Hero FinCorp: A Comprehensive Data-Driven Analysis

Perform a comprehensive analysis of Hero FinCorp's loan portfolio and customer behavior to understand:

1. **Default Risks and Factors:**
   - Identifying customer and loan attributes contributing to defaults.
   - Trends in defaults across branches, regions, and loan types.
2. **Branch and Regional Efficiency:**
   - Evaluating branch performance in loan disbursement, processing time, and recovery rates.
3. **Customer Insights:**
   - Analyzing demographics, credit scores, and repayment patterns.
   - Segmenting customers to identify high-value and high-risk groups.
4. **Profitability:**
   - Understanding key drivers of profitability through loan disbursement trends, interest income, and recovery efficiency.

All tasks and analyses will collectively help Hero FinCorp optimize its loan approval process, reduce defaults, and improve branch efficiency, enabling data-driven decision-making.

Dataset Descriptions

Customer Demographics

➢ Contains details about Hero FinCorp's customers.
➢ Key Columns:
○ Customer_ID: Unique identifier for each customer.
○ Full_Name: Name of the customer.
○ Credit_Score: Numerical credit rating of the customer.
○ Annual_Income: Yearly income of the customer.
○ Employment_Status: Employment type (e.g., Salaried, Self-Employed).

Loan Data

➢ Details of all active and closed loans.

➢ Key Columns:
- Loan_ID: Unique identifier for each loan.
- Customer_ID: Customer associated with the loan.
- Loan_Amount: Principal loan amount.
- Interest_Rate: Annual interest rate on the loan.
- Loan_Term: Loan repayment duration in months.

Loan Applications

➢ Tracks the loan application process.
➢ Key Columns:
- Application_ID: Unique identifier for each loan application.
- Approval_Status: Whether the application was approved or rejected.
- Rejection_Reason: Reason for rejection (if applicable).
- Processing Fee: Fee charged during the application process.
- Application_Date: Date of application submission.

Transactions

➢ Record all financial transactions related to loans.
➢ Key Columns:
- Transaction_ID: Unique identifier for each transaction.
- Loan_ID: Associated loan for the transaction.
- Transaction_Type: Type of transaction (e.g., EMI Payment, Penalty).
- Transaction_Amount: Monetary amount of the transaction.
- Transaction_Date: Date of the transaction.

Default Records

➢ Details of customers who defaulted on their loans.
➢ Key Columns:
○ Default_ID: Unique identifier for each default case.
○ Loan_ID: Associated loan for the default.
○ Default_Amount: Amount remaining unpaid.
○ Recovery_Amount: Amount recovered post-default.
○ Default_Date: Date when the default occurred.

Branch Information

➢ Contains information about Hero FinCorp branches.
➢ Key Columns:
○   Branch_ID: Unique identifier for each branch.
○   Region: Geographic location of the branch.
○   Total_Active_Loans: Number of loans currently active in the branch.
○   Delinquent_Loans: Number of overdue loans in the branch.
○   Loan_Disbursement_Amount: Total loan disbursement volume.

Analysis Tasks

1. Data Quality and Preparation

- Validate and clean the datasets.

- Check for missing values, duplicate entries, and inconsistent data.

- Standardize date formats and remove irrelevant columns.

- Handle outliers in numeric columns like Loan_Amount, Interest_Rate, and Default_Amount.

---

2. Descriptive Analysis

- Summarize and visualize key metrics:

  o Distribution of Loan_Amount, EMI_Amount, and Credit_Score.

  o Regional trends in loan disbursement and defaults.

  o Monthly trends in loan approvals and disbursements.

---

3. Default Risk Analysis

- **Correlation Between Loan Attributes and Defaults**:

  o Calculate correlations between Loan_Amount, Interest_Rate, Credit_Score, and Default_Flag (a binary indicator for default).

- **Pairwise Correlation Analysis**:

  - o Create a heatmap to visualize the correlations between key variables, such as EMI_Amount, Overdue_Amount, and Default_Amount.

- **Correlation Between Branch Metrics and Defaults**:

  - o Analyze the relationship between branch performance metrics (e.g., Delinquent_Loans, Loan_Disbursement_Amount) and default rates.

---

4. Branch and Regional Performance

- Rank branches by:

  - o Loan disbursement volume.

  - o Processing time efficiency.

  - o Default rates and recovery rates.

- Compare branch performance across regions.

---

5. Customer Segmentation

- Segment customers by **income, credit score, and loan status**.

- Identify high-risk and high-value customer groups.

- Analyze repayment behavior across segments.

---

6. Advanced Statistical Analysis

1. **Correlation Analysis for Default Risks**:

   o   Examine the correlation between Credit_Score, Loan_Amount, Interest_Rate, Overdue_Amount, and Default_Flag.

2. **Pairwise Correlation Heatmap**:

   o   Generate a heatmap to visualize correlations among key variables like EMI_Amount, Recovery_Rate, and Default_Amount.

3. **Branch-Level Correlation**:

   o   Explore the relationship between branch performance metrics (Delinquent_Loans, Loan_Disbursement_Amount, Recovery_Rate) and overall efficiency.

---

7. Transaction and Recovery Analysis

- Analyze penalty payments and overdue trends.

- Evaluate recovery rates by Default_Reason and Legal_Action.

- Compare recovery rates across regions and branches.

---

8. EMI Analysis

- Analyze the relationship between EMI amounts and default probabilities.

- Identify thresholds for EMI amounts where defaults are most likely.

- Compare EMI trends across loan types.

---

9. Loan Application Insights

- Calculate approval and rejection rates for loan applications.

- Identify the most common reasons for loan rejection.

- Compare application processing fees between approved and rejected applications.

10. Recovery Effectiveness

- Determine the effectiveness of recovery efforts by calculating the ratio of Recovery_Amount to Default_Amount.

- Compare recovery rates for defaults with and without legal actions.

- Analyze branch-wise recovery performance.

11. Loan Disbursement Efficiency

- Analyze the time from application to loan disbursement and identify bottlenecks.

- Compare average processing times across branches.

- Evaluate disbursement trends by loan purpose and region.

12. Profitability Analysis

- Calculate the total interest income generated across all loans.

- Identify the most profitable loan purposes based on interest earnings.

- Compare profitability metrics for branches across regions.

13. Geospatial Analysis

- Map the distribution of active loans across regions.

- Compare default rates across different geographic regions.

- Visualize the loan disbursement trends for rural vs. urban areas.

---

14. Default Trends

- Analyze the number of defaults over time to identify patterns.

- Calculate the average default amount for different loan purposes.

- Compare default rates across customer income categories.

---

15. Branch Efficiency

- Calculate the average loan disbursement time for each branch.

- Identify branches with the highest number of rejected applications.

- Compare branch efficiency based on customer satisfaction metrics (if available).

---

16. Time-Series Analysis

- Analyze monthly loan disbursement trends over the last 5 years.

- Identify seasonal patterns in loan applications and disbursements.

- Compare monthly default rates across regions.

---

17. Customer Behavior Analysis

- Categorize customers based on their repayment behavior (e.g., always on time, occasional defaulters, frequent defaulters).

- Analyze patterns in loan approval and rejection reasons segmented by customer demographics.

- Identify high-value customers with consistent repayment histories.

18. Risk Assessment

- Develop a risk matrix for loan products based on Default_Amount, Loan_Term, and Interest_Rate.

- Rank loan types by risk level and suggest mitigation strategies.

- Analyze high-risk customer segments by credit score and income.

19. Time to Default Analysis

- Calculate the average time from loan disbursement to default for overdue loans.

- Identify loan purposes with the shortest time to default.

- Compare the time to default across customer demographics.

20. Transaction Pattern Analysis

- Identify customers with irregular repayment patterns.

- Analyze penalty payments as a proportion of total transactions.

- Compare transaction amounts for overdue vs. non-overdue loans.

Collective Agenda

The **collective agenda** is to provide Hero FinCorp with actionable recommendations to:

1. Minimize loan defaults by identifying high-risk customers and regions.

2. Optimize branch operations by improving processing time and recovery rates.

3. Enhance profitability through better customer segmentation and interest income strategies.

Deliverables

1. **Key Insights Report**:

   o Summarize findings for each task, including key metrics and insights.

2. **Visualizations**:

   o Provide plots, charts, and heatmaps to support findings.

3. **Recommendations**:

   o Strategies for reducing defaults, optimizing branch performance, and improving profitability.