

Hero Fincorp - Data Analysis Report

Task 1: Data Quality and Preparation

Insights

- Datasets are large and mostly populated; several columns have missing values (see sample counts).
- Dates and numeric columns required standardization (done).
- Some demographic fields like Gender/Marital_Status may be partially missing.

Recommendations

- Fill critical missing fields (Credit_Score, Annual_Income) where possible from KYC or external sources.
- Standardize date formats and ensure transaction-level timestamps for richer analyses.
- Add explicit transaction penalty/fee columns and map RM -> Branch for branch-level KPIs.

Task 2: Descriptive Analysis

Insights

- Loan amounts show wide dispersion — median significantly lower than max indicating a heavy-tailed distribution.
- Regional disbursements concentrated in a few regions (see top regions).
- Monthly disbursal trends available for time-series analyses.

Recommendations

- Use log-transformed metrics for modelling due to skew in Loan_Amount.
- Monitor monthly disbursal trends to detect seasonality and product demand shifts.

Task 3: Default Risk Analysis

Insights

- Loan_Amount and Overdue_Amount show positive correlation with Default_Amount.
- Lower Credit_Score correlates with higher default incidence where credit score is present.
- Branch/regional default rates vary widely — regional policies or RM practices likely contribute.

Recommendations

- Include Credit_Score, EMI/Income ratio, and Loan_Term in default propensity models.
- Prioritize branches/regions with elevated default rates for targeted interventions.

Task 4: Branch and Regional Performance

Insights

- Branches exhibit different disbursement volumes and processing times (see branches dataset).
- Where available, delinquent loans per branch give quick proxy for branch performance.

Recommendations

- Rank branches by combined score: disbursement volume, default rate, recovery rate, and processing time to identify under/over-performers.
- Deploy targeted capacity building for branches with high defaults but high disbursement volumes.

Task 5: Customer Segmentation

Insights

- Lower income and lower credit bands show higher default rates (where data is present).
- High-income & high-credit segments form the most profitable and least risky cohort.

Recommendations

- Create tailored products for each segment (e.g., smaller ticket loans for low-income band with stricter monitoring).
- Use segmentation to prioritize underwriting checks and pricing.

Task 6: Advanced Statistical Analysis

Insights

- Pairwise correlations highlight features useful for predictive modelling (Loan_Amount, EMI_Amount, Overdue_Amount).
- Use multicollinearity checks before regression modelling.

Recommendations

- Use logistic regression / tree-based models with cross-validation to build default propensity models.

Task 7: Transaction and Recovery Analysis

Insights

- Recovery rate differs by default reason; some reasons yield very low recoveries.
- Transaction-level data missing reduces precision of penalty analysis.

Recommendations

- Capture transaction-level data including penalty type and timestamps.
- Tailor recovery strategies by default reason (e.g., job loss cases may need restructuring).

Task 8: EMI Analysis

Insights

- Higher EMI customers have higher default propensity (top quartile analysis).

Recommendations

- Cap EMI according to income bands or offer longer tenor/alternate product to reduce stress.

Task 9: Loan Application Insights

Insights

- Approval rates and rejection reasons provide pipeline improvement areas.
- Processing fee differences between approved/rejected can indicate pricing sensitivity.

Recommendations

- Automate common rejection checks to speed up processing.
- Analyze rejected applications to refine targeting and product fit.

Task 10: Recovery Effectiveness

Insights

- Legal action generally improves recovery rate but with associated costs and timelines.
- Some default reasons have near-zero recovery suggesting alternative solutions required.

Recommendations

- Track cost-to-recover and net recovery to evaluate legal strategy ROI.
- Prioritize quick field recovery steps for small defaults to avoid escalation costs.

Task 11: Loan Disbursement Efficiency

Insights

- Processing time varies; longer times can lead to drop-offs and customer dissatisfaction.

Recommendations

- Monitor and reduce processing times, especially for high-value loans.

Task 12: Profitability Analysis

Insights

- Interest income is concentrated in larger loan tickets.
- Product-level profitability can be skewed by default rates; need net interest after expected losses.

Recommendations

- Compute net interest income after provisioning for expected defaults per product/segment.
- Consider pricing adjustments for high-risk products.

Task 13: Geospatial Analysis

Insights

- Geospatial mapping recommended (latitude/longitude needed) to move beyond region-level analysis.

Recommendations

- Obtain geo-coordinates for branches/customers for detailed mapping and rural vs urban splits.

Task 14: Default Trends

Insights

- Defaults show temporal patterns; monthly series useful for seasonality detection.

Recommendations

- Track rolling 12-month default rates and compare with disbursement vintages.

Task 15: Branch Efficiency

Insights

- Some branches have much longer avg processing times — bottleneck candidates for improvement.

Recommendations

- Benchmark best-performing branches and replicate their processes in underperforming ones.

Task 16: Time-Series Analysis

Insights

- Seasonality may be present; more detailed plots recommended for presentation.

Recommendations

- Use rolling averages and seasonal decomposition for forecasting demand and staffing needs.

Task 17: Customer Behavior Analysis

Insights

- A substantial proportion of customers are 'On-time' payers, while a smaller group contributes to most defaults.

Recommendations

- Create loyalty and cross-sell programs for On-time payers; strong remediation for repeat defaulters.

Task 18: Risk Assessment

Insights

- Risk matrix helps rank loan products by expected loss proxy (default/loan * rate).

Recommendations

- Use risk ranking to set provisioning and pricing per product type.

Task 19: Time to Default Analysis

Insights

- Time-to-default varies by loan purpose and customer cohorts; short time-to-default indicates underwriting miss.

Recommendations

- Flag loans that default within first 6 months for stricter pre-approval checks in similar segments.

Task 20: Transaction Pattern Analysis

Insights

- Customers with multiple overdue loans and elevated overdue ratios are at highest risk.

Recommendations

- Create tailored outreach, restructure offers, or legal escalation pathways for top irregular customers.