Analyzing and Enhancing the UPI Ecosystem

MTH209 Project Report

Tathagata Banerjee, Ahana Bose, Rohit Karwa, Ashirvad Pawar, Gunavant Thakare

Indian Institute of Technology, Kanpur

October 18, 2024

Contents

1 Introduction

Unified Payments Interface (UPI) has revolutionized digital payments in India, providing an efficient and popular platform for transferring funds between bank accounts. It was developed by the National Payments Corporation of India (NPCI), and has gained immense popularity over the last few years. UPI enables users to link multiple bank accounts to a single application on their phones, eliminating the need for traditional payment methods like cash or cheques. PhonePe,Google Pay and Paytm are some of the leading applications, catering to users of most of the popular banks.

Users can utilize the UPI applications available in their devices to initiate transactions in real-time, 24/7, from anywhere with an internet connection. The system utilizes virtual payment addresses (VPAs), also known as UPI IDs, or mobile numbers linked to bank accounts, ensuring secure and instant fund transfers.

Since its launch, UPI has witnessed massive growth, with millions of transactions being undertaken daily. Today, UPI is used in almost all dimensions of financial transactions, ranging from payment of bank loans,account-to-account transfers to everyday purchases such as food and stationary,just by scanning QR codes provided by the merchants. It is used for bill payments and online purchases as well.

1.1 About the Dataset

In this study, we will be analyzing the UPI transactions of the most popular banks over the last few years. The data has been collected from the NPCI official website, since it is considered to be a reliable and official source of information. The dataset contains the information of the top 50 beneficiary and remittance banks and their total volume of transactions, approval amount, business decline, technical decline, total reversal count, debit reversal success amount, deemed approved amount from August 2021 to January 2024. The banks have been ranked on the basis of the total volume remitted and total volume received, respectively.

The different variables of the dataset have been described in the following list. **R** and **B** indicate whether the variable considered belongs to the remittance bank data or beneficiary bank data, respectively.

- **Remitter Bank(R)**: The bank of the account holder who is sending the money.
- **Beneficiary Bank(B)**: The bank of the account holder who is receiving money.
- **Total Volume(R/B)**: Total quantity of transactions (in millions) processed in a given month.

- **Approved Transaction Volume(R/B)**: A transaction marked as approved indicates that it has passed all necessary checks and has been successfully authorized by the sender's bank and recipient's bank.
- **Business Decline (BD)(R/B)**: Transaction decline due to a customer entering an invalid PIN, incorrect beneficiary account, or due to other business reasons such as exceeding per transaction limit, exceeding permitted count of transactions per day, exceeding amount limit for the day, etc.
- **Technical Decline (TD)(R/B)**: Transaction decline due to technical reasons, such as unavailability of systems and network issues on bank or NPCI side.
- **Total Debit Reversal Count(R)**: It refers to the total number of transactions (in millions) where a debit has been reversed, which means that the initial debit transaction has been undone and the funds have been returned to the account.
- **Debit Reversal Success Amount(R)**: Indicates the volume of transactions where a customer account may be debited and their bank is unable to confirm instantly about the status of reversal of such a debit.
- **Deemed Approved Amount(B)**: Indicates the total volume of transactions, where credit confirmations are not received online from the beneficiary banks for the credit.

2 Exploratory Data Analysis

2.1 Principal Component Analysis

In order to proceed with the analysis, we consider only the banks which are common in both the top rankings of beneficiary and remitter banks. Applying Principal Component Analysis on the updated dataset, we obtain the following findings.

Table 1: Principal Component Analysis Summary

Variable	Proportion of Variance Explained
Remittance Total Volume	0.58
Beneficiary Total Volume	0.78
DRA	0.9
Remittance Approved Volume	0.94
DRS Remittance	0.98
Approved Beneficiary Volume	0.987
Remittance BD Volume	0.996
Beneficiary BD Volume	1

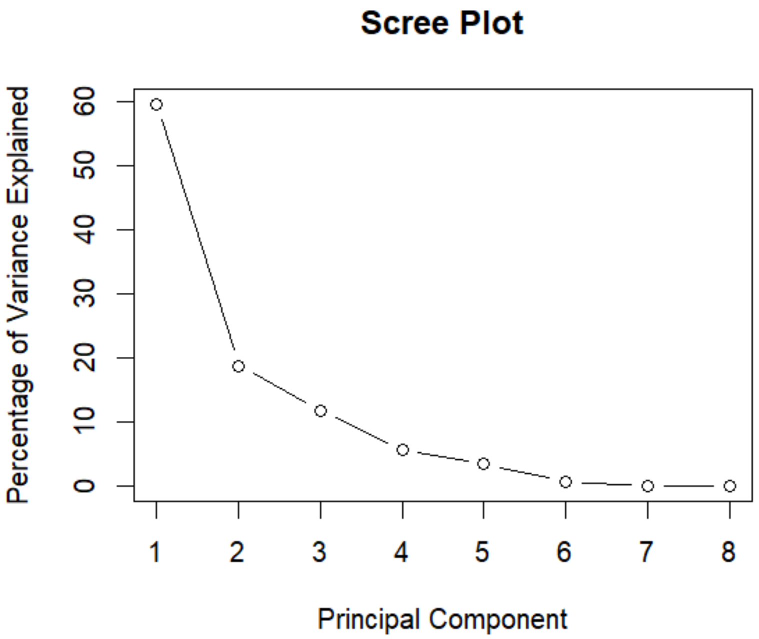


Figure 1: Scree Plot

Since first two variables explain most of the variability, we will focus our analysis mostly on the total volumes in this study, and analyzing the other variables wherever possible.

2.2 Some Visualizations

In this study, we are analyzing the beneficiary and remittance behaviour of the most popular banks. Since the banks are ranked according to their total volume(remittance/beneficiary), those variables are of special interest. We plot the histograms of these variables in the following plots.

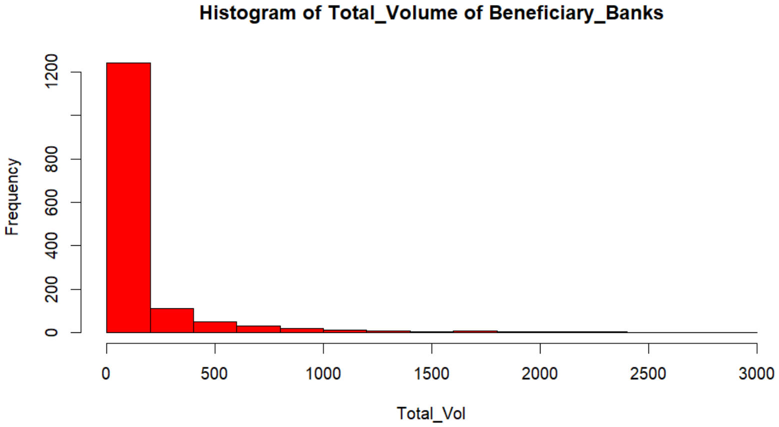


Figure 2: Histogram of Total Beneficiary Volume

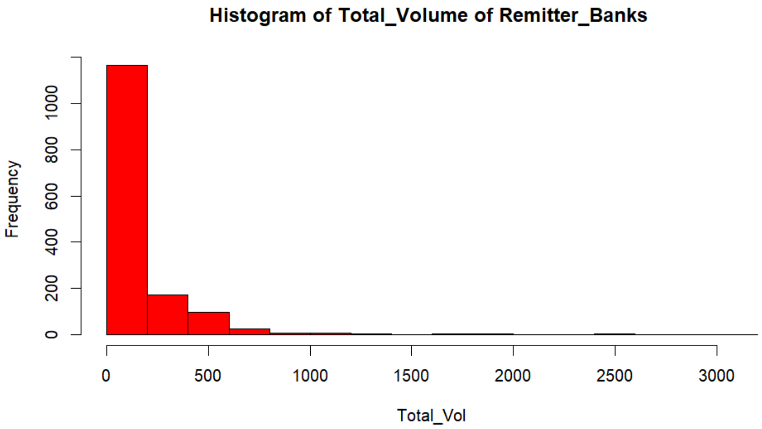
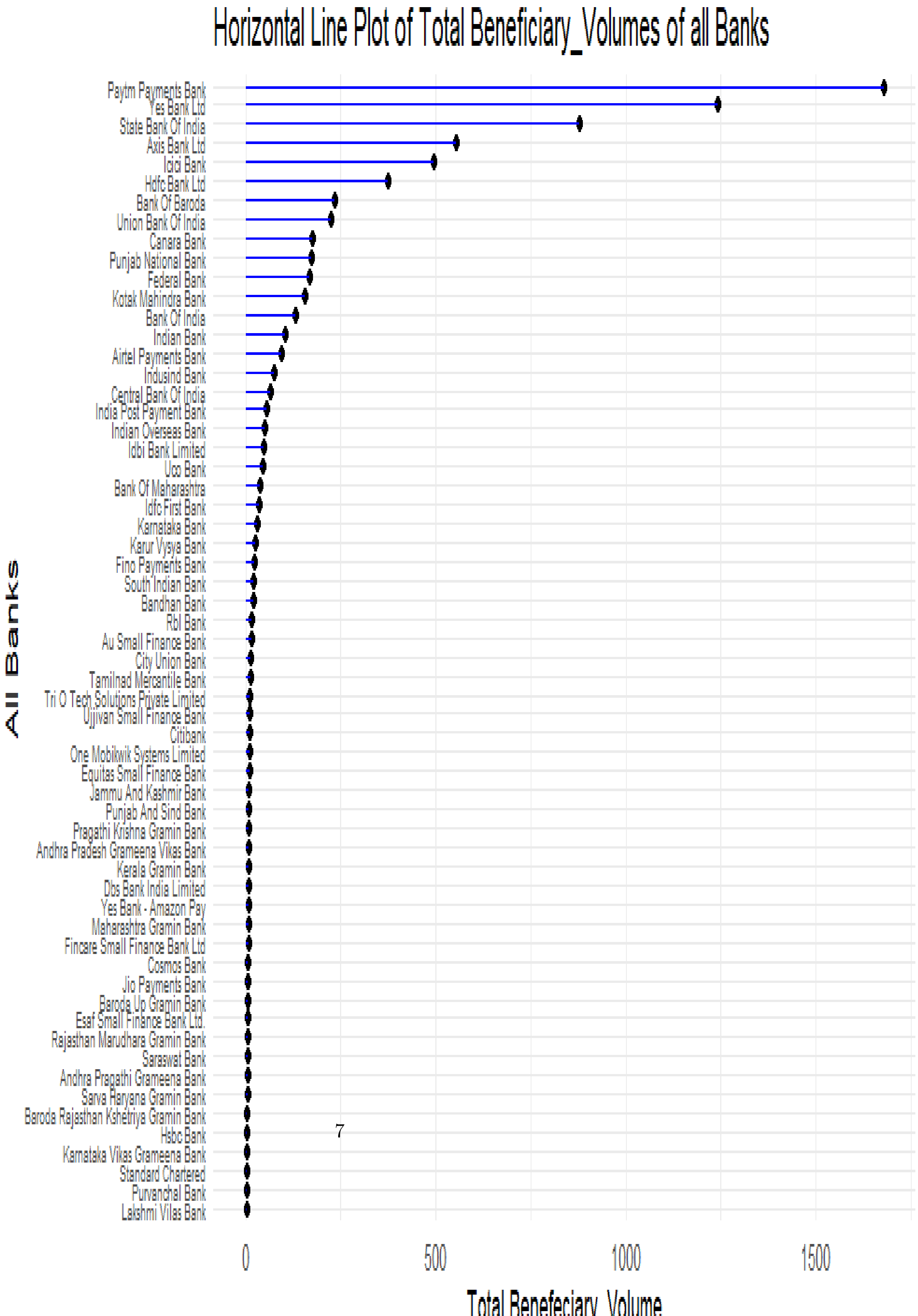
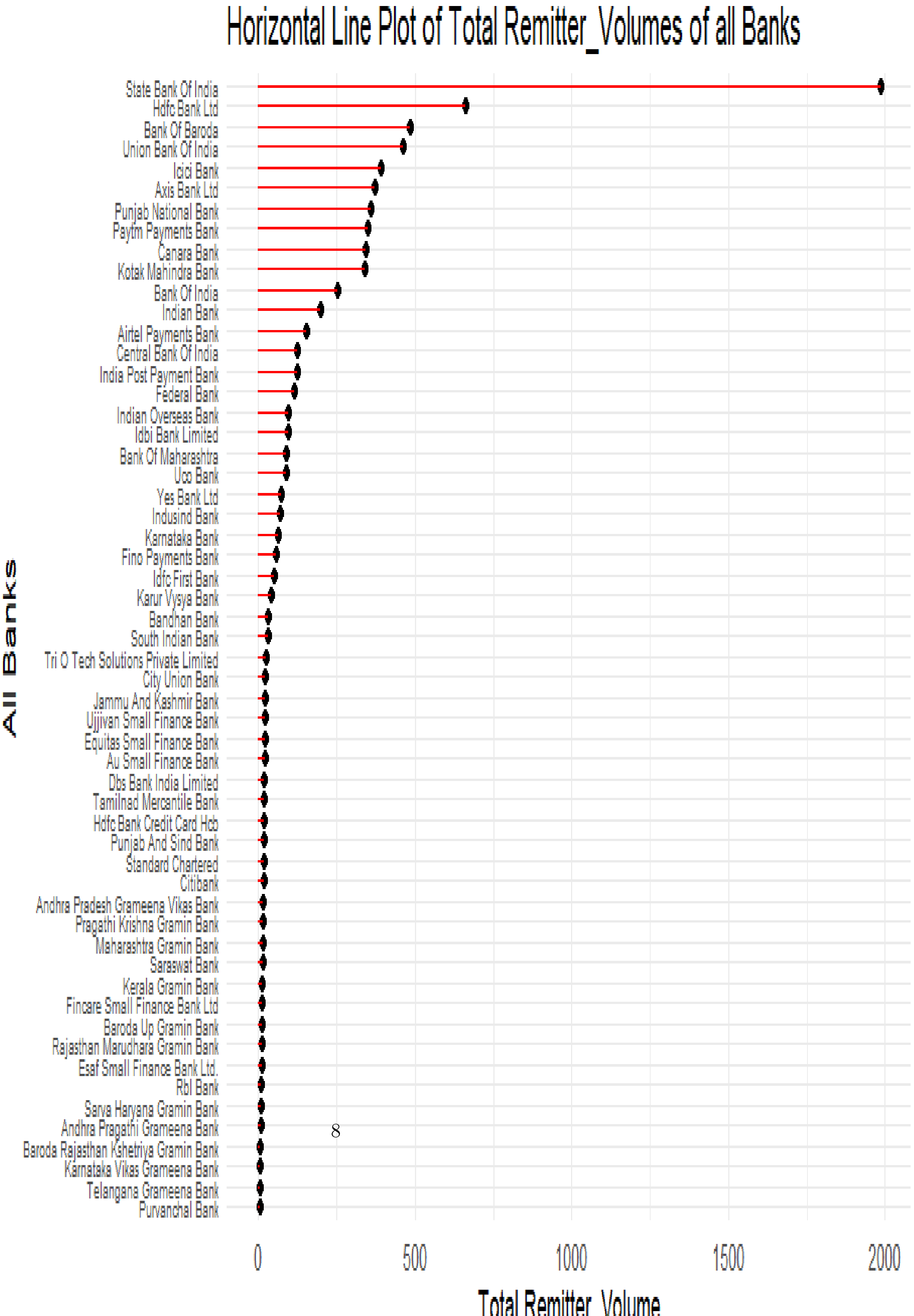


Figure 3: Histogram of Total Remittance Volume

From these plots, it is evident the total beneficiary and remittance volume are both highly positively skewed, possibly due to the presence of few banks with very high remittance and beneficiary volumes as compared to the others.

To verify this idea and get a better understanding of the distribution of total volumes across banks, we obtain the mean remittance and beneficiary volume for all banks, which are plotted in the following plots.


Horizontal Line Plot of Total Beneficiary_Volumes of all Banks
All Banks
Paytm Payments Bank
Yes Bank Ltd
State Bank Of India
Axis Bank Ltd
Icici Bank
Hdfc Bank Ltd
Bank Of Baroda
Union Bank Of India
Canara Bank
Punjab National Bank
Federal Bank
Kotak Mahindra Bank
Bank Of India
Indian Bank
Airtel Payments Bank
Indusind Bank
Central Bank Of India
India Post Payment Bank
Indian Overseas Bank
Idbi Bank Limited
Uco Bank
Bank Of Maharashtra
Idfc First Bank
Karnataka Bank
Karur Vysya Bank
Fino Payments Bank
South Indian Bank
Bandhan Bank
Rbl Bank
Au Small Finance Bank
City Union Bank
Tamilnad Mercantile Bank
Tri O Tech Solutions Private Limited
Ujjivan Small Finance Bank
Citibank
One Mobikwik Systems Limited
Equitas Small Finance Bank
Jammu And Kashmir Bank
Punjab And Sind Bank
Pragathi Krishna Gramin Bank
Andhra Pradesh Grameena Vikas Bank
Kerala Gramin Bank
Dbs Bank India Limited
Yes Bank - Amazon Pay
Maharashtra Gramin Bank
Fincare Small Finance Bank Ltd
Cosmos Bank
Jio Payments Bank
Baroda Up Gramin Bank
Esaf Small Finance Bank Ltd.
Rajasthan Marudhara Gramin Bank
Saraswat Bank
Andhra Pragathi Grameena Bank
Sarva Haryana Gramin Bank
Baroda Rajasthan Kshetriya Gramin Bank
Hsbc Bank
Karnataka Vikas Grameena Bank
Standard Chartered
Purvanchal Bank
Lakshmi Vilas Bank
0
500
1000
1500
Total Benefeciary_Volume


Horizontal Line Plot of Total Remitter_Volumes of all Banks
All Banks
State Bank Of India
Hdfc Bank Ltd
Bank Of Baroda
Union Bank Of India
Icici Bank
Axis Bank Ltd
Punjab National Bank
Paytm Payments Bank
Canara Bank
Kotak Mahindra Bank
Bank Of India
Indian Bank
Airtel Payments Bank
Central Bank Of India
India Post Payment Bank
Federal Bank
Indian Overseas Bank
Idbi Bank Limited
Bank Of Maharashtra
Uco Bank
Yes Bank Ltd
Indusind Bank
Karnataka Bank
Fino Payments Bank
Idfc First Bank
Karur Vysya Bank
Bandhan Bank
South Indian Bank
Tri O Tech Solutions Private Limited
City Union Bank
Jammu And Kashmir Bank
Ujjivan Small Finance Bank
Equitas Small Finance Bank
Au Small Finance Bank
Dbs Bank India Limited
Tamilnad Mercantile Bank
Hdfc Bank Credit Card Hcb
Punjab And Sind Bank
Standard Chartered
Citibank
Andhra Pradesh Grameena Vikas Bank
Pragathi Krishna Gramin Bank
Maharashtra Gramin Bank
Saraswat Bank
Kerala Gramin Bank
Fincare Small Finance Bank Ltd
Baroda Up Gramin Bank
Rajasthan Marudhara Gramin Bank
Esaf Small Finance Bank Ltd.
Rbl Bank
Sarva Haryana Gramin Bank
Andhra Pragathi Grameena Bank
Baroda Rajasthan Kshetriya Gramin Bank
Karnataka Vikas Grameena Bank
Telangana Grameena Bank
Purvanchal Bank
0
500
1000
1500
2000
Total Remitter_Volume

From the beneficiary volume plot, we can conclude that Paytm Payments Bank dominates the beneficiary volume list,followed by Yes Bank Ltd and State Bank of India. Also,the lower ranked banks have similar beneficiary volumes,which are very low as compared to the top banks.

From the remittance volume plot, we can conclude that State Bank of India dominates in remittance volumes . In fact, it's total volume is significantly higher than that of the second highest which is HDFC Bank. Also,similar to the beneficiary volume plot, we observe that lower ranked banks have similar and very low total remittance volumes.

Interestingly, SBI, a nationalised bank, has the highest remittance volume, while two private banks dominate the beneficiary volume charts. It is suggested that banks of these two categories have different beneficiary and remittance behaviour, which we hope to explore.

2.3 k-means Clustering

In order to get a better understanding of the remittance and beneficiary behaviour of banks, we apply K-means clustering to analyze remittance (total sent) and beneficiary (total received) volume and group banks. We only consider the banks common in both remittance and beneficiary top rankings for this purpose.

To identify the optimal number of clusters, we use **wss**,**silhouette** and **gap stat** methods, and obtain the optimal number to be 5.

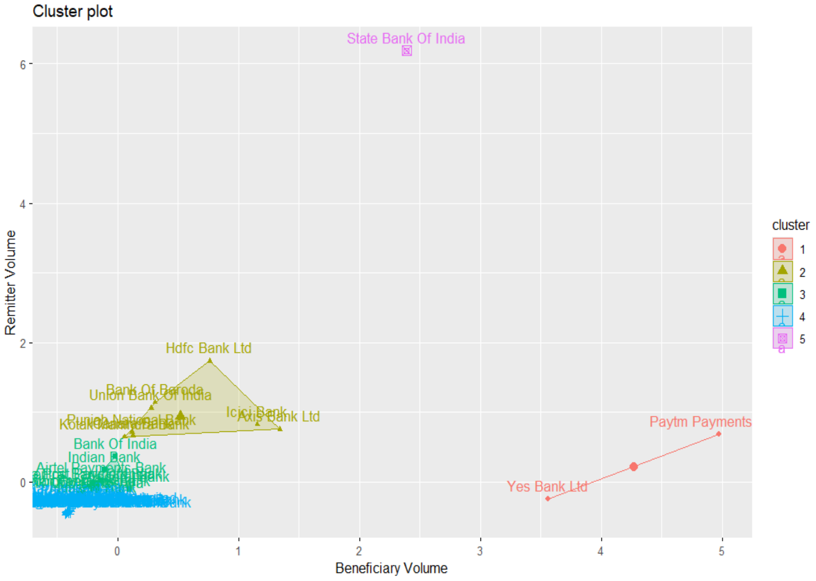


Figure 6: k-means clustering of banks with 5 clusters

From the cluster plot, we make the following observations-

- 47 of the 50 banks are in three large clusters, having lower remittance and beneficiary volumes compared to the other 3 banks.
- **Paytm Payments Bank** and **Yes Bank Ltd** show similar behaviour, having higher beneficiary volume compared to remittance volume. We also note that both of them are private banks,placed 3^{rd} and 2^{nd} in the combined rankings.
- **State Bank of India** has very high remittance volume, and moderately high beneficiary volume, hence forming a cluster of it's own, as suggested by it's placement at the top of the combined rankings.

3 Analysis

3.1 Linear Regression

Observing the dataset, we aim to predict the total beneficiary volume of a bank at a particular month given the other variables. We apply linear regression on the dataset, and the findings are summarized in the following table.

Table 2: Coefficients of the Linear Regression Model(Multiple $R^2 = 0.61$)

Variable	**Estimate**	**Std. Error**	**Test Statistic**	**p-value**
Intercept	34.51278	8.93857	3.861	0.000118
Total Remittance Volume	0.16144	0.06827	2.365	0.018185
Beneficiary Approved Volume	0.89832	0.13336	6.736	0
Beneficiary BD	-5.92846	1.73493	-3.417	0.000651
DRA	-35.05593	8.14788	-4.302	0
DRS Remittance	0.15771	0.31965	0.493	0.621813

We observe that though the multiple R^2 is moderate, almost all of the predictors are significant at 0.05 level of significance. Further, total remittance volume, approved beneficiary volume and DRS all have positive signs, indicating that as these quantities increase, total beneficiary volume increases, which is intuitive for popular banks. Also, BD and DRA, have negative signs, indicating that users may not prefer banks with these poor qualities.

We have also applied quadratic regression with same predictors to this data, but the multiple R^2 just increased to 0.63,which is not a considerable increase. So, we do not discuss that model.

3.2 Rank Analysis

In this analysis, the ranks of both the remittance banks and beneficiary banks have been determined based on their transaction volumes. Next,we have computed the mean rank of all banks with respect to both remittance and transaction behaviour. Subsequently, the mean rank has been calculated for each bank, serving as a composite measure to assess the overall rank of the banks. By computing the mean rank, the analysis aims to offer a consolidated representation of the banks' performance, considering both their roles as remittance and beneficiary entities in transactions. We use this mean rank to get a new ranking of the banks.

After finding the mean rank, we identify the top 10 banks and the total volume of average transactions attributed to each of these leading banks were analyzed.

Table 3: Beneficiary and Remittance Ranks of Banks

Banks	Mean beneficiary Rank	Mean remittance Rank	Combined Rank
State Bank Of India	2.70	1.00	1.85
Hdfc Bank Ltd	6.00	2.00	4.00
Paytm Payments Bank	1.00	7.70	4.35
Bank Of Baroda	7.00	3.03	5.02
Icici Bank	4.73	5.50	5.12
Axis Bank Ltd	4.27	6.83	5.55
Union Bank Of India	8.00	4.30	6.15
Punjab National Bank	10.07	7.23	8.65
Canara Bank	9.67	8.63	9.15
Kotak Mahindra Bank	11.87	8.80	10.33

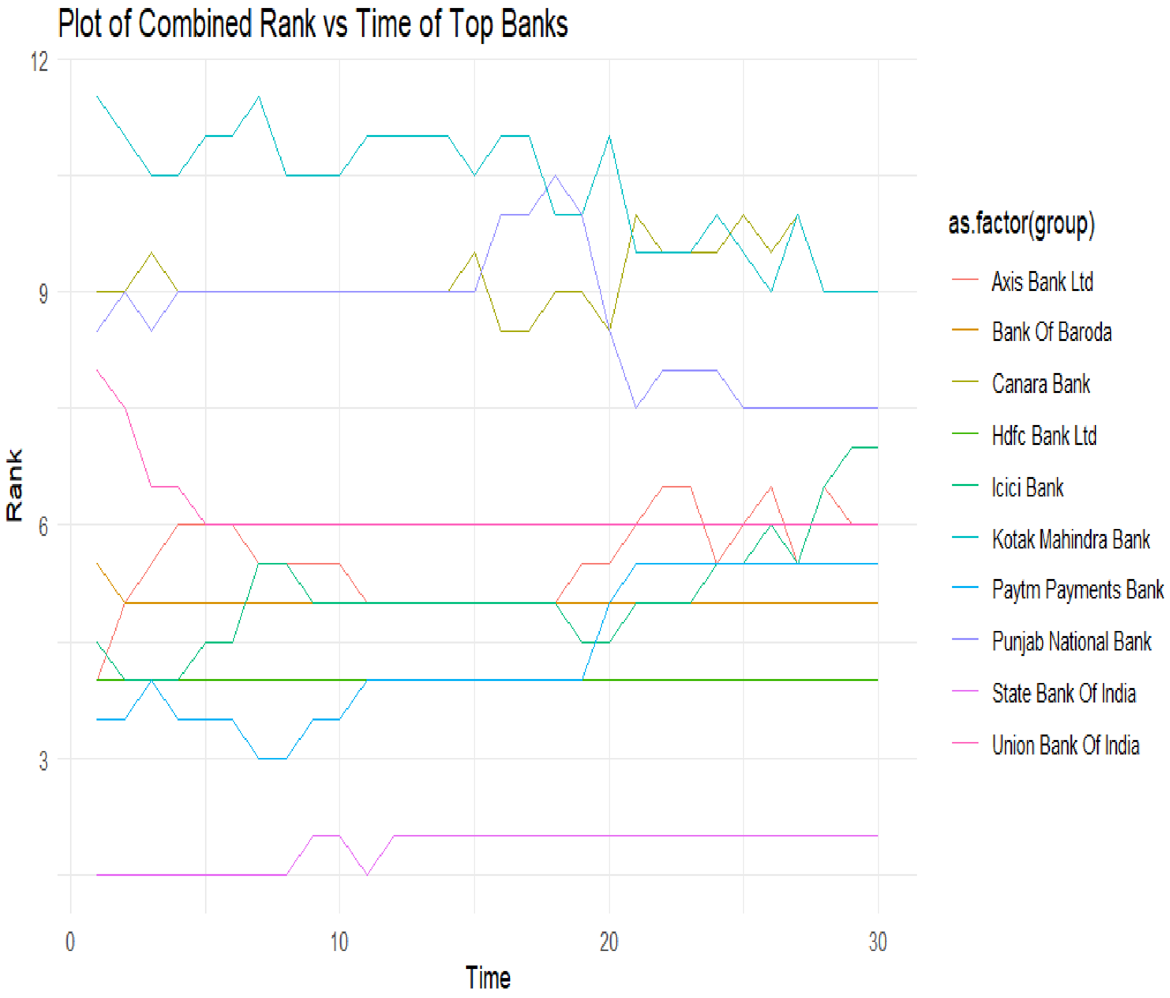


Figure 7: Combined Rank v/s Time Plot of top 10 banks

From the above plot, we can make the following observations-

- In all 30 months, State Bank of India has been very popular, with an average rank lower than all other banks. This supports the fact that it has the highest mean remittance volume and very high beneficiary volume as well.
- HDFC bank shows no rank fluctuations, having equal average rank in all 30 months.
- ICICI bank and Union bank of India are moderately popular Throughout.
- The lower ranked banks have lower average ranking throughout, indicating no major changes in their transcation amounts with time.

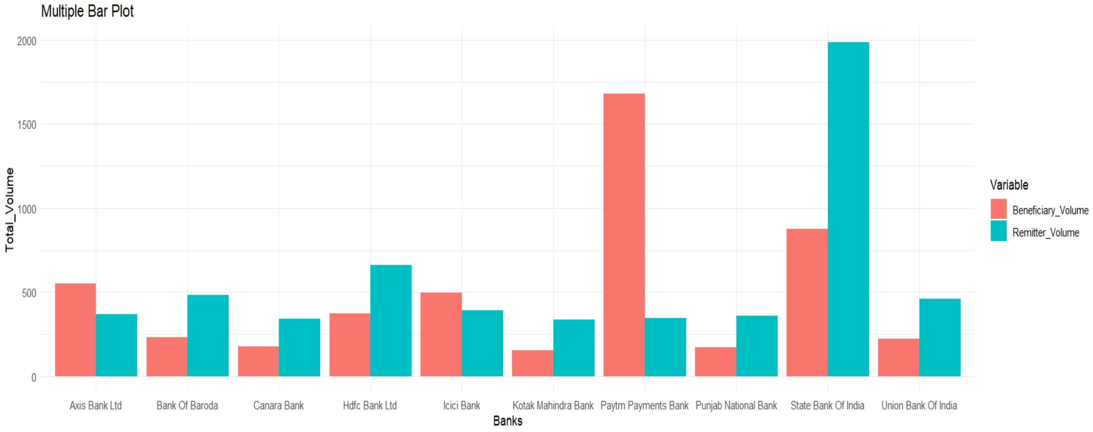


Figure 8: Beneficiary Volume v/s Remittance Volume Plot of top 10 banks

From this plot, we observe that SBI has a very high remittance volume compared to all other banks. Paytm Payments Bank has the highest beneficiary volume but low remittance volume. All other top banks have similar remittance and beneficiary volumes, but in some of them, remittance is greater, and less in some. We aim to analyze this further in this study, trying to find the factor influencing this difference in behaviour.

3.3 Time Series Analysis

In the dataset, we have time series data of beneficiary and remittance volume of most popular banks. However, analyzing all 50 banks individually for trends within 30 months of data can be overwhelming and potentially misleading due to external factors affecting each bank differently.
Hence, to gain insights regarding the nature of the total volume at varying time, we focus on the top bank according to the combined rankings, which is **State bank of India**.
Time series analysis on this bank's data can reveal its underlying trends, seasonal patterns, and irregular variations. However, with only 30 months of data, cyclical variations cannot be captured in the data. Also,identifying infrequent or irregular variations might be less reliable. So, we focus on analyzing the trend and seasonality of the beneficiary and remittance volume only.

Here, we have 30(monthwise) observations (Aug 2021 - Jan 2024) for both beneficiary and remittance volume of **State Bank of India**.
Considering the beneficiary volume, classical decomposition separates the series into:

- Trend: Long-term increase/decrease Our data shows a strong linear increasing trend,showing that UPI transactions depositing money at SBI has consistently increased with time.

- Seasonality: Though the data shows a dominating increasing trend, after removal of trend, the data reveals repeating patterns across years. It suggests that strong seasonality exists, likely due to the financial year impacting UPI transactions. Irregular: After removal of trend and seasonality, the data shows unpredictable fluctuations,possibly indicating the presence of other factors in play, which affect the total beneficiary volume. Limited data makes conclusions difficult.

The plots in the next pages show the original data, each decomposed component, and their contribution to the overall behavior.
We note that the model is additive (trend + seasonality + irregularities). Augmented Dickey-Fuller test indicates non-stationarity in the irregular components. Differencing can address this for future forecasting.
The remittance volume data for **State Bank of India** gives us similar observations, which is plotted in . It suggests that the several components of the remittance volume can be analyzed similarly.

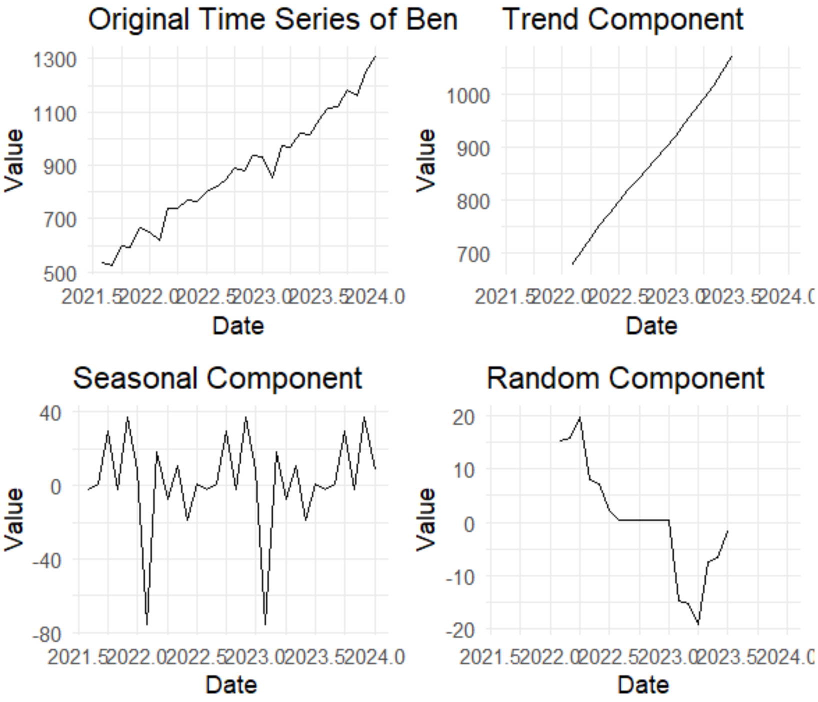


Figure 9: Decomposition of Beneficiary Volume Time Series Plot

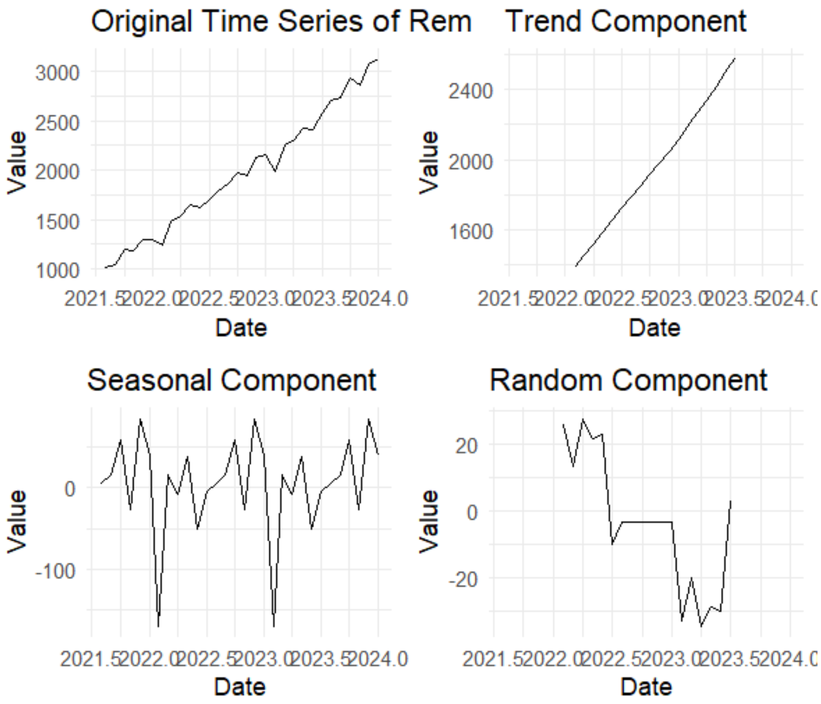


Figure 10: Histogram of Total Remittance Volume

3.4 Nationalized v/s Private Analysis

Banks in India can be broadly classified into two categories,**Nationalized** and **Private** banks.

Nationalized banks are government-owned entities, typically serving broader socio-economic objectives, often prioritizing financial inclusion and rural development. Private banks, on the other hand, are owned by private individuals or corporations, focusing on profitability and innovation, often offering specialized services and catering to specific market segments with competitive products and services.

Category	Remitter Banks	Beneficiary Banks	Total Banks
Combined	56	60	62
Nationalized	24	23	24
Private	30	30	36

Table 4: Remitter, Beneficiary, and Total Banks Data

Table 5 summarizes the distribution of nationalized and private banks among the beneficiary and remitter banks. We note that number of private banks is greater in both cases.

Now, we subset the dataset into nationalized and private banks, and obtain the following plots.

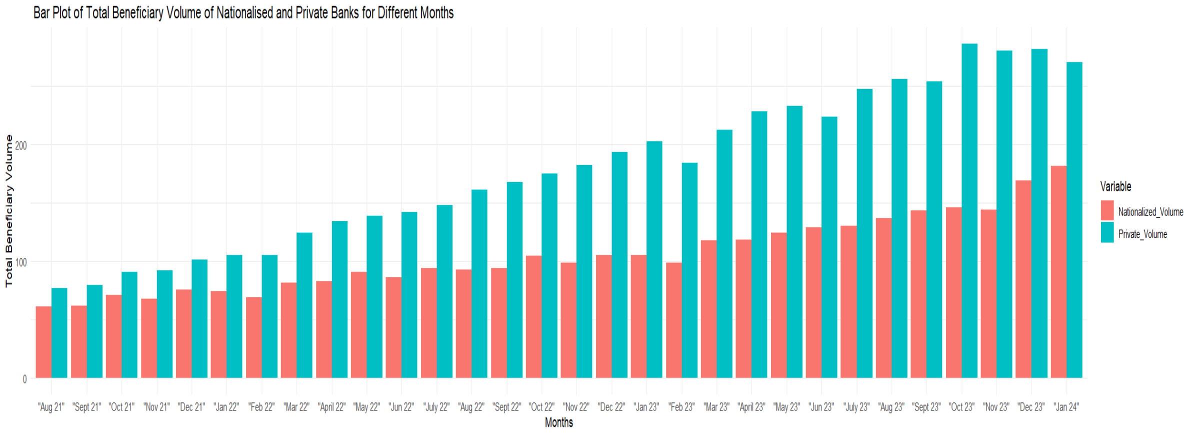


Figure 11: Bar plot of Beneficiary Volume v/s Time:X-axis is time from "Aug 21" to "Jan 24"

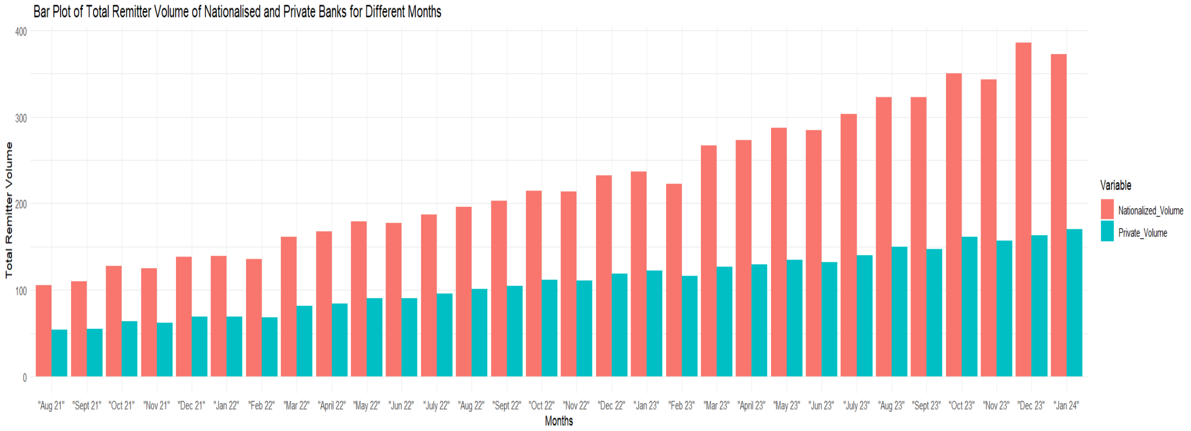


Figure 12: Bar plot of Remittance Volume v/s Time:X-axis is time from "Aug 21" to "Jan 24"

Note that the bars in blue denote the private banks, while the bars in red denote nationalised banks.

The barplot for the beneficiary volume shows that private banks are always dominating over nationalised banks and for both the total beneficiary volume is increasing. This could be possibly explained by the facts that merchants receive a huge portion of the money sent through UPI transactions, and they use private banks to a large extent.

The barplot for the remittance volume between nationalized and private banks shows an opposite picture, that the nationalised banks have higher remittance volume than private banks throughout. One possible reason could be that nationalized banks are used more by the mass, who usually send most of the money through UPI transactions. Also,similar to the previous plot,the transaction amounts for both banks are increasing with time.

Now, to analyze the difference between nationalized and private banks, we perform the following tests-

- Test for equality of mean total remittance volumes of nationalized and private banks
- Test for equality of mean total beneficiary volumes of nationalized and private banks
- Test for equality of proportion of approved remittance amount of nationalized and private banks
- Test for equality of proportion of approved beneficiary amount of nationalized and private banks
- Test for equality of debit reversal success proportion of nationalized and private banks

We assume asymptotic normality for all the tests, since sample size is large. The test summary is as follows-

Test	Nationalized Banks Estimates	Sample Private Banks Estimates	Test Statistic	p-value
Total Remittance Volume	104.41	179.73	-4.65	0
Total Beneficiary Volume	224.33	109.79	6.01	0
Proportion of Approved Remittance Amount	0.968	0.976	-5.05	0.001
Proportion of Approved Beneficiary Amount	0.891	0.902	-3.48	0.005
Debit Reversal Success Proportion	0.663	0.73	-5.35	0

Table 5: t-test Summary

From the t-test summary table, we can note that all of the tests have been rejected at 0.05 level of significance, so the remittance and beneficiary behaviour of the nationalized and private banks seem to vary considerably.

3.5 Approximate Distributions

We are interested in analyzing the top 5 banks, examining their financial performance to gain insights into industry trends. Since we have 30 data points for all 5 banks,to test the hypothesis of normality, we use Shapiro-Wilk test, the output of which is summarized in table 6.
The Shapiro-Wilk test calculates a test statistic based on the correlation between the observed data and the expected values from a normal distribution. If the p-value of the test is below a certain significance level, the null hypothesis of normality is rejected, indicating that the data is not normally distributed.

Table 6: Shapiro–Wilk Test Summary

Rank	**Bank**	**p-value for Remittance Volume**	**p-value for Beneficiary Volume**
1	State Bank of India	0.27	0.66
2	HDFC Bank Ltd	0.17	0.08
3	Paytm Payments Bank	0.07	0.23
4	Bank of Baroda	0.21	0.36
5	ICICI Bank	0.32	0.02

All the total volumes, except beneficiary volume for the fifth ranked bank, can be considered to be normally distributed at 0.05 level of significance. In order to control the variance(which is high since the data values are itself large), we scale the data by 1/1000 for the next table. We tabulate the means and variances of these normal distributions in the next table,which can be used for further analysis.

Table 7: Top 5 Banks Approximate Distribution Summary

Bank	**Remittance Mean**	**Remittance Variance**	**Beneficiary Mean**	**Beneficiary Variance**
State Bank Of India	1.98	0.4	0.87	0.05
HDFC Bank Ltd	0.66	0.05	0.37	0.01
Paytm Payments Bank	0.35	0.005	1.68	0.44
Bank Of Baroda	0.48	0.03	0.23	0.004
ICICI Bank	0.39	0.01	-	-

3.6 Tests

We conduct various t-tests and sign tests to analyze the data. These tests help us explore relationships, differences, and patterns within datasets, providing valuable insights.

We have already identified the top 10 banks on the basis of the average rank. We are interested in the difference between the top 10 banks and the others, based on the factors other than total volume,which is higher for top banks anyway, such as approved proportion,BD proportion and debit reversal success proportion. To test this, we use Welch's two-sample t-test, which has a null hypothesis of equal means of both samples(here,the top 10 banks and the remaining banks), assuming normality, which can be assumed due to the large sample size of both categories.

Table 8: Summary of Welch's Two Sample t-tests

Test for Equality	**Top 10 Banks Estimate**	**Other Banks Estimate**	**Test Statistic**	**p-value**
Beneficiary Approved Proportion	0.989	0.97	17.118	0
Beneficiary BD Proportion	0.005	0.009	-11.479	0
Remittance Approved Proportion	0.939	0.887	21.49	0
Remittance BD Proportion	0.024	0.057	-20.11	0
Debit Reversal Success Proportion	0.859	0.669	17.17	0

We note the following points from the above table-

- The top banks have significantly higher approved proportion for both beneficiary and remittance volume, indicating they are better in this aspect.
- The top banks also have significantly lower BD proportion in both cases, indicating that their accuracy is also high along with their total volume of transactions.
- The top banks also have significantly higher debit reversal success proportion, suggesting they are safer for the users as well.
- Thus, the top banks are capable of handling transactions with high accuracy even though their total volume of transactions is high, indicating they are indeed "top" banks.

We know from past data that 0.95 and 0.05 can be considered to be "high" and "low" value for Approved and BD proportion respectively, which are "good" and "poor" characteristics of a bank. So, we test whether our top 10 banks have these qualities. For this purpose, we apply the non-parametric sign test, not assuming any distribution.

The sign test works by comparing the number of observations that are greater than the hypothesized median to the number that are less than the hypothesized median. The hypothesized median in this case is 0.95 and 0.05 for approved and BD proportion respectively. If this difference is sufficiently large, the null

Table 9: Summary of Sign tests

Test	**Alternative**	**Sample Estimate**	**p-value**
Beneficiary Approved Proportion	> 0.95	0.989	0
Beneficiary BD Proportion	<0.05	0.004	0.013
Remittance Approved Proportion	>0.95	0.938	1
Remittance BD Proportion	<0.05	0.024	0.038

hypothesis is rejected, suggesting a significant difference between the median and the hypothesized value.

From the sign test summary, we note the following points-

- For the top 10 banks, beneficiary approved proportion is significantly greater than 0.95,which is considered to be a high approved proportion.
- However,the remittance approved proportion can't be concluded to be higher than 0.95.
- The BD proportion for beneficiary and remittance is significantly lower than 0.05, which indicates a very low decline rate.
- So, the top banks seem to have a very high accuracy in transactions, supporting the results of the previous t-tests.

4 Conclusion and Further Scope

In the study, we have analyzed the most popular banks based on their remittance and beneficiary behavior, examining factors such as total volume, approved proportion, BD proportion, and debit reversal success proportion of the top 50 banks for of the last 30 months. Additionally, we have investigated the difference between nationalized, and private banks in terms of their beneficiary and remittance behavior.Also, we have performed extensive analysis of the top 10 banks ranked based on the average beneficiary and remittance ranking.

In this study, we have obtained valuable insights regarding the beneficiary and remittance behavior of the most popular banks used in UPI transactions in India over the last 30 months. However, the dataset at hand is relatively small in size, partially due to the fact that UPI transactions have been popular only recently. Additionally, a large number of external factors are at play when studying UPI transactions, most of which cannot be explained using the small dataset and limited scope of study.

We aim to study this data more extensively in the future when UPI transactions would likely be more popular, as showcased by the increasing trend,which would give us a more extensive dataset. By doing so, we can possibly analyze the remittance and beneficiary behavior of top banks better and gain a deeper understanding of the financial landscape of the country.

5 Acknowledgments

We would like to express our sincere appreciation to **Prof. Subhajit Dutta**, for his valuable guidance, support, and encouragement throughout this project. We also thank the TAs for this course. Their guidance was invaluable in shaping our work.

Additionally, we would like to thank **IIT Kanpur** for providing resources and facilities that were essential for the completion of this project.

We are grateful for the opportunity to work on this project, as it has been a great learning experience for all of us.

Lastly, we extend our thanks to our classmates and friends for their understanding and encouragement during this project.

6 References

Books:

- Tukey, John. *Exploratory Data Analysis.* Addison-Wesley, 1977.
- James, Gareth, et al. *An Introduction to Statistical Learning: with Applications in R.* Springer, 2013.
- Shumway, Robert H., and Stoffer, David S. *Time Series Analysis and Its Applications: With R Examples.* Springer, 2017.

Websites:

- Introduction to k-means clustering. Towards Data Science. `https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa`
- A gentle introduction to exploratory data analysis. Towards Data Science. `https://towardsdatascience.com/a-gentle-introduction-to-exploratory-data-analysis-f11d8`
- Introduction to simple linear regression. Towards Data Science. `https://towardsdatascience.com/introduction-to-simple-linear-regression-f1df70b404b8`
- Towards Data Science - Time Series Analysis. `https://towardsdatascience.com/tagged/time-series-analysis`