[RISK MANAGEMENT MUDRA LOAN DATASET PROJECT]

PROGRAM-DATA ANALYTICS

November 18, 2024

TEAM LEADER- TANISHKA MISHRA

ID-BC2023196

Team Members-

STUDENT NAME	STUDENT ID	ROLL NO	PROJECT STATUS
MOHD WASEEM	BC2023032	2310201357	CONTRIBUTION IN CODING
TANISHKA MISHRA	BC2023196	2310201582	CONTRIBUTION IN PROJECT REPORT
HARMENDRA SINGH	BC2023324	2310201217	ACTIVE
DARSHIKA RATHORE	BC202367	2310101126	ACTIVE
MONU PRAKASH	BC2023458	2310201366	ACTIVE
GURUPAL SINGH	BC2023308	2310201215	ACTIVE
PRADEEP SINGH	BC2023598	2310201407	ACTIVE
WAHEED HUSAIN	BC2023055	2310201628	ACTIVE
ZAIB ILAHI SHAMSI	BC2023047	2310201634	ACTIVE
SAURABH KUMAR	BC2023476	2310201514	ACTIVE
SHUBHAM GANGWAR	BC2023721	2310201677	ACTIVE
AMIT	BC2023112	2310201077	ACTIVE
SACHIN VERMA	BC2023365	2310201486	ACTIVE
ANSHIKA AGARWAL	BC2023035	2310201109	ACTIVE

Sr. No.	Content
1.	**Introduction**
2.	**Dataset Description**
3.	**Methodology**
4.	**Exploratory Data Analysis**
5.	**Predictive Analysis**
6.	**Key Findings and Discussion**
7.	**Recommendations**
8.	**Conclusion**
9.	**Reference**
10.	**Appendix**

Introduction

- **<u>OVERVIEW OF MUDRA LOANS:-</u>**
  Mudra (Micro Units Development and Refinance Agency) Loans are a Government of India initiative to provide financial assistance to small businesses. Highlight its importance for MSMEs.

- **<u>Purpose of the Project:-</u>**
  To analyze the dataset to uncover patterns, trends, and insights about loan distributions, defaults, and sectoral allocations.

- **<u>Scope of the Analysis:-</u>**
  Describe the dataset's scope, such as timeframe, geographical coverage, and data points (e.g., loan amount, type, beneficiary).

Dataset Description

☐ **Source of the Data-**
Mention where you obtained the dataset

☐ **Features and Attributes-**
Provide a table summarizing key columns

☐ **Size and Structure- Number of rows and columns. Missing values and data types.**

☐ **Data Limitations- Cleaning techniques Transformation steps**

☐ **Tools and Technologies Used-**

> ✦ **PANDAS- FOR DATA MANIPULATION AND ANALYSIS**
> ✦ **NUMPY- FOR NUMERICAL COMPUTATIONS**
> ✦ **MATPLOT AND SEABORN- FOR DATA VISUALIZATION**

Methodology

- The Project appears to be an `.ipynb` file, which is a Jupyter Notebook.

- its content and review it to provide a methodology section if it contains relevant information.

- The file contains a structured Jupyter Notebook with code cells, markdown, and possibly output.

Analytical Techniques-

- Exploratory Data Analysis (EDA).
- Statistical summaries, correlations, and visualizations.

Exploratory Data Analysis (EDA)

1.Load and Analyze Dataset-I'll extract relevant code snippets and execute them if the dataset is included or referenced.

2. **Inspect Outputs- Review any outputs, plots, or summaries already present.**

3. **Perform Additional EDA- If needed, I can run additional EDA steps using the provided data.**

The Project contains various steps likely aimed at analyzing a dataset named `Project.csv`.

⚒ Here's an outline of the key initial cells:

1. **Cell 1: Mounts Google Drive (to access files).**

2. **Cell 2: Imports libraries (`pandas`, `numpy`, `matplotlib.pyplot`, and `seaborn`) for data analysis and visualization.**

3. **Cell 3: Loads a dataset (`Project.csv`) into a DataFrame named `df` and displays it.**

4. **Cell 4-5: Explores the dataset using `.shape` and `.info()`.**

5. **Cell 6: Checks for missing values with `.isnull().sum()`.**

6. **Cell 7-8: Drops missing values and verifies the absence of null entries.**

Predictive Analysis

🗲 **The document's goal is likely to explore and analyze loanrelated data to identify trends, correlations, or insights. Let me know if you'd like a deeper analysis or specific aspects highlighted!**

☐ **Data Loading and Cleaning-**
  Loading a CSV file and exploring its structure (e.g., using info() and checking for missing values).
    - Dropping rows with missing data.

☐ **Data Analysis and Visualization-**
☐ **Plotting histograms for columns such as Classification_Code, Loan_Term, and Low_Documentation_Loan.**
☐ **Creating bar plots and box plots for visual insights.**
☐ **Performing numeric conversions on specific columns (e.g., Low_Documentation_Loan and Count_Employees).**
☐ **Loan terms**
☐ **Employee count and demographic relationships.**
☐ **Focus Areas-**
☐ **Possible classification-related analysis (e.g., Classification_Code).**

Key Findings and Discussion

The content of the Project seems to involve data analysis and visualization, focusing on tasks such as-

<u>Keywords:</u>

- **Google Colab- References using Google Colab for analysis.**
- **Drive Mount-Accessing files through mounted drives.**

- **<u>Pandas</u>- Handling and manipulating data.**

- **<u>NumPy</u>- Numerical computations.**

- **<u>Matplotlib</u>- Plotting and visualization.**

- **<u>Seaborn</u>- Advanced data visualization.**
- **<u>Data Cleaning</u>- Tasks like handling null values (df.isnull().sum() and dropna).**
- **<u>Histograms</u>-Visualization of distributions (plot(kind='hist')).**
- **<u>Dataframe Operations</u>-**
- **References to df indicate significant work with DataFrames.**

Discussion

1. <u>**nbformat and nbformat minor:**</u>
   Specify the version of the Jupyter Notebook format used.

2. <u>**metadata:**</u>
   Includes meta information about the notebook.

3. <u>**cells**</u>**:**
   Contains the actual content, including code, text (Markdown), and outputs.

<u>**21 code cells (all cells are of type "code")-**</u> **The first few cells perform the following actions**:

1. **Initial setup** (e.g., mounting Google Drive using drive.mount).

2. **Importing libraries such as pandas and numpy**.

3. **Loading a CSV file (Project.csv) into a DataFrame**.

4. **Displaying basic DataFrame attributes like its shape**.

Recommendations

1.Enhance Documentation:

- Add markdown cells to describe the purpose of each code block.
- Explain findings and insights from visualizations.

2. **Improve Code Readability:**

- Use meaningful variable names (e.g., filtered_df could be filtered_loans).
- Add comments explaining key steps, especially the logic behind outlier detection.

3. **Deepen Analysis:**

- Provide statistical summaries or correlations between variables.
- Perform hypothesis testing or modeling based on the dataset's goal.

4. ***Data Validation*:**

- Validate data preprocessing steps and explain why specific transformations are applied.

5. **Interactive Visualizations:**

- Consider using libraries like Plotly or Dash for more interactive exploration.

6. **File Path Flexibility:**

- Dynamically fetch the file path to avoid hardcoding (path="/Project.csv").

Structure of the Notebook-

- **Environment Setup-**
  Mounting Google Drive (drive.mount) to access files. Importing libraries like pandas, numpy, matplotlib, and seaborn.

- **Data Loading-**
Reads a CSV file named Project.csv into a DataFrame.

- **Exploratory Data Analysis (EDA)- Checks:**
  Shape of the dataset (df.shape).
  Dataset info (df.info()).
  Null values (df.isnull().sum()).
  Drops rows with null values (df.dropna()).

- **Data Visualization-**
  Histograms for variables like:
  Classification_Code.
  loan_Term.
  Low_Documentation_Loan.
  Box plots for:
  Loan_Term.
  Business_Numeric.
  Bar chart showing Count_Employees against Demography.

- **Outlier Detection and Handling-**
  Calculates interquartile range (IQR) for Loan_Term.
  Filters the dataset based on IQR boundaries.
  Visualizes filtered data with box plots.

Conclusion

- The Mudra Loan dataset project provides valuable insights into the financial inclusion and credit patterns of micro, small, and medium enterprises (MSMEs) in India.

- By analyzing the data, we identified key trends in loan disbursements, borrower demographics, and repayment behaviors.

- These findings can inform policymakers and financial institutions to better tailor their strategies for promoting economic growth and supporting small businesses.

- Future work can include integrating additional datasets for a more comprehensive analysis and leveraging machine learning for predictive modeling of loan defaults.

References

- **nbformat:** Specifies the major version of the notebook format.

- **nbformat_minor**: Specifies the minor version of the notebook format.

- **metadata**: Contains metadata about the notebook, such as authorship, tools, or environment details

- **cells:** A list of cells, which include the content (code, text, or markdown) and associated outputs or metadata.

Appendix

General Overview-

- **Total Cells: 21**
- **Code Cells: 21**
- **Markdown Cells: 0**
- **Raw Cells: 0**

Details of the First 10 Cells:

- **Cell 1: Code, 0 characters (empty)**

- **Cell 2: Code, 60 characters**

- **Cell 3: Code, 92 characters**

- **Cell 4: Code, 43 characters**

- **Cell 5: Code, 8 characters**

- **Cell 6: Code, 9 characters**

- **Cell 7: Code, 17 characters**

- **Cell 8: Code, 16 characters**

- **Cell 9: Code, 19 characters**