Unveiling UPI App Dynamics:

Market Analysis in India

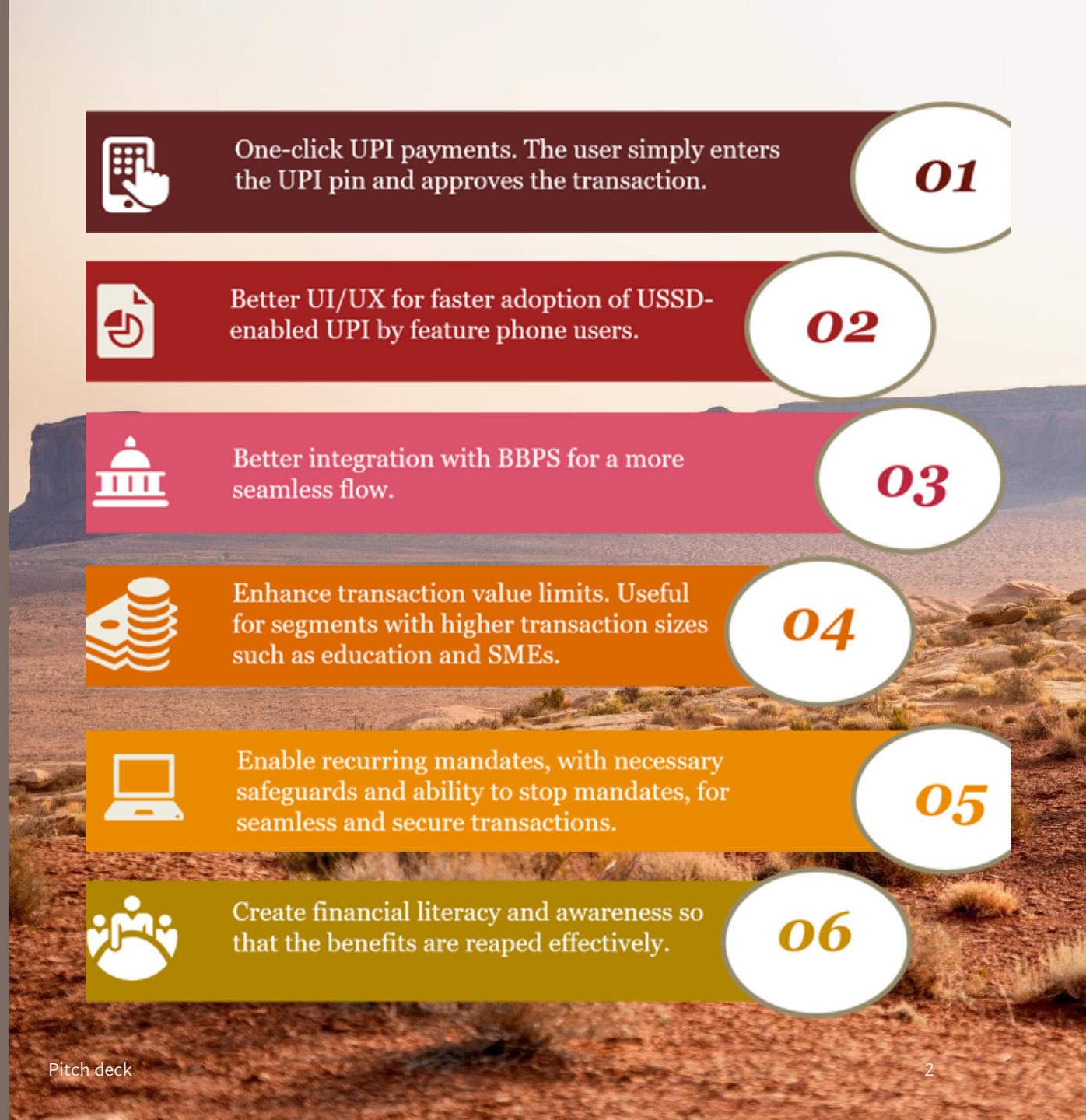
Chavi

M.Sc. Statistics

Unified Payments Interface (UPI)

Unified Payments Interface (UPI) is a system that powers multiple bank accounts into a single mobile application (of any participating bank), merging several banking features, seamless fund routing & merchant payments into one hood. It also caters to the “Peer to Peer” collect request which can be scheduled and paid as per requirement and convenience.

With the above context in mind, NPCI conducted a pilot launch with 21 member banks. The pilot launch was on 11th April 2016 by Dr. Raghuram G Rajan, Governor, RBI at Mumbai. Banks have started to upload their UPI enabled Apps on Google Play store from 25th August, 2016 onwards.



ABSTRACT

This project provides a thorough market analysis of Unified Payments Interface (UPI) apps in India, focusing on their evolution, impact, and pivotal role in the nation's digital financial landscape. The study investigates the historical development of UPI, analyzes prominent UPI apps, examines regulatory frameworks, and explores collaborative efforts shaping the ecosystem.

OBJECTIVE OF THE STUDY:

The main objective of the study is to analyze the Indian market of Unified Payments Interface (UPI) Apps.

Guide:

About Data

Source of data, Sample Data, Data description.

Literature

Literature related to UPI and other project reviews

Analysis

Methodology, statistical techniques

Summary

ABOUT DATA

Source of data, Sample Data, Data description.

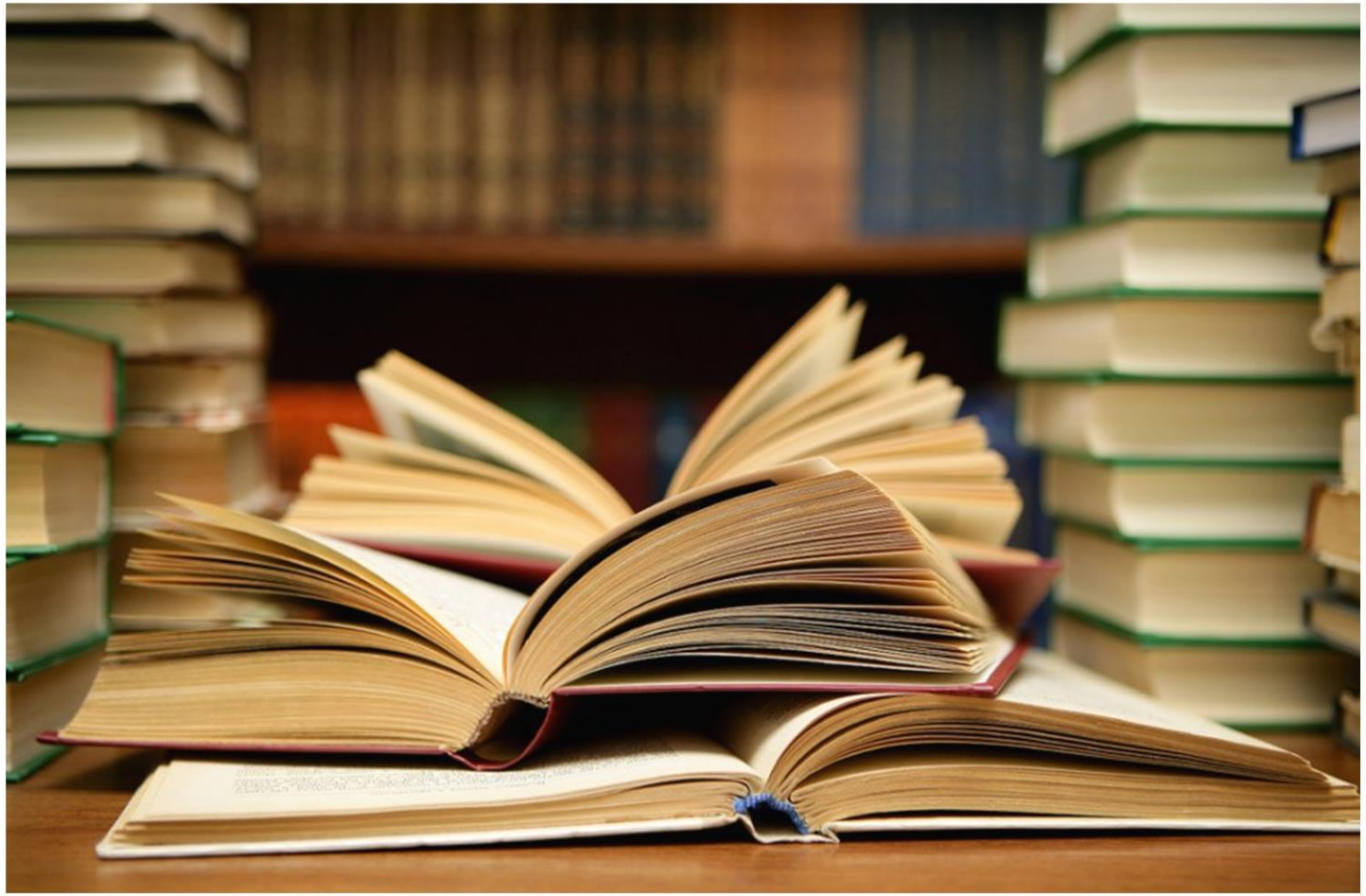
UPI Banks	Volume (Mn) by Customers	Value (Cr) by Customers	Volume (Mn)	Value (Cr)	Month	Year	Month
Airtel Payments Bank App	8.53	2047.45	13.22	4729.77	1	2022	Jan
Airtel Payments Bank App	5.8	1199.46	7.58	2210.17	2	2022	Feb
Airtel Payments Bank App	8.33	1934.41	13.16	4492.25	3	2022	March
Airtel Payments Bank App	5.29	454.64	5.3	460.9	4	2022	April
Airtel Payments Bank App	6.1	486.55	6.11	486.62	5	2022	May
Airtel Payments Bank App	6.38	506.14	6.38	506.2	6	2022	June
Airtel Payments Bank App	7.4	555.09	7.41	555.26	7	2022	July
Allahabad Bank App	0.02	7.36	0.02	7.36	1	2022	Jan
Allahabad Bank App	0.02	6.28	0.02	6.28	2	2022	Feb
Allahabad Bank App	0.02	6.15	0.02	6.15	3	2022	March
Allahabad Bank App	0.02	6.25	0.02	6.25	5	2022	May
Allahabad Bank App	0.02	6.79	0.02	6.79	6	2022	June
Allahabad Bank App	0.03	6.95	0.03	6.95	7	2022	July
Amazon Pay	73.5	6729.66	73.5	6729.66	1	2022	Jan
Amazon Pay	63.49	6044.47	63.49	6044.47	2	2022	Feb
Amazon Pay	76.34	6894.78	76.34	6894.78	3	2022	March
Amazon Pay	73.21	6699.57	73.21	6699.57	4	2022	April
Amazon Pay	74.83	6954.18	74.83	6954.18	5	2022	May
Amazon Pay	68.4	6541.84	68.4	6541.84	6	2022	June
Amazon Pay	68.77	6751.8	68.77	6751.8	7	2022	July
AU Small Finance Bank App	0.23	81.15	0.23	81.15	1	2022	Jan
AU Small Finance Bank App	0.25	89.36	0.25	89.36	2	2022	Feb
AU Small Finance Bank App	0.29	104.48	0.29	104.48	3	2022	March
AU Small Finance Bank App	0.22	109.89	0.22	109.89	4	2022	April

Sample Dataset, Source and Description:

The dataset was downloaded from Kaggle after confirming that the dataset was previously scrapped from the official site of NPCI (National Payments Corporation of India)

The dataset has 8 variables namely – UPI Banks/Apps, Volume of transaction by Customers in Millions, Value of transactions by Customers in Crores, Total Volume of transactions in Millions, Total Values of transactions in Crores , Months (in Numbers), Year and Months (in Words).

Here we have data of quantity and amount of UPI transactions of 75 banks and apps in the months of January to July for the year 2022.



Literature

Literature related to UPI and other project reviews

DATA, DATA, DATA, DATA, DATA, DATA, DATA, DATA, DATA, DATA

About Unified Payments Interface (UPI)

Introduction and Background:

UPI was introduced by the National Payments Corporation of India (NPCI) in 2016 to facilitate instant money transfers between bank accounts through mobile devices. It operates on a real-time basis and has gained widespread popularity for its simplicity and convenience.

Growth and Adoption:

UPI has witnessed rapid adoption since its inception, with a significant increase in transaction volumes and user registrations. The ease of use and interoperability across various banks and financial institutions contributed to its success.

Key Features:

UPI allows users to link multiple bank accounts to a single mobile application, making it easy to manage finances. It supports peer-to-peer transactions, bill payments, merchant payments, and other financial activities.

Key Players:

Several UPI apps have gained prominence in the Indian market, with major banks and third-party providers offering UPI-enabled applications. Prominent players include PhonePe, Google Pay, Paytm, and others.

Market Share and Competition:

The market is characterized by intense competition among UPI app providers. Different apps have vied for market share through innovative features, user incentives, and strategic partnerships.

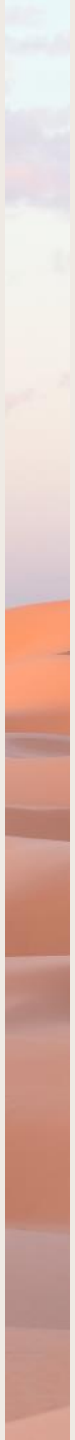
Transaction Volumes and Value:

The literature survey may present data on the growth of transaction volumes and the total value of transactions processed through UPI apps. This data is crucial for understanding the overall market dynamics.

User Demographics and Behavior:

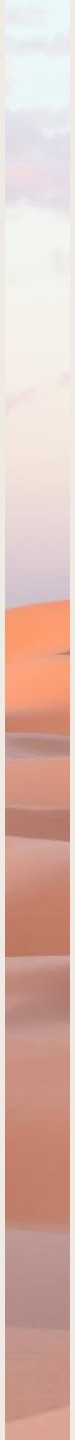
Studies might delve into user demographics, preferences, and behavior regarding the usage of UPI apps. This information is valuable for both app providers and policymakers.

Sunny Gupta and Dinesh Chand (2021)



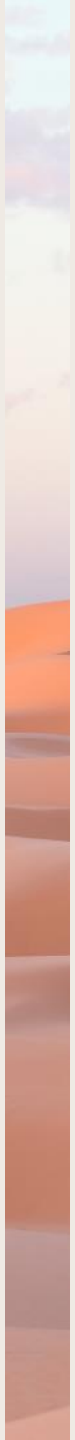
- Study on consumers perception towards UPI.
- Maximum number of people use UPI for fund transfer , mobile recharge , or cashback.
- It was concluded that mobile phones were most used device for transferring and accepting payments.
- Respondents were having positive attitude towards UPI transactions

Arvind and Deepak Chaudhari (2019)



- To study the consumer satisfaction on UPI with special reference to Hyderabad and suburbs.
- In their study they revealed that customer has a positive attitude towards UPI services and highlighted that there is relationship between education of the respondent and usage of UPI services.
- They studied the impact of the UPI services in customer satisfaction.

Poorna Pushkal C and Pappeswari C (2021)



- Study on awareness and customer satisfaction.
- To identify the level of awareness and satisfaction among people about UPI.
- It is based on select mode of digital payment.
- The work was done to identify the adoption of UPI in money transfer system.

Analysis

DATA, DATA, DATA, DATA, DATA, DATA, DATA, DATA, DATA, DATA, DATA, DATA

Methodology, statistical techniques

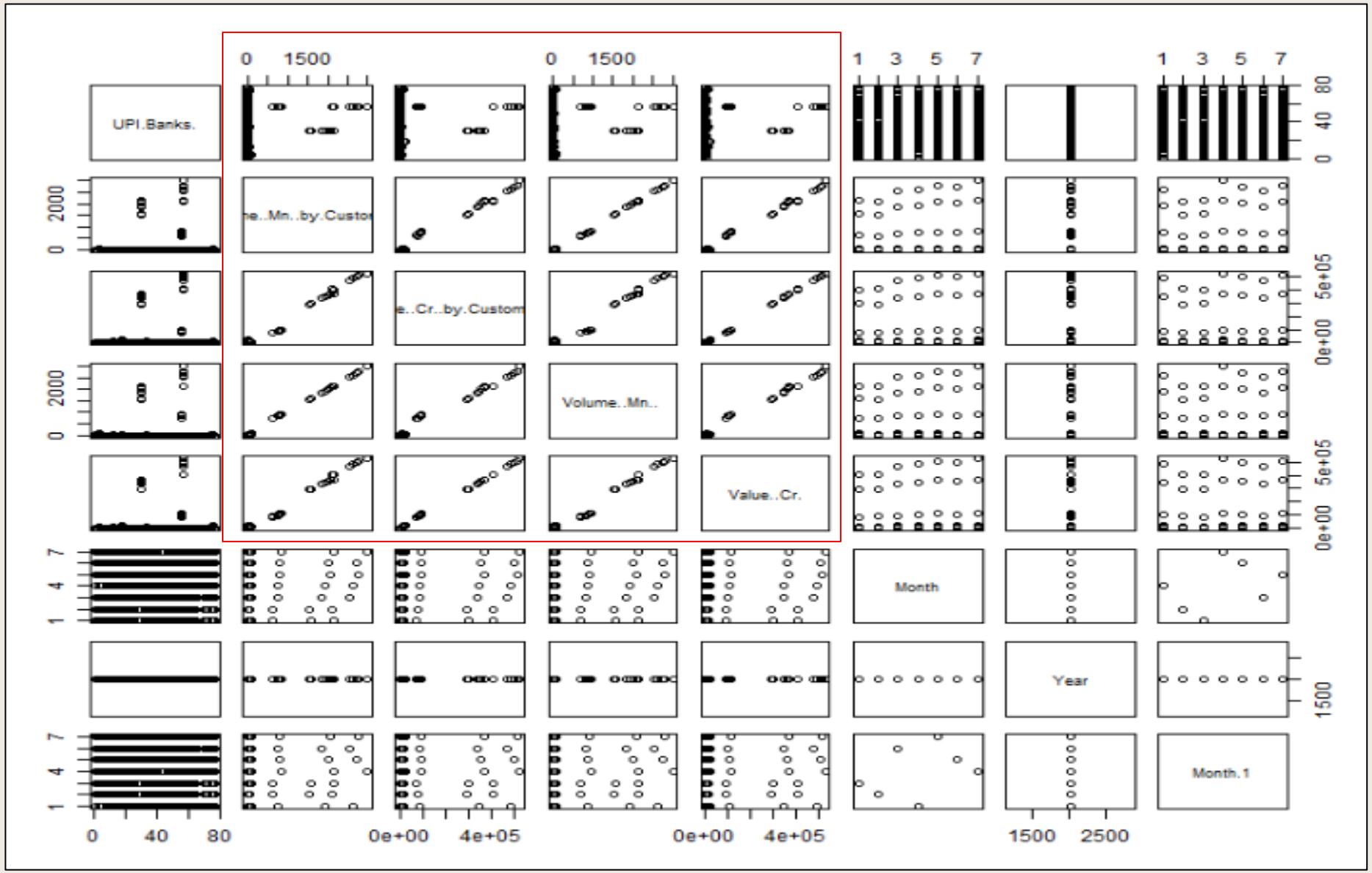
Methodology and Statistical Techniques used

The secondary dataset was collected from Kaggle after confirming that the dataset was previously scrapped from the official site of NPCI (National Payments Corporation of India). The dataset was then cleaned in Excel and loaded into Power BI for data visualization by creating dashboards. In addition, the dataset was loaded into R Studio for further analysis so as to know more about the Indian market as in the year 2022 for the months of January to July. The statistical techniques of correlation, linear regression, multiple linear regression, selection of best fit model using forward regression approach was taken into consideration to know more about the variables in the dataset.

Analysis Using R Studio

```
> #Loading data set:
>
> data <- read.csv("C:/Users/Dell/Documents/Statistics/Data Analysis/Data set/UPI/UPI cleaned dataset.csv")
> #Understanding data set:
>
> str(data)
'data.frame':   453 obs. of  8 variables:
 $ X  : chr  "UPI Banks\n" "Airtel Payments Bank App" "Airtel Payments Bank App" "Airtel Payments Bank App" .
 $ X.1: chr  "Volume (Mn) by Customers\n" "8.53" "5.8" "8.33" ...
 $ X.2: chr  "Value (Cr) by Customers\n" "2047.45" "1199.46" "1934.41" ...
 $ X.3: chr  "Volume (Mn)\n" "13.22" "7.58" "13.16" ...
 $ X.4: chr  "Value (Cr)" "4729.77" "2210.17" "4492.25" ...
 $ X.5: chr  "Month" "1" "2" "3" ...
 $ X.6: chr  "Year" "2022" "2022" "2022" ...
 $ X.7: chr  "Month" "Jan" "Feb" "March" ...
> summary(data)
      X                  X.1                X.2                X.3                X.4
 Length:453         Length:453         Length:453         Length:453         Length:453
 Class :character   Class :character   Class :character   Class :character   Class :character
 Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character
     X.5                X.6                X.7
 Length:453         Length:453         Length:453
 Class :character   Class :character   Class :character
 Mode  :character   Mode  :character   Mode  :character
>
```

```
#Data Visualization:

plot(data)
```

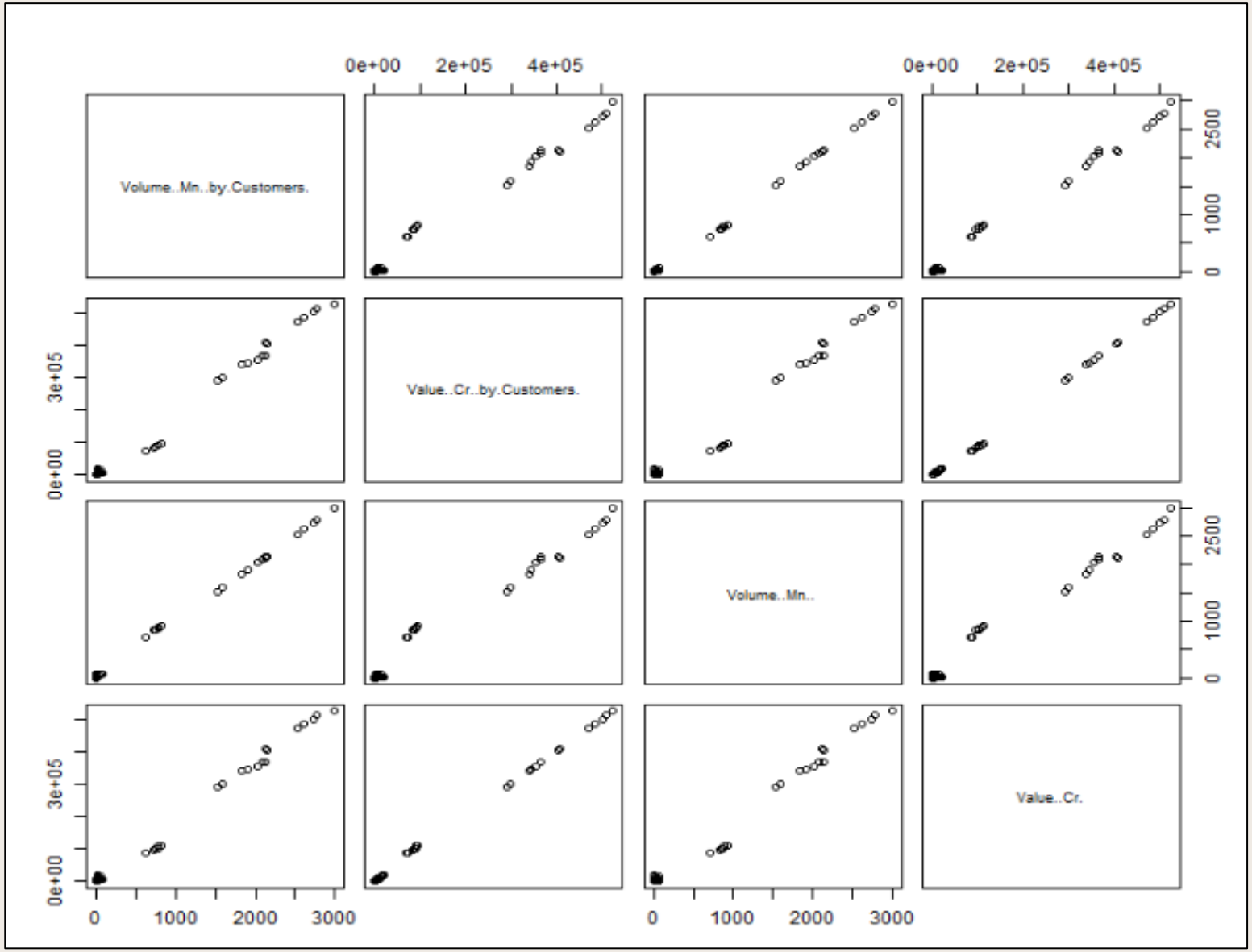


```r
#Extracting required data for analysis:

data1 <- data[,c(2,3,4,5)]
View(data1)
```

Volume..Mn..by.Customers.	Value..Cr..by.Customers.	Volume..Mn..	Value..Cr.
8.53	2047.45	13.22	4729.77
5.80	1199.46	7.58	2210.17
8.33	1934.41	13.16	4492.25
5.29	454.64	5.30	460.90
6.10	486.55	6.11	486.62
6.38	506.14	6.38	506.20
7.40	555.09	7.41	555.26
0.02	7.36	0.02	7.36
0.02	6.28	0.02	6.28
0.02	6.15	0.02	6.15
0.02	6.25	0.02	6.25
0.02	6.79	0.02	6.79
0.03	6.95	0.03	6.95
73.50	6729.66	73.50	6729.66
63.49	6044.47	63.49	6044.47

```
> #Understanding extracted data set:
>
> str(data1)
'data.frame':   452 obs. of  4 variables:
 $ Volume..Mn..by.Customers.: num  8.53 5.8 8.33 5.29 6.1 6.38 7.4 0.02 0.02 0.02 ...
 $ Value..Cr..by.Customers. : num  2047 1199 1934 455 487 ...
 $ Volume..Mn..             : num  13.22 7.58 13.16 5.3 6.11 ...
 $ Value..Cr.               : num  4730 2210 4492 461 487 ...
> summary(data1)
 Volume..Mn..by.Customers. Value..Cr..by.Customers.  Volume..Mn..        Value..Cr.
 Min.   :   0.010          Min.   :     0.3          Min.   :   0.010   Min.   :     0.3
 1st Qu.:   0.050          1st Qu.:    14.2          1st Qu.:   0.050   1st Qu.:    14.2
 Median :   0.250          Median :    68.4          Median :   0.250   Median :    68.8
 Mean   :  83.662          Mean   : 14837.8          Mean   :  86.117   Mean   : 15107.7
 3rd Qu.:   2.987          3rd Qu.:   472.5          3rd Qu.:   2.987   3rd Qu.:   472.5
 Max.   :2993.830          Max.   :524742.5          Max.   :2993.830   Max.   :524742.5
>
```

```
#Visualizing the extracted data:

plot(data1)
```



```
> attach(data1)
>
> #Multiple correlation:
> #Correlation Matrix: using Pearson, Spearman and Kendall correlation
>
> cor(data1,method = "pearson")
                          Volume..Mn..by.Customers. Value..Cr..by.Customers. Volume..Mn.. Value..Cr.
Volume..Mn..by.Customers.                 1.0000000                0.9956103    0.9994140  0.9973286
Value..Cr..by.Customers.                  0.9956103                1.0000000    0.9924936  0.9996804
Volume..Mn..                              0.9994140                0.9924936    1.0000000  0.9949731
Value..Cr.                                0.9973286                0.9996804    0.9949731  1.0000000
> cor(data1,method = "spearman")
                          Volume..Mn..by.Customers. Value..Cr..by.Customers. Volume..Mn.. Value..Cr.
Volume..Mn..by.Customers.                 1.0000000                0.9478183    0.9990929  0.9480920
Value..Cr..by.Customers.                  0.9478183                1.0000000    0.9464731  0.9998675
Volume..Mn..                              0.9990929                0.9464731    1.0000000  0.9468124
Value..Cr.                                0.9480920                0.9998675    0.9468124  1.0000000
> cor(data1,method = "kendall")
                          Volume..Mn..by.Customers. Value..Cr..by.Customers. Volume..Mn.. Value..Cr.
Volume..Mn..by.Customers.                 1.0000000                0.8105251    0.9931997  0.8115576
Value..Cr..by.Customers.                  0.8105251                1.0000000    0.8059062  0.9978611
Volume..Mn..                              0.9931997                0.8059062    1.0000000  0.8073555
Value..Cr.                                0.8115576                0.9978611    0.8073555  1.0000000
```

```
> #Fitting linear regression models:
> #Y ~ a + b*X
>
> #(1) <- Y = Total Volume and X <- Total Value
>
> model <- lm(Volume..Mn..~Value..Cr.)
> model

Call:
lm(formula = Volume..Mn.. ~ Value..Cr.)

Coefficients:
(Intercept)   Value..Cr.
   2.380212     0.005543

> #a = 2.380212, b =  0.005543
```

```
> anova(model)
Analysis of Variance Table

Response: Volume..Mn..
            Df   Sum Sq  Mean Sq F value    Pr(>F)
Value..Cr.   1 72169250 72169250   44422 < 2.2e-16 ***
Residuals  450   731080     1625
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> summary(model)

Call:
lm(formula = Volume..Mn.. ~ Value..Cr.)

Residuals:
     Min       1Q   Median       3Q      Max
-141.588   -2.679   -2.411   -2.327  315.436

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.3802122  1.9370447   1.229     0.22
Value..Cr.  0.0055426  0.0000263 210.766   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 40.31 on 450 degrees of freedom
Multiple R-squared:   0.99,     Adjusted R-squared:  0.9899
F-statistic: 4.442e+04 on 1 and 450 DF,  p-value: < 2.2e-16

> #Adjusted R-squared:  98.99%
```

```
> #Probability of F statistic is very small(less than 5%)leading to rejection of null hypothesis.
> #Rejecting H0: b = 0
```

```
> #(2)<- Y = Total Volume by Customers and X <- Total Value by Customers
>
> model1 <- lm(Volume..Mn..by.Customers.~Value..Cr..by.Customers.)
> model1

Call:
lm(formula = Volume..Mn..by.Customers. ~ Value..Cr..by.Customers.)

Coefficients:
            (Intercept)  Value..Cr..by.Customers.
               1.640995                  0.005528

> #a = 1.640995 , b =  0.005528
```

```
> anova(model1)
Analysis of Variance Table

Response: Volume..Mn..by.Customers.
                         Df   Sum Sq  Mean Sq F value    Pr(>F)
Value..Cr..by.Customers.  1 71313786 71313786   50919 < 2.2e-16 ***
Residuals                450   630243     1401
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> summary(model1)

Call:
lm(formula = Volume..Mn..by.Customers. ~ Value..Cr..by.Customers.)

Residuals:
     Min       1Q   Median       3Q      Max
-134.809   -1.942   -1.672   -1.594  304.565

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)              1.6409951  1.7974038   0.913    0.362
Value..Cr..by.Customers. 0.0055278  0.0000245 225.652   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 37.42 on 450 degrees of freedom
Multiple R-squared:  0.9912,    Adjusted R-squared:  0.9912
F-statistic: 5.092e+04 on 1 and 450 DF,  p-value: < 2.2e-16

> #Adjusted R-squared:  99.12%
```

```
> #Probability of F statistic is very small(less than 5%)leading to rejection of null hypothesis.
> #Rejecting H0: b = 0
```

```
> #(3) <- Y = Total Volume and X <- Total Volume by Customers
>
> model2 <- lm(Volume..Mn..~Volume..Mn..by.Customers.)
> model2

Call:
lm(formula = Volume..Mn.. ~ Volume..Mn..by.Customers.)

Coefficients:
             (Intercept)  Volume..Mn..by.Customers.
                   1.950                     1.006

> # a = 1.950 and b = 1.006
```

```
> anova(model2)
Analysis of Variance Table

Response: Volume..Mn..
                          Df   Sum Sq  Mean Sq F value    Pr(>F)
Volume..Mn..by.Customers.  1 72814918 72814918  383629 < 2.2e-16 ***
Residuals                450    85413      190
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> summary(model2)

Call:
lm(formula = Volume..Mn.. ~ Volume..Mn..by.Customers.)

Residuals:
    Min      1Q  Median      3Q     Max
-20.016  -1.961  -1.951  -1.950 104.430

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)                1.949756   0.662110   2.945   0.0034 **
Volume..Mn..by.Customers. 1.006034   0.001624 619.378   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.78 on 450 degrees of freedom
Multiple R-squared:  0.9988,    Adjusted R-squared:  0.9988
F-statistic: 3.836e+05 on 1 and 450 DF,  p-value: < 2.2e-16

> #Adjusted R-squared:  99.88%
```

```
> #Probability of F statistic is very small(less than 5%)leading to rejection of null hypothesis.
> #Rejecting H0: b = 0
```

```
> #(4) <- Y = Total Value and X <- Total Value by Customers
>
> model3 <- lm(Value..Cr.~Value..Cr..by.Customers.)
> model3

Call:
lm(formula = Value..Cr. ~ Value..Cr..by.Customers.)

Coefficients:
             (Intercept)  Value..Cr..by.Customers.
                 225.752                     1.003

> # a = 225.752  and b = 1.003
```

```
> anova(model3)
Analysis of Variance Table

Response: Value..Cr.
                          Df     Sum Sq    Mean Sq F value    Pr(>F)
Value..Cr..by.Customers.   1 2.3477e+12 2.3477e+12  703701 < 2.2e-16 ***
Residuals                450 1.5013e+09 3.3362e+06
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> summary(model3)

Call:
lm(formula = Value..Cr. ~ Value..Cr..by.Customers.)

Residuals:
    Min      1Q  Median      3Q     Max
-1784.5  -226.7  -225.9  -225.8 16232.6

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)               2.258e+02  8.773e+01   2.573   0.0104 *
Value..Cr..by.Customers. 1.003e+00  1.196e-03 838.869   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1827 on 450 degrees of freedom
Multiple R-squared:  0.9994,    Adjusted R-squared:  0.9994
F-statistic: 7.037e+05 on 1 and 450 DF,  p-value: < 2.2e-16

> #Adjusted R-squared:  99.94%
```

```
#Probability of F statistic is very small(less than 5%)leading to rejection of null hypothesis.
#Rejecting HO: b = 0
```

```
> #Fitting a Multiple linear regression model of best fit:
> #Using Forward Selection Method to get the model of best fit to the data.
> #It is selected based on Adjusted R square (Adj R sq) value:
>
> m0 <- lm(Volume..Mn..~1) #Null Model
> summary(m0)

Call:
lm(formula = Volume..Mn.. ~ 1)

Residuals:
    Min      1Q  Median      3Q     Max
 -86.11  -86.07  -85.87  -83.13 2907.71

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    86.12      18.91   4.554 6.79e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 402 on 451 degrees of freedom
```

```
> m1 <- update(m0,.~.+Value..Cr.)
> summary(m1)

Call:
lm(formula = Volume..Mn.. ~ Value..Cr.)

Residuals:
     Min       1Q   Median       3Q      Max
-141.588   -2.679   -2.411   -2.327  315.436

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.3802122  1.9370447   1.229     0.22
Value..Cr.  0.0055426  0.0000263 210.766   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 40.31 on 450 degrees of freedom
Multiple R-squared:   0.99,     Adjusted R-squared:  0.9899
F-statistic: 4.442e+04 on 1 and 450 DF,  p-value: < 2.2e-16

> #Adj R sq <- 98.99%
>
```

```
> m2 <- update(m1,.~.+Volume..Mn..by.Customers.)
> summary(m2)

Call:
lm(formula = Volume..Mn.. ~ Value..Cr. + Volume..Mn..by.Customers.)

Residuals:
    Min      1Q  Median      3Q     Max
-47.441  -2.013  -1.989  -1.843  54.720

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)               2.006e+00  4.679e-01   4.287 2.22e-05 ***
Value..Cr.               -1.849e-03  8.696e-05 -21.265  < 2e-16 ***
Volume..Mn..by.Customers.  1.339e+00  1.571e-02  85.234  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.735 on 449 degrees of freedom
Multiple R-squared:  0.9994,    Adjusted R-squared:  0.9994
F-statistic: 3.844e+05 on 2 and 449 DF,  p-value: < 2.2e-16

> #Adj R sq <- 99.94% (Increased)
```

```
> m3 <- update(m2,.~.+Value..Cr..by.Customers.)
> summary(m3)

Call:
lm(formula = Volume..Mn.. ~ Value..Cr. + Volume..Mn..by.Customers. +
    Value..Cr..by.Customers.)

Residuals:
    Min      1Q  Median      3Q     Max
-17.621  -0.764  -0.760  -0.747  55.796

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                0.7612154  0.3133682   2.429   0.0155 *
Value..Cr.                 0.0058861  0.0003261  18.052   <2e-16 ***
Volume..Mn..by.Customers.  1.0485779  0.0159137  65.892   <2e-16 ***
Value..Cr..by.Customers.  -0.0061529  0.0002553 -24.101   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.431 on 448 degrees of freedom
Multiple R-squared:  0.9997,    Adjusted R-squared:  0.9997
F-statistic: 5.874e+05 on 3 and 448 DF,  p-value: < 2.2e-16

> #Adj R sq <- 99.97% (Increased)
```

```
> #Adding Interaction between exploratory variables:
>
> m4 <- update(m3,.~.+Value..Cr.:Volume..Mn..by.Customers.)
> summary(m4)

Call:
lm(formula = Volume..Mn.. ~ Value..Cr. + Volume..Mn..by.Customers. +
    Value..Cr..by.Customers. + Value..Cr.:Volume..Mn..by.Customers.)

Residuals:
    Min      1Q  Median      3Q     Max
-17.142  -0.739  -0.734  -0.723  55.794

Coefficients:
                                        Estimate Std. Error t value Pr(>|t|)
(Intercept)                            7.353e-01  3.141e-01   2.341   0.0197 *
Value..Cr.                             5.722e-03  3.557e-04  16.087   <2e-16 ***
Volume..Mn..by.Customers.              1.053e+00  1.627e-02  64.681   <2e-16 ***
Value..Cr..by.Customers.              -5.983e-03  2.945e-04 -20.316   <2e-16 ***
Value..Cr.:Volume..Mn..by.Customers. -1.136e-08  9.832e-09  -1.156   0.2485
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.429 on 447 degrees of freedom
Multiple R-squared:  0.9997,    Adjusted R-squared:  0.9997
F-statistic: 4.409e+05 on 4 and 447 DF,  p-value: < 2.2e-16

> #Adj R sq <- 99.97% (No Change)
>
```

```
> m5 <- update(m3,.~.+Value..Cr.:Value..Cr..by.Customers.)
> summary(m5)

Call:
lm(formula = Volume..Mn.. ~ Value..Cr. + Volume..Mn..by.Customers. +
    Value..Cr..by.Customers. + Value..Cr.:Value..Cr..by.Customers.)

Residuals:
    Min      1Q  Median      3Q     Max
-17.318  -0.740  -0.736  -0.725  55.798

Coefficients:
                                      Estimate Std. Error t value Pr(>|t|)
(Intercept)                          7.369e-01  3.145e-01   2.343   0.0196 *
Value..Cr.                           5.772e-03  3.490e-04  16.538   <2e-16 ***
Volume..Mn..by.Customers.            1.050e+00  1.598e-02  65.714   <2e-16 ***
Value..Cr..by.Customers.            -6.024e-03  2.914e-04 -20.675   <2e-16 ***
Value..Cr.:Value..Cr..by.Customers. -4.981e-11  5.423e-11  -0.918   0.3589
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.432 on 447 degrees of freedom
Multiple R-squared:  0.9997,    Adjusted R-squared:  0.9997
F-statistic: 4.404e+05 on 4 and 447 DF,  p-value: < 2.2e-16

> #Adj R sq <- 99.97% (No Change)
```

```
> m6 <- update(m3,.~.+Volume..Mn..by.Customers. :Value..Cr..by.Customers.)
> summary(m6)

Call:
lm(formula = Volume..Mn.. ~ Value..Cr. + Volume..Mn..by.Customers. +
    Value..Cr..by.Customers. + Volume..Mn..by.Customers.:Value..Cr..by.Customers.)

Residuals:
    Min      1Q  Median      3Q     Max
-17.105  -0.739  -0.734  -0.723  55.794

Coefficients:
                                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                                      7.348e-01  3.140e-01   2.340   0.0197 *
Value..Cr.                                       5.710e-03  3.581e-04  15.948   <2e-16 ***
Volume..Mn..by.Customers.                        1.053e+00  1.626e-02  64.732   <2e-16 ***
Value..Cr..by.Customers.                        -5.972e-03  2.977e-04 -20.062   <2e-16 ***
Volume..Mn..by.Customers.:Value..Cr..by.Customers. -1.163e-08  9.822e-09  -1.184   0.2370
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.428 on 447 degrees of freedom
Multiple R-squared:  0.9997,    Adjusted R-squared:  0.9997
F-statistic: 4.409e+05 on 4 and 447 DF,  p-value: < 2.2e-16

> #Adj R sq <- 99.97% (No Change)
```

```
> m7 <- lm(Volume..Mn..~ Value..Cr.*Volume..Mn..by.Customers.*Value..Cr..by.Customers.)
> #Saturated model
> summary(m7)

Call:
lm(formula = Volume..Mn.. ~ Value..Cr. * Volume..Mn..by.Customers. *
    Value..Cr..by.Customers.)

Residuals:
    Min      1Q  Median      3Q     Max
-12.791  -0.740  -0.660  -0.657  55.632

Coefficients:
                                                      Estimate Std. Error t value Pr(>|t|)
(Intercept)                                          6.577e-01  3.161e-01   2.081  0.03804 *
Value..Cr.                                           2.225e-03  8.552e-04   2.602  0.00959 **
Volume..Mn..by.Customers.                            1.102e+00  2.979e-02  37.003  < 2e-16 ***
Value..Cr..by.Customers.                            -2.457e-03  8.717e-04  -2.819  0.00503 **
Value..Cr.:Volume..Mn..by.Customers.                 2.895e-06  1.289e-06   2.247  0.02514 *
Value..Cr.:Value..Cr..by.Customers.                  6.626e-10  4.564e-10   1.452  0.14725
Volume..Mn..by.Customers.:Value..Cr..by.Customers.  -3.310e-06  1.247e-06  -2.654  0.00824 **
Value..Cr.:Volume..Mn..by.Customers.:Value..Cr..by.Customers.  3.455e-13  1.133e-13   3.049  0.00243 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.281 on 444 degrees of freedom
Multiple R-squared:  0.9998,    Adjusted R-squared:  0.9998
F-statistic: 2.639e+05 on 7 and 444 DF,  p-value: < 2.2e-16
```

```
> #We can see that adding interaction terms to the model is not increasing the Adj R sq value at all.
> #Hence we do not add any interaction terms to the model.
> #Therefore we can see that m3 is the model of best fit with the Adj R sq values of 99.97%
> #Multiple Linear Regression Model of Best Fit is: Y ~ a + b1X1 + b2X2 + b3X3
> #Y = Total Volume, X1 = Total Value, X2 = Volume by customers, X3 = Value by customers
```

Summary

From this Project we get to know that in the year 2022 for the months of Jan to July, PhonePe, Google pay, and Paytm are the top 3 UPI Apps being used in India both in terms of Value and Volume of transactions. We also see the growth in the use of UPI apps by every month. Then we see that there is high correlation between volume and value of transactions to which we were able to fit linear and multiple linear regression models which at the end gave us the best fit model having 99.97% of Adjusted R square.

With this we can say that India is growing rapidly and we as Indian can proudly say India is a country having highest digital transactions.

DATASET

QR to Kaggle showing the dataset

Link to the NPCI site:

https://www.npci.org.in/what-we-do/upi/upi-ecosystem-statistics#innerTabTwoDec23

Sunny Gupta and Dinesh Chand (2021)

https://drive.google.com/file/d/1Zhqbr9-Xs6Gl28MdN4e4t1vnOiTQau_s/view?usp=share_link

Arvind And Deepak Chaudhari (2019)

https://acsbodwadcollege.org/wp-content/uploads/2020/01/207-A_Arvind-Dipak-2019.pdf

PornaPushpak C and Pappsewari (2021)

https://www.annalsofrscb.ro/index.php/journal/article/download/7898/5834/14071

Thank you

Chavi

M.Sc Statistics
(2nd year)

MENTOR:

Swapna Mam