Case Study and Dashboards Report

Introduction:
The objective of this project is to analyze loan portfolios, borrower demographics, and associated risk factors using various Power BI dashboards. These dashboards offer insights into loan distributions, default rates, and borrower profiles, allowing us to identify high-risk segments and propose data-driven recommendations. By applying analytical techniques to historical loan data, we aim to improve loan approval processes and reduce default risks.

Methodology:

1. Data Cleaning and Preprossessing:

With Python Programming Language and Power Bi to visualize these data

- Handling Missing Values:Missing values were identified and handled using appropriate techniques (filling with mean/median values or removing records with significant missing data).
- Outliers Identification:Outliers in key financial columns like loan amount, interest rate, and annual income were detected using box plots and handled through transformation or removal to improve model accuracy.
- Duplicate Records:Duplicate entries were removed to avoid redundant calculations.
- Loan Status Categorization:Loan status was transformed into distinct categories such as “Fully Paid,” “Charged Off,” and “Current.”
- Employment Length Standardization:Employment lengths were standardized (e.g., merging “10+ years” and “10 years”) to avoid inconsistencies.
- Data Encoding:Categorical variables such as “Grade” and “Home Ownership” were converted into numerical values using label encoding to facilitate analysis.

2. Data Preparation and Dashboard Development

Each dashboard was built using the following KPIs and filters:

- First Dashboard: Loan Portfolio Overview

- **KPIs:** Total Loan Amount, Total no of Loans, Sum of last Pymnt amnt, Avg of Int_rate,Loan Default Rate.
- **Visualizations:**
  1. Bar chart: Sum of loan amnt by Grade
  2. Dount chart: Sum of loan amnt by Term
  3. Gauge chart: Avg of loan amnt by Avg int_rate
  4. Pie chart: Loan status by Sum of loan amnt

- Second Dashboard: Loan Performance Analysis

- Similar to the first, this dashboard further focused on geographic distribution and borrower characteristics such as state and FICO score range.

- **KPIs**: Prepayment Rate, Avg Time to Default, Recovery Rate, Delinquency_Rate
- Visualizations:
  5. Bar chart: Delinquency_Rate by Grade
  6. Bar chart: Emp_Length by Avg Time to Default
  7. Pie chart: Purpose by Recovery Rate
  8. Line chart : Term by Prepayment Rate
  9. Heat Map: addr_state by Count of Dayes to Default

Some of calculations:

```
1. PrepaymentRate = DIVIDE(
    [FullyPaidOffLoans],
    [TotalLoansCount],
    0
)
```

```
2.AverageTimeToDefault =
CALCULATE(
    AVERAGEX(
        Sheet1,
        IF(
            Sheet1[Loan_Status] = "Charged Off",
            DATEDIFF(
                Sheet1[issue_d],
                Sheet1[last_pymnt_d],
                DAY
            ),
            BLANK()
        )
```

```
3.RecoveryRate = DIVIDE(
    [TotalRecoveredAmount],
    [TotalChargedOffAmount],
    0
)
```

- Third Dashboard: Loan Distribution by Customer Demographics:

- **KPIs**:Default rate by income range, loan approval rate by employment length, average loan amount by age group, default rate by education level, State of Borrower.
- **Visualizations:**

1. Line chart: Income Range by Default Rate
2. Clustered bar chart: Education Level by Default Rate
3. Stacked column chart: Age Group and Default Rate in axis and Avg loan amnt as a values
4. Dount chart: Emp_Length by Loan Approval Rate by Education levels
5. Heat Map: addr_state by Education levels and count of Age group

-Some of calculations:

```
1.AgeGroup =
SWITCH(
    TRUE(),
    YEAR(Sheet1[earliest_cr_line]) >= YEAR(TODAY()) - 25 && YEAR(Sheet1[earliest_cr_line]) <=
YEAR(TODAY()) - 18, "18-25",
    YEAR(Sheet1[earliest_cr_line]) >= YEAR(TODAY()) - 35 && YEAR(Sheet1[earliest_cr_line]) <=
YEAR(TODAY()) - 26, "26-35",
    YEAR(Sheet1[earliest_cr_line]) >= YEAR(TODAY()) - 45 && YEAR(Sheet1[earliest_cr_line]) <=
YEAR(TODAY()) - 36, "36-45",
    YEAR(Sheet1[earliest_cr_line]) >= YEAR(TODAY()) - 55 && YEAR(Sheet1[earliest_cr_line]) <=
YEAR(TODAY()) - 46, "46-55",
    YEAR(Sheet1[earliest_cr_line]) >= YEAR(TODAY()) - 65 && YEAR(Sheet1[earliest_cr_line]) <=
YEAR(TODAY()) - 56, "56-65",
    YEAR(Sheet1[earliest_cr_line]) < YEAR(TODAY()) - 65, "65+",
    "Unknown"
)
```

```
2.LoanApprovalRateByEmploymentLength =
DIVIDE(
    CALCULATE(
        COUNTROWS(Sheet1),
        Sheet1[loan_status] IN {"Current", "Fully Paid"},
        ALLEXCEPT(Sheet1, Sheet1[Emp_Length])
    ),
    CALCULATE(
        COUNTROWS(Sheet1),
        ALLEXCEPT(Sheet1, Sheet1[Emp_Length])
    ),
```

```
    0)

3.EducationLevel =
SWITCH(
    TRUE(),
    CONTAINSSTRING(LOWER(Sheet1[emp_title]), "engineer"), "Bachelor's Degree",
    CONTAINSSTRING(LOWER(Sheet1[emp_title]), "manager"), "Master's Degree",
    CONTAINSSTRING(LOWER(Sheet1[emp_title]), "director"), "Master's Degree",
    CONTAINSSTRING(LOWER(Sheet1[emp_title]), "analyst"), "Bachelor's Degree",
    CONTAINSSTRING(LOWER(Sheet1[emp_title]), "technician"), "Associate's Degree",
    CONTAINSSTRING(LOWER(Sheet1[emp_title]), "clerk"), "High School Diploma",
    CONTAINSSTRING(LOWER(Sheet1[emp_title]), "assistant"), "Associate's Degree",
    CONTAINSSTRING(LOWER(Sheet1[emp_title]), "consultant"), "Bachelor's Degree",
    CONTAINSSTRING(LOWER(Sheet1[emp_title]), "supervisor"), "Bachelor's Degree",
    CONTAINSSTRING(LOWER(Sheet1[emp_title]), "advisor"), "Master's Degree",
    CONTAINSSTRING(LOWER(Sheet1[emp_title]), "teacher"), "Master's Degree",
    CONTAINSSTRING(LOWER(Sheet1[emp_title]), "lieutenant"), "Bachelor's Degree",
    CONTAINSSTRING(LOWER(Sheet1[emp_title]), "lead"), "Bachelor's Degree",
    CONTAINSSTRING(LOWER(Sheet1[emp_title]), "police"), "Bachelor's Degree",
    CONTAINSSTRING(LOWER(Sheet1[emp_title]), "mechanic"), "High School Diploma",
    CONTAINSSTRING(LOWER(Sheet1[emp_title]), "coach"), "Master's Degree",
    "Not specified"
)

4.PrepaymentRate = DIVIDE(
    [FullyPaidOffLoans],
    [TotalLoansCount],
    0
)

5.RecoveryRate = DIVIDE(
    [TotalRecoveredAmount],
    [TotalChargedOffAmount],
    0
)

6.Estimated_Age = 22 + MAX(Sheet1[Emp_Length])
7.FullyPaidOffLoans = CALCULATE(
    [TotalLoansCount],
    Sheet1[Loan_Status] = "Fully Paid",
    Sheet1[last_pymnt_d] < EDATE(Sheet1[issue_d], Sheet1[NumericScheduledTerm])
)
```

- **Fifth Dashboard: Loan Delinquency Analysis**

- Defining Default Criteria: Loans were flagged as “default” if delinquency exceeded 30% of the loan amount or remained delinquent for more than 90 days.

- **KPIs:** Total number of default loans, default rate, total default amount, average default amount,Default rate over time.

- **Visualizations:**

1.Line chart: Issue_d by Default rate over time
2.Clustered bar chart: Count of Default rate by Purpose
3.Clustered bar chart: Count of Default rate by loan amnt
4.Pie chart: Home ownership by Income Range with values Total of default loans
5.Heat Map: addr state by Default rate, and Age group

-Some of calculations:

1.Average_Default_Amount = AVERAGEX(
```
    FILTER(Sheet1, Sheet1[is_default] = 1),
    Sheet1[loan_amnt]
)
```

2.AverageTimeToDefault =
```
CALCULATE(
    AVERAGEX(
        Sheet1,
        IF(
            Sheet1[Loan_Status] = "Charged Off",
            DATEDIFF(
                Sheet1[issue_d],
                Sheet1[last_pymnt_d],
                DAY
            ),
            BLANK()
        )
    )
)
```

3.Default_Rate = DIVIDE(
```
    CALCULATE(COUNT(Sheet1[id]), Sheet1[is_default] = 1),
    COUNT(Sheet1[id])
)
```

4.Default_Rate_Over_Time =
```
DIVIDE(
    CALCULATE(
        COUNT(Sheet1[id]),
        FILTER(Sheet1, Sheet1[is_default] = 1)
    ),
    COUNT(Sheet1[id])
)
```

```
5.is_default = IF(Sheet1[Delinquent_Loans] > 0, 1, 0)
6.TotalDefaultLoans =
CALCULATE(
    COUNT(Sheet1[id]),
    FILTER(Sheet1, Sheet1[is_default] = 1)
)
```

Conclusions:

1. Loan Portfolio Insights:
   The analysis of loan portfolios showed that higher loan amounts and extended loan terms (60 months) correlate with higher interest rates. However, default rates were found to be disproportionately high in loans with lower credit grades.

2. Borrower Demographic Patterns:
   Borrowers with higher incomes and longer employment histories tended to have lower default rates, while younger borrowers and those with shorter employment lengths displayed higher default risks.

3. Loan Default Trends:
   Default trends showed that loans delinquent for more than 90 days accounted for a significant portion of overall defaults, especially in regions with lower income levels and poorer credit scores.

Recommendations:

1. Improving Loan Approval Processes:
   - Implement stricter loan approval criteria for high-risk segments such as young borrowers and individuals with shorter employment histories.
   - Introduce more detailed background checks and enhanced verification for lower credit grade loans to reduce the default rate.

2. Interest Rate Adjustments:
   - Reassess interest rates for loans in the 36-month term bracket, as they exhibit lower default rates. This could incentivize borrowers to opt for shorter-term loans, reducing risk for the lender.

3. Targeted Financial Products:
   - Develop loan products tailored for high-income and long-term employed individuals, offering them lower interest rates and flexible payment plans to encourage higher borrowing from low-risk segments.

4. Risk-Based Lending Policies:
   - Adopt a more dynamic risk-based pricing model, where interest rates are determined not just by credit grade but also by demographic factors such as age, employment length, and income level.