The University Of Calcutta

Vivekananda College

A project on

Time Series Analysis Of India's UPI Growth

By

Debaditya Chakraborty

Roll.No: 213544-21-0103
Reg.No: 544-1111-0507-21

In partial fulfillment of the requirement
for the degree of B.Sc with Honours
in Statistics

Submitted On- 2024-06-22

Contents

Keywords

Acronym	Definition
UPI	Unified Payment Interface,A fast and secure digitial pyment system
PSP	Payment Service Provider,can be a bank or other institution
MSME	Micro and Small & Medium Enterprises
NEFT	National Electronic Funds Transfer is an electronic funds transfer system maintained by the Reserve Bank of India
RTGS	Real-time gross settlement systems are specialist funds transfer systems where the transfer of money or securities takes place from one bank to any other bank
IMPS	Immediate Payment Service (IMPS) is an instant interbank electronic fund transfer service through mobile phones.
Aeps	Aadhaar Enabled Payment System (AePS) is a banking service developed by the National Payments Corporation of India(NPCI).
CTS	Cheque Truncation System is an online image-based cheque clearing system
AR(p)	Autoregressive Process of order p
MA(q)	Moving Average Process of order q
ARMA(p,q)	Combination of AR & MA process of order p and q
ARIMA(p,d,q)	A more complex version of ARMA where d represents the order of differencing
SARIMA	ARIMA model that implements Seasonality
AIC	Akaike Information Criterion is a mathematical method for evaluating how well a model fits the data it was generated from
MAPE	Mean absolute percentage error,this measures the average magnitude of error produced by a model, or how far off predictions are on average
MASE	Mean Absolute Scaled Error, A more reliable accuracy measure to compare different models.
P2M & P2P	Person to Merchant and Peer to Peer Payments

Acknowledgements

Firstly, I am profoundly grateful to **Prof. Nilkanta Mukherjee**, Head of the Department of Statistics, Vivekananda College, for his invaluable leadership and support. His insightful feedback and encouragement have been instrumental in shaping the direction of my research for the project. I would also like to extend my heartfelt thanks to **Dr. Junaid Khan**, whose expertise and dedication provided me with the tools and confidence needed to navigate through complex statistical concepts. A special note of appreciation goes to **Prof. Sutapa Biswas**, her detailed critiques and constructive suggestions significantly enhanced the quality of my work. Her guidance was crucial in helping me to refine my analysis and methodology. I am also deeply thankful to **Prof. Riddhi Das Majumdar** for her continuous support and motivation.I am grateful to my alma mater, *Vivekananda College* and the Department of Statistics, for providing a conducive academic environment and the necessary resources for my studies. Lastly, I would like to thank my friends, whose camaraderie and moral support have been indispensable. Their encouragement and understanding helped me stay focused and motivated during challenging times.This project report would not have been possible without the collective efforts and support of all these individuals.
Thank you all.

Abstract

This paper presents a detailed analysis of the Unified Payments Interface (UPI) in India, focusing on its rapid adoption and transformative impact on digital payments from 2016 to 2024. The study employs advanced statistical techniques, including time series analysis, stochastic modeling, and forecasting, to investigate monthly and daily UPI transaction data. Key insights reveal significant trends in transaction volume and value growth, the influence of seasonal patterns, and the stabilization of average transaction values over time. The analysis also delves into the categorization of payments and merchant types, identifying sectors with the highest UPI adoption. Additionally, the paper addresses critical issues such as security concerns and transaction errors, offering a comprehensive overview of the challenges and prospects of UPI. Through rigorous data analysis, this study underscores UPI's pivotal role in enhancing financial accessibility and its potential to drive future innovations in India's digital economy.

Methodology

1. Data Collection-

- Sources: Data for the project is secondary and has been collected via web scraping or downloadable files from RBI and NPCI which are two government bodies, these are totally publicly available data and there are no underlying ethical issues.
- Time Period: The time period for which the data was collected is from 2016 to 2024 as per the latest release.
- Data Types: Data has been collected for monthly UPI transaction values, volumes, number of banks live in UPI, daily transaction volumes and values, payment categories, and merchant categories.

2. Data Reprocessing-

- Cleaning: The available data had many inconsistencies which had to be taken care .
- Transformation: To account for difference for volume and values log transformation has been used in some cases, monthly values have also been aggregated to calculate business and transaction decline, etc.

3. Exploratory Data Analysis (EDA)

- Techniques Used: Since the data is Time Series plots have been done for different data sets that are used, although summary statistics aren't that useful for time series they still have been added.
- Tools: All of the visualizations have been done using the `ggplot` and base `graphics` packages in `R`,additional customization were done using `ggthemes`, `hrbrthemes`, `ggextras` etc.

4. Time Series Analysis -

- Monthly UPI Value and Volume Analysis:
  Trend Fitting: For the monthly data used moving averages and logistic curves to fit the data and check the general trend for the future.
- Per Transaction Value Analysis:
  Forecasting Methods: Used forecasting techniques such as Holt-Winters, ARIMA, and exponential smoothing. to make predictions for future per transaction values.
- Monthly Growth Rate Calculation:
  Growth Rate Function: Built a function in R to calculate the monthly growth rate.
- Forecasting Models: Used forecasting models such as ARIMA ,SES, Holt-winters to Make future predictions after finding a suitable framework.

5. Daily Time Series Analysis

- Seasonality Handling: Explained the challenges faced due to monthly seasonality and how Fourier terms were incorporated into the ARIMA model to address these issues.

6. Exploratory Data Analysis for Payment and Merchant Categories

- Payment Categories: Made charts to understand the distribution and trends in different payment categories over the months.
- Merchant Categories: Analyzed the presence of certain merchant categories in high and medium transaction range , identifying key sectors and their growth patterns.

7. Addressing UPI Issues

- Security Concerns: Methodologies used to analyze security issues have been- literature review and case studies.
- Decline Analysis: Used available NPCI data to study declines in business and transaction volumes via basic EDA measures.

8. Software and Tools Most of the project has been done using `R` infact the report itself has been generated using `Rmarkdown` . Excel has been used in some data sets(such as the daily time series) to incorporate with the available data ,some `python` scripts were used as well to combine some excel sheets . This was done for the ease of data handling.

A list of used `R` packages is added in the reference section.

Introduction

The introduction of UPI has revolutionized the digital space. UPI usage has exponentially increased since its inception in 2016, with its growth outpacing all other modes of digital payments. UPI is an instant, real-time payment network built, owned, and operated by the National Payments Corporation of India (NPCI). This payment system is built as an inter-operable protocol and allows third-party vendors to build apps to provide payments as a service to all customers of participating banks. Due to interchangeability, customers with an account in Bank "A" can use a payments app built by PSP "X" to send money from their account in one bank to self or other party accounts of any other bank or PSP participating in UPI via QR codes, mobile numbers, or other identifiers, with instant settlement of payments (NPCI, 2016). UPI is used by multiple stake Holders, including individuals, micro, small, and medium enterprises (MSMEs), and especially smaller merchants. It is easily accessible through mobile devices, provides convenient payment initiation methods, such as users registered mobile numbers, QR codes, etc., and ensures universal interperability between financial institutions. These design choices have helped enhance digital and financial literacy and included the portion of the population that was formerly underserved or unserved by financial institutions.

Impact of UPI in India's Economy

In about eight years, India's indigenously developed UPI, has evolved into the default option to transact—from small ticket purchases at roadside shops to settling utility bills to restaurant bills, to now IPO stock purchases and mutual fund payments.

This transformation, which has now become a global template that many other countries are emulating, is founded on multiple edifices powered by a behavioral change among hundreds of millions.While UPI has made sending and receiving money at the tap of a mobile phone app, the bigger question is how has it added to India's broader economy? Importantly, what has been the specific incremental contribution of UPI or India's rapid digitization of payments to India's gross domestic product (GDP).The answer to this is two-fold. One is the opportunity cost. Two is through enabling easier credit-driven spending.

UPI has had a profound impact on financial access in India by enhancing the ease and convenience of digital transactions, especially for those who were previously underserved by traditional banking services. Here are several ways UPI has contributed to improving financial access:

1. **Accessibility**: UPI can be accessed through smartphones, making it available to a wide range of individuals, including those in remote areas where traditional banking infrastructure is limited.

2. **Inclusion of Unbaked Population**: UPI has facilitated financial inclusion by allowing unbaked individuals to open a bank account digitally and link it to UPI, enabling them to participate in digital transactions.

3. **Simplified Transactions**: UPI simplifies the process of making payments and transferring money, even for those with limited literacy or familiarity with banking proce-

dures, thus lowering the barrier to entry for digital financial services.

4. **Cost-Effective Transactions:** UPI transactions are often low-cost or free, making it an affordable option for individuals and businesses alike, which reduces the financial burden associated with traditional banking fees.

5. **Real-Time Transactions:** UPI enables instant, real-time transactions, which enhances the efficiency of financial operations for both consumers and businesses, allowing for quick and seamless money transfers.

6. **Security and Fraud Prevention:** UPI incorporates robust security measures such as two-factor authentication and encryption, which build trust among users and encourage the adoption of digital transactions by mitigating the risk of fraud.

7. **Integration with Various Financial Services:** UPI's integration with multiple financial services, including mobile wallets, online banking, and third-party payment apps, provides users with a versatile and comprehensive digital payment ecosystem.

8. **Merchant Adoption:** The widespread adoption of UPI by merchants, ranging from small roadside vendors to large retail chains, has significantly expanded the acceptance of digital payments across various sectors, enhancing the overall digital economy.

9. **Support for Government Initiatives:** UPI supports government initiatives aimed at promoting digital payments and financial inclusion, such as the Direct Benefit Transfer (DBT) scheme, which directly deposits subsidies and benefits into recipients' bank accounts.

10. **Enhanced Transparency:** By digitizing transactions, UPI promotes transparency in financial dealings, reducing the reliance on cash and helping to curb the shadow economy.

11. **Boost to Digital Literacy:** The widespread use of UPI has encouraged more people to become digitally literate, as they learn to navigate and utilize mobile banking apps and other digital financial services.

12. **Economic Formalization:** UPI contributes to the formalization of the economy by bringing more transactions into the digital space, which aids in better tax compliance and economic monitoring by the authorities.

13. **Financial Empowerment:** By providing a user-friendly and accessible platform, UPI empowers individuals to manage their finances more effectively, track their spending, and make informed financial decisions.

14. **Innovation and Competition:** The success of UPI has spurred innovation in the fintech sector, leading to the development of new financial products and services that cater to the diverse needs of the Indian population, fostering competition and improving service quality.

15. **Reduced Reliance on Cash:** UPI has significantly reduced the reliance on cash transactions, promoting a shift towards a cashless economy, which is more efficient and less prone to issues such as theft and counterfeiting.

1 Analyzing Monthly UPI Transaction(From 2016 to 2024)

Here are the first few rows of the dataset, to get an idea of the -

Table 1: Monthly UPI Metrics

	Month	No. of Banks live on UPI	Volume(In Mn)	Value(In Cr)	Volume(In Cr)
V95	2016-04-01	21	0	0.00	0.000
V94	2016-05-01	21	0	0.00	0.000
V93	2016-06-01	21	0	0.00	0.000
V92	2016-07-01	21	0.09	0.38	0.009
V91	2016-08-01	21	0.09	3.09	0.009
V90	2016-09-01	25	0.09	32.64	0.009
V89	2016-10-01	26	0.1	48.57	0.010
V88	2016-11-01	30	0.29	100.46	0.029
V87	2016-12-01	35	1.99	707.93	0.199
V86	2017-01-01	36	4.46	1696.22	0.446
V85	2017-02-01	44	4.38	1937.71	0.438
V84	2017-03-01	44	6.37	2425.14	0.637
V83	2017-04-01	48	7.2	2271.24	0.720
V82	2017-05-01	49	9.36	2797.07	0.936
V81	2017-06-01	52	10.35	3098.36	1.035
V80	2017-07-01	53	11.63	3411.35	1.163
V79	2017-08-01	55	16.8	4156.62	1.680
V78	2017-09-01	57	30.98	5325.81	3.098
V77	2017-10-01	60	76.96	7057.78	7.696
V76	2017-11-01	61	105.02	9669.33	10.502

Only the first 10 rows are shown for convenience.The volumes have been converted to crore from million.

1.1 Exploratory Analysis

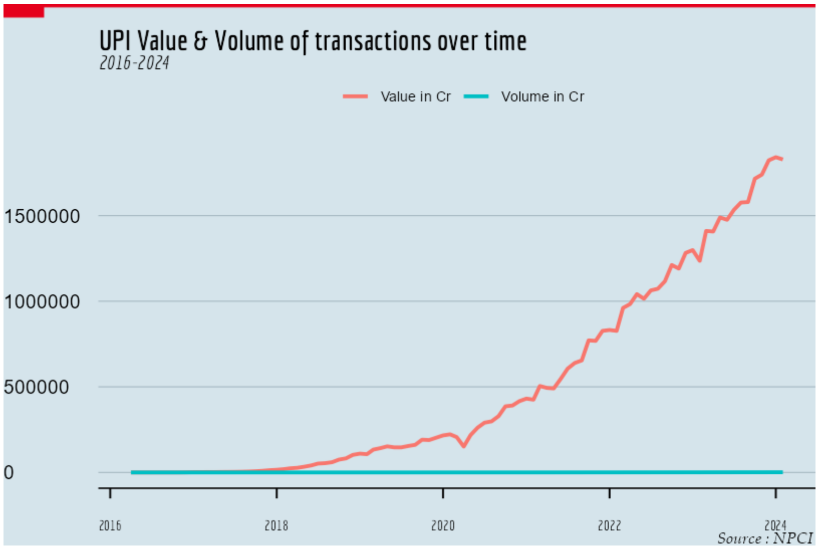


Figure 1: Plot showing the growth of Upi Transaction volume and Transaction amount overtime(2016-2024)

As it can be seen the transaction amounts are much larger than the volume so it is not possible to contain them in the same graph and compare their growth simultaneously. To solve this problem log transformation has been used for both variables to bring them in a comparable range.

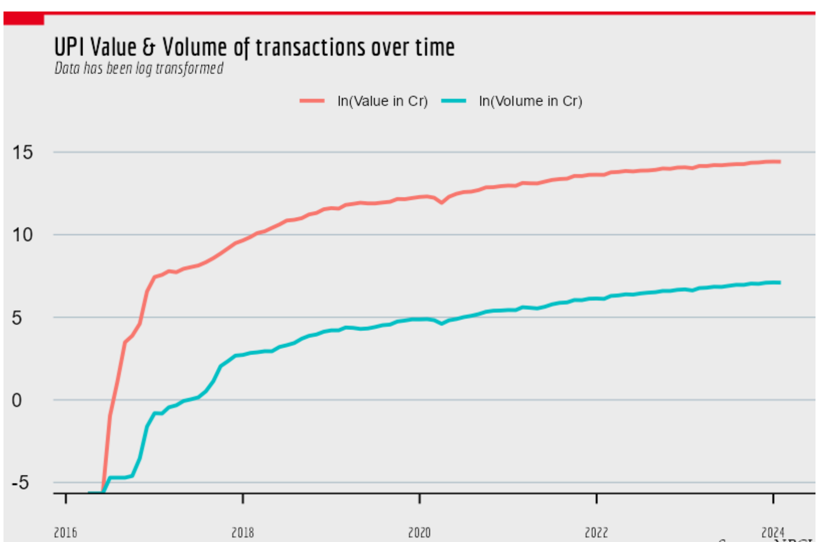


Figure 2: Plot after log transform

The simultaneous growth of both factors is evident, and as anticipated, the growth of UPI transaction volume and transaction amount is nearly identical, despite the significant difference in their values. This is due to each transaction resulting in some amount of money being transferred, ranging from very small to very high values. Observing the presence of trends and seasonal components is difficult in this combined plot, so value and volume are plotted separately for clearer analysis.

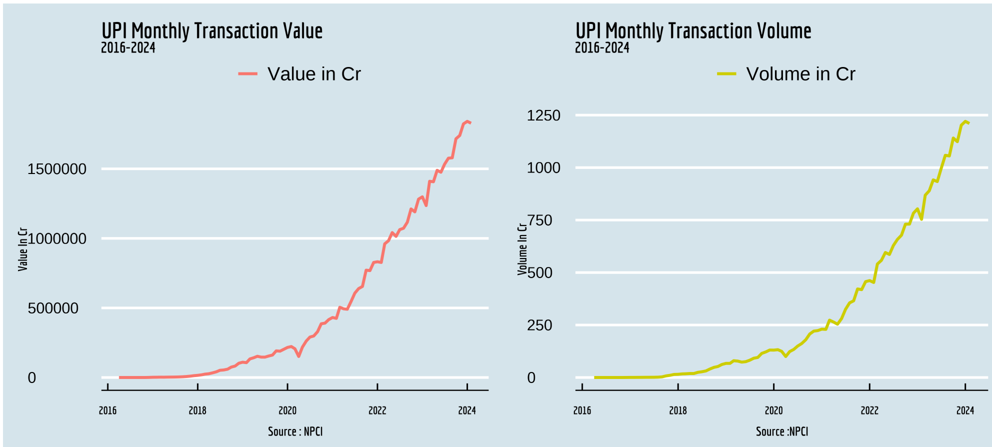


Figure 3: Value and Volume of transactions per month

A strong secular trend is visible with some random fluctuations in both plots. There isn't a very strong seasonal effect identified from the plot. Notably, the growth of UPI in terms of transaction volume and value has been exponential rather than linear. This information will assist in identifying a proper model for analysis. The impact of COVID-19 is evident, as there is a significant change in both metrics (value and volume) at the start of the pandemic. Post-pandemic, there has been a rapid increase in transaction value, possibly influenced by inflation, which may be analyzed further.

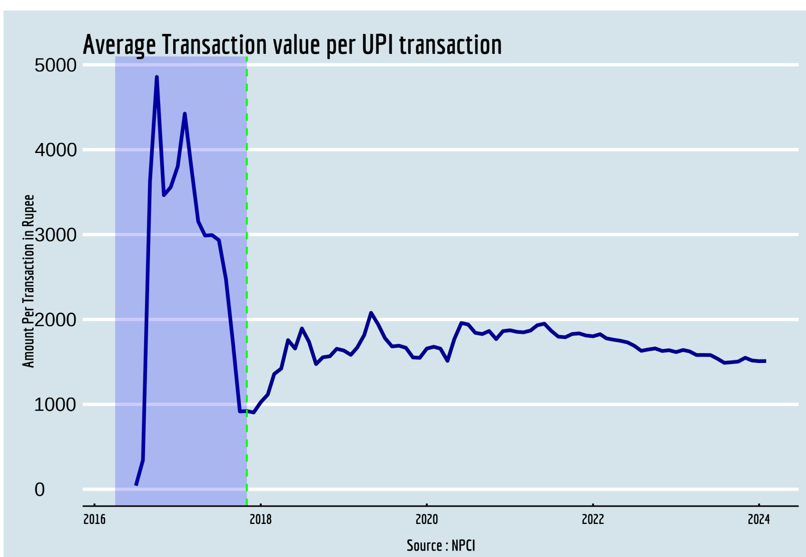


Figure 4: Plot of Average Transaction value

The plot indicates that the average transaction value between the period of initialization (2016) and up to 2018 (marked by blue area) was extremely high, as UPI was initially used by a select few individuals. Over time, with easier access to the internet, UPI became a mainstream method of payment, and the average payment value stabilized around 2019. Currently, the data shows the average transaction value per transaction is slowly decreasing, indicating that over time, people are using UPI for more smaller transactions.

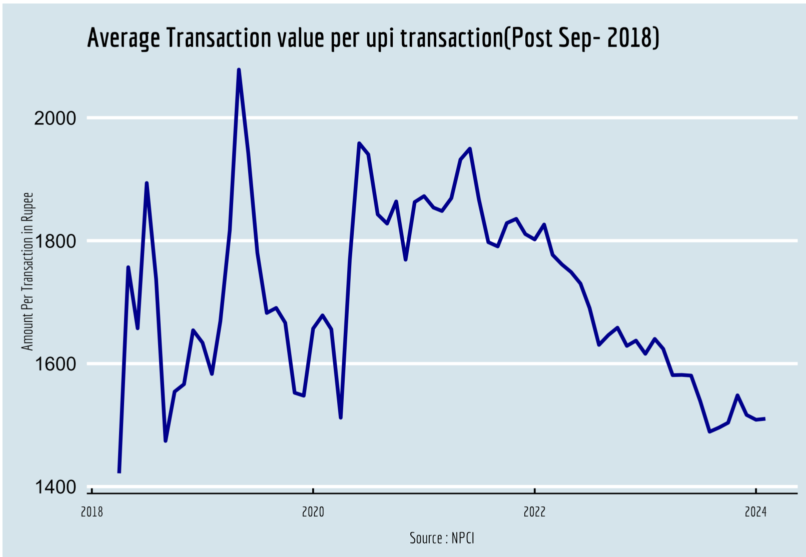


Figure 5: Average transaction value post September 2018

During this period, the average transaction value stabilizes around 1300-1500, with a slow decline also observed. This time series holds particular interest as it is unique in not showing a strong secular trend and lacking visible seasonal fluctuations, yet containing a significant amount of random fluctuation. To account for this, deterministic procedures should be avoided, and stochastic analysis methods will be utilized for this object.

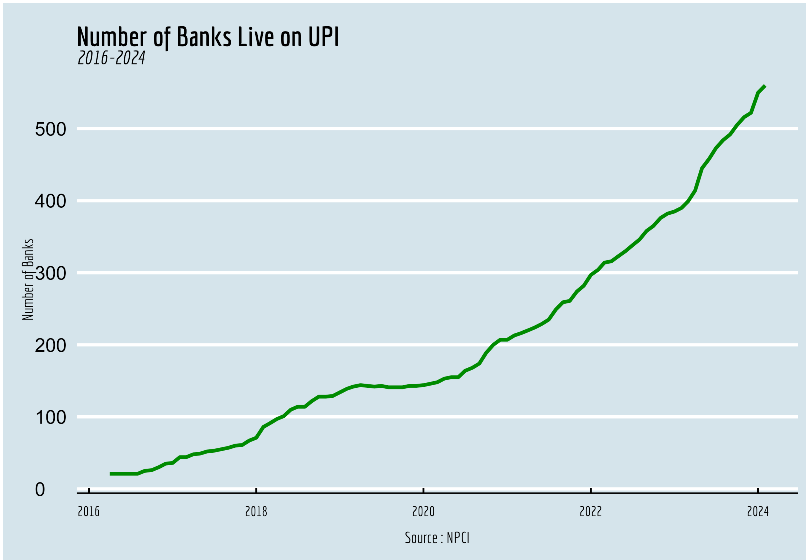


Figure 6: Number of Banks live in UPI

Number banks allowing UPI registration is growing rapidly this indicates at the growth of financial inclusion among the population of India the more banks especially regional banks allow UPI registrations the better will be the penetration of digitization of payments throughout the country.

Table 2: Summary Statistics for UPI Monthly Metrics

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
Value	0.00000	25597.2250	216242.970	500797.1573	896287.385	1841083.970	0.000
Volume	0.00000	18.3765	130.502	303.8076	501.140	1220.302	0.000
Avg. Transaction Val	42.22222	1577.0483	1704.861	1865.8310	1863.037	4857.000	3.000
Avg. Transaction Val(Post 2018 Sept)	1421.60406	1582.4160	1682.418	1707.4222	1827.031	2078.268	1421.604

1.2 Time Series Analysis of Monthly UPI Metrics.

Initially some classical methods such as smoothing procedures like Moving Average or filters are performed to dampen the fluctuations and then proceed to decompose the time series into several components.After this stochastic models maybe used, such as AR(Auto Regressive), MA(Moving Average) and if needed $ARMA$(Auto Regressive Moving Average Process) AND $ARIMA$(Auto Regressive Integrated Moving Average Process) to model the data, given the conditions to assume these models HOLD such as *stationarity* etc.

1.2.1 Analysing Time series with Trend and no Seasonal Variation(Monthly UPI Value & Volume of transaction)

From the exploratory analysis it was found that that monthly value and volume of UPI transactions contained a significant amount of secular positive trend with some underlying random component,there is no visible seasonal fluctuation in these data.

Trend- From (Kendall and Stuart 1966) "The concept of trend is more difficult to define. Generally, one thinks of it as a smooth broad motion of the system over a long term of years, but" long" in this connexion is a relative term, and what is long for one purpose may be short for another."
The simplest type of trend is the familiar 'linear trend + noise', for which the observation at time t is a random variable X_t, given by

$$X_t = \alpha + \beta_t + \varepsilon_t .....(1)$$

where α, β are constants and ϵ_t denotes a random error term with zero mean. The mean level at time t is given by

$$m_t = (\alpha + \beta_t).....(2)$$

this is sometimes called *'the trend term'*. Other writers prefer to describe the slope β as the trend, so that trend is the change in the mean level per unit time.The trend in Equation (1) is a deterministic function of time and is sometimes called a *global linear trend*. In practice, this generally provides an unrealistic model, and nowadays there is more emphasis on models that allow for local linear trends.This could be done deterministically, but it is more common to assume that α and β evolve stochastically giving rise to what is called a stochastic trend.So far the models considered have been linear,another possibility, depending on how the data look, is that the trend has a nonlinear form, such as quadratic growth.(Chatfield 2016)

- **Filtering**- One of the most used procedure for dealing with a trend is to use a linear filter, which converts one time series, $x_t$into another y_t, by the linear operation

$$y_t = \sum_{r=-q}^{+s} a_r x_{t+r}$$

  where a_r is a set of weights. In order to smooth out local fluctuations and estimate the local mean, one should clearly choose the weights so that $\sum a_r = 1$, and then the operation is known as **Moving Average.** (Chatfield 2016)

There are many different choices for the weights of the moving average such as Spencer's 15 Point Moving average weights , Henderson's Moving average weights etc. Here the data is relatively small, so undertaking the end effects, the simple moving average with a 6 month order to smooth the data is used. This can be easily done using the `ma()` function in `stats` package .

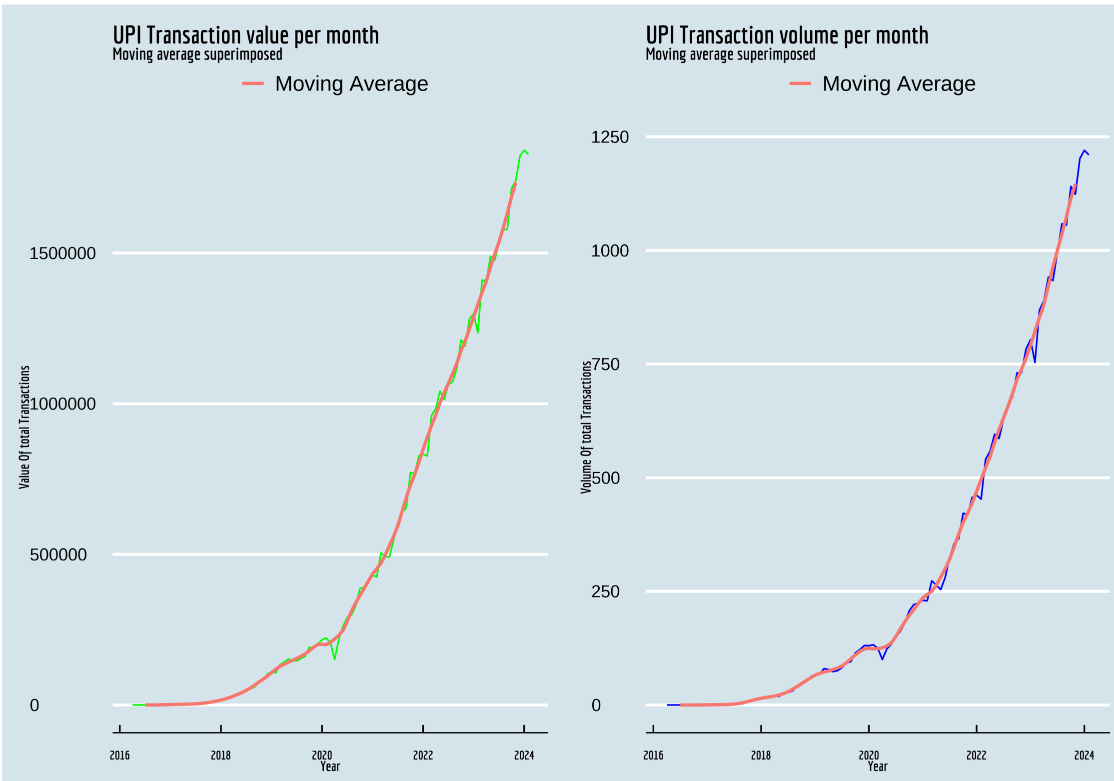


Figure 7: Moving Average Trend

Both the plots are identical and it can be observed that the moving average has successfully removed most of the random fluctuations within the series .Considering the deterministic model to be $Y_t = T_t + e_t$ where Y_t,T_t and e_t represents the original series ,Trend component and the random component respectively then the moving Average values can be considered to be a good representative of the trend component.

Thus the calculated trend values are-

Month	Smoothed Value	Smoothed Volume
2016-04-01	NA	NA
2016-05-01	NA	NA
2016-06-01	NA	NA
2016-07-01	1.006583e+01	5.333333e-03
2016-08-01	2.248500e+01	8.583333e-03
2016-09-01	8.985083e+01	2.758333e-02

2016-10-01	2.901650e+02	8.058333e-02
2016-11-01	5.927033e+02	1.527500e-01
2016-12-01	9.532967e+02	2.408333e-01
2017-01-01	1.337894e+03	3.523333e-01
2017-02-01	1.747834e+03	4.870833e-01
2017-03-01	2.171754e+03	6.323333e-01
2017-04-01	2.513884e+03	7.617500e-01
2017-05-01	2.841721e+03	9.250000e-01
2017-06-01	3.268352e+03	1.233583e+00
2017-07-01	3.908953e+03	2.020000e+00
2017-08-01	4.880520e+03	3.398500e+00
2017-09-01	6.292865e+03	5.323083e+00
2017-10-01	8.145842e+03	7.618833e+00
2017-11-01	1.040663e+04	1.007550e+01
2017-12-01	1.322466e+04	1.258942e+01
2018-01-01	1.645890e+04	1.475767e+01
2018-02-01	2.009083e+04	1.640417e+01
2018-03-01	2.436408e+04	1.794742e+01
2018-04-01	2.969173e+04	1.980283e+01
2018-05-01	3.563823e+04	2.199067e+01
2018-06-01	4.153396e+04	2.506100e+01
2018-07-01	4.850223e+04	2.939517e+01
2018-08-01	5.657724e+04	3.462633e+01
2018-09-01	6.580261e+04	4.053683e+01
2018-10-01	7.579012e+04	4.697683e+01
2018-11-01	8.500796e+04	5.331992e+01
2018-12-01	9.552048e+04	5.961858e+01
2019-01-01	1.072439e+05	6.539442e+01
2019-02-01	1.186834e+05	6.962800e+01
2019-03-01	1.281991e+05	7.248608e+01
2019-04-01	1.349012e+05	7.485200e+01
2019-05-01	1.419197e+05	7.813283e+01
2019-06-01	1.482334e+05	8.146317e+01
2019-07-01	1.546768e+05	8.581358e+01
2019-08-01	1.618523e+05	9.291192e+01
2019-09-01	1.695801e+05	1.015710e+02
2019-10-01	1.800643e+05	1.102092e+02
2019-11-01	1.915534e+05	1.176265e+02
2019-12-01	2.009715e+05	1.234528e+02
2020-01-01	2.013704e+05	1.246447e+02
2020-02-01	2.004490e+05	1.235359e+02

2020-03-01	2.078221e+05	1.239047e+02
2020-04-01	2.189562e+05	1.257453e+02
2020-05-01	2.314633e+05	1.297910e+02
2020-06-01	2.479930e+05	1.368447e+02
2020-07-01	2.777872e+05	1.503892e+02
2020-08-01	3.117517e+05	1.674541e+02
2020-09-01	3.389974e+05	1.830621e+02
2020-10-01	3.635795e+05	1.972504e+02
2020-11-01	3.858628e+05	2.095791e+02
2020-12-01	4.110806e+05	2.229592e+02
2021-01-01	4.346986e+05	2.354673e+02
2021-02-01	4.519650e+05	2.429572e+02
2021-03-01	4.712014e+05	2.504796e+02
2021-04-01	4.967260e+05	2.631332e+02
2021-05-01	5.291555e+05	2.815311e+02
2021-06-01	5.594488e+05	2.997417e+02
2021-07-01	5.950527e+05	3.205767e+02
2021-08-01	6.413509e+05	3.474476e+02
2021-09-01	6.877903e+05	3.758284e+02
2021-10-01	7.298892e+05	4.018961e+02
2021-11-01	7.643424e+05	4.214067e+02
2021-12-01	8.055054e+05	4.441007e+02
2022-01-01	8.486793e+05	4.700653e+02
2022-02-01	8.890911e+05	4.961747e+02
2022-03-01	9.274760e+05	5.217177e+02
2022-04-01	9.623538e+05	5.464486e+02
2022-05-01	1.002099e+06	5.774768e+02
2022-06-01	1.035583e+06	6.060376e+02
2022-07-01	1.067595e+06	6.318502e+02
2022-08-01	1.099041e+06	6.574887e+02
2022-09-01	1.133770e+06	6.851637e+02
2022-10-01	1.175720e+06	7.161239e+02
2022-11-01	1.208953e+06	7.386541e+02
2022-12-01	1.247041e+06	7.624843e+02
2023-01-01	1.287827e+06	7.916278e+02
2023-02-01	1.328991e+06	8.224483e+02
2023-03-01	1.369988e+06	8.525426e+02
2023-04-01	1.405682e+06	8.811533e+02
2023-05-01	1.453650e+06	9.226448e+02
2023-06-01	1.496098e+06	9.636586e+02
2023-07-01	1.535885e+06	1.000167e+03

2023-08-01	1.582498e+06	1.036257e+03
2023-09-01	1.632338e+06	1.073801e+03
2023-10-01	1.686915e+06	1.114831e+03
2023-11-01	1.733480e+06	1.146123e+03
2023-12-01	NA	NA
2024-01-01	NA	NA
2024-02-01	NA	NA

The downside of this method is that it can not be used to make future prediction and also there's an effect of missing end values due to moving average. It can be seen that the estimated trend values are very close the original values which falls with our initial assumption that this time series is made up of trend and random error only.

- **Curve Fitting**- While fitting a deterministic function of time as a curve the intital goal is to figure out what kind of a function might properly represent our time series. Everett Rogers in his book Diffusion of Innovations(2003) mentions “The logistic function can be used to illustrate the progress of the diffusion of an innovation through its life cycle” ,historically, when new products are introduced there is an intense amount of research and development which leads to dramatic improvements in quality and reductions in cost. This leads to a period of rapid industry growth. Some of the more famous examples are: railroads, incandescent light bulbs, electrification, cars and air travel. Eventually, dramatic improvement and cost reduction opportunities are exhausted, the product or process are in widespread use with few remaining potential new customers, and markets become saturated. UPI is a modern innovation which has revolutionized the way payments are done it may be a good idea to fit a logistic growth curve to the monthly value and volume data for UPI transactions.
  The Logistic Function in terms of time is given as-

  $$y_t = \frac{k}{1 + \exp(\frac{b-t}{a})}$$

  where y_t is the value of the time series at time t and a , b , k are constants.

There are many different methods to fit a logistic curve to our data most of these include long calculations for ease of calculations the `SSlogis()` function from `stats` package along with the `nls()` function in R may be used, `SSlogis()` employs a self starting logistic function using the input data(Period of time) and calculates constants k(*Asymptote*), b(*point of inflexion*) and a (*Scaling constant*) , while `nls()` uses the model given by `SSlogis` to fit the data using non linear least squares.

Plotting the Calculated model-

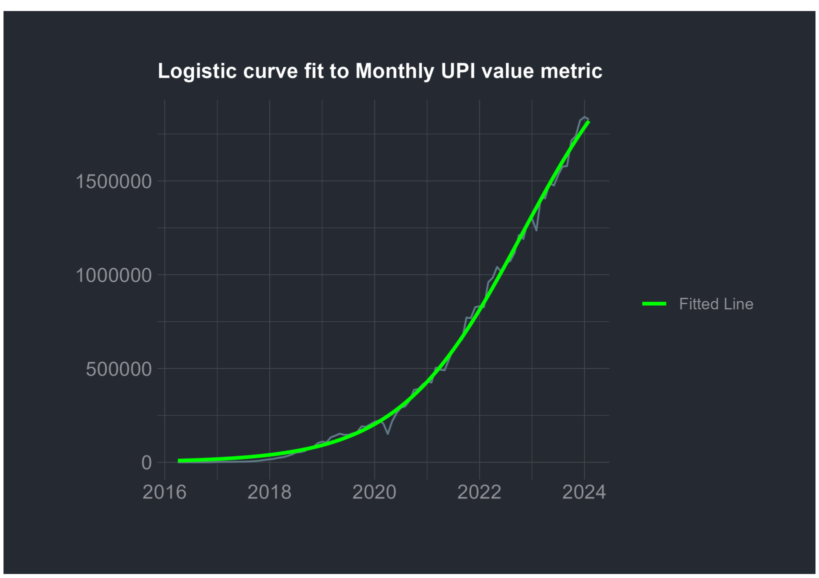


Figure 8: Fitting logistic curve to Monthly UPI value metric

From the fitted model it can be seen that the model choice was decent as the data seems to be very close to the fitted line. Here The estimated values for the constants are given -

Table 4: Estimates for Logistic Fit

term	estimate	std.error	statistic	p.value
k	2.436074e+06	6.826906e+04	35.68342	0
b	7.974670e+01	8.953216e-01	89.07046	0
a	1.408733e+01	3.238176e-01	43.50392	0

from this the calculated equation becomes -

$$y_t = \frac{2.436074e+06}{1+\exp(\frac{7.974670e+01-t}{1.408733e+01})}$$

based on the equation the fitted values are

Table 5: Logistic Curve fit for UPI Monthly Transaction Value

Time	Original Value	Fitted Value
2016-04-01	0.00	9065.907
2016-05-01	0.00	9730.185
2016-06-01	0.00	10442.926
2016-07-01	0.38	11207.636
2016-08-01	3.09	12028.065
2016-09-01	32.64	12908.232
2016-10-01	48.57	13852.438
2016-11-01	100.46	14865.287
2016-12-01	707.93	15951.705
2017-01-01	1696.22	17116.961
2017-02-01	1937.71	18366.692
2017-03-01	2425.14	19706.925
2017-04-01	2271.24	21144.100
2017-05-01	2797.07	22685.100
2017-06-01	3098.36	24337.278
2017-07-01	3411.35	26108.484
2017-08-01	4156.62	28007.097
2017-09-01	5325.81	30042.056
2017-10-01	7057.78	32222.894
2017-11-01	9669.33	34559.772
2017-12-01	13174.24	37063.512
2018-01-01	15571.20	39745.638

2018-02-01	19126.20	42618.409
2018-03-01	24172.60	45694.862
2018-04-01	27021.85	48988.846
2018-05-01	33288.51	52515.066
2018-06-01	40834.03	56289.117
2018-07-01	51843.14	60327.531
2018-08-01	54212.26	64647.806
2018-09-01	59835.36	69268.451
2018-10-01	74978.27	74209.017
2018-11-01	82232.21	79490.135
2018-12-01	102594.82	85133.538
2019-01-01	109932.43	91162.096
2019-02-01	106737.12	97599.832
2019-03-01	133460.72	104471.936
2019-04-01	142034.39	111804.777
2019-05-01	152449.29	119625.901
2019-06-01	146566.35	127964.018
2019-07-01	146386.64	136848.980
2019-08-01	154504.89	146311.748
2019-09-01	161456.56	156384.338
2019-10-01	191359.94	167099.754
2019-11-01	189229.09	178491.902
2019-12-01	202520.76	190595.477
2020-01-01	216242.97	203445.834
2020-02-01	222516.95	217078.833
2020-03-01	206462.31	231530.647
2020-04-01	151140.66	246837.551
2020-05-01	218391.60	263035.680
2020-06-01	261835.00	280160.743
2020-07-01	290537.86	298247.717
2020-08-01	298307.61	317330.502
2020-09-01	329027.66	337441.537
2020-10-01	386106.74	358611.397
2020-11-01	390999.15	380868.345
2020-12-01	416176.21	404237.864
2021-01-01	431181.89	428742.169
2021-02-01	425062.76	454399.691
2021-03-01	504886.44	481224.557
2021-04-01	493663.68	509226.067
2021-05-01	490638.65	538408.170
2021-06-01	547373.17	568768.964

2021-07-01	606281.14	600300.220
2021-08-01	639116.95	632986.946
2021-09-01	654351.81	666807.003
2021-10-01	771444.98	701730.797
2021-11-01	768436.11	737721.034
2021-12-01	826848.22	774732.583
2022-01-01	831993.11	812712.437
2022-02-01	826843.00	851599.782
2022-03-01	960581.66	891326.194
2022-04-01	983302.27	931815.955
2022-05-01	1041520.00	972986.503
2022-06-01	1014384.00	1014748.995
2022-07-01	1062991.00	1057009.002
2022-08-01	1072792.68	1099667.300
2022-09-01	1116438.10	1142620.768
2022-10-01	1211582.51	1185763.359
2022-11-01	1190593.39	1228987.138
2022-12-01	1282055.01	1272183.354
2023-01-01	1298726.62	1315243.531
2023-02-01	1235846.62	1358060.559
2023-03-01	1410443.01	1400529.748
2023-04-01	1407007.55	1442549.834
2023-05-01	1489145.44	1484023.915
2023-06-01	1475464.27	1524860.300
2023-07-01	1533645.20	1564973.249
2023-08-01	1576536.56	1604283.603
2023-09-01	1579133.18	1642719.293
2023-10-01	1715768.34	1680215.717
2023-11-01	1739740.61	1716716.000
2023-12-01	1822949.42	1752171.123
2024-01-01	1841083.97	1786539.937
2024-02-01	1827869.33	1819789.073

Based on this curve fitting the future estimates for the next 12 months will be

Table 6: Prediction for Monthly Values

Month	Predicted Value
2024-03-01	1851893
2024-04-01	1882832
2024-05-01	1912597
2024-06-01	1941181
2024-07-01	1968585
2024-08-01	1994817
2024-09-01	2019888
2024-10-01	2043815
2024-11-01	2066618
2024-12-01	2088321
2025-01-01	2108951
2025-02-01	2128537

Applying the same steps for transaction volume gives us -

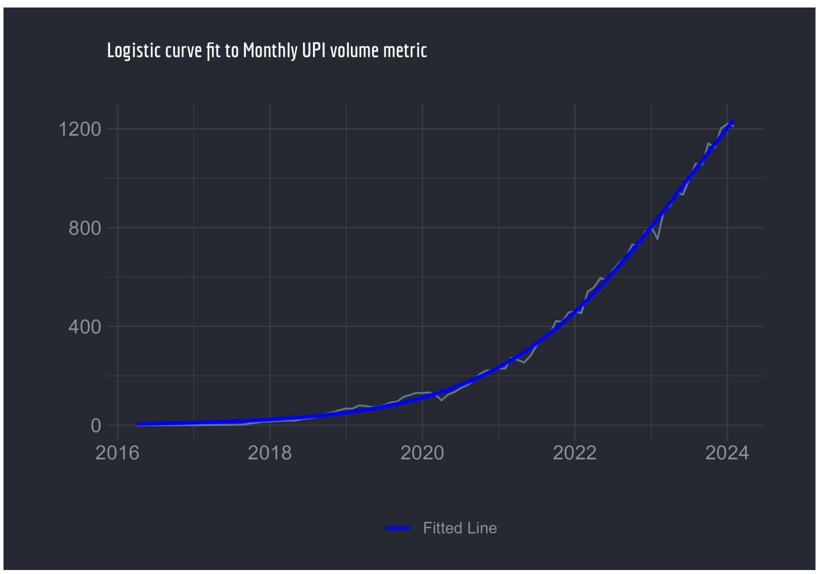


Figure 9: Logistic curve fit to Monthly UPI volume metric

Table 7: Estimates from Logistic -Curve Fit for Volume

term	estimate	std.error	statistic	p.value
k	2006.45018	81.6561480	24.57194	0
b	88.00355	1.2140952	72.48488	0
a	14.83489	0.3258464	45.52724	0

from this the calculated equation becomes -

$$y_t = \frac{2006.45018}{1 + \exp(\frac{88.00355 - t}{14.83489})}$$

based on the equation the fitted values are

Table 8: Logistic Curve fit for UPI Monthly Transaction Volume

Time	Original Volume	Fitted Volume
2016-04-01	0.000	5.677418
2016-05-01	0.000	6.072121
2016-06-01	0.000	6.494175
2016-07-01	0.009	6.945462
2016-08-01	0.009	7.427994
2016-09-01	0.009	7.943916
2016-10-01	0.010	8.495520
2016-11-01	0.029	9.085252
2016-12-01	0.199	9.715722
2017-01-01	0.446	10.389716
2017-02-01	0.438	11.110205
2017-03-01	0.637	11.880361
2017-04-01	0.720	12.703563
2017-05-01	0.936	13.583418
2017-06-01	1.035	14.523768
2017-07-01	1.163	15.528709
2017-08-01	1.680	16.602605
2017-09-01	3.098	17.750105
2017-10-01	7.696	18.976158
2017-11-01	10.502	20.286035
2017-12-01	14.564	21.685343
2018-01-01	15.183	23.180048
2018-02-01	17.140	24.776491
2018-03-01	17.805	26.481416
2018-04-01	19.008	28.301985
2018-05-01	18.948	30.245804
2018-06-01	24.637	32.320946
2018-07-01	27.375	34.535974
2018-08-01	31.202	36.899966
2018-09-01	40.587	39.422537
2018-10-01	48.236	42.113871
2018-11-01	52.494	44.984737
2018-12-01	62.017	48.046521

2019-01-01	67.275	51.311246
2019-02-01	67.419	54.791601
2019-03-01	79.954	58.500959
2019-04-01	78.179	62.453402
2019-05-01	73.354	66.663742
2019-06-01	75.454	71.147536
2019-07-01	82.229	75.921106
2019-08-01	91.835	81.001549
2019-09-01	95.502	86.406745
2019-10-01	114.836	92.155365
2019-11-01	121.877	98.266865
2019-12-01	130.840	104.761483
2020-01-01	130.502	111.660224
2020-02-01	132.569	118.984836
2020-03-01	124.684	126.757779
2020-04-01	99.957	135.002187
2020-05-01	123.450	143.741812
2020-06-01	133.693	153.000958
2020-07-01	149.736	162.804403
2020-08-01	161.883	173.177307
2020-09-01	180.014	184.145099
2020-10-01	207.162	195.733348
2020-11-01	221.023	207.967620
2020-12-01	223.416	220.873315
2021-01-01	230.273	234.475475
2021-02-01	229.290	248.798585
2021-03-01	273.168	263.866345
2021-04-01	264.106	279.701425
2021-05-01	253.957	296.325199
2021-06-01	280.751	313.757464
2021-07-01	324.782	332.016138
2021-08-01	355.555	351.116945
2021-09-01	365.430	371.073098
2021-10-01	421.865	391.894956
2021-11-01	418.648	413.589699
2021-12-01	456.630	436.160993
2022-01-01	461.715	459.608666
2022-02-01	452.749	483.928401
2022-03-01	540.565	509.111448
2022-04-01	558.305	535.144370
2022-05-01	595.520	562.008822

2022-06-01	586.275	589.681375
2022-07-01	628.840	618.133395
2022-08-01	657.963	647.330975
2022-09-01	678.080	677.234942
2022-10-01	730.542	707.800923
2022-11-01	730.945	738.979491
2022-12-01	782.949	770.716385
2023-01-01	803.689	802.952808
2023-02-01	753.476	835.625798
2023-03-01	868.530	868.668673
2023-04-01	889.814	902.011534
2023-05-01	941.519	935.581838
2023-06-01	933.506	969.305012
2023-07-01	996.461	1003.105105
2023-08-01	1058.602	1036.905470
2023-09-01	1055.569	1070.629458
2023-10-01	1140.879	1104.201112
2023-11-01	1123.529	1137.545846
2023-12-01	1202.023	1170.591099
2024-01-01	1220.302	1203.266954
2024-02-01	1210.268	1235.506702

Based on this curve fitting the future estimates for the next 12 months will be-

Table 9: Prediction for Monthly UPI Volume of Transaction

Month	Predicted Value
2024-03-01	1267.247
2024-04-01	1298.430
2024-05-01	1329.001
2024-06-01	1358.909
2024-07-01	1388.112
2024-08-01	1416.570
2024-09-01	1444.248
2024-10-01	1471.118
2024-11-01	1497.157
2024-12-01	1522.346
2025-01-01	1546.672
2025-02-01	1570.126

A long term prediction for both value and volume can be given via a plot as -

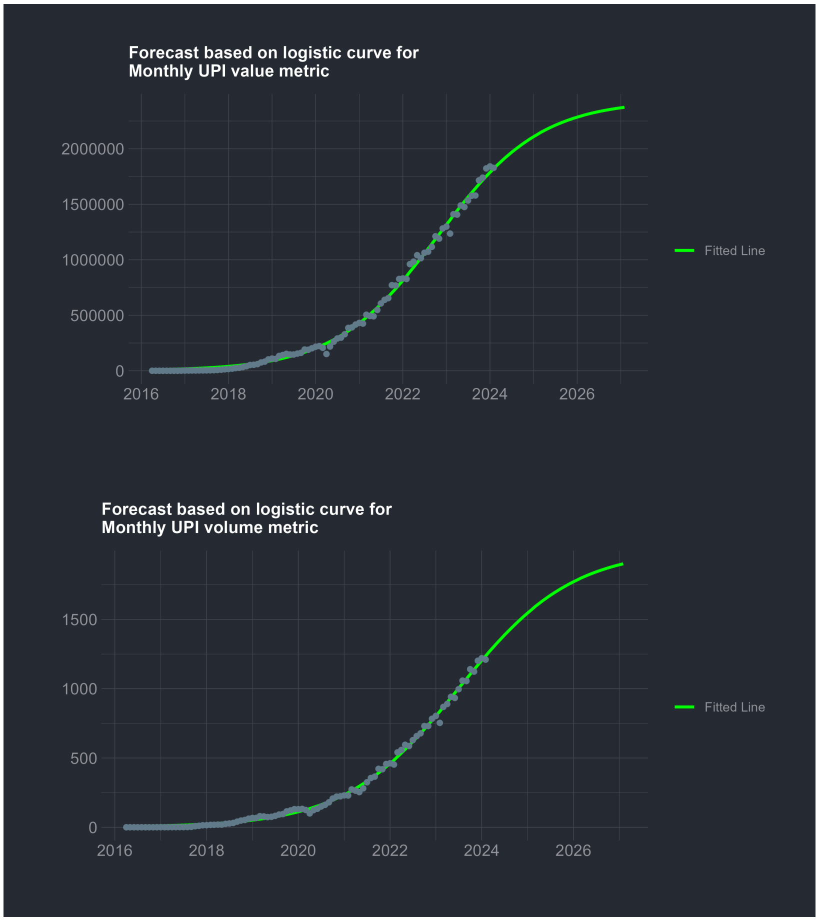
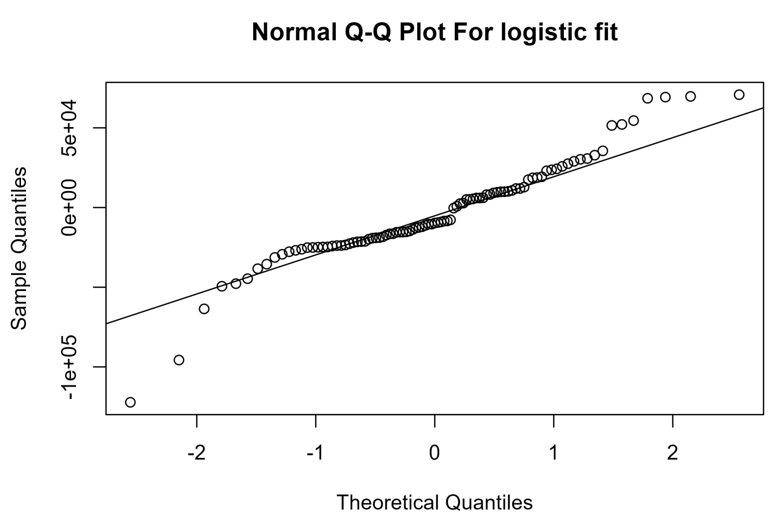


Figure 10: Long Term Forecast Via Logisitic Curve

Since the assumed model is non linear so R^2 is not suitable as a model adequacy checker,to overcome this the residuals are checked via a normal qqplot



From the previous qqplot, The residuals are kind of linear with some significant tail values drifting outside the line this usually indicates a fat tail .Due to the extremely large values some deviations may have been too big, there is also a possibility of some outliers which could've caused this.

Foundings-

- It is found that by the year 2025 Volume of monthly transactions will cross 1500 crores and value of monthly transactions will cross 2000000 crores.
- It is expected that by the year 2027 both of these metrics will start to stabilize, although this is dependent on many other factors which have not been considered in the study these include availability of smartphones and fast internet connection for the percentage of population. But never the less this is an expectable figure.

1.3 Stochastic Analysis of monthly Per Transaction Value

Previous exploratory data analysis (EDA) revealed that the monthly per transaction value is decreasing over time. It was also found that the data appears visually stationary when ignoring the initial instability. For forecasting and analyzing this time series, the next step involves plotting the time series along with its autocorrelation function (ACF) and partial autocorrelation function (PACF).

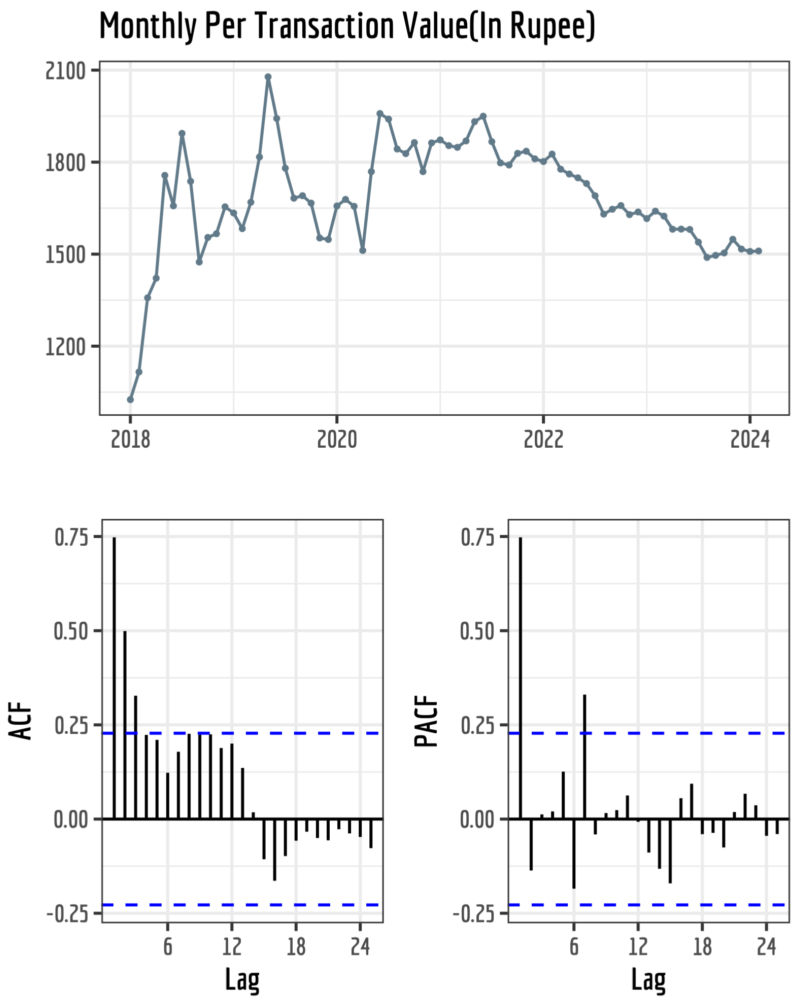


Figure 11: Displaying Per Transaction and Differenced Time Series

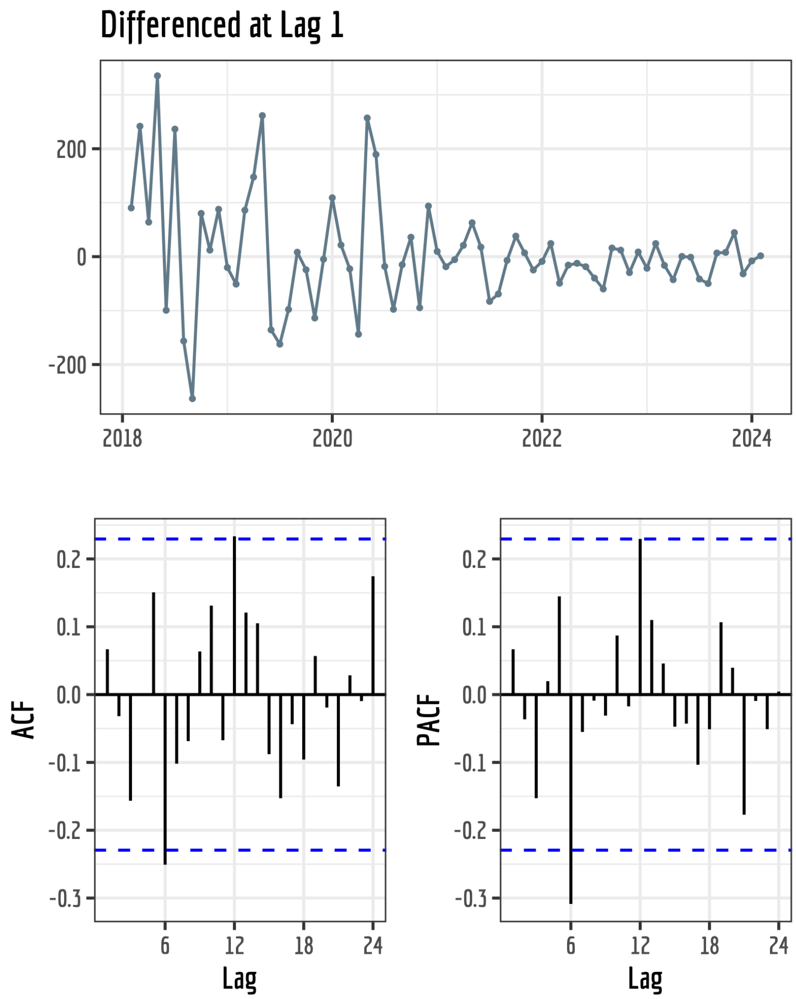


Figure 12: Displaying Per Transaction and Differenced Time Series

The ACF cuts off at lag 1, and the PACF shows a significant value at lag 1. After differencing, there are no significant autocorrelation values in the time series. Assuming no seasonal effects, the differenced time series can be defined as a random walk model.

$$y_t = y_{t-1} + \varepsilon_t$$

Which in ARIMA terms is written as ARIMA(0,1,0). Checking the `auto.arima()` output and see if the model selection aligns with the previous analysis is

```
## Series: .
## ARIMA(0,1,0)
##
## sigma^2 = 9998:  log likelihood = -439.75
## AIC=881.51   AICc=881.56   BIC=883.8
```

Table 10: Accuracy Measures for ARIMA(0,1,0) fit

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	6.564308	99.31256	65.35913	0.3437234	3.870788	0.3991829	0.0661837

The Mean Absolute Percentage Error (MAPE) value is less than 10, indicating a very good model fit. Since the data does not contain zero values, MAPE serves as a reliable model accuracy indicator. Next, the forecast is plotted using the ARIMA(0,1,0) model.

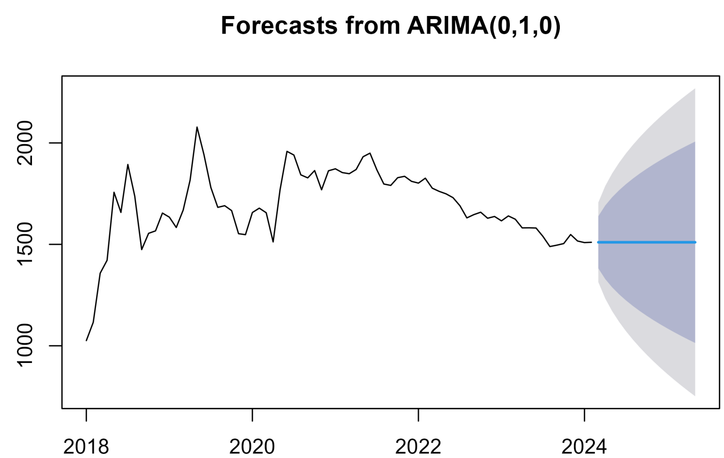


Table 11: Forecasts from ARIMA(0,1,0)

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Mar 2024	1510.301	1382.158	1638.444	1314.3236	1706.279
Apr 2024	1510.301	1329.080	1691.523	1233.1470	1787.456
May 2024	1510.301	1288.351	1732.251	1170.8579	1849.745
Jun 2024	1510.301	1254.015	1766.587	1118.3459	1902.257
Jul 2024	1510.301	1223.765	1796.838	1072.0818	1948.521
Aug 2024	1510.301	1196.416	1824.186	1030.2559	1990.347
Sep 2024	1510.301	1171.267	1849.336	991.7930	2028.810
Oct 2024	1510.301	1147.858	1872.744	955.9926	2064.610
Nov 2024	1510.301	1125.872	1894.730	922.3682	2098.234
Dec 2024	1510.301	1105.078	1915.525	890.5654	2130.037
Jan 2025	1510.301	1085.299	1935.303	860.3168	2160.286
Feb 2025	1510.301	1066.401	1954.201	831.4146	2189.188
Mar 2025	1510.301	1048.275	1972.327	803.6936	2216.909
Apr 2025	1510.301	1030.834	1989.768	777.0199	2243.583
May 2025	1510.301	1014.006	2006.597	751.2829	2269.320

It appears that the predicted forecast remains constant, specifically matching the last observation. This outcome arises because random walks permit only naive predictions, lacking discernible patterns. Additional forecasts, such as Simple Exponential Smoothing and Holt-Winters Exponential Smoothing, could be plotted for comparison.

The Holt-Winters Exponential smoothing which is also known as Triple Exponential Smoothing, As the name suggests it applies the general Exponential Smoothing Algorithm Thrice to account for reccuring patterns.It is also a part of ETS state space models.

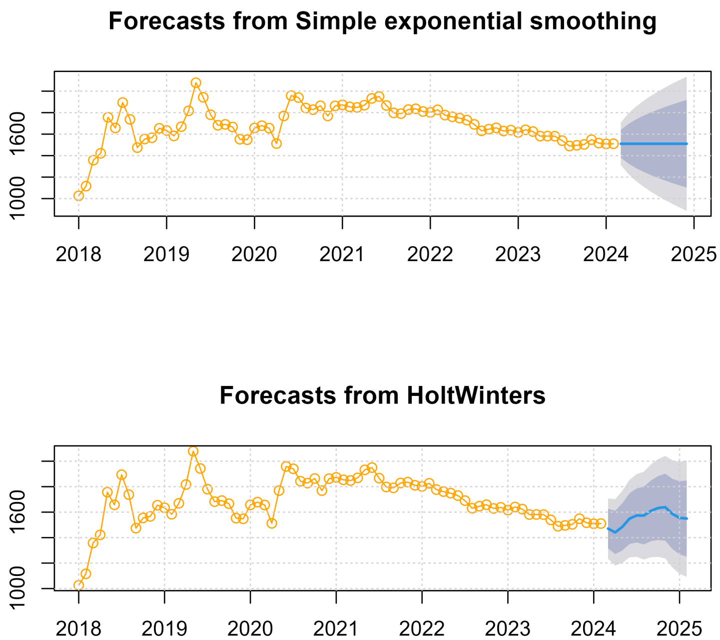


Figure 13: Forecasts Based on SES and Holtwinters

As it can be seen SES gives a naive constant forecast which is the same as the ARIMA forecast. Holt-Winters on the other hand gives a rather interesting looking prediction, the predicted values are given as -

Table 12: Forecasts from Holt-Winters Exponential Smoothing

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Mar 2024	1470.844	1315.831	1625.857	1233.773	1707.915
Apr 2024	1440.896	1270.556	1611.236	1180.383	1701.409
May 2024	1484.525	1299.628	1669.421	1201.750	1767.299
Jun 2024	1548.835	1349.976	1747.693	1244.706	1852.963
Jul 2024	1574.016	1361.663	1786.369	1249.250	1898.781
Aug 2024	1572.806	1347.336	1798.276	1227.979	1917.632
Sep 2024	1610.208	1371.928	1848.487	1245.791	1974.625
Oct 2024	1632.474	1381.639	1883.309	1248.855	2016.093
Nov 2024	1638.886	1375.707	1902.065	1236.388	2041.383
Dec 2024	1586.350	1311.004	1861.697	1165.244	2007.456
Jan 2025	1555.521	1268.155	1842.886	1116.033	1995.008
Feb 2025	1549.546	1250.286	1848.805	1091.868	2007.223

Here is the accuracy measure for this model.

Table 13: Accuracy Measures for Holt-Winters Method

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	-23.45331	122.2484	92.50533	-1.429825	5.393558	0.5649791	0.4701516

Findings

- The per transaction value shows a declining trend over time, suggesting it will continue to decrease until stabilizing at a certain point. This data is crucial for understanding UPI user behavior evolution over the years. UPI usage is increasingly prevalent in smaller transactions, indicating its integration into daily life and its role as a viable alternative to cash, thus enhancing financial inclusivity. The convenience of UPI transactions is particularly beneficial for MSMEs, presenting them with an opportunity to leverage UPI-specific offers to attract more customers.

1.4 Analyzing monthly growth rate for transaction volume.

The monthly growth rate relative to past month is calculated using this function-

```
#Calculating growth rate####
dat<-list()
#This Function Calculates the growth rate#
month_growth<-
  function(data,returndat)
  {
    returndat[1]=0;
```

```
    for(i in 2:length(data)){
      if(data[i-1]>0)
      {
        returndat[i]=(((data[i]-data[i-1])/data[i-1])*100)
      }
      else if(data[i-1]==0)
        returndat[i]=0
    }
    returndat
  }
growth<-matrix(month_growth(as.numeric(data1$`Volume(In Cr)`),dat),ncol=1)
growth<-data.frame(as.numeric(growth))
colnames(growth)<-c("GrowthRate")
```

Here is the plot of monthly growth rate-

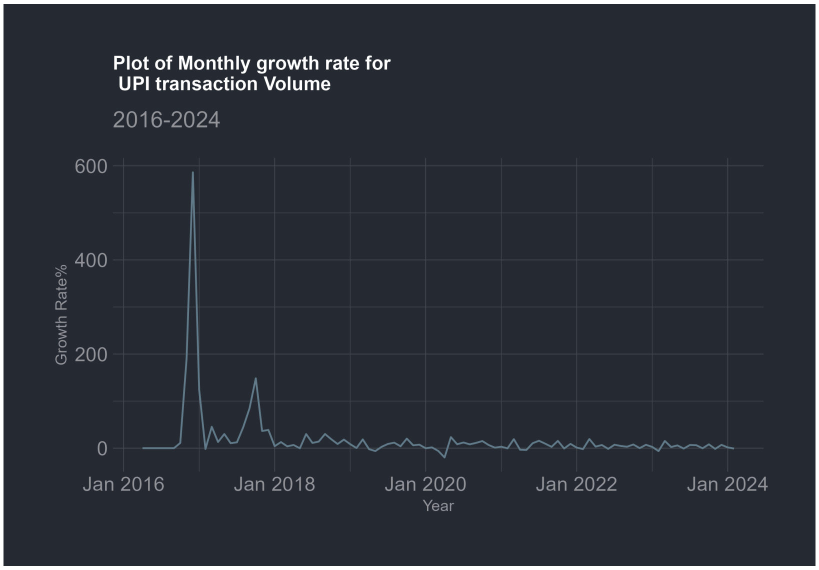


Figure 14: Monthly Growth Rate

Here is the first and last few rows of the growth rate-

Table 14: Monthly Growth Rate for UPI Volume of Transaction

	GrowthRate
Apr 2016	0
May 2016	0
Jun 2016	0
Jul 2016	0
Aug 2016	0
Sep 2016	0

Table 15: Monthly Growth Rate for UPI Volume of Transaction

	GrowthRate
Sep 2023	-0.2865099
Oct 2023	8.0818971
Nov 2023	-1.5207572
Dec 2023	6.9863795
Jan 2024	1.5206864
Feb 2024	-0.8222555

Plotting the monthly growth rate reveals initial values that are exceptionally high, hindering the visualization of subsequent changes. Assuming these outliers stem from the anticipated initial high growth rate, it is better to exclude these values and re-plot the data, focusing on observations following the initial period

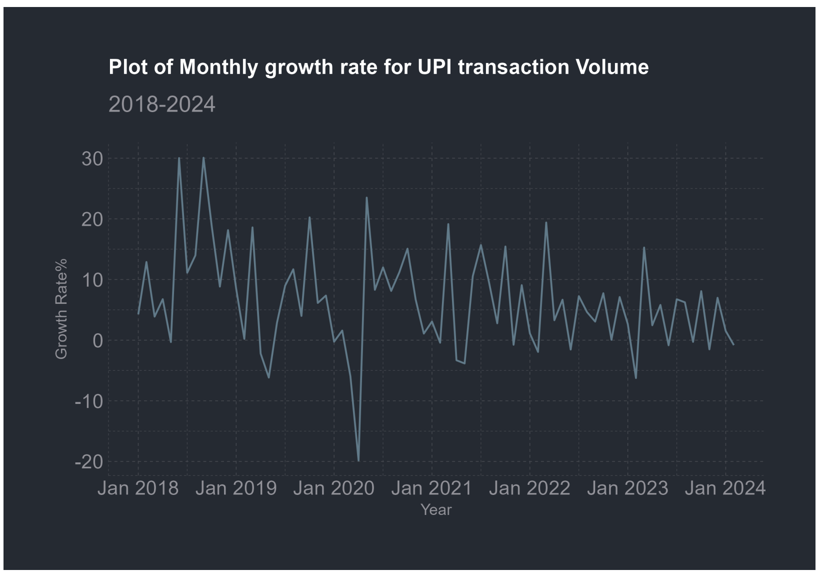


Figure 15: Monthly Growth Rate discarding intital volatility

Visually the data looks kind of stationary , to confirm this assumption **Augmented Dickey Fueller** Test is used to look for unit roots and find if the data is truly stationary or not and also find the lag order.Here the assumed significance level is 0.05.

```
## Warning in adf.test(growth1): p-value smaller than printed p-value
```

Table 16: ADF-Test Results

statistic	p.value	parameter	method	alternative
-4.753066	0.01	4	Augmented Dickey-Fuller Test	stationary

It can be observed that the p value for the test is less than the assumed significance level of 0.05 .So the null hypothesis is rejected and it can be concluded that the data is stationary.

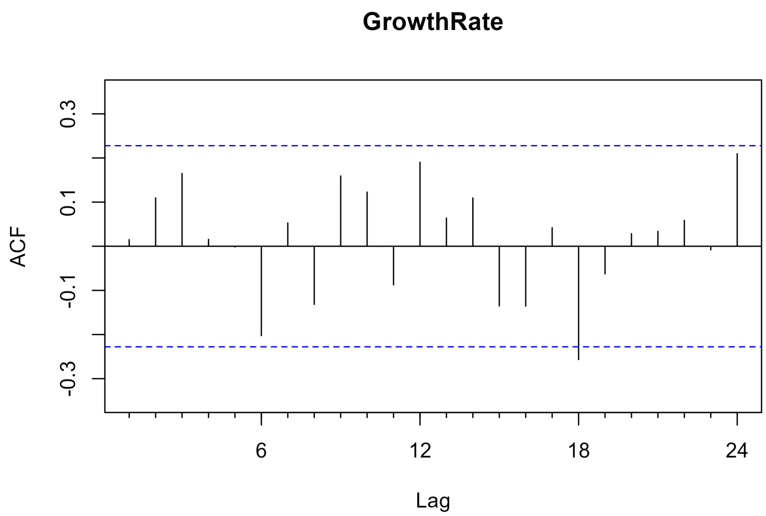


Figure 16: ACF For Monthly growth rate

from the plot it is found that the process acf is identical to a **white noise** process.Some short term predictions for the monthly growth rate can be made using a **Simple Exponential Smoothing** forecast this is done using the `ses()` function in `forecast` package.

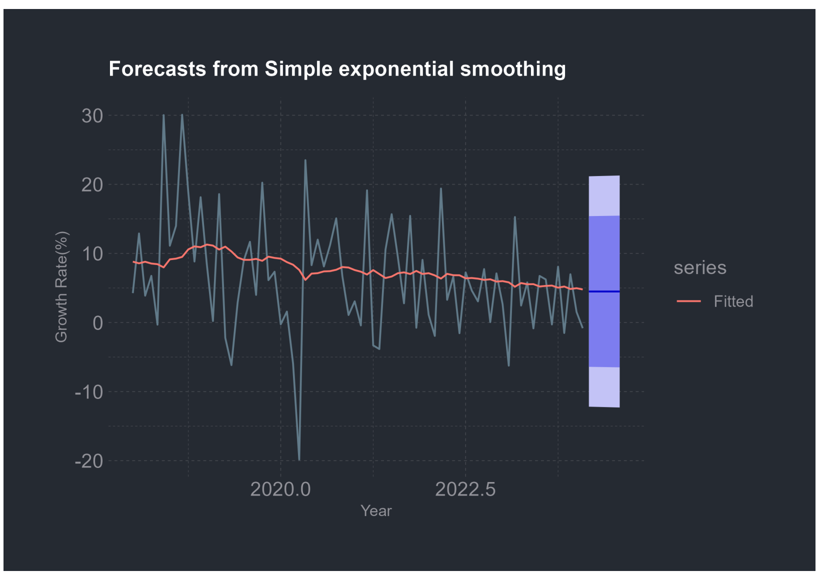


Figure 17: SES Forecast

It can be seen that the forecast model is naive and the fit isn't very identical. The forecasts for future growth rate are given around 5% positive growth for the next months.

This is a simple and naive forecast so it won't be absolutely perfect.But it does provide some idea.

Table 17: Accuracy Measure for Simple Exponential Smoothing

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	-1.11	8.39	6.49	-32	503.39	0.96	-0.06

The accuracy measures show less than satisfactory results[1]. Applying the **Holt-Winters Exponential smoothing** using the `HoltWinters()` function might improve the forecast.

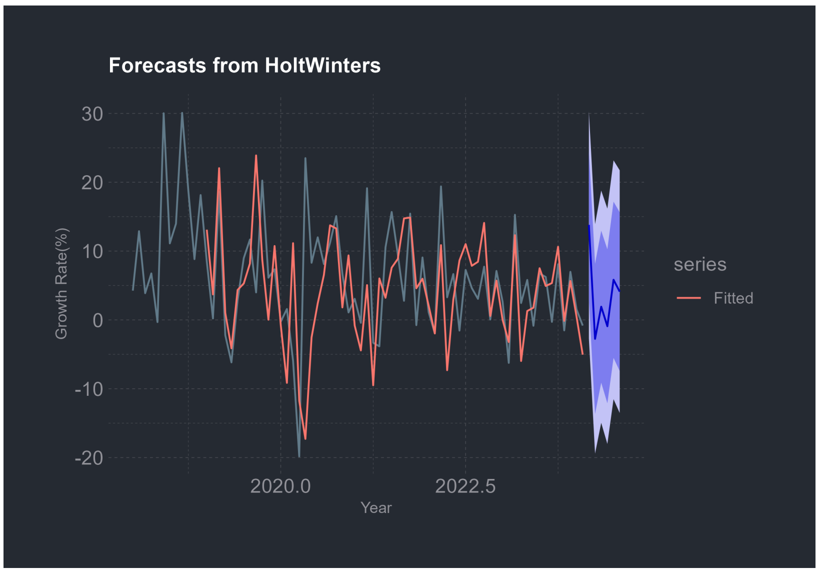


Figure 18: Holt-Winters Forecast

The forecast occasionally lags behind actual values within the sample. Looking ahead, predictions suggest approximately 15% growth for the next month, followed by a stabilization around 5% to 8% positive growth thereafter.

Table 18: Accuracy Measure for Simple Exponential Smoothing

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	0.85	8.39	5.73	9.63	228.76	0.84	0.04

From the accuracy measures we can see that the Holt-Winters Model turned out to be better than the SES model, based on accuracy measures like MAE,MASE etc.

[1]MAPE should be ignored here since there are 0 values which skews the MAPE index

2 Effect of Inflation

There are many underlying variables which have considerable effect in this study, one such example is inflation.Inflation is the rate of increase in prices over a given period of time.A simple example can be used to show what effect does inflation play in this study, say person **X** buys object **A** regularly using UPI, if due to inflation this object A's price keeps increasing then despite the volume of UPI transactions staying same, the value of UPI transactions will keep rising, This could lead to unreliable forecasts since there would be an underlying effect of inflation which the forecasts wouldn't be able to predict.

2.1 Inlfation in India

The most well-known indicator of inflation is the Consumer Price Index (CPI), which measures the percentage change in the price of a basket of goods and services consumed by households. In India the general Consumer Price Index is shared by the Ministry of Statistics And Programme implementation on a monthly basis, via a press release the latest of such is this

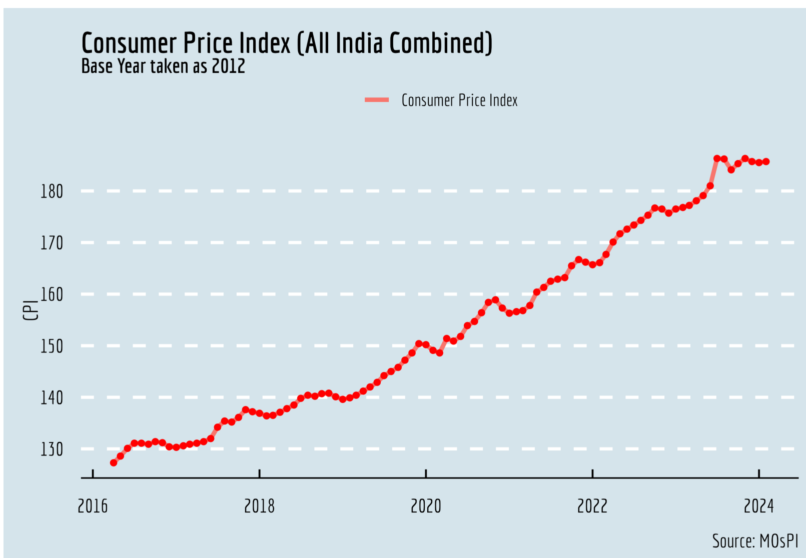


Figure 19: Monthly Aggregated CPI

As it can be seen there is a steady increase of CPI in this period with some mild dips in some certain sections,this means there is a linear increase in inflation through the years.To overcome the effect of inflation in our Value of Transaction data we can use the index numbers for deflation, but even then the figures might not be a true representation of the actual situation because there might be even more such variables which have an underlying significant effect in value of transactions.

This suggests it will be better to analyzye volume of transactions- Since

1. This is independent of inflation since prices increasing doesn't mean number of transactions have to increase,it doesn't imply reduction in number of transactions also, since even though increasing prices may lead consumers to stop buying certain objects but th requirement for that object still needs to be fulfilled a transaction has to happen.

2. It was seen that the growth of both value and volume of payments have been nearly identical so forecasting one can give idea of future forecasts for the other.

3 Analyzing Daily UPI transactions(2016-2018)

This data has been collected from RBI daily payment system indicators.This data is daily updated by RBI and is provided in the form of a excel workbook with multiple sheets where each sheet contains data about every months data from 2020 to the most latest data available. The data was in a format with multiple sub-columns within each column ,`R` isn't well suited for handling this kind of data so first a power query was run through the excel file to combine multiple sheets into a single sheet . The original file contained more columns and data about other digital payment metrics as well but since these data were added in different intervals of time so some of them were scrapped .Some of the columns contain 0 values these are bank dependent payment methods so they are turned off during bank holidays(some Saturday's and Sunday's and other bank holidays) .

Issues in Analysing Daily Data

The main issues that arise while analyzing daily data are the effects of multiple seasonality since it is hard to model such a component, more issues arise if these components follow some irregular pattern.

3.1 Exploratory Data Analysis

The first few rows of the data set is –

Date	UPI_Vol	RTGS_Vol	NEFT_Vol	IMPS_Vol	AePS_Vol	CTS_Vol
2020-06-01	476.9671	4.85000	172.11000	76.80648	0.43618	17.5486
2020-06-02	476.7818	4.54340	100.06772	72.24891	0.44138	18.2500
2020-06-03	456.2593	4.30157	100.36426	68.14805	0.43952	16.7600
2020-06-04	463.0496	4.35152	94.65655	70.68543	0.44828	17.3900
2020-06-05	464.7940	4.56267	111.26259	72.99507	0.47535	18.2500
2020-06-06	458.6493	3.78611	77.05000	70.34825	0.53671	17.5600
2020-06-07	427.2591	0.00000	8.35691	54.24646	0.43795	0.0000
2020-06-08	469.9929	5.32742	121.32275	71.29805	0.61689	20.4500
2020-06-09	466.9834	4.94615	95.19347	69.54556	0.63000	20.4600
2020-06-10	461.5806	4.78815	87.89294	69.47982	0.69205	21.4000
2020-06-11	449.6500	4.68150	79.90802	68.71000	0.63000	20.2000
2020-06-12	453.4291	5.23362	78.75303	68.23533	0.63961	20.5000
2020-06-13	289.0025	0.00000	16.07575	47.24816	0.58489	0.0000
2020-06-14	435.8700	0.00000	10.82748	57.40624	0.39485	0.0000
2020-06-15	463.9147	6.73663	95.28147	72.58162	0.53362	29.0000
2020-06-16	469.2435	5.31426	82.53249	67.36546	0.54833	25.7100
2020-06-17	446.5830	4.96661	72.76998	68.21495	0.53836	22.2400
2020-06-18	433.3174	4.78605	70.91951	66.09145	0.58279	20.6900
2020-06-19	440.2921	4.71551	65.82844	65.63303	0.56548	19.0900

2020-06-20	437.7465	3.87281	53.97209	64.78842	0.48062	18.7400

Plotting the daily UPI transaction volume

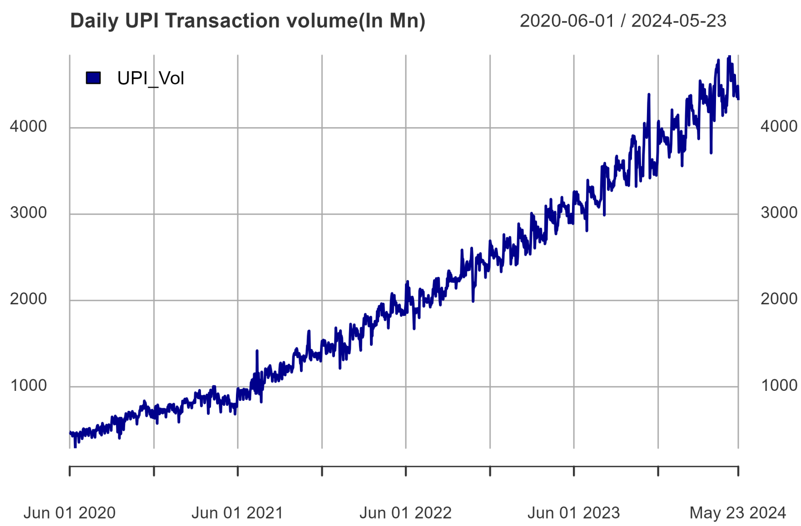


Figure 20: Daily Time Series

There appears to be a discernible repeating pattern in the data, a new observation. Unlike the previously analyzed monthly data, which showed no seasonality or cyclic behavior, further inspection reveals patterns emerging within specific monthly periods. To gain a clearer understanding of the pattern, zooming into a specific portion of the graph for detailed analysis would be beneficial.

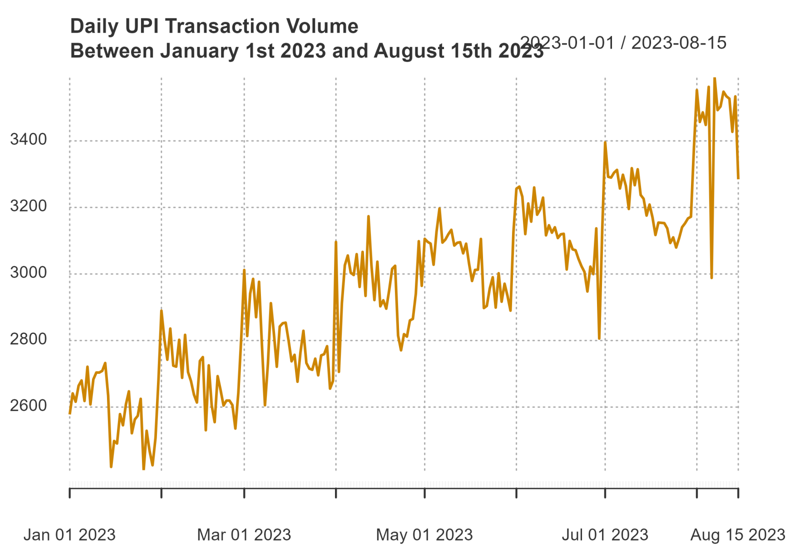


Figure 21: Zoomed Graph

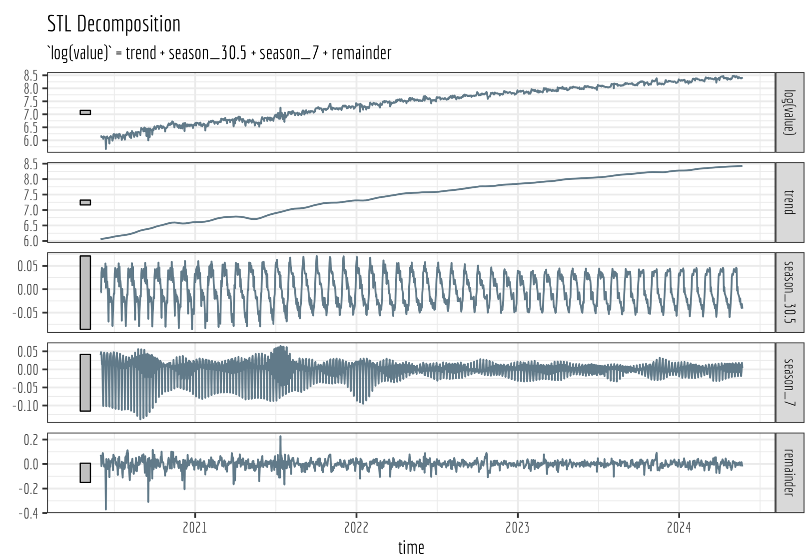
From the plot, it's evident that at the start of each month, the transaction volume reaches its peak, which then gradually decreases throughout the month until another peak is reached at the beginning of the following month. This pattern represents a monthly seasonal component. Additionally, this seasonal component appears to be increasing along with a trend in the data. Therefore, it would be appropriate to consider a multiplicative decomposition when decomposing the data.

3.1.1 Decomposition

The main problem with decomposing a daily time series is that monthly seasonal patterns are hard to catch since their period of occurrence although is technically seasonal but is irregular patterns since all months don't have the same number of days. So a decomposition may be performed based on an assumed model of -

$$Data = Season_m * Season_w * Trend * Error$$

where $Season_m$ &$Season_w$ are monthly and weekly seasonality respectively. Since the data is daily so classical decomposition is not really a option, since classical decomposition is unable to catch seasonality within the months and there is no provision for multiple seasonality.To overcome this issue the **STL** decomposition method can be used, here STL stands for "Seasonal and Trend decomposition using LOESS(locally estimated scatterplot smoothing)" ,This method was developed by R. B. Cleveland et al. (Cleveland et al. 1990).STL has several advantages over the classical decomposition or more specific seasonal decomposition methods like Ratio to trend , Ratio to Moving Average (Gupta and Kapoor 1994) etc, such as it considers multiple seasonal components, it allows the seasonal component to change with time unlike the classical method and most importantly there is no loss of data , i.e decomposed values for all observations are available.A log_e transform is applied to the data to reduce the variance, later to get the individual components an inverse transformation can be done.It is also being done since STL does not allow for a direct multiplicative model.



It can be seen from the decomposition plot that the seasonal component is increasing with time,the trend component is fairly smooth and shows a upward growth as it was seen in the monthly data.The seasonality shows an increasing trend towards the end of the year this can be attributed to increase in festivities during the later part of the Year .

A sample of The decomposed data is given as -

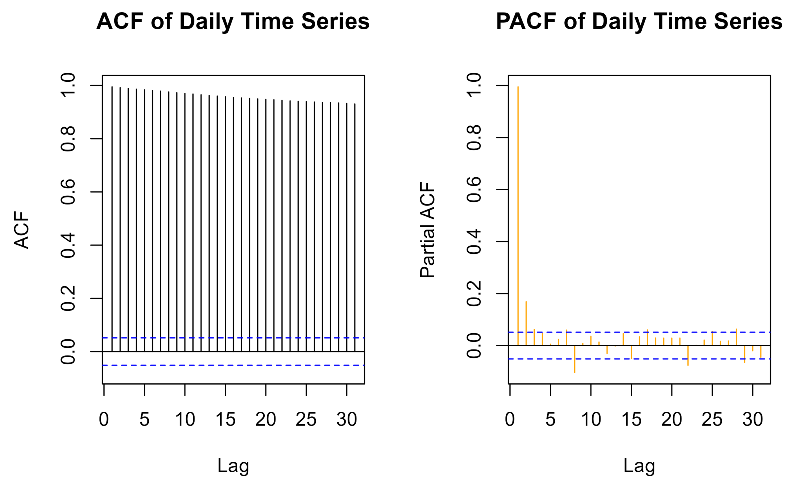
Table 20: Decomposition of Daily UPI Volume of Transactions

time	log(value)	trend	season_7	season_30.5	remainder	season_adjust
2024-04-24	8.365649	8.406829	0.0160306	-0.0595049	0.0022948	8.409123
2024-04-25	8.370281	8.407321	0.0039775	-0.0495043	0.0084875	8.415808
2024-04-26	8.337756	8.407813	-0.0069799	-0.0373691	-0.0257081	8.382105
2024-04-27	8.383310	8.408305	0.0146986	-0.0221473	-0.0175461	8.390758
2024-04-28	8.356160	8.408797	-0.0002167	-0.0039483	-0.0484713	8.360325
2024-04-29	8.378448	8.409289	-0.0207066	0.0250443	-0.0351780	8.374111
2024-04-30	8.410265	8.410046	-0.0063633	0.0325978	-0.0260157	8.384030
2024-05-01	8.477348	8.410803	0.0169964	0.0315253	0.0180227	8.428826
2024-05-02	8.470523	8.411561	0.0034836	0.0308137	0.0246647	8.436226
2024-05-03	8.457874	8.412318	-0.0072195	0.0392099	0.0135655	8.425884
2024-05-04	8.484850	8.413076	0.0139053	0.0433224	0.0145464	8.427622
2024-05-05	8.457802	8.413833	-0.0012610	0.0402728	0.0049570	8.418790
2024-05-06	8.448901	8.414614	-0.0198640	0.0387705	0.0153807	8.429995
2024-05-07	8.421587	8.415395	-0.0062774	0.0370250	-0.0245558	8.390840
2024-05-08	8.464079	8.416177	0.0174366	0.0398965	-0.0094302	8.406746
2024-05-09	8.457689	8.416958	0.0035429	0.0461836	-0.0089946	8.407963
2024-05-10	8.463862	8.417739	-0.0073782	0.0441136	0.0093879	8.427127
2024-05-11	8.434972	8.418520	0.0136971	-0.0030539	0.0058087	8.424328
2024-05-12	8.426270	8.419355	-0.0018747	0.0052486	0.0035410	8.422896
2024-05-13	8.382088	8.420191	-0.0191279	-0.0086318	-0.0103432	8.409848
2024-05-14	8.410697	8.421026	-0.0062857	-0.0114661	0.0074219	8.428448
2024-05-15	8.436376	8.421862	0.0178117	-0.0141736	0.0108756	8.432738
2024-05-16	8.410289	8.422697	0.0035745	-0.0178628	0.0018800	8.424577
2024-05-17	8.404365	8.423533	-0.0075278	-0.0264419	0.0148015	8.438335
2024-05-18	8.406344	8.424407	0.0135276	-0.0196329	-0.0119573	8.412450
2024-05-19	8.394246	8.425281	-0.0024315	-0.0329114	0.0043079	8.429588
2024-05-20	8.380035	8.426155	-0.0187864	-0.0308771	0.0035437	8.429698
2024-05-21	8.396112	8.427028	-0.0064093	-0.0396927	0.0151856	8.442214
2024-05-22	8.408543	8.427902	0.0181283	-0.0318203	-0.0056672	8.422235
2024-05-23	8.371374	8.428776	0.0036388	-0.0414512	-0.0195894	8.409186

3.2 Stochastic Modelling & Forecasting.

So far stochastic models have been rarely used to analyze our dataset, now moving to a more sophisticated analysis and forecast using Stochastic Models like **AR** ,**MA**,**ARMA**,**ARIMA** &**SARIMA** .To validate our models the data may be split into test and training parts to check for accuracy measures. The Auto Correlation function and the Partial Autocorrelation Function maybe plotted to see if the process can be identified.

Interpretation of ACF & PACF-



1. **ACF**- The ACF shows significant autocorrelations at all lags, slowly decreasing. This pattern is characteristic of a non-stationary series, typically one that might be differenced to achieve stationarity.
2. **PACF** - The PACF has a significant spike at lag 1 & 2 and then cuts off quickly, interestingly there is some cyclic pattern where significant lags can be seen in lags of multiples of 7,A weekly effect maybe playing effect her. The general suggestion is that the time series might follow an autoregressive process of order 1 - $AR(1)$, but that would undermine the seasonal effects .

Using the `auto.arima()` function from the `forecast` package an optimal ARIMA model based on the lowest AIC values can be found.This function is based on the Hyndman-Khandakar algorithm.(Rob J. Hyndman and Khandakar 2008)

3.2.1 ARIMA Modeling -

The General ARIMA(p,d,q) model is defined as

$$Wt = \alpha_1 W_{t-1} + \cdots + \alpha_p W_{t-p} + Z_t + \cdots + \beta_q Z_{t-q}$$

Where the process is a combination of $AR(p)$ &$MA(q)$ terms. One of the primary assumptions of stochastic modelling is stationarity, although ARIMA does not explicitly requires stationarity since it uses the d parameter as number of differences required to achieve stationarity, but even then the general ARIMA model is not suited for Seasonal data, infact the Extended Seasonal ARIMA model can only take seasonality for weeks or years but there aren't really any such general models for monthly seasonality as it can seen already talked about in the Issues paragraph.If these are ignored then using the `auto.arima` function to fit an ARIMA model the results are -

```
##
##  Fitting models using approximations to speed things up...
##
##  ARIMA(2,1,2) with drift         : 17161.58
##  ARIMA(0,1,0) with drift         : 17410.9
##  ARIMA(1,1,0) with drift         : 17284.3
##  ARIMA(0,1,1) with drift         : 17242.15
##  ARIMA(0,1,0)                    : 17409.97
##  ARIMA(1,1,2) with drift         : 17159.98
##  ARIMA(0,1,2) with drift         : 17231.3
##  ARIMA(1,1,1) with drift         : 17191.76
##  ARIMA(1,1,3) with drift         : 17161.99
##  ARIMA(0,1,3) with drift         : 17227.03
##  ARIMA(2,1,1) with drift         : 17161.67
##  ARIMA(2,1,3) with drift         : 17161.96
##  ARIMA(1,1,2)                    : 17200
##
##  Now re-fitting the best model(s) without approximations...
##
##  ARIMA(1,1,2) with drift         : 17168.56
##
##  Best model: ARIMA(1,1,2) with drift
```

```
## Series: UPI Volume of Daily Transactions
## ARIMA(1,1,2) with drift
##
## Coefficients:
##          ar1      ma1     ma2   drift
##       0.7930  -1.2040  0.2273  2.8383
## s.e.  0.0267   0.0394  0.0369  0.2677
##
## sigma^2 = 7953:  log likelihood = -8579.26
## AIC=17168.52   AICc=17168.56   BIC=17194.92
##
## Training set error measures:
##                      ME     RMSE      MAE       MPE     MAPE      MASE
## Training set -0.2706182 89.02712 59.52952 -0.653771 3.613357 0.9234873
##                       ACF1
## Training set -0.0004667928
```

Here the chosen model is ARIMA(1,1,2).

$$W_t = 0.79W_{t-1} - 1.2040Z_{t-1} + 0.2273Z_{t-2}$$

This is a non seasonal model[2], but our data definitely has some seasonal pattern within it , so the model chosen through `auto.arima()` doesn't seem really ideal this time.
Looking at the forecasts for the next 30 days-

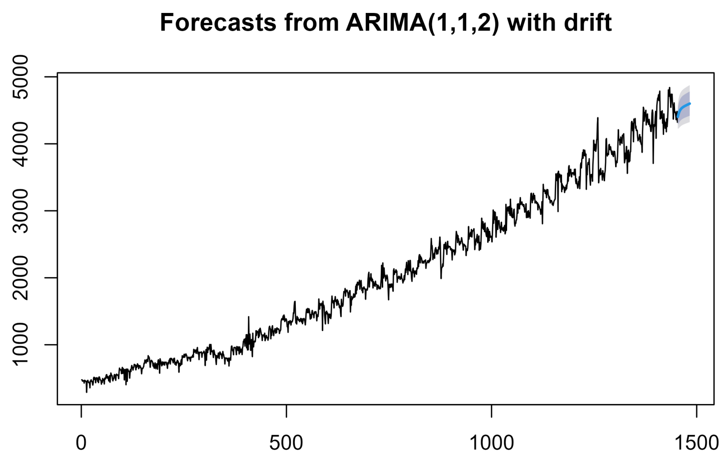


Figure 22: Forecast using ARIMA without considering seasonality

The forecast seems naive and shows no seasonal pattern.

Table 21: Acuuracy Measures for ARIMA

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	-0.2789604	72.3474	50.05477	-0.6376543	3.977516	0.8948011	0.0006335
Test set	320.7306098	412.2910	335.39521	7.7353686	8.212111	5.9956724	NA

The Accuracy measures may be noted for future comparison.

3.2.2 Using a Dynamic Rgeression Model

To find a way to incorporate the seasonality, a dynamic regression model with ARIMA errors where the explanatory variables are fourier terms (where each term is a sin cos pair) may be used, since fourier terms contain a wave pattern they could be useful to simulate the effect of seasonality, the dynamic regression model is given as -

$$y_t = \beta_0 + \beta_1 x_{1,t} + \cdots + \beta_k x_{k,t} + \eta_t$$

[2]in general even SARIMA models can't handle monthly seasonality

Except in this case $\sum \beta_k x_{k,t}$ is replaced with$\phi_t(k)$ where $\phi_t(k)$ is a linear combination of k pairs of sin cos terms each having seperate coefficients, This is also known as Dynamic Harmonic Regression.Here η_t is an ARIMA error term.

```
## Series: Daily UPI Value of Transaction
## Regression with ARIMA(2,1,2) errors
##
## Coefficients:
##          ar1      ar2      ma1     ma2   drift     S1-30    C1-30    S2-30
##       0.7868  -0.0099  -1.2163  0.2416  2.8361  -43.8546  -6.0005  -0.4362
## s.e.  0.1597   0.1050   0.1574  0.1510  0.2672    8.5865   8.5642   5.9145
##         C2-30    S3-30   C3-30    S4-30    C4-30
##       -1.6658  -3.3898  1.9874  -4.2037  -3.0016
## s.e.   5.9083   4.4865  4.4854   3.6927   3.6933
##
## sigma^2 = 7852:  log likelihood = -8565.42
## AIC=17158.84   AICc=17159.13   BIC=17232.77
```

Here 4 fourier terms have been added to the ARIMA model to simulate the seasonality ,it can be seen that the main ARIMA model is a (2,1,2) model i.e it has an AR order of 2 , MA order of 2 and the times the data has been differenced is equal to 1.There is a drift component as well, which is usually the case for data with trend.

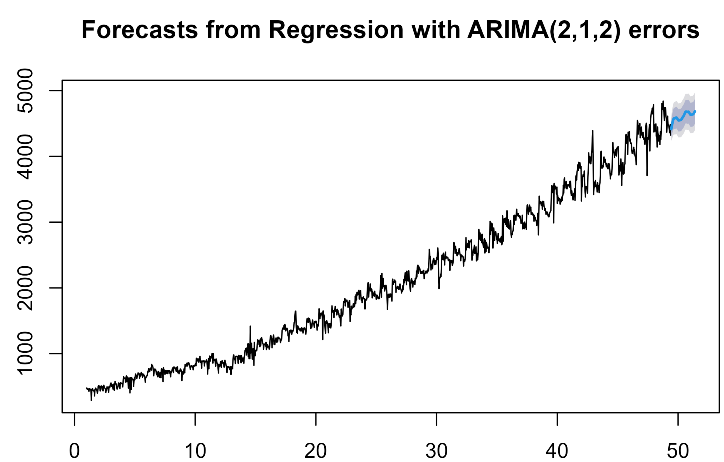


Figure 23: Forecasts after adding fourier terms to the ARIMA model

From the plot it can be seen that adding fourier terms was a good idea since now the forecast does account for the monthly seasonality.But even then the problem persists since the period of months isn't really equal and for that reason the result is a smoothed curve as forecast.The data is now split to check for accuracy measures.

Table 22: Acuuracy Measures for Dynamic Harmonic Regression

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	-0.3696803	73.24721	50.74154	-0.6451053	3.923337	0.8953204	-0.0002667
Test set	135.5477912	302.71755	243.15452	2.8056679	6.109612	4.2903940	NA

The accuracy measures show improvement from the previous model.
Here are the forecasted values-

Table 23: Forecast of Daily UPI Volume of Transactions

	Point.Forecast	Lo.80	Hi.80	Lo.95	Hi.95
2024-05-24	4413.169	4299.611	4526.728	4239.497	4586.842
2024-05-25	4447.406	4316.666	4578.146	4247.456	4647.355
2024-05-26	4478.260	4337.291	4619.229	4262.666	4693.853
2024-05-27	4507.741	4360.151	4655.330	4282.022	4733.460
2024-05-28	4534.722	4382.629	4686.816	4302.116	4767.329
2024-05-29	4556.540	4401.251	4711.830	4319.045	4794.035
2024-05-30	4570.935	4413.287	4728.584	4329.833	4812.038
2024-05-31	4577.773	4418.318	4737.229	4333.907	4821.639
2024-06-01	4579.453	4418.563	4740.342	4333.393	4825.512
2024-06-02	4579.698	4417.631	4741.765	4331.837	4827.559
2024-06-03	4581.472	4418.407	4744.536	4332.086	4830.857
2024-06-04	4585.313	4421.381	4749.245	4334.601	4836.026
2024-06-05	4589.168	4424.462	4753.874	4337.273	4841.063
2024-06-06	4589.829	4424.419	4755.239	4336.856	4842.802
2024-06-07	4585.101	4419.037	4751.164	4331.129	4839.072
2024-06-08	4575.332	4408.654	4742.010	4320.421	4830.243
2024-06-09	4563.388	4396.126	4730.651	4307.583	4819.194
2024-06-10	4553.105	4385.280	4720.929	4296.439	4809.771
2024-06-11	4547.222	4378.852	4715.591	4289.723	4804.721
2024-06-12	4546.092	4377.192	4714.993	4287.781	4804.403
2024-06-13	4547.902	4378.481	4717.323	4288.794	4807.009
2024-06-14	4550.167	4380.235	4720.100	4290.278	4810.057
2024-06-15	4551.470	4381.032	4721.907	4290.808	4812.132
2024-06-16	4552.283	4381.346	4723.220	4290.857	4813.709
2024-06-17	4554.431	4382.999	4725.862	4292.248	4816.613
2024-06-18	4559.624	4387.702	4731.547	4296.691	4822.557
2024-06-19	4568.155	4395.745	4740.566	4304.477	4831.834
2024-06-20	4578.653	4405.758	4751.548	4314.233	4843.073
2024-06-21	4589.070	4415.693	4762.448	4323.912	4854.228

2024-06-22	4598.168	4424.311	4772.026	4332.276	4864.060

3.3 Forecast Based on Decomposition

STL decomposition and then a state-space exponential smoothing can be applied to the decomposed data to find forecasts for future values.This is done using the `stlf()` function. A Box-Cox Transformation with lambda value of 0.4 has been applied in the data to reduce the effect of multiplicative seasonality.(Robin John Hyndman and Athanasopoulos 2018)

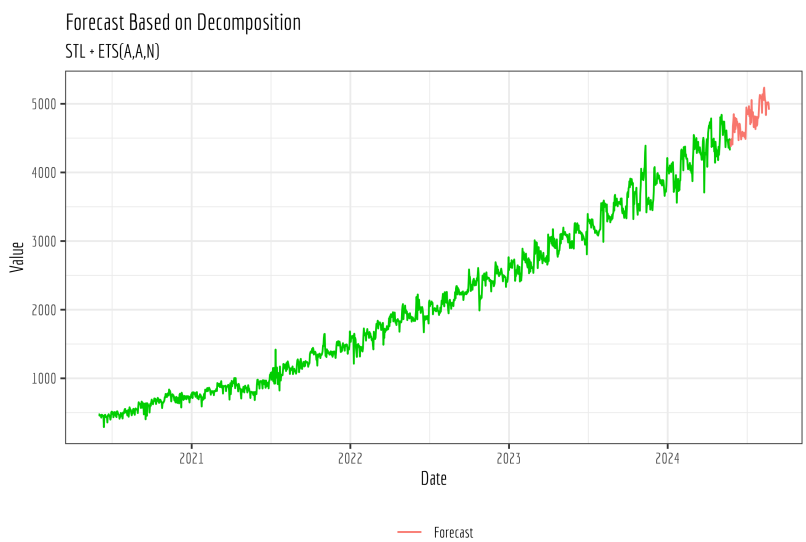


Figure 24: Forecast Based on STL and Exponential Smoothing

Visually the forecast seems much better than the previous ones as it seems to incorporate the seasonal pattern much better.The Model is given by-

```
## ETS(A,A,N)
##
## Call:
##  ets(y = na.interp(x), model = etsmodel, allow.multiplicative.trend = allow.multiplic
##
##   Smoothing parameters:
##     alpha = 0.2036
##     beta  = 1e-04
##
##   Initial states:
##     l = 27.1008
##     b = 0.0293
##
##   sigma:  0.6889
##
```

```
##       AIC     AICc      BIC 
## 9502.829 9502.870 9529.236
```

Here the model is ETS(A,A,N) which is defined as Holt's linear method with additive errors. This model consists of a measurement equation that describes the observed data, and some state equations that describe how the unobserved components or states (level, trend, seasonal) change over time. Hence, this is referred to as state space models. For this model, we assume that the one-step-ahead training errors are given by $\varepsilon_t = y_t - \ell_{t-1} - b_{t-1} \sim NID(0, \sigma^2)$ Substituting this into the error correction equations for Holt's linear method we obtain
$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t,$
$\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t,$
$b_t = b_{t-1} + \beta\varepsilon_t$
where, for simplicity, we have set $\beta = \alpha\beta^*$, here y_t is the forecast equation, and l_t and b_t are the two smoothing equations.

For our data The Smoothing parameter α is equal to 0.2036 and β is equal to 0.0001, showing there's less effect of the second smoothing equation.

Table 24: Acuuracy Measures for STL +ETS

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	0.6300264	53.12098	38.00708	-0.1125392	2.955232	0.0425318	-0.0129705	NA
Test set	33.1809920	156.77699	124.31132	0.9962058	3.378156	0.1391106	0.5053524	1.202261

Acuuracy measures MAPE,MASE [3] etc can be used, Hyndman in his book(Robin John Hyndman and Athanasopoulos 2018) suggests MASE is the best measure for accuracy for comparison between different models based on seasonal data.From the tables(21,22,24) MASE for the STL +ETS model is the lowest compared to the others,so it can be concluded that the model based on STL decomposition and Exponential Smoothing gives the best results.

[3] AIC is not suitable for comparison because for different category of models it is not a comparable measure since it is based on the likelihood function

4 Payment Category Analysis

In this section analysis of how the 3 different payment categories under Peer to Peer and person to merchant payments have seen changed through the time is done suing EDA, The three payment categories are-

1. Less Than 500
2. Greater than 500 but less than 2000
3. Greater than 2000

The dataset is too large to show in a page, a link to the data is given here

4.1 EDA

Plotting the the transaction volumes in different payment categories as percentage of total volume of transaction instead of the raw values-

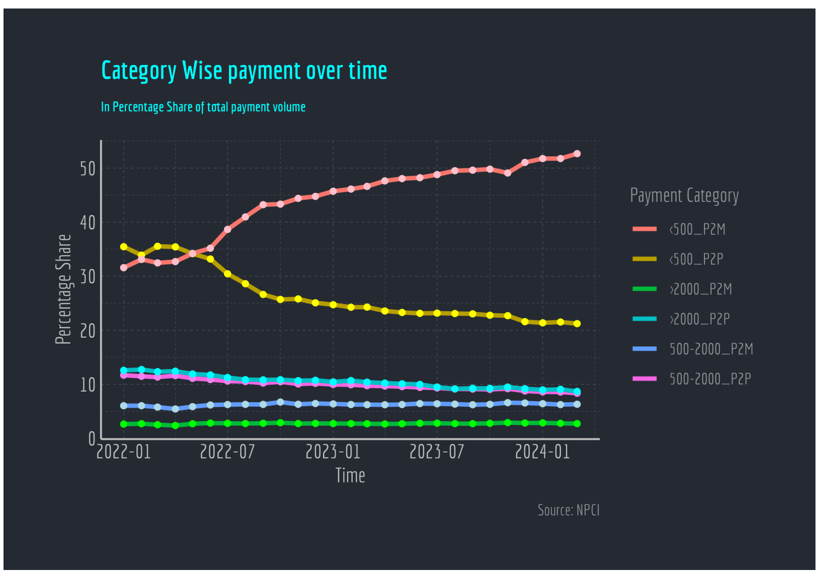


Figure 25: Comparsion Between different Payment categories

It can be seen that except the less than 500 category of Peer to Peer and person to merchant payments the other categories have not shown any significant changes throughout this period,it may also be noted that while less than 500 Peer to Peer payments are decreasing peson to merchant are increasing in a quite inverse proportion.Infact the correlation between the two is given as ρ= -0.9934348,which is nearly a perfect negative correlation.There is a slight change in greater than 2000 Peer to Peer and (500-2000) Peer to Peer payments as it can be seen, there is a downward trend for both of them.

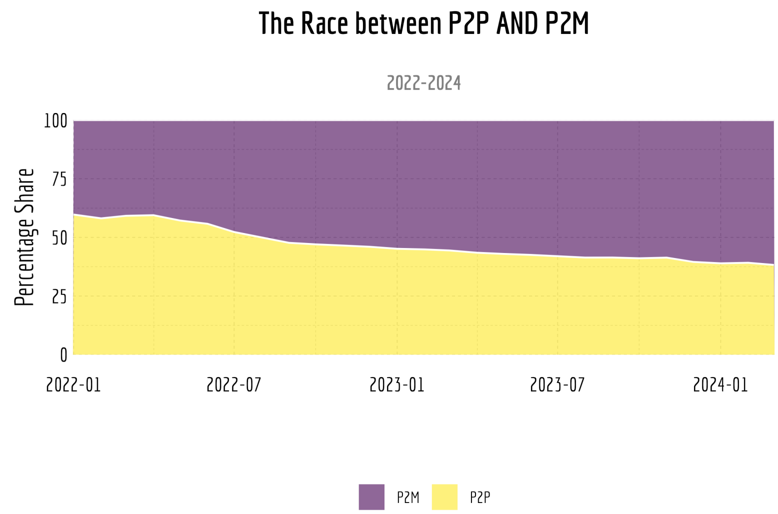


Figure 26: Comparison between p2p and p2m overtime

From the plot it can be seen that over time P2M(Person to Merchant) payments have overtaken P2P(Peer to Peer) payments , this indicates the wide-scale acceptance of UPI by merchants throughout india. In a country where a substantial number of people and businessmen are skeptical about digital payments this is a remarkable achievement since this implies a growing trust towards UPI and a much wider acceptance.

5 Analysing Different Merchant Categories Under UPI

The growth of UPI has been extremely helpful for businesses in our country, in this section an analysis to see which merchant categories fall under high transacting categories and medium transacting categories has been done.

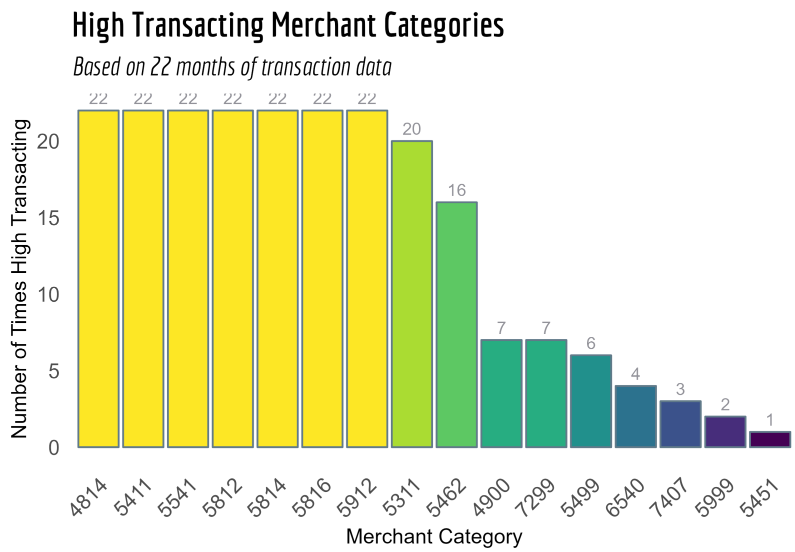


Figure 27: High Transacting Categories

The plot as of it self does not explain the data well since the merchant codes in the x axis are not self explanatory,so a table with attached description for these codes is shared-

MCC	count	Description
4814	**22**	Telecommunication Services
5411	**22**	Groceries And Supermarkets
5541	**22**	Service Stations (With Or Without Ancillary Services)
5812	**22**	Eating Places And Restaurants
5814	**22**	Fast Food Restaurants
5816	**22**	Digital Goods – Games
5912	**22**	Drug Stores And Pharmacies
5311	**20**	Department stores
5462	**16**	Bakeries
4900	**7**	Utilities electric, gas, water and sanitary
7299	**7**	Miscellaneous Personal Services Not Elsewhere Classified
5499	**6**	Miscellaneous Food Shops Convenience And Speciality Retail Outlets
6540	**4**	Debit card to wallet credit (Wallet top up)
7407	**3**	P2PM CHANGES
5999	**2**	Miscellaneous And Speciality Retail Outlets
5451	**1**	Dairies

As it can be seen the categories which are among the High transaction categories are mostly MSME's that directly provide to the public with their services,i.e they don't include very high cost businesses , this shouldn't come out as a surprising result since the introduction of UPI was made to account for the digitization of day to day cash payments for the indian population. This further shows that UPI's main user base includes a rather young aged people. Since the consumers for some of the high transacting business are mostly young people, such as Digital Goods , Fast Food Restaurants etc and this is expected since the majority userbase of smartphones in india is a rather young population.

Most of the merchant categories mentioned here are in general the most important ones, these are businesses which a average person has to deal with every month or week atleast once.Some rather unexpected categories which are worth of interest are Digital Goods(Games) & Bakeries.For digital goods such as live service application and in game objects the introduction of UPI has made it very easy to buy these(Ex:Digital Subscription ,In game currency etc),earlier one had to use credit or debit cards to buy these and the hassle attached to that was a hindrance in the growth of the digital service market in India .

As of 2023 the bakery business in india is worth US$ 12.6 billion and has seen an annual growth rate of 9.6% and the presence of Bakeries in the high transacting category further establishes that bakeries cater very well to the young population.

Here are the merchant categories that fall under medium number of transactions-

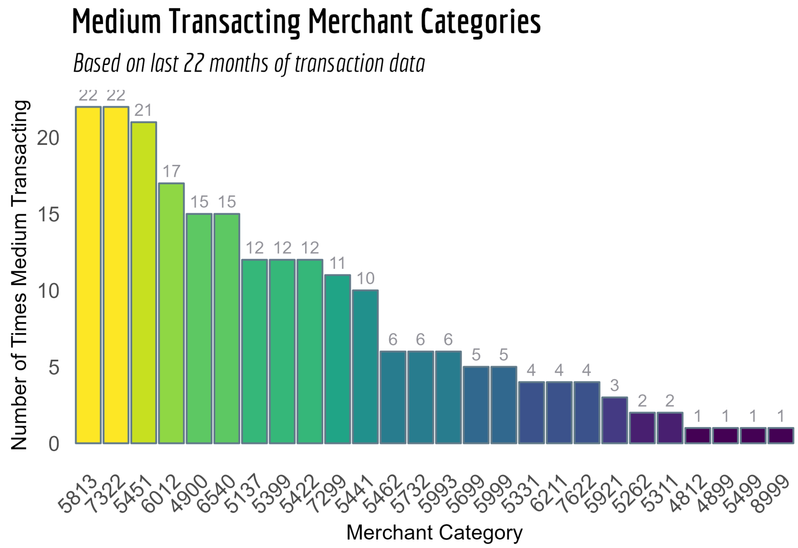


Figure 28: Medium Transaction Categories

And the associated description table is -

MCC	count	Description
5813	**22**	Drinking Places(Alcoholic Beverages) Bars, Pubs etc
7322	**22**	Debt Collection Agencies
5451	**21**	Dairies
6012	**17**	Financial Institutions - Merchandise And Services
4900	**15**	Utilities - Electric, Gas, Water And Sanitary
6540	**15**	Debit Card To Wallet Credit (Wallet Top Up)*
5137	**12**	Mens, womens and childrens uniforms and commercial clothing
5399	**12**	Miscellaneous General Merchandise
5422	**12**	Freezer and locker meat provisioners
7299	**11**	Miscellaneous personal services not elsewhere classified
5441	**10**	Candy, nut and confectionery shops
5462	**6**	Bakeries
5732	**6**	Electronics shops
5993	**6**	Cigar shops and stands
5699	**5**	Miscellaneous Apparel And Accessory Shops
5999	**5**	Miscellaneous and speciality retail outlets
5331	**4**	Variety stores
6211	**4**	Securities - brokers and dealers
7622	**4**	Electronics repair shops
5921	**3**	Package shops beer, wine and liquor
5262	**2**	Online Marketplaces
5311	**2**	Department Stores
4812	**1**	Telecommunication equipment and telephone sales
4899	**1**	Cable and other pay television services
5499	**1**	Miscellaneous food shops convenience and speciality retail outlets
8999	**1**	Professional services not elsewhere classified

In this section there are many categories the one which is of specific interest is “Debt Collection Agencies” , Digital debt collection has seen an unprecedented growth in recent years especially post COVID. The young indian population under the need of quick money are easy targets for digital loan apps which are quick but have rather high interest rates. The fact that with UPI these are accessible within a touch has made them very popular , but this category doesn’t only include these online debt applications but also general debt collecting banks and institutions.

6 Issues Related to UPI

With this exponential growth of UPI comes different concerns the primary ones are Security related concerns and Transaction Faults[4]

6.1 Security Concerns-

What's at stake?

- Virtual payment addresses
- Digital identity of individuals
- UPI ecosystem built and integrated for provisioning services
- Security of the identity, transaction information and data over the network
- Time to respond—transaction speed is the highest
- Customer confidence in the service, market trust and faster adoption by the customer
- Regulatory compliance
- Financial and reputational aspects

Whats being done?

According to the Reserve Bank's Annual Report 2023–24, the number of frauds in the banking sector increased year-on-year to 36,075 out of which 29,082 digital/Card payment related.(Note that this only includes frauds of worth 1 lakh or above) . Specifically for the case of UPI there has been over 95000 frauds reported in the year 2023 This number is higher than the previous one since these include smaller frauds as well. Most of these frauds aren't done through hacking or other technical means rather these are done via different scams and exploitative measures such as vishing[5],fake UPI Id's ,fake QR codes etc. to gain from innocent civillians and their trust in other human beings.

How to safeguard ?

The bedrock of UPI's security architecture lies in its multi-factor authentication system. When you set up a UPI account, you link it to your mobile number, bank account, and a UPI Personal Identification Number (UPI-PIN). This multifactor authentication ensures that only the account HOLDer can initiate transactions, thus mitigating the risk of unauthorised access. Moreover, each transaction requires a UPI-PIN, acting as an additional layer of security.

[4] here Faults mean different technical and banking errors for which transactions are not completed.
[5] scamsters posing as bank representatives

- Digital literacy and educating individuals on the UPI payment methodology is indeed key in preventing users from falling prey to many of such frauds and the Indian government is regularly working on increasing awareness among the individuals.
- Besides this, many fintechs are also working to innovate newer Ways of Securing and Authenticating UPI payments in India. Some of the possible innovative ways in which UPI payments can be made more secure in India:
- Using biometric authentication for UPI payments such as multi-factor authentication for UPI payments could include biometric authentication, possibly through mobile devices, though the challenge at hand is that iris biometrics is not supported by a majority of smartphones in India.Adding on voice, finger, or face biometric authentication by validating the user's voice/face with pre-saved audio/pic samples can be a breakthrough in enhancing the security of UPI payments in India.

6.2 Transaction Faults-

With a rise in number of transactions comes the obvious issue of load handling and digital capacity, there's a limit to how many transactions the bank can process and not only the bank for a UPI payment there are multiple parties involved who need to be working simultaneously these include the concerned UPI app(G-Pay,PhonePe etc) senders bank and receivers bank also banks associated with creating the UPI Id etc.

Here is a sample data for the month of March 2024 about transaction declines,business declines and total volume of transactions -

Table 25: For the Month of March,2024

UPI Remitter Banks	Total Volume(In Mn)	Approved %	BD %	TD%
State Bank Of India	3557.54	0.9499	0.0470	0.0032
HDFC BANK LTD	1158.51	0.9613	0.0386	0.0001
Bank of Baroda	141.51	0.9330	0.0625	0.0045
Union Bank of India	848.59	0.9347	0.0514	0.0140
Punjab National Bank	728.08	0.9422	0.0554	0.0023
Kotak Mahindra Bank	703.56	0.9366	0.0620	0.0014
Canara Bank	642.81	0.9372	0.0591	0.0037
Axis Bank Ltd.	630.13	0.9607	0.0393	0.0001
ICICI Bank	603.62	0.9640	0.0352	0.0009
Bank of India	434.19	0.9463	0.0503	0.0034
Indian Bank	400.98	0.9439	0.0540	0.0021
Airtel Payments Bank	387.91	0.9125	0.0863	0.0012
India Post Payment Bank	312.23	0.8017	0.1385	0.0598
Federal Bank	204.83	0.9479	0.0501	0.0020
Central Bank Of India	200.37	0.8598	0.1069	0.0333
Indian Overseas Bank	188.01	0.9266	0.0492	0.0242

UCO Bank	171.87	0.9319	0.0590	0.0091
Bank of Maharashtra	153.39	0.8977	0.0541	0.0483
IDBI Bank Limited	151.70	0.9327	0.0526	0.0147
INDUSIND BANK	142.64	0.9276	0.0676	0.0047
Fino Payments Bank Limited	136.77	0.8898	0.0953	0.0148
Karnataka Bank	106.23	0.9283	0.0607	0.0110
Yes Bank Ltd	103.43	0.9443	0.0540	0.0017
Paytm Payments Bank	98.18	0.8973	0.1025	0.0003
IDFC FIRST Bank	83.30	0.9299	0.0672	0.0028
Karur Vysya Bank	67.68	0.9162	0.0587	0.0251
Bandhan Bank	54.03	0.9313	0.0590	0.0097
South Indian Bank	53.06	0.9566	0.0430	0.0004
Jammu and Kashmir Bank	52.39	0.9419	0.0501	0.0080
AU small Finance Bank	45.20	0.9376	0.0501	0.0123
Tri O Tech Solutions Private Limited	43.69	1.0000	0.0000	0.0000
Ujjivan Small Finance Bank	42.64	0.9416	0.0562	0.0022
City Union Bank	40.70	0.9212	0.0757	0.0031
Punjab and Sind Bank	36.91	0.8897	0.0685	0.0418
Tamilnad Mercantile Bank	36.68	0.9140	0.0686	0.0175
Equitas Bank	32.26	0.9246	0.0726	0.0028
Andhra Pradesh Grameena Vikas Bank	28.83	0.8418	0.1267	0.0315
Kerala Gramin Bank	27.13	0.9284	0.0620	0.0096
DBS Bank India Limited	26.67	0.9470	0.0504	0.0026
Pragathi Krishna Gramin Bank	26.20	0.8982	0.0808	0.0210
Maharashtra Gramin Bank	25.45	0.8757	0.1161	0.0083
Standard Chartered	25.13	0.9712	0.0269	0.0019
Rajasthan Marudhara Gramin Bank	24.86	0.8335	0.1187	0.0478
Saraswat Bank	23.28	0.9574	0.0394	0.0032
Baroda UP Gramin Bank	22.31	0.7422	0.1056	0.1522
ESAF Small Finance Bank Ltd.	21.76	0.9386	0.0508	0.0106
Fincare Small Finance Bank Ltd	21.32	0.8510	0.1443	0.0047
Citibank	20.78	0.9661	0.0306	0.0032
Sarva Haryana Gramin Bank	17.71	0.9250	0.0700	0.0050
Andhra Pragathi Grameena Bank	17.33	0.8452	0.1338	0.0210

Business Decline (BD) - Transaction decline due to a customer entering an invalid pin, incorrect beneficiary account etc. 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. Such declined transactions are termed as Business Decline. Any decline which is not because of a technical reason of the bank or NPCI is termed as business Decline.
Technical Decline (TD) - Transaction decline due to technical reasons, such as unavailability of systems and network issues on bank or NPCI side.

The main interest lies in how business declines and transaction declines have changed through the years and if the large number of upi transactions have resulted in more faults. To do this a weighted average is calculated where the weights are the ratio of total volume of transaction for a specific bank in a month and the total volume of transactions for all the banks for that month.The reason for such a choice of weight is, since some the banks such as SBI , Axis Bank etc. have very high user base while some have very low.

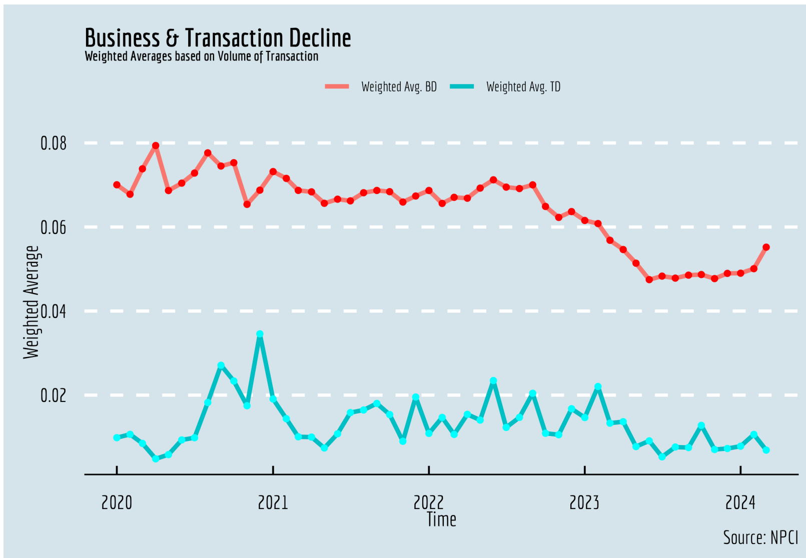


Figure 29: Transaction and Business decline overtime(Weighted averages)

It can be seen throughout the period that the transaction decline which is associated with upi servers and bank servers have been fairly low and is slowly decreasing which is an optimistic news since this means the technology handling these transactions are robust and it is suitable to assume that it will get even better with time.

Business Declines have always been higher than transaction decline but over time it can be seen that the numbers are decreasing showing increased literacy about UPI in the userbase further proving the wider acceptability of UPI.

Conclusion

- The growth of Unified Payments Interface (UPI) in India has been nothing short of transformative, revolutionizing the digital payments landscape with remarkable efficiency and convenience. This project has delved into multiple dimensions of UPI's growth and provided a comprehensive analysis of various factors influencing its trajectory.

- Through the Monthly UPI Value and Volume Time Series Analysis,employed basic Exploratory Data Analysis (EDA) and trend fitting using moving averages, followed by a logistic curve fitting. The results demonstrated that logistic growth is a reliable estimate for UPI's expansion, highlighting a rapid adoption phase followed by a stabilization, indicative of a maturing market.

- In the analysis of Per Transaction Value for UPI payments per month, basic EDA revealed critical insights about how UPI is being used for more and more low cost transactions. Forecasting using methods such as Holt-Winters, ARIMA, and exponential smoothing provided robust predictions, showcasing the decreasing average transaction value, which reflects increasing consumer trust and reliance on UPI for small value transactions.

- The Monthly Growth Rate of UPI was calculated and forecasted using ARIMA and other statistical methods. This section highlighted the dynamic growth rate of UPI, underlining its rapid adoption and the factors contributing to its fluctuating growth rates over time.

- The effect of inflation in value of UPI transactions was considered, and analysis showed it is better to forecast volume of transactions since it is independent of such outside variables.

- For the Daily Time Series Analysis of UPI daily transaction volume, EDA was conducted to address the seasonality issues inherent in the ARIMA model. By incorporating Fourier terms, the model's forecasting accuracy was enhanced.

- A final forecasting for the daily time series was done using decomposition and exponential smoothing which showed much better results than the previous ones, incorporating monthly seasonal patterns.

- Additionally, the EDA for payment categories and merchant categories provided granular insights into the diverse applications of UPI, from peer-to-peer transfers to merchant payments. This diversity underscores UPI's versatility and broad acceptance across various sectors.

- The study also addressed pertinent issues such as security concerns and transaction faults.

- Further insights into transaction faults showed an optimistic future for UPI, with better infrastructure and widescale education about UPI usage instructions.

In conclusion, UPI's growth trajectory in India has been characterized by rapid adoption, increasing transaction values, and broad application across payment categories. While the growth forecasts are promising, it is imperative to address the security issues and operational declines to ensure the continued success and stability of the UPI ecosystem. This project provides a comprehensive understanding of UPI's growth dynamics and serves as a valuable resource for stakeHOLDers aiming to enhance the digital payment infrastructure in India.

Some End Notes

1. The project is publicly available in the github repository named UPI_Analysis all the associated dataset and codes are available there , there are also some guides for future reference of others to help them in creating a project report using `Rmarkdown`.

2. The sources for the different datasets are-

- NPCI Website
- RBI Website

References

Chatfield, C. 2016. *The Analysis of Time Series: An Introduction, Sixth Edition*. Chapman & Hall/CRC Texts in Statistical Science. CRC Press. https://books.google.co.in/books?id=qKzyAbdaDFAC.

Cleveland, Robert B., William S. Cleveland, Jean E. McRae, and Irma Terpenning. 1990. "STL: A Seasonal-Trend Decomposition Procedure Based on Loess (with Discussion)." *Journal of Official Statistics* 6: 3–73.

Gupta, S. C., and V. K. Kapoor. 1994. *Fundamentals of Applied Statistics*. Sultan Chand & Sons. https://books.google.co.in/books?id=4wm8YgEACAAJ.

Hyndman, Rob J., and Yeasmin Khandakar. 2008. "Automatic Time Series Forecasting: The Forecast Package for r." *Journal of Statistical Software* 27 (3): 1–22. https://doi.org/10.18637/jss.v027.i03.

Hyndman, Robin John, and George Athanasopoulos. 2018. *Forecasting: Principles and Practice*. 2nd ed. Australia: OTexts.

Kendall, M. G., and A. Stuart. 1966. *The Advanced Theory of Statistics: Vol. 3, Design and Analysis, and Time-Series*. v. 3. Griffin. https://books.google.co.in/books?id=lvlgzwEACAAJ.

Rogers, E. M. 2003. *Diffusion of Innovations, 5th Edition*. Free Press. https://books.google.co.in/books?id=9U1K5LjUOwEC.