**District-wise Analysis of Direct Benefit Transfer (DBT) in India**

**EDA REPORT**

*Submitted for*

*Data Science C.A – 1*

*EDA Report Analysis*

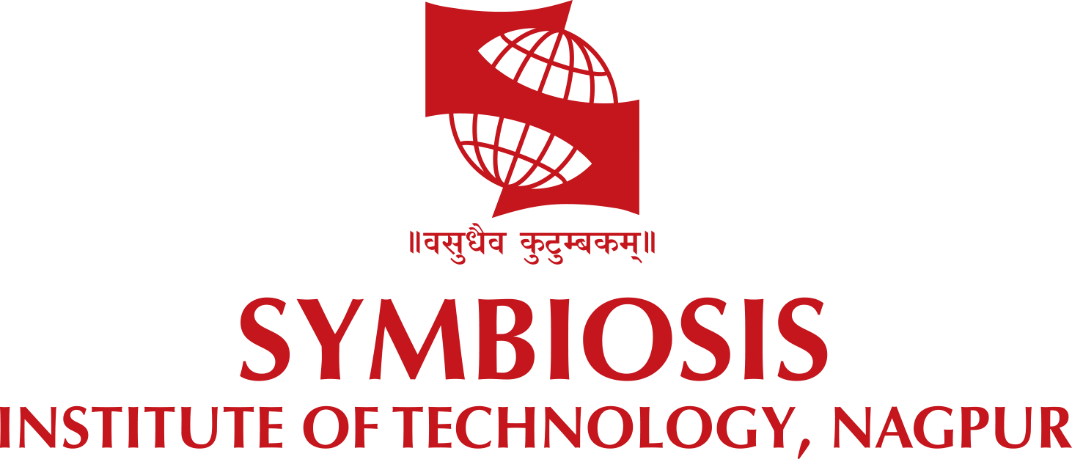
*Submitted by*

**Shrishti Prasad, 22070521035**

**B. Tech Computer Science and Engineering**

*Under the Guidance of*

**Dr. Piyush Chauhan**



1. Introduction

**What is DBT?**

**Direct Benefit Transfer (DBT)** is a government initiative in India that aims to transfer subsidies and financial assistance directly into the bank accounts of beneficiaries. It reduces leakages, ensures transparency, and promotes faster delivery of services by eliminating intermediaries. DBT is widely used in schemes related to education, health, LPG, pensions, and various welfare programs.

**Importance of DBT in Governance and Public Schemes**

DBT plays a crucial role in strengthening **good governance** by:

* **Reducing corruption** and fraudulent claims
* **Improving targeting accuracy** of welfare schemes
* **Ensuring real-time transfer** of benefits to eligible citizens
* **Bringing accountability** through digital trails and verification
* **Enhancing efficiency** by replacing physical cash handling with electronic systems

It has helped India move toward a **transparent, inclusive, and digitally empowered welfare infrastructure**.

**Why This Analysis is Important?**

Analysing DBT data at the **district and state levels** helps uncover:

* Regions with high or low financial disbursement
* Patterns in government spending across years
* Disparities in transaction volume versus monetary value
* Outliers or anomalies that may indicate inefficiencies or data issues

Such insights can help **policy-makers, analysts, and administrators**:

* Improve the design and delivery of welfare schemes
* Reallocate funds more equitably
* Identify under-served regions needing more attention
* Monitor the effectiveness of DBT implementation over time

1. Problem Statement

"To explore and analyze DBT transfer patterns across Indian districts and states to identify regional trends, anomalies, and areas needing policy attention. This project aims to explore and analyze Direct Benefit Transfer (DBT) patterns across Indian districts and states using a publicly available dataset. By examining parameters such as the total transfer amount and the number of transactions, the goal is to uncover regional trends, detect disparities, and identify anomalies in fund allocation. Through this analysis, we seek to support data-driven policy-making by highlighting under-served regions and suggesting areas that may require more focused administrative attention for improved DBT implementation."

1. Dataset Information

* **Source**: <https://indiadataportal.com/p/direct-benefit-transfer/r/cabinet-dbtbharat_dbt_performance-dt-ot-aaa>
* **Size**: Number of rows and columns – 3420 rows and 8 columns
* **Columns Description**:

|  |  |
| --- | --- |
| Column | Description |
| fy | Financial year |
| state name | State name |
| district name | District name |
| total dbt transfer | Total transfer amount |
| no\_of\_dbt\_transactions | Number of DBT Transactions |
| State\_code | Numeric code of the state |
| district\_code | Numeric code of the district |
| id | Unused identifier column (dropped later) |

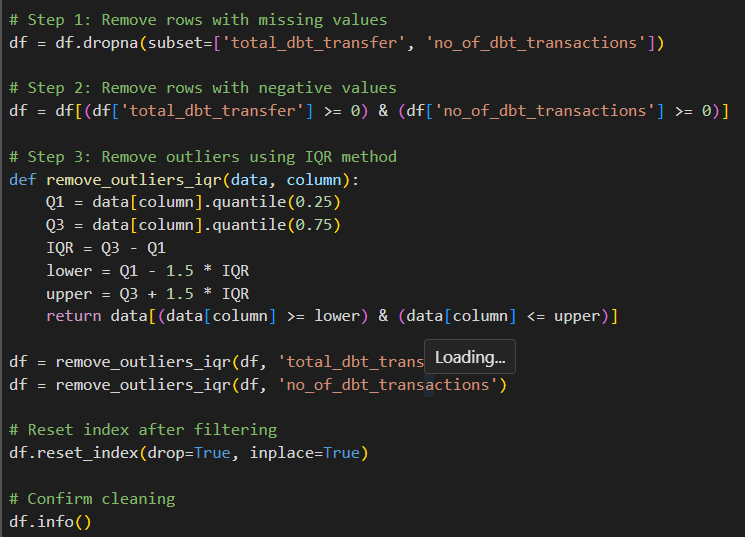
1. **Data Preprocessing**

* Loaded data
* Handled missing values
* Removed duplicates
* Removed negative or invalid values
* Outlier handling (IQR filtering if used)
* Extracted start\_year, end\_year from fy
* Final cleaned dataset shape

1. **Data Cleaning Steps**

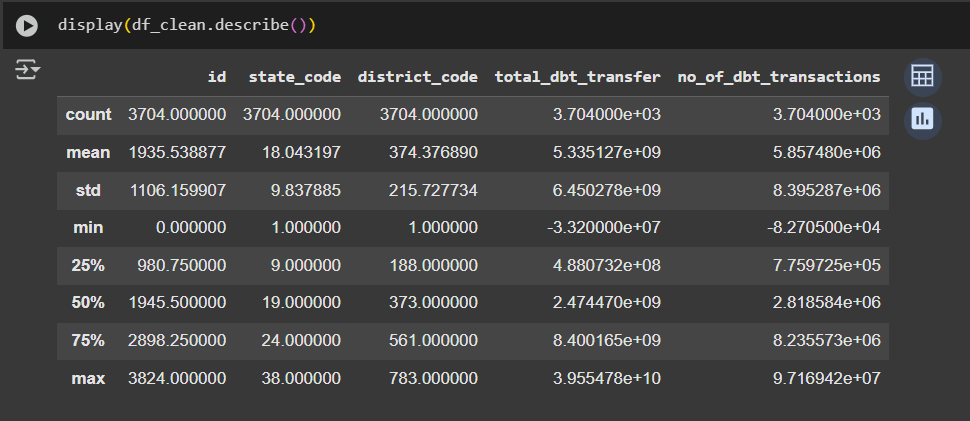
**Step-by-step cleaning:**

1. **Read and copy original data** using Pandas.
2. **Drop rows** with missing values in crucial columns (total\_dbt\_transfer, no\_of\_dbt\_transactions).
3. **Split fy** into two separate columns: start\_year and end\_year.
4. **Drop id** column — considered irrelevant for analysis.
5. **Remove duplicates**.
6. **Create state\_district** column to combine state\_name and district\_name for grouped analysis.
7. **Convert data types** to appropriate formats for better memory and correctness.
8. **Remove negative values** in numerical columns.
9. **Remove outliers** using IQR method.
10. **Reset index** after cleaning.



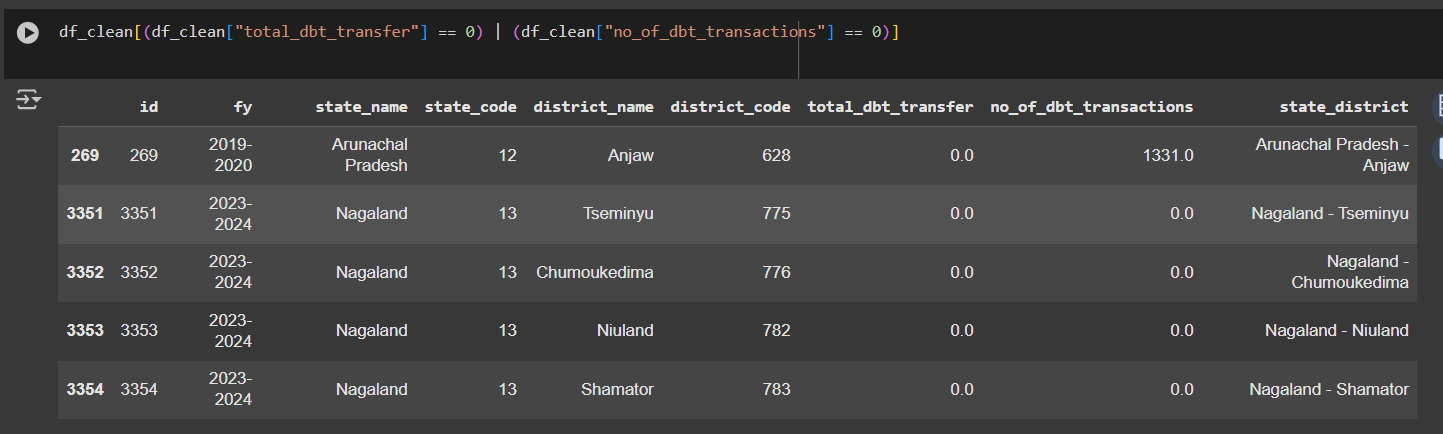
📊 Interpretation

Removed missing and negative values in key columns to ensure clean data. Then applied the IQR method to remove extreme outliers from total\_dbt\_transfer and no\_of\_dbt\_transactions. Finally, reset the index to keep the DataFrame organized. The .info() confirms the dataset is now clean and ready for analysis.



📊 Interpretation:

Showed summary statistics (like count, mean, min, max) for numerical columns. Helps understand the central tendency, spread, and detect any anomalies in the data.

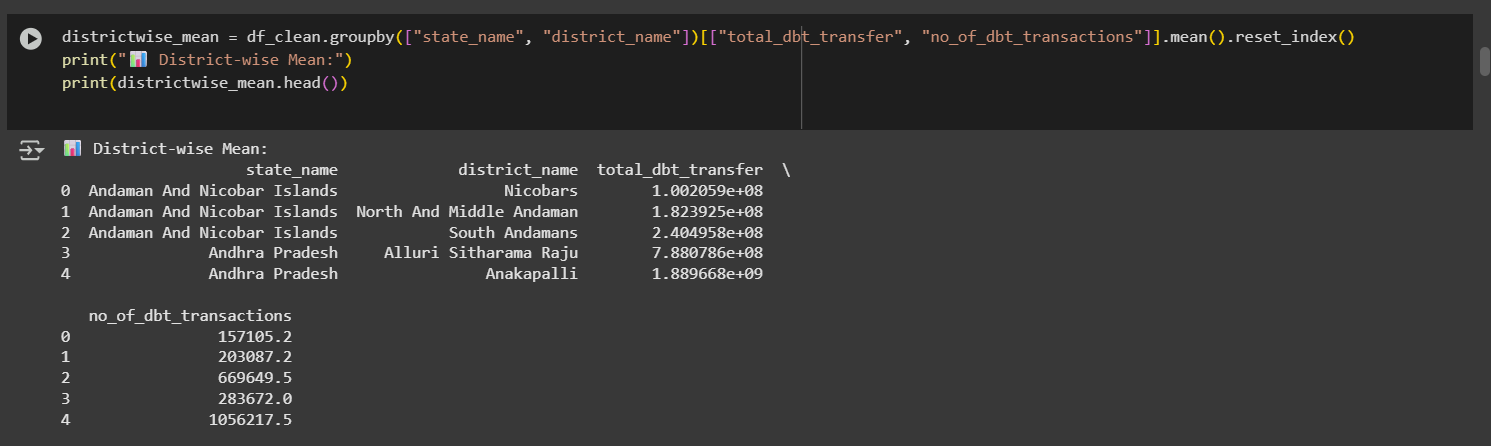


📊 Interpretation:

Filtered rows where either total DBT transfer or number of transactions is zero. These records may indicate missing implementation or data errors and can be flagged or removed to maintain analysis accuracy.

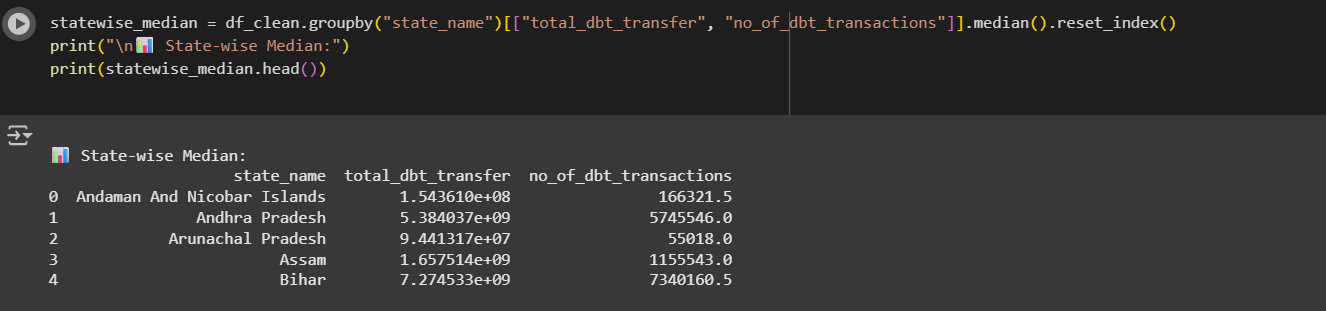
1. Descriptive and Statistical Summary

* df.describe() provides:
* Mean, min, max, standard deviation for all numeric columns.
* Categorical value counts were shown for:
  + state\_name
  + district\_name (top and bottom)
  + fy (Financial year distribution)
  + state\_district combinations



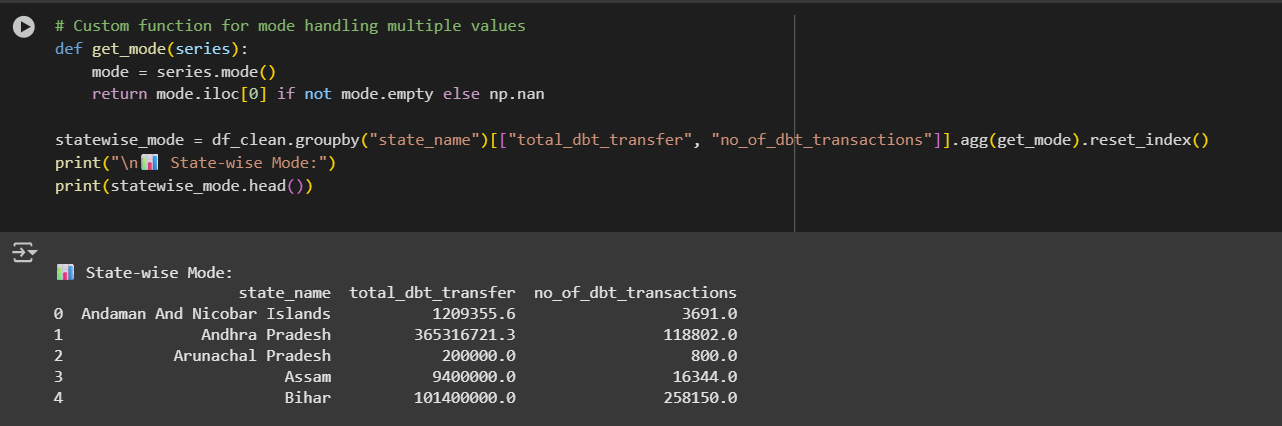
📊 Interpretation:

Computed the average DBT transfer and transaction count for each district within every state. Useful for detailed regional comparisons and identifying high- or low-performing districts.



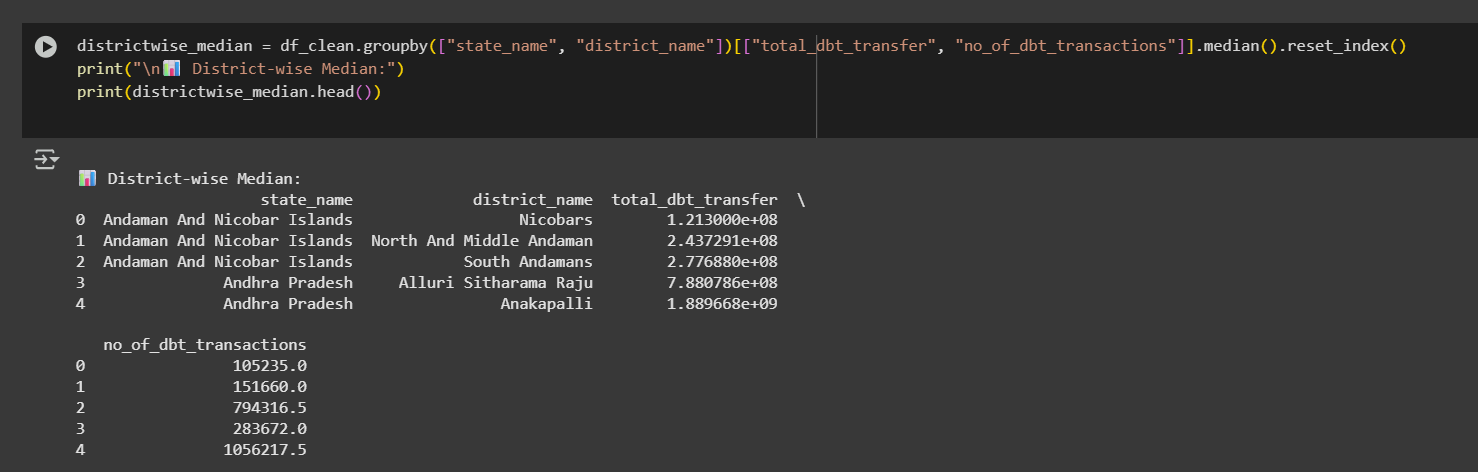
📊 Interpretation:

Computed the median DBT transfer and transaction count per state to reduce the impact of outliers. This gives a better central tendency measure for skewed data distributions.



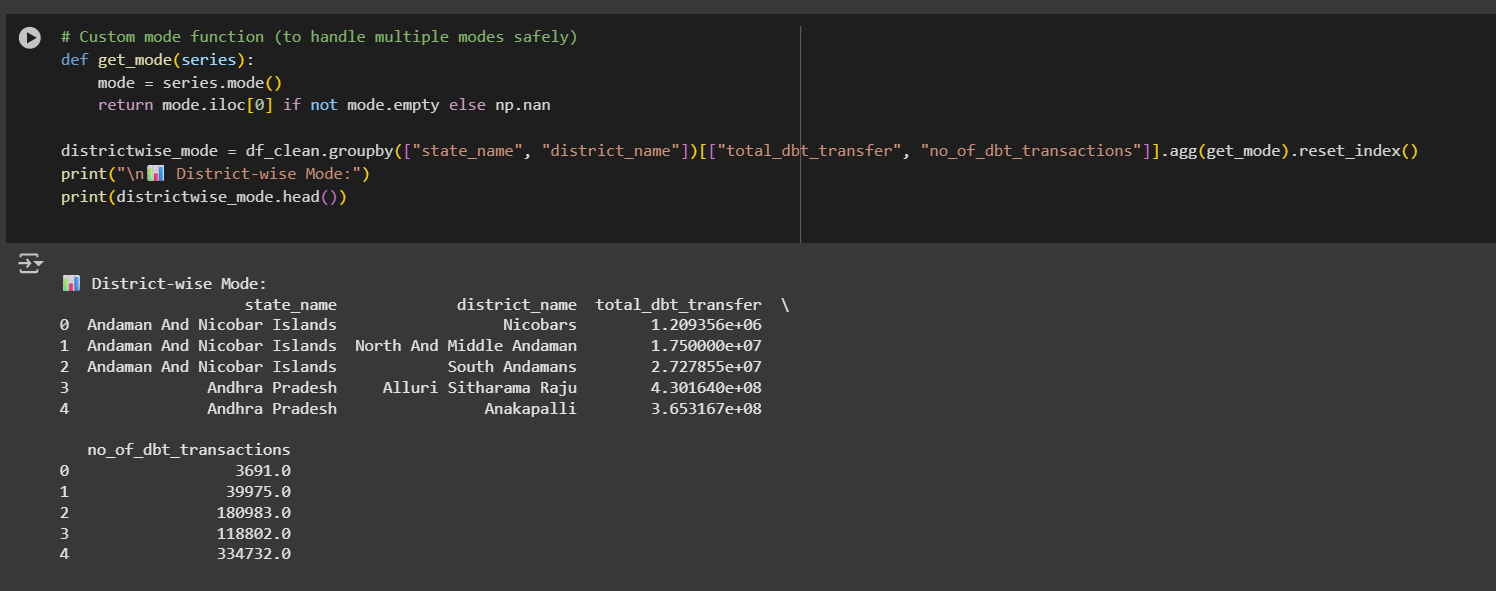
📊 Interpretation:

Calculated the mode (most frequent value) for DBT transfers and transactions per state. Helps identify the most commonly occurring values and capture repetitive patterns across districts in each state.



📊 Interpretation:

Calculated the median DBT transfer and transaction count for each district, reducing the effect of extreme values. Helps identify typical district-level performance within each state.

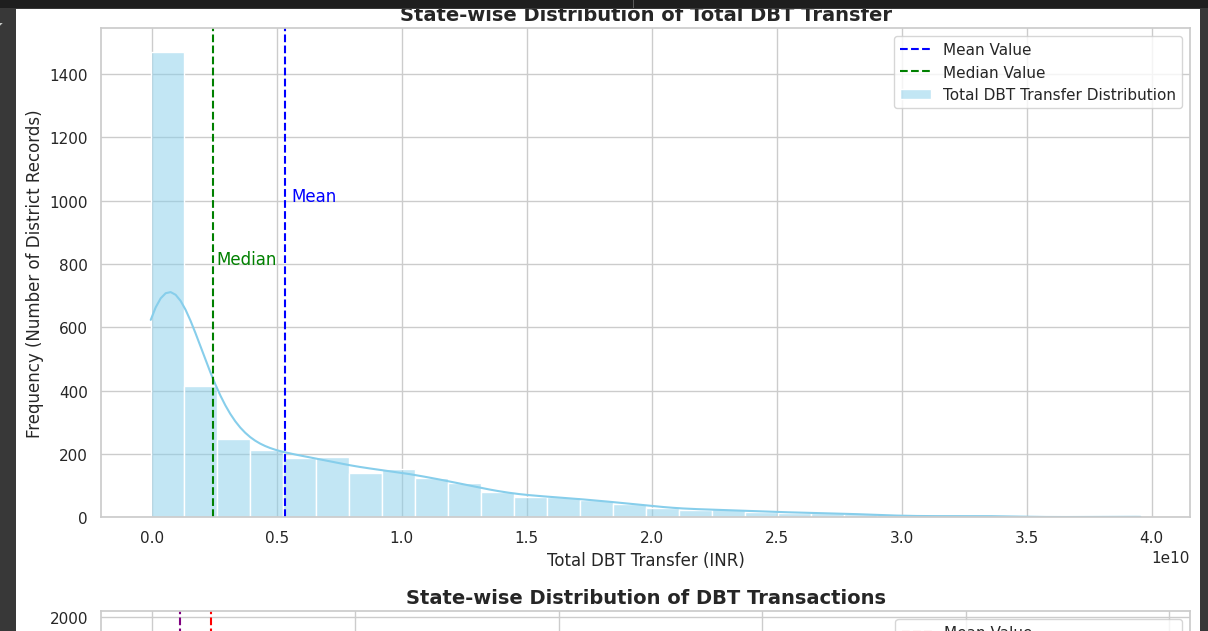


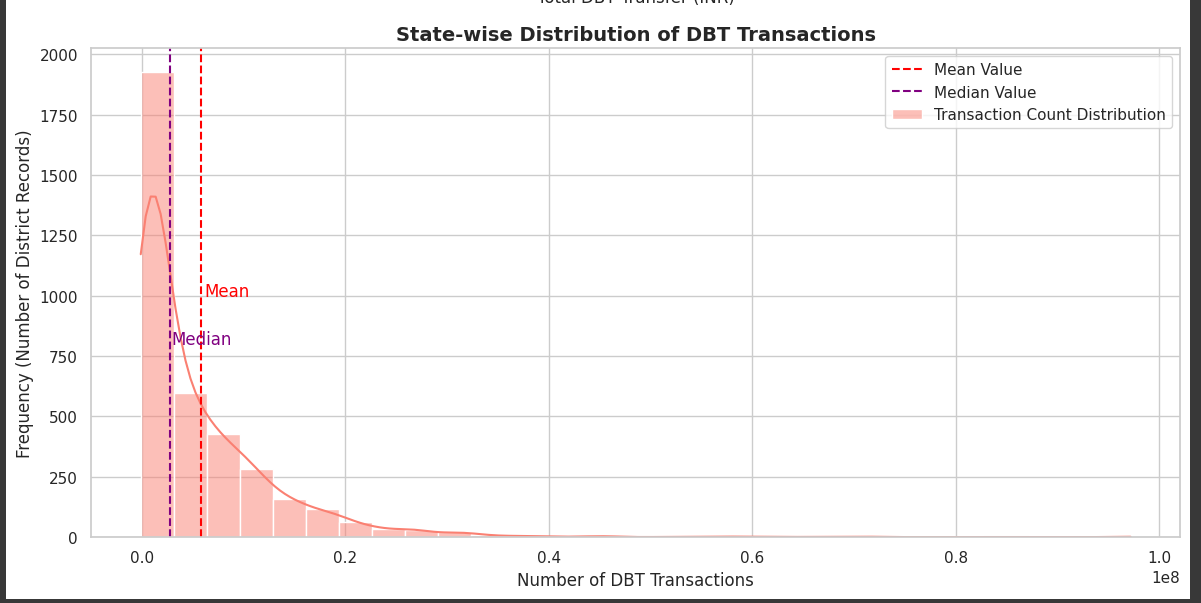
📊 Interpretation:

Found the most frequent (mode) values of DBT transfers and transactions for each district. Helps detect typical transaction behaviors and frequently repeated fund patterns across regions.

1. Univariate Analysis

State - wise





📊 Interpretation:

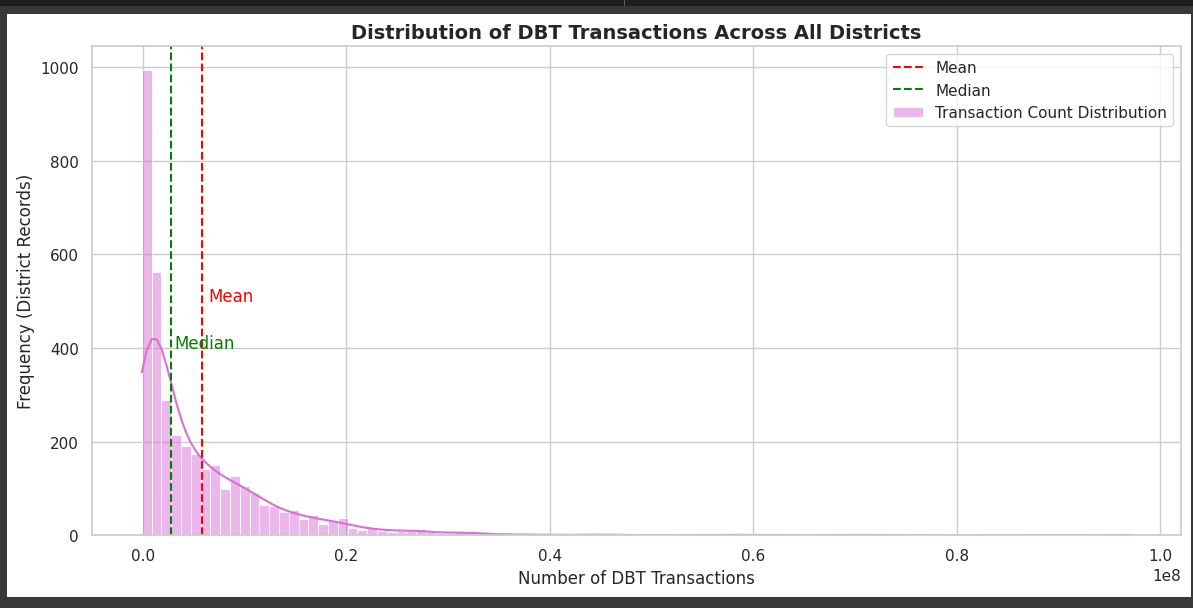
Two histograms were plotted to show the distribution of:

Total DBT Transfer

Number of DBT Transactions

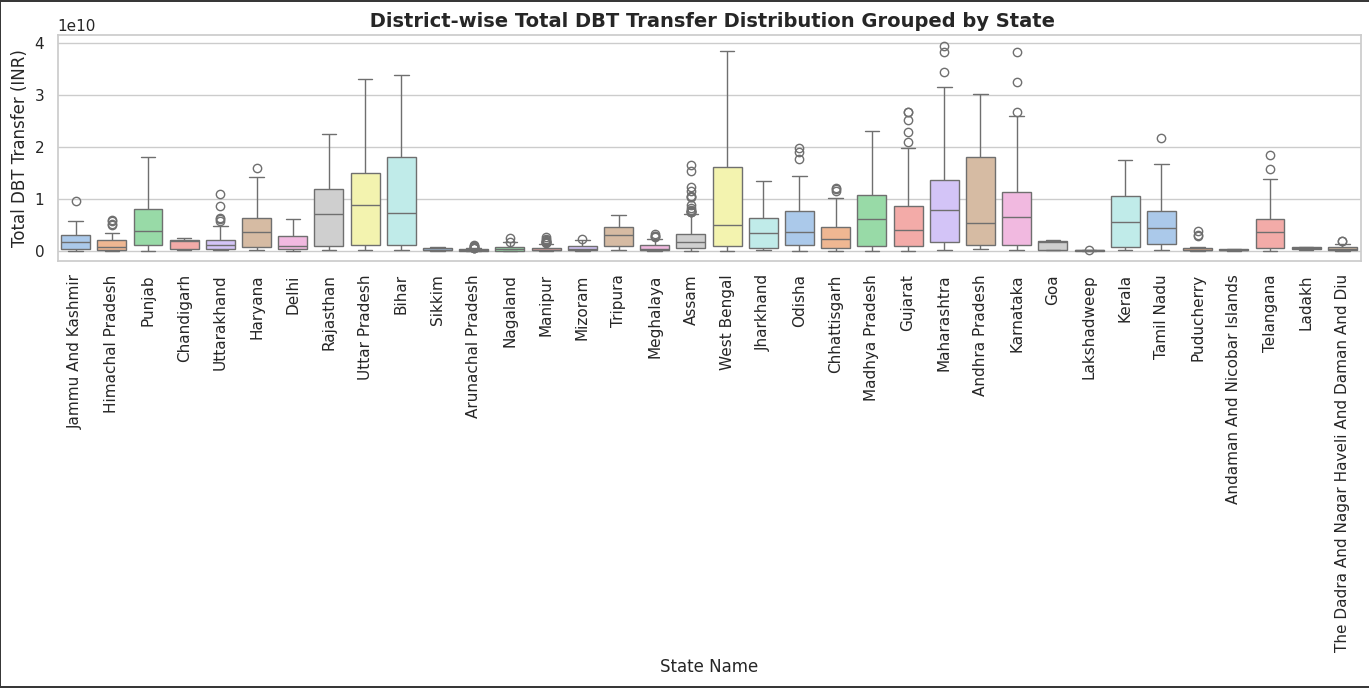
Both include KDE curves to show the shape of distribution.

District wise univariate analysis



📊 Interpretation:

This histogram shows how DBT transactions are distributed across all districts. The mean and median are marked to highlight central values. The skewness and spread help identify whether most districts have low, average, or high transaction volumes. Useful for spotting districts with exceptionally high or low activity.

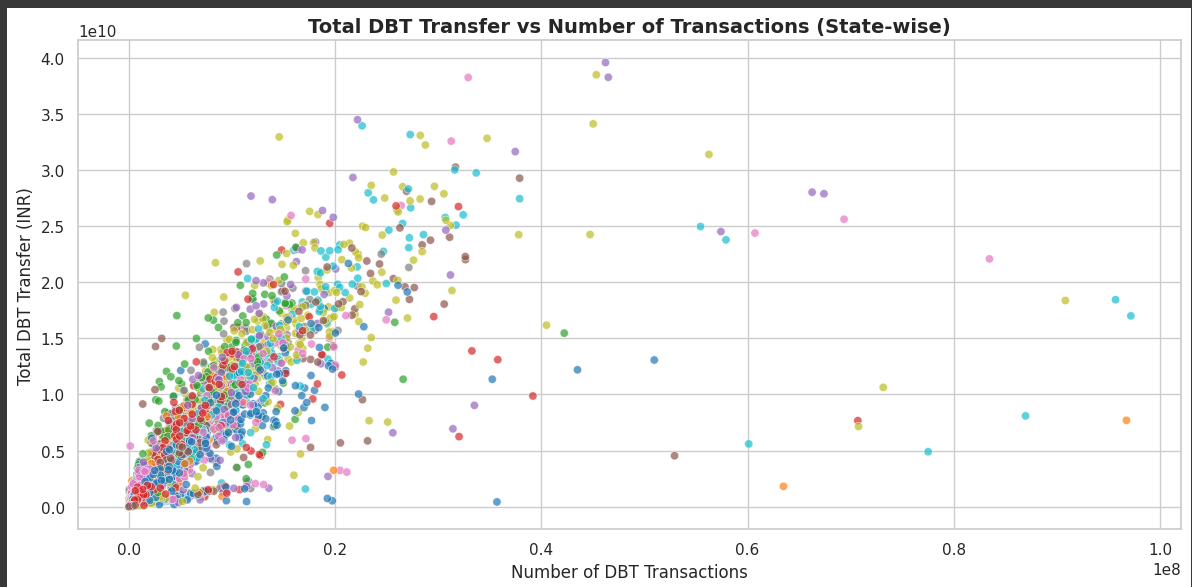


📊Interpretation:

Boxplot: District-wise Total DBT Transfer Distribution by State States like West Bengal, Maharashtra, and Andhra Pradesh exhibit high variability in DBT transfers, with several districts having very high transfer amounts (shown by outliers).

Smaller states/UTs such as Sikkim, Goa, and Chandigarh show low and consistent DBT disbursements across districts.

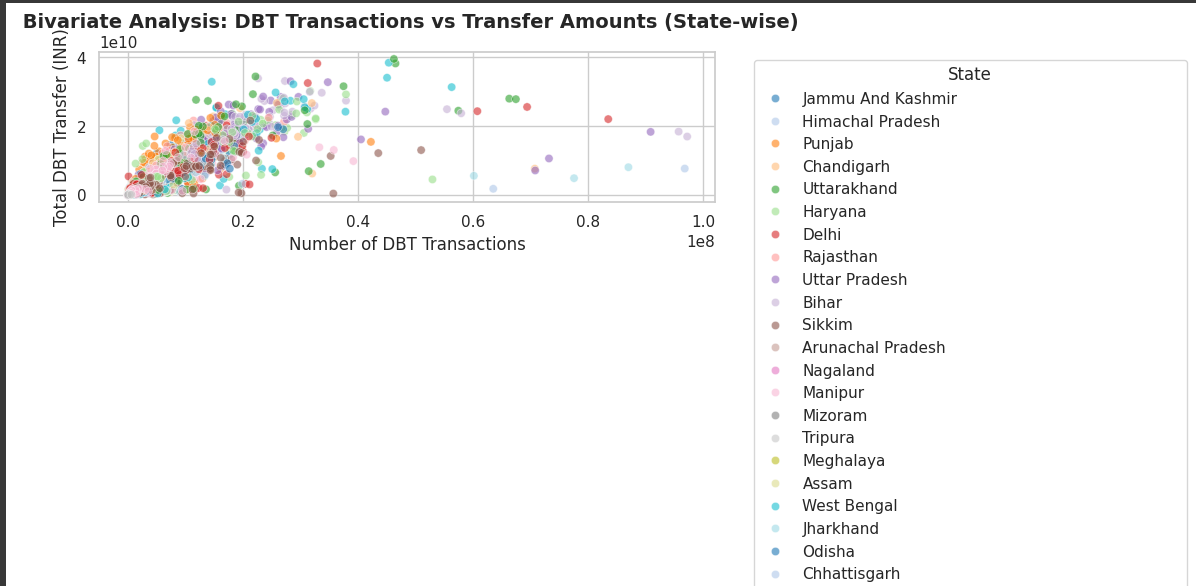
1. Bivariate Analysis



📈 Interpretation

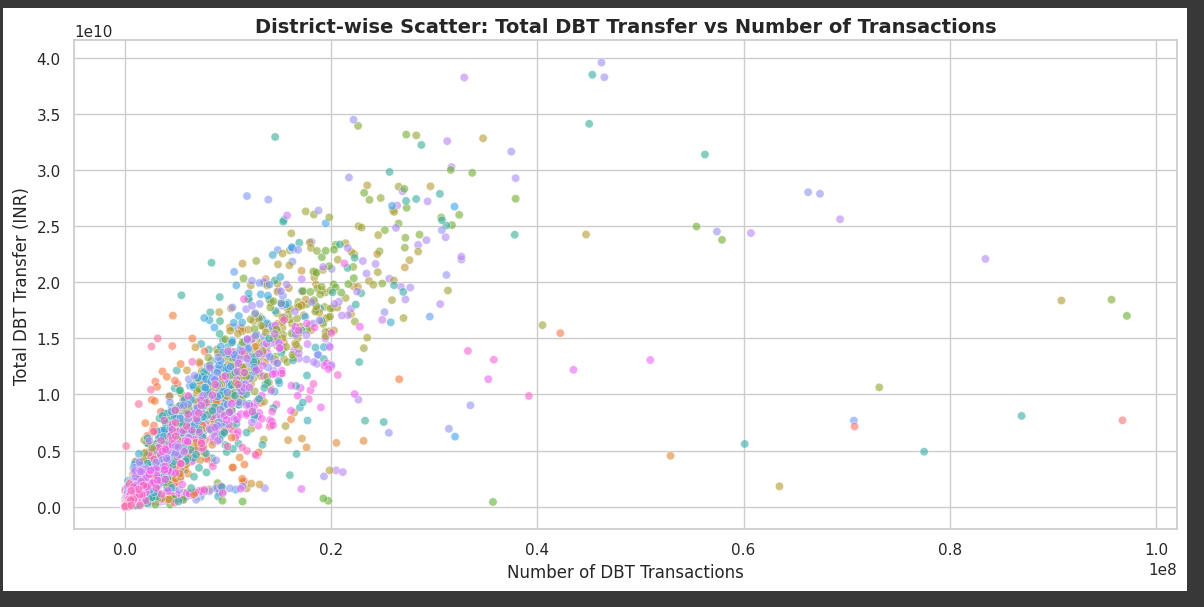
Scatter Plot: Total DBT Transfer vs. Number of Transactions There's a strong positive correlation: as the number of DBT transactions increases, the total amount transferred also tends to rise.

A few outliers suggest cases with unusually high amounts for relatively fewer transactions, indicating possible high-value schemes in certain districts.



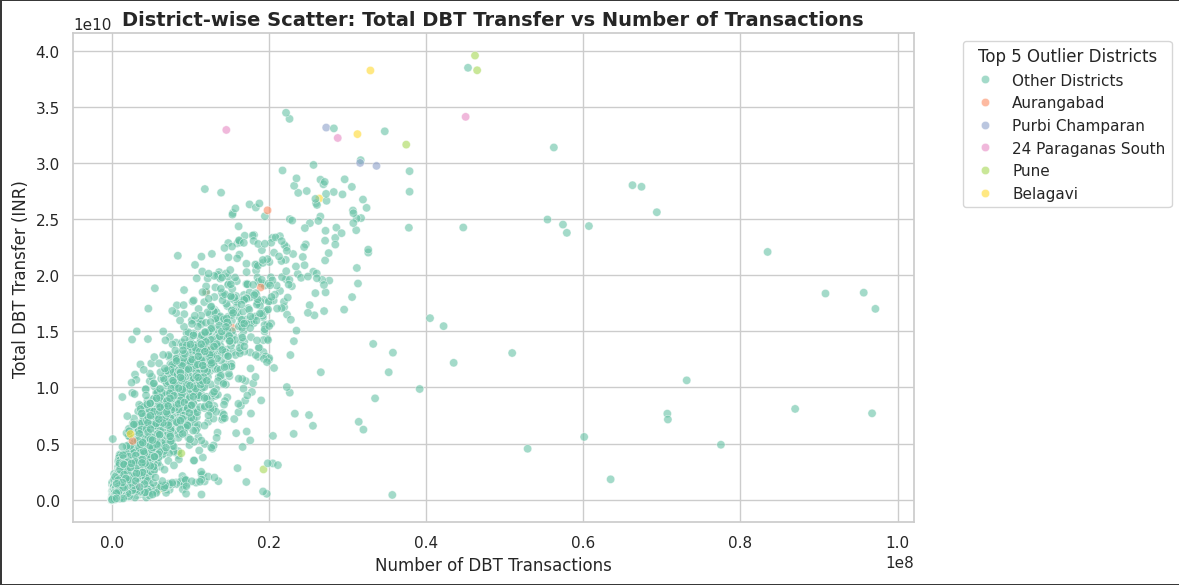
📊 Interpretation:

This scatter plot shows the relationship between number of DBT transactions and total transfers, colored by state. It helps identify how different states behave — whether higher transactions consistently lead to higher transfers, and spot state-level clusters or outliers. Disabling the legend improves readability with many states



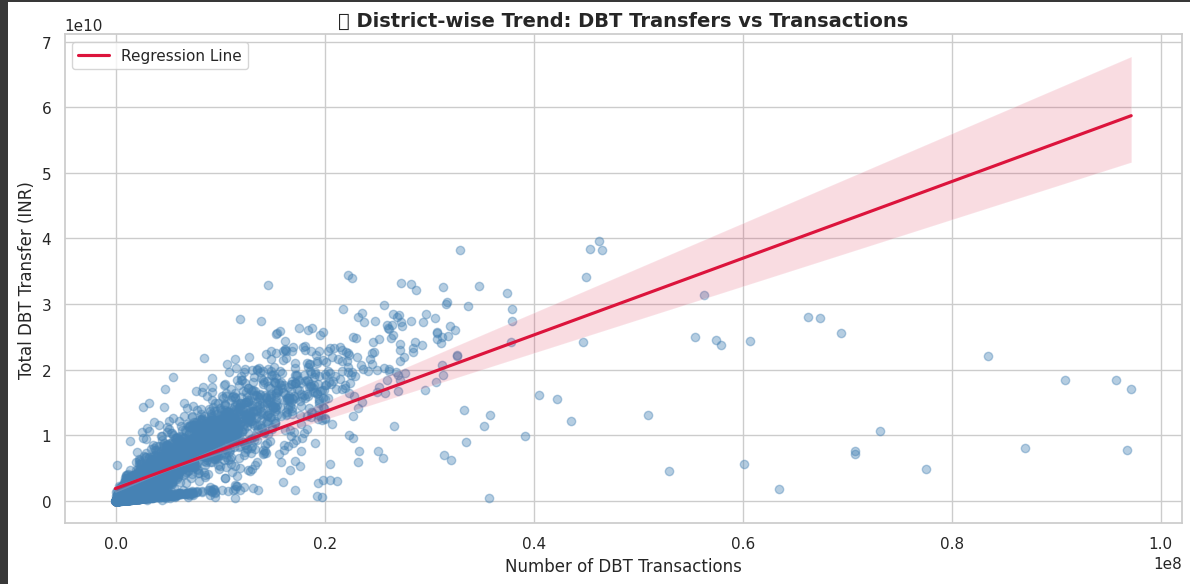
📊 Interpretation:

This district-level scatter plot shows the relationship between transactions and total DBT transfer for each district. Each point is a district (colored uniquely). It helps visualize patterns, clusters, and outliers in performance across all districts. Legend is disabled for clarity due to large number of districts.



📊 Interpretation:

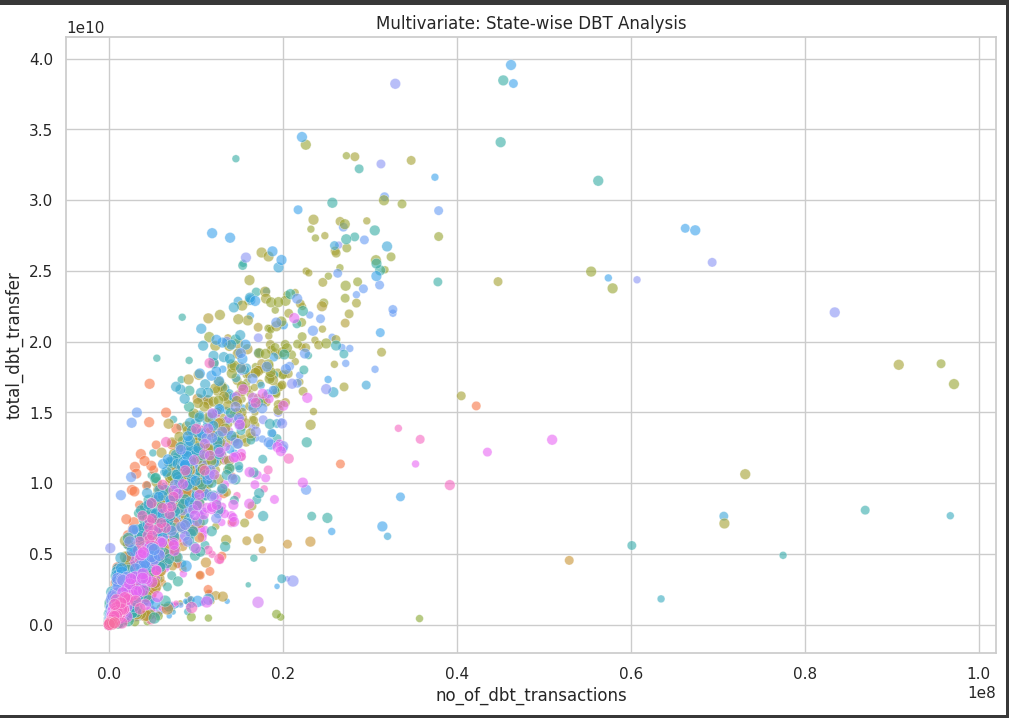
This scatter plot highlights the top 5 outlier districts with the highest total DBT transfers. These are labeled individually in the legend, while all others are grouped as "Other Districts." This helps focus analysis on extreme performers without cluttering the plot, making outliers stand out clearly for further investigation.



📊 Interpretation:

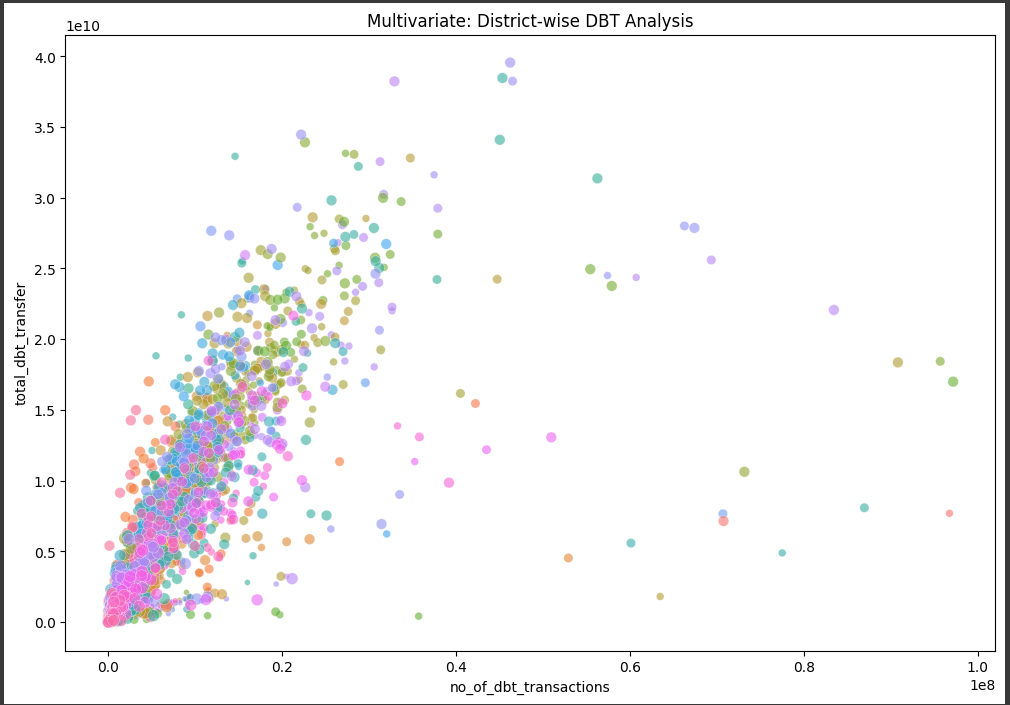
This regression plot shows the overall trend between number of DBT transactions and total transfer amount across districts. The red regression line indicates a positive relationship, suggesting that districts with more transactions tend to have higher fund transfers. Scatter transparency helps visualize point density.

1. Multivariate Analysis



📊 Interpretation:

This bubble plot analyzes DBT performance by plotting transactions vs. transfers, with bubble size representing the financial year (end\_year). It shows how DBT metrics vary over time and across states, helping identify growth patterns or shifts in performance year-wise. The plot captures three variables simultaneously for richer insights.

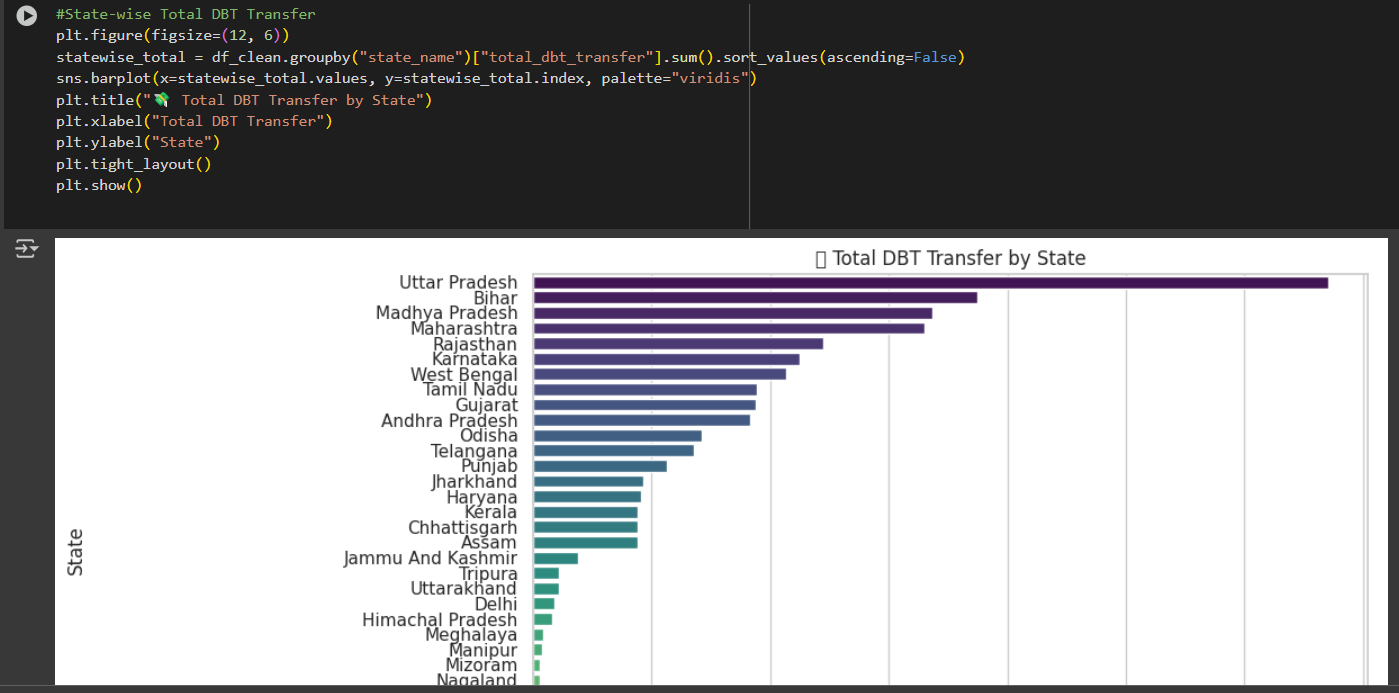


📊 Interpretation:

This multivariate scatter plot shows the relationship between transactions and transfers for each district, with bubble size representing the start year. It helps visualize how DBT activity varies over time and across districts, while also spotting growth trends or consistently high-performing regions.

1. Exploratory Data Analysis (EDA)

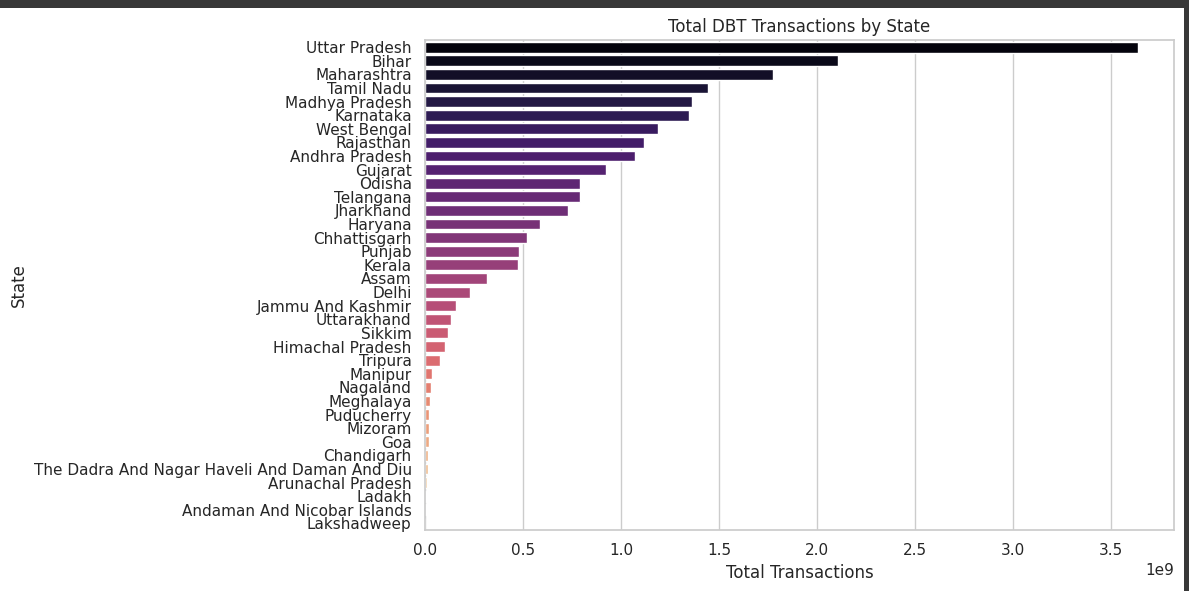
EDA State wise



📌 Interpretation:

This graph shows which states have received highest total DBT amounts.

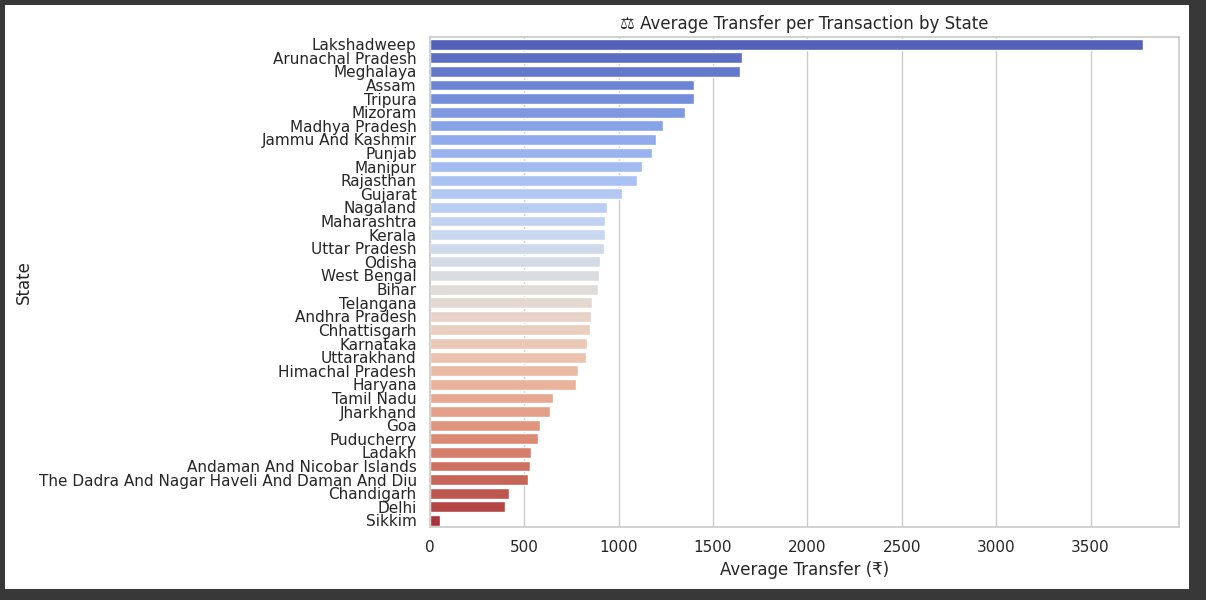
Helps identify regions with maximum financial assistance.



📌 Interpretation:

Shows volume of DBT activities.

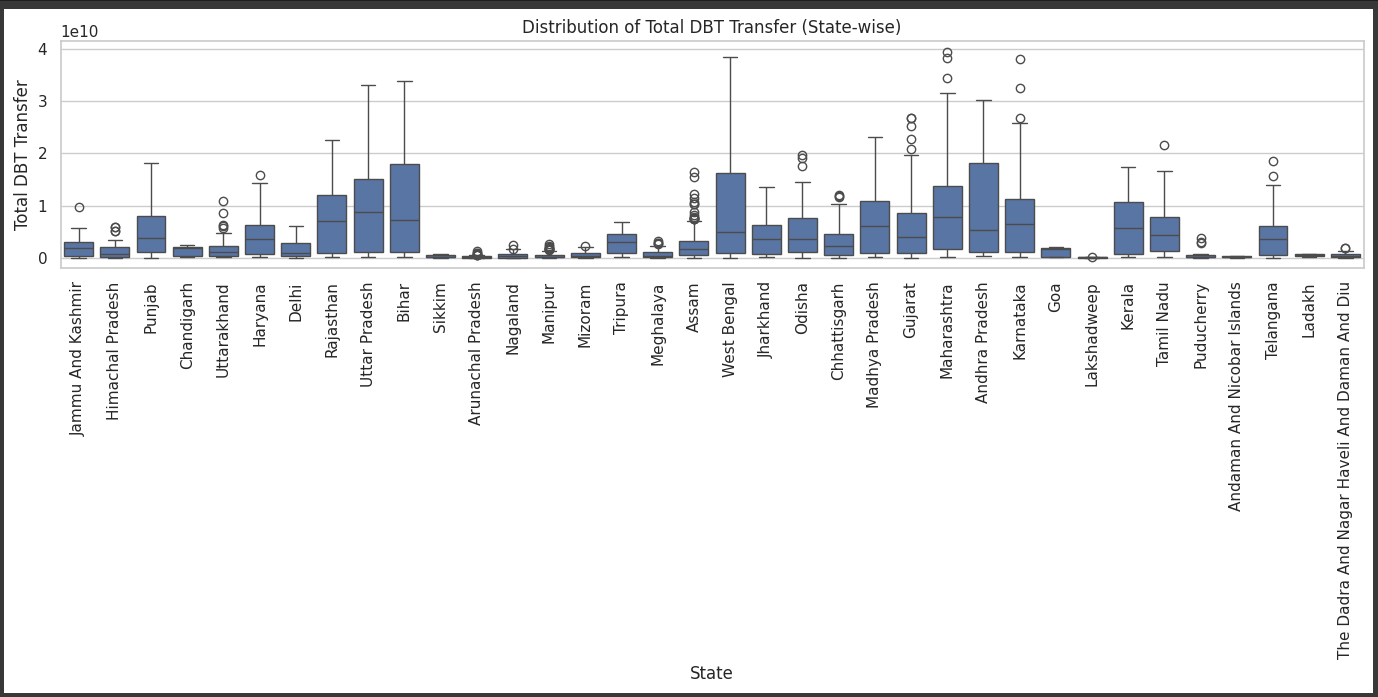
Some states may have more transactions but lower transfer amounts (e.g., many small-value transfers).



📌 Interpretation:

This shows how much money (on average) is being transferred per transaction.

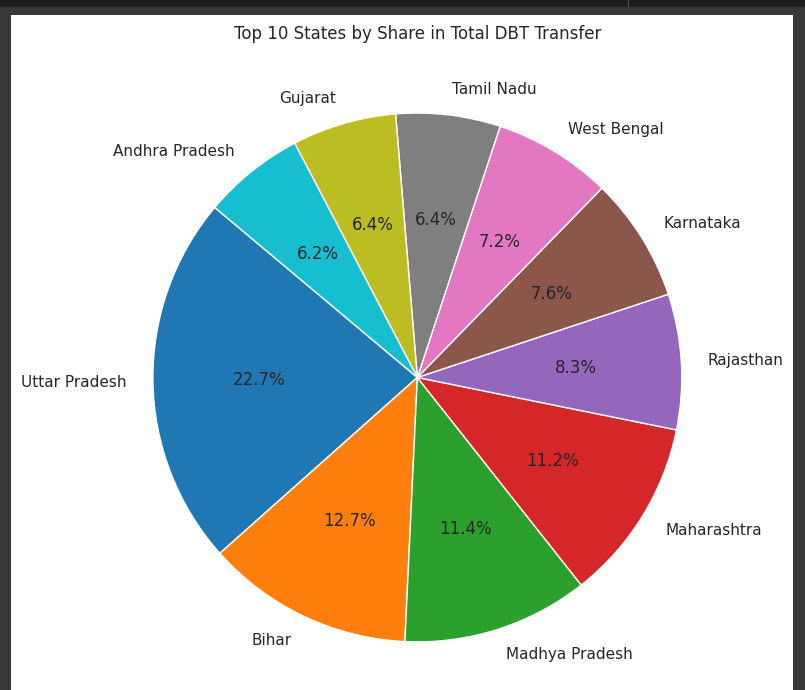
Higher values may indicate larger welfare schemes or one-time payouts.



📌 Interpretation:

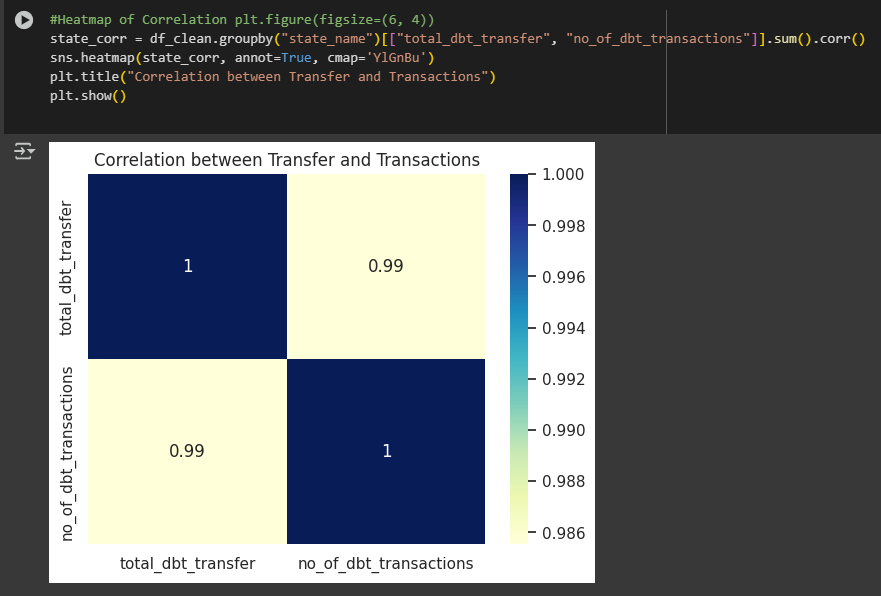
Detects outliers or skewed distributions in transfer values.

Useful to check variation within each state.



📌 Interpretation:

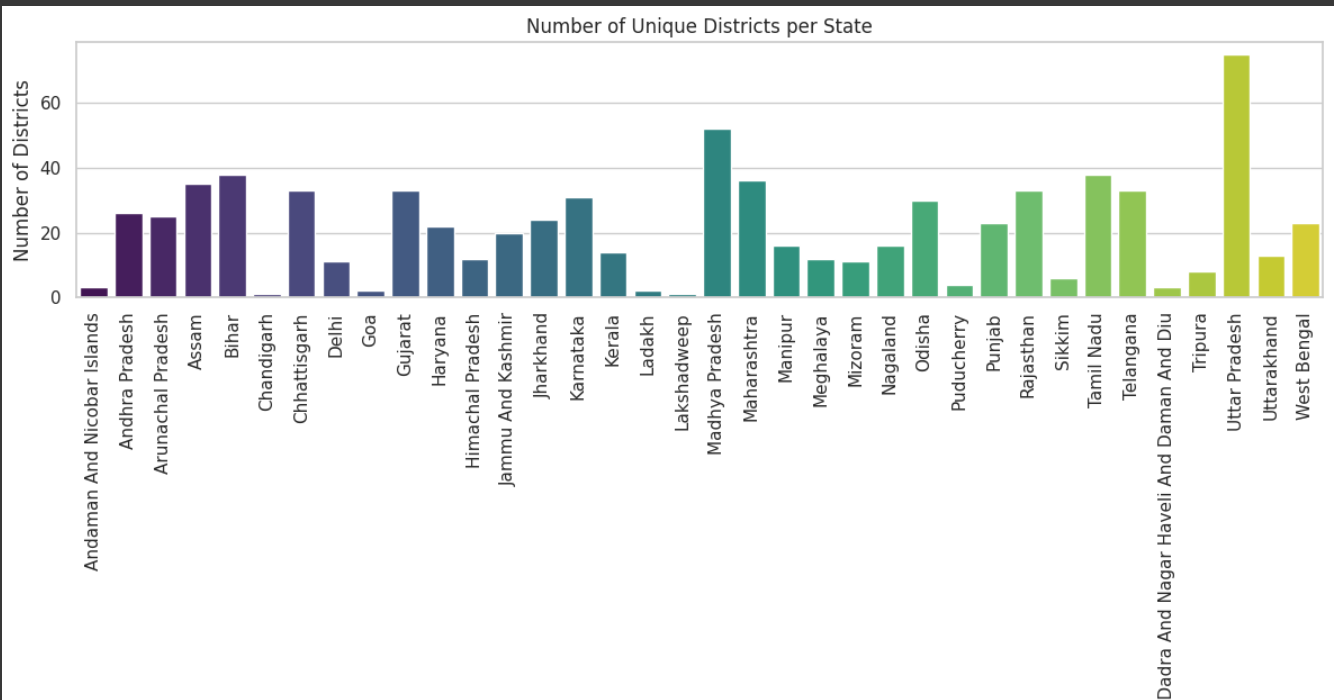
Gives quick % share of DBT burden across leading states.



 📌 Interpretation:

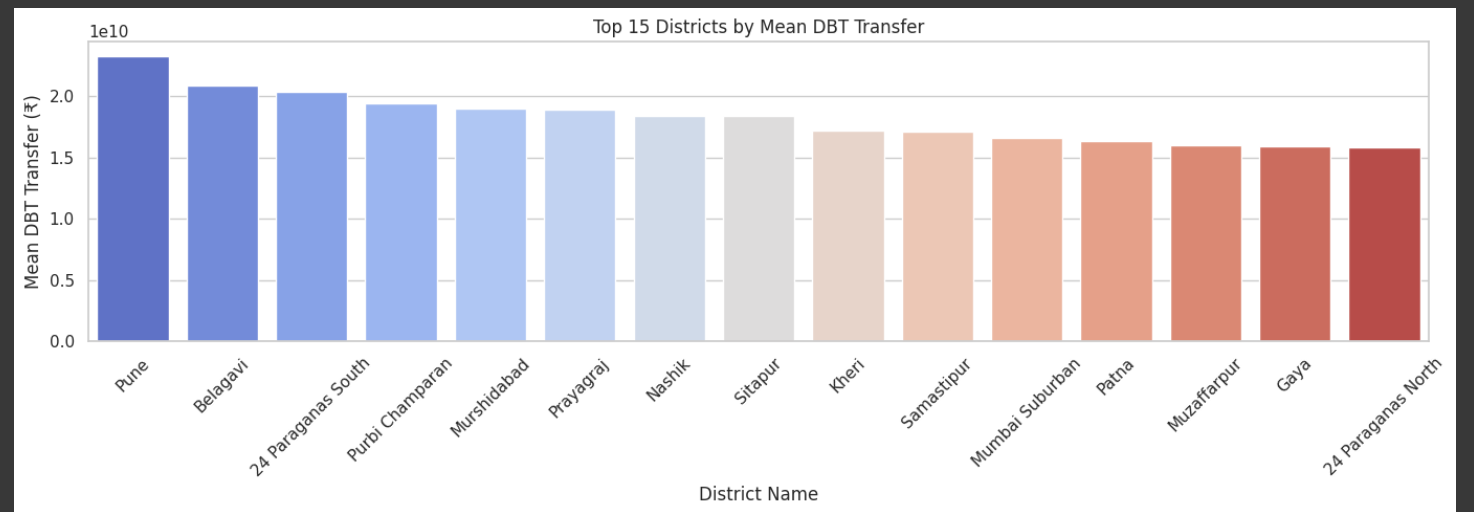
Shows how strongly total DBT transfer correlates with number of transactions.

EDA District wise



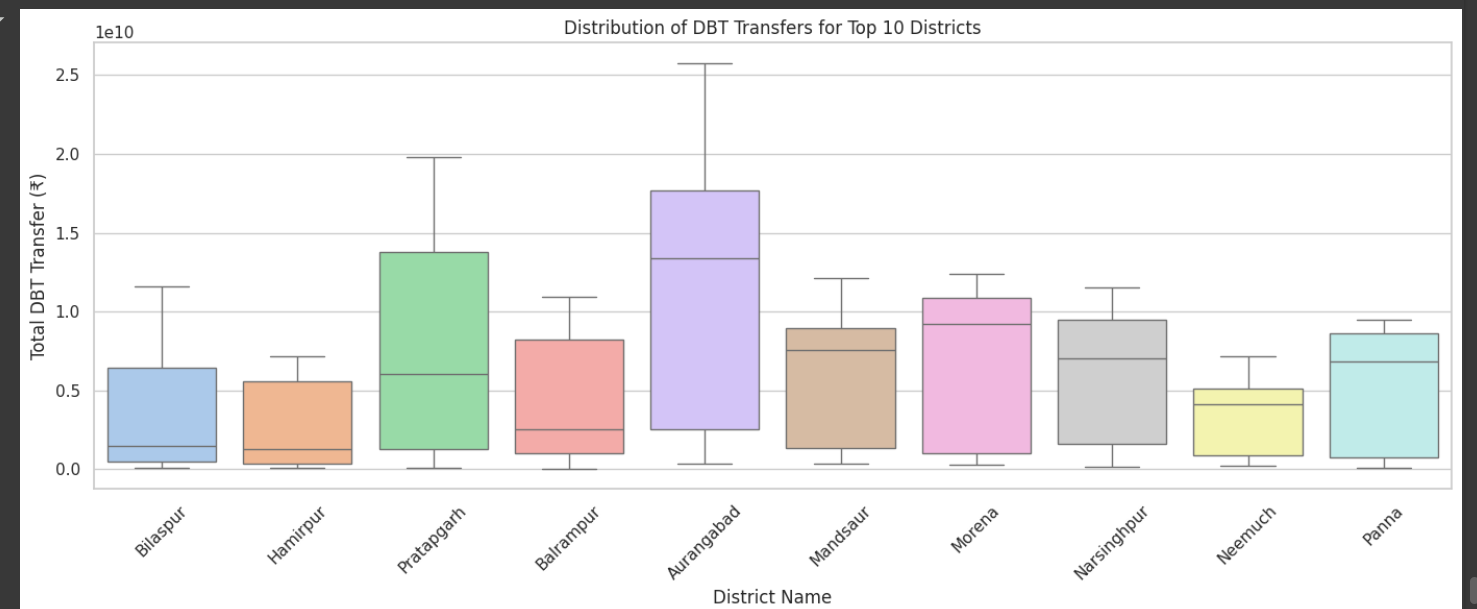
📌 Interpretation:

* This **bar plot** represents the **number of unique districts** reported in your dbt-district-wise.csv dataset for each **state/UT**.
* It uses unique () to count **distinct district names** per state.
* Top states by Number of Districts in Dataset: **Uttar Pradesh** has the highest count -~70 districts. Followed by **Madhya Pradesh**, **Tamil Nadu**, **Maharashtra**, **Bihar**, and **West Bengal**.
* Smallest Counts: **Chandigarh**, **Ladakh**, **Lakshadweep**, **Delhi**, and **Dadra & Nagar Haveli and Daman & Diu** show **very few or only 1 district** (as expected — these are small UTs or unionized areas).



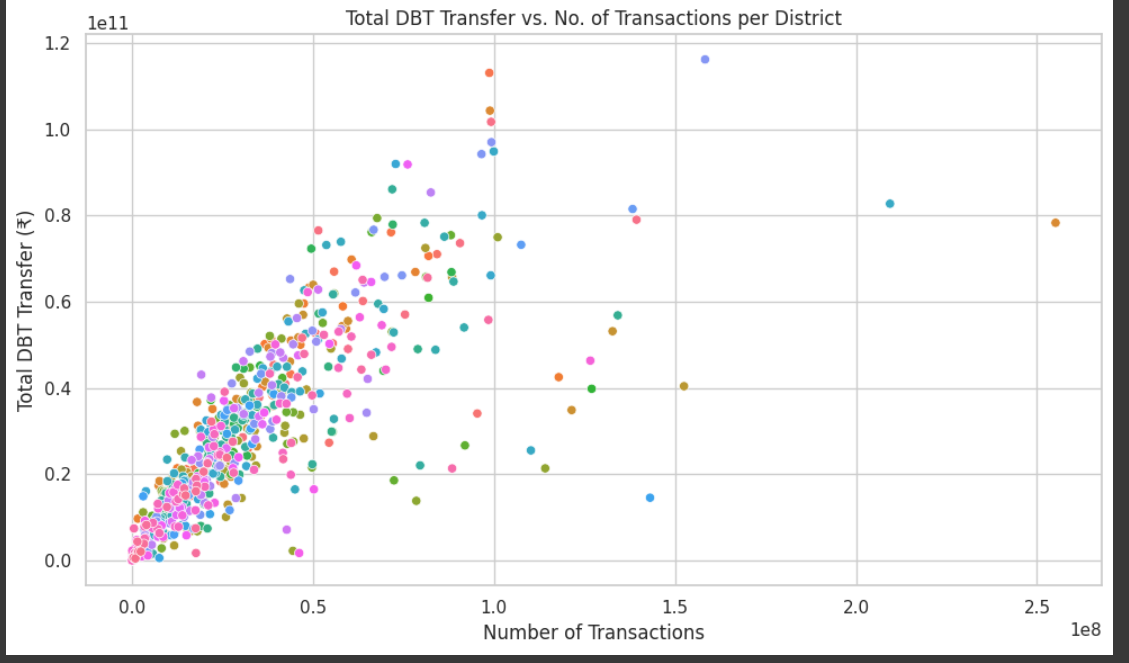
📌 Interpretation:

These districts received higher average transfer amounts per record, possibly indicating higher-value schemes or fewer but larger transactions.



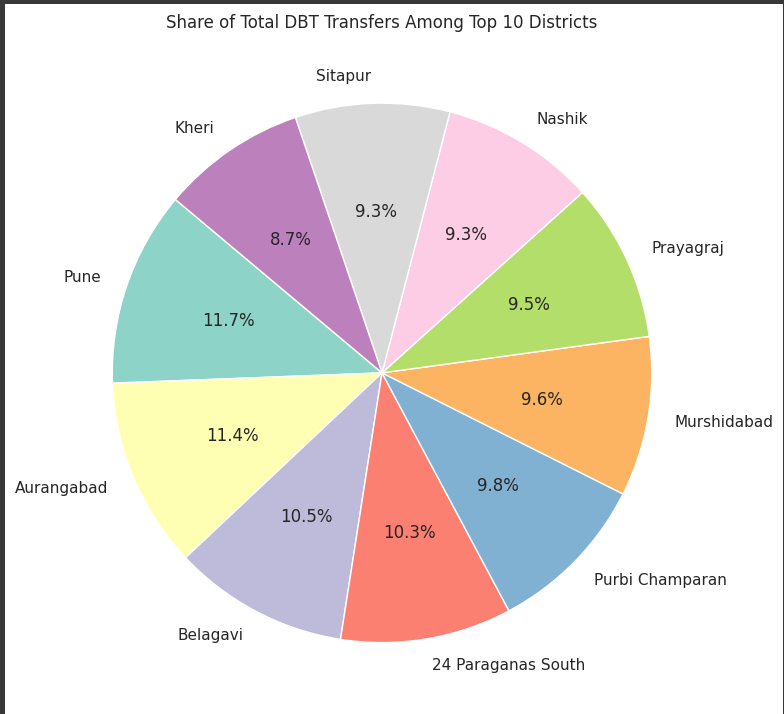
📌 Interpretation:

This helps to spot outliers and variability in transfer amounts within top districts.



📌 Interpretation:

Shows whether districts with more transactions also have higher total transfers. Helps identify inefficiencies or outliers.



📌 Interpretation:

Visualizes concentration of benefits- whether a few districts dominate the total DBT transfer.

1. **Observations & Insights**

Based on the exploratory data analysis, the following key observations were made:

1. **State-wise Transfer Patterns:**
   * States like **Uttar Pradesh, Maharashtra, and West Bengal** received the highest DBT amounts.
   * Some smaller states and union territories had significantly lower DBT volumes, indicating either fewer schemes or a smaller beneficiary population.
2. **Transaction Volume vs Amount:**
   * States with higher transaction volumes do not always correspond to higher DBT amounts, indicating variation in per-transaction value.
3. **District Disparities:**
   * Certain districts consistently appeared in the **top 10 in terms of both amount and volume**, suggesting efficient DBT rollout in those areas.
   * A few districts had **very low DBT values**, possibly due to poor infrastructure, lack of awareness, or data reporting issues.
4. **Yearly Trends:**
   * Most states show a **year-on-year increase** in both transaction count and amount, aligning with increased digitization and DBT penetration.
5. **Outlier Detection:**
   * A few entries with unusually high transaction amounts were filtered out using **IQR method**, confirming the need for preprocessing before analysis.
6. **Average DBT per Transaction:**
   * The average DBT value per transaction helped identify whether funds were being disbursed in large or small amounts across different regions.
7. **Conclusion**

The district-wise and state-wise analysis of Direct Benefit Transfer (DBT) data reveals several important insights into the geographic and financial distribution of welfare funds in India. The study highlights that DBT implementation is active nationwide, though not uniformly distributed across states and districts. Certain regions dominate in terms of both fund allocation and transaction volumes, while others lag behind. Through proper preprocessing—such as handling missing data, removing outliers, and standardizing values—and visual exploratory analysis, the project successfully identified patterns, outliers, and disparities. These insights can guide more equitable and efficient delivery of financial assistance and lay a strong foundation for further advanced analytics, such as predictive modeling or fraud detection in welfare schemes.

Colab Link:

<https://colab.research.google.com/drive/1SOn4D7zi6vBxpD_wa8IQV60PAD4KG_h6#scrollTo=BT7L4rzX_ROn>

Git hub Link:

<https://github.com/shrishti1108/Direct-Benefit-Transfer-DBT-in-India>