upi_transactions_2024_rcode

UTKARSH NARULA

2025-07-09

Ask

This project analyzes UPI transactions to answer:

- **When do people transact most?** (hour, day)
- **Which merchant categories see the highest payments?**
- **Which device types have the highest fraud rates?**
- **Which age group does the most UPI transactions?**

The goal is to find actionable insights for marketing and fraud prevention.

Prepare

We load the dataset and check structure & summary stats.

```r
library(dplyr)
library(ggplot2)
library(readr)
```

```r
upi <- read_csv("upi_transactions_2024.csv") %>%
    rename( amount = `amount (INR)`)
```

```
## Rows: 250000 Columns: 17
## -- Column specification -------------------------------------------------
## Delimiter: ","
## chr  (12): transaction id, transaction type, merchant_category, transaction_...
## dbl   (4): amount (INR), fraud_flag, hour_of_day, is_weekend
## dttm  (1): timestamp
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Process

Filter to keep only successful transactions with positive amounts.

```r
upi_clean <- upi %>%
filter(transaction_status=="SUCCESS")
```

Analyze

1. Transactions by hour of day

```r
hour_summary <- upi_clean %>%
  group_by(hour_of_day) %>%
  summarise(total_transactions = n(),
            avg_amount = mean(amount))

ggplot(hour_summary, aes(x = hour_of_day, y = total_transactions)) +
  geom_line(color = "steelblue", size = 1.2) +
  geom_point(color = "darkblue", size = 2) +
  theme_minimal() +
  labs(
    title = "Transactions by Hour of Day",
    subtitle = "Line plot to show hourly transaction trends",
    x = "Hour of Day",
    y = "Total Transactions"
  )
```
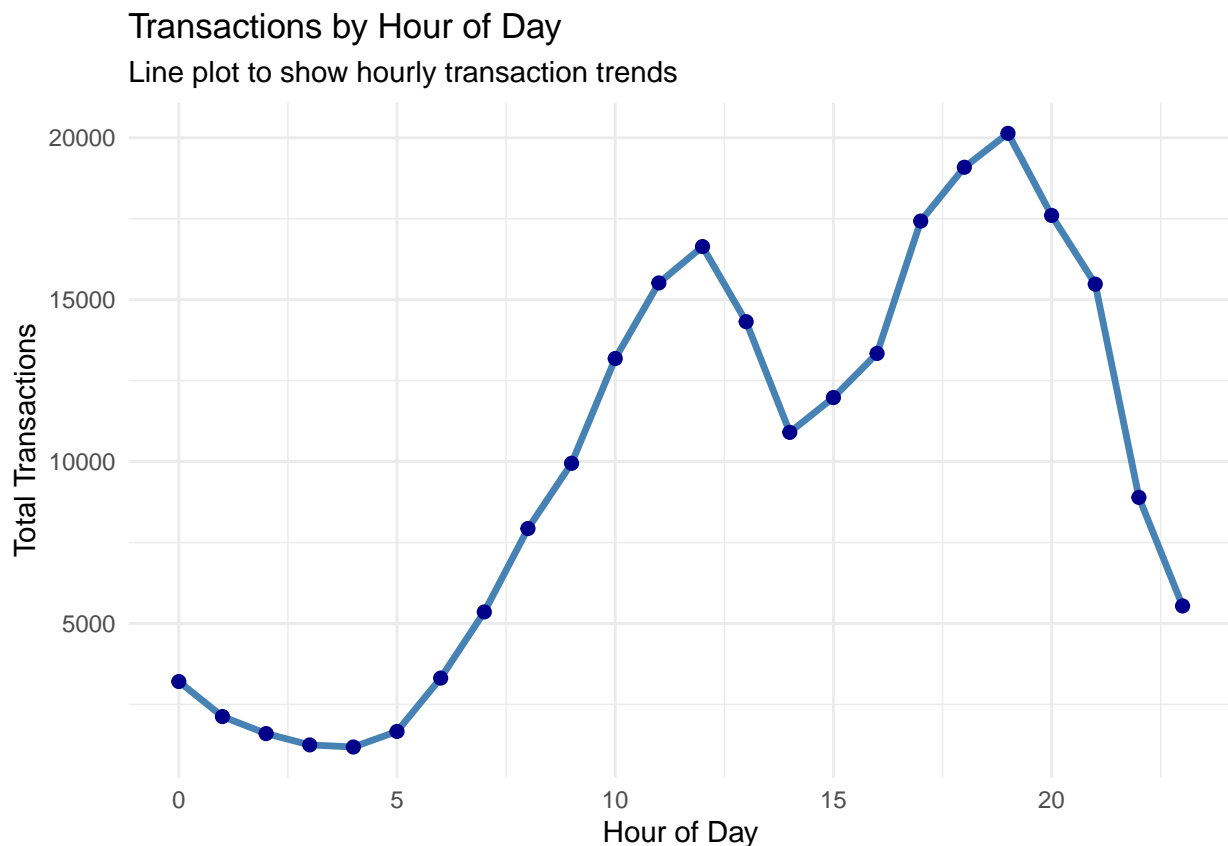

2. Merchant category totals

```r
merchant_summary <- upi_clean %>%
  group_by(merchant_category) %>%
  summarise(total_transactions = n(),
            total_amount = sum(amount)) %>%
  arrange(desc(total_amount))
```

```r
ggplot(merchant_summary, aes(x = reorder(merchant_category, -total_amount), y = total_amount)) +
  geom_col(fill = "darkgreen") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(title = "Total Payments by Merchant Category",
       x = "Merchant Category",
       y = "Total Amount (INR)")
```

Total Payments by Merchant Category
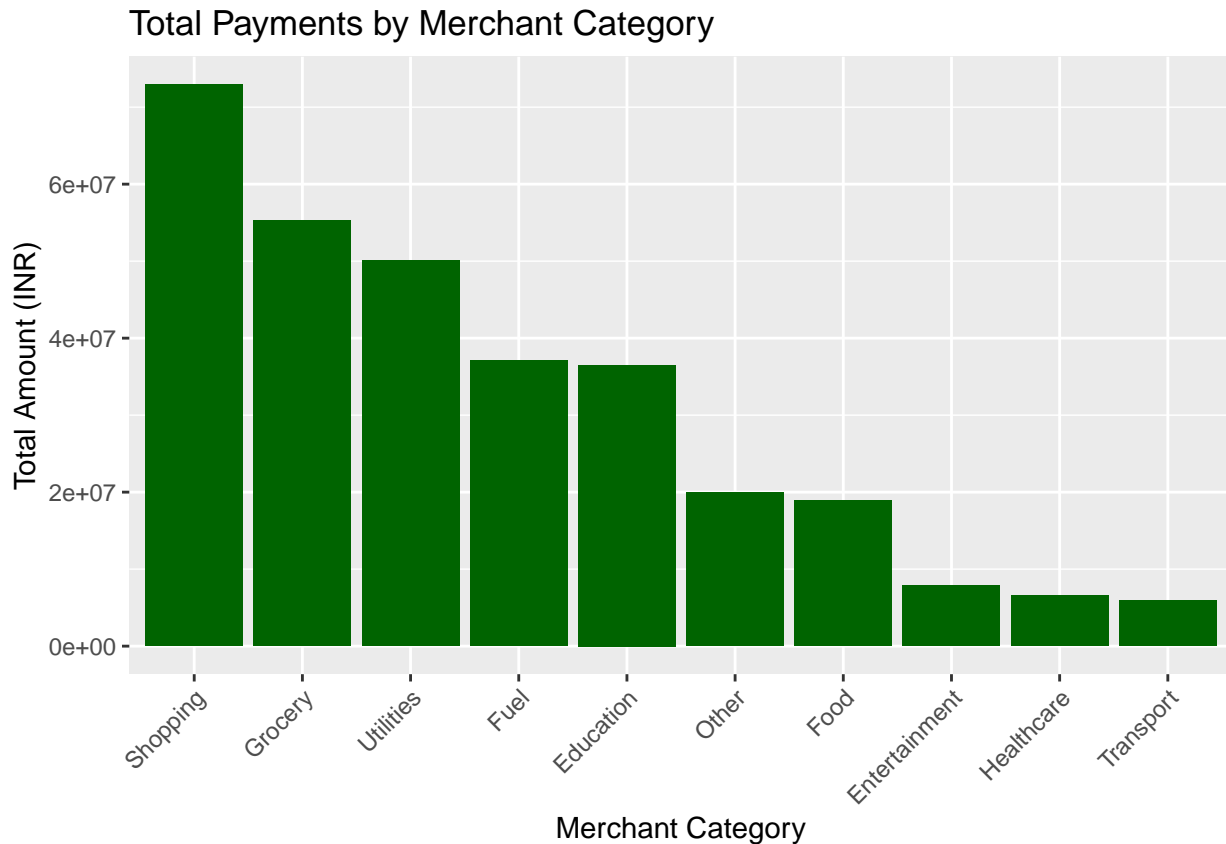
3. Fraud % by device type

```r
fraud_summary <- upi_clean %>%
  group_by(device_type) %>%
  summarise(total_txns = n(),
            fraud_txns = sum(fraud_flag == 1),
            fraud_rate = round((fraud_txns / total_txns) * 100, 2))

ggplot(fraud_summary, aes(x = 2, y = fraud_rate, fill = device_type)) +
  geom_bar(stat = "identity", width = 1, color = "white") +
  coord_polar(theta = "y") +
  xlim(0.5, 2.5) +
  geom_text(aes(label = paste0(fraud_rate, "%")),
            position = position_stack(vjust = 0.5)) +
  theme_void() +
  labs(title = "Fraud Rate by Device Type") +
  theme(legend.title = element_blank())
```

Fraud Rate by Device Type
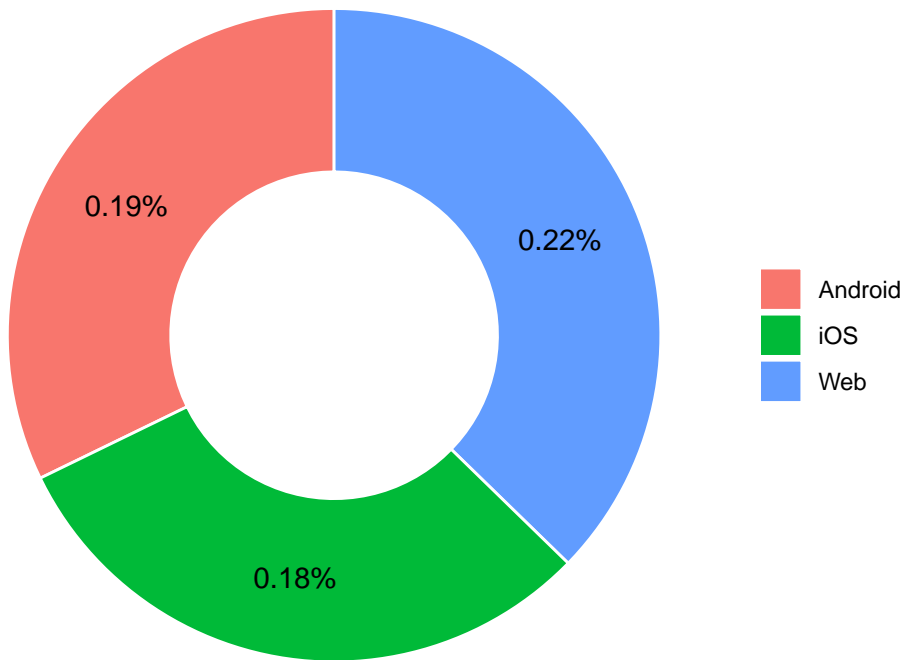
4. Transactions by sender age group

```r
age_summary <- upi_clean %>%
  group_by(sender_age_group) %>%
  summarise(total_transactions = n())

ggplot(age_summary, aes(x = reorder(sender_age_group, -total_transactions), y = total_transactions)) +
  geom_col(fill = "purple") +
  labs(title = "Transactions by Sender Age Group",
       x = "Age Group",
       y = "Total Transactions")
```

Transactions by Sender Age Group
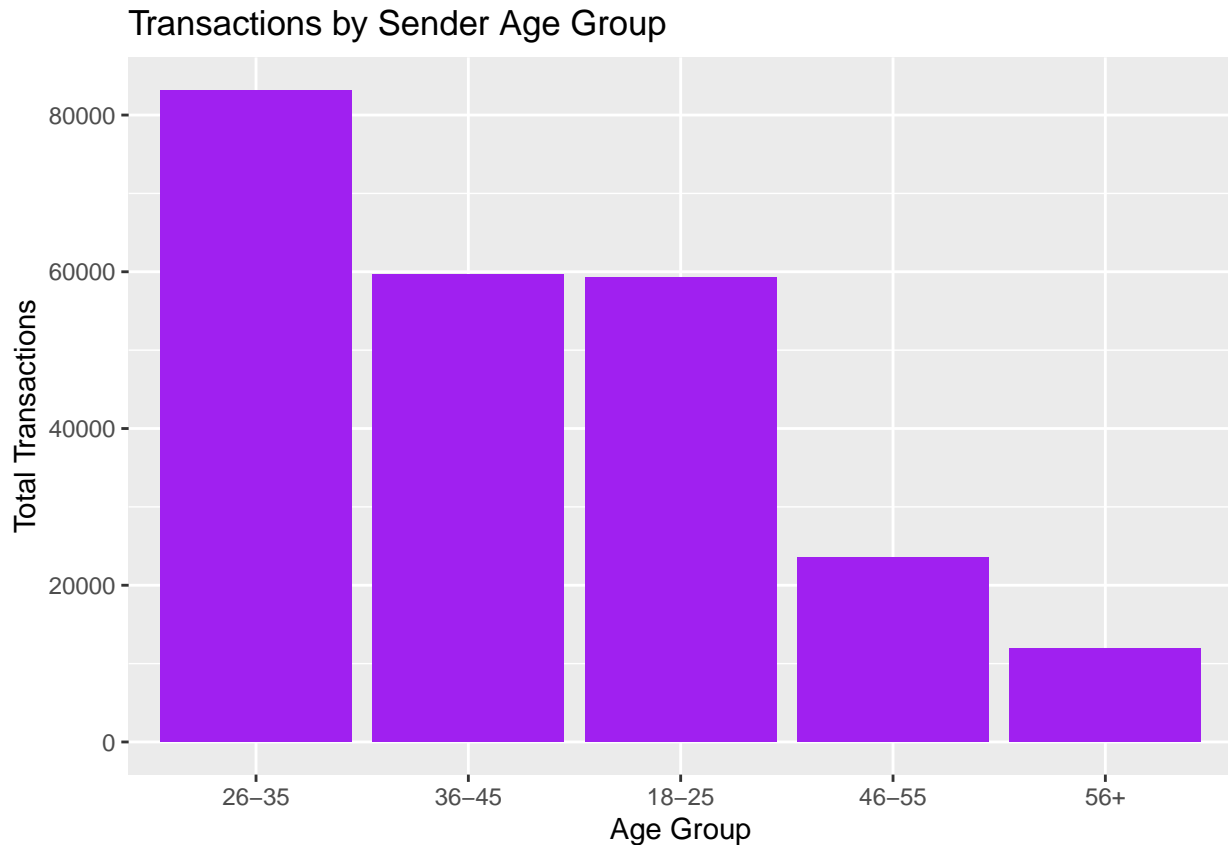

5. Transactions on weekdays vs weekends

```
week_summary <- upi_clean %>%
  group_by(is_weekend) %>%
  summarise(total_transactions = n(),
            avg_amount = mean(amount))


week_summary$is_weekend <- ifelse(week_summary$is_weekend == 1, "Weekend", "Weekday")

ggplot(week_summary, aes(x = is_weekend, y = total_transactions, fill = is_weekend)) +
  geom_col() +
  labs(title = "Transactions: Weekdays vs Weekends",
       x = "",
       y = "Total Transactions") +
  theme_minimal()
```

Transactions: Weekdays vs Weekends
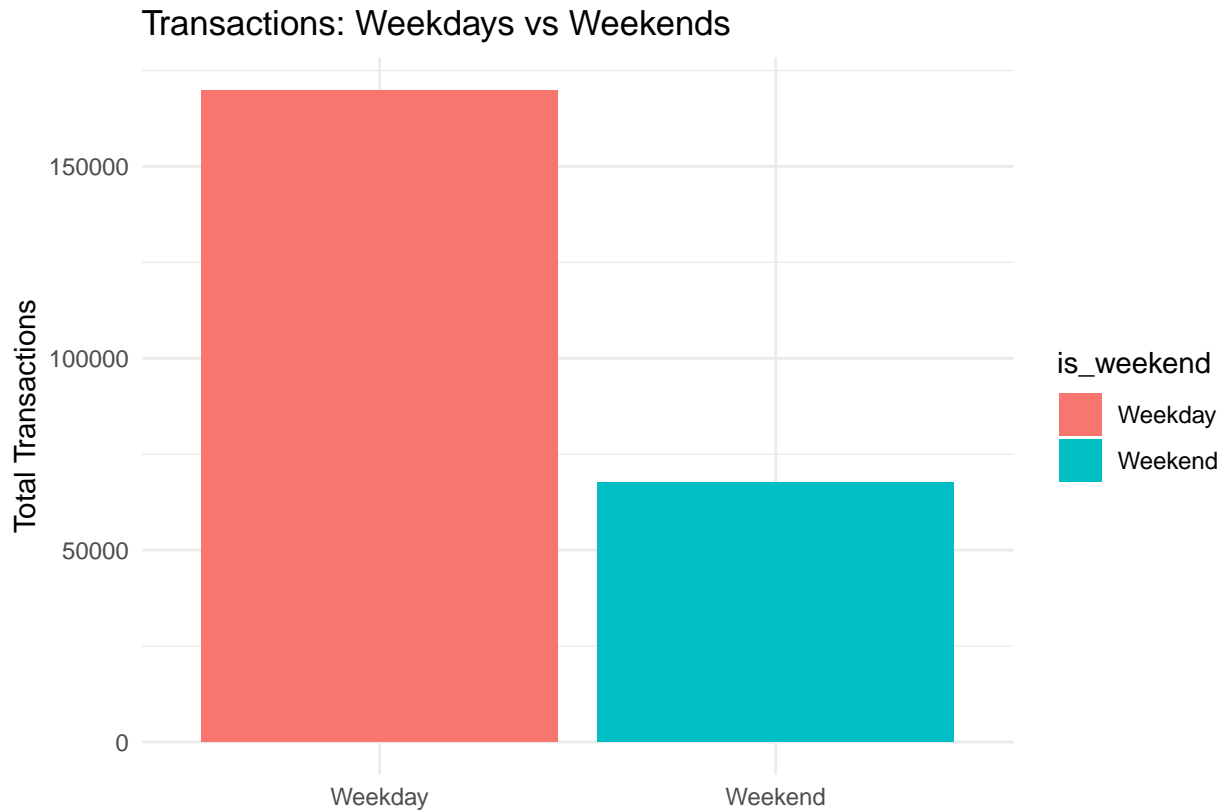
Conclusions

- **Transaction Timing:** Most UPI transactions occur between **12PM and 8PM**, with a peak around early evenings. Suggests users transact after work hours or during breaks.

- **Merchant Patterns: Shopping** and **Grocery** dominate both transaction counts and values. Education, while fewer in transactions, has high total amounts — indicating large ticket payments.

- **Fraud Trends:** Overall fraud rates are low (~0.2%), but slightly higher on **Web transactions**, signaling a need for stronger browser session controls.

- **Customer Segments:** The **26-35 age group** is the most active, followed by 18-25 and 36-45. Usage drops sharply after 45.

- **Weekdays vs Weekends:** Majority (~71%) of transactions happen on **weekdays**, with similar average amounts across all days.

Business recommendations:

- Focus cashback & loyalty programs on **26-35 year olds** and on **Shopping/Grocery categories** to maximize impact.

- Enhance fraud monitoring for **Web transactions** despite their lower volume.

- Launch special weekend campaigns to drive usage during lower-volume periods.

This analysis helps prioritize operational readiness during peak hours and informs targeted marketing to drive growth while managing risk.