**Description**

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

Records 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:

Distribution of Loan\_Amount, EMI\_Amount, and Credit\_Score.

Regional trends in loan disbursement and defaults.

Monthly trends in loan approvals and disbursements.

3. Default Risk Analysis

Correlation Between Loan Attributes and Defaults:

Calculate correlations between Loan\_Amount, Interest\_Rate, Credit\_Score, and Default\_Flag (a binary indicator for default).

Pairwise Correlation Analysis:

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:

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:

Loan disbursement volume.

Processing time efficiency.

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:

Examine the correlation between Credit\_Score, Loan\_Amount, Interest\_Rate, Overdue\_Amount, and Default\_Flag.

1. Pairwise Correlation Heatmap:

Generate a heatmap to visualize correlations among key variables like EMI\_Amount, Recovery\_Rate, and Default\_Amount.

1. Branch-Level Correlation:

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:

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

Optimize branch operations by improving processing time and recovery rates.

Enhance profitability through better customer segmentation and interest income strategies.

Deliverables

Key Insights Report:

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

Visualizations:

Provide plots, charts, and heatmaps to support findings.

Recommendations:

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

# Clean Branches dataset

branches\_df.fillna({"Delinquent\_Loans": 0,

"Loan\_Disbursement\_Amount": branches\_df["Loan\_Disbursement\_Amount"].median()}, inplace=True)

branches\_df.drop\_duplicates(inplace=True)

# Clean Customers dataset

customers\_df.fillna({

"Credit\_Score": customers\_df["Credit\_Score"].median(),

"Annual\_Income": customers\_df["Annual\_Income"].median(),

"Employment\_Status": "Unknown"

}, inplace=True)

customers\_df.drop\_duplicates(inplace=True)

# Clean Defaults dataset

defaults\_df.fillna({"Recovery\_Amount": 0}, inplace=True)

defaults\_df.drop\_duplicates(inplace=True)

# Clean Loans dataset

loans\_df.fillna({

"Interest\_Rate": loans\_df["Interest\_Rate"].median(),

"Loan\_Term": loans\_df["Loan\_Term"].median(),

"Loan\_Status": "Unknown"

}, inplace=True)

loans\_df.drop\_duplicates(inplace=True)

# Clean Transactions dataset

transactions\_df.fillna({"Transaction\_Amount": 0}, inplace=True)

transactions\_df.drop\_duplicates(inplace=True)

# Summary of cleaned data

cleaned\_data\_quality = {

"branches": {

"Missing Values": branches\_df.isnull().sum(),

"Duplicates": branches\_df.duplicated().sum(),

},

"customers": {

"Missing Values": customers\_df.isnull().sum(),

"Duplicates": customers\_df.duplicated().sum(),

},

"defaults": {

"Missing Values": defaults\_df.isnull().sum(),

"Duplicates": defaults\_df.duplicated().sum(),

},

"loans": {

"Missing Values": loans\_df.isnull().sum(),

"Duplicates": loans\_df.duplicated().sum(),

},

"transactions": {

"Missing Values": transactions\_df.isnull().sum(),

"Duplicates": transactions\_df.duplicated().sum(),

}

}

# Print the summary

for dataset, info in cleaned\_data\_quality.items():

print(f"Dataset: {dataset}")

print(f"Missing Values:\n{info['Missing Values']}")

print(f"Duplicate Rows: {info['Duplicates']}\n")