SYMBIOSIS
INSTITUTE OF TECHNOLOGY, NAGPUR
॥वसुधैव कुटुम्बकम्॥

DATA SCIENCE Presentation

DISTRICT BENEFIT TRANSFER

PARIKSHIT ABUJ
22070521009
Sem:7 SEC A
Under the guidance of
DR PIYUSH CHAUHAN

Symbiosis Institute of Technology, Nagpur Campus

SYMBIOSIS
INSTITUTE OF TECHNOLOGY, NAGPUR

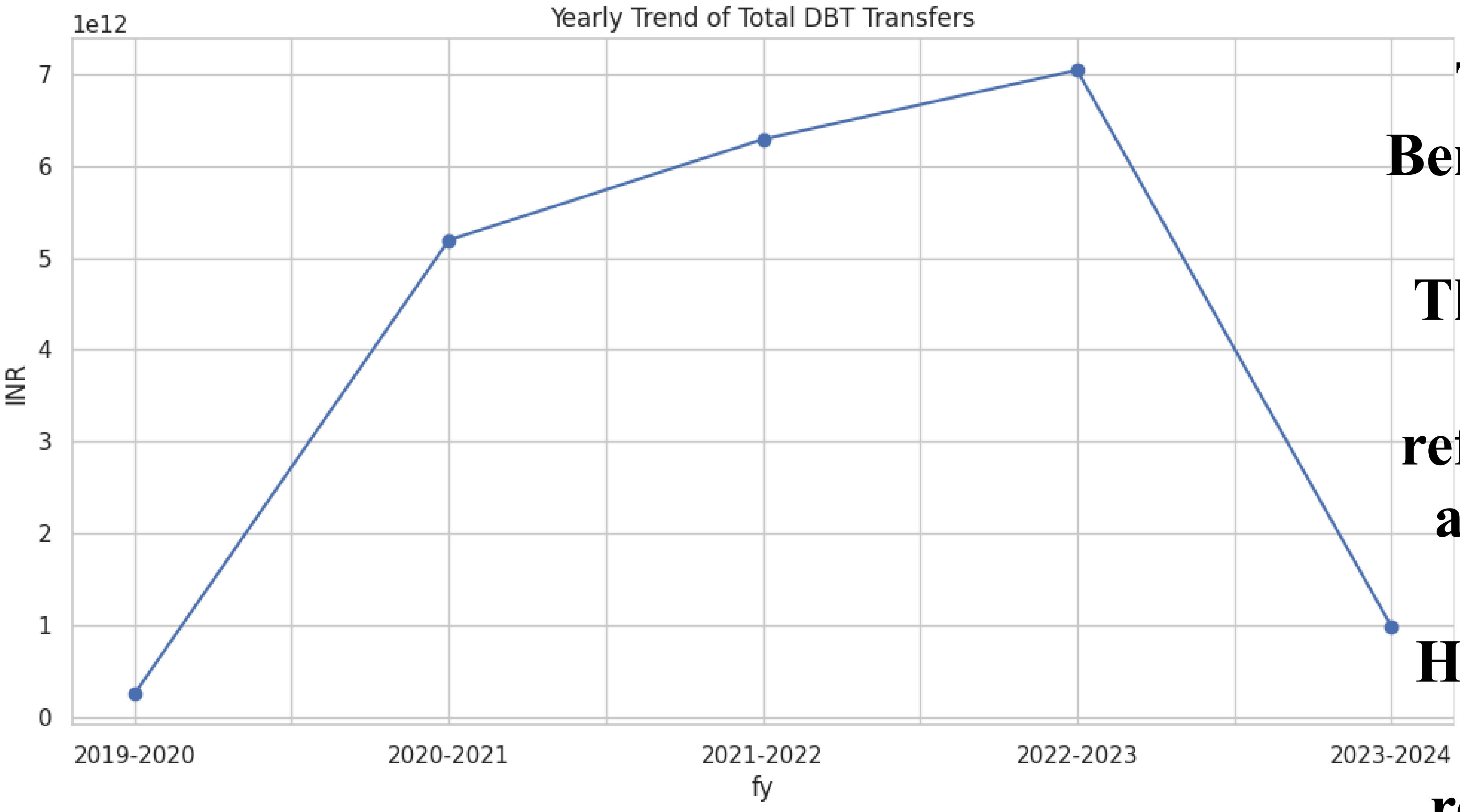
INTRODUCTION

- **The Direct Benefit Transfer (DBT) system is a flagship initiative of the Government of India aimed at delivering subsidies and welfare benefits directly into beneficiaries' bank accounts.**
- **It eliminates intermediaries, reduces delays, and promotes transparency in public fund distribution.**
- **This project analyzes DBT transaction data to uncover patterns, trends, and insights at both state and district levels.**
- **The study also highlights yearly progress, top-performing regions, and the overall impact of DBT on financial inclusion.**
- **The goal is to understand how data-driven governance enhances efficiency, accountability, and citizen empowerment.**

DATA CLEANING AND PREPROCESSING

Data cleaning and preprocessing are crucial steps in preparing a dataset for analysis, ensuring accuracy, consistency, and readiness for exploration. In this process, several key steps were undertaken to refine the dataset. Firstly, column names were standardized by converting them to lowercase and replacing spaces with underscores, enhancing readability and consistency across the dataset. Missing data was addressed by identifying **121 rows** with **absent values** in critical columns like total_dbt_transfer and no_of_dbt_transactions, which were subsequently dropped to avoid calculation errors. A thorough check for duplicate rows confirmed the dataset's uniqueness, as **no duplicates were found**. Additionally, data types were verified and adjusted, with numeric columns set to float and the fy column converted to a string for easier categorical analysis. As a result, the dataset now comprises 3,704 entries, providing a clean and reliable foundation for further exploration and analysis.

Yearly Trend of Total DBT Transfers

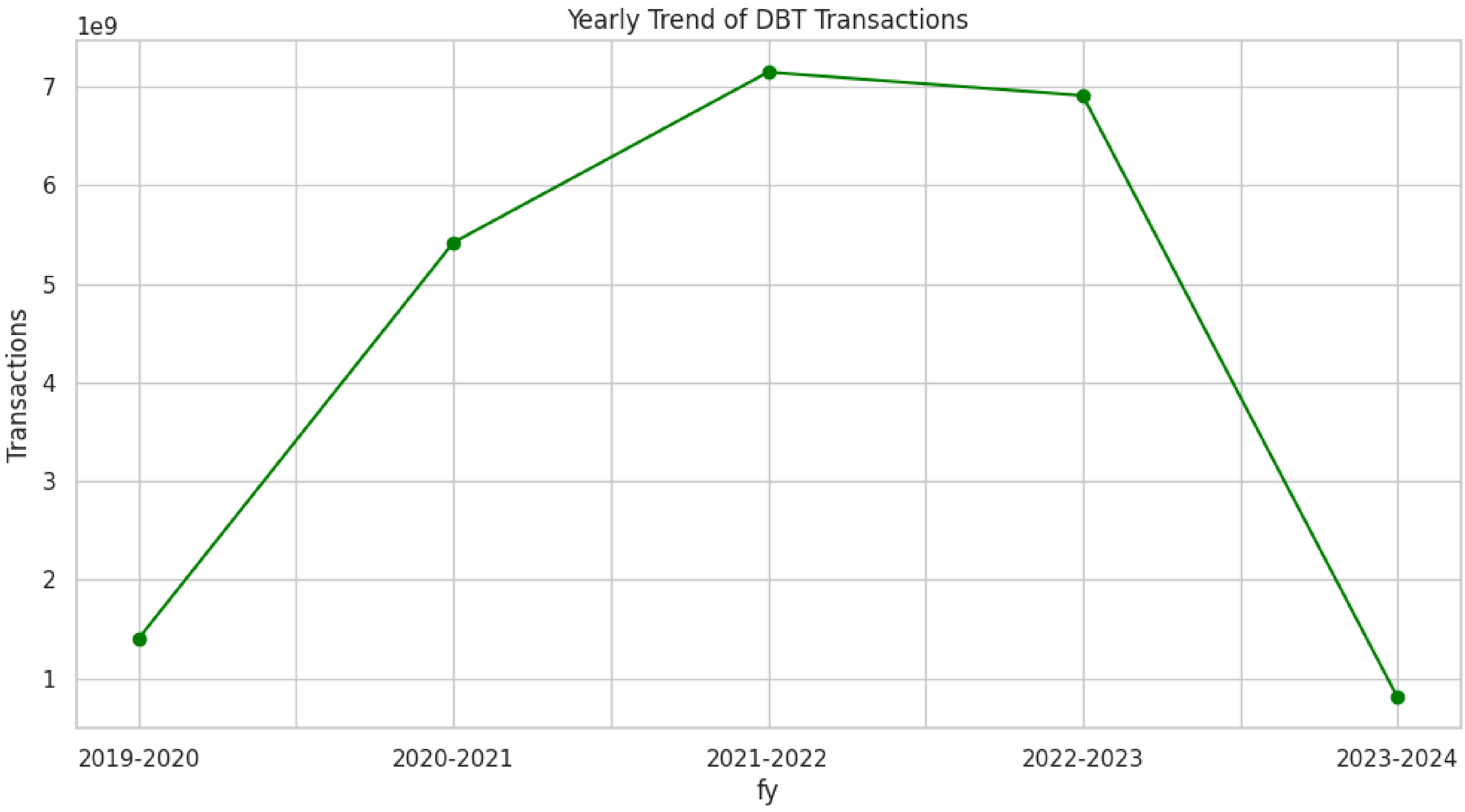


The graph depicts the total value of Direct Benefit Transfers (DBT) from 2019–20 to 2023–24.

There was a sharp increase from ₹0.25 trillion in 2019–20 to over ₹7 trillion in 2022–23, reflecting the government's major digital push and welfare expansion during the pandemic years.

However, in 2023–24, the total transfers saw a significant decline, possibly due to scheme restructuring and improved data validation.

Yearly Trend of DBT Transactions

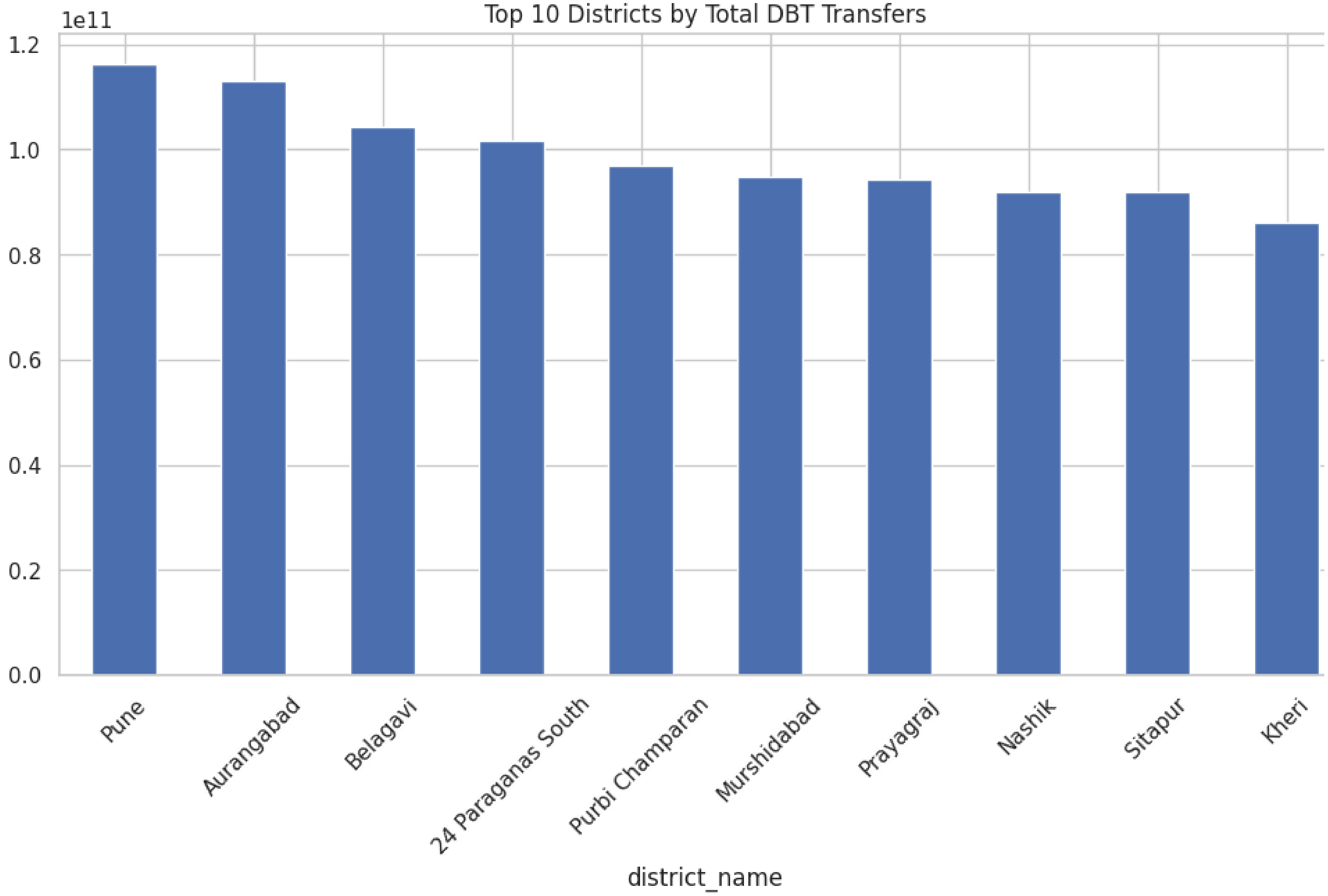


The chart illustrates the yearly trend of Direct Benefit Transfer (DBT) transactions from 2019–20 to 2023–24.

There was a sharp rise from 2019–20 to 2021–22, peaking at over 7 billion transactions, reflecting major digital adoption and welfare expansion during and after the pandemic.

A slight decline was seen in 2022–23, followed by a steep drop in 2023–24, likely due to restructuring of schemes and data consolidation.

Top 10 Districts by Total DBT Transfers

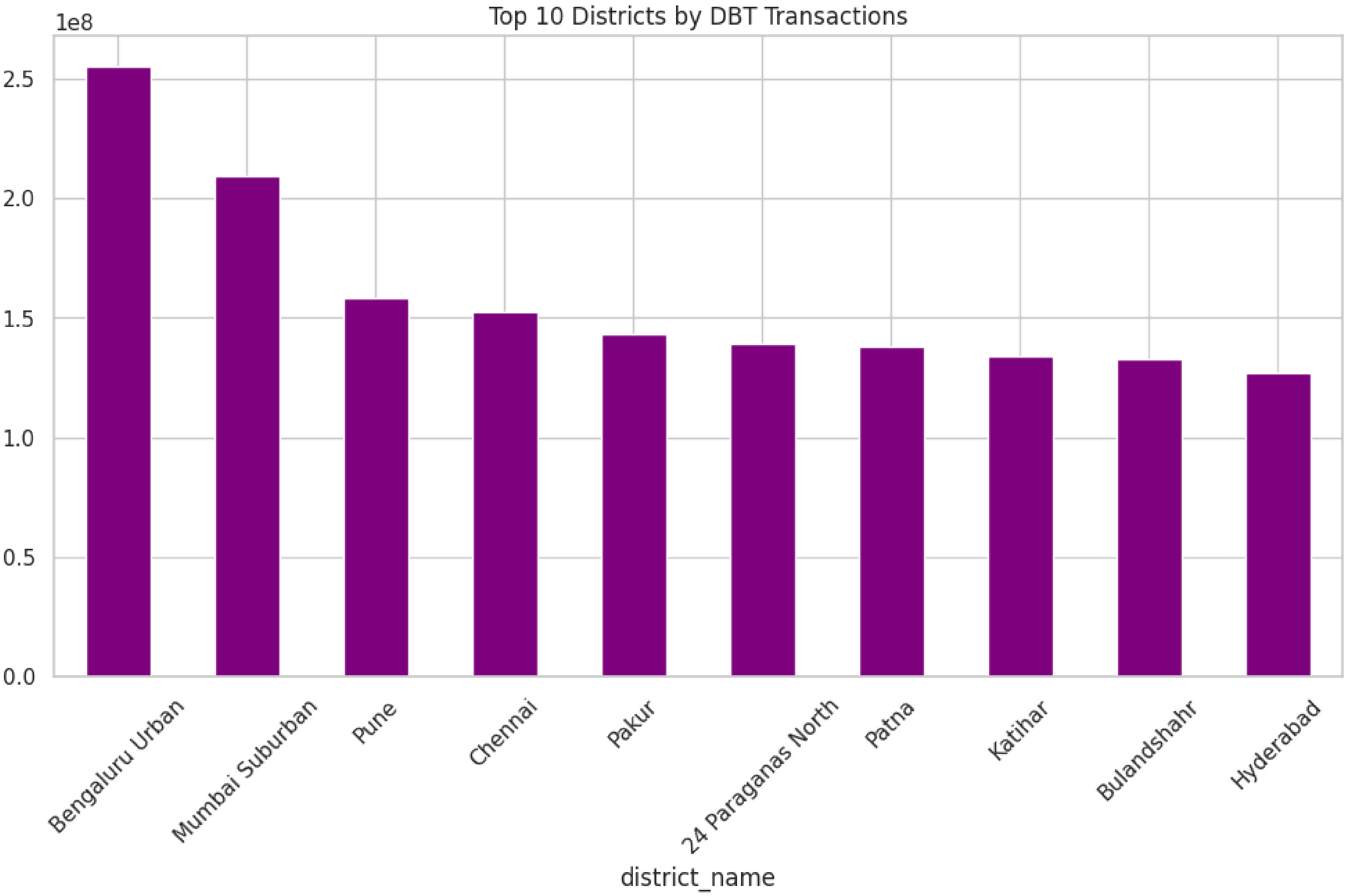


This chart displays the districts that have received the highest total amount of Direct Benefit Transfers (DBT) in India.

Pune leads the list, followed by Aurangabad, Belagavi, and South 24 Parganas, indicating that economically significant and populous districts handle larger volumes of welfare fund transfers.

The high transfer amounts reflect strong DBT coverage, effective digital infrastructure, and inclusion under multiple government schemes, showing the concentration of welfare disbursement in major regional hubs.

Top 10 Districts by DBT Transactions

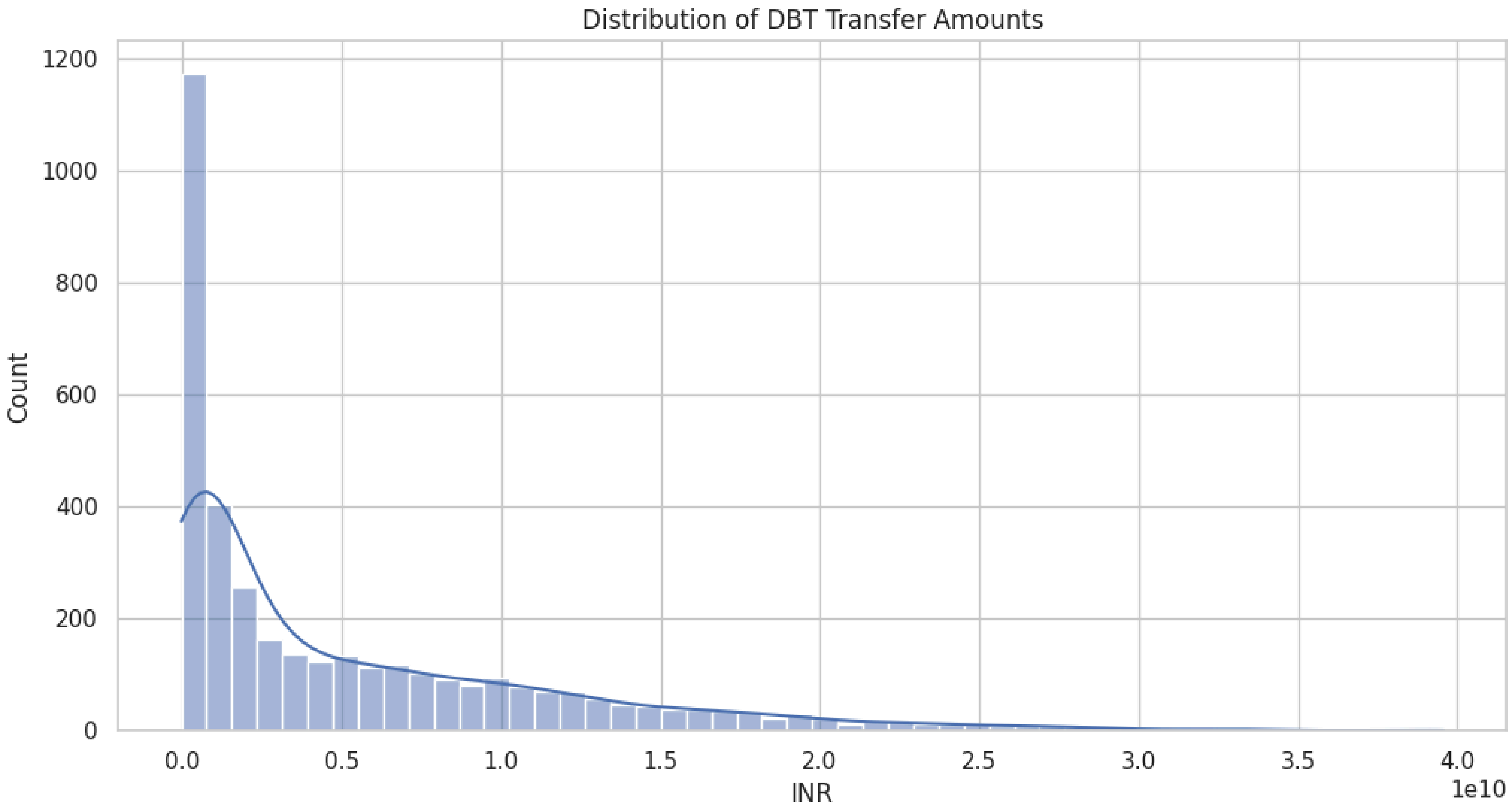


This chart highlights the districts with the highest number of Direct Benefit Transfer (DBT) transactions across India.

Bengaluru Urban leads with the maximum transactions, followed by Mumbai Suburban, Pune, and Chennai, indicating that urban and economically active regions have a higher concentration of DBT activity.

The trend reflects that digitally advanced and densely populated districts manage a large share of benefit disbursements, showing the strong adoption of DBT platforms in urban centers.

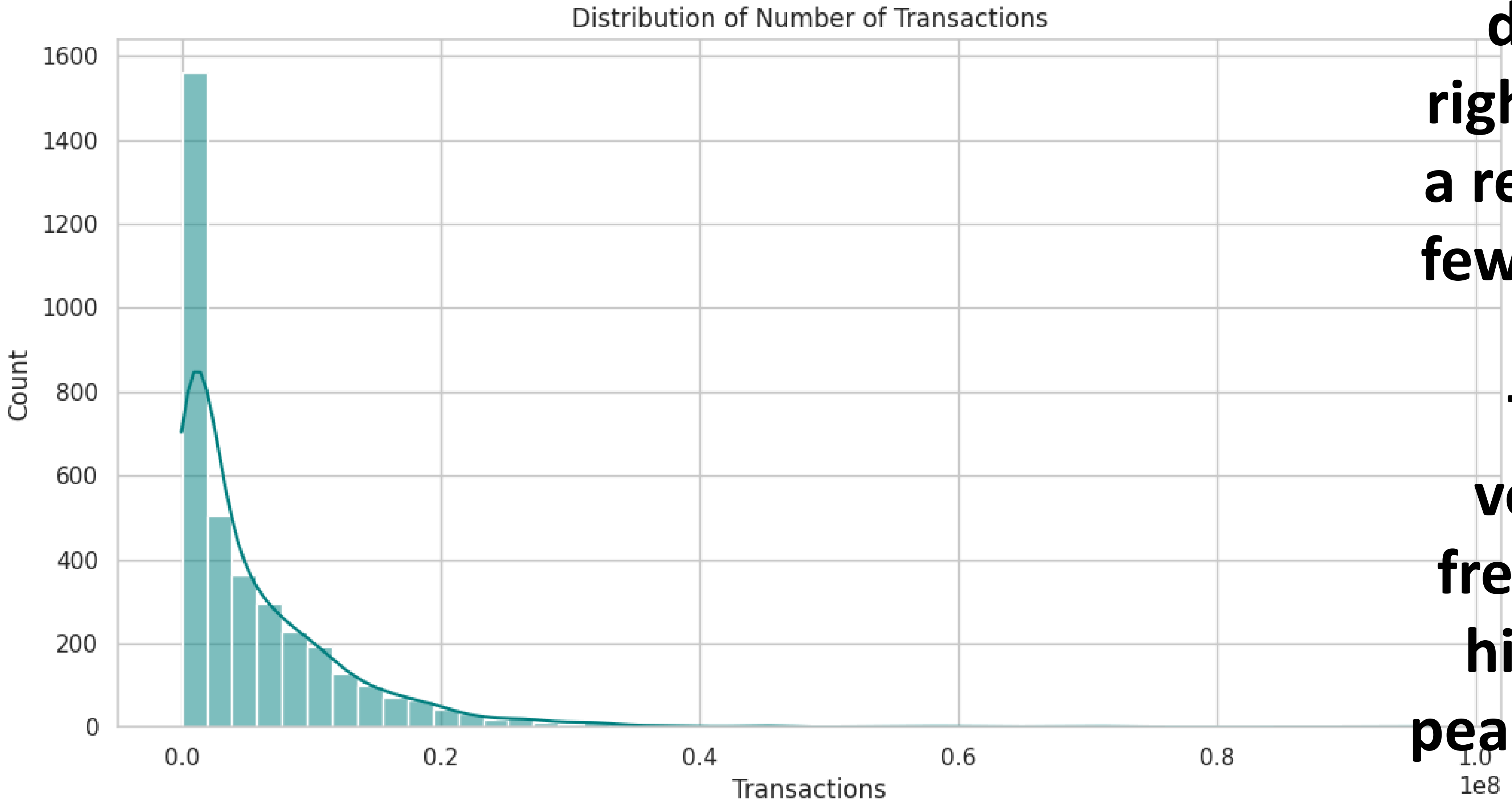
Distribution of DBT Transfer Amounts



This chart illustrates the distribution of total Direct Benefit Transfer (DBT) amounts across districts or states. The distribution is highly right-skewed, showing that most regions handle relatively smaller total DBT amounts, while a few regions receive or distribute exceptionally large sums.

The tall bar on the left represents areas with low overall fund transfers, possibly due to smaller populations or limited scheme coverage. The long tail on the right signifies a few high-value regions, where large-scale welfare disbursements are concentrated.

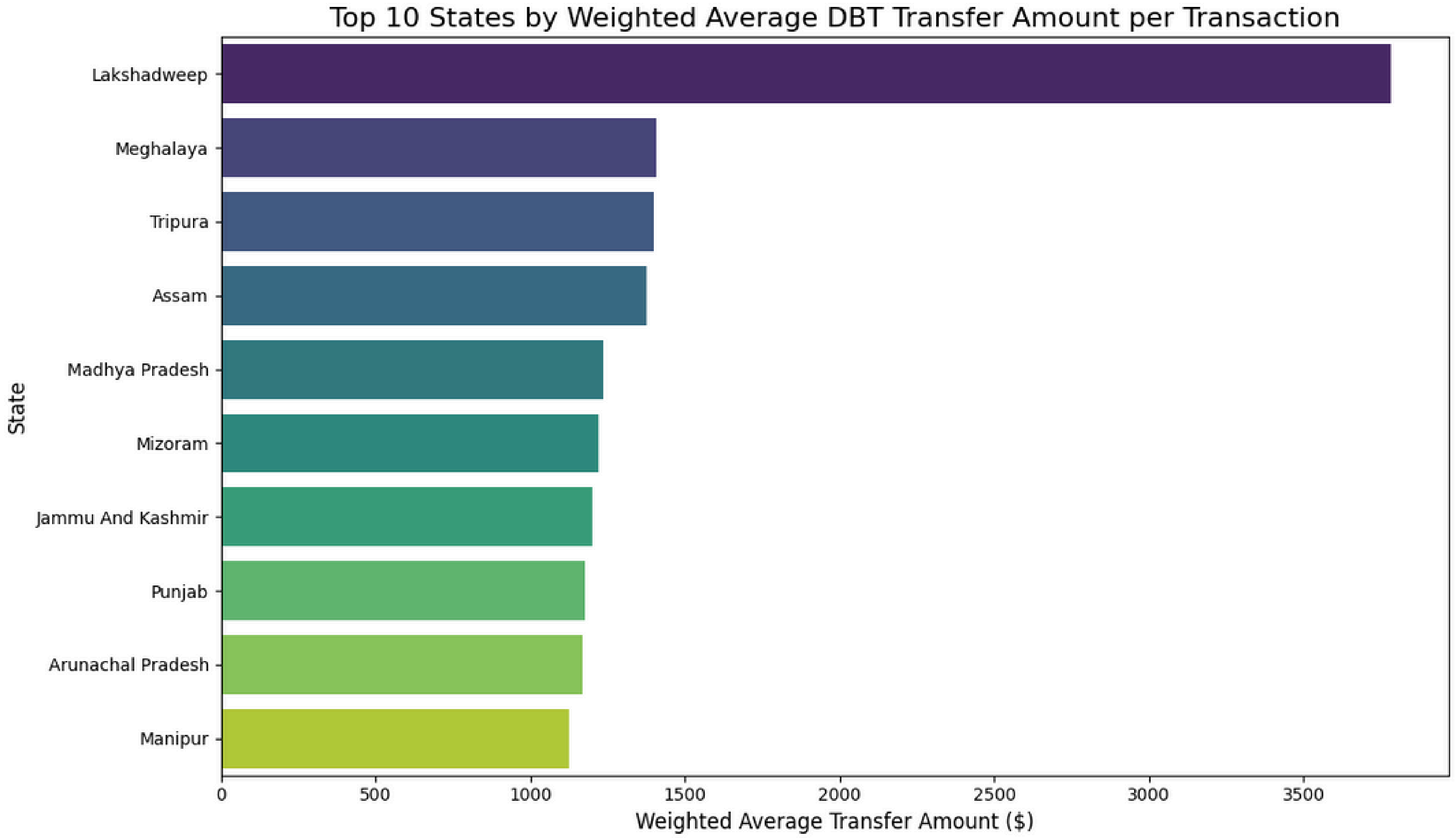
Distribution of Number of Transactions



This graph shows the distribution of the total number of DBT transactions across different districts or states. The distribution is highly right-skewed, indicating that most regions have a relatively low number of transactions, while a few regions have an exceptionally large number of transactions.

The long tail on the right represents high-volume areas where DBT transfers are more frequent, possibly due to larger populations or higher coverage under welfare schemes. The peak on the left shows that a majority of regions engage in smaller-scale DBT activity.

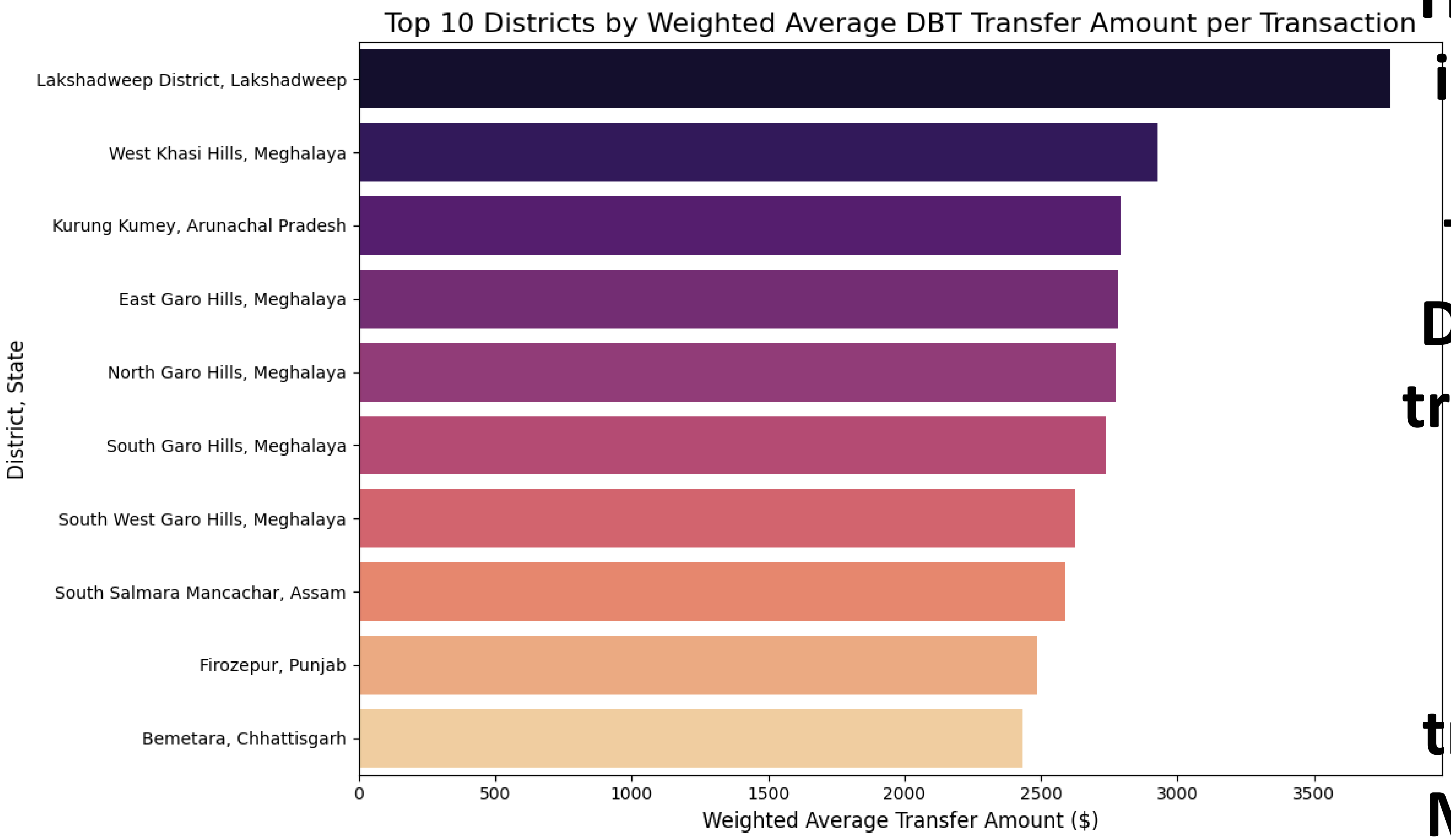
Top 10 States by Weighted Average DBT Transfer Amount per Transaction



The chart illustrates the top 10 Indian states and union territories ranked by their weighted average DBT transfer amount per transaction. This metric indicates the average value of each DBT transaction, providing insight into the scale and intensity of benefit disbursements.

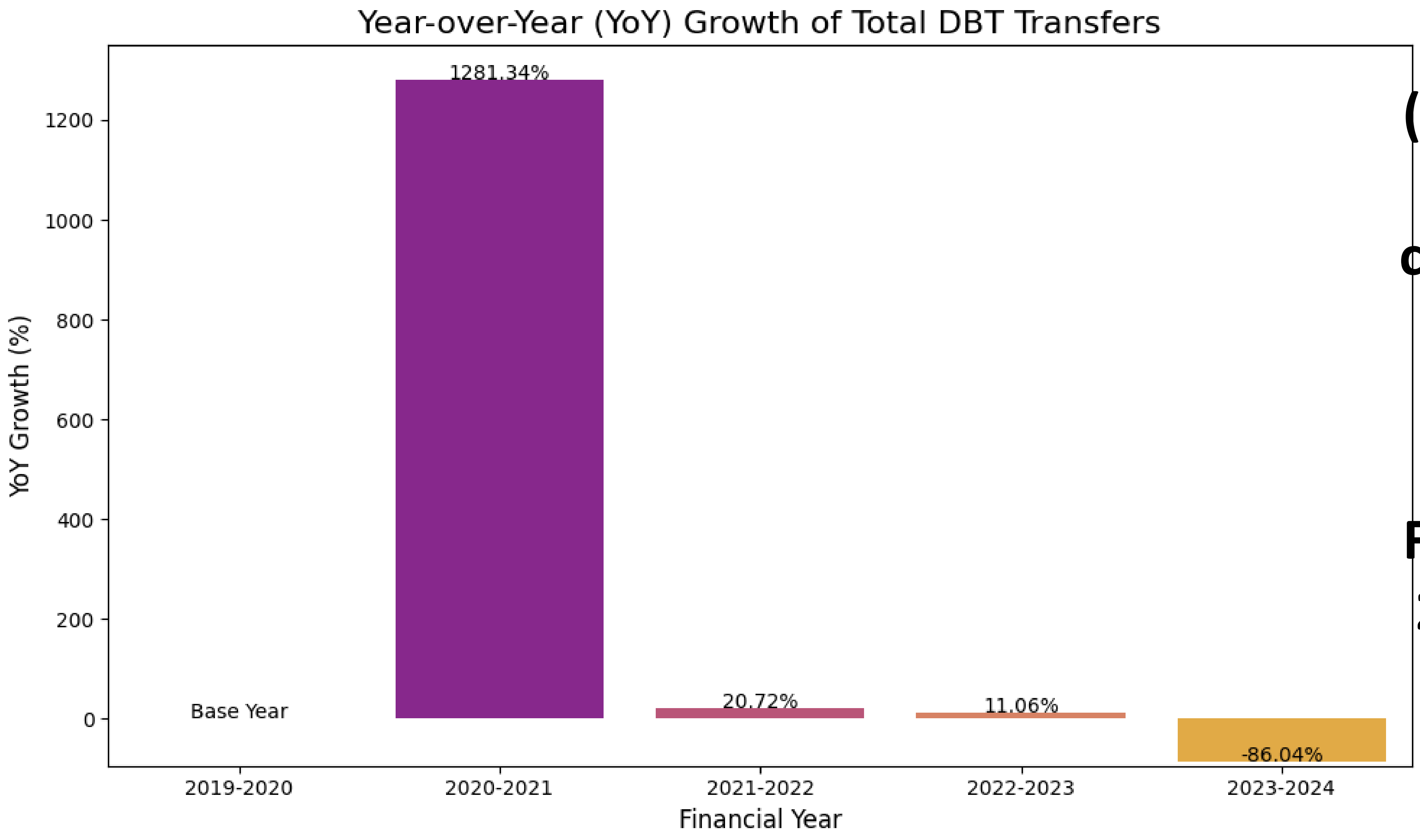
Lakshadweep leads by a significant margin, suggesting a small number of high-value transactions. Meghalaya, Tripura, and Assam follow, reflecting substantial average transfers in the northeastern region. Other states like Madhya Pradesh, Mizoram, Punjab, and Jammu & Kashmir also feature prominently, indicating a balanced mix of transaction volume and value.

Top 10 Districts by Weighted Average DBT Transfer Amount per Transaction



The bar chart above highlights the top 10 districts in India ranked by their weighted average Direct Benefit Transfer (DBT) amount per transaction. This measure reflects the average value of each DBT transaction, accounting for the total amount transferred and the number of transactions made. The Lakshadweep District stands out with the highest average transfer amount, significantly above other regions, indicating large-value transactions with fewer recipients. Districts from Meghalaya — such as West Khasi Hills, East Garo Hills, North Garo Hills, and South Garo Hills — also feature prominently, suggesting concentrated DBT activities with substantial average payouts.

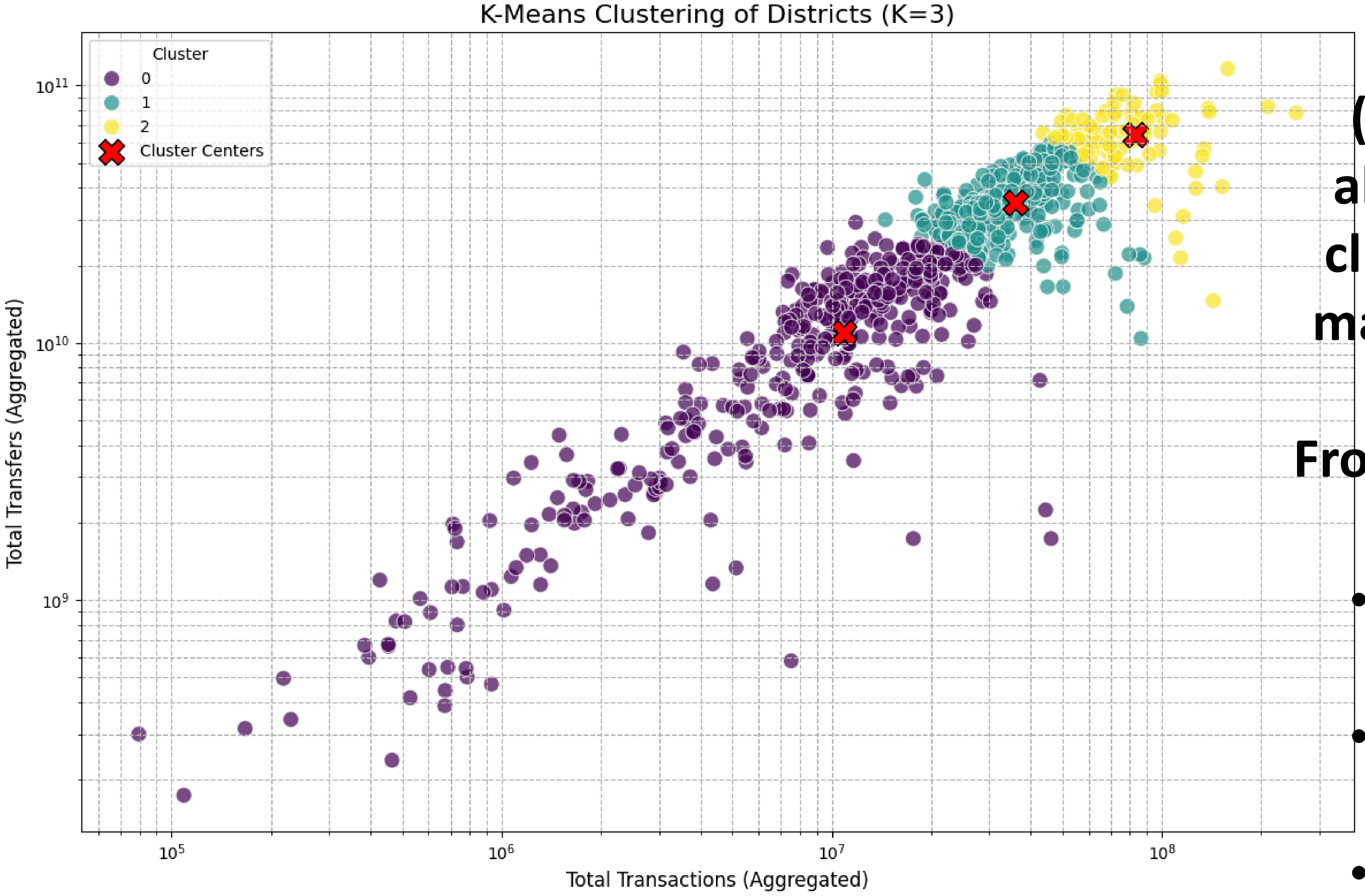
Year-over-Year (YoY) Growth of Total DBT Transfers



The bar chart above illustrates the Year-over-Year (YoY) growth rate of Total Direct Benefit Transfer (DBT) Transfers across financial years. The base year (2019–2020) marks the starting point for comparison. A significant surge is observed in 2020–2021, with a remarkable 1281.34% increase, likely due to large-scale DBT initiatives implemented during the pandemic period.

Following this, growth stabilized to 20.72% in 2021–2022 and 11.06% in 2022–2023, indicating a period of steady but moderate expansion. However, in 2023–2024, the trend reversed, showing a sharp decline of -86.04%, possibly due to reduced fund allocation or normalization of post-pandemic programs.

K-Means Clustering of Districts (K=3)

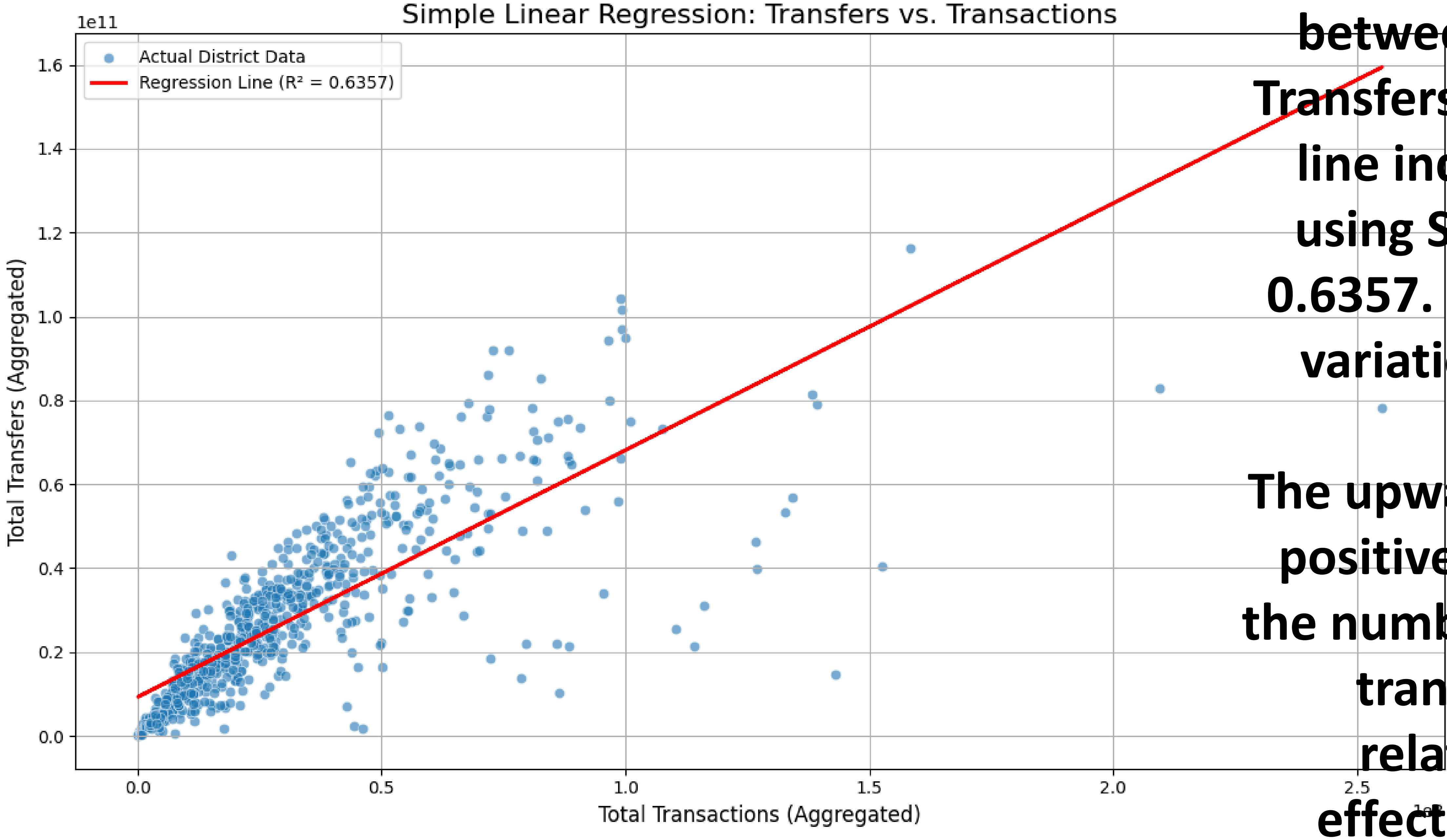


The scatter plot above shows the K-Means clustering results of districts based on Total Transactions (Aggregated) and Total Transfers (Aggregated). The algorithm was set to K = 3, resulting in three distinct clusters represented by different colors. Each red "X" marks the cluster center, indicating the mean position of data points within that cluster.

From the visualization, we can observe a clear grouping pattern:

- **Cluster 0 (Purple): Districts with low transaction and transfer volumes.**
- **Cluster 1 (Teal): Districts with moderate transactions and transfers.**
- **Cluster 2 (Yellow): Districts with high transaction and transfer activity.**

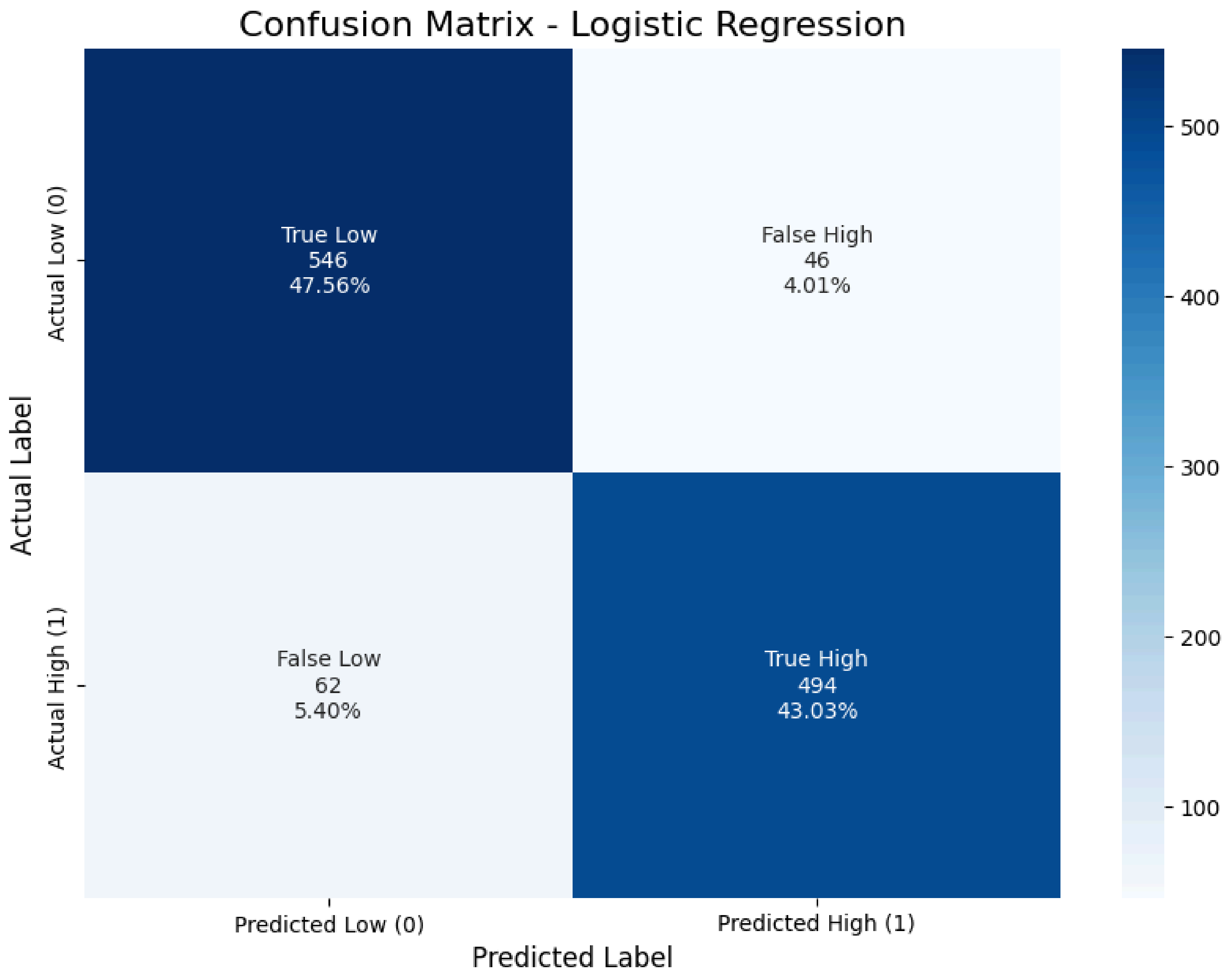
Simple Linear Regression – Transfers vs. Transactions



The scatter plot above represents the relationship between Total Transactions (Aggregated) and Total Transfers (Aggregated) across various districts. The red line indicates the best-fit regression line generated using Simple Linear Regression, with an R^2 value of 0.6357. This means that approximately 63.57% of the variation in total transfers can be explained by the number of transactions.

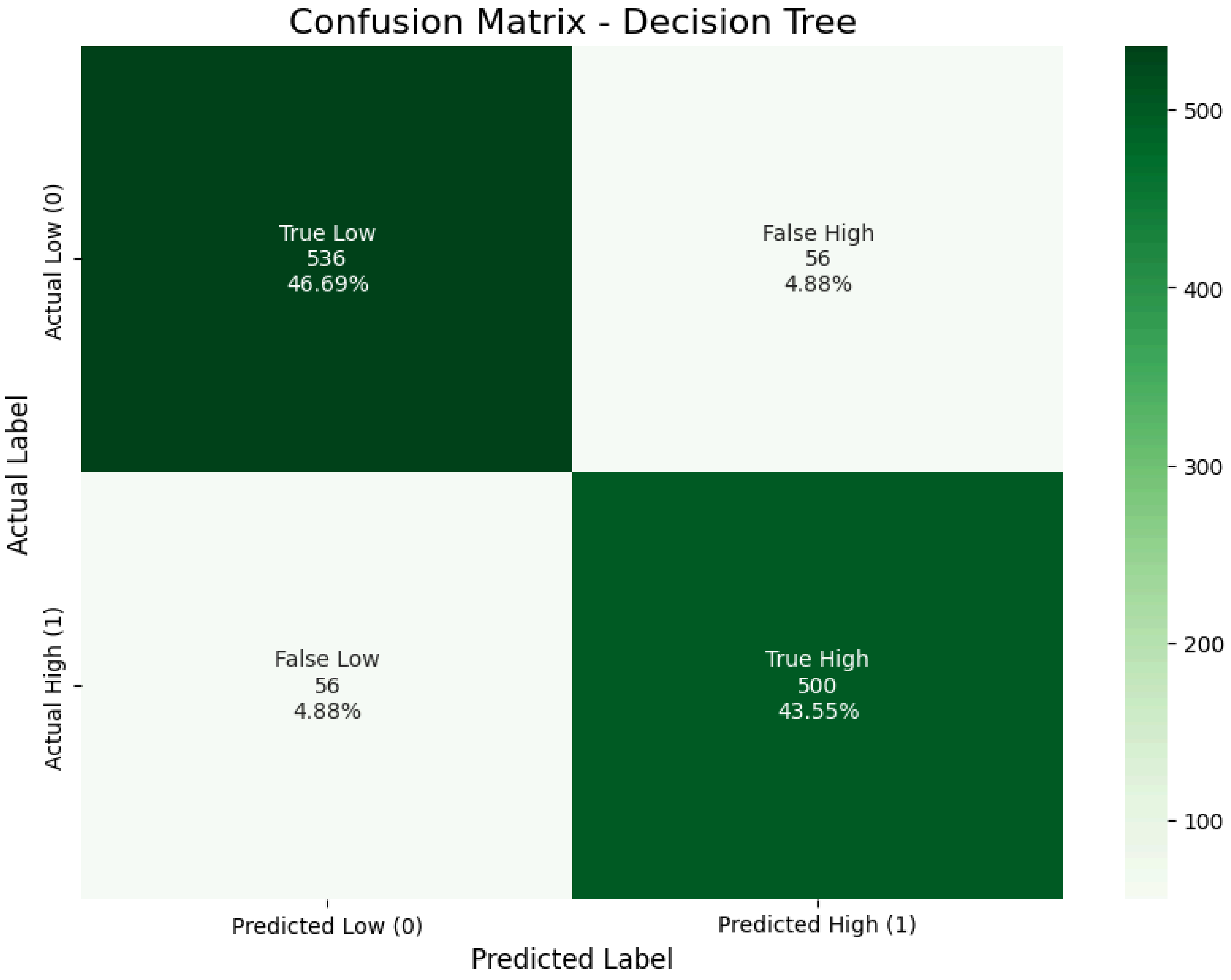
The upward trend in the regression line shows a strong positive correlation between the two variables — as the number of transactions increases, the total value of transfers also tends to increase. Although the relationship is not perfectly linear, the model effectively captures the overall trend in the data, making it a good fit for understanding the financial interaction pattern among districts.

Confusion Matrix Interpretation – Logistic Regression



- **The confusion matrix above represents the performance of the Logistic Regression model in classifying the data into two categories: Low (0) and High (1). Out of the total predictions, the model correctly identified 546 True Lows and 494 True Highs, while it made 46 False High and 62 False Low predictions. This indicates that the model achieved an overall accuracy of approximately 90.6%, showing strong predictive capability.**
- **The precision for the High class is around 91.5%, meaning that when the model predicts "High," it is correct most of the time. The recall for the High class is about 88.9%, showing that the model successfully detects most of the actual High cases. The F1-score, which balances precision and recall, is approximately 90.2%, indicating consistent and reliable performance across both classes.**
- **Overall, the confusion matrix demonstrates that the Logistic Regression model performs effectively, with minimal misclassifications, and can be considered a good fit for the given dataset.**

Confusion Matrix Interpretation – Decision Tree



The confusion matrix above illustrates the performance of the Decision Tree model in classifying the data into Low (0) and High (1) categories. The model correctly predicted 536 True Lows and 500 True Highs, while it made 56 False High and 56 False Low predictions. This shows that the Decision Tree achieved a balanced performance between both classes.

The model's overall accuracy is approximately 90.2%, which is slightly lower but comparable to the Logistic Regression model. The precision for the High class is around 89.9%, and the recall is about 89.9% as well, indicating that the model maintains a good balance between correctly identifying and predicting the "High" cases.

CONCLUSION

- The analysis provided valuable insights into Direct Benefit Transfer (DBT) trends across India.
- Uttar Pradesh, Bihar, and Maharashtra emerged as top-performing states in terms of transaction volume.
- Over the years, DBT transfers have shown significant growth, highlighting the government's commitment to transparency and efficiency.
- The data revealed how technology-driven governance has streamlined welfare delivery to beneficiaries.
- Clean and preprocessed data enabled accurate visualizations and meaningful conclusions.