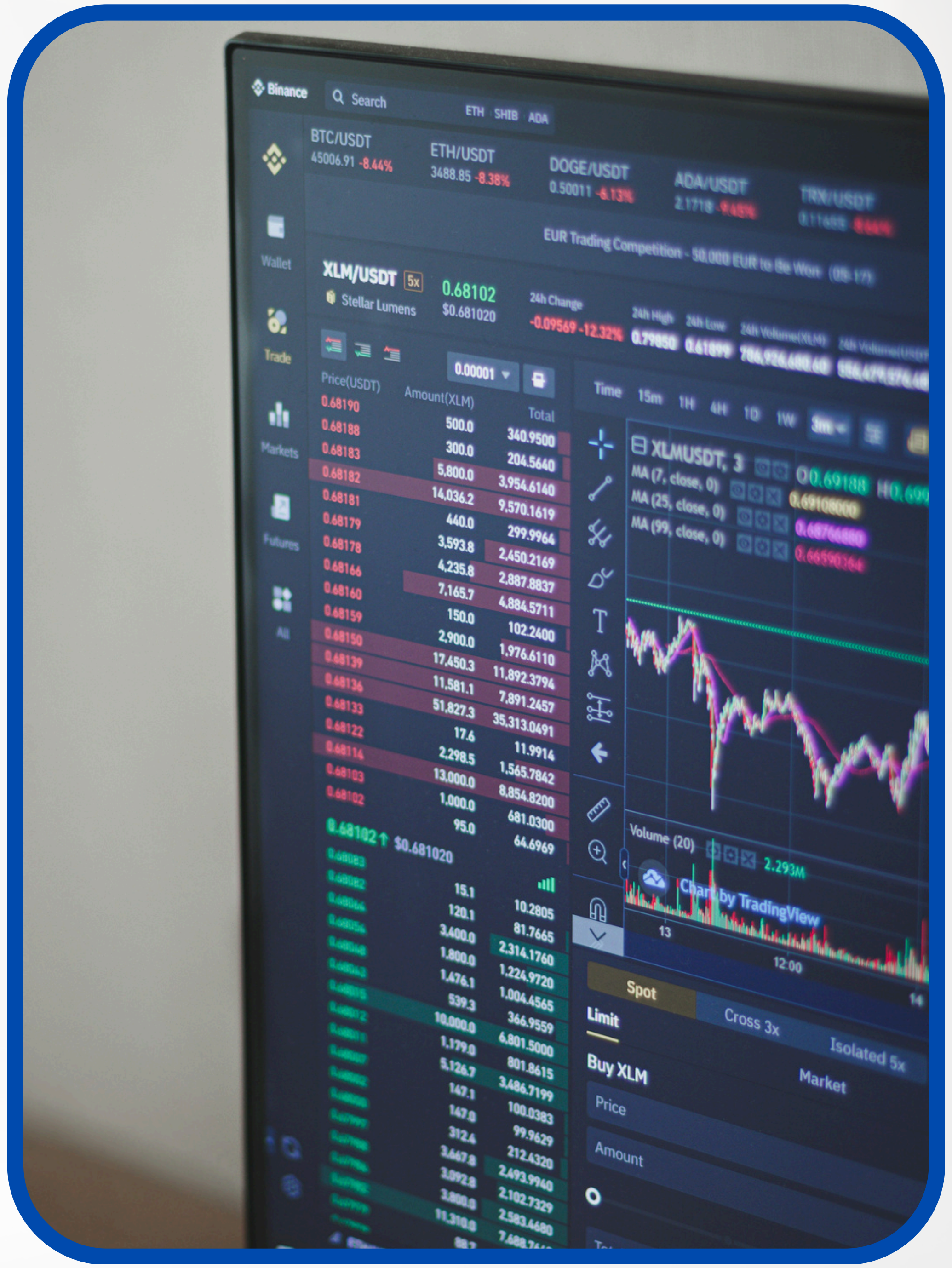
Using Data-Driven Insights to Enhance Loan Risk Management, Prevent Fraud, and Understand Borrower Behavior

PROJECT
OVERVIEW

Purpose:
Leverage Exploratory Data Analysis (EDA) to uncover actionable insights that improve loan decision-making processes for banks and financial institutions.

Scope Includes:
- Risk Assessment
- Fraud Detection
- Customer Behavior Analysis





PROJECT **OBJECTIVES**

1 Assess Financial Risk

Evaluate applicant creditworthiness by analyzing financial indicators like CreditScore, RiskScore, DebtToIncomeRatio, and PaymentHistory. This helps identify high-risk profiles and supports more accurate loan approval decisions, reducing default exposure.

2 Detect Fraud

Identify inconsistencies between reported income, credit behavior, and financial history to flag suspicious applications. Detect patterns such as high income with poor credit or unrealistic asset declarations, helping prevent financial fraud.

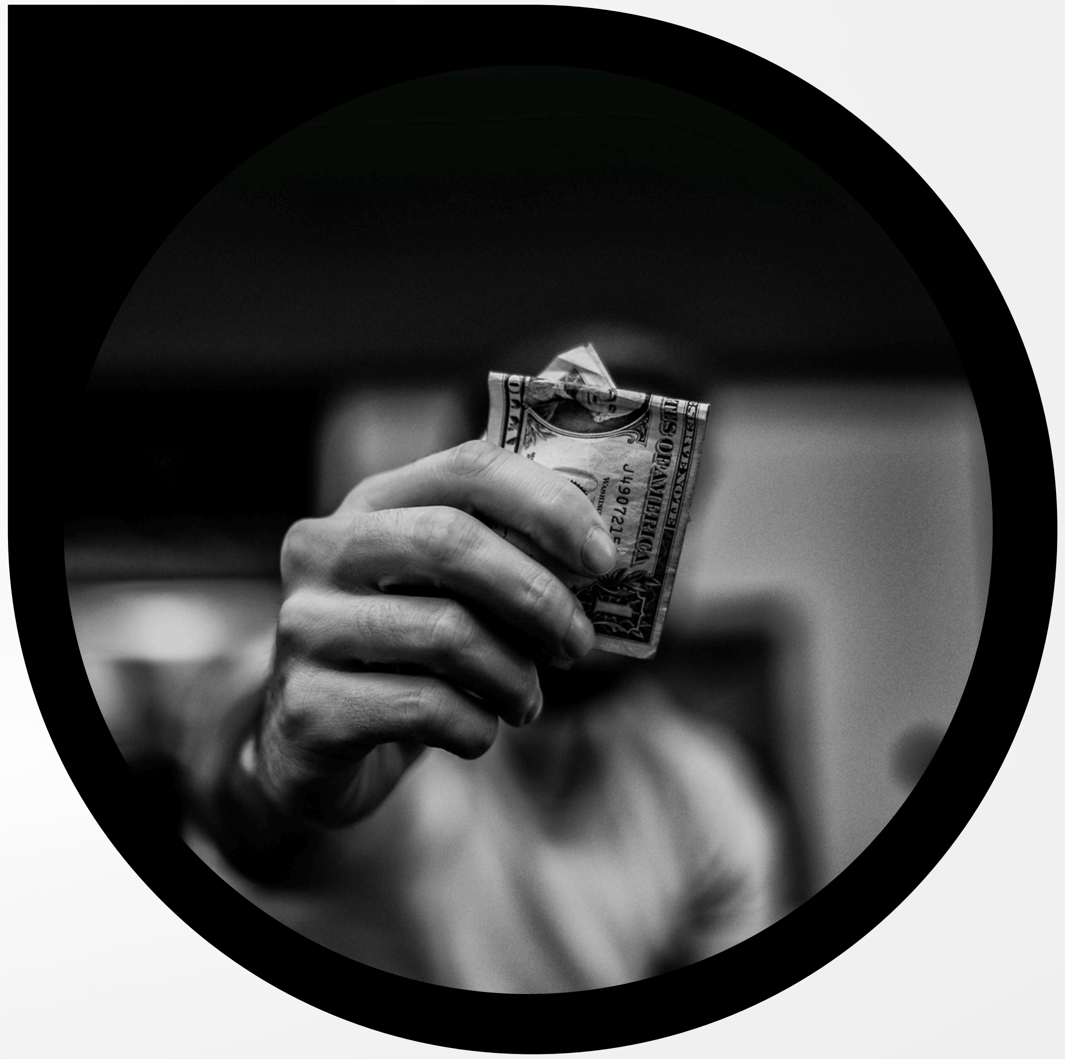
3 Understand Customer Behavior

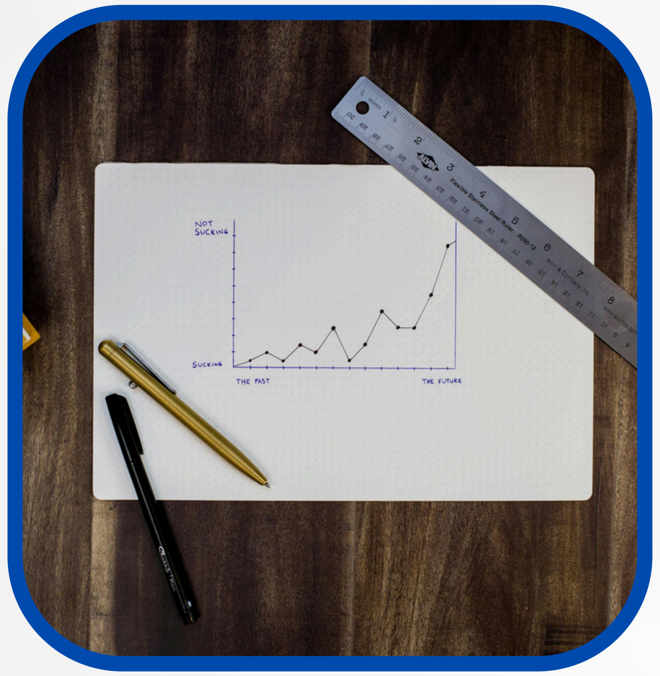
Segment applicants based on demographics, income, employment, and financial habits. Recognize patterns that define reliable borrowers vs. higher-risk applicants, enabling tailored loan products and risk-adjusted interest strategies.

PROJECT GOAL &
BUSINESS DOMAIN

Goal:
To minimize financial risk, improve approval accuracy, and personalize banking products using data-driven decision-making.

Business Domain:
Retail lending and loan underwriting within banking and financial services.





DATASET **SUMMARY**

- **Source:** Kaggle – Financial Risk for Loan Approval
- **Rows:** 20,000
- **Columns:** 36
- No missing values
- Covers personal, financial, and behavioral features of applicants

TOOLS & TECHNOLOGIES

EXCEL

Used for initial data inspection, basic cleaning, and formatting before deeper analysis.

JUPYTER NOTEBOOK

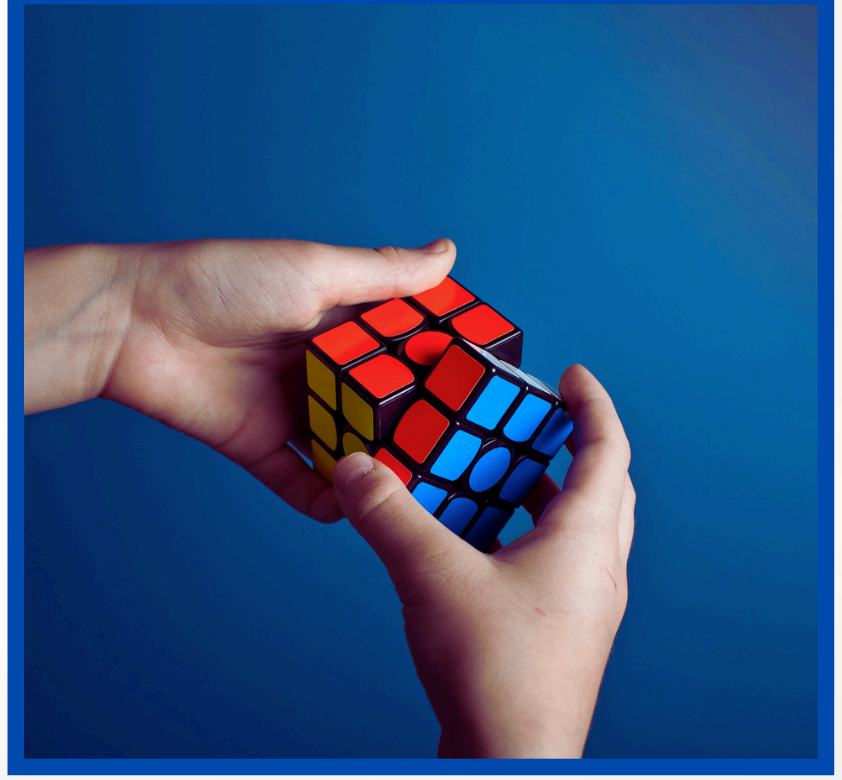
Provided an interactive environment to write, test, and visualize the data exploration process step-by-step.

PYTHON & LIBRARIES

- **pandas** – For data wrangling, filtering, aggregation, and manipulation of tabular data.
- **numpy** – For numerical operations and efficient handling of arrays and mathematical functions.
- **matplotlib & seaborn** – To create clear and insightful visualizations such as histograms, boxplots, scatter plots, and heatmaps for data exploration.

BUSINESS PROBLEMS
SOLVED THROUGH EDA

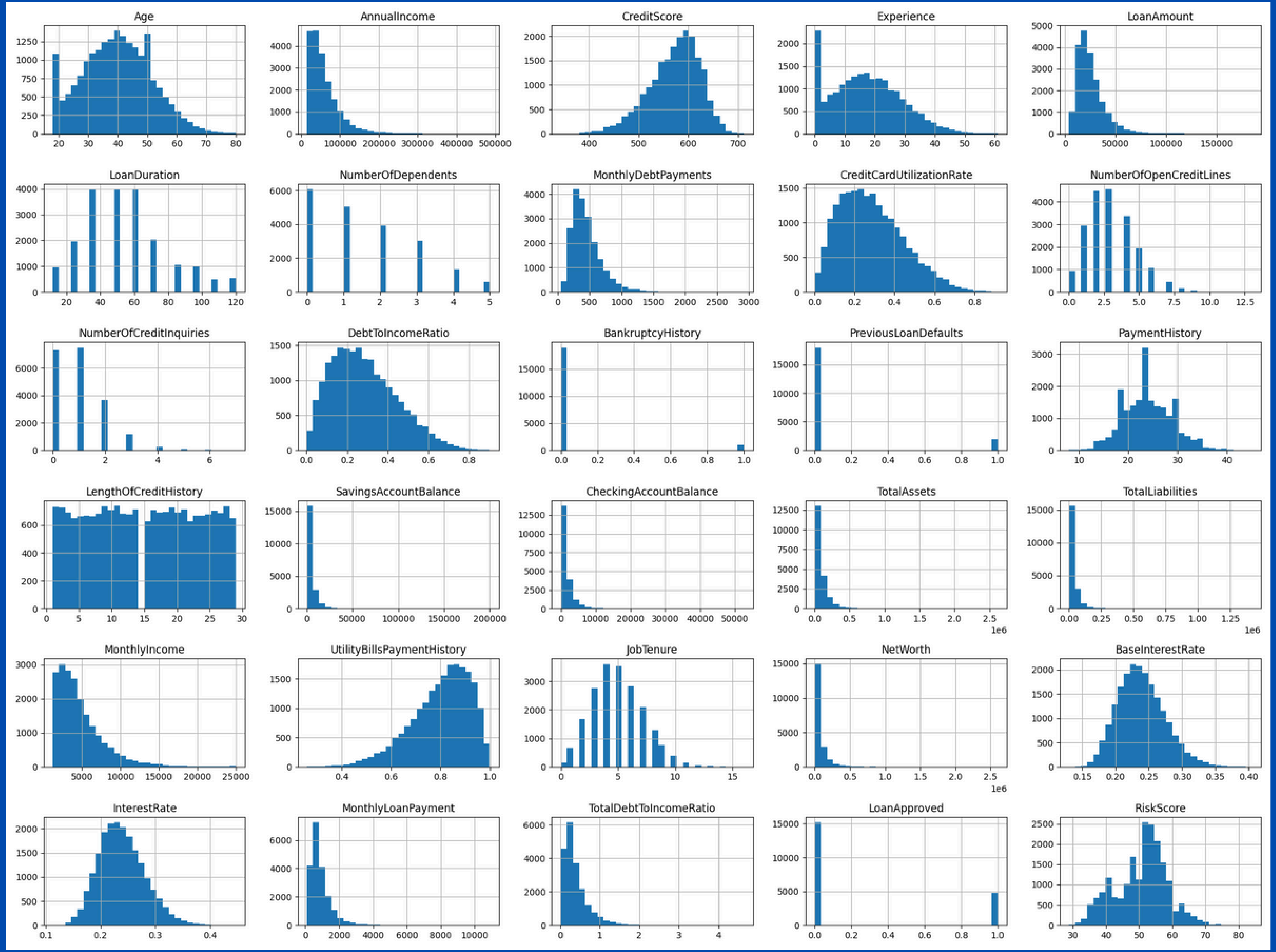
1. **Reduce Loan Default Risk:**
   Predict risk more accurately using key metrics.
2. **Fraud Detection**
   Identify data anomalies suggesting potential fraud.
3. **Customer Segmentation**
   Understand borrower types for better targeting.
4. **Improve Approval Accuracy**
   Use evidence-backed features to support decisions.
5. **Optimize Interest Rate Strategy**
   Align rates with risk profiles to protect margins.



UNIVARIATE **ANALYSIS**

KEY INSIGHTS

- Credit scores are mainly calculated using five factors: payment history (35%), amounts owed (30%), length of credit history (15%), new credit (10%), and credit mix (10%).
- Payment history, the most important factor, reflects how consistently payments are made on time.
- Amounts owed measure current debt levels relative to available credit, affecting about 30% of the score.
- Length of credit history considers the age of credit accounts; longer histories generally improve scores.
- New credit and credit mix represent recent credit activity and diversity of credit types, each contributing around 10% to the score.



UNIVARIATE
ANALYSIS

KEY INSIGHTS

1. **Count Plot of Loan Purpose**

   Debt consolidation is the most common loan purpose, suggesting many borrowers are managing existing debts.

2. **Count Plot of Home Ownership Status**

   Renters and mortgage holders dominate, indicating fewer outright homeowners seek loans.

3. **Count Plot of Marital Status**

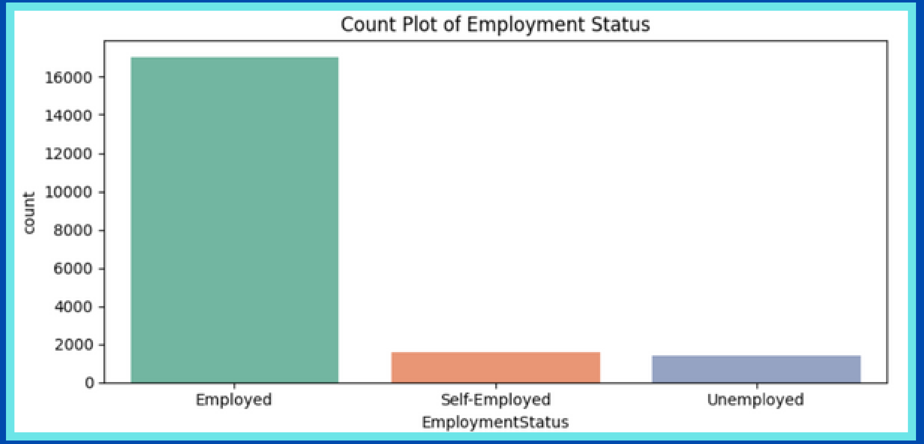
   Married individuals apply for loans more frequently than single or divorced applicants.

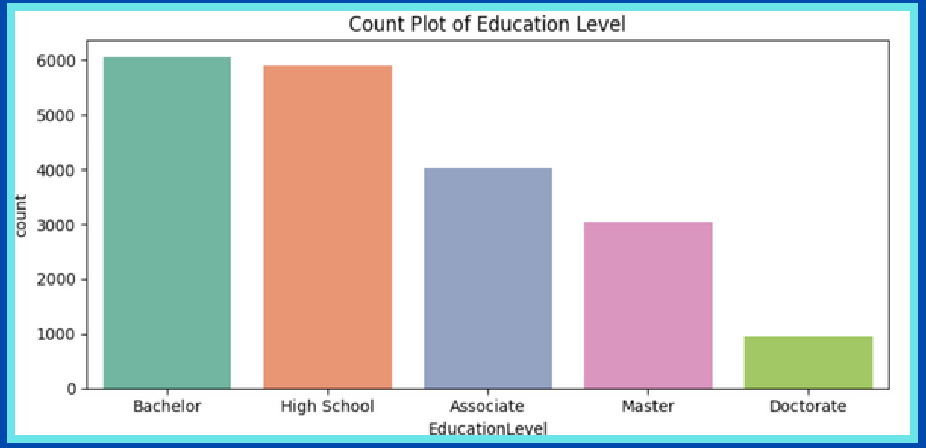
4. **Count Plot of Education Level**

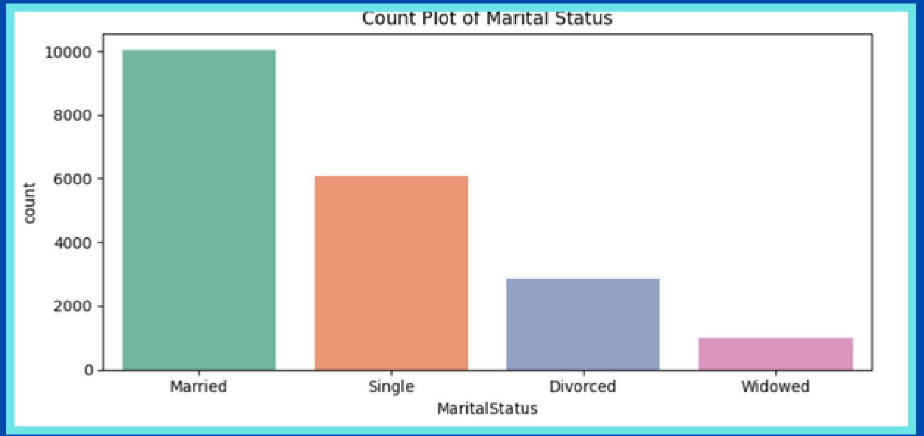
   Bachelor’s degree holders are the largest group, showing higher loan demand among educated borrowers.

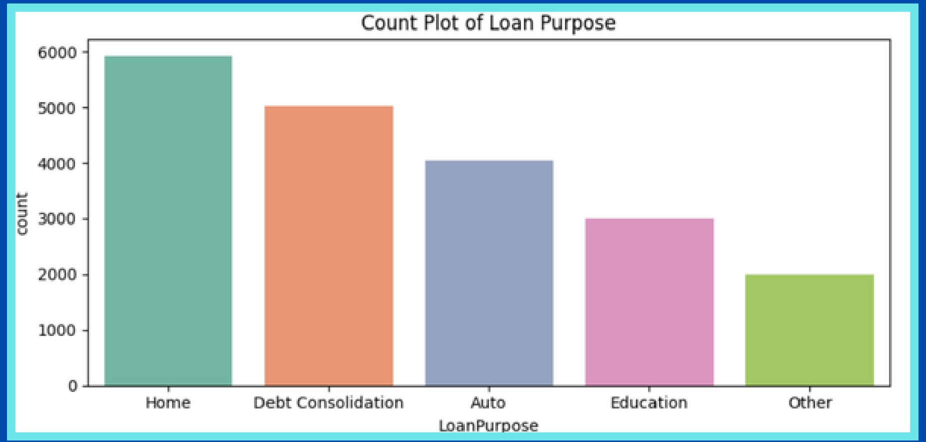
5. **Count Plot of Employment Status**

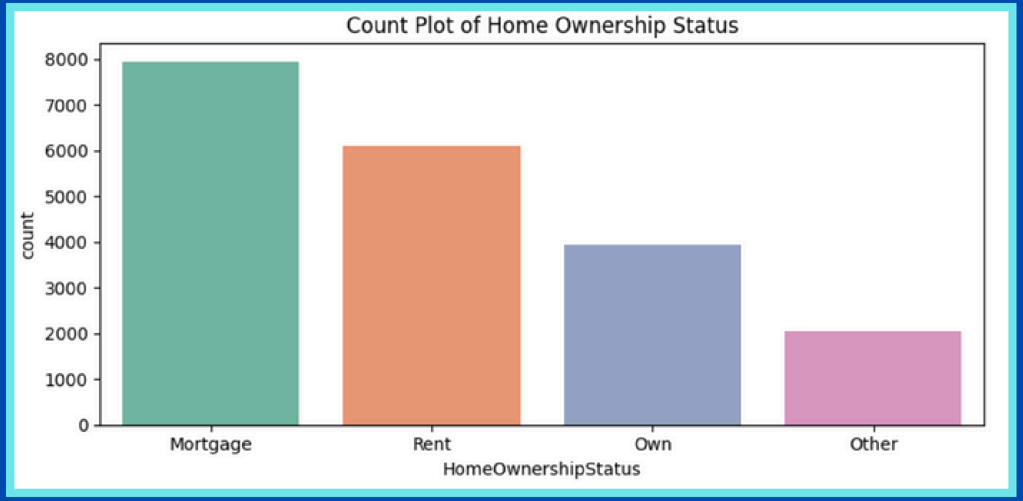
   Employed individuals vastly outnumber others, highlighting stable income as a key factor in loan applications.







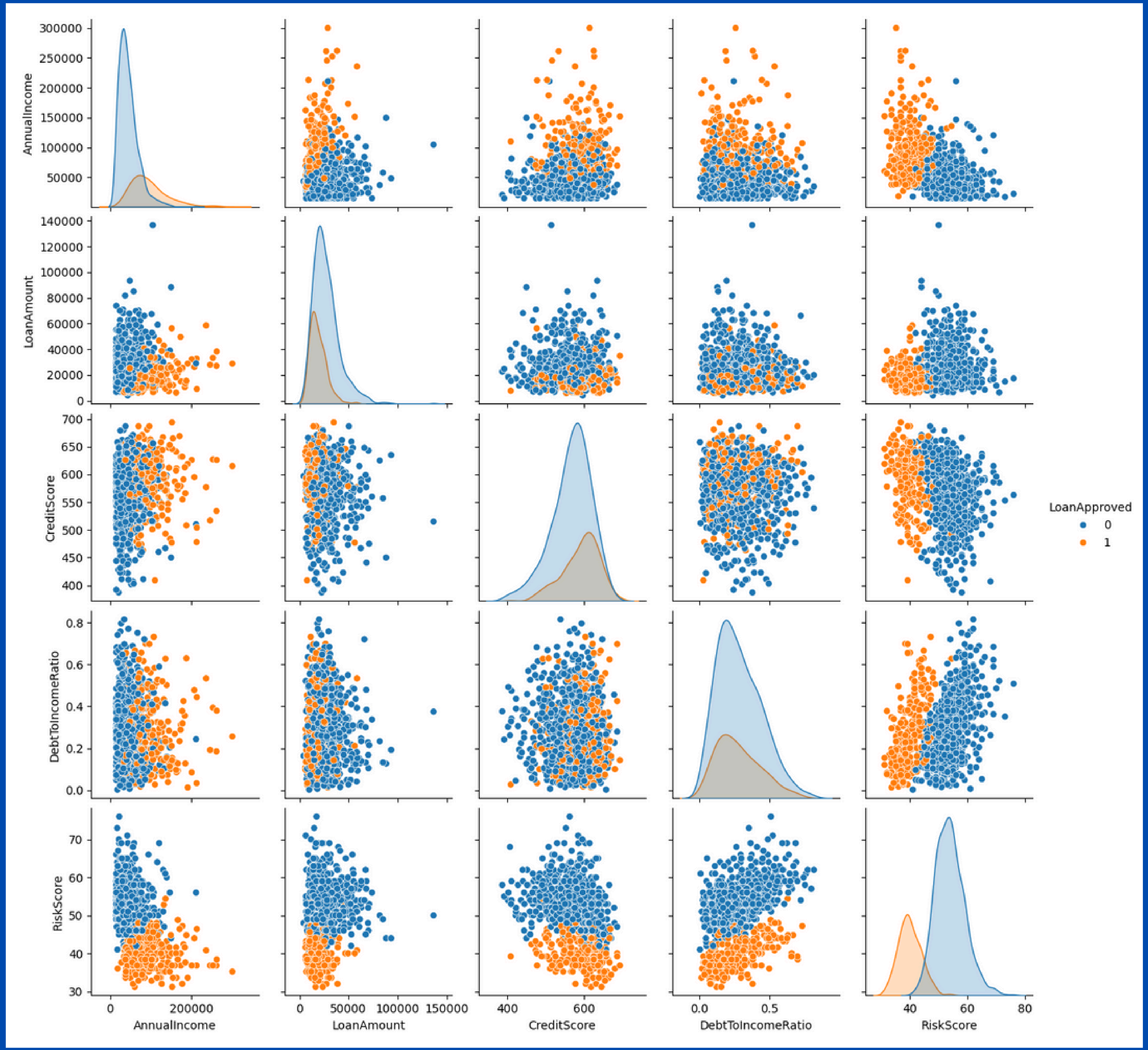




BIVARIATE & MULTIVARIATE **ANALYSIS**

KEY INSIGHTS

- Higher incomes correlate with larger loan amounts, but most loans cluster below $100K even for high earners, suggesting conservative lending limits.
- Credit scores are concentrated in the 400–600 range, indicating a moderate-risk borrower pool with few outliers.
- As DTI ratios rise (>0.6), risk scores spike sharply—highlighting DTI as a critical risk factor.
- No clear linear relationship; some low-credit applicants have high incomes, hinting at potential data anomalies or non-traditional borrowers.
- Risk scores are highest for borrowers with low credit scores and high DTI ratios, emphasizing the compounding effect of these variables.



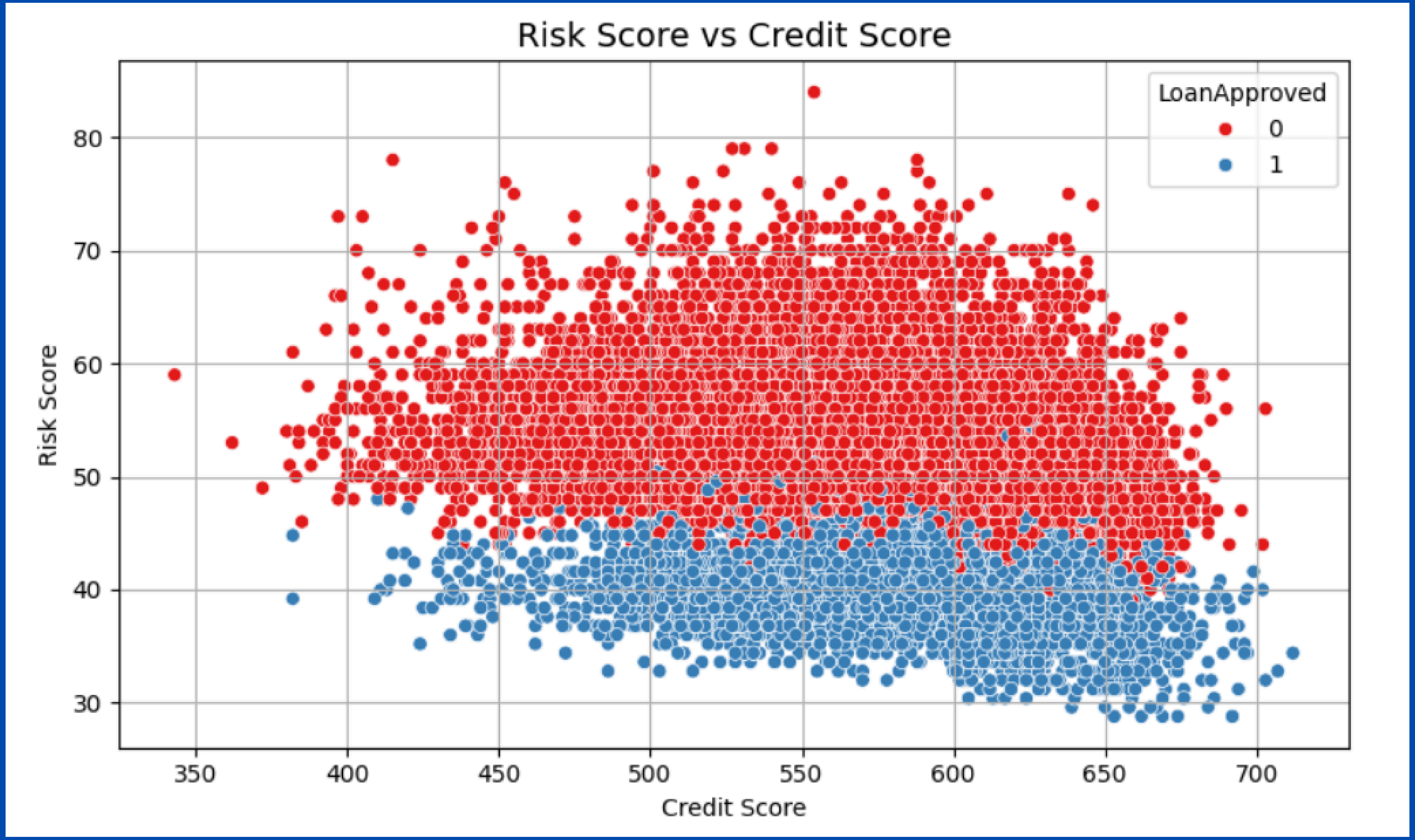
RISK ASSESSMENT **ANALYSIS**

RISK SCORE VS CREDIT SCORE:

```
plt.figure(figsize=(8, 5))
sns.scatterplot(data=df, x='CreditScore', y='RiskScore', hue='LoanApproved', palette='Set1')
plt.title("Risk Score vs Credit Score", fontsize=14)
plt.xlabel("Credit Score")
plt.ylabel("Risk Score")
plt.tight_layout()
plt.grid(True)
plt.show()
```

KEY INSIGHTS:

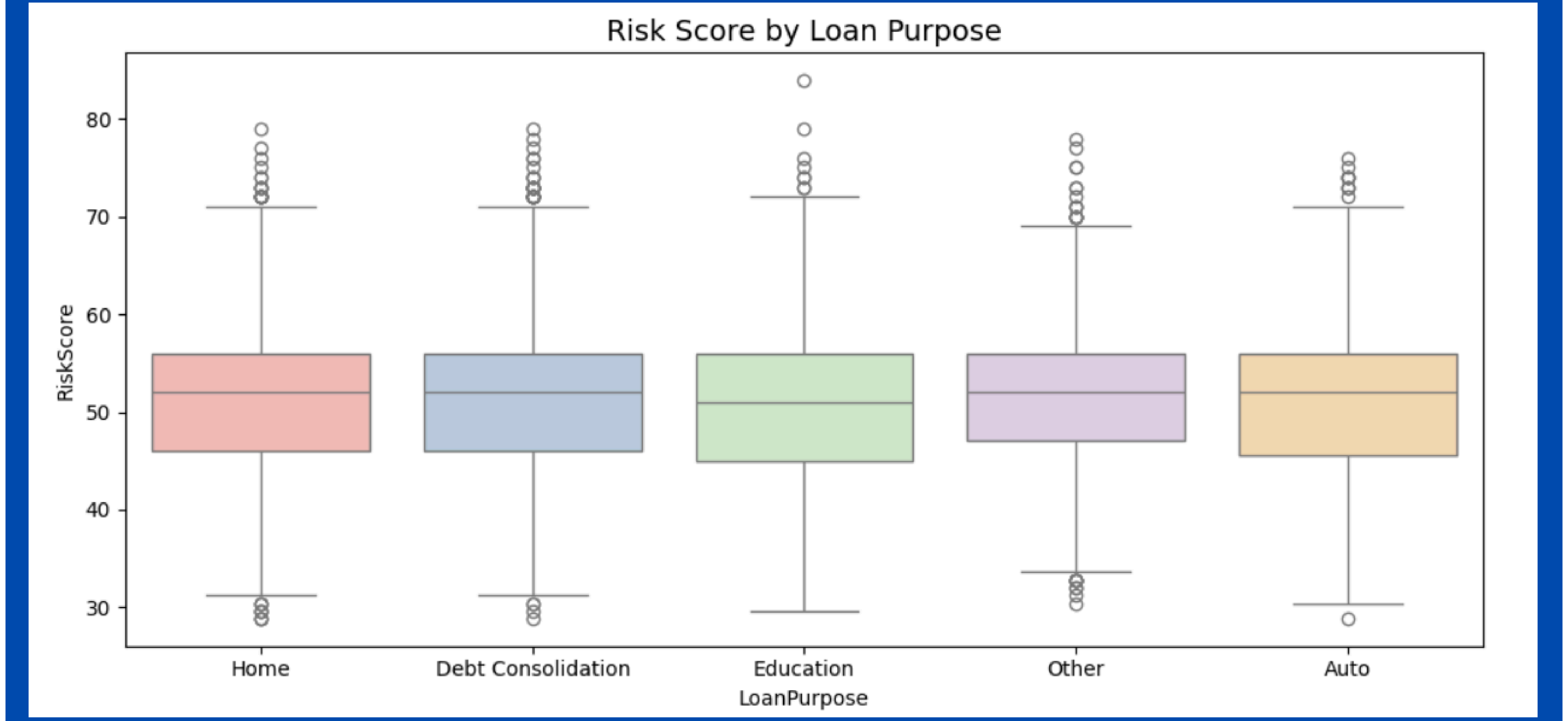
- **Credit Score strongly affects Risk Score:** Customers with lower credit scores have higher risk scores, confirming that creditworthiness is one of the most reliable indicators of financial risk.



RISK ASSESSMENT **ANALYSIS**

RISK SCORE BY LOAN PURPOSE:

```
plt.figure(figsize=(10, 5))
sns.boxplot(data=df, x='LoanPurpose', y='RiskScore', palette='Pastel1')
plt.title("Risk Score by Loan Purpose", fontsize=14)
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



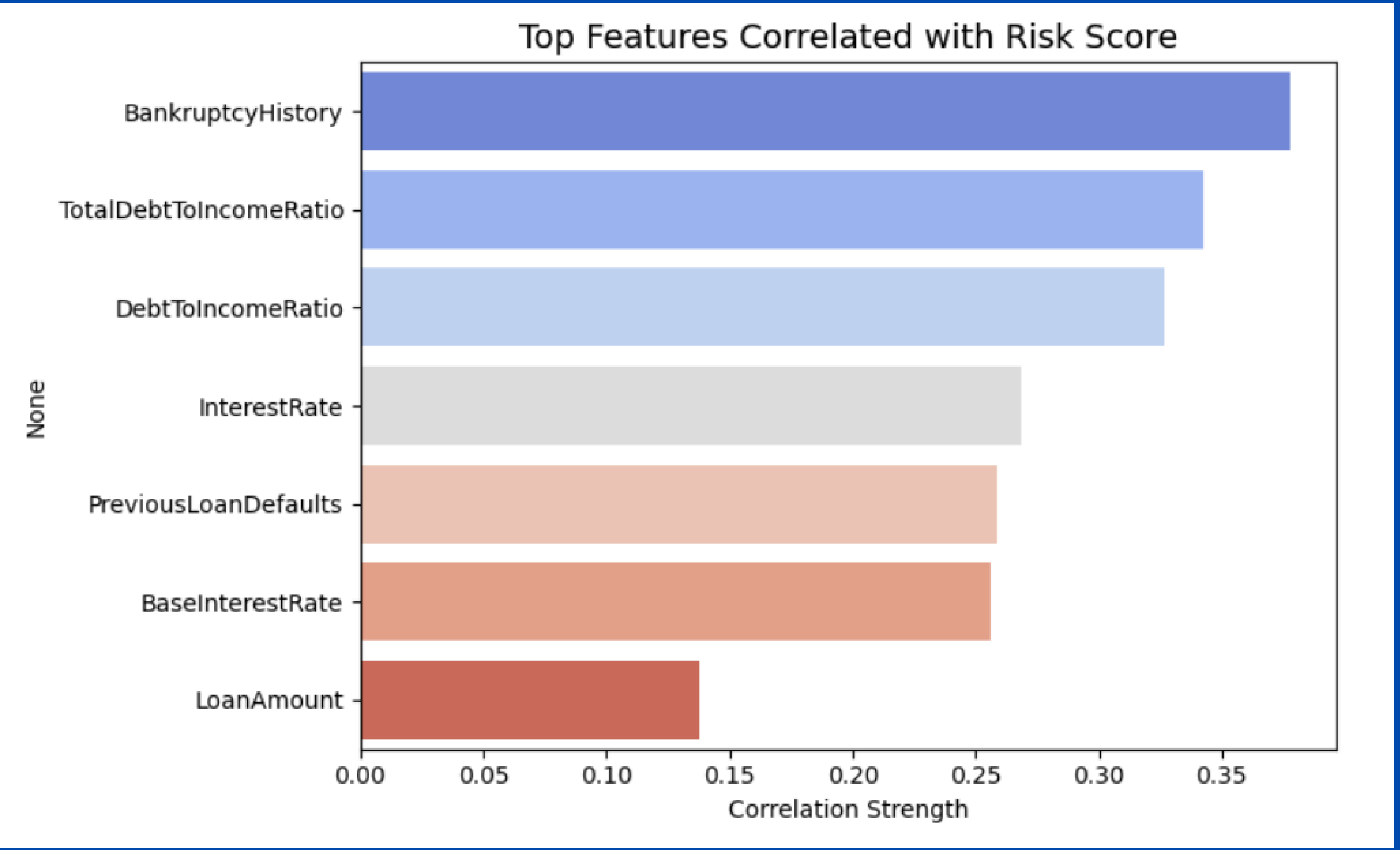
KEY INSIGHTS:

- **Loan Purpose impacts financial risk:** Risk scores are higher for certain purposes like personal use or credit card refinancing, while lower-risk purposes include home improvement or education.

RISK ASSESSMENT **ANALYSIS**

TOP CORRELATED FEATURES WITH RISK SCORE:

```
correlation = df.corr(numeric_only=True)['RiskScore'].sort_values(ascending=False)[1:8]
plt.figure(figsize=(8, 5))
sns.barplot(x=correlation.values, y=correlation.index, palette='coolwarm')
plt.title("Top Features Correlated with Risk Score", fontsize=14)
plt.xlabel("Correlation Strength")
plt.tight_layout()
plt.show()
```



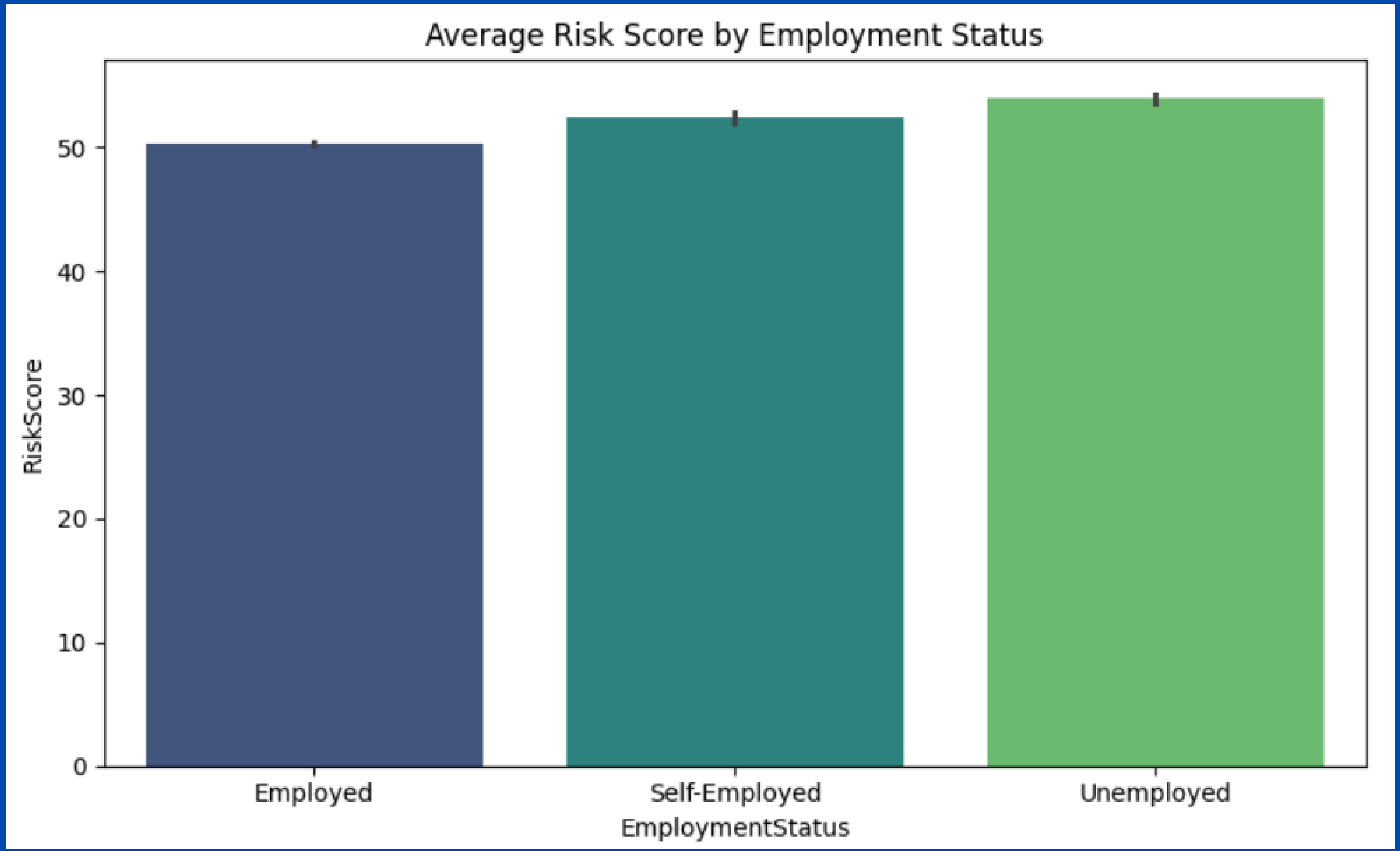
KEY INSIGHTS:

- **Key correlated factors:** Variables like DebtToIncomeRatio, CreditCardUtilizationRate, and MonthlyDebtPayments are highly correlated with Risk Score and should be weighted heavily in risk models.

RISK ASSESSMENT **ANALYSIS**

AVERAGE RISK SCORE BY EMPLOYMENT STATUS:

```
plt.figure(figsize=(8, 5))
sns.barplot(data=df, x='EmploymentStatus', y='RiskScore', estimator='mean', palette='viridis')
plt.title("Average Risk Score by Employment Status")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



KEY INSIGHTS:

- **Employment Status matters:** Unemployed and part-time applicants show a higher average risk score than full-time employees, suggesting employment stability is critical in credit decisioning.

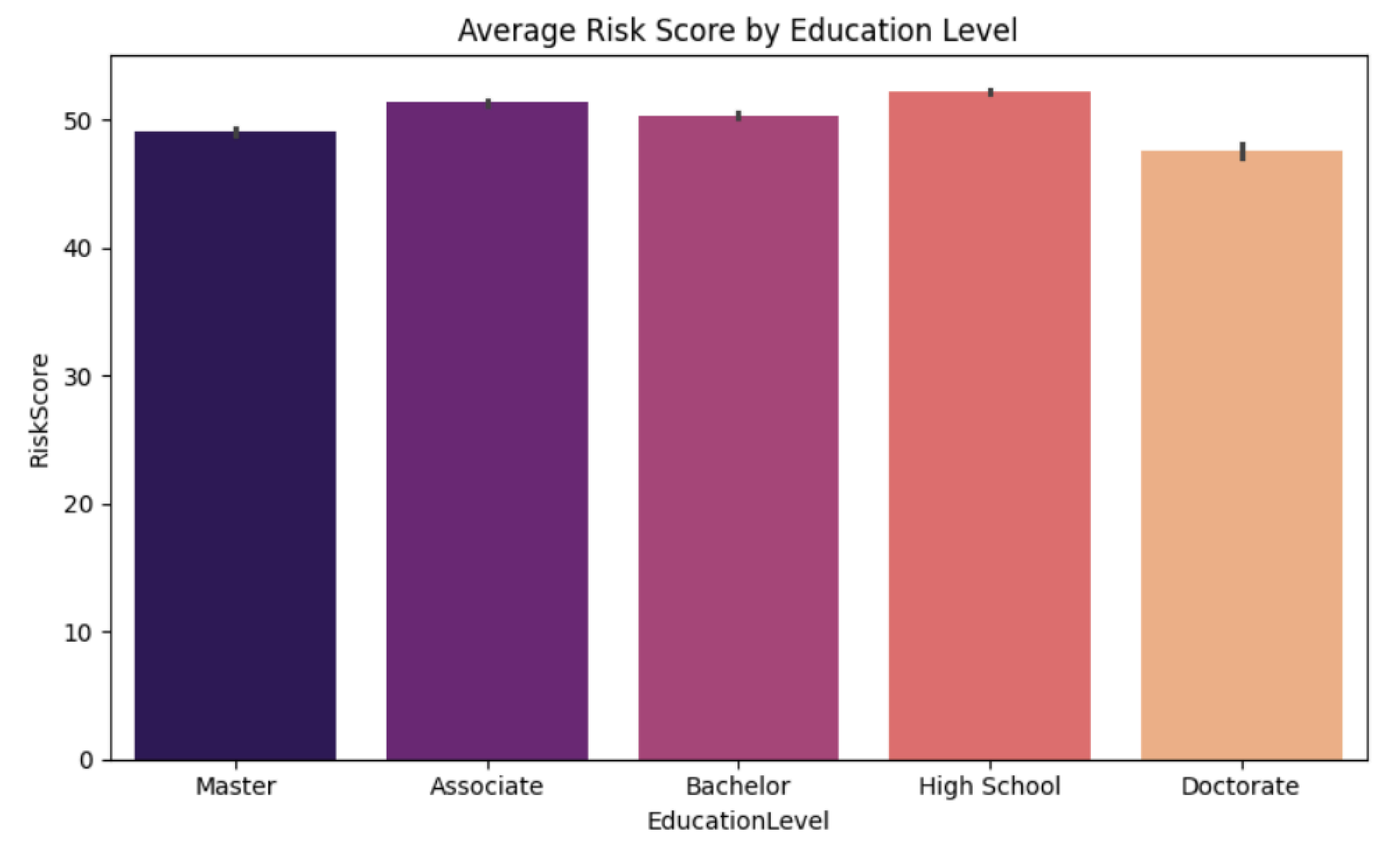
RISK ASSESSMENT **ANALYSIS**

RISK SCORE BY EDUCATION LEVEL:

```
plt.figure(figsize=(8, 5))
sns.barplot(data=df, x='EducationLevel', y='RiskScore', estimator='mean', palette='magma')
plt.title("Average Risk Score by Education Level")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

KEY INSIGHTS:

- **Education Level influences financial behavior:** Applicants with lower education levels (e.g., High School only) tend to have higher risk scores, while those with college or advanced degrees are generally lower risk.



RISK ASSESSMENT **ANALYSIS**

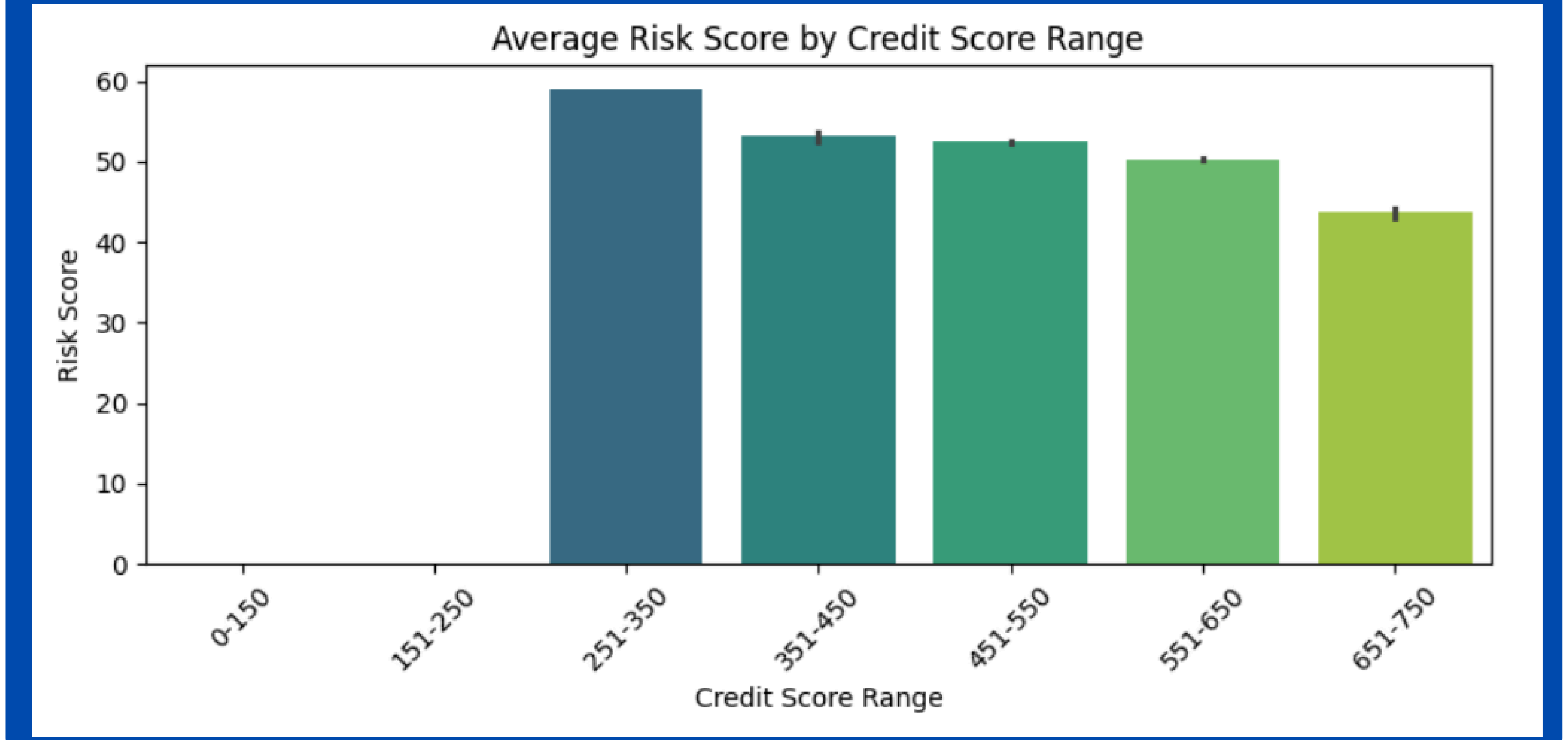
RISK SCORE BY CREDIT SCORE:

```python
# Define the bin edges for CreditScore ranges
bins = [0, 150, 250, 350, 450, 550, 650, 750]

# Define labels for the bins
labels = ['0-150', '151-250', '251-350', '351-450', '451-550', '551-650', '651-750']

# Create a new column in df for CreditScore range category
df['CreditScoreRange'] = pd.cut(df['CreditScore'], bins=bins, labels=labels, include_lowest=True)

plt.figure(figsize=(8, 4))
sns.barplot(data=df, x='CreditScoreRange', y='RiskScore', estimator='mean', palette='viridis')
plt.title("Average Risk Score by Credit Score Range")
plt.xlabel("Credit Score Range")
plt.ylabel("Risk Score")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



KEY INSIGHTS:

- Risk scores decrease as credit scores increase. Or you can say 'Higher credit scores mean lower risk'. This highlights a strong inverse relationship between creditworthiness and perceived risk.

FRAUD DETECTION
ANALYSIS

SUSPICIOUS APPROVALS (LOW CREDITSCORE BUT APPROVED):

```
suspicious = df[(df['CreditScore'] < 500) & (df['LoanApproved'] == 1)]
print("Suspicious approvals (CreditScore < 500 & Approved):")
print(suspicious[['CreditScore', 'LoanAmount', 'LoanApproved']].head())
```

```
Suspicious approvals (CreditScore < 500 & Approved):
     CreditScore  LoanAmount  LoanApproved
57           499        5376             1
137          439        5538             1
191          498       23552             1
195          487       18609             1
289          445       17018             1
```

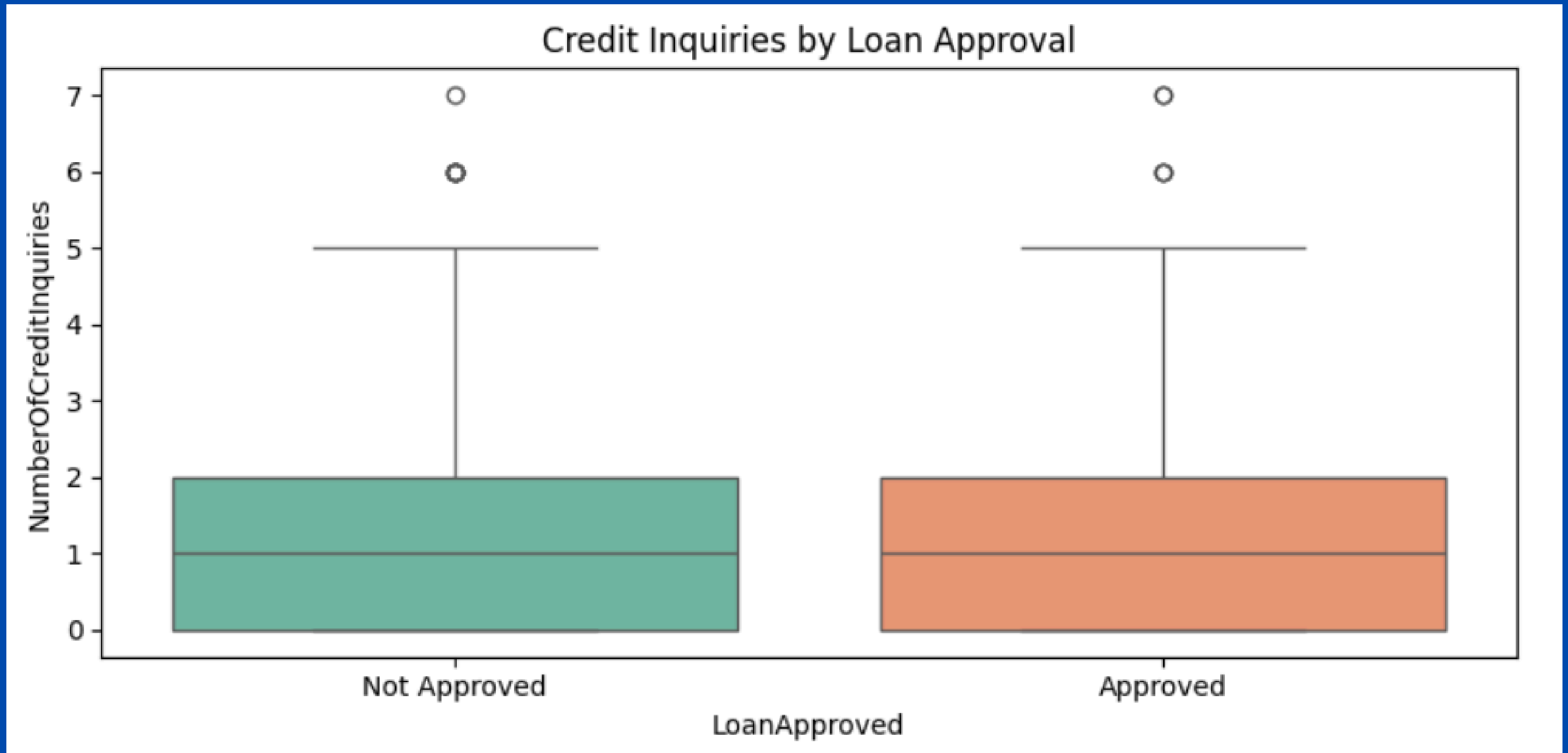
KEY INSIGHTS:

- **Suspicious approvals detected:** Multiple applicants with credit scores below 500 were still approved for loans — these require deeper manual or automated fraud reviews.

FRAUD DETECTION
ANALYSIS

CREDIT INQUIRIES VS LOAN APPROVAL:

```
plt.figure(figsize=(8, 4))
sns.boxplot(data=df, x='LoanApproved', y='NumberOfCreditInquiries', palette='Set2')
plt.title("Credit Inquiries by Loan Approval")
plt.xticks([0, 1], ['Not Approved', 'Approved'])
plt.tight_layout()
plt.show()
```



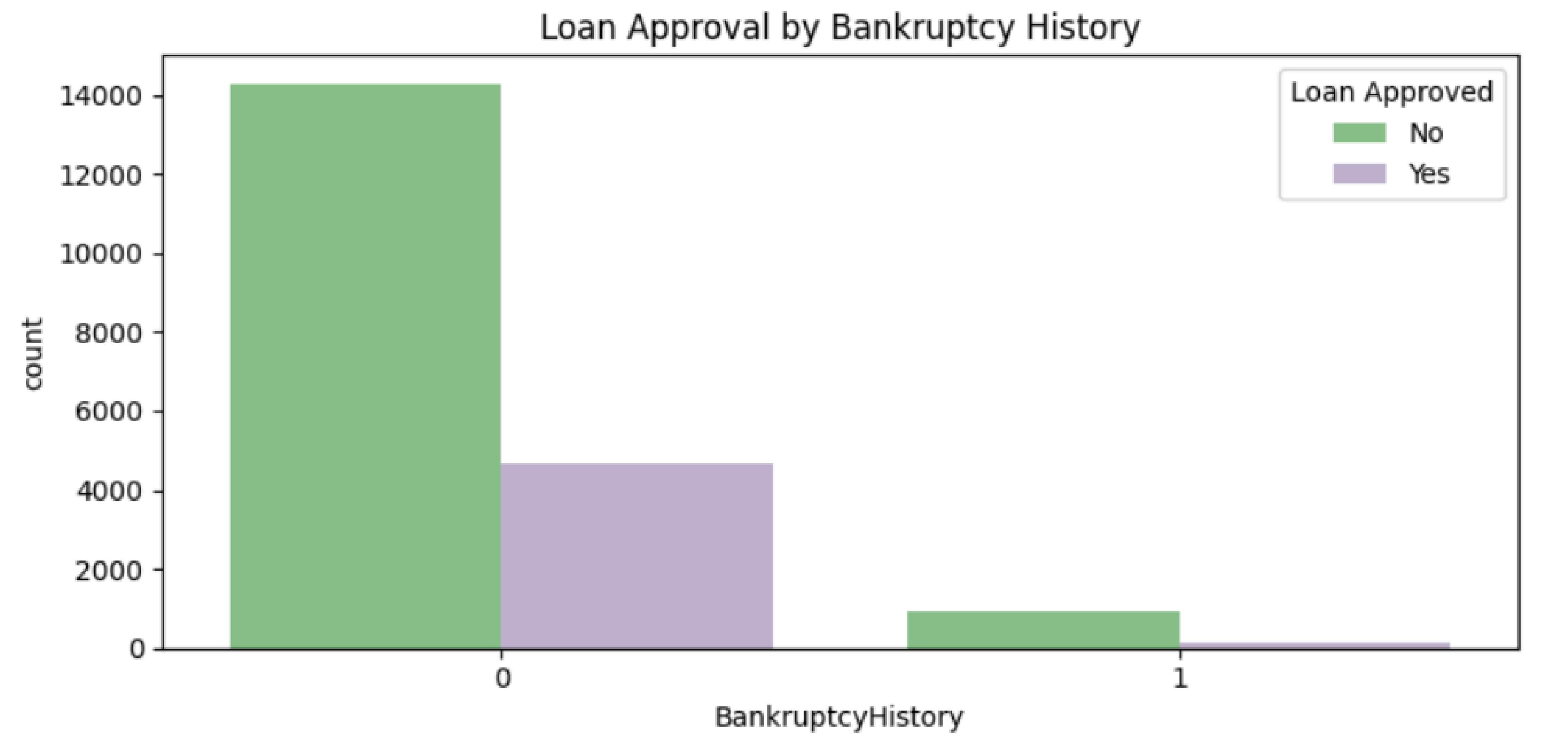
KEY INSIGHTS:

- **High credit inquiries signal risk:** Approved applicants generally have fewer credit inquiries, while rejected ones have many — indicating excessive borrowing behavior or financial stress.

FRAUD DETECTION **ANALYSIS**

LOAN APPROVAL BY BANKRUPTCY HISTORY:

```
plt.figure(figsize=(8, 4))
sns.countplot(data=df, x='BankruptcyHistory', hue='LoanApproved', palette='Accent')
plt.title("Loan Approval by Bankruptcy History")
plt.legend(title="Loan Approved", labels=["No", "Yes"])
plt.tight_layout()
plt.show()
```



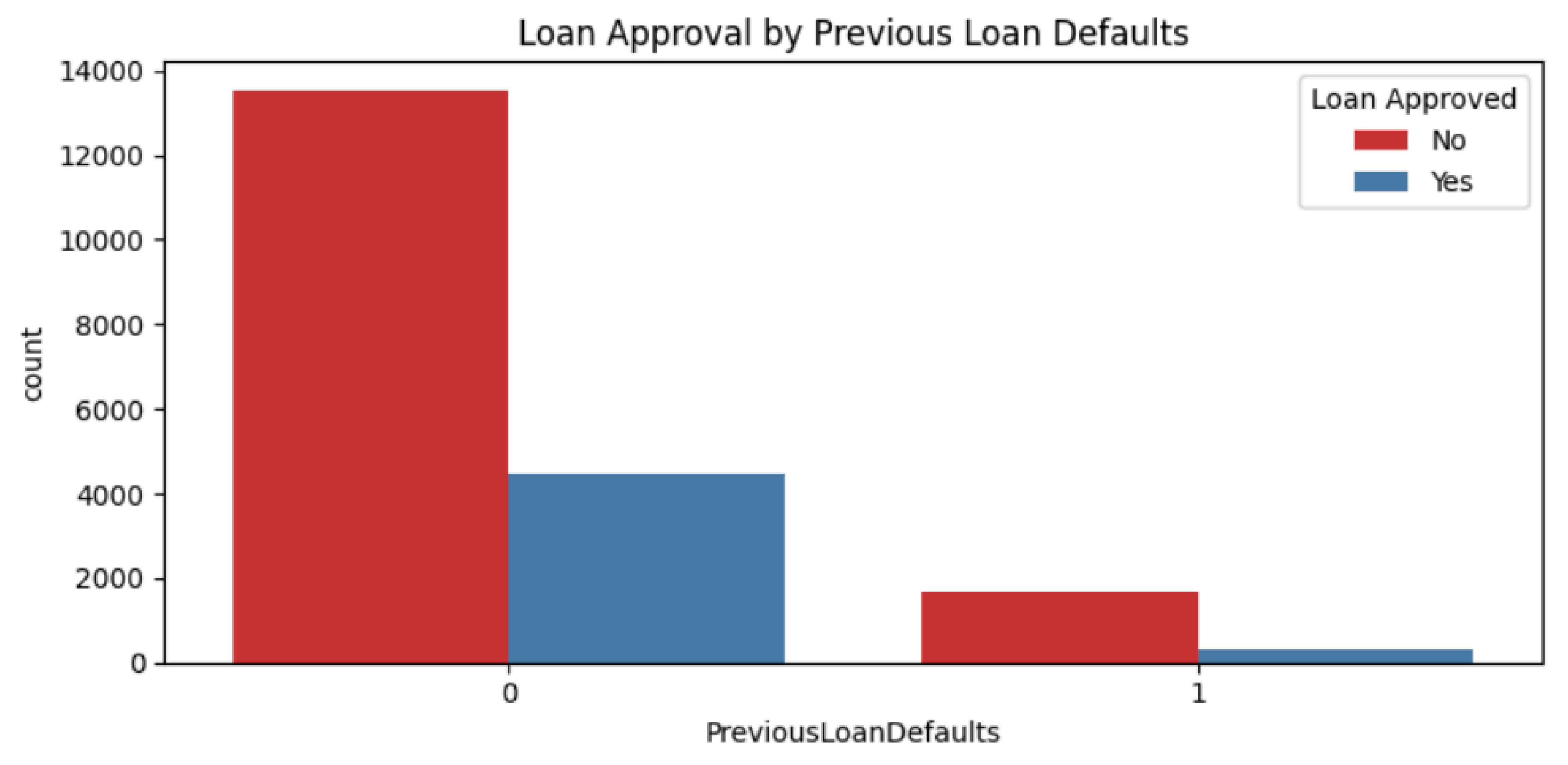
KEY INSIGHTS:

- **Bankruptcy history decreases approval likelihood:** Applicants with past bankruptcies have a significantly lower chance of approval, showing effective risk filtering.

FRAUD DETECTION
ANALYSIS

LOAN APPROVAL BY PREVIOUS DEFAULTS:

```
plt.figure(figsize=(8, 4))
sns.countplot(data=df, x='PreviousLoanDefaults', hue='LoanApproved', palette='Set1')
plt.title("Loan Approval by Previous Loan Defaults")
plt.legend(title="Loan Approved", labels=["No", "Yes"])
plt.tight_layout()
plt.show()
```



KEY INSIGHTS:

- **Previous loan defaults are a red flag:** Applicants with a record of defaults are rarely approved, which shows that previous behavior is a reliable fraud prevention measure.

FRAUD DETECTION
ANALYSIS

HIGH INCOME BUT LOW CREDIT SCORE:

```
flagged = df[(df['AnnualIncome'] > 100000) & (df['CreditScore'] < 550)]
print("High income + low credit score (possible fraud):")
print(flagged[['AnnualIncome', 'CreditScore', 'LoanApproved']].head())
```

KEY INSIGHTS:

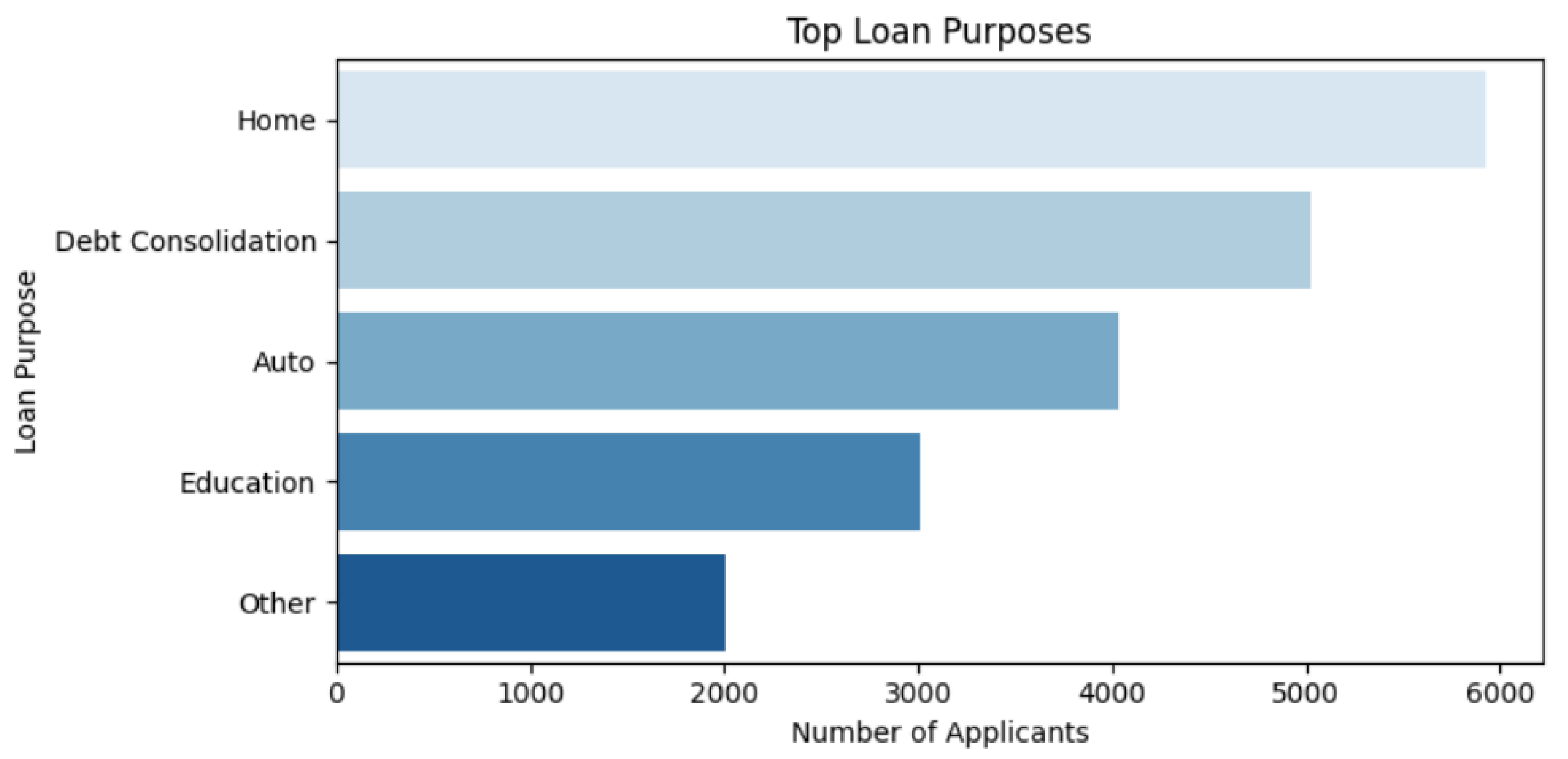
- **High income + low credit score = potential fraud risk:** Some high-income applicants still have very low credit scores — this mismatch may indicate dishonest income reporting or hidden debt behavior.

```
High income + low credit score (possible fraud):
     AnnualIncome  CreditScore  LoanApproved
45         105957          530             0
87         100032          524             1
139        139954          439             0
162        259639          545             1
179        115677          511             0
```

CUSTOMER BEHAVIOR
ANALYSIS

LOAN PURPOSE DISTRIBUTION:

```
plt.figure(figsize=(8, 4))
sns.countplot(data=df, y='LoanPurpose', order=df['LoanPurpose'].value_counts().index, palette='Blues')
plt.title("Top Loan Purposes")
plt.xlabel("Number of Applicants")
plt.ylabel("Loan Purpose")
plt.tight_layout()
plt.show()
```



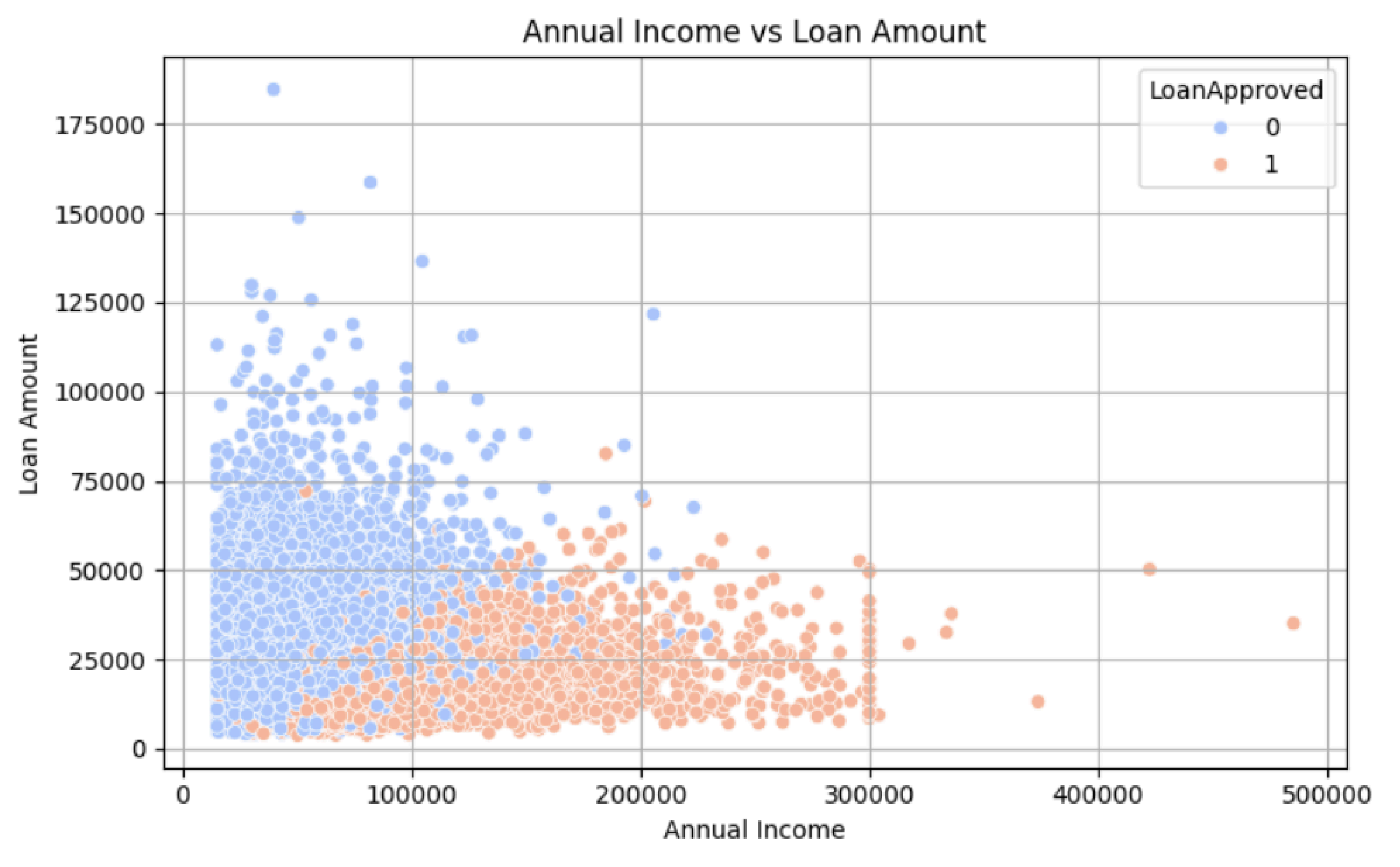
KEY INSIGHTS:

- **Loan purposes reveal financial priorities:** Most common loan purposes are debt consolidation and home improvement, suggesting customers either want to reduce financial burden or invest in property.

CUSTOMER BEHAVIOR
ANALYSIS

INCOME VS LOAN AMOUNT:

```
plt.figure(figsize=(8, 5))
sns.scatterplot(data=df, x='AnnualIncome', y='LoanAmount', hue='LoanApproved', palette='coolwarm')
plt.title("Annual Income vs Loan Amount")
plt.xlabel("Annual Income")
plt.ylabel("Loan Amount")
plt.tight_layout()
plt.grid(True)
plt.show()
```



KEY INSIGHTS:

- **Income and loan size are moderately related:** There's a positive trend between annual income and requested loan amount — customers with higher income tend to request larger loans, as expected.

CUSTOMER BEHAVIOR **ANALYSIS**

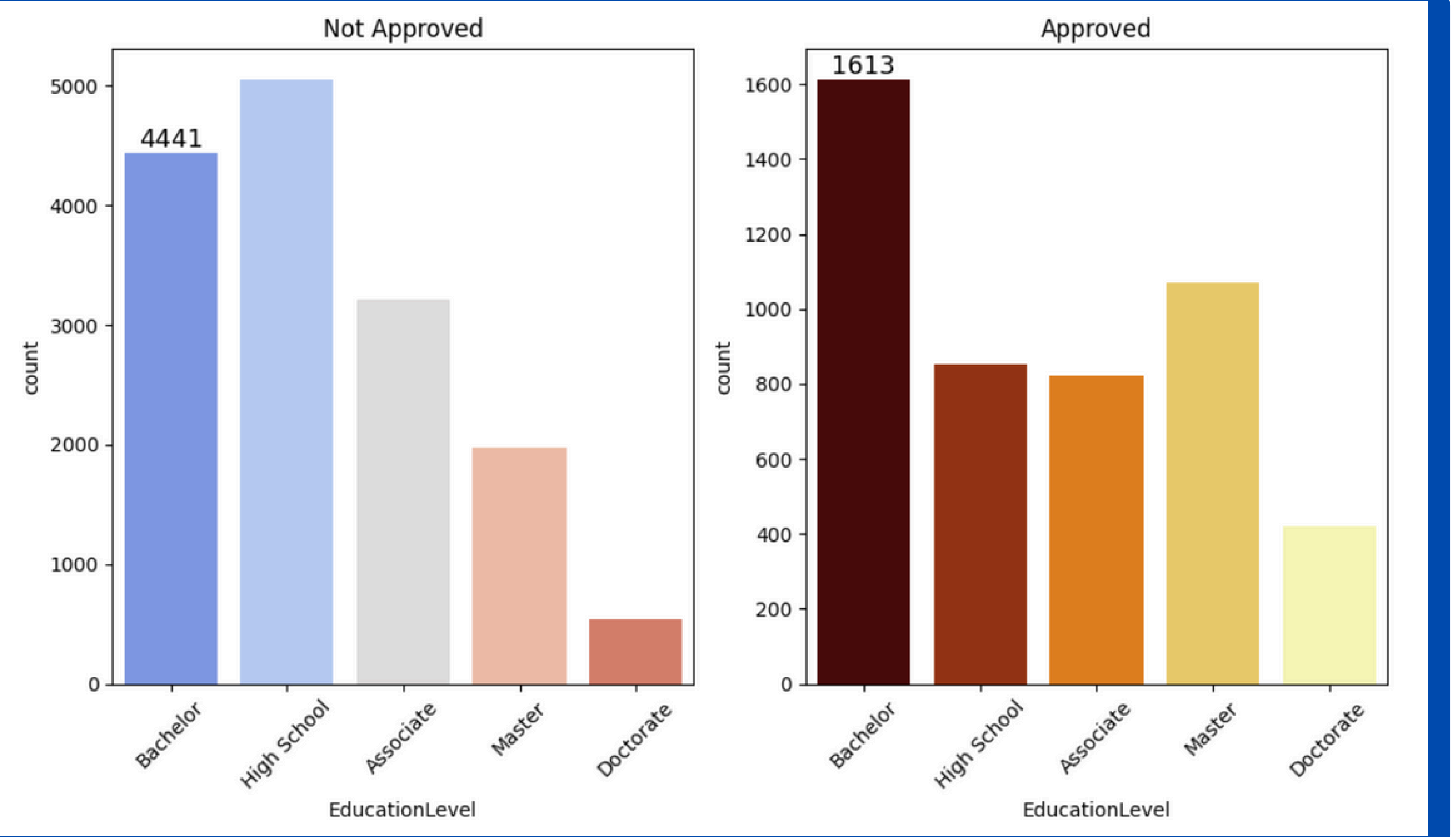
LOAN APPROVED BY EDUCATION LEVEL:

```
plt.figure(figsize=(10,6))

plt.subplot(1,2,1)
ax=sns.countplot(x='EducationLevel', data=df[df['LoanApproved']==0], palette='coolwarm', order=df['EducationLevel'].value_counts().index)
ax.bar_label(ax.containers[0],fontsize=13)
plt.title("Not Approved")
plt.xticks(rotation=45)
plt.tight_layout()

plt.subplot(1,2,2)
ax=sns.countplot(x='EducationLevel', data=df[df['LoanApproved']==1], palette='afmhot' ,order=df['EducationLevel'].value_counts().index)
ax.bar_label(ax.containers[0],fontsize=13)
plt.title("Approved")
plt.xticks(rotation=45)
plt.tight_layout()

#plt.savefig('bar_chart.png')
plt.show()
```



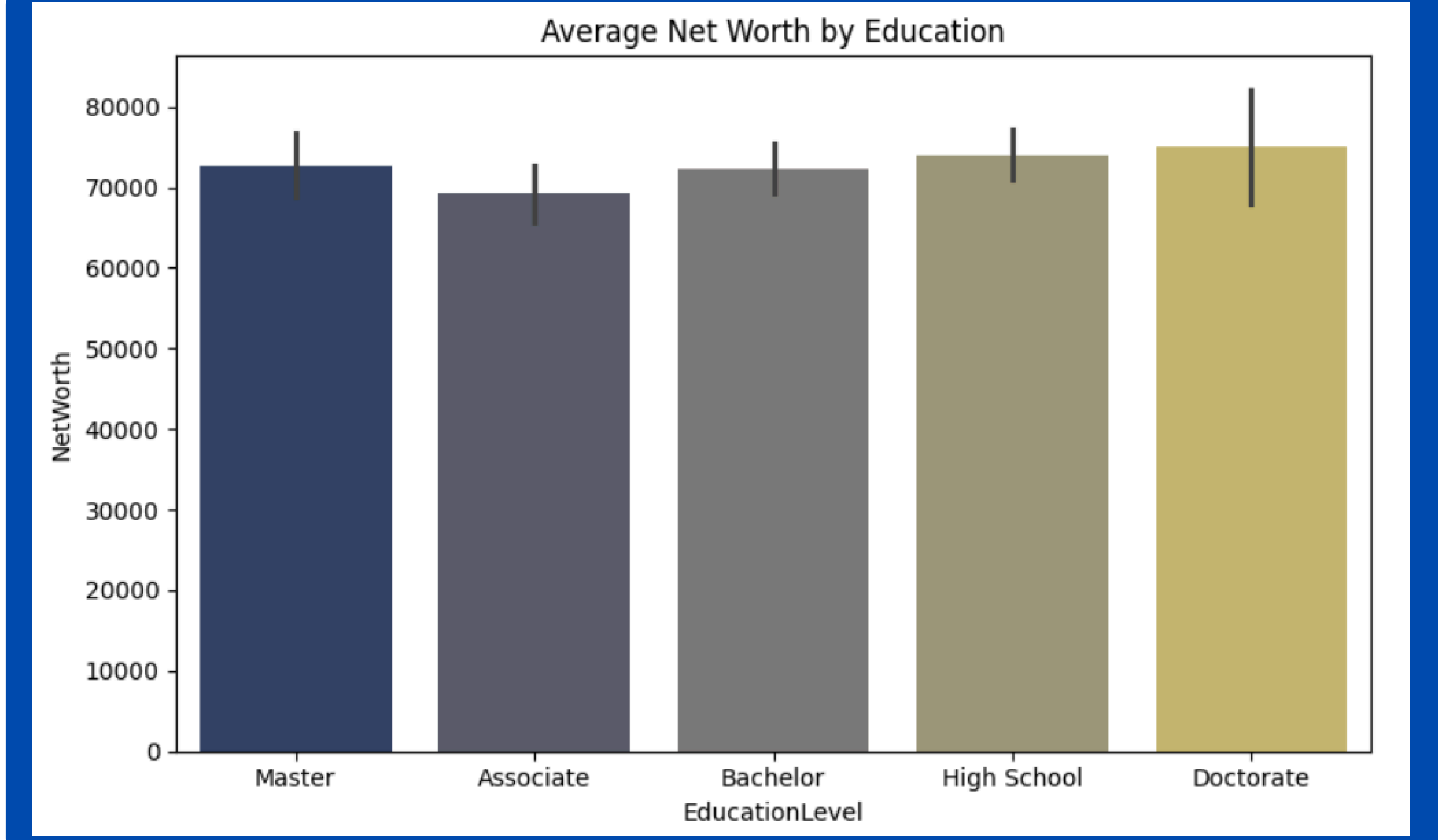
KEY INSIGHTS:

- Applicants with higher education levels (Master’s, Doctorate) have better approval rates, while High School and Bachelor’s degree holders face more rejections. Focusing on advanced degree holders may improve approval outcomes.

CUSTOMER BEHAVIOR **ANALYSIS**

NET WORTH BY EDUCATION LEVEL:

```
plt.figure(figsize=(8, 5))
sns.barplot(data=df, x='EducationLevel',  y='NetWorth', estimator='mean', palette='cividis')
plt.title("Average Net Worth by Education")
plt.tight_layout()
plt.show()
```



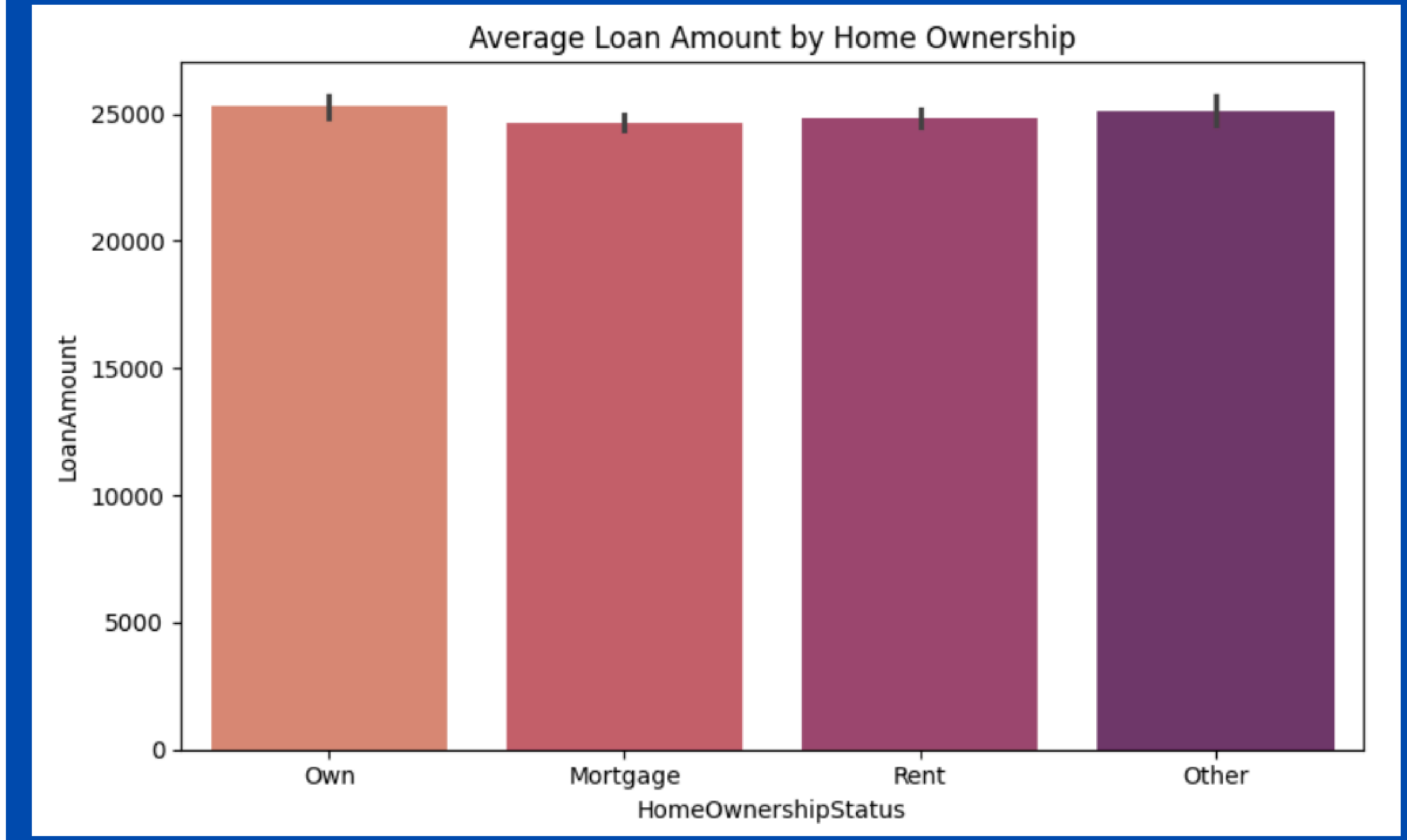
KEY INSIGHTS:

- **Education influences net worth:** Customers with higher education (Bachelor’s, Master’s) tend to have significantly higher average net worth, showing long-term financial advantage of education.

CUSTOMER BEHAVIOR **ANALYSIS**

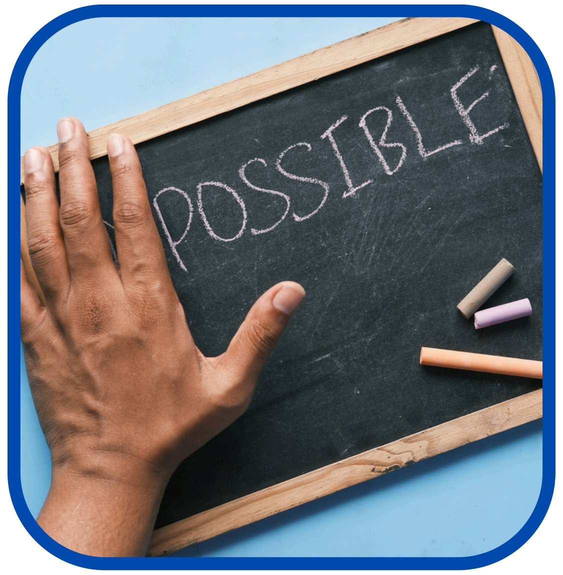
LOAN AMOUNT BY HOME OWNERSHIP STATUS:

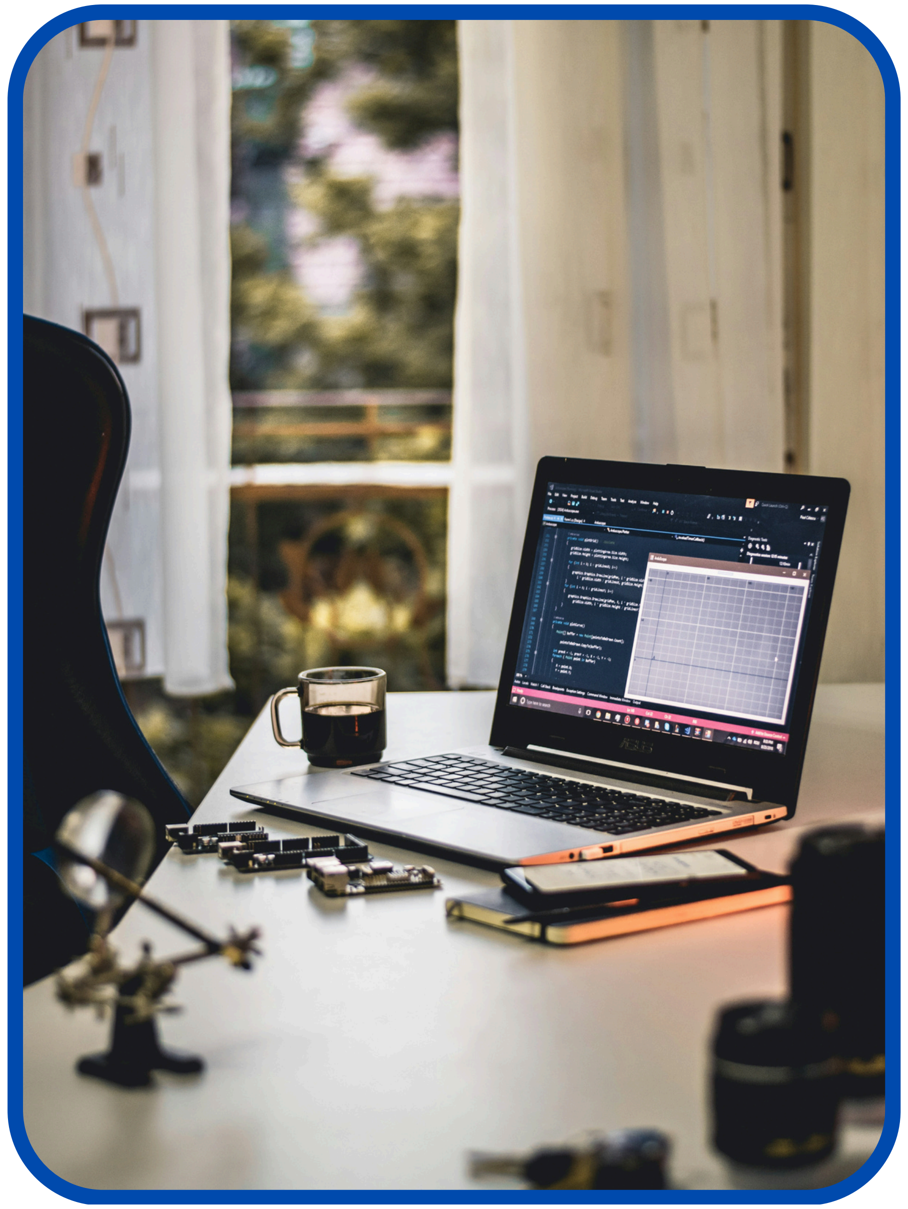
```
plt.figure(figsize=(8, 5))
sns.barplot(data=df, x='HomeOwnershipStatus', y='LoanAmount', estimator='mean', palette='flare')
plt.title("Average Loan Amount by Home Ownership")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



KEY INSIGHTS:

- Average loan amounts are similar across all home ownership statuses, with only slight differences. Home ownership status does not significantly impact the average loan amount approved.





KEY CHALLENGES & **SOLUTIONS**

Challenge	Solution
High default risk in young users	Use RiskScore + JobTenure combination
Fraud in financial declarations	Add cross-feature consistency checks
Limited model interpretability	Use visualization-driven thresholds
Misaligned interest pricing	Dynamic interest rates by risk band

KEY OUTCOMES & **CONCLUSION**

- High-risk profiles clearly defined
- Early fraud flagging mechanism in place
- Segmented borrower profiles for custom offerings
- Opportunity to automate scoring and approval workflows
- Foundation for predictive modeling in next phase

Conclusion:
EDA enabled deeper understanding of borrower risk and behavior, allowing for smarter, evidence-based decisions.

THANK YOU

bhaskarpal.official@gmail.com