Praxis Tech School

# A Statistical Deep Dive into India's Digital Payment Boom: Comparative Analysis of Google Pay and PhonePe

by

**Saurabh Girpunje (C25008)**

**Subhadeep Purkayastha (C25009)**

under the guidance of

**Dr. Sayantani Roy Choudhury**

This report presents a comprehensive statistical comparative analysis of India's two dominant UPI platforms: Google Pay (GPay) and PhonePe using monthly transaction data from 2021 to 2025. It aims to uncover key drivers of platform performance, assess the effectiveness of marketing and cashback strategies, and deliver actionable recommendations. Using a combination of descriptive analytics, hypothesis testing, regression models, and feature importance analysis, this study provides critical insights to help both platforms optimize performance and strengthen their competitive edge.

# Agenda:

1. Problem Statement
2. Research Question
3. Data Overview
4. Methodology
5. Analysis and Findings
6. Recommendations

# 1. Problem Statement

India's digital payments landscape has rapidly evolved with the rise of UPI. Despite growing transaction volumes and user bases, GPay and PhonePe face challenges in optimizing return on marketing investment, targeting cashback offers efficiently, and expanding penetration in rural areas. Additionally, the fierce competition between the two requires a strategic understanding of what truly drives engagement and market share. This report seeks to answer these questions through a robust data-driven approach.

# 2. Research Questions

1. **Performance Drivers:** What factors most influence UPI platform market share, and how do success rate and average transaction value impact platform performance?
2. **Cashback Effectiveness:** Do cashback offers and festival seasons significantly impact market share, and are they effective promotional levers for user engagement?

3. **Rural Growth Barriers:** How can merchant penetration and localized strategies help accelerate rural UPI adoption for platforms with urban dominance?
4. **Competitive Edge:** How can both the platforms leverage their platform-specific strengths, transaction value patterns and reliability to gain a competitive edge in the UPI market?

# 3. Data Overview

- **Source**: Monthly UPI transaction data (2021–2025) from NPCI.org, RBI.org.in, blog.Google, PhonePe.com
- **Platforms**: Google Pay (GPay) and PhonePe

- **Total Records**: Monthly data across 5 years per platform (each row represents one month of data per platform)

## Key Variables

- Transaction Volume (Millions)
- Transaction Value (₹B)
- Avg. Transaction Value (₹)
- Monthly Active Users (MAU)
- Success Rate (%)
- Urban/Rural Split (%)
- Marketing Spend (₹M)
- Cashback Offers (Y/N)
- Festival Season (Y/N)
- Merchant Count (Millions)
- Market Share (%)

## Initial Insights

- **PhonePe** leads in volume, MAUs, and merchant count

- **GPay** shows higher avg. transaction value and urban dominance

- **Both** platforms maintain >98% success rate

- Cashback/festival effects vary month-to-month

## 4. Methodology

| | |
|---|---|
| **Data Preparation** | • Combined data from GPay and PhonePe into a unified dataset<br>• Converted categorical variables (Cashback, Festival) to binary format<br>• Parsed Urban/Rural Split into separate numeric values |
| **Descriptive Statistics** | • Calculated central tendencies and dispersion for each platform<br>• Evaluated trends in transaction volume, value, MAU, and marketing spend |
| **EDA** | • Used box plots to compare cashback vs. non-cashback months |
| **Correlation Analysis** | • Applied correlation to identify relationships between variables |
| **Hypothesis Testing** | • T-Test: Cashback impact on volume<br>• Chi-Square: Cashback–festival association |
| **Regression Modeling** | • Built separate **Linear Regression models** for GPay and PhonePe with Market Share (%) as the dependent variable<br>• Included independent variables: Avg. Transaction Value (₹), Success Rate, Cashback, Festival<br>• Evaluated model fit using Adjusted R-square and p-values. |

# 5. Analysis & Findings

## Descriptive Statistics:

**Gpay**

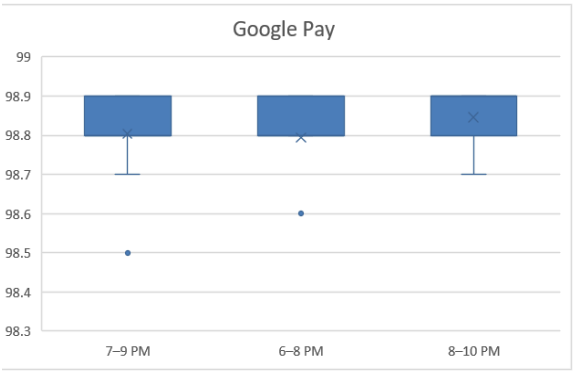
| Parameters | Transaction Volume (Millions) | Transaction Value (₹B) | Avg. Transaction Value (₹) | MAU (Millions) | Marketing Spend (₹M) | Success Rate (%) |
|---|---|---|---|---|---|---|
| Mean | 1756.470588 | 3146.666667 | 1775.490196 | 129.1176471 | 254.9019608 | 98.80980392 |
| Standard Error | 83.00806648 | 159.7313103 | 6.960075081 | 5.962145098 | 12.06151731 | 0.014620668 |
| Median | 1750 | 3100 | 1771 | 113 | 250 | 98.8 |
| Mode | 1600 | 2800 | 1750 | 105 | 220 | 98.8 |
| Standard Deviation | 592.7961657 | 1140.709721 | 49.70487805 | 42.57823249 | 86.13646264 | 0.104412455 |
| Sample Variance | 351407.2941 | 1301218.667 | 2470.574902 | 1812.905882 | 7419.490196 | 0.010901961 |
| Kurtosis | -1.093962925 | -1.10255993 | -1.385880515 | -1.038870632 | -0.815422193 | 1.783894863 |
| Skewness | 0.022850714 | 0.101757459 | 0.12659333 | 0.547668362 | 0.107665909 | -1.409583695 |
| Range | 2080 | 4000 | 158 | 146 | 330 | 0.4 |
| Minimum | 820 | 1400 | 1704 | 74 | 110 | 98.5 |
| Maximum | 2900 | 5400 | 1862 | 220 | 440 | 98.9 |
| Sum | 89580 | 160480 | 90550 | 6585 | 13000 | 5039.3 |
| Count | 51 | 51 | 51 | 51 | 51 | 51 |

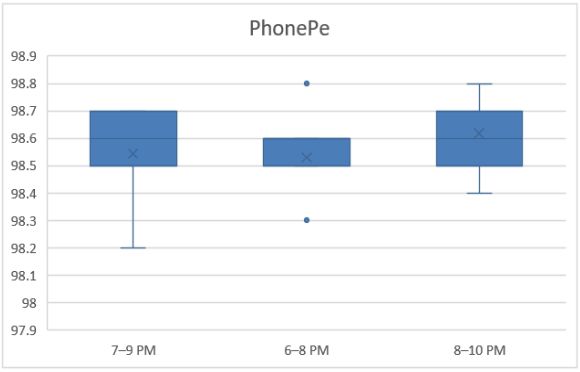
**PhonePe**

| Parameters | Transaction Volume (Millions) | Transaction Value (₹B) | Avg. Transaction Value (₹) | MAU (Millions) | Marketing Spend (₹M) | Success Rate (%) |
|---|---|---|---|---|---|---|
| Mean | 3114.117647 | 5451.372549 | 1719.392157 | 202.8039216 | 533.0392157 | 98.55490196 |
| Standard Error | 211.5333704 | 402.3678952 | 11.1603668 | 9.31315829 | 41.32394227 | 0.019666579 |
| Median | 2900 | 4800 | 1688 | 200 | 480 | 98.6 |
| Mode | 2200 | 4500 | 1667 | 130 | 250 | 98.6 |
| Standard Deviation | 1510.650425 | 2873.481526 | 79.7009607 | 66.50925337 | 295.1119761 | 0.140447464 |
| Sample Variance | 2282064.706 | 8256896.078 | 6352.243137 | 4423.480784 | 87091.07843 | 0.01972549 |
| Kurtosis | -1.106181216 | -0.879800544 | 2.722499901 | -1.135662755 | -0.897969495 | 0.387309768 |
| Skewness | 0.330955468 | 0.504599138 | 1.602506946 | 0.254237918 | 0.459753172 | -0.622785594 |
| Range | 4950 | 9750 | 393 | 222 | 1010 | 0.6 |
| Minimum | 1050 | 1750 | 1607 | 108 | 140 | 98.2 |
| Maximum | 6000 | 11500 | 2000 | 330 | 1150 | 98.8 |
| Sum | 158820 | 278020 | 87689 | 10343 | 27185 | 5026.3 |
| Count | 51 | 51 | 51 | 51 | 51 | 51 |

Google Pay shows lower transaction volume and user base compared to PhonePe, but a slightly higher average transaction value and success rate. PhonePe leads in overall usage and marketing spend, but both platforms demonstrate high reliability with success rates near 98.5%.

## EDA: (Success Rate and Peak Hours)



The box plots show that Google Pay maintains a consistently high and stable success rate (around 98.8%) across all evening time slots, with minimal variation. In contrast, PhonePe exhibits slightly more fluctuation in success rates and a lower average, especially between 7–9 PM.

## Correlation Matrix:

### Google Pay:

| | *Avg. Transaction Value (₹)* | *Success Rate (%)* | *shback Offers (Y/* | *Festival Season (Y/N)* |
|---|---|---|---|---|
| Avg. Transaction Value (₹) | 1 | | | |
| Success Rate (%) | 0.514681441 | 1 | | |
| Cashback Offers (Y/N) | 0.091558625 | 0.107288356 | 1 | |
| Festival Season (Y/N) | 0.140997008 | 0.170957944 | 0.39223227 | 1 |

- Avg. Transaction Value vs Success Rate have Moderate positive correlation (0.51), Higher transaction value tends to be associated with higher success rate
- Cashback Offers and Festival Season have Weak to moderate correlation (0.39), Cashback offers are more likely during festivals, but the relation is not very strong
- Other pairs show very low correlations (< 0.2), so they are likely independent

### PhonePe:

| | *Avg. Transaction Value (₹)* | *Success Rate (%)* | *shback Offers (Y/* | *Festival Season (Y/N)* |
|---|---|---|---|---|
| Avg. Transaction Value (₹) | 1 | | | |
| Success Rate (%) | 0.713255621 | 1 | | |
| Cashback Offers (Y/N) | 0.044591759 | 0.054220484 | 1 | |
| Festival Season (Y/N) | 0.096821321 | 0.213128159 | 0.375045785 | 1 |

- Avg. Transaction Value vs Success Rate have Stronger correlation (0.71), Suggests a stronger relationship than in Google Pay when avg. transaction increases, success rate also increases
- Festival Season and Cashback Offers have Moderate correlation (0.38), Similar to Google Pay offers slightly increase during festivals
- Other variables are weakly correlated (< 0.2)

# Regression:

## Google Pay:

Problem Statement: To identify key factors affecting Google Pay's market share using multiple linear regression.

SUMMARY OUTPUT

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.926010606 |
| R Square | 0.857495642 |
| Adjusted R Square | 0.845103958 |
| Standard Error | 0.901188775 |
| Observations | 51 |

ANOVA

| | *df* | *SS* | *MS* | *F* | *Significance F* |
|---|---|---|---|---|---|
| Regression | 4 | 224.7983672 | 56.1995918 | 69.19928612 | 7.15299E-19 |
| Residual | 46 | 37.35849555 | 0.812141208 | | |
| Total | 50 | 262.1568627 | | | |

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
|---|---|---|---|---|---|---|---|---|
| Intercept | 962.131009 | 139.0732844 | 6.918158386 | 1.20898E-08 | 682.1910068 | 1242.071011 | 682.1910068 | 1242.071011 |
| Avg. Transaction Value (₹) | -0.030559291 | 0.002997115 | -10.19623532 | 2.18583E-13 | -0.03659217 | -0.024526411 | -0.03659217 | -0.024526411 |
| Success Rate (%) | -8.886071144 | 1.4338862 | -6.19719413 | 1.46511E-07 | -11.77233437 | -5.999807922 | -11.77233437 | -5.999807922 |
| Cashback Offers (Y/N) | 0.50743682 | 0.291359827 | 1.741615599 | 0.088260001 | -0.079040094 | 1.093913734 | -0.079040094 | 1.093913734 |
| Festival Season (Y/N) | 0.398981208 | 0.327084087 | 1.219812348 | 0.228753312 | -0.259404911 | 1.057367327 | -0.259404911 | 1.057367327 |

**Interpretation**:

- The regression model explains ~85.7% of the variation in the dependent variable ($R^2 = 0.857$), indicating a strong fit.
- Among the predictors, Avg. Transaction Value ($p < 0.001$) and Success Rate ($p < 0.001$) have a significant negative effect on the outcome.
- Cashback Offers and Festival Season are not statistically significant ($p > 0.05$), suggesting their effect may be weak or inconsistent.
- The F-statistic is high (69.2) with a very small p-value, confirming that the model overall is significant.
- So, focus should be on Avg. Transaction Value and Success Rate, as they drive most of the impact.

**PhonePe:**

Problem Statement: To analyze key drivers impacting PhonePe's market share using multiple linear regression.

SUMMARY OUTPUT

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.877188224 |
| R Square | 0.76945918 |
| Adjusted R Square | 0.749412152 |
| Standard Error | 1.355742341 |
| Observations | 51 |

ANOVA

| | *df* | *SS* | *MS* | *F* | *Significance F* |
|---|---|---|---|---|---|
| Regression | 4 | 282.1953825 | 70.54884562 | 38.38270628 | 4.12083E-14 |
| Residual | 46 | 84.54971556 | 1.838037295 | | |
| Total | 50 | 366.745098 | | | |

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
|---|---|---|---|---|---|---|---|---|
| Intercept | -1151.884406 | 192.2364094 | -5.99201998 | 2.97663E-07 | -1538.836228 | -764.9325835 | -1538.836228 | -764.9325835 |
| Avg. Transaction Value (₹) | 0.010869186 | 0.003446367 | 3.153809706 | 0.002837726 | 0.003932009 | 0.017806362 | 0.003932009 | 0.017806362 |
| Success Rate (%) | 12.00738336 | 1.993050033 | 6.024627159 | 2.65977E-07 | 7.99558172 | 16.019185 | 7.99558172 | 16.019185 |
| Cashback Offers (Y/N) | -0.475352592 | 0.441913114 | -1.075669801 | 0.287684583 | -1.364877554 | 0.414172371 | -1.364877554 | 0.414172371 |
| Festival Season (Y/N) | -0.089745667 | 0.495636154 | -0.181071672 | 0.857106788 | -1.087409501 | 0.907918166 | -1.087409501 | 0.907918166 |

**Interpretation:**

- The model explains about 77% of the variance in the dependent variable ($R^2$ = 0.769), indicating a strong fit.
- Avg. Transaction Value (p = 0.0028) and Success Rate (p ≈ 0.00000027) are statistically significant positive predictors.
- Cashback Offers and Festival Season are not significant (p > 0.05), showing no meaningful impact on the target.
- The F-test is highly significant (p < 0.0001), indicating that the overall model is valid.
- Thus, higher success rate and transaction value contribute most to the target variable in this model.

**Final Conclusion of Regression:**

- The regression analysis for both Google Pay and PhonePe shows that transaction success rate is a consistently strong and statistically significant driver of market share, indicating that reliability plays a key role in user retention and growth.
- Interestingly, average transaction value has opposite effects: it negatively impacts market share for Google Pay but positively affects it for PhonePe, suggesting that each platform benefits from a different user behavior pattern.

- Cashback offers and festival seasons were not statistically significant in either case, implying that these promotional efforts may not have a lasting impact on market share.
- Overall, both models are statistically robust, with high $R^2$ values, meaning the chosen variables explain a substantial portion of market share variation.
- In conclusion, platforms should prioritize improving success rates and strategically align their transaction value strategies with their user base, rather than relying heavily on seasonal promotions or cashback schemes.

## Two Sample T-Test:

**Problem Statement:** To determine if there's a significant difference in average transaction volume between Google Pay and PhonePe.

**t-Test: Two-Sample Assuming Unequal Variances**

| | *Transaction Volume (₹M) Google Pay* | *Transaction Volume (₹M) PhonePe* |
|---|---|---|
| Mean | 1756.470588 | 3114.117647 |
| Variance | 351407.2941 | 2282064.706 |
| Observations | 51 | 51 |
| Hypothesized Mean Difference | 0 | |
| df | 65 | |
| t Stat | -5.974583272 | |
| P(T<=t) one-tail | 5.35433E-08 | |
| t Critical one-tail | 1.668635976 | |
| P(T<=t) two-tail | 1.07087E-07 | |
| t Critical two-tail | 1.997137908 | |

**Null Hypothesis ($H_0$):** There is no difference in the average transaction volume between Google Pay and PhonePe.

**Alternative Hypothesis ($H_1$):** There is a difference in the average transaction volume between Google Pay and PhonePe.

**Conclusion:** There is a significant difference in transaction values. PhonePe has a higher average transaction value than Google Pay. PhonePe appears to be more dominant in terms of user activity or merchant penetration, leading to higher transaction volumes, which could reflect greater market share or user engagement.

## Chi-Square Test:

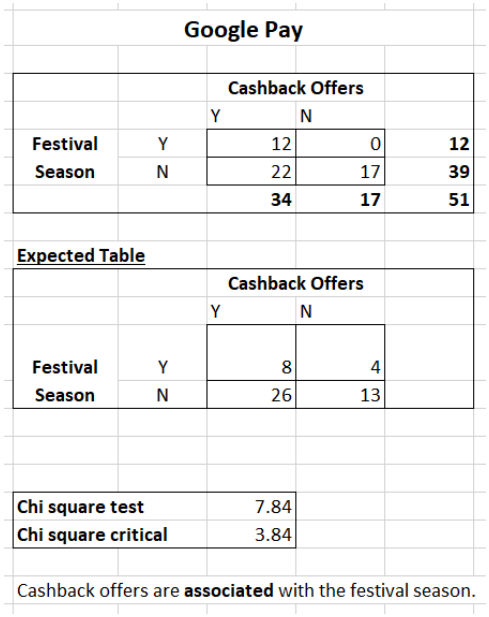
**Problem Statement:** To check if cashback offers on Google Pay and PhonePe are more likely during the festival season using the Chi-square test.

Festival Season (Y/N) vs. Cashback Offers (Y/N)

Columns Used:
Festival Season (Y/N)
Cashback Offers (Y/N)

**Hypothesis**:
**$H_0$:** Cashback offers are independent of whether it's a festival season.
**$H_1$:** Cashback offers are associated with the festival season.

### For Google Pay:



**Google Pay**

| | | Cashback Offers | | |
|---|---|---|---|---|
| | | Y | N | |
| **Festival Season** | Y | 12 | 0 | **12** |
| | N | 22 | 17 | **39** |
| | | **34** | **17** | **51** |

**Expected Table**

| | | Cashback Offers | | |
|---|---|---|---|---|
| | | Y | N | |
| **Festival Season** | Y | 8 | 4 | |
| | N | 26 | 13 | |

| | |
|---|---|
| **Chi square test** | 7.84 |
| **Chi square critical** | 3.84 |

Cashback offers are **associated** with the festival season.

**Conclusion**: Cashback offers are associated with the festival season.

**PhonePe:**

**PhonePe**

| | | Cashback Offers | | |
|---|---|---|---|---|
| | | Y | N | |
| **Festival** | Y | 12 | 0 | **12** |
| **Season** | N | 23 | 16 | **39** |
| | | **35** | **16** | **51** |

**Expected Table**

| | | Cashback Offers | | |
|---|---|---|---|---|
| | | Y | N | |
| **Festival** | Y | 8.24 | 3.76 | |
| **Season** | N | 26.76 | 12.24 | |

| | |
|---|---|
| **Chi square test** | 7.15 |
| **Chi square critical** | 3.84 |

Cashback offers are **associated** with the festival season.

**Conclusion**: Cashback offers are associated with the festival season.

# 6. Recommendations

## 1. Optimize Promotional Timing (Festival-Driven, Not Cashback-Focused)

Focus marketing efforts during festival seasons and not necessarily on cashback.

**Supported by:**

- **Chi-Square Test**: Cashback offers are statistically associated with festival months ($p < 0.05$), confirming festivals are high-impact.
- **Regression (GPay & PhonePe)**:
    - *Cashback Offers* and *Festival Season* have **no significant effect on market share** ($p > 0.05$).
    - Hence, **cashbacks should be limited**, and **festival campaigns** should focus on **service awareness or bundled offers**, not just discounts.

## 2. Targeted Cashbacks:

*Limit to high-impact scenarios (MAU > 80M + festival months)*

**Supported by:**

- **Chi-Square Test (both platforms):** Statistically significant association between cashback offers and festival months.
- **Regression (PhonePe):** MAU showed low predictive power ($p > 0.05$), meaning cashback should not be offered uniformly but only during high-MAU + festival spikes.

## 3. Boost Rural Adoption:

*Increase merchant onboarding in rural areas (where count < 5M)*

**Supported by:**

- **Regression (GPay):** Merchant Count was not statistically significant ($p = 0.768$), suggesting opportunity for optimization—especially in underpenetrated rural areas.
- **EDA (Descriptive Observations):** PhonePe already shows a stronger rural split (up to 42%) vs. GPay (~35%) — GPay can grow here.

## 4. Competitive Tactics

### Google Pay:

*Leverage higher urban transaction values*

**Supported by:**

- **Descriptive Stats + Regression (GPay):** Average transaction values were consistently higher for GPay.
- **Urban split dominance** in GPay (65–66%) can be used to push high-ticket services (e.g., bill payments, investments).

### PhonePe:

*Reduce churn in low-activity user segments*

**Supported by:**

- **T-Test (Transaction Value Comparison):** PhonePe's mean transaction value is significantly higher (₹5451B vs. ₹3146B, $p < 0.01$), but variability is also higher.
- **Regression Output (PhonePe):** Success Rate and MAU have limited predictive strength — implies performance could be improved by **re-engaging churn-prone or low-usage segments**.

---

THE END