Loan Prediction Analysis - Detailed Documentation

Chapter: 1 Introduction

This document provides an in-depth analysis of the Loan Default Prediction dataset and its associated dashboard. The goal of the analysis is to understand patterns in loan defaults, identify key risk factors, and provide actionable insights based on data visualization.

Chapter 2: Dataset Overview

- **Total Records:** 255,347
- **Total Features:** 18
- **Target Variable:** Default (0 = No Default, 1 = Default)

Key Features:

1. **Demographics:** Age, Education, Marital Status, Employment Type
2. **Financial Factors:** Income, Credit Score, Loan Amount, Interest Rate, DTI Ratio
3. **Loan Details:** Loan Term, Num Credit Lines, Loan Purpose, Months Employed
4. **Additional Factors:** Has Mortgage, Has Dependents, Has Co-Signer

Chapter: 3 Dashboard Overview

The Power BI dashboard contains several visualizations that analyze various aspects of loan defaults. The key charts are numbered for clarity.

Chart 1 : Total Defaulters Card:

Observation:

- The total number of defaulters is **30,000.**
- This metric represents the count of individuals who failed to repay their loans.
- The number is displayed prominently, indicating a significant portion of the loan portfolio is at risk.

Insights:

- A high number of defaulters suggests a potential issue in loan approval processes, borrower risk assessment, or economic conditions affecting repayment.
- Further analysis should explore **default rates by income group, loan amount, credit score, and employment status** to identify key risk factors.
- Financial institutions should consider **debt counseling programs** for high-risk borrowers to improve repayment behavior.
- If the total loan portfolio is relatively small, 30K defaulters could indicate a severe risk; however, if the portfolio is large, the percentage of defaults needs to be assessed for a clearer risk evaluation.

Chart 2: Total Loan Amount Card:

Observation:

- The total loan amount disbursed is **33 billion**.
- This represents the cumulative sum of all loans granted to borrowers.

Insights:

- A high total loan amount indicates **significant financial exposure** for the lending institution.

- If a large portion of this amount is linked to high-risk borrowers (e.g., low credit scores, high DTI ratios), there could be **potential liquidity concerns**.
- Further breakdown by **loan purpose, borrower demographics, and loan tenure** can help identify areas where risk is concentrated.

Chart 3: Total Loans Card

Observation:

- The total number of loans disbursed is **255.35K (i.e., 255,350 loans).**
- This re presents the total count of loan agreements issued by the lender.

Insights:

- With 255.35K total loans and **30K defaulters**, the default rate is approximately **11.75%**.
- The lender's portfolio size is substantial, indicating significant lending activity.
- If a high number of loans come from high-risk segments, this could indicate a need for stricter credit evaluation.

Chart 4: Education wise Loan Count (Bar Chart)

Observation:

- The loan count is evenly distributed across different education levels, with 64K loans for each category:
    - Bachelor’s 64.37 k loans.
    - High School 63.9 k loans.
    - Master’s 63.54k loans.
    - PhD 63.54 k loans.

Insights:

- Bachelor’s degree holders have the highest number of loans (64.37K), while Master’s and PhD holders have the lowest (63.54K each).
- **The difference in loan count is minimal**, suggesting that education level does not play a significant role in loan approval.
- However, the slight **decline in loan count from Bachelor's to PhD** may indicate:
    - Higher education individuals might **require fewer loans** due to better job prospects or financial stability.

- Fewer people pursue **higher education degrees**, leading to lower loan applications.

Chart 5: Default Rate Summary (Bar Chart)

Observation:

- **225.69K** borrowers have repayed their loan (Default = 0).
- **29.65K** borrowers have no repayed their loan (Default = 1).
- The majority of borrowers **(around 88.4%) are non-defaulters**, while **11.6% are defaulters**.

Insights:

- **Low Default Rate (11.6%)**: The proportion of defaulters is relatively small, indicating that most borrowers are managing their loan repayments effectively.
- **Credit Risk Management Appears Strong**: The financial institution seems to have a robust screening process, ensuring loans are given to individuals with a lower probability of defaulting.

Chart 6: Marital Status wise Loans (Bar Chart)

Observationx

- Married individuals have the highest number of loans **(85.30K).**
- **Divorced individuals** have **85.03K** loans.
- **Single individuals** have **85.01K** loans.
- The loan distribution across marital statuses is almost equal, with a minor variation.

Insights:

- **Marital Status Has No Significant Impact on Loan Distribution**: The number of loans is fairly uniform across all marital statuses, indicating that banks or financial institutions do not heavily differentiate based on marital status when granting loans.
- **Slightly Higher Loans for Married Individuals**: This might suggest that married individuals have more financial responsibilities or better loan eligibility (possibly due to dual income or stability).
- **Loan Approval Criteria May Depend on Other Factors**: Since marital status alone does not show a strong difference, other factors like income, credit score, or employment type might play a more significant role in loan approvals.

Chart 7: Loan Purpose Distribution (Pie Chart)

Observation:

- The loan distribution across different purposes is fairly balanced, with each category having around **19.91% to 20.09%** of total loans.
- **Business Loans** have the highest percentage (**20.09%**, 51.3K loans).
- **Auto Loans** have the lowest percentage (**19.91%**, 50.84K loans).
- Other categories such as **Home, Education, and Other Loans** also hold nearly equal shares.

Insights:

- **Even Loan Distribution**: The near-equal distribution suggests that financial institutions provide loans equally across different needs, indicating a balanced demand.
- **Slightly Higher Business Loans**: Business loans being the highest might indicate a growing interest in entrepreneurship or business expansion.
- **Auto Loans Being Slightly Lower**: This could mean fewer people are taking car loans compared to other loan purposes, possibly due to economic conditions or alternative financing options.

Chart 8: Loan Term wise Default (Line Chart)

Observation:

- The default count fluctuates slightly between loan terms **10 to 40 months**, remaining between **5,907 to 5,922**.
- A **significant increase in defaults** is observed for the **60-month term**, reaching **5,983**.
- The lowest default count occurs at **40 months (5,907)**.

Insights:

- Longer Loan Terms Lead to Higher Defaults:
    - Borrowers with **longer repayment periods (60 months)** tend to default more.
- Stability Between 10-40 Months:
    - Default counts remain almost stable for loan terms **10-40 months**, indicating a **safer repayment period**.

Chart 9: Loan Amount Distribution by Employment Type (Tree Map)

Observation:

- Unemployed individuals received the highest total loan amount: **8,175M**
- Part-time employees followed closely: **8,170M**.
- Self-employed individuals received **8,118M**, which is slightly lower.
- Full-time employees received the lowest loan amount: **8,114M.**

Insights:

- Unemployed individuals receiving the highest loan amount (8,175M) is unusual:
    - This could indicate government-backed loans, emergency financial aid, or loans based on assets rather than income.
    - It would be useful to check their **default rate** to assess repayment risks.
- Part-time employees received a higher loan amount than full-time employees:
    - This might suggest they apply for loans more frequently due to lower income stability.
    - Banks may still consider them relatively safe borrowers.
- Self-employed individuals received slightly more loans than full-time employees:
    - Likely due to business expansion or operational needs.
    - But not as much as unemployed/part-time, possibly due to variability in income.
- Full-time employees received the lowest loan amount:
    - This may indicate they rely less on loans due to stable monthly earnings.
    - They might qualify for loans with lower amounts due to better financial planning.

Chart 10: Default Rate (in %) by Employment Type (Bar Chart)

Observation:

- **Unemployed: 13.6%** (Highest)
- **Part-time: 12.0%**
- **Self-employed: 11.5%**
- **Full-time: 9.5%** (Lowest)

Insights:

- Higher Loan Amount ≠ Lower Default Risk:
  - The unemployed group received the highest loan amount (8,175M) but has the highest default rate (13.6%).
  - This suggests that financial institutions may be issuing large loans to unemployed individuals, possibly due to government-backed schemes or collateral-based approvals, but they also pose the highest risk.
- Full-Time Employees Have the Lowest Default Rate (9.5%):
  - They also received the lowest loan amount (8,114M), which aligns with their **higher financial stability** and better ability to repay loans.
- Part-Time and Self-Employed Borrowers Show Moderate Risk:
  - Both groups received substantial loan amounts (8,170M for part-time, 8,118M for self-employed).
  - However, their default rates (**12.0% and 11.5%**) are still significantly higher than full-time employees, indicating moderate financial uncertainty.

Chart 11: Default Rate Analysis by Marital Status (Bar Chart)

Observation:

- Divorced: 12.5% (Highest)
- **Single: 11.9%**
- **Married: 10.4%** (Lowest)

Insights:

- Divorced individuals have the highest default rate (12.5%):
  - This may indicate financial instability due to **single-income dependency**, **legal expenses**, or **loss of shared financial responsibilities.**
- Single individuals have a slightly lower default rate (11.9%):
  - They might have fewer financial obligations compared to divorced individuals, but still lack the stability of a dual-income household.
- Married individuals have the lowest default rate (10.4%):
  - Likely due to **dual-income households**, **shared financial responsibilities**, and **better financial planning**.

Chart 12: Table:

Insights:

- Business loans have the highest default rate (12.33%)
    - Higher risk due to **uncertain revenue streams** and **market fluctuations**.
- Home loans have the lowest default rate (10.23%)
    - **Lower risk** due to **collateral-backed nature** and **stable repayment patterns**.
- **Education, Auto, and Other loans** have similar default rates (**11.79% - 11.88%**)
    - Education loans may be riskier due to **delayed income sources**.
    - Auto loans may default due to **vehicle depreciation and financial instability**.

Chart 13: Age wise Defaulters (Line Chart)

Observation:

- Younger individuals (below 25 years) have the highest number of defaulters (above 1000).
- Default rate gradually decreases with age as financial stability improves.
- Defaults are significantly lower for individuals aged 40 and above.
- **Elderly borrowers (60+) have the lowest defaults**, likely due to better financial management and lower loan exposure.

Chart 14: Slicer

Features:

- **Dropdown Filters:**
    - Employment Type: Allows selection of specific employment categories (Full-time, Part-time, Self-employed, Unemployed).
    - Loan Purpose: Filters data based on different loan types (Business, Home, Auto, etc.).
- **Navigation Button:**
    - **"Towards Dashboard 2"** with a left arrow button, likely used for navigation to another page.

Chapter 4: DAX Measure Used

To enhance data analysis, the following DAX queries were used to create key performance indicators:

1. Default Rate Percentage:

```
VAR TotalLoans = COUNT(Loan_default_prediction[LoanID])
VAR DefaultedLoans = COUNTROWS(FILTER(Loan_default_prediction,
Loan_default_prediction[Default] = 1))
RETURN
DIVIDE(DefaultedLoans, TotalLoans, 0) * 100
```

2. Number of Defaulters:

```
COUNTROWS(FILTER(Loan_default_prediction, Loan_default_prediction[Default] = 1))
```

3. Total Loan Amount (in Million):

```
SUM(Loan_default_prediction[LoanAmount]) / 1000000
```

Chapter 5: Key Insights

1. Default Rate by Employment Type

- Unemployed individuals have the highest default rate (13.6%), followed by part-time employees (12.0%) and self-employed individuals (11.5%).
- **Full-time employees have the lowest default rate (9.5%)**, indicating that stable employment reduces default risk.

2. Default Rate by Marital Status

- Divorced individuals have the highest default rate (12.5%), followed by single individuals (11.9%).
- **Married individuals have the lowest default rate (10.4%)**, suggesting financial stability might be a factor in lower defaults.

3. Default Rate by Loan Purpose

- **Business loans have the highest default rate (12.33%)**, likely due to the risks involved in entrepreneurship.

- **Home loans have the lowest default rate (10.23%)**, possibly because they are secured by collateral.
- **Education, Auto, and Other loan categories** have default rates between **11.79% and 11.88%**, indicating moderate risk levels.

4. Age-wise Default Trend

- The number of defaulters is highest among younger individuals (below 25 years) and gradually decreases with age.
- This suggests that **financial responsibility increases with age**, or older individuals have better credit management

5. Slicer Functionality:

- The dashboard includes slicers for Employment Type and Loan Purpose, allowing dynamic filtering of data.
- A navigation button is present for easy access to Dashboard 1.

Chapter 6: Conclusion

- Employment stability plays a crucial role in loan defaults. Full-time employees have the lowest default rate, while unemployed individuals are at the highest risk.
- **Marital status impacts default rates.** Divorced and single individuals are more likely to default compared to married individuals, indicating that financial stability might be better in married households.
- **Loan purpose influences default risk.** Business loans have the highest default rate due to the uncertainty of business success, while home loans have the lowest, likely due to collateral backing.
- **Younger individuals are at a higher risk of defaulting,** possibly due to lower financial experience and unstable income sources. Default rates gradually decrease with age.
- **The dashboard enables dynamic data exploration** through slicers, allowing stakeholders to analyze default patterns based on employment type and loan purpose efficiently.