Loan Management Analytics Project

End-to-End Workflow: Python → SQL → Power BI

Prepared by: Krushant Shah

Role: Data Analyst | Python | SQL | Power BI

1. Introduction

This document provides a detailed overview of the Loan Management Analytics Project, which demonstrates
a complete data analysis lifecycle starting from raw data extraction to dashboard development in Power BI.
The project includes the following major activities:

1. Data Cleaning & Transformation using Python
2. Data Loading, Modeling, and Aggregation using MySQL
3. Developing an interactive Power BI dashboard for business insights
4. Visual representation of loan portfolio performance, collections, branch metrics, and DPD analysis

The primary objective of the project is to help financial institutions track loan performance,
understand borrower behavior, measure collection efficiency, and identify risk patterns.

2. Dataset Description

The project uses multiple datasets collected over different months, including:

• Attrition_By_Branch_Data.csv – Tracks branch-level joining, exit, and promotions
• Collection_May_2024 to Collection_Aug_2024 – Monthly collection data
• Portfolio_Report_May_2024 to Portfolio_Report_Aug_2024 – Monthly loan portfolio performance

These datasets include fields such as loan amount, collection amount, DPD (Days Past Due),
branch name, state, customer details, and loan type.

3. Data Cleaning & Transformation (Python)

The data cleaning process was performed in Python (Jupyter Notebook) using Pandas and NumPy libraries.
The major cleaning steps included:

• Removing missing and duplicate records
• Standardizing date formats across multiple datasets
• Converting numeric columns (loan amount, collection amount) into consistent formats
• Merging monthly datasets for a unified structure
• Renaming inconsistent column names
• Generating derived fields such as Total Collection, OD Demand, and Month-Year

4. SQL Database Modeling & Transformations

After cleaning, the data was loaded into a MySQL database for structured storage and advanced querying.
The SQL workflow included:

• Creating tables for each dataset
• Normalizing the data for better relational modeling
• Using JOINS to merge portfolio and collection tables
• Aggregating collection totals, OD (Overdue Demand), and principal balance
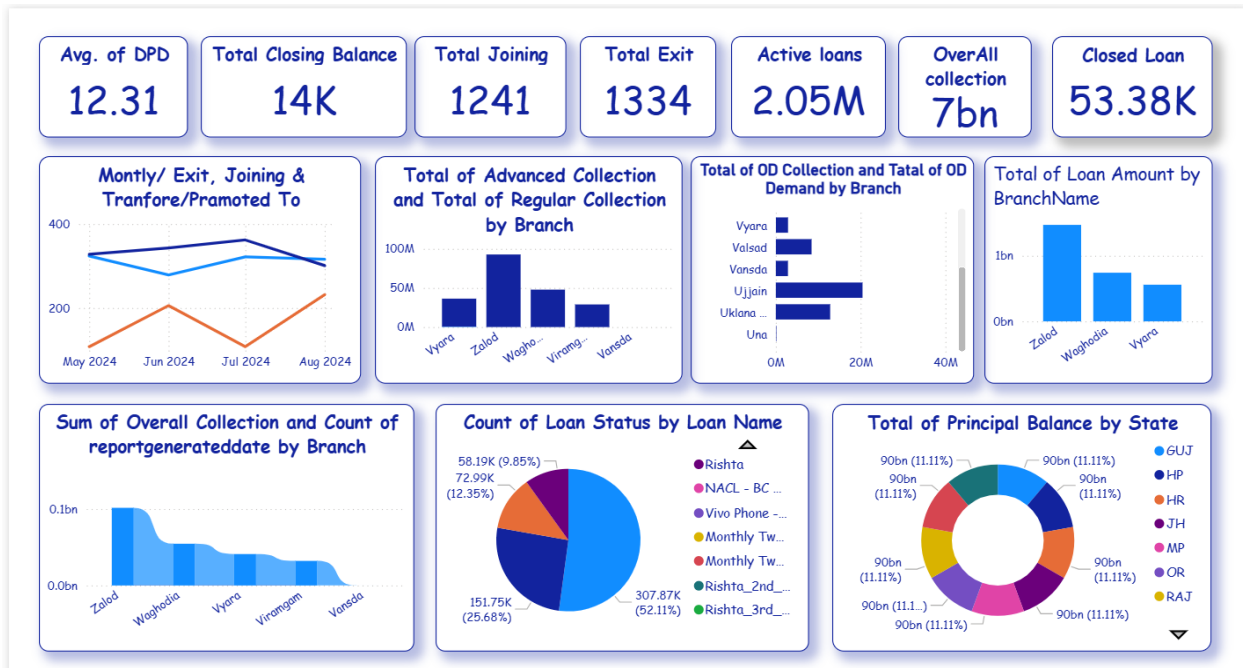• Creating SQL Views for Power BI connectivity

Example SQL query:

```
SELECT branch_name,
SUM(collection_amount) AS total_collection,
SUM(od_demand) AS total_od
FROM collection_data
GROUP BY branch_name;
```

5. Power BI Dashboard & Insights

The Power BI dashboard represents the final analytical output of the project.
It includes visual insights such as:

• Total Closing Balance
• Avg. DPD
• Total Joining & Total Exit metrics
• Overall Collection vs Demand
• Branch-wise Loan Amount
• OD Collection Performance
• Loan Status Distribution (Pie Chart)
• State-wise Principal Balance (Donut Chart)

These visuals help management take data-driven decisions and evaluate loan risk and collection
performance.

6. Key Insights & Business Value

Based on the Power BI analysis:

• Average DPD is 12.31 – within acceptable limits
• Overall collections reached 7bn
• Zalod branch recorded the highest loan amount and collection numbers
• Rishta loan type holds the largest share of loan accounts (52%)
• States such as GUJ, HP, MP show balanced principal distribution

These insights can help refine loan policies and improve collection strategies.

7. Conclusion

This project demonstrates a robust data pipeline integrating Python, SQL, and Power BI to deliver actionable insights for loan management. The workflow enables organizations to track loan performance,
improve collections, minimize risk, and make informed lending decisions.

The modular structure allows easy scaling for new datasets and additional business metrics.