

1. Understanding the data

Convert raw Files into FlatTables

- Understanding the type of data available.
- Examining the meaning of each column.
- Identifying data types (Integer, Text, Date, etc.).
- Analyzing relationships between different tables.

2. Wireframing & Data flow design

- **INTRODUCTION**

Wireframing is the process of creating a visual blueprint for dashboards before actual development. It helps structure data presentation, define user interaction, and ensure clarity in design.

- **Why Use Wireframing for Dashboards?**
- Clarity
- Efficiency
- Collaboration

- **Define Key Metrics**

Key Metrics (KPIs) are quantifiable values that measure performance, helping organizations track progress and make data-driven decisions.

- **Key Metrics Overview**

Total Loan Application :- Total number of loan request. (use card)

Total Loan Sanction approved :- Total number of loan application that have been officially approved. (use card)

Total Loan Application:- Total number of loan application that have been officially.(use card)

Average Loan TAT Processing : - Average time from application to sanction. (use line charts)

Average TAT Disbursement : - average time from sanction to disbursement.(use clustered Bar charts)

Branch-wise TAT performance : - comparison of the time taken to complete a process or across different branches of an organization. (use Bar charts).

Center-Wise TAT performance : - Time taken to complete a process or across different branches of an organization. (use line charts)

Average TAT Sanction :- average time taken for loan sanction. (use line charts)

Average TAT trend over Time : - average time taken to complete a process day, month, quarter, years. (use line charts)

Geographic performance : - Analyze loan Disbursement by state. (use map charts)

Loan processing TAT :- Time taken from submitted date to sanction date. (use Bar charts)

Disbursement TAT :- Time taken from Sanction date to disbursement date. (use Bar charts)

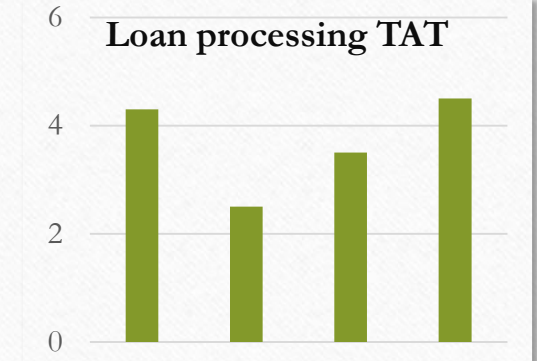
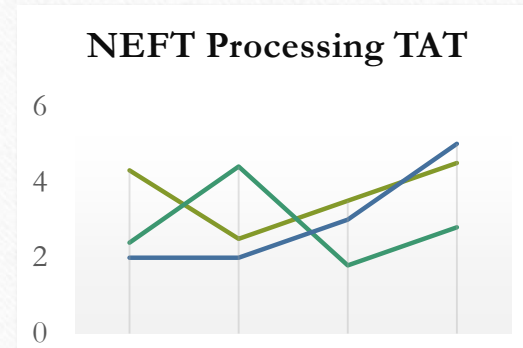
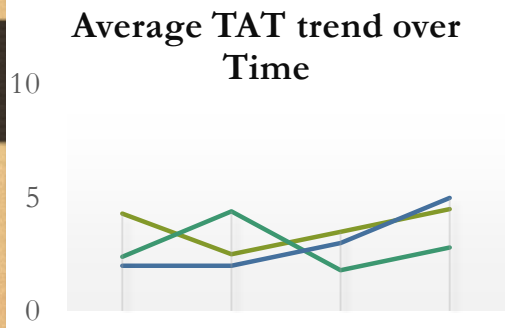
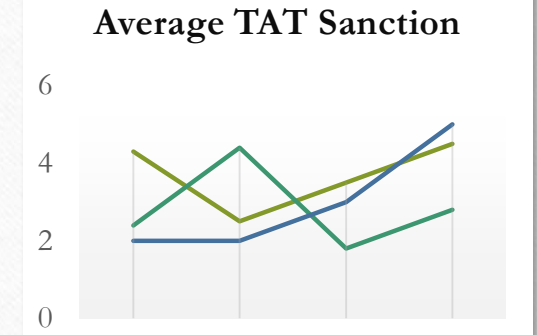
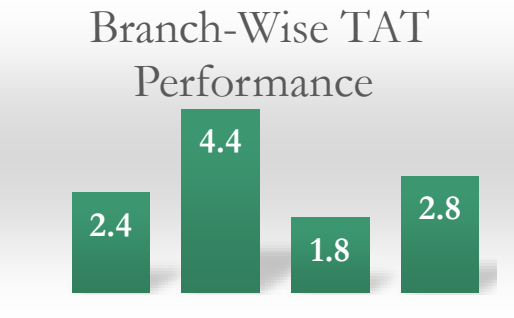
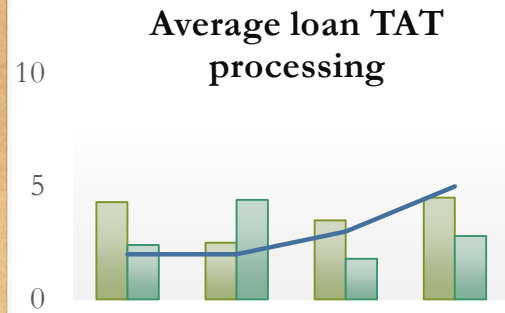
NEFT Processing TAT :- Time from Disbursement date to Transaction date. (use line charts)

Total Loan Application

Total Loan sanction Approved

Average sanction Time

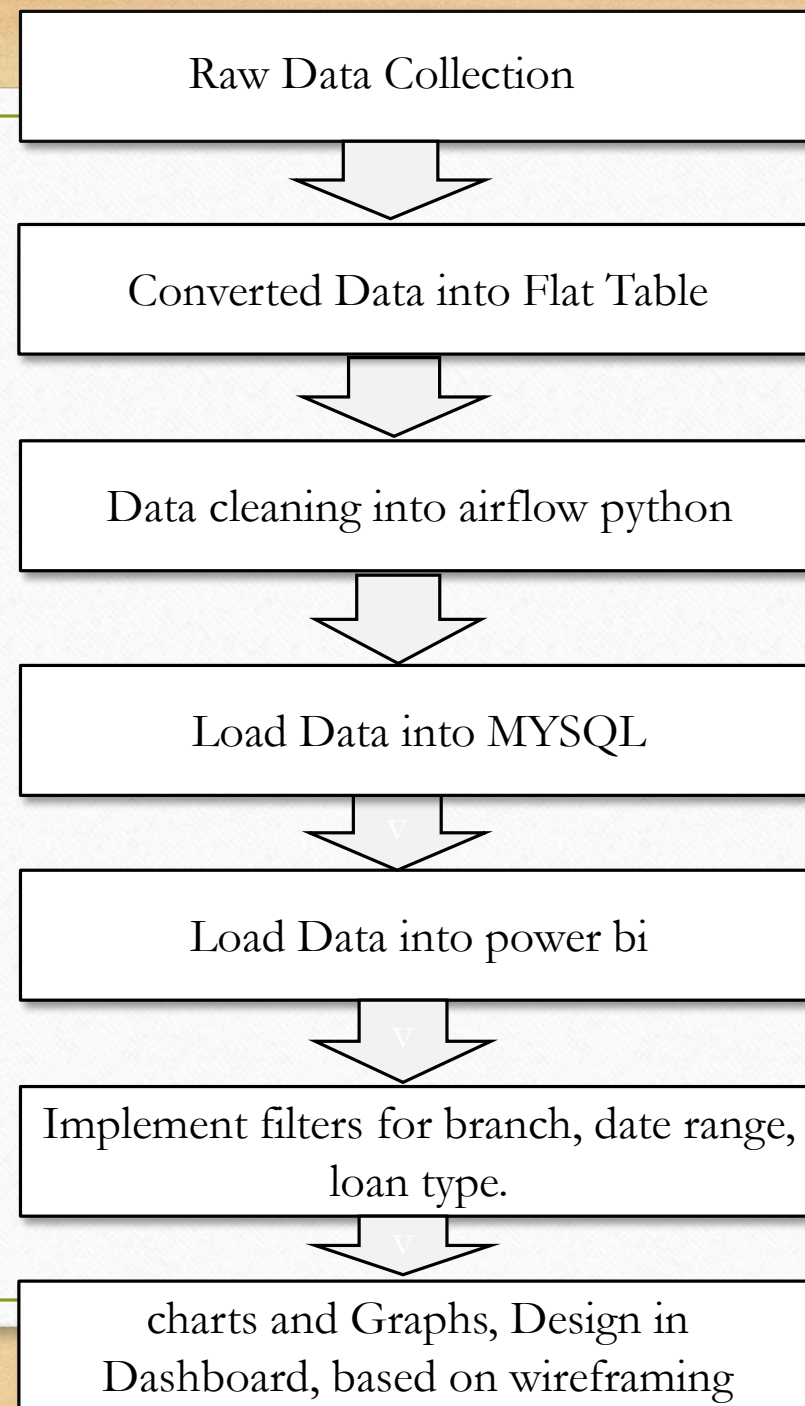
Average Disbursement Time



Disbursement TAT



Data Flow : -



3. data structure

- We created a data flow system and converted raw data into a structured format. The key steps included:
-
- Identifying the primary key for linking multiple datasets. The primary key chosen was **LoanID**.
 - Handling seven different datasets from July, August, and September, including:
 - Branch Info
 - Center Info
 - Client Disbursement Data
 - Eligible and Not Eligible Data
 - High Ticker Data
 - NEFT Data
 - Sanction Data

- Based on the primary key, we structured the data into three main tables:
- **Client Table:** Contains all client-related information.
- **Loan Table:** Includes loan-related details such as Branch Name, State, District, Center Name, Loan Status, Submission Date, Sanction Date, Disbursement Date, Record Date, Transaction Date, Loan Approval Status, and Loan Amount.
- **NEFT Table:** Stores transaction details, including transaction dates, amounts paid, and any additional financial considerations.
- For this stage, **VLOOKUP** was extensively used to ensure data integrity, with **LoanID** serving as the unique identifier.

4. Data Cleaning and Processing in Apache Airflow

A Python script was developed to clean and transform the data using **DAGs in Apache Airflow**. The main operations included:

- Dropping unnecessary columns.
- Using **fillna()** to handle missing values, especially in critical fields where blank values could not be removed.
- Cleaning string data by removing unnecessary characters.

Libraries Used: Pandas, NumPy

Here a snap shot of data successfully running in airflow:-

localhost:8080/home

Airflow

DAGs

Cluster Activity

Datasets

Security

Browse

Admin

Docs

09:34 UTC

AA

DAGs

All 9Active 9Paused 0

Running 0Failed 1

Filter DAGs by tag

Search DAGs

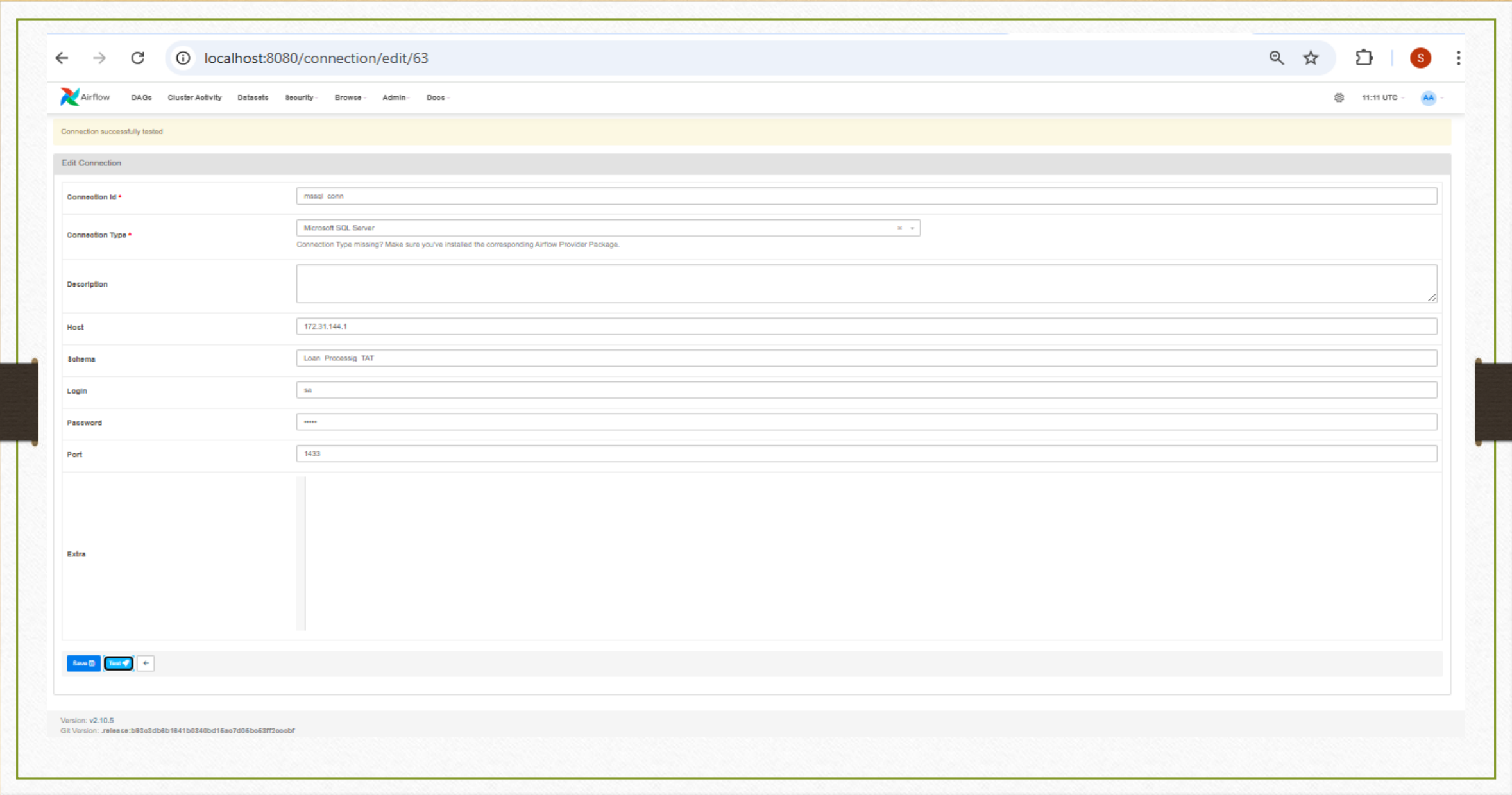
Auto-refresh

DAG	Owner	Runs	Schedule	Last Run	Next Run	Recent Tasks	Actions
<input checked="" type="checkbox"/> client_data_aug24	airflow	<div><div>10</div></div>	@daily	2025-03-12, 00:00:00	2025-03-13, 00:00:00	<div><div>1</div></div>	<div><div></div><div></div><div></div></div>
<input checked="" type="checkbox"/> client_data_processing_jul24	airflow	<div><div>1</div></div>	None	2025-03-03, 07:35:00		<div><div>2</div></div>	<div><div></div><div></div><div></div></div>
<input checked="" type="checkbox"/> client_data_processing_september_2024	airflow	<div><div>10</div></div>	@daily	2025-03-12, 00:00:00	2025-03-13, 00:00:00	<div><div>3</div></div>	<div><div></div><div></div><div></div></div>
<input checked="" type="checkbox"/> loan_data_jul24_processing	airflow	<div><div>2</div></div>	None	2025-03-08, 10:59:46		<div><div>1</div><div>1</div></div>	<div><div></div><div></div><div></div></div>
<input checked="" type="checkbox"/> loan_data_processing	airflow	<div><div>1</div></div>	None	2025-03-08, 10:41:04		<div><div>1</div></div>	<div><div></div><div></div><div></div></div>
<input checked="" type="checkbox"/> loan_data_processing_aug24	airflow	<div><div>10</div></div>	@daily	2025-03-12, 00:00:00	2025-03-13, 00:00:00	<div><div>3</div></div>	<div><div></div><div></div><div></div></div>
<input checked="" type="checkbox"/> loan_data_processing_september_2024	airflow	<div><div>10</div></div>	@daily	2025-03-12, 00:00:00	2025-03-13, 00:00:00	<div><div>3</div></div>	<div><div></div><div></div><div></div></div>
<input checked="" type="checkbox"/> mssql_test_dag	airflow	<div><div>1</div></div>	None	2025-03-08, 07:55:25		<div><div>1</div></div>	<div><div></div><div></div><div></div></div>
<input checked="" type="checkbox"/> my_first_dag	airflow	<div><div></div></div>	None			<div><div></div></div>	<div><div></div><div></div><div></div></div>

5. MSSQL Step up and connection

- Installed the **MSSQL**
- Created the **database schema** and wrote a Python script with DAGs in **VS Code**.
- Established a connection between **MSSQL** (installed on Windows) and **Apache Airflow**.
- Configured the **Airflow connection** through the Admin panel.
- Successfully ran the script, ensuring the data was loaded correctly into **MSSQL**.
- Verified the number of rows and columns to confirm accurate data ingestion.

Here are some snap shot of how we made connections for building the pipeline



6. Power BI Integration with MSSQL

To visualize the data in Power BI, we established a connection with MSSQL sever by following these steps:

1. Downloaded and installed **MSSQL**
2. Open **SQL Server Configuration Manager**
3. Navigate to **SQL Server Network Configuration**
4. Enable **TCP/IP** and restart the SQL Server..
5. Added **MSSQL Server Connection** and configured a connection using:
 - Server Name
 - port
 - Database Username
 - IP Address
 - Password

Opened **Power BI**, navigated to **Get Data** and go to sql server put server name , Database name click ok and
Verified that all data was successfully imported into **Power BI**.

The data pipeline successfully integrates:

Navigator

