MTH209 PROJECT

GROUP NAME: STATISTICOS

ANALYZING AND ENHANCING THE UPI ECOSYSTEM

SUMMARY AND OUTLINE OF THE PROJECT-

In this study, we are aiming to analyse the performance of the top 50 most popular beneficiary and remitter banks in terms of the month-wise UPI transactions conducted by these banks across the last few years.

The data has been collected from the NPCI UPI Ecosystem Statistics website. The dataset contains the information of the top 50 beneficiary and remitter 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.

IMPORTANT TERMINOLOGIES –

1. **REMITTER BANK** - The bank of the account holder who is sending the money
2. **BENEFICIARY BANK** - The bank of the account holder who is receiving money
3. **TOTAL VOLUME**- Total quantity of transactions (in millions) processed in a given timeframe.
4. **APPROVED TRANSACTION VOLUME-** 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.
5. **BUSINESS DECLINE (BD)** - 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.
6. **TECHNICAL DECLINE (TD)** - Transaction decline due to technical reasons, such as unavailability of systems and network issues on bank or NPCI side.
7. **TOTAL DEBIT REVERSAL COUNT** - 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.
8. **DEBIT REVERSAL SUCCESS AMOUNT-** 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.
9. **DEEMED APPROVED AMOUNT** - Indicates the total volume of transactions, where credit confirmations are not received online from the beneficiary banks for the credit.

In this study, we plan to apply the following methods of statistics, probability and linear algebra to analyse the data.

METHODOLOGY-

- **STATISTICS-**

    1. RANK ANALYSIS – Comparing the ranks of the banks based on the given variables to study whether the preferences of the people changed over time.

2. TIME SERIES ANALYSIS - Performing time series analysis for the most popular remitter and beneficiary banks , to see the trend in total volume of transactions over time and seasonality if any.
3. HYPOTHESIS TESTING –
    a) To statistically test whether the top banks of both remitter and beneficiary banks have significant difference in their means.
    b) To test for significance of average difference between the means of private and government banks.
4. PLOTS – Bar plots, scatterplots, boxplots and histograms of the dataset to check the variability and identify the trend and outliers in the dataset.

- **PROBABILITY-**

1. REGRESSION ANALYSIS – Aiming to fit a regression model to predict the total amount received by a particular bank using it's remittance data as predictors.
2. ASYMPTOTIC DISTRIBUTIONS - Finding the asymptotic distributions of total volume of transactions by the most popular remitter and beneficiary banks.

- **LINEAR ALGRBRA-**

1. PRINCIPAL COMPONENT ANALYSIS – Using principal component analysis to remove some of the variables in the data, while not compromising on the variability of the dataset.
2. K-MEAN CLUSTERING – Using K-Mean Clustering of volume of transactions by beneficiary and remitter banks to identify banks with similar debit and/or credit behaviour.