

EDA REPORT - COMPREHENSIVE EDA REPORT: EMI ELIGIBILITY & RISK ASSESSMENT

Project Overview

Domain: FinTech – EMI Eligibility & Risk Assessment

Dataset Size: ~400,000 customer financial profiles

Input Features: 25

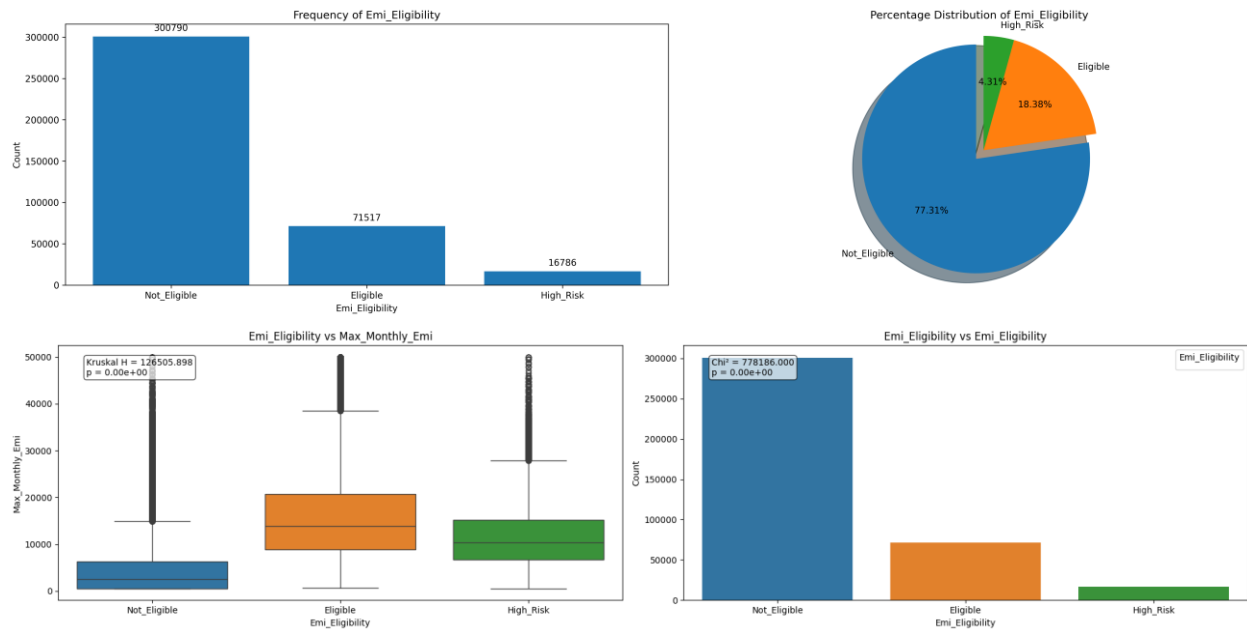
Target Variables: 2

- **Emi_Eligibility** - Classification (Eligible / High-Risk / Not Eligible)
- **Max_Monthly_Emi** - Regression (maximum EMI repayment capacity)

Target Variables:

Classification Target:

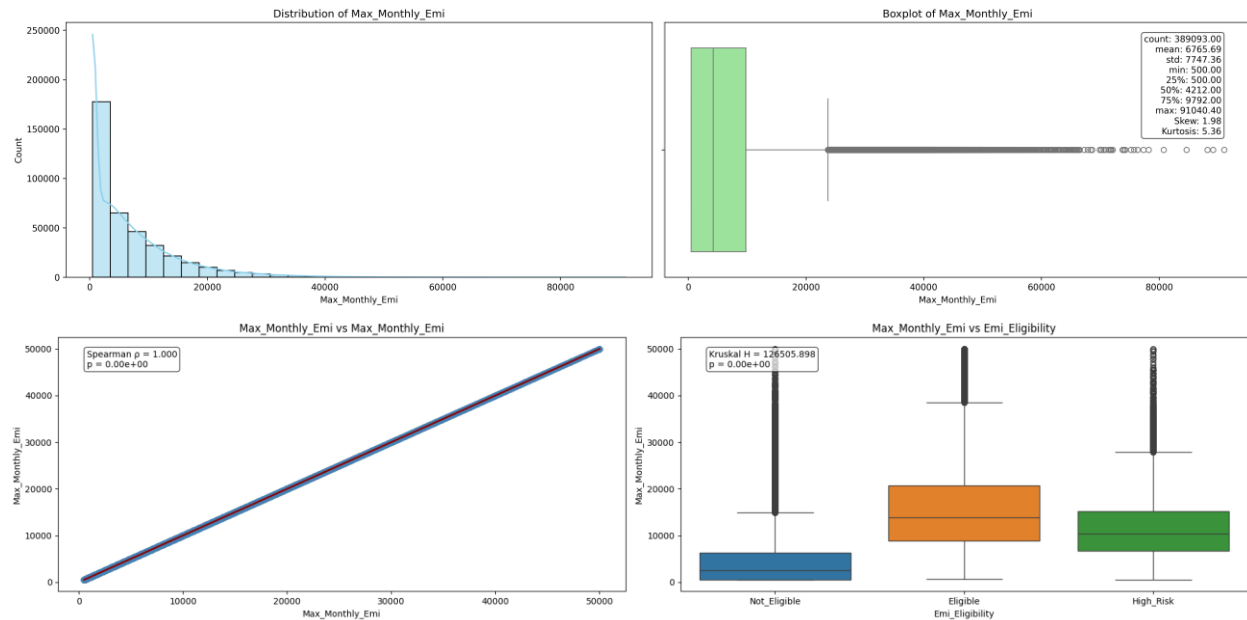
- Emi_Eligibility: Primary classification target with 3 classes** (**Eligible:** Low risk, comfortable EMI affordability; **High_Risk:** Marginal case, requires higher interest rates; **Not_Eligible:** High risk, loan not recommended)



Aspect	Observation	Insight/Interpretation
Variable Type	Categorical (3 classes)	EMI approval decision levels
Distribution Shape	Highly imbalanced	~ 77% Not Eligible (Majority class) ~18% Eligible ~4% High_Risk
Skew & Kurtosis	Not applicable (categorical)	None
Outliers	Not applicable (categorical)	None
Relationship with Max_Monthly_Emi (Regression target)	- Eligible group has highest EMI capability, Not Eligible has lowest - Statistical Test - Kruskal $p=0.00$ (Significant)	- Higher EMI capacity → higher approval - EMI capacity differs meaningfully across eligibility groups
Relationship with Emi_Eligibility (Classification target)	Not applicable	None
Modeling Impact	Must handle class imbalance Important variable interactions with income, expenses, credit score	Use class weights, sampling methods

Regression Target:

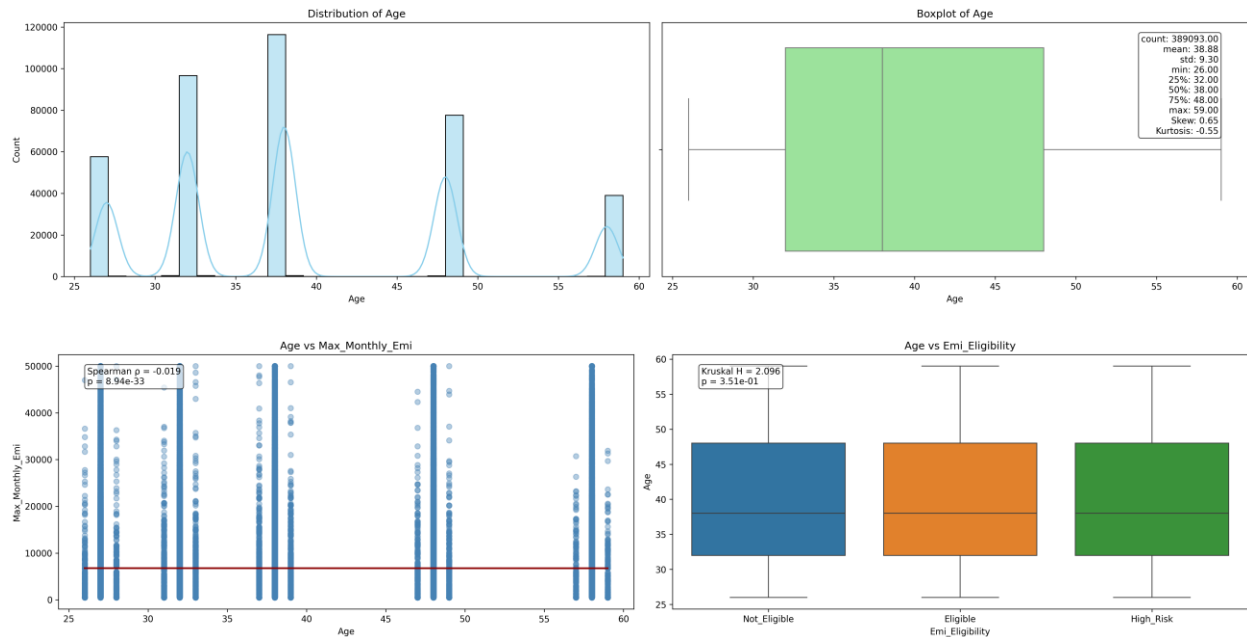
- **Max_Monthly_Emi:** Primary regression target (Continuous variable representing **maximum safe monthly EMI amount (500-50000 INR)**)



Aspect	Observation	Insight/Interpretation
Variable Type	Numerical (continuous)	Maximum EMI repayment capacity
Distribution Shape	Single-peaked near lower EMI amounts, Long tail	Most customers have low EMI repayment capacity; a few have very high
Skew & Kurtosis	1.98 & 5.36	Strong right-skew & Leptokurtic
Outliers	Large number of high extreme values visible	Clipped based on domain limit (500-50000). Skew = 1.89
Relationship with Max_Monthly_Emi (Regression target)	Not Applicable	None
Relationship with Emi_Eligibility (Classification target)	- Eligible > High_Risk > Not Eligible (clear separation in medians) - Statistical Test - Kruskal $p = 0.00$ (Significant)	- Higher EMI capability → greater chance of approval - EMI capacity differs significantly across eligibility classes
Modeling Impact	Strong predictive power for classification target	Key variable — should be prioritized

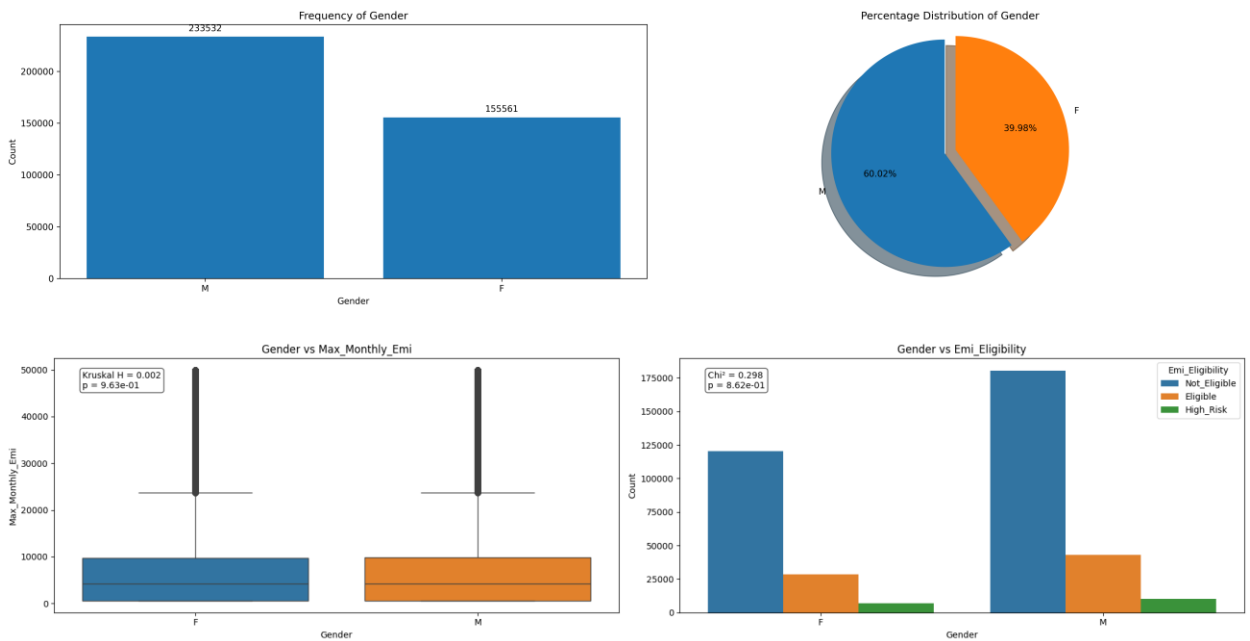
Personal Demographics:

Age:



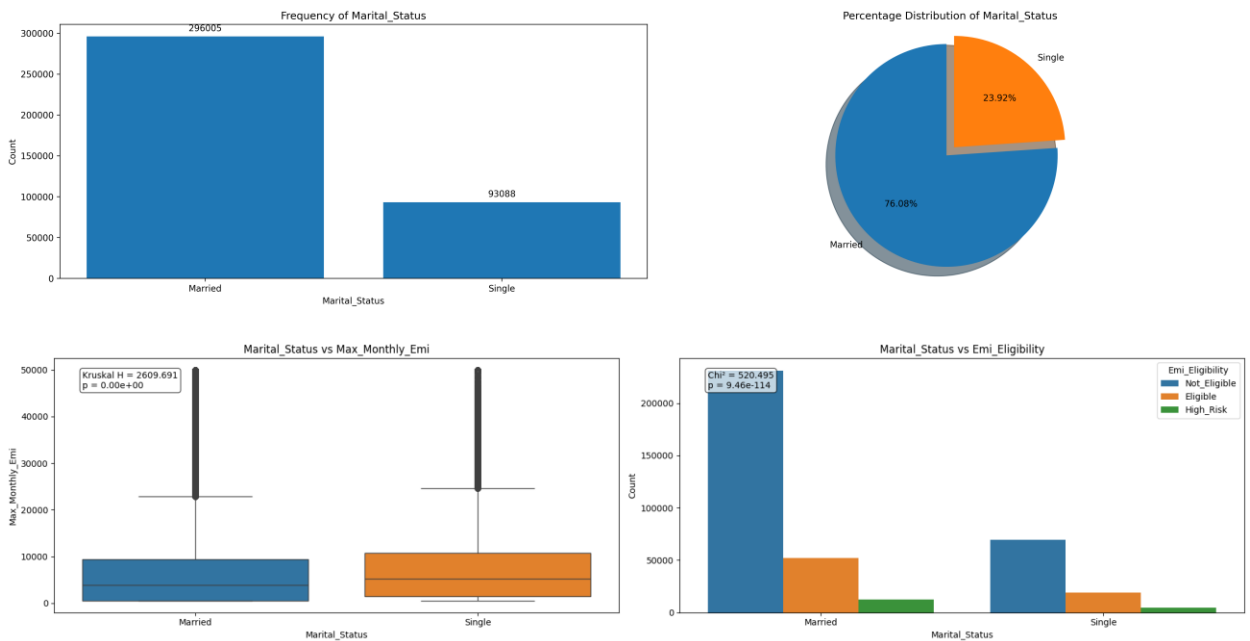
Aspect	Observation	Insight/Interpretation
Variable Type	Numerical (continuous)	Customer Age
Distribution Shape	Multi-modal spikes at specific ages; Slight right skew	Age appears generated in banded groups, not smooth continuous distribution
Skew & Kurtosis	0.65 & -0.55	Moderate right skew & Platykurtic
Outliers	No outliers	Age values mostly between 26 to 59, consistent with working population – no capping
Relationship with Max_Monthly_Emi (Regression target)	- Points widely scattered across all ages - Statistical Test - Spearman correlation ≈ -0.019 (very weak)	- EMI capacity varies independently of age - Age has negligible effect on EMI repayment capacity; no clear increasing/decreasing trend
Relationship with Emi_Eligibility (Classification target)	- Boxplot medians nearly identical across groups - Statistical Test - Kruskal $p \approx 0.35$ (Insignificant)	- EMI approval does not significantly differ by age
Modeling Impact	Low predictive importance expected	Safe to keep but may not lift accuracy

Gender:



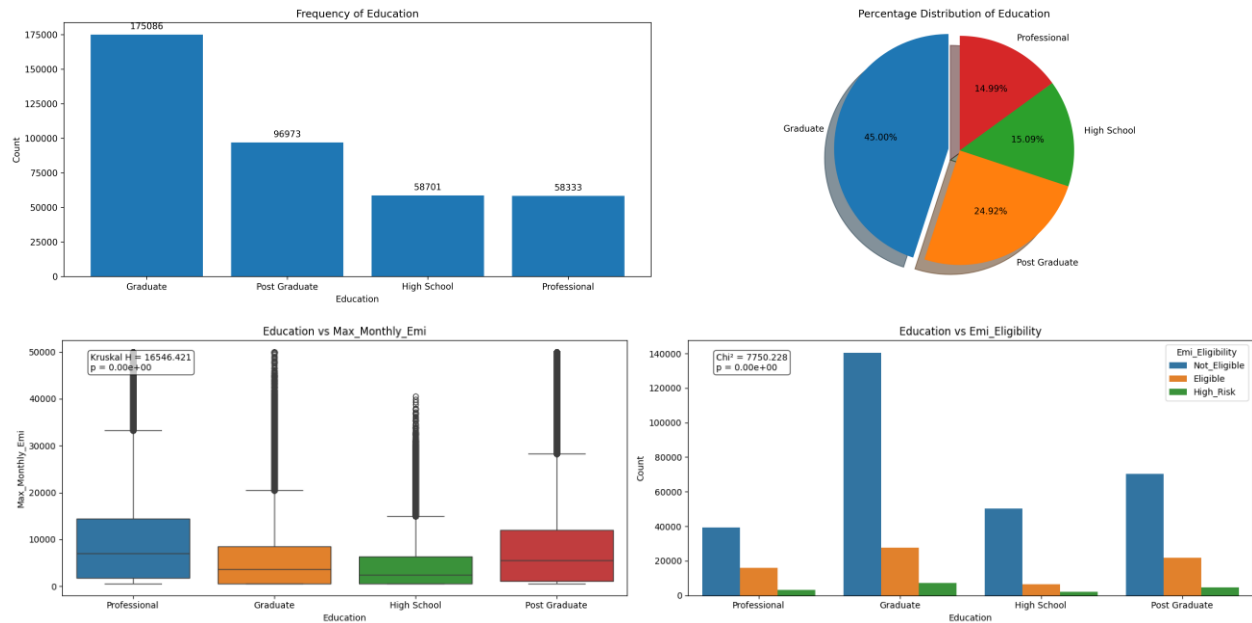
Aspect	Observation	Insight/Interpretation
Variable Type	Categorical (2 classes)	Demographic feature
Distribution Shape	Male ≈ 60% Female ≈ 40%	Slight male dominance — common in lending datasets due to more male applicants
Skew & Kurtosis	Not applicable (categorical)	None
Outliers	Not applicable (categorical)	None
Relationship with Max_Monthly_Emi (Regression target)	- Boxplots for Male & Female nearly identical - Statistical Test - Kruskal p ≈ 0.96 (Insignificant)	- EMI repayment capacity appears independent of gender - No statistically meaningful difference
Relationship with Emi_Eligibility (Classification target)	- Eligibility distribution similar across genders - Statistical Test - Chi-Square p ≈ 0.86 (Insignificant)	- Approval behavior appears similar for Male and Female - No meaningful association between gender and EMI eligibility
Modeling Impact	Likely low predictive power	Safe to keep, but not a driving feature (Binary encoding)

Marital Status:



Aspect	Observation	Insight/Interpretation
Variable Type	Categorical (2 classes)	Demographic feature
Distribution Shape	Married $\approx 76\%$ Single $\approx 24\%$	Majority of customers are married — expected in adult working population
Skew & Kurtosis	Not applicable (categorical)	None
Outliers	Not applicable (categorical)	None
Relationship with Max_Monthly_Emi (Regression target)	- Boxplots show higher median EMI capacity for married customers - Statistical Test - Kruskal $p = 0.00$ (Significant)	- Married individuals appear to have higher repayment capability , possibly due to dual-income or higher financial stability - EMI capacity differs meaningfully between married and single customers
Relationship with Emi_Eligibility (Classification target)	- Married group has more eligible approvals in count plot - Statistical Test - Chi-Square $p \approx 0.00$ (Significant)	- Married individuals tend to be more eligible for EMI approvals - Marital status has statistical association with eligibility classification
Modeling Impact	Potentially useful categorical predictor	Binary encoding sufficient

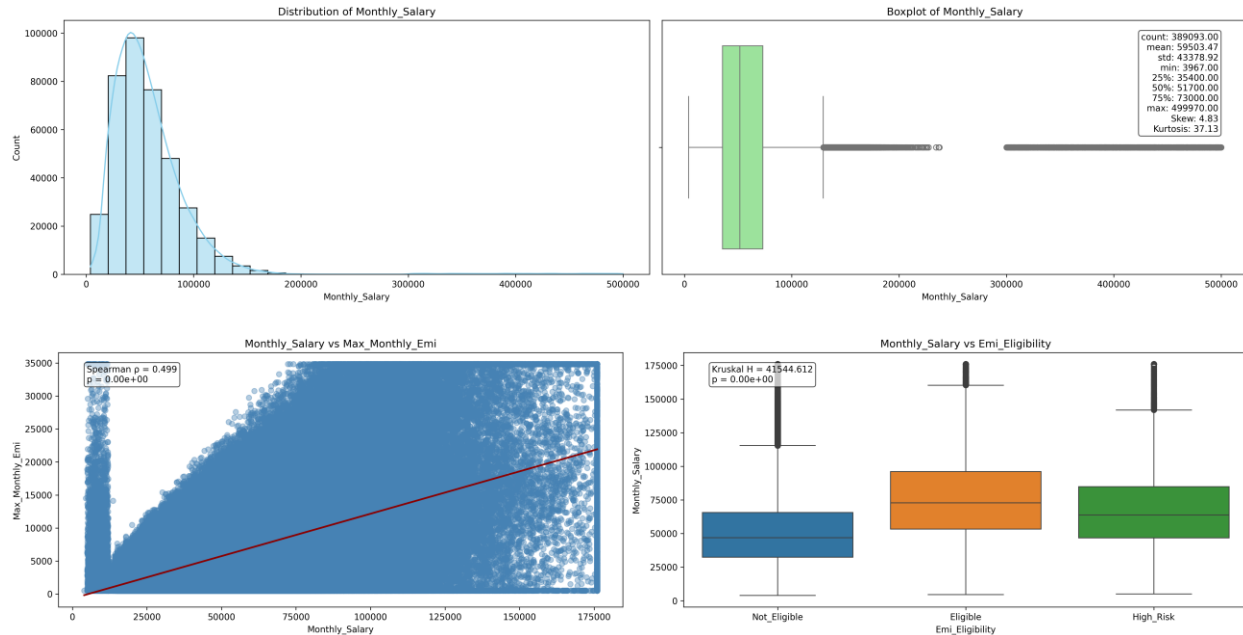
Education:



Aspect	Observation	Insight/Interpretation
Variable Type	Categorical (4 classes)	Education qualification level
Distribution Shape	Graduate ≈ 45% (largest) Post Graduate ≈ 25% High School ≈ 15% Professional ≈ 15%	Higher education levels dominate — with graduate category being most common among EMI applicants
Skew & Kurtosis	Not applicable (categorical)	None
Outliers	Not applicable (categorical)	None
Relationship with Max_Monthly_Emi (Regression target)	- Higher education → higher EMI capacity - Statistical Test - Kruskal p = 0.00 (Significant)	- More educated customers likely have better income and paying capacity - EMI capacity differs significantly across education levels
Relationship with Emi_Eligibility (Classification target)	- Higher education categories show higher eligible counts - Statistical Test - Chi-Square p = 0.00 (Significant)	- Education level positively influences EMI approval likelihood - Education has strong association with EMI eligibility
Modeling Impact	Strong predictor — keep	One-hot encoding recommended

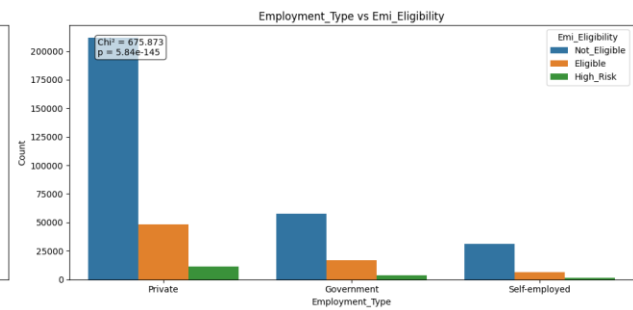
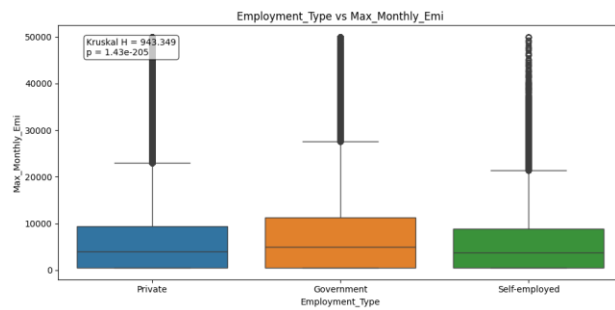
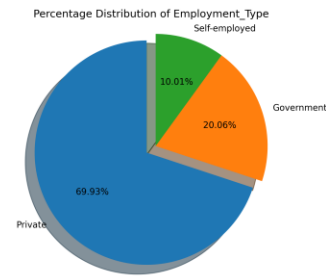
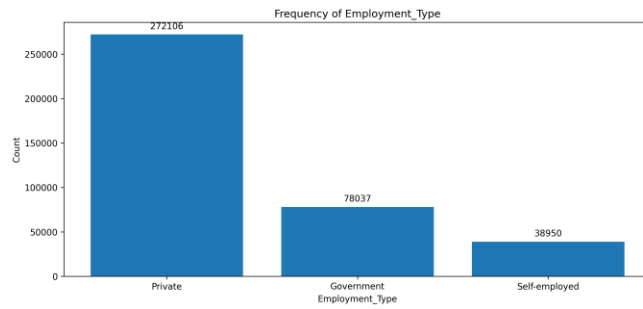
Employment and Income:

Monthly Salary:



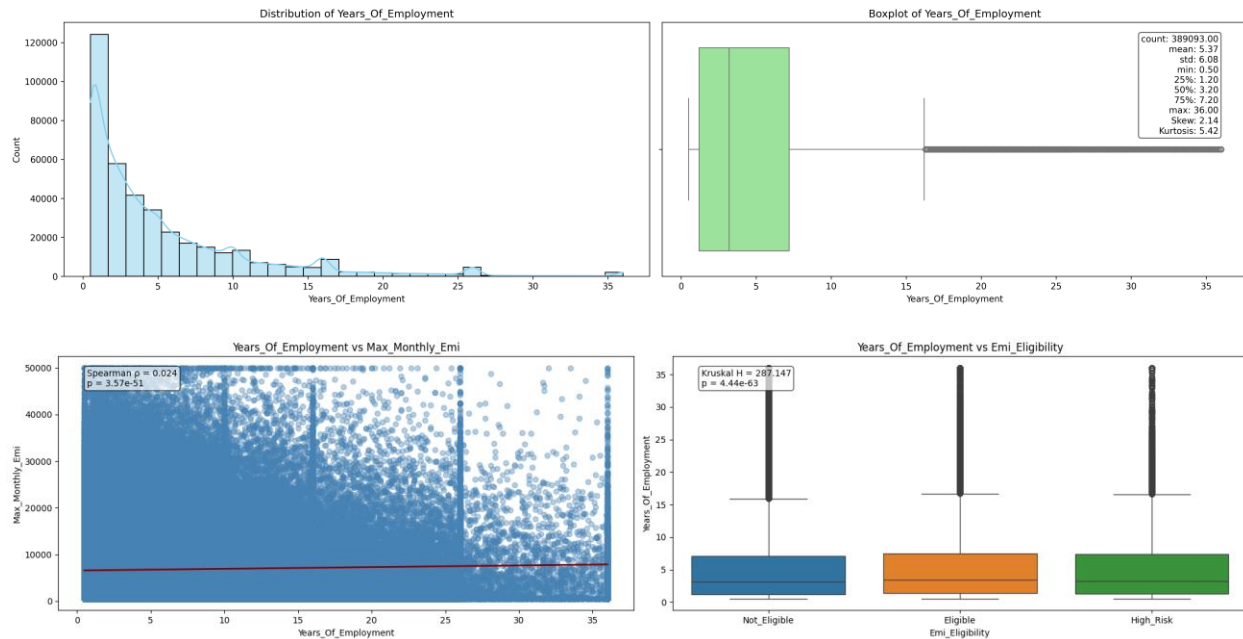
Aspect	Observation	Insight/Interpretation
Variable Type	Numerical (continuous)	Education qualification level
Distribution Shape	Right-skewed long tail; Most values clustered between ~20K–80K	Majority of applicants earn moderate salaries; a few high-income customers extend tail
Skew & Kurtosis	4.83 & 37.13	Extremely right-skewed & Ultra-leptokurtic
Outliers	Visible extreme highs	- Represents high-earning segment capped at 99 th percentile - Skew = 1.20
Relationship with Max_Monthly_Emi (Regression target)	- Positive upward trend - Statistical Test - Spearman ≈ 0.499 (Significant)	- Higher salary \rightarrow higher EMI capacity - Moderately strong correlation
Relationship with Emi_Eligibility (Classification target)	- Eligible group has noticeably higher median salaries; Not Eligible group lower - Statistical Test – Kruskal $p = 0.00$ (Significant)	- Salary strongly influences eligibility — higher earners more likely approved - Salary differences across groups are statistically meaningful
Modeling Impact	Strong predictor — keep	Salary is a major factor in credit risk & EMI affordability modeling

Employment Type:



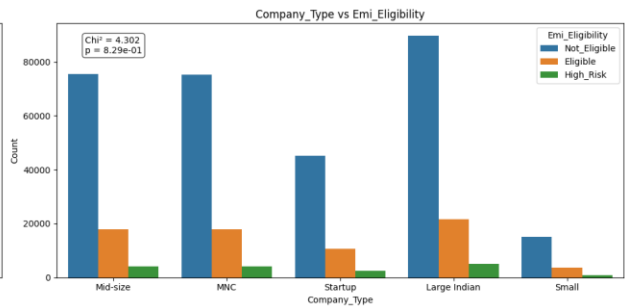
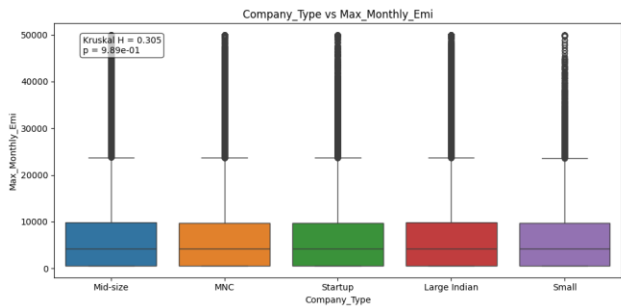
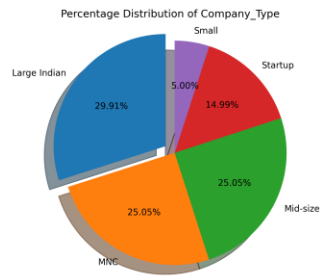
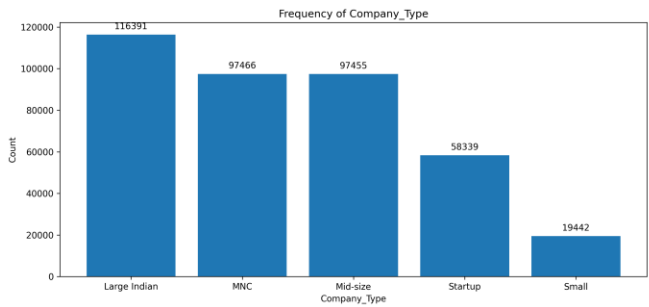
Aspect	Observation	Insight/Interpretation
Variable Type	Categorical (3 classes)	Employment classification
Distribution Shape	Private ≈ 70% (majority) Government ≈ 20% Self-employed ≈ 10%	Loan applicants are predominantly private-sector employees
Skew & Kurtosis	Not applicable (categorical)	None
Outliers	Not applicable (categorical)	None
Relationship with Max_Monthly_Emi (Regression target)	- Government > Private > Self-employed - Statistical Test - Kruskal p ≈ 0.00 (Significant)	- Government workers have higher and more stable EMI affordability self-employed lowest capacity - Employment type significantly impacts EMI capacity
Relationship with Emi_Eligibility (Classification target)	- Government > Private > Self-employed in approval share - Statistical Test - Chi-Square p ≈ 0.00 (Significant)	- Higher approval likelihood for government employees ; self-employed face stricter approval - Employment type significantly impacts EMI eligibility
Modeling Impact	Strong predictor — keep	One-hot encoding recommended

Years Of Employment:



Aspect	Observation	Insight/Interpretation
Variable Type	Numerical (continuous)	Represents years of work experience
Distribution Shape	Highly right-skewed; Most values between 0–7 years; few with >20 years	Majority of applicants are early-career, Small segment with long experience (tail)
Skew & Kurtosis	2.14 & 5.42	Very strong right-skew & Highly leptokurtic
Outliers	High-end experience values visible	- Represents senior professionals (domain-valid) no capping - Skew = 2.14
Relationship with Max_Monthly_Emi (Regression target)	- Very weak positive pattern - Statistical Test – Spearman \approx 0.024 (Significant)	- Experience slightly increases EMI capacity but effect is minimal
Relationship with Emi_Eligibility (Classification target)	- Eligible group has slightly higher median experience - Statistical Test - Kruskal $p \approx$ 0.00 (Significant)	- More experienced applicants marginally more likely to get approved - Differences exist statistically but real-world impact small
Modeling Impact	Weak predictor — keep but low weight expected	Experience may support other income-related variables rather than drive eligibility alone

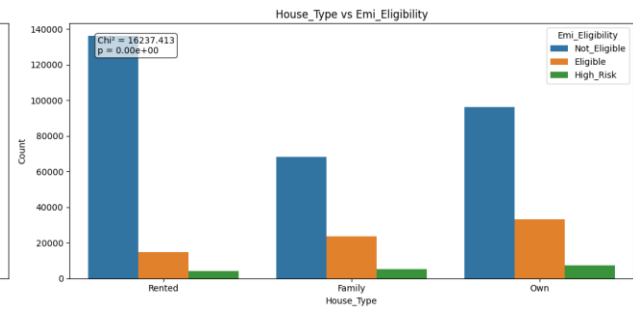
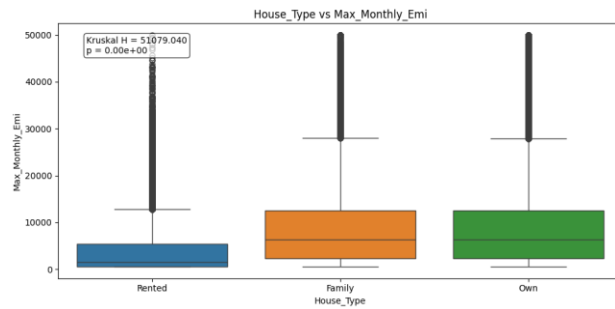
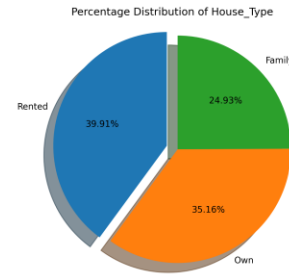
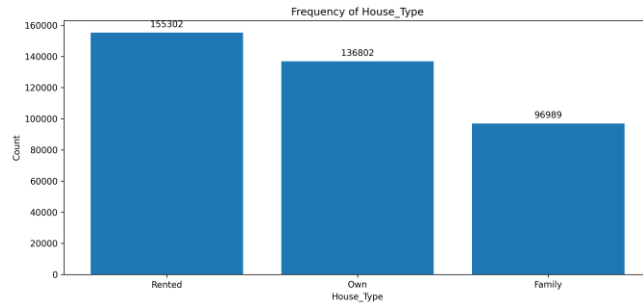
Company Type:



Aspect	Observation	Insight/Interpretation
Variable Type	Categorical (5 classes)	Company classification
Distribution Shape	Large Indian ~30% (highest) MNC & Mid-size ~25% each Startup ~15% Small ~5%	Majority employed in established firms Small startups least represented
Skew & Kurtosis	Not applicable (categorical)	None
Outliers	Not applicable (categorical)	None
Relationship with Max_Monthly_Emi (Regression target)	- Boxplots nearly identical across company types - Statistical Test - Kruskal $p \approx 0.99$ (Insignificant)	- Employer type does not materially affect EMI capacity - No statistically significant relationship with EMI capacity
Relationship with Emi_Eligibility (Classification target)	- Eligibility distribution visually similar across categories - Statistical Test - Chi-Square $p \approx 0.83$ (Insignificant)	- Approval rates do not vary meaningfully by employer type - No statistically significant relationship with EMI eligibility
Modeling Impact	Weak predictor — low value	Low priority for feature importance

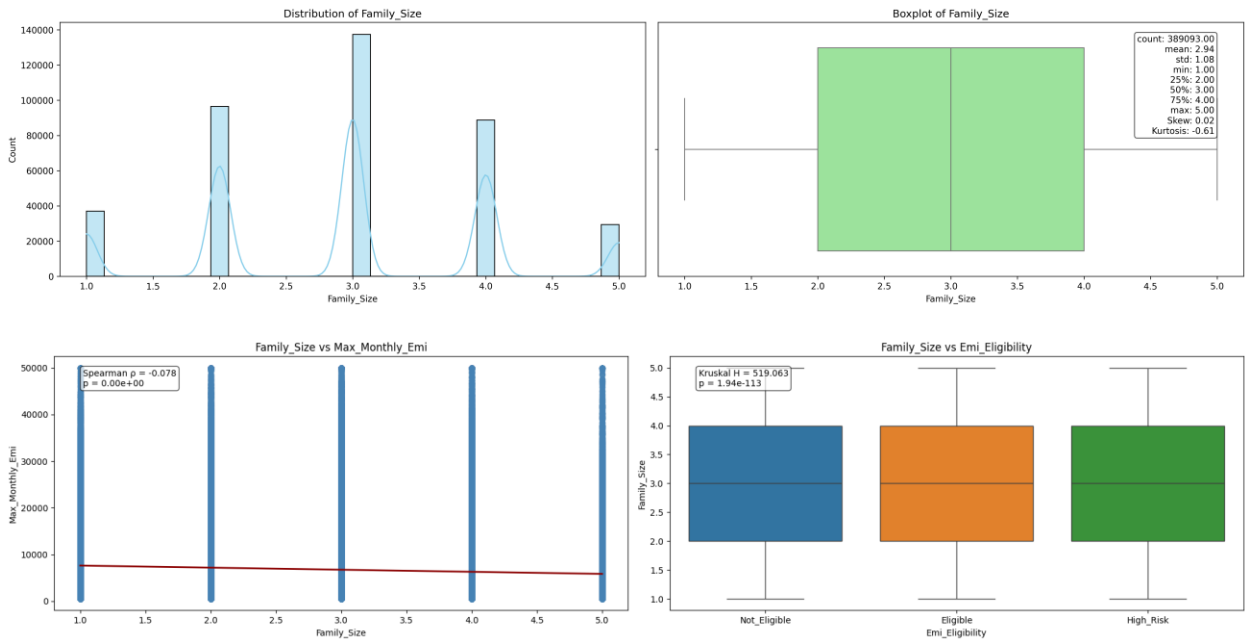
Housing and Family:

House Type:



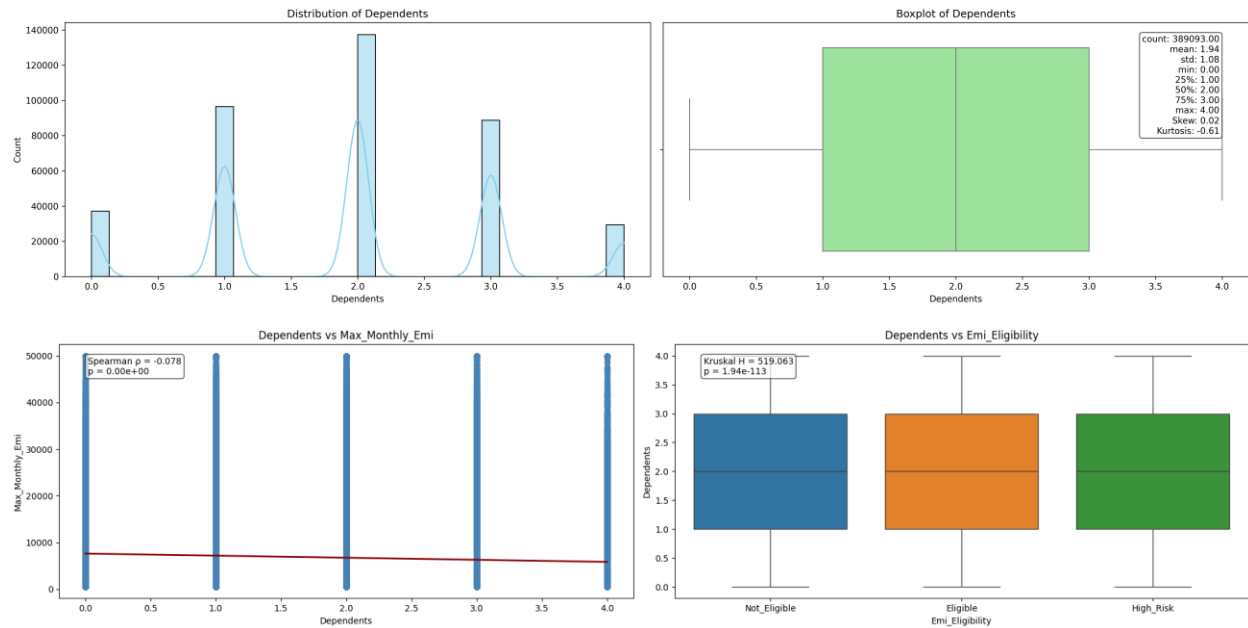
Aspect	Observation	Insight/Interpretation
Variable Type	Categorical (3 classes)	Housing situation indicator
Distribution Shape	Rented ≈ 40% (highest) Own ≈ 35% Family ≈ 25%	Majority live in rented or owned homes; smaller share lives with family
Skew & Kurtosis	Not applicable (categorical)	None
Outliers	Not applicable (categorical)	None
Relationship with Max_Monthly_Emi (Regression target)	- Own > Family > Rented (clear increasing pattern) - Statistical Test - Kruskal p = 0.00 (Significant)	- Owning a home indicates higher financial stability , ability to take larger EMI - House type significantly impacts EMI capacity
Relationship with Emi_Eligibility (Classification target)	- Higher approvals for Own homes; lowest for Rented - Statistical Test - Chi-Square p = 0.00 (Significant)	- Rented customers more likely to be rejected - House type significantly impacts EMI approval likelihood
Modeling Impact	Strong predictive feature— keep	One-hot encoding recommended

Family Size:



Aspect	Observation	Insight/Interpretation
Variable Type	Numerical (discrete)	Number of family members
Distribution Shape	Most common: $3 > 2 > 4 > 1 > 5$ Slight left skew	Majority of applicants belong to medium-sized families (2–4 members)
Skew & Kurtosis	0.02 & -0.61	Approximately symmetric & Platykurtic
Outliers	None — bounded realistic values (1–5)	All values are domain-valid
Relationship with Max_Monthly_Emi (Regression target)	- Very weak negative correlation - Statistical Test - Spearman $\rho \approx -0.078$ (Significant)	- Larger family size slightly reduces EMI capacity, but effect is minimal - Worth noting statistically
Relationship with Emi_Eligibility (Classification target)	- Eligible group shows slightly fewer dependents on average - Statistical Test - Kruskal $p = 0.00$ (Significant)	- Smaller families may manage expenses better, improving approval odds
Modeling Impact	Low to moderate importance	Useful as secondary socio-economic indicator

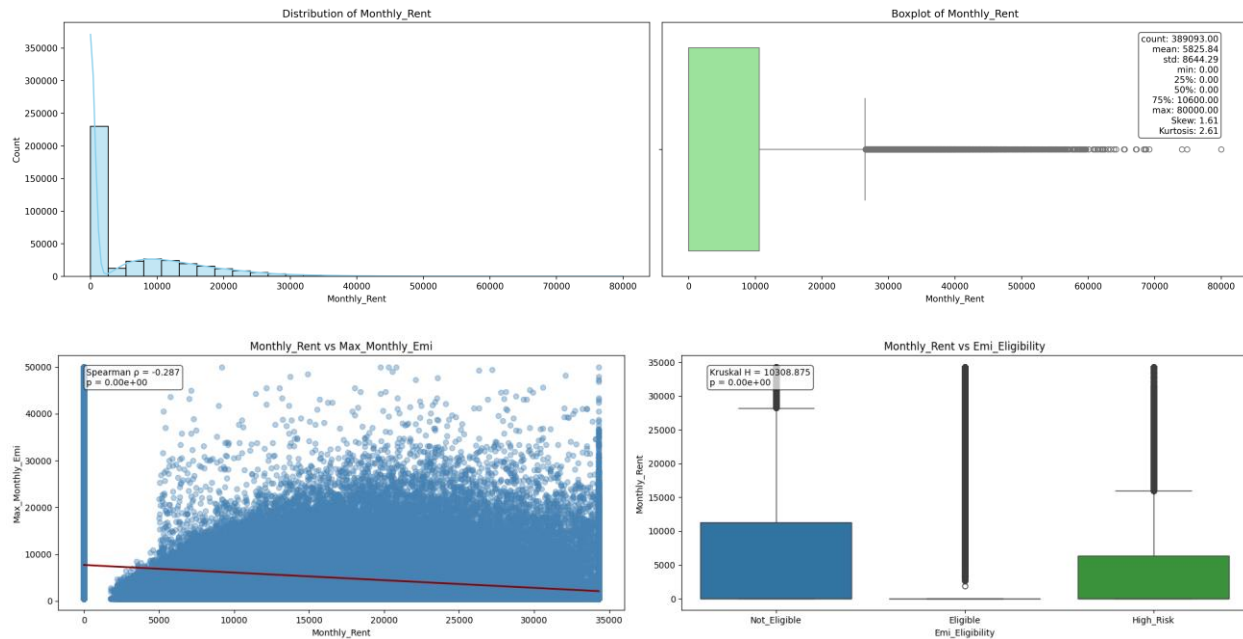
Dependents:



Aspect	Observation	Insight/Interpretation
Variable Type	Numerical (discrete)	Number of financial dependents
Distribution Shape	Most common: $2 > 1 > 3 > 0 > 4$ Slight left skew	Majority have 1–3 dependents; very few extreme dependent counts
Skew & Kurtosis	0.02 & -0.61	Approximately symmetric & Platykurtic
Outliers	None — bounded realistic values	Values reflect typical family structures
Relationship with Max_Monthly_Emi (Regression target)	- Very weak negative correlation - Statistical Test - Spearman $\rho \approx -0.078$ (Significant)	- More dependents slightly reduce EMI capacity due to higher living expenses; effect very small - Differences exist statistically, but real-world effect is weak
Relationship with Emi_Eligibility (Classification target)	- Eligible group has slightly lower family size than Not Eligible group - Statistical Test - Kruskal $p = 0.00$ (Significant)	- Fewer dependents may improve disposable income → better approval chances
Modeling Impact	Low importance , supportive variable	Helps capture financial burden context; not a primary driver

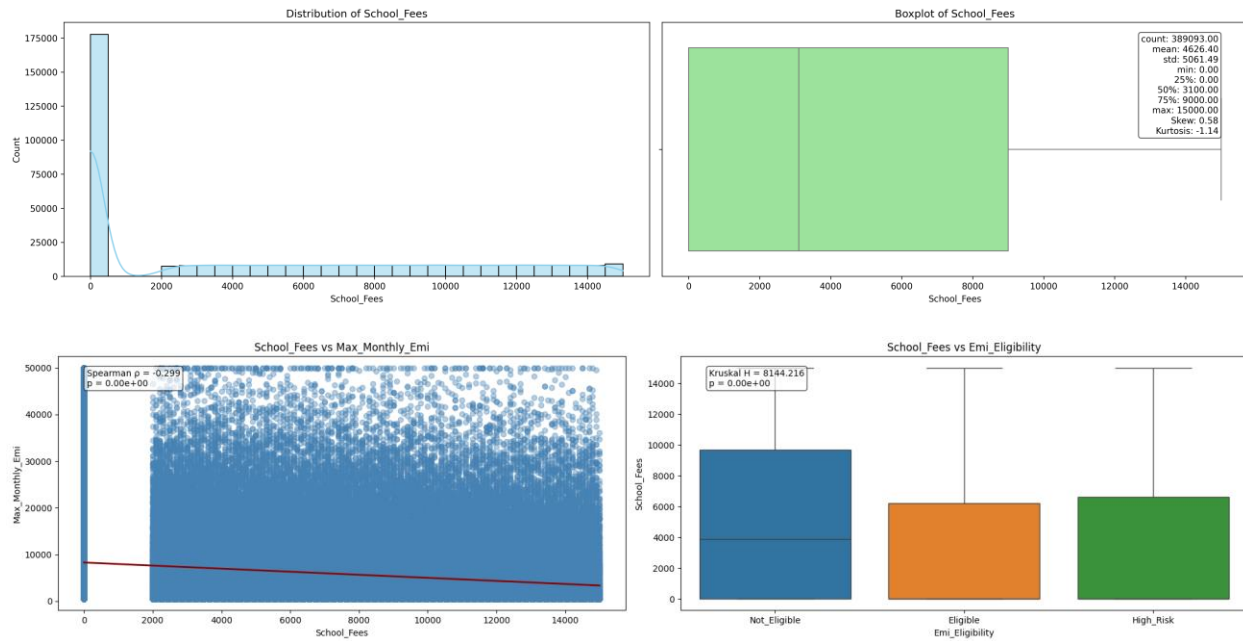
Monthly Financial Obligations:

Monthly Rent:



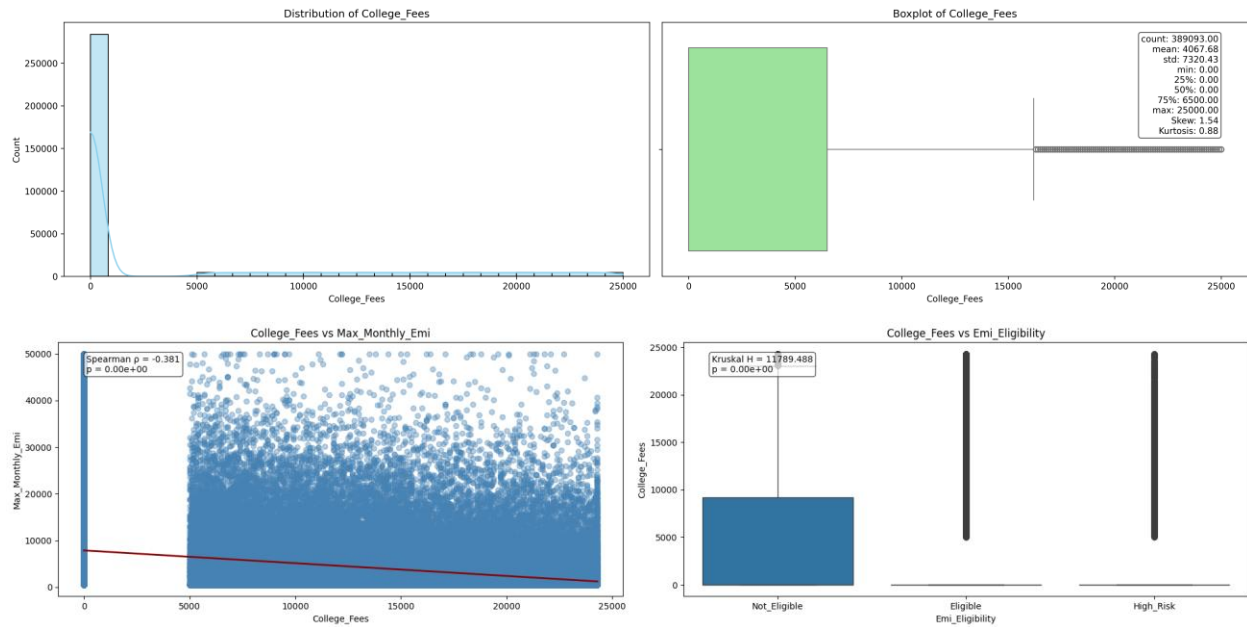
Aspect	Observation	Insight/Interpretation
Variable Type	Numerical (continuous)	Monthly rental expense
Distribution Shape	Highly right-skewed , majority at 0 rent (living in own/family homes); long tail up to ~80,000	Most applicants do not pay rent , renters form minority with wide cost variation
Skew & Kurtosis	1.61 & 2.61	Strong right skew & Leptokurtic
Outliers	Very high rent values present; long upper tail	- Likely high-income urban segments, capped at 99th percentile - Skew = 1.40
Relationship with Max_Monthly_Emi (Regression target)	- Moderate negative trend - Statistical Test - Spearman ≈ -0.287 (Significant)	- Higher rent \rightarrow lower EMI capacity due to reduced disposable income - Rent significantly affects EMI affordability
Relationship with Emi_Eligibility (Classification target)	- Not Eligible group shows higher median rent - Statistical Test - Kruskal $p = 0.00$ (Significant)	- Renting increases fixed monthly burden \rightarrow lowers EMI approval likelihood - Rent significantly affects EMI eligibility
Modeling Impact	Strong expense-related predictor — keep	Captures discretionary income limitation; key affordability factor

School Fees:



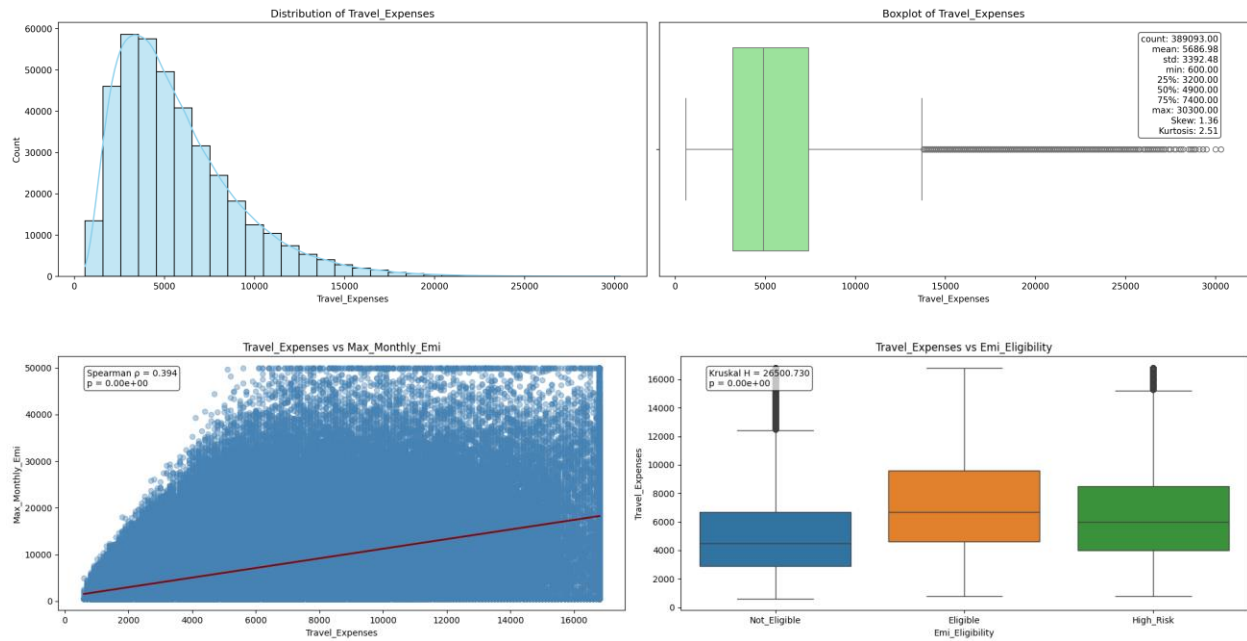
Aspect	Observation	Insight/Interpretation
Variable Type	Numerical (continuous)	Monthly school fee expenditure
Distribution Shape	Right-skewed ; majority near zero; long tail up to ~15,000	Most applicants have zero or low school fee burden
Skew & Kurtosis	0.58 & -1.14	Mild right-skew & Strongly platykurtic
Outliers	Present – but bounded realistic values	- High-fee cases represent wealthier families or more children – no capping - Skew = 0.58
Relationship with Max_Monthly_Emi (Regression target)	- Moderate negative trend - Statistical Test - Spearman ≈ -0.299 (Significant)	- Higher school fees reduce EMI capacity — indicates higher fixed obligations - School fees significantly affects EMI affordability
Relationship with Emi_Eligibility (Classification target)	- Not Eligible group has higher school fees than Eligible group - Statistical Test - Kruskal $p = 0.00$ (Significant)	- Higher child education expenses reduce loan approval likelihood - School fees significantly affects EMI eligibility
Modeling Impact	Strong expense-related predictor — keep	Important indicator of household burden and net disposable income

College Fees:



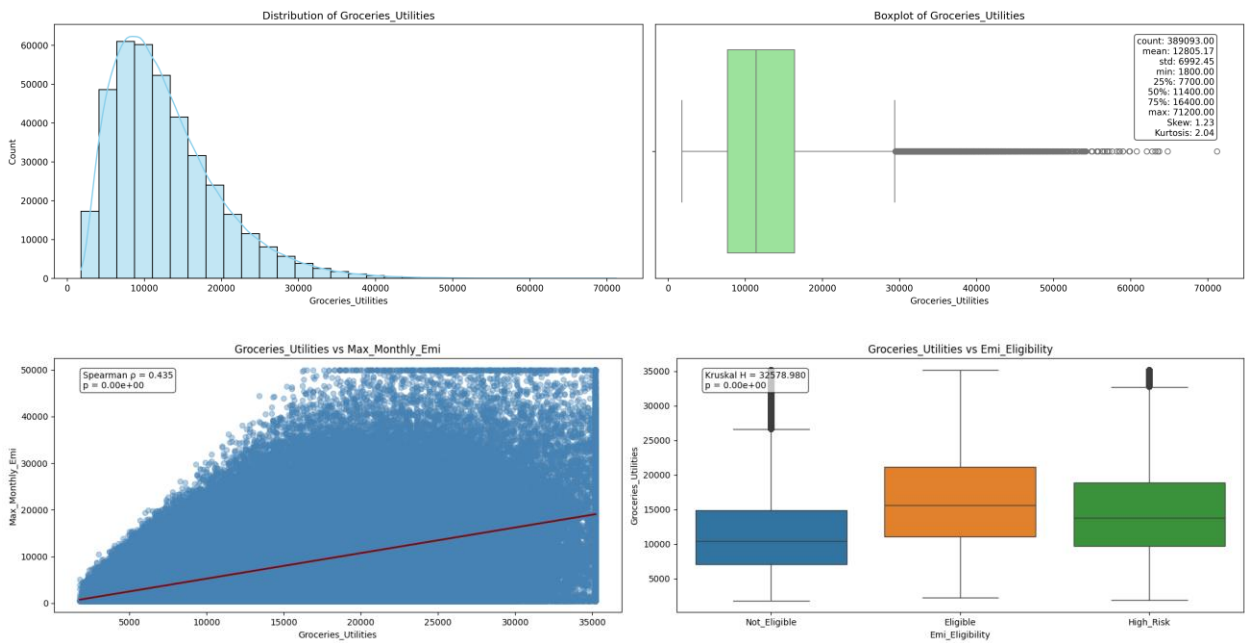
Aspect	Observation	Insight/Interpretation
Variable Type	Numerical (continuous)	Monthly college-related educational expense
Distribution Shape	Highly right-skewed ; majority at 0 , long tail up to ~₹25,000	Most applicants likely have no college-going dependents
Skew & Kurtosis	1.54 & 0.88	Right skew & Platykurtic
Outliers	Present	<ul style="list-style-type: none"> - High-fee cases reflect households with high financial load; capped at 99th percentile - Skew = 1.53
Relationship with Max_Monthly_Emi (Regression target)	<ul style="list-style-type: none"> - Moderate negative correlation - Statistical Test - Spearman ≈ -0.381 (Significant) 	<ul style="list-style-type: none"> - Higher college fees → lower EMI affordability due to reduced disposable income - Spending on higher education significantly impacts EMI capacity
Relationship with Emi_Eligibility (Classification target)	<ul style="list-style-type: none"> - Not Eligible group shows higher median college fees - Statistical Test - Kruskal $p = 0.00$ (Significant) 	<ul style="list-style-type: none"> - Higher education expenses increase default risk perception → lower approval probability - Spending on higher education significantly impacts loan approval
Modeling Impact	Strong expense-related predictor — keep	Helps model understand financial stress and limited repayment bandwidth

Travel Expenses:



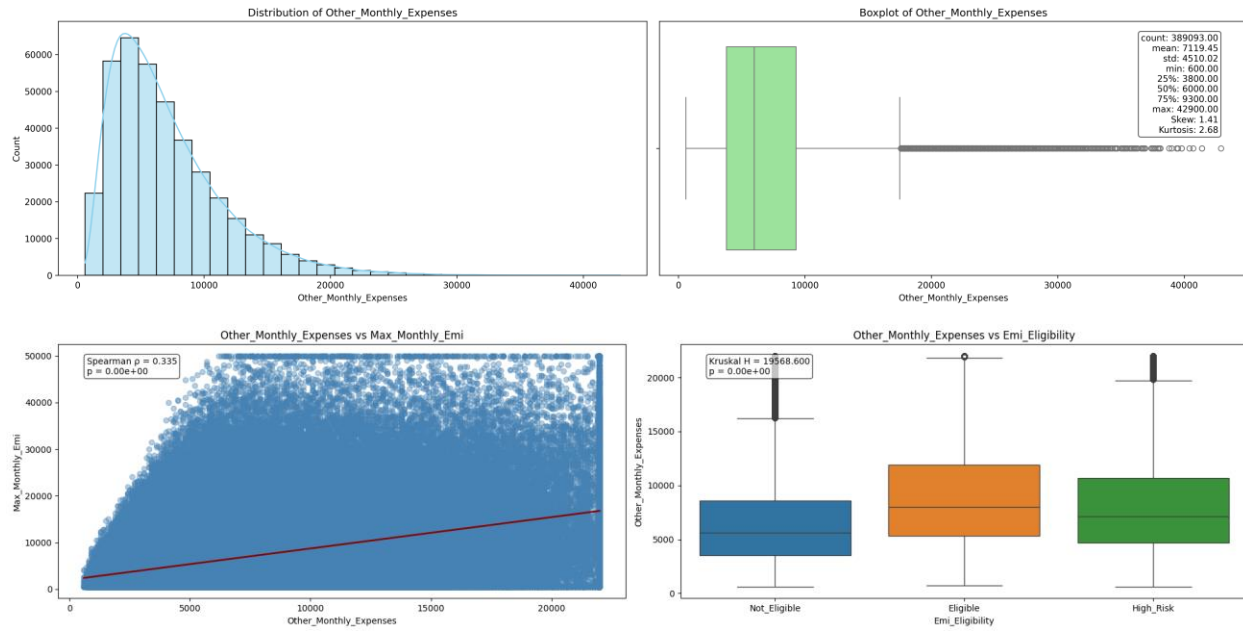
Aspect	Observation	Insight/Interpretation
Variable Type	Numerical (continuous)	Monthly travel or commute expenditure
Distribution Shape	Right-skewed , long tail; common values around ₹3K–₹7K	Most applicants have moderate commute expenses.
Skew & Kurtosis	1.36 & 2.51	Right-skewed & Leptokurtic
Outliers	Present	- Represents high-cost travel profiles; capped at 99th percentile - Skew = 1.13
Relationship with Max_Monthly_Emi (Regression target)	- Moderate positive correlation - Statistical Test - Spearman ≈ 0.394 (Significant)	- Higher travel expenses → indicates higher income / lifestyle → higher EMI capability - travel expenses significantly impacts EMI capacity
Relationship with Emi_Eligibility (Classification target)	- Eligible group shows higher travel expenses than others - Statistical Test - Kruskal $p \approx 0.00$ (Significant)	- Travel expense levels differ significantly across EMI eligibility classes - travel expenses significantly impacts loan approval
Modeling Impact	Useful feature — keep	Helps capture lifestyle affordability signals

Groceries Utilities:



Aspect	Observation	Insight/Interpretation
Variable Type	Numerical (continuous)	Monthly groceries expense
Distribution Shape	Right-skewed, long tail; most values \approx ₹6K–₹15K	Majority spend at moderate household levels; a few have very high living expenses
Skew & Kurtosis	1.23 & 2.04	Right-skewed & Leptokurtic
Outliers	Present	High-spending households — likely high-income families; capped at 99 th percentile - Skew = 1.02
Relationship with Max_Monthly_Emi (Regression target)	- Moderate positive correlation - Statistical Test - Spearman \approx 0.435 (Significant)	- Higher household spending \rightarrow higher disposable income \rightarrow stronger EMI capability - Spending on higher education significantly impacts EMI capacity
Relationship with Emi_Eligibility (Classification target)	- Eligible group shows higher grocery & utility expenses than others - Statistical Test - Kruskal $p = 0.00$ (Significant)	- Higher spending capacity aligns with better credit approval likelihood - Spending patterns differ significantly by eligibility category
Modeling Impact	Useful feature — keep	Acts as a proxy for income & lifestyle affordability, helps model identify strong profiles

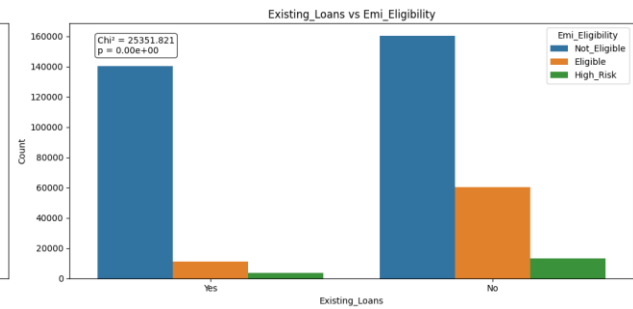
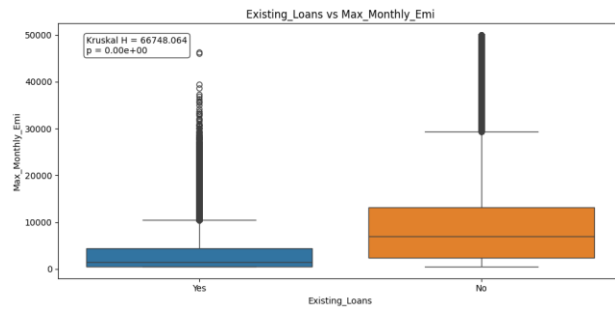
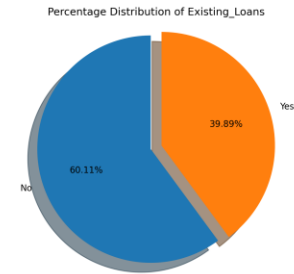
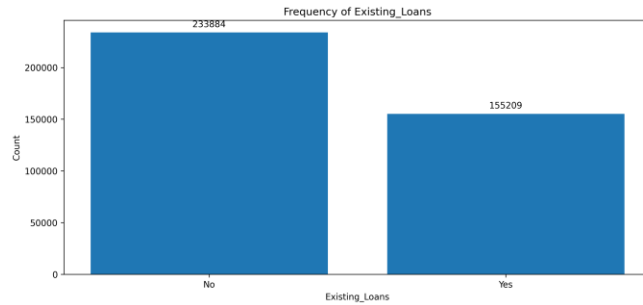
Other Monthly Expenses:



Aspect	Observation	Insight/Interpretation
Variable Type	Numerical (continuous)	Monthly miscellaneous living expenses
Distribution Shape	Right-skewed, long tail; majority values ~₹3K–₹12K	Most households spend moderate amounts; few have very high monthly miscellaneous expenses
Skew & Kurtosis	1.41 & 2.68	Right-skewed & Leptokurtic
Outliers	Present	High-expense households; outliers capped at 99th percentile - Skew = 1.17
Relationship with Max_Monthly_Emi (Regression target)	- Moderate positive correlation - Statistical test - Spearman ≈ 0.335 (Significant)	- Higher miscellaneous spending \rightarrow typically higher disposable income \rightarrow higher EMI capacity - Spending significantly impacts EMI capacity
Relationship with Emi_Eligibility (Classification target)	- Eligible group has higher median other expenses - Statistical Test - Kruskal $p = 0.00$ (Significant)	- People approved for EMI tend to spend more \rightarrow higher income/lifestyle affordability - Spending patterns vary significantly across eligibility classes
Modeling Impact	Useful feature — keep	Helps capture lifestyle spending behavior & financial capacity

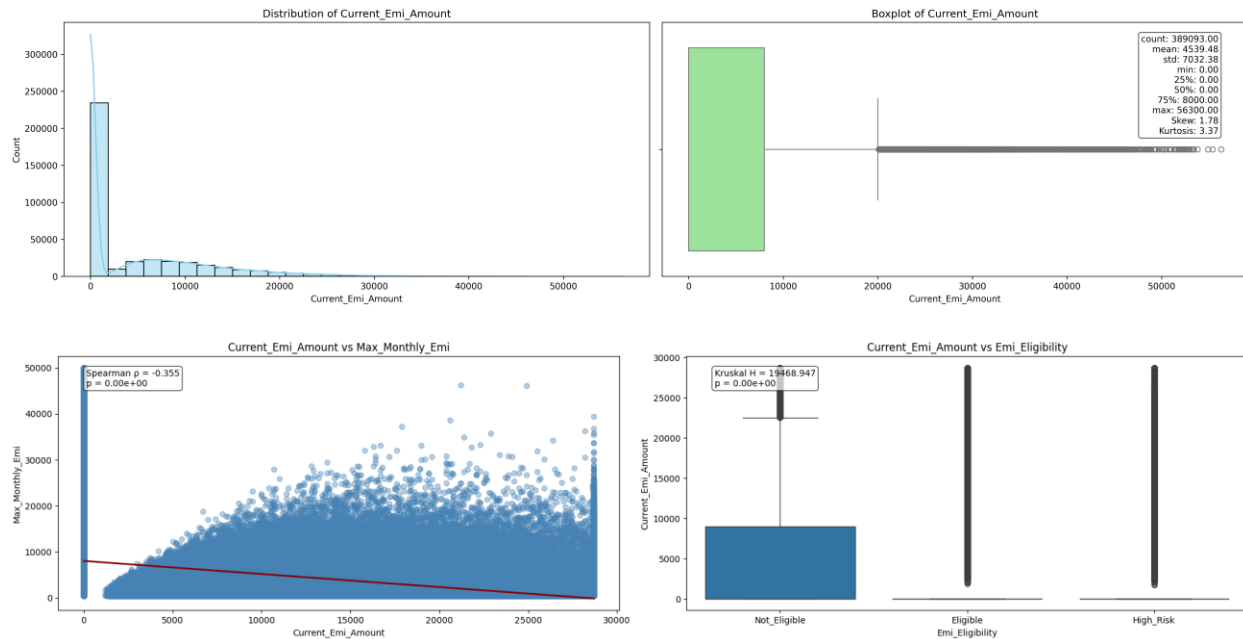
Financial Status and Credit History:

Existing Loans:



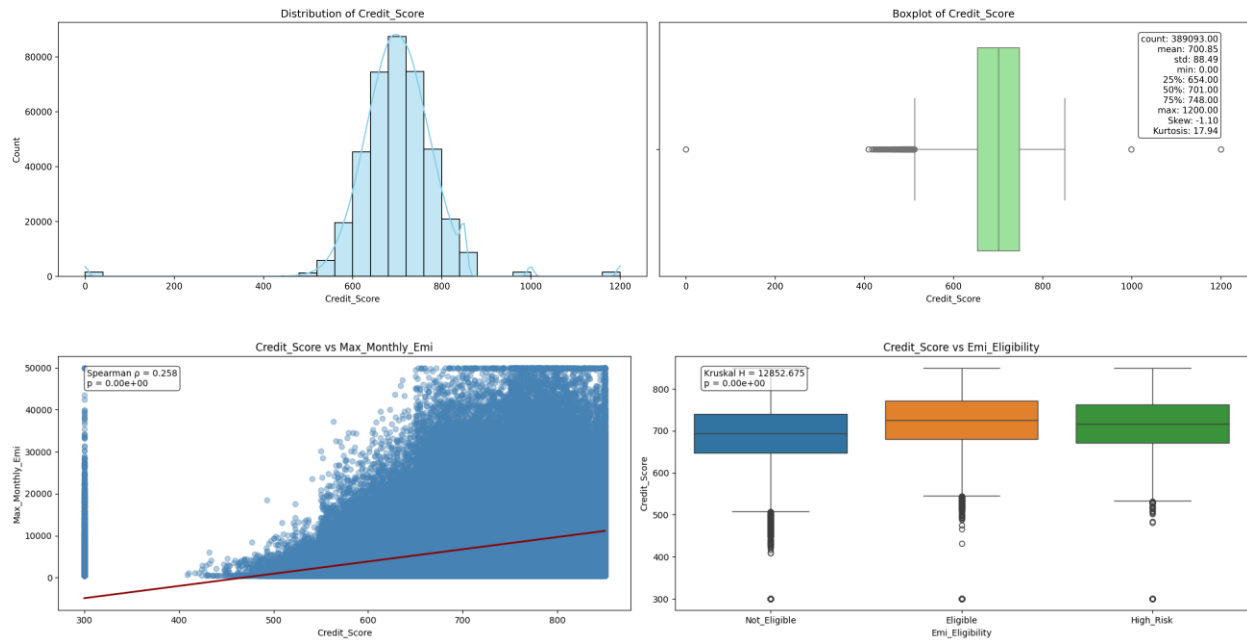
Aspect	Observation	Insight/Interpretation
Variable Type	Categorical (2 classes)	Indicates whether customer has active loans
Distribution Shape	No ≈ 60% Yes ≈ 40%	Majority applicants do not have existing loans
Skew & Kurtosis	Not applicable (categorical)	None
Outliers	Not applicable (categorical)	None
Relationship with Max_Monthly_Emi (Regression target)	- Customers with No existing loans have higher EMI capacity - Statistical Test - Kruskal p = 0.00 (Significant)	- No current loan burden → higher disposable income → better EMI affordability
Relationship with Emi_Eligibility (Classification target)	- Higher approval rates in No loans group - Statistical Test - Chi-Square p = 0.00 (Significant)	- Existing loans increase perceived credit risk → lower EMI eligibility
Modeling Impact	Strong predictor — keep	Important for credit risk assessment; signals debt burden & repayment potential

Current Emi Amount:



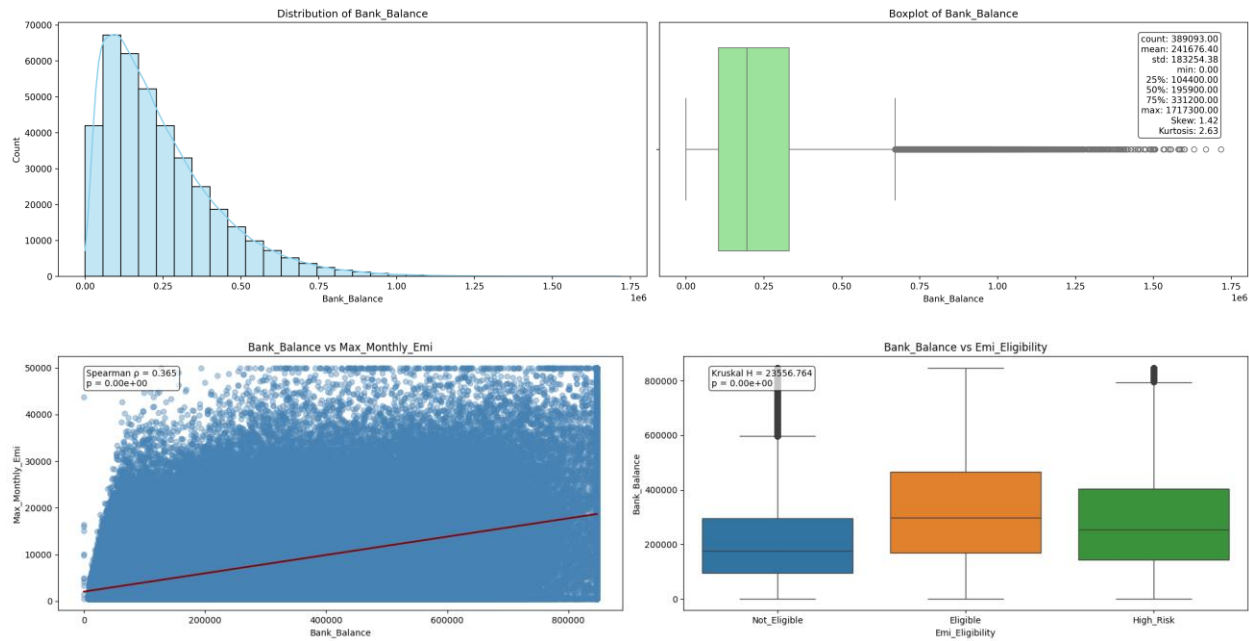
Aspect	Observation	Insight/Interpretation
Variable Type	Numerical (continuous)	Amount customer is currently paying toward other EMIs
Distribution Shape	Highly right-skewed , long tail; many at ₹0, gradual drop until ~₹50K	Most customers have no or low existing EMI load ; few have very high EMI burden
Skew & Kurtosis	1.78 & 3.37	Right-skewed & Leptokurtic
Outliers	Present	A minority with very high EMI payments ; capped at 99 th percentile - Skew = 1.55
Relationship with Max_Monthly_Emi (Regression target)	- Moderate negative correlation - Statistical Test - Spearman $\rho \approx -0.355$, $p = 0.00$ (Significant)	- Higher EMI burden reduces disposable income → lower future EMI affordability
Relationship with Emi_Eligibility (Classification target)	- Not-Eligible group has higher current EMI payments - Statistical Test - Kruskal $p = 0.00$ (Significant)	- Higher EMI commitments increase credit risk → lower approval probability
Modeling Impact	Strong risk indicator — keep	Represents existing financial obligations

Credit Score:



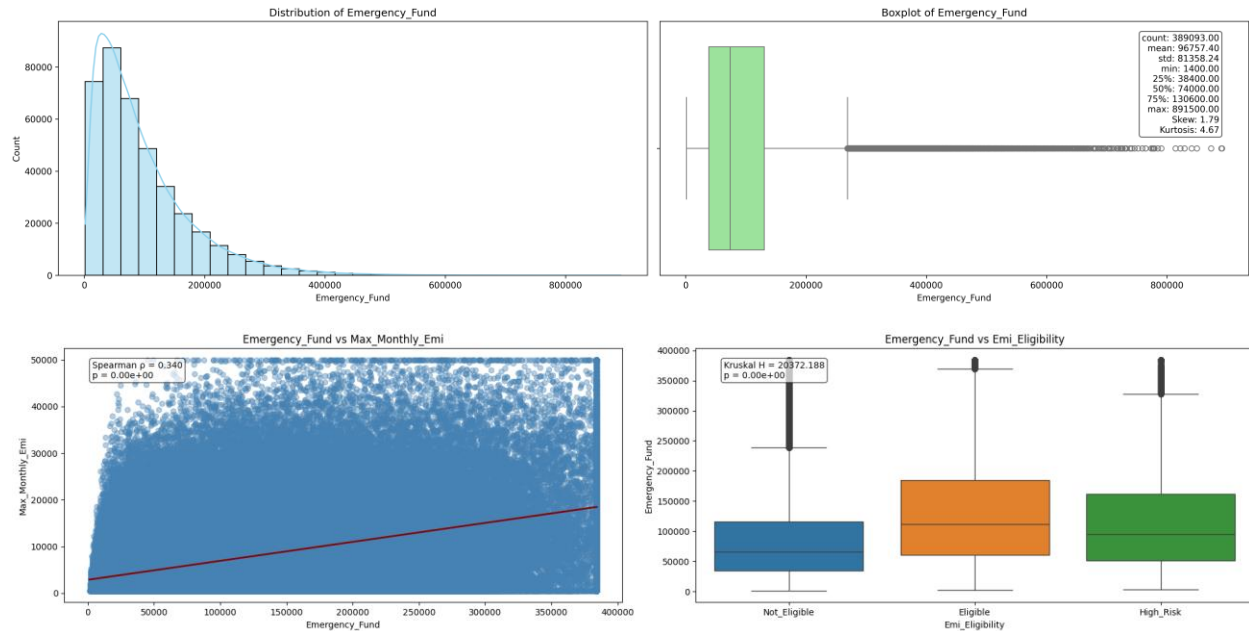
Aspect	Observation	Insight/Interpretation
Variable Type	Numerical (continuous)	Customer credit score
Distribution Shape	Roughly bell-shaped, centered around ~700 ; a few extreme low & very high values	Most customers have average to good credit scores; small group with poor scores
Skew & Kurtosis	-1.10 & 17.94	Slight Right-skewed & Leptokurtic
Outliers	Present	Few customers with extremely low or unusually high credit scores; capped at domain limit (300 – 850) - Skew = -0.60
Relationship with Max_Monthly_Emi (Regression target)	- Weak-to-moderate positive correlation - Statistical Test - Spearman ≈ 0.258 (Significant)	- Higher creditworthiness increases lending confidence → higher EMI affordability
Relationship with Emi_Eligibility (Classification target)	- Eligible group has higher credit scores - Statistical Test - Kruskal $p = 0.00$ (Significant)	- Higher credit scores improve EMI approval likelihood; poor scores → more Not Eligible
Modeling Impact	Strong risk signal — keep	Critical credit underwriting feature; helps identify reliable borrowers

Bank Balance:



Aspect	Observation	Insight/Interpretation
Variable Type	Numerical (continuous)	Customer's total bank balance
Distribution Shape	Right-skewed, long tail ; majority balances ₹0–₹3L	Most people have modest bank balances; few customers hold very high amounts
Skew & Kurtosis	1.42 & 2.63	Right-skewed & Leptokurtic
Outliers	Present	High-wealth customers significantly affect distribution tail; capped at 99th percentile - Skew = 1.18
Relationship with Max_Monthly_Emi (Regression target)	- Moderate positive correlation - Statistical Test - Spearman ≈ 0.365 (Significant)	- Stronger savings/balance indicates higher financial stability & repayment capability - Higher bank balance → higher Max EMI capacity
Relationship with Emi_Eligibility (Classification target)	- Eligible group has significantly higher bank balances - Statistical Test - Kruskal $p = 0.00$ (Significant)	- Customers with higher savings are more likely to get approved ; low-balance customers are more likely to be rejected
Modeling Impact	Strong predictor — keep	Represents financial security & liquidity ; key parameter for credit decisioning

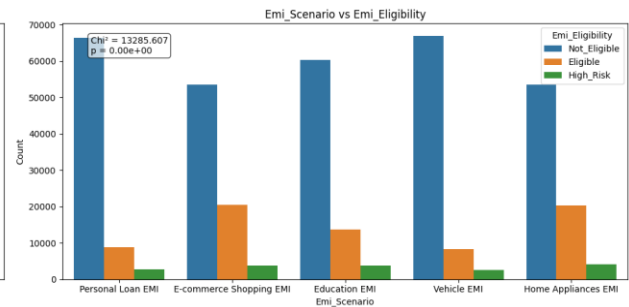
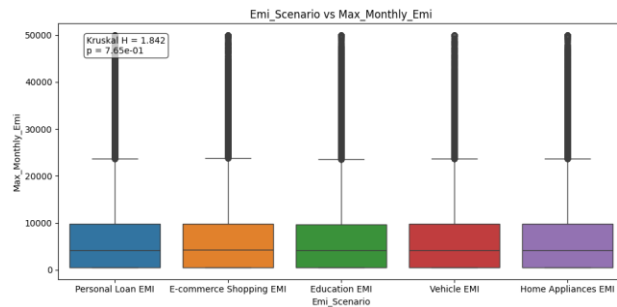
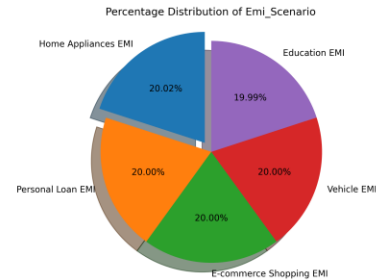
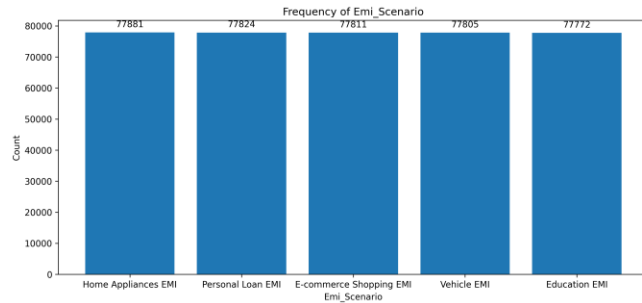
Emergency_Fund:



Aspect	Observation	Insight/Interpretation
Variable Type	Numerical (continuous)	Customer's dedicated emergency reserve
Distribution Shape	Right-skewed , long tail; majority savings $\approx ₹20K-₹1.5L$	Most customers maintain moderate emergency funds; few have very large reserves
Skew & Kurtosis	1.79 & 4.67	Strong right-skew & highly leptokurtic
Outliers	Present	High-wealth customers with very large emergency funds; capped at 99th percentile - Skew = 1.44
Relationship with Max_Monthly_Emi (Regression target)	- Moderate positive correlation - Statistical Test - Spearman ≈ 0.258 (Significant)	- Customers with stronger liquidity cushions can afford higher EMIs confidently
Relationship with Emi_Eligibility (Classification target)	- Eligible customers have significantly higher emergency funds - Statistical Test - Kruskal $p = 0.00$ (Significant)	- Higher financial buffer \rightarrow lower credit risk, greater approval likelihood
Modeling Impact	Strong predictor — keep	Represents financial stability & risk tolerance ; improves model's ability to detect safe borrowers

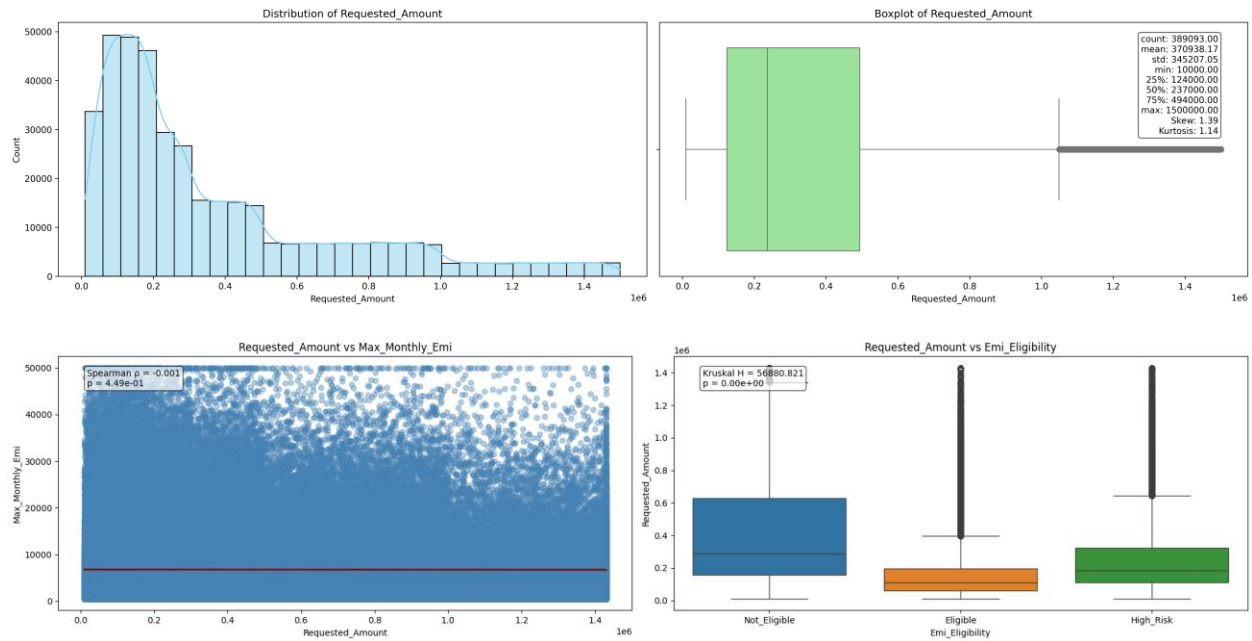
Loan Application Details:

Emi_Scenario:



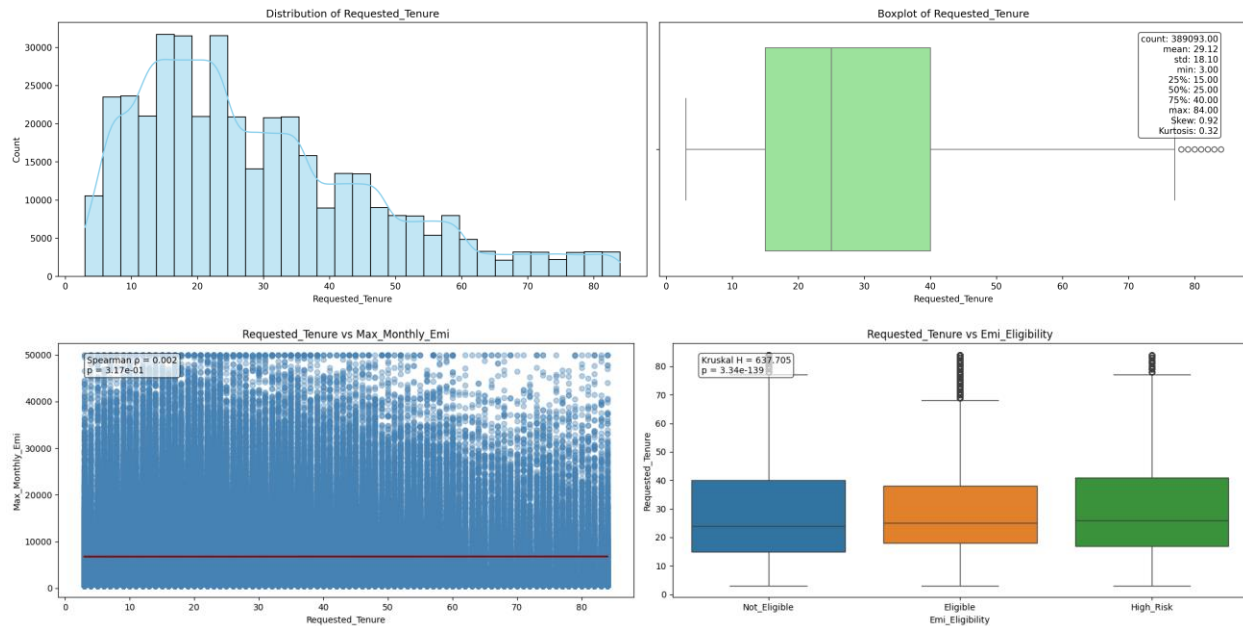
Aspect	Observation	Insight/Interpretation
Variable Type	Categorical (5 classes)	Type of EMI / loan purpose
Distribution Shape	Balanced — each category $\approx 20\%$ (Home Appliances, Personal Loan, E-commerce, Vehicle, Education EMI)	Dataset is evenly distributed — no dominant EMI type
Skew & Kurtosis	Not applicable (categorical)	None
Outliers	Not applicable (categorical)	None
Relationship with Max_Monthly_Emi (Regression target)	- Boxplots show very similar EMI capacity across all EMI types - Statistical Test - Kruskal $p \approx 0.765$ (Insignificant)	- EMI scenario does not influence maximum EMI capacity — spending purpose doesn't change affordability
Relationship with Emi_Eligibility (Classification target)	- Pattern looks similar across categories - Statistical Test - Chi-Square $p = 0.00$ (Significant)	- While statistically significant due to large sample size, practical effect is weak — approval likelihood is not strongly driven by EMI type
Modeling Impact	Weak predictor - keep but low weight	Minor contribution — helps capture spending category pattern but not a key financial capacity signal

Requested Amount:



Aspect	Observation	Insight/Interpretation
Variable Type	Numerical (continuous)	Loan amount requested by applicant
Distribution Shape	Right-skewed ; long tail Most values $\approx ₹1L - ₹5L$	Majority request moderate loan sizes; few applicants request very large loans
Skew & Kurtosis	1.39 & 1.14	Right-skewed & moderately leptokurtic
Outliers	Present	Some applicants request abnormally high loan amounts; capped at 99th percentile — Skew = 1.37
Relationship with Max_Monthly_Emi (Regression target)	- Statistical Test - Spearman $\rho \approx -0.001$ (Insignificant)	Loan request amount does not correlate with EMI capacity — applicants request amounts not aligned with affordability
Relationship with Emi_Eligibility (Classification target)	- Eligible group has lower median request ; High-Risk group requests slightly higher amounts - Statistical Test - Kruskal $p = 0.00$ (Significant)	- Applicants asking for higher loans face lower approval ; large loan size increases credit risk
Modeling Impact	Useful feature — keep	Important indicator of borrowing intent and risk appetite; helps model judge affordability gap

Requested Tenure:



Aspect	Observation	Insight/Interpretation
Variable Type	Numerical (continuous)	Loan duration requested (in months)
Distribution Shape	Slight right-skew , Peak around ~12–36 months	Most borrowers prefer short-to-medium loan tenures (1–3 years)
Skew & Kurtosis	0.92 & 0.32	Mild right skew & mild leptokurtosis
Outliers	Present	Few borrowers request very long terms; may reflect lower affordability / higher loan need ; capped at 99 th percentile – Skew = 0.91
Relationship with Max_Monthly_Emi (Regression target)	- No meaningful correlation - Statistical Test - Spearman $\rho \approx 0.002$ (Insignificant)	Tenure request isn't driven by EMI affordability
Relationship with Emi_Eligibility (Classification target)	- Eligible borrowers show slightly lower tenure preference - Statistical Test - Kruskal $p \approx 0.00$ (Significant)	- Shorter requested tenure → higher approval likelihood Long tenure = perceived higher credit risk
Modeling Impact	Moderately useful— keep	Reflects repayment behavior & risk profile; helps classify loan risk