**Consolidated-Disbursement-01-Apr-2024-to-31-Mar-2025-Sample-data**

**Data Cleaning and Transformation**

**A. Handle Missing Values**

* **Columns with Missing Values**:
  + Center Name: 164 missing.
  + Group Type: 225 missing.
  + Disbursement Mode: 739 missing.
  + Product Name, Product Code, Loan Purpose: 1,653 missing each.
  + Customer Caste, Customer Religious: 1,653 missing each.
  + Officer Emp Code Number: 164 missing.
* **Problem:** The dataset had missing values in several categorical columns. Missing values can cause issues in data analysis, modeling, and reporting, as they result in incomplete or skewed insights.
* **Affected Columns:**
  + Center Name
  + Group Type
  + Disbursement Mode
  + Product Name
  + Product Code
  + Loan Purpose
  + Customer Caste
  + Customer Religious
  + Officer Emp Code Number

**Cleaning Action Taken:**

* **Approach:**  
  For missing categorical values, we replaced them with the placeholder 'Unknown'. This ensures that the data remains consistent and analyzable without introducing bias or misrepresenting the information.
* **Why 'Unknown'?**
  + It clearly identifies the records where information was unavailable.
  + The placeholder value ensures transparency and flexibility for future updates.

**B. Variations in capitalization and extra whitespace**

* **Problem:**  
  Inconsistent formatting of text data across multiple columns was observed. Variations in capitalization and extra whitespace caused inefficiencies in data analysis and made it challenging to standardize insights.
* **Affected Columns:**
  + Center Name
  + Group Type
  + Customer Name
  + Disbursement Mode
  + Loan Purpose
  + Installment Frequency
  + Customer Caste
  + Customer Religious
  + Type
  + Officer Name
  + Branch
  + Region
  + State
* **Only whitespace removed, No lowercase is applied on the below colunms:**
* Product Name
* Product Code
* Center Code

**Cleaning Action Taken:**

* **Approach:**  
  For the specified columns, we applied two standardization techniques:
  1. **Stripping Whitespaces:**
     + Removed leading and trailing spaces from text data to eliminate formatting inconsistencies.
  2. **Converting to Lowercase:**
     + Transformed all text values to lowercase to ensure uniformity in capitalization.

**Why Was This Necessary?**

1. **Uniformity in Data:**  
   Standardizing the text values ensures that rows with the same logical value (e.g., New York and new york) are treated identically.
2. **Improved Analysis:**  
   Consistent data formatting avoids redundant categories during grouping, filtering, or reporting.

**C.** **Some rows had values such as 12 or 24 (numeric), while others had 12 Months or 24 Months (textual)**.

**Problem:**  
The column **'loan tenure'** contained inconsistent data formatting. Some rows had values such as 12 or 24 (numeric), while others had 12 Months or 24 Months (textual). This inconsistency can cause errors in analysis, aggregation, and visualization, especially in tools like Power BI.

**Cleaning Action Taken:**

1. **Standardizing Column Names:**
   * Stripped extra spaces from column names to ensure uniformity and prevent errors during referencing or processing.
   * Example: A column named 'loan tenure ' was cleaned to loan tenure.
2. **Standardizing the 'loan tenure' Column:**
   * Converted the **'loan tenure'** column to string format to handle mixed numeric and textual values consistently.
   * Removed any leading or trailing spaces in the column to eliminate formatting inconsistencies.
   * Replaced numeric values (12 and 24) with their standardized textual equivalents (12 Months and 24 Months) to ensure uniform representation.

**Why Was This Necessary?**

1. **Data Uniformity:**  
   By having all values in the same format (e.g., 12 Months), we ensure consistent analysis, reporting, and visualization.
2. **Avoid Confusion:**  
   Mixed data formats can cause errors in filters, groupings, and calculations, especially in tools that depend on clean, structured data.
3. **Enhanced Readability:**  
   Representing loan tenure as 12 Months or 24 Months is clearer and more intuitive for stakeholders.

**Key Benefits of the Cleaning Process:**

1. **Improved Data Quality:**  
   Ensures that all entries in the **'loan tenure'** column are standardized, free of ambiguity, and ready for further use.
2. **Ease of Reporting:**  
   Power BI and similar tools can now process this column without errors or manual adjustments.

**D. DateTime Format**

**Problem:**  
The dataset contains columns such as **'disbursement date'** and **'first repayment date'**, which store date values. However, these columns might not be in a proper datetime format and could include invalid or inconsistent entries. This can cause challenges during analysis or reporting, especially when performing date-based calculations, filtering, or visualizations.

**Cleaning Action Taken:**

1. **Identified Date Columns:**
   * Targeted the following columns for cleaning:
     + **'disbursement date'**
     + **'first repayment date'**
2. **Converted Columns to Datetime Format:**
   * Applied pd.to\_datetime() to these columns to convert them into a consistent datetime format.

**Why Was This Necessary?**

1. **Data Consistency:**  
   Standardizing date formats ensures uniformity, making it easier to perform date-based operations like sorting, filtering, or aggregations.
2. **Facilitating Analysis:**  
   Date columns in proper datetime format are essential for calculations such as the time difference between two dates, identifying trends over time, or generating visualizations in tools like Power BI.

**Key Benefits of the Cleaning Process:**

1. **Improved Usability:**  
   Standardized date columns are now ready for direct use in analytics, dashboards, or reports.
2. **Enhanced Functionality:**  
   Ensures compatibility with date-related operations and functions in Python, Power BI, or other analytics tools.

**E**. **Disbursement Year** **and** **Effective Interest**

* **Objective:**  
  To enrich the dataset by adding two derived columns, **'disbursement year'** and **'effective interest'**, which provide additional insights and facilitate better analysis.

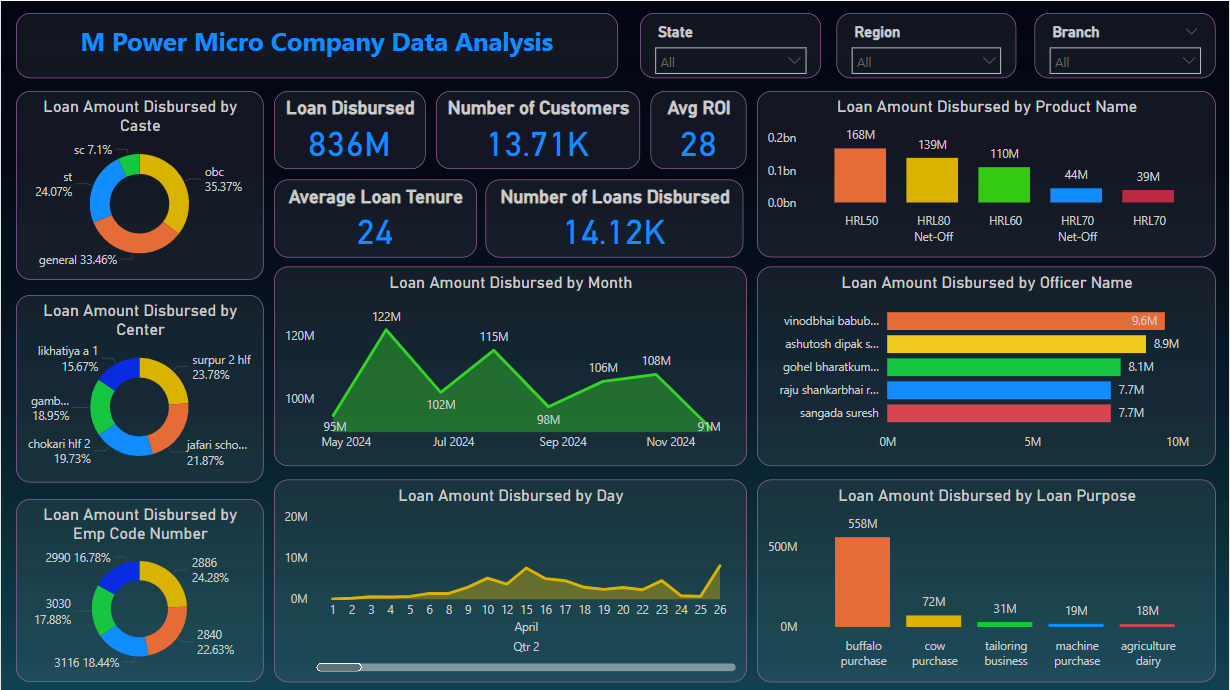
**Enhancement Actions Taken:**

1. **Derived the Year of Disbursement:**
   * **Column Added:**
     + **'disbursement year'**
2. **Calculated Effective Interest:**
   * **Column Added:**
     + **'effective interest'**
   * **Purpose:**
     + Provides a direct monetary value of the interest amount for each loan. This helps in understanding the financial impact of loans and evaluating revenue from interest.

**Why Was This Necessary?**

1. **Actionable Insights:**
   * The **'disbursement year'** allows for time-series analysis, enabling the identification of patterns or shifts in loan distribution over years.
   * The **'effective interest'** provides clarity on the financial value of interest generated for each loan.
2. **Power BI Integration:**
   * Including these columns ensures easier visualization and deeper insights in the dashboard, especially for revenue analysis or trend visualization.

**Dashboard page – 1**

****

**Explanation of Dashboard page -1**

### ****1. Loan Amount Disbursed by Caste (Pie Chart)****

This pie chart displays the distribution of the total loan amount disbursed across different caste categories:

* **SC**: 7.1% of the total loans.
* **ST**: 24.07%.
* **OBC**: 35.37%.
* **General**: 33.46%. This provides insights into which caste groups received the majority of loans, indicating the company's inclusivity or outreach strategy.

### ****2. Loan Disbursed, Number of Customers, Avg ROI (KPI Metrics)****

* **Loan Disbursed**: 836M (total loan amount disbursed by the company).
* **Number of Customers**: 13.71K (total number of customers who received loans).
* **Avg ROI**: 28 (average return on investment for the loans provided). These KPIs give a quick snapshot of the company's loan disbursement performance.

### ****3. Average Loan Tenure and Number of Loans Disbursed (KPI Metrics)****

* **Average Loan Tenure**: 24 months (average duration of loans).
* **Number of Loans Disbursed**: 14.12K (total number of loans provided). These metrics highlight the tenure and frequency of loan disbursements.

### ****4. Loan Amount Disbursed by Product Name (Bar Chart)****

The bar chart shows the loan amount disbursed for different loan products:

* **HRL50**: 168M.
* **HRL80 Net-Off**: 139M.
* **HRL60**: 110M.
* **HRL70 Net-Off**: 44M.
* **HRL70**: 39M. This chart highlights the popularity or demand for specific loan products.

### ****5. Loan Amount Disbursed by Center (Pie Chart)****

This chart breaks down the loan amounts disbursed across different centers:

* **Surpur 2 Hlf**: 23.78%.
* **Jafari School**: 21.87%.
* **Gamb.**: 18.95%.
* **Chokari Hf 2**: 19.73%.
* **Likhatiya A 1**: 15.67%. This helps in understanding the regional performance of loan disbursement.

### ****6. Loan Amount Disbursed by Month (Line Chart)****

This line chart depicts the trend of loan amounts disbursed across months:

* **May 2024**: 95M.
* **July 2024**: 122M (peak).
* **September 2024**: 115M.
* **November 2024**: 108M. The chart reveals seasonal or monthly variations in loan disbursement trends.

### ****7. Loan Amount Disbursed by Officer Name (Bar Chart)****

This bar chart lists officers who disbursed the highest loan amounts:

* **Vinodbhai Baubhai**: 9.6M.
* **Ashutosh Dipak S**: 8.9M.
* **Gohel Bharatkumar**: 8.1M.
* **Raju Shankarbhai**: 7.7M.
* **Sangada Suresh**: 7.7M. This chart provides insights into employee performance.

### ****8. Loan Amount Disbursed by Emp Code Number (Pie Chart)****

This pie chart shows the loan amounts distributed by employee codes:

* **2986 (Top Employee)**: 24.28%.
* Other notable contributors include employees with codes **3030 (17.89%)**, **3116 (18.44%)**, and **2840 (22.63%)**. It allows identification of high-performing employees.

### ****9. Loan Amount Disbursed by Day (Line Chart)****

This line chart visualizes daily trends in loan disbursements:

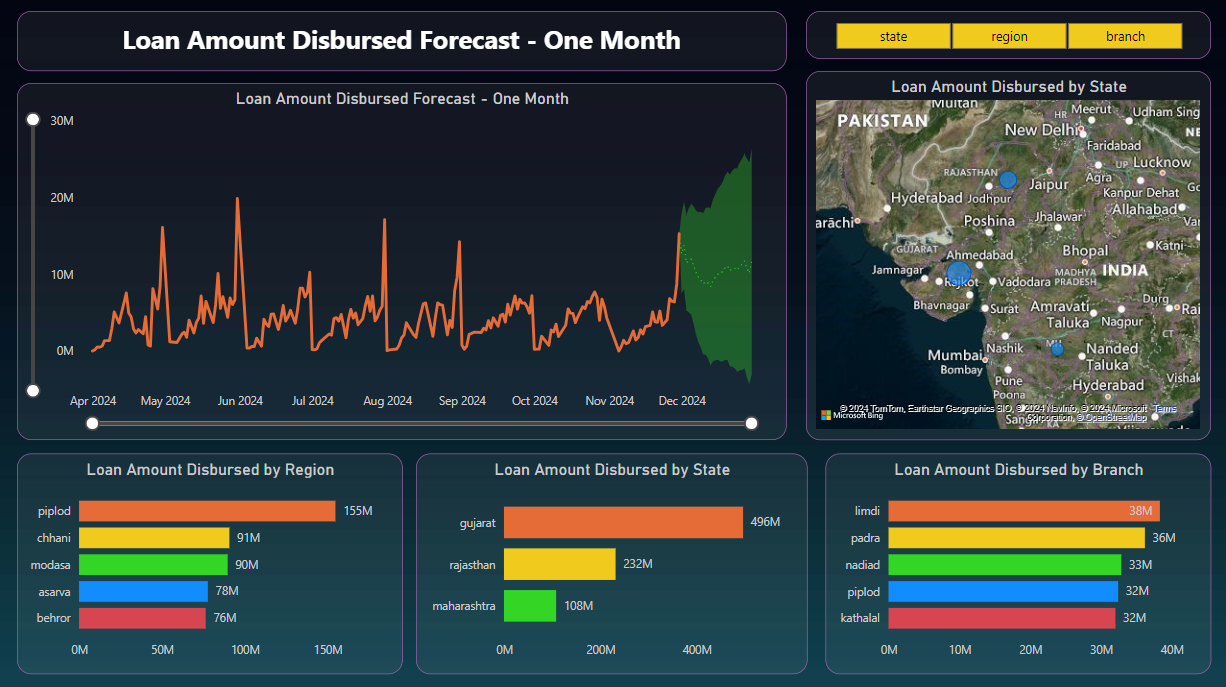
* Loan amounts start low in early days of the quarter, with peaks around specific days.
* April shows more uniform disbursement trends with slight spikes. This can indicate patterns tied to customer behavior or operational workflows.

### ****10. Loan Amount Disbursed by Loan Purpose (Bar Chart)****

The bar chart represents loan disbursements by purpose:

* **Buffalo Purchase**: 558M (majority of the loans).
* **Cow Purchase**: 72M.
* **Tailoring Business**: 31M.
* **Machine Purchase**: 19M.
* **Agriculture Dairy**: 18M. It reveals the primary sectors or purposes for which customers seek loans.

**Dashboard page – 2**

****

**Explanation of Dashboard page -2**

### ****1. Loan Amount Disbursed Forecast - One Month (Line Chart)****

* **Purpose**: This chart shows the loan disbursement trend over time and includes a forecast for the next month (December 2024).
* **Details**:
  + The orange line represents the historical loan disbursement from April 2024 to November 2024.
  + Peaks are noticeable in May, June, and November, indicating higher loan disbursements during these months.
  + The shaded green area represents the forecasted range for December 2024. The dotted green line shows the most likely trajectory for loan disbursement in the coming month.

### ****2. Loan Amount Disbursed by State (Map)****

* **Purpose**: This map visualizes loan disbursements geographically across states in India.
* **Details**:
  + The size of the blue bubbles corresponds to the total loan amount disbursed in each state.
  + Larger bubbles indicate states with higher loan disbursements, such as **Gujarat**.
  + This chart highlights the regional distribution of loans and allows for geographic analysis of loan performance.

### ****3. Loan Amount Disbursed by Region (Bar Chart)****

* **Purpose**: Displays the total loan disbursement for specific regions.
* **Details**:
  + **Piplod** leads with the highest disbursement at **155M**, followed by **Chhani (91M)** and **Modasa (90M)**.
  + Other notable regions include **Asarva (78M)** and **Behror (76M)**.
  + This chart provides a granular look at loan disbursements at a regional level.

### ****4. Loan Amount Disbursed by State (Bar Chart)****

* **Purpose**: Highlights the total loan disbursement for different states.
* **Details**:
  + **Gujarat**: 496M (highest disbursement).
  + **Rajasthan**: 232M.
  + **Maharashtra**: 108M.
  + This chart reinforces the map visualization and offers exact figures for each state.

### ****5. Loan Amount Disbursed by Branch (Bar Chart)****

* **Purpose**: Displays the loan disbursement at the branch level.
* **Details**:
  + The **Limdi branch** disbursed the most loans (38M), followed by **Padra (36M)** and **Nadiad (33M)**.
  + Branches such as **Piplod (32M)** and **Kathlal (32M)** also show significant contributions.
  + This chart provides insights into branch-level performance and allows for comparisons among branches.

### Key Insights from the Dashboard:

1. **Gujarat** is the leading state in terms of loan disbursements, contributing significantly to the overall total.
2. The **Piplod region** is a high-performing region, with the most loan disbursements across regions.
3. The **Limdi branch** emerges as the top-performing branch, with the highest loan disbursements among all branches.
4. Loan disbursement shows periodic spikes, with a forecast indicating potential growth in December 2024.
5. The geographic map adds a visual layer for identifying high-performing states and areas needing improvement.

**Consolidated Pos as on 30 Nov 2024 (Sample Data)**

**A. Handling Missing Value  
 Columns with Missing Value:**

* **Officer Employee Number : 9,553**
* **Disb Type : 9,553**
* **Cycle Number : 295**

**Problem:** The Disb Type column contained missing values, which could lead to inconsistencies and inaccuracies in data analysis. Having incomplete data made it difficult to properly classify and analyze the distribution of disbursement types, potentially skewing results and insights.

**Affected Column:**

* **Disb Type**

**Cleaning Action Taken:**

* **Approach:** For the Disb Type column, the following action was taken:
  1. **Filling Missing Values:**
     + The missing values (NaN) in the Disb Type column were replaced with the most frequent value, "Normal", to ensure completeness of the data.

**Why Was This Necessary?**

1. **Data Completeness:**
   * Missing values in the dataset can cause gaps in analysis, making it difficult to draw accurate insights. By filling in the missing values, we ensured that each row in the dataset had a valid Disb Type, allowing for more reliable analysis.
2. **Consistency in Reporting:**
   * The "Normal" category is the most frequent disbursement type, representing the majority of the data. By assigning "Normal" to missing values, we maintained consistency with the distribution of disbursement types.
3. **Accurate Insights:**
   * With missing values filled, any analysis, aggregation, or grouping by the Disb Type column can now be performed without the risk of missing data affecting results, leading to more accurate insights.

**B. Handling Missing Value**

**Problem:** The Cycle Number column had missing values (NaN), which could lead to incomplete data analysis and potentially misclassify entries. The missing values needed to be addressed to maintain consistency and accuracy across the dataset.

**Affected Column:**

* **Cycle Number**

**Cleaning Action Taken:**

* **Approach:** For the Cycle Number column, the following action was taken:
  1. **Filling Missing Values:**
     + The missing values (NaN) in the Cycle Number column were replaced with the most common cycle number, 1.0, to ensure completeness of the data and maintain consistency.

**Why Was This Necessary?**

1. **Data Completeness:**
   * Missing values in the Cycle Number column could lead to gaps in the analysis, making it difficult to classify and evaluate the data accurately. By filling the missing values with 1.0, we ensured that all rows had a valid cycle number, allowing for a more complete dataset.
2. **Consistency in Analysis:**
   * The cycle number 1.0 was chosen because it is the most frequent value in the dataset, appearing in the majority of the records. Using this value for missing entries ensures that the dataset reflects the most common cycle number, minimizing the impact of missing data.
3. **Improved Accuracy in Reporting:**
   * With missing values addressed, any analysis or calculations based on the Cycle Number column can now be carried out without worrying about the influence of missing data, resulting in more reliable and accurate insights.

**C. Handling Missing Value**

**Problem:** The Officer Employee Number column contained missing values (NaN), which could lead to gaps in data analysis. It was crucial to handle these missing values to maintain the integrity and consistency of the dataset.

**Affected Column:**

* **Officer Employee Number**

**Cleaning Action Taken:**

* **Approach:** For the Officer Employee Number column, the following action was taken:
  1. **Filling Missing Values:**
     + The missing (NaN) values in the Officer Employee Number column were replaced with the string 'Unknown' to fill the gaps and ensure that the column does not have any missing data.

**Why Was This Necessary?**

1. **Data Completeness:**
   * Missing values in the Officer Employee Number column could cause inconsistencies in analysis or make it difficult to identify specific officers. By replacing NaN values with 'Unknown', we ensured that every entry has a valid value, avoiding incomplete data.
2. **Handling Null Values Consistently:**
   * In cases where an officer's employee number is not available, the value 'Unknown' serves as a placeholder, signaling the absence of a specific officer. This placeholder ensures that missing data does not disrupt the analysis process and provides a consistent approach to handling null values.
3. **Enabling Accurate Reporting:**
   * With no missing values, it becomes easier to generate reports, analyze officer-related metrics, and perform any further operations on the Officer Employee Number column without concerns about missing data affecting the outcomes.

**D. Handling Officer Name and Id**

**Problem:** The Officer Name column contained both valid officer names and potential officer IDs, which could cause confusion and affect data analysis. Officer names and IDs follow different patterns, so it was important to identify and filter out rows where the Officer Name was not a valid name.

**Affected Column:**

* **Officer Name**

**Why Was This Necessary?**

1. **Data Accuracy:**
   * Officer IDs are different from officer names, and mixing the two could result in misleading data analysis or incorrect reporting. Filtering out rows with IDs ensures that the data only contains valid officer names, making it more accurate and reliable for analysis.
2. **Consistency in Data:**
   * By ensuring that the Officer Name column only contains names and not IDs, we standardized the data. This consistency improves the quality of reporting and reduces the likelihood of errors in downstream analysis.
3. **Improved Data Usability:**
   * Keeping only rows with officer names enhances the usability of the dataset, especially when generating reports or conducting detailed analysis involving officer-related data. The dataset becomes more focused and relevant to specific use cases, like identifying officer performance or creating officer-specific reports.

**Result:** This data cleaning step ensures that the Officer Name column now contains only valid officer names, making the dataset cleaner and ready for accurate and meaningful analysis.

**E. Date & Time**

**Problem:** The dataset contained several columns with date values stored as strings, which can create inconsistencies and errors when performing date-related analysis or calculations. These string values could contain incorrect formats or invalid entries, leading to potential issues when working with time-based analysis in downstream processes.

**Affected Columns:**

* **Disbursement Date**
* **First Installment Date**
* **Last Installment Date**

**Cleaning Action Taken:**

* **Approach:** To standardize and ensure proper handling of date values, the following steps were implemented:
  1. **Converting String Dates to Datetime Format**
  2. **Handling Invalid Date Entries**
  3. **Ensuring Data Consistency**

**Why Was This Necessary?**

1. **Consistency in Date Format:**
2. **Efficient Date Handling:**
3. **Handling Invalid Data:**

**Result:** This data cleaning step standardizes the date columns, ensuring that all date values are in the correct datetime format, and invalid date entries are safely handled. This improves the dataset’s consistency, facilitates accurate date-based analysis, and prevents errors related to improperly formatted date values.

**F. Stripping and Lowercasing**

**Problem:** The dataset contains multiple columns with text-based data, and inconsistencies in formatting were observed. These inconsistencies include leading or trailing whitespaces and variations in capitalization. Such formatting discrepancies can hinder analysis, cause inconsistencies in grouping or filtering, and result in inaccurate insights.

**Affected Columns:**

* **Officer Name**
* **Center Name**
* **Type**
* **OG Branch Name**
* **State**
* **Region**
* **Urban/Rural**
* **District**
* **Own/ Managed**
* **Agri / Non Agri**
* **Haryali / Samrudhi**
* **Disb Type**
* **Pre-Post**
* **Billing / Matured**

**Cleaning Action Taken:**

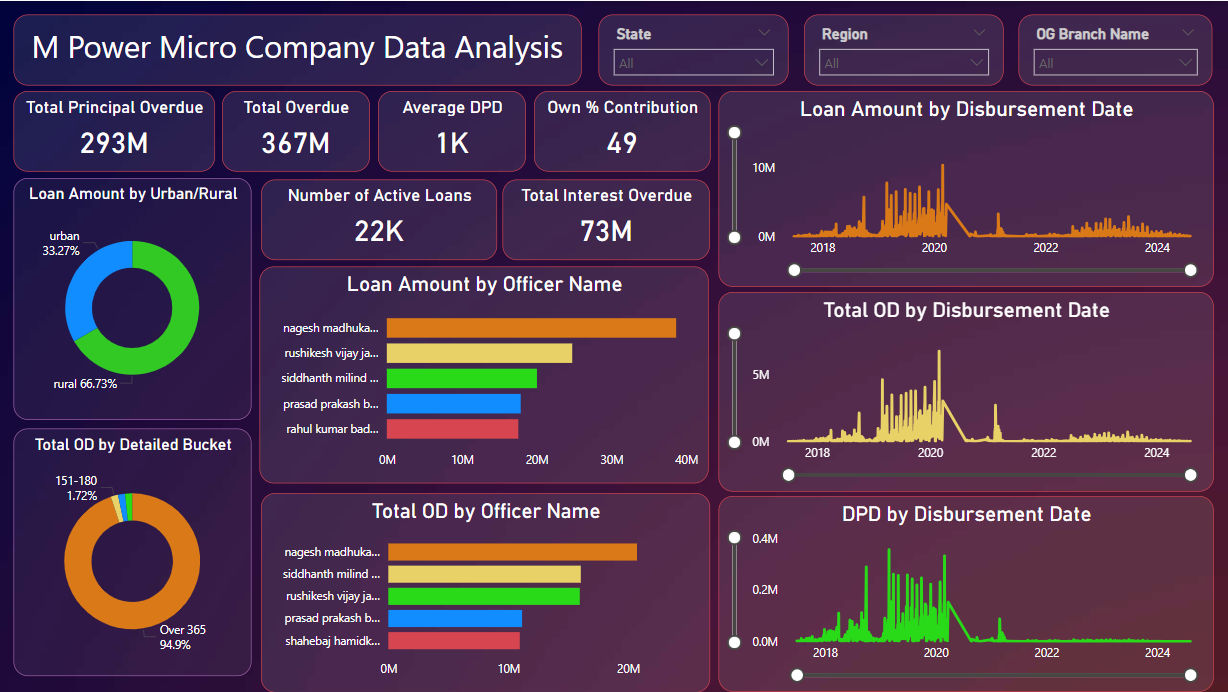
* **Approach:** To standardize the text values in the specified columns, the following steps were applied:
  1. **Stripping Leading and Trailing Whitespaces:**
     + The str.strip() method was used to remove any leading or trailing spaces from the text values in the specified columns. This ensures that any extra spaces that may have been inadvertently added during data entry or data extraction are eliminated.
  2. **Converting to Lowercase:**
     + The str.lower() method was applied to transform all text values to lowercase. This standardizes the capitalization, ensuring that variations like "New York" and "new york" are treated as the same value.
  3. **Ensuring Data Consistency:**
     + Both the stripping of whitespaces and the conversion to lowercase were applied to each of the affected columns, ensuring uniformity across the dataset and removing inconsistencies that could affect further analysis.

**Why Was This Necessary?**

1. **Consistency in Formatting:**
   * Text data with extra spaces or inconsistent capitalization can result in multiple entries for what should be the same value. For example, "New York" and " new york " would be treated as different values unless these issues are addressed. By stripping the whitespaces and converting text to lowercase, all values are standardized, improving data consistency.
2. **Improved Data Grouping and Analysis:**
   * Inconsistent formatting (such as variations in capitalization or extra spaces) could lead to issues when grouping, filtering, or aggregating data. For instance, grouping by state name might create multiple groups for the same state due to differences in formatting. Standardizing the text ensures accurate grouping and analysis.
3. **Data Quality Improvement:**
   * Cleaning the data by removing extra spaces and ensuring consistent capitalization improves the overall quality of the dataset. It reduces the chance of errors and ensures that insights derived from the dataset are based on accurate, uniform data.

**Result:** This cleaning step standardizes the text columns by removing extraneous spaces and ensuring all text values are in lowercase. The result is a cleaner, more consistent dataset, which enables accurate grouping, filtering, and analysis of categorical variables.

**Dashboard page – 1**



**Explanation of Dashboard page -1**

### ****1. Overview Metrics (Top Row Cards)****

1. **Total Principal Overdue**:
   * Shows the amount of principal overdue, totaling **293M**, indicating loans not paid back on time.
2. **Total Overdue**:
   * Represents the total overdue amount, which is **367M**, combining principal and interest.
3. **Average DPD (Days Past Due)**:
   * Indicates the average number of days borrowers have delayed their payments, which is **1K days**.
4. **Own % Contribution**:
   * A metric that quantifies a specific contribution, with a value of **49%**. This could represent the company’s direct involvement in loan disbursement or collections.
5. **Number of Active Loans**:
   * Indicates the total number of ongoing loans, which is **22K**.
6. **Total Interest Overdue**:
   * Shows the overdue interest amount, totaling **73M**, highlighting unpaid interest across active and overdue loans.

### ****2. Loan Amount by Urban/Rural (Pie Chart)****

* **Purpose**: Displays the loan distribution across urban and rural regions.
* **Details**:
  + Urban areas account for **33.27%**, while rural areas dominate with **66.73%**.
  + This shows a higher focus on rural lending, aligning with financial inclusion goals.

### ****3. Total OD by Detailed Bucket (Pie Chart)****

* **Purpose**: Visualizes overdue loans categorized by days past due.
* **Details**:
  + Loans overdue for **over 365 days** make up **94.9%**, indicating long-term overdue loans dominate the portfolio.
  + Loans overdue for **151–180 days** constitute only **1.72%**, showing fewer loans in this medium-term category.

### ****4. Loan Amount by Officer Name (Bar Chart)****

* **Purpose**: Shows the loan disbursements handled by each officer.
* **Details**:
  + **Nagesh Madhukar** disbursed the most loans, nearly **40M**.
  + Other officers like **Rushikesh Vijay** and **Siddhant Milind** follow with approximately **20M** each.
  + This chart highlights individual officer contributions to the overall disbursement.

### ****5. Total OD by Officer Name (Bar Chart)****

* **Purpose**: Displays overdue loans attributed to each officer.
* **Details**:
  + **Nagesh Madhukar** also has the highest overdue loans, reflecting either high-risk lending or larger loan portfolios.
  + Other officers like **Siddhant Milind** and **Rushikesh Vijay** show moderate overdue amounts.

### ****6. Loan Amount by Disbursement Date (Line Chart)****

* **Purpose**: Shows the trend of loan disbursements over time.
* **Details**:
  + A steady increase in loan disbursement is visible from 2018 to 2020, peaking during that period.
  + There is a noticeable decline after 2020, with smaller disbursements in subsequent years.

### ****7. Total OD by Disbursement Date (Line Chart)****

* **Purpose**: Displays overdue amounts over time.
* **Details**:
  + Overdue loans show a similar trend, peaking around 2020 and declining thereafter.
  + The spike aligns with higher disbursement periods, indicating challenges in repayment during those years.

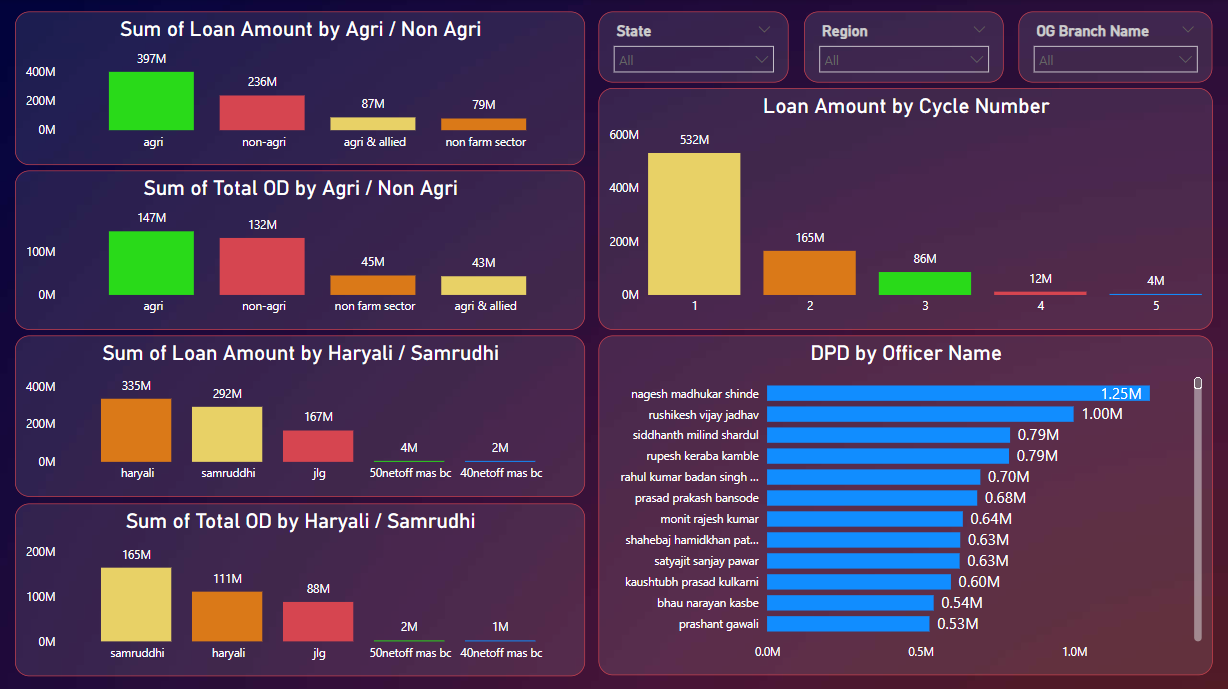
### ****8. DPD by Disbursement Date (Line Chart)****

* **Purpose**: Tracks the number of days past due (DPD) over time.
* **Details**:
  + Peaks in DPD occur around 2020, consistent with high overdue amounts during this period.
  + The decline in later years suggests improved repayment patterns or reduced disbursement volumes.

### Key Insights:

1. **High Rural Focus**: Loans are predominantly disbursed in rural areas, making up two-thirds of the total portfolio.
2. **Overdue Challenges**: A significant portion of overdue loans (94.9%) exceeds 365 days, indicating long-term repayment issues.
3. **Officer Performance**: While officers like Nagesh Madhukar contribute significantly to disbursement, they also handle a large share of overdue loans.
4. **Peak Periods**: The years around 2020 witnessed high disbursement and overdue amounts, possibly due to economic conditions.
5. **Repayment Trends**: The decline in overdue amounts and DPD post-2020 indicates an improvement in loan recovery processes.

**Dashboard page – 2**

****

**Explanation of Dashboard page -2**

**1. Sum of Loan Amount by Agri / Non Agri**

* **Categories**:
  + **Agri**: Represented in green with a loan amount of **397M**.
  + **Non-Agri**: Represented in red with a loan amount of **236M**.
  + **Agri & Allied**: Represented in yellow with a loan amount of **87M**.
  + **Non-Farm Sector**: Represented in orange with a loan amount of **79M**.
* **Insights**:
  + Agricultural loans dominate the portfolio, with a significantly higher loan amount compared to Non-Agri.
  + The Agri & Allied and Non-Farm Sector categories are subcategories with smaller contributions.

**2. Sum of Total OD by Agri / Non Agri**

* **Categories**:
  + **Agri**: Represented in green with an Overdue (OD) amount of **147M**.
  + **Non-Agri**: Represented in red with an OD amount of **132M**.
  + **Non-Farm Sector**: Represented in orange with an OD amount of **45M**.
  + **Agri & Allied**: Represented in yellow with an OD amount of **43M**.
* **Insights**:
  + Total OD for Agri loans is higher than Non-Agri loans, showing that overdue payments are more concentrated in the agricultural sector.
  + Subcategories (Non-Farm Sector and Agri & Allied) have relatively lower overdue amounts.

**3. Sum of Loan Amount by Haryali / Samrudhi**

* **Categories**:
  + **Haryali**: Represented in orange with a loan amount of **335M**.
  + **Samrudhi**: Represented in yellow with a loan amount of **292M**.
  + **Jig**: Represented in red with a loan amount of **167M**.
  + Minor categories such as **50netoff mas bc** and **40netoff mas bc** have minimal contributions of **4M** and **2M**, respectively.
* **Insights**:
  + Loans under the Haryali scheme have the highest share, followed closely by Samrudhi.
  + The Jig category has a moderate contribution, while others are negligible.

**4. Sum of Total OD by Haryali / Samrudhi**

* **Categories**:
  + **Samrudhi**: Represented in yellow with an OD amount of **165M**.
  + **Haryali**: Represented in orange with an OD amount of **111M**.
  + **Jig**: Represented in red with an OD amount of **88M**.
  + Minor categories such as **50netoff mas bc** and **40netoff mas bc** have minimal OD contributions of **2M** and **1M**, respectively.
* **Insights**:
  + Samrudhi loans exhibit the highest overdue payments, indicating potential risks in this scheme.
  + Haryali loans also have significant overdue amounts, but less than Samrudhi.

**5. Loan Amount by Cycle Number**

* **Categories**:
  + **Cycle 1**: Represented in yellow with a loan amount of **532M**.
  + **Cycle 2**: Represented in orange with a loan amount of **165M**.
  + **Cycle 3**: Represented in green with a loan amount of **86M**.
  + **Cycle 4**: Represented in red with a loan amount of **12M**.
  + **Cycle 5**: Represented in purple with a loan amount of **4M**.
* **Insights**:
  + The majority of loans are in Cycle 1, reflecting new loans or loans in the initial stages.
  + Loan amounts decrease significantly as cycle numbers increase, indicating lower retention or refinancing in subsequent cycles.

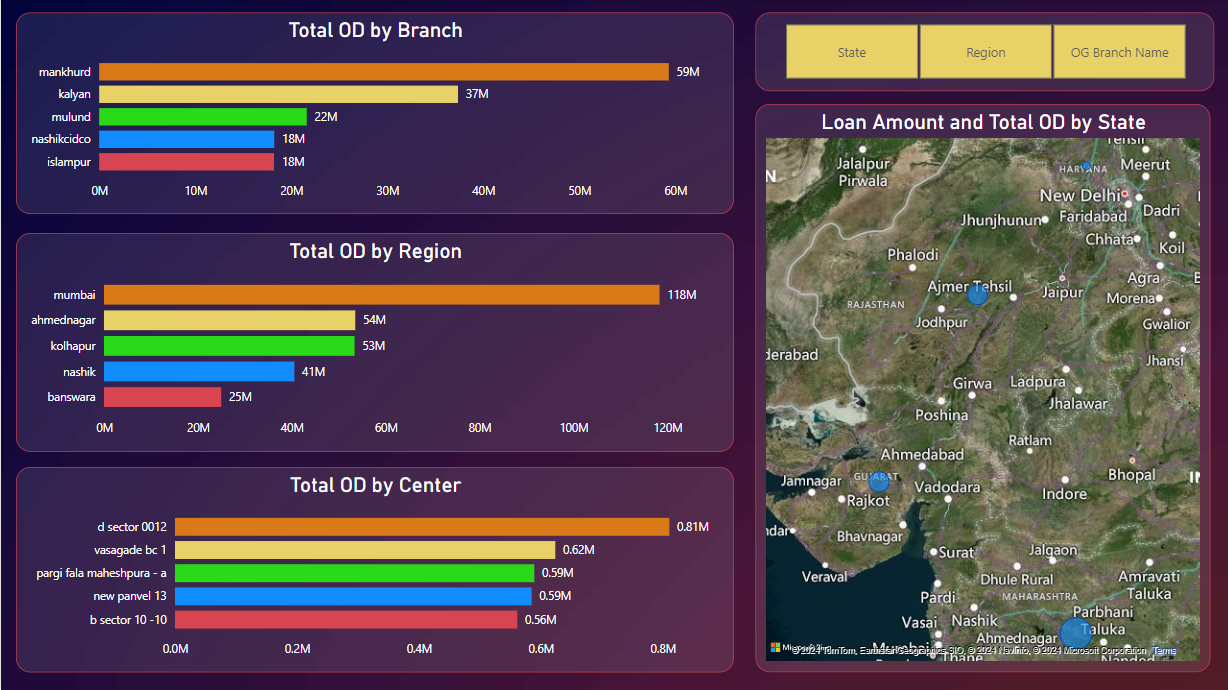
**6. DPD by Officer Name**

* **Details**:
  + The chart lists the Days Past Due (DPD) values by officer name, with the highest being **Nagesh Madhukar Shinde (1.25M)**.
  + Other officers such as Rishikesh Vijay Jadhav and Siddhanth Milind Shardul also have significant DPD values (1.00M and 0.79M, respectively).
* **Insights**:
  + Nagesh Madhukar Shinde is associated with the highest overdue amount, indicating the need for closer monitoring.
  + Officers with lower DPD values may reflect better performance or less risky loan portfolios.

**Filters (State, Region, OG Branch Name)**

* These dropdown filters allow users to refine the data based on:
  + **State**: Select loans or overdue data for a specific state.
  + **Region**: Filter data by geographic region.
  + **OG Branch Name**: Drill down into data at the branch level.

**Dashboard page – 3**

****

**Explanation of Dashboard page -3**

**1. Total OD by Branch**

* **Bars**:
  + **Mankhurd**: Highest overdue amount at **59M** (orange).
  + **Kalyan**: Second highest at **37M** (yellow).
  + **Mulund**: Moderate overdue amount of **22M** (green).
  + **Nashikcidco** and **Islampur**: Equal overdue amounts of **18M** (blue and red, respectively).
* **Insights**:
  + The **Mankhurd** branch has the most overdue payments, which may indicate a need for stricter credit checks or follow-ups.
  + The branches **Nashikcidco** and **Islampur** show the least overdue amounts among the listed branches.

**2. Total OD by Region**

* **Bars**:
  + **Mumbai**: Highest OD amount at **118M** (orange).
  + **Ahmednagar**: Second highest at **54M** (yellow).
  + **Kolhapur**: Very close to Ahmednagar at **53M** (green).
  + **Nashik**: Moderate OD amount of **41M** (blue).
  + **Banswara**: Lowest OD amount at **25M** (red).
* **Insights**:
  + The **Mumbai region** has a significantly higher overdue amount compared to others, indicating higher credit risk.
  + Regions like **Banswara** have lower overdue amounts, which might suggest better repayment behavior or smaller loan portfolios.

**3. Total OD by Center**

* **Bars**:
  + **D Sector 0012**: Highest OD amount at **0.81M** (orange).
  + **Vasagade BC 1**: Second highest at **0.62M** (yellow).
  + **Pargi Fala Maheshpura - A** and **New Panvel 13**: Both at **0.59M** (green and blue, respectively).
  + **B Sector 10 - 10**: Lowest OD amount at **0.56M** (red).
* **Insights**:
  + The **D Sector 0012** center stands out with the highest overdue amount, indicating a focus area for collections.
  + The OD amounts for other centers are closely clustered, showing relatively uniform overdue levels in this category.

**4. Loan Amount and Total OD by State (Map)**

* **Map Overlay**:
  + Overdue amounts are represented geographically across states.
  + Larger markers indicate higher overdue amounts in specific states or regions.
* **Insights**:
  + States with larger markers have higher overdue amounts, showing where most repayment risks are concentrated.
  + The visualization helps identify geographical patterns and target areas for improvement.

**Key Observations**

1. **Branch-Level OD**:
   * **Mankhurd** and **Kalyan** require attention due to their high overdue amounts.
2. **Region-Level OD**:
   * **Mumbai** contributes the largest share of overdue payments, which may signify higher loan disbursement in this area.
3. **Center-Level OD**:
   * **D Sector 0012** has the most overdue payments among the listed centers.
4. **State-Level OD**:
   * Geographical hotspots for overdue amounts can be targeted for interventions such as collection drives or financial education.

**Dashboard page – 4**

****

**Explanation of Dashboard page -4**

**1. Loan Amount and Total OD by Region**

* **Blue Bars**: Represent the total loan amount for each region.
* **Orange Bars**: Represent the total overdue (OD) amount for each region.
* **Highlights**:
  + **Mumbai**:
    - Highest loan amount (~0.2bn).
    - Also has a significantly high OD amount, though lower than the total loans.
  + **Ahmednagar** and **Kolhapur**:
    - Both show relatively high loan amounts (~0.1bn each) with OD amounts proportionately lower than Mumbai.
  + **Nashik** and **Banswara**:
    - Moderate loan and OD amounts compared to the top three regions.
  + **Piplod**, **Ajmer**, and others:
    - Lower loan and OD amounts, indicating smaller-scale operations or better repayment performance.
* **Insights**:
  + Regions like **Mumbai** require attention due to high overdue payments.
  + Smaller regions (e.g., **Modasa**) may have fewer loans and lower overdue risks.

**2. Loan Amount and Total OD by Branch**

* **Blue Bars**: Represent the total loan amount for each branch.
* **Orange Bars**: Represent the total overdue (OD) amount for each branch.
* **Highlights**:
  + **Mankhurd**:
    - Highest loan amount (~90M) and OD amount (~60M).
  + **Kalyan** and **Mulund**:
    - High loan and OD amounts, but lower than Mankhurd.
  + Branches like **Ajmer**, **Pushkar**, and others toward the right of the chart:
    - Minimal loan and OD amounts, likely due to smaller operations or better collections.
* **Insights**:
  + Branches like **Mankhurd** and **Kalyan** need better overdue management.
  + Branches with lower loans and OD amounts may indicate fewer financial activities or strong repayment performance.

**3. Loan Amount and Total OD by Officer**

* **Blue Bars**: Represent the total loan amount handled by each officer.
* **Orange Bars**: Represent the total overdue (OD) amount for each officer.
* **Highlights**:
  + **Nagesh Madhukar Shinde**:
    - Highest loan amount (~40M) and a significant OD amount (~20M).
  + **Rushikesh Vijay Jadhav**, **Siddhanth Milind Shardul**, and others:
    - High loan amounts with varying OD amounts, generally lower than Nagesh.
  + Officers toward the right (e.g., **Prashant Gawali**):
    - Smaller loan and OD amounts, indicating lower responsibility or better performance.
* **Insights**:
  + Officers like **Nagesh Madhukar Shinde** may require support to manage overdue accounts effectively.
  + Officers with smaller loan portfolios may have a better handle on repayments or fewer responsibilities.

**Key Takeaways**

1. **Region-Level Observations**:
   * **Mumbai** has the highest financial activity and overdue risks.
   * Smaller regions like **Modasa** have minimal overdue risks.
2. **Branch-Level Observations**:
   * **Mankhurd**, **Kalyan**, and **Mulund** require attention for overdue management.
   * Branches with minimal loans and ODs may not require immediate intervention.
3. **Officer-Level Observations**:
   * Top officers like **Nagesh Madhukar Shinde** need to focus on overdue accounts.
   * Officers with lower loan amounts generally show better performance.