Paytm End-to-End Data Analysis Project Report

1. Executive Summary

This project presents an end-to-end data analytics case study on Paytm customer behaviour, simulating a real-world business problem faced by a digital payments company. Using Python, SQL, and Power BI, the analysis converts raw transactional data into actionable business insights that support strategic decision-making, customer experience improvement, and revenue optimization.

---

2. Business Problem Statement

Paytm offers multiple digital financial services such as Recharge & Bills, Money Transfer, Loans, and Insurance. Despite a large user base, the company faces challenges in understanding customer behaviour, transaction failures, service adoption patterns, and revenue contribution across services.

Objective:

To analysed customer behaviour across Paytm services and identify key insights related to usage patterns, transaction failures, customer segmentation, and opportunities to improve engagement, retention, and operational efficiency.

---

3. Dataset Overview

The analysis is based on multiple structured datasets representing Paytm services:

- **All Users**: Customer demographics, activity status, and engagement indicators
- **Recharge & Bills**: Recharge transactions, failure reasons, and bill categories
- **Money Transfer**: Transfer amounts, success/failure status, and channels
- **Loans**: Loan types, approval status, and customer eligibility
- **Insurance**: Policy types, purchase behaviour, and premium trends

Each dataset was analysed individually and later connected for cross-functional insights.

---

4. Tools & Technologies Used

- **Python**: Pandas, NumPy, Matplotlib, Seaborn (Data Cleaning & EDA)
- **SQL (MySQL)**: Data modeling, joins, aggregation, and business queries
- **Power BI**: Interactive dashboards and KPI visualization
- **Jupyter Notebook**: Analysis workflow
- **GitHub**: Version control and project hosting

---

5. Data Cleaning & Preparation (Python)

Key data preparation steps included:

- Handling missing and null values
- Dropping irrelevant columns (e.g., rating fields not used for analysis)
- Standardizing categorical values
- Creating derived metrics such as transaction success rate
- Ensuring consistent data types across datasets

This step ensured reliable and analysis-ready data.

```
all_users.head(10)
```

	User_ID	Name	Age	Join_Date	Gender	Profile_Age	Country
0	PTM0000001	Holly Rivera	56	2025-06-17	Male	Kid	USA
1	PTM0000002	Kevin Lopez	46	2023-12-10	Male	Kid	Germany
2	PTM0000003	Douglas Roberts	32	2024-09-14	Male	Adult	Australia
3	PTM0000004	Walter Davila	60	2023-10-19	Male	Adult	USA
4	PTM0000005	Grace Blake	25	2025-03-10	Male	Adult	Brazil
5	PTM0000006	Nicholas Adkins	38	2024-09-08	Male	Teen	Brazil
6	PTM0000007	James Campbell	56	2023-09-10	Male	Teen	Brazil
7	PTM0000008	Melissa Hall	36	2023-12-26	Male	Adult	Canada
8	PTM0000009	Robert Williamson	40	2025-02-11	Male	Adult	Germany
9	PTM0000010	James Golden	28	2025-01-23	Male	Adult	India

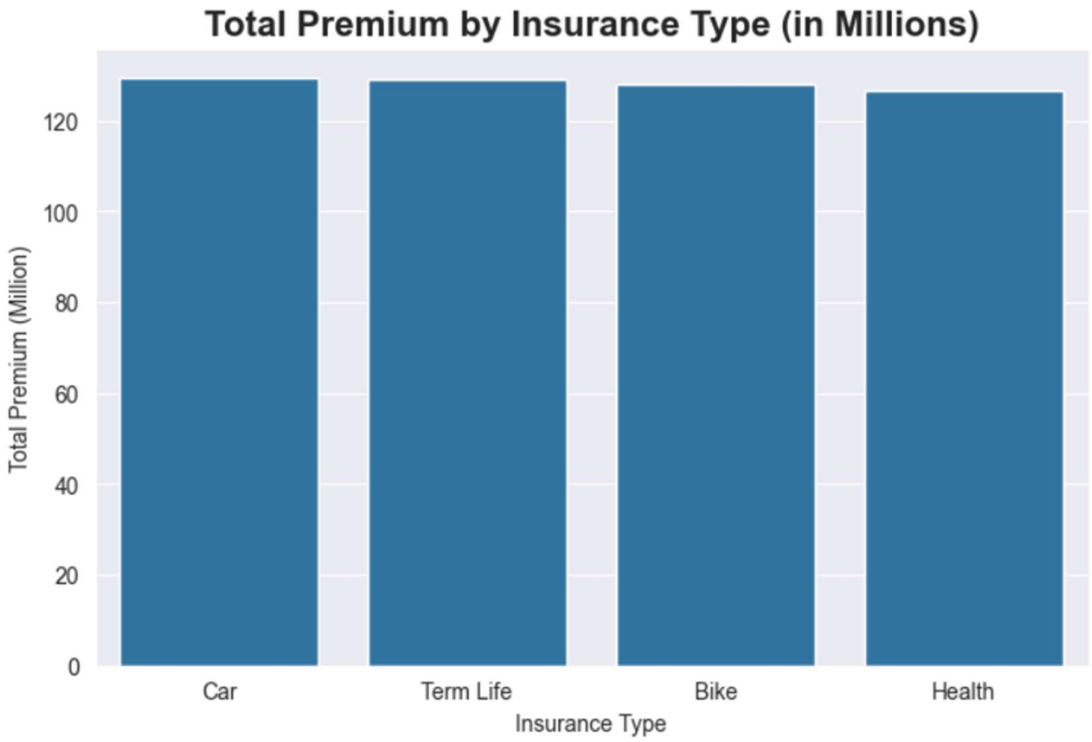
---

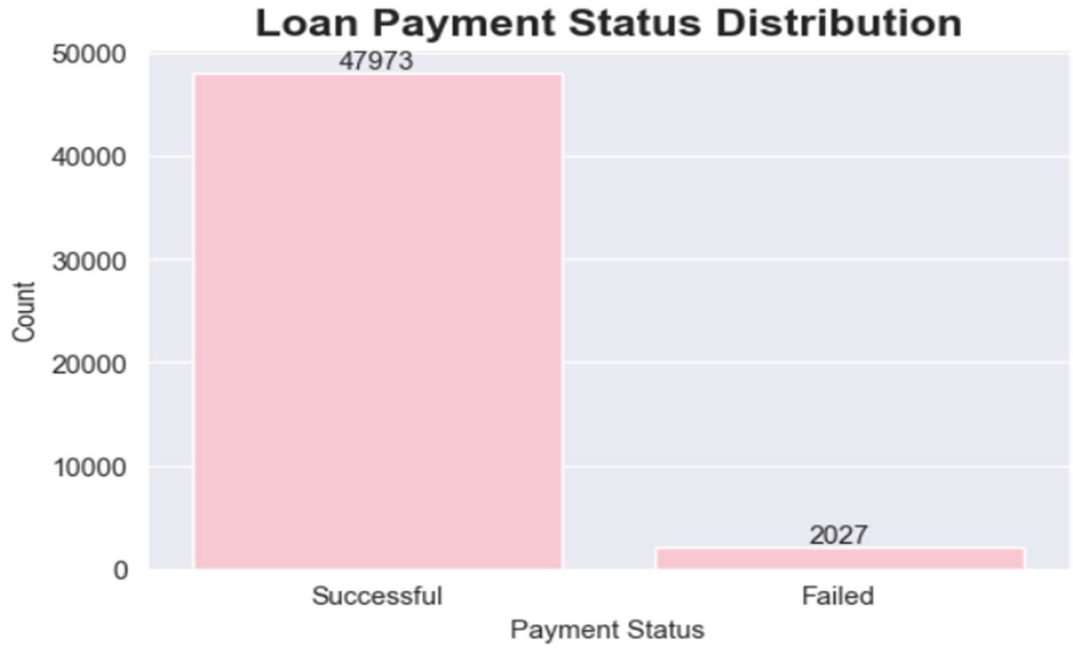
6. Exploratory Data Analysis (EDA)

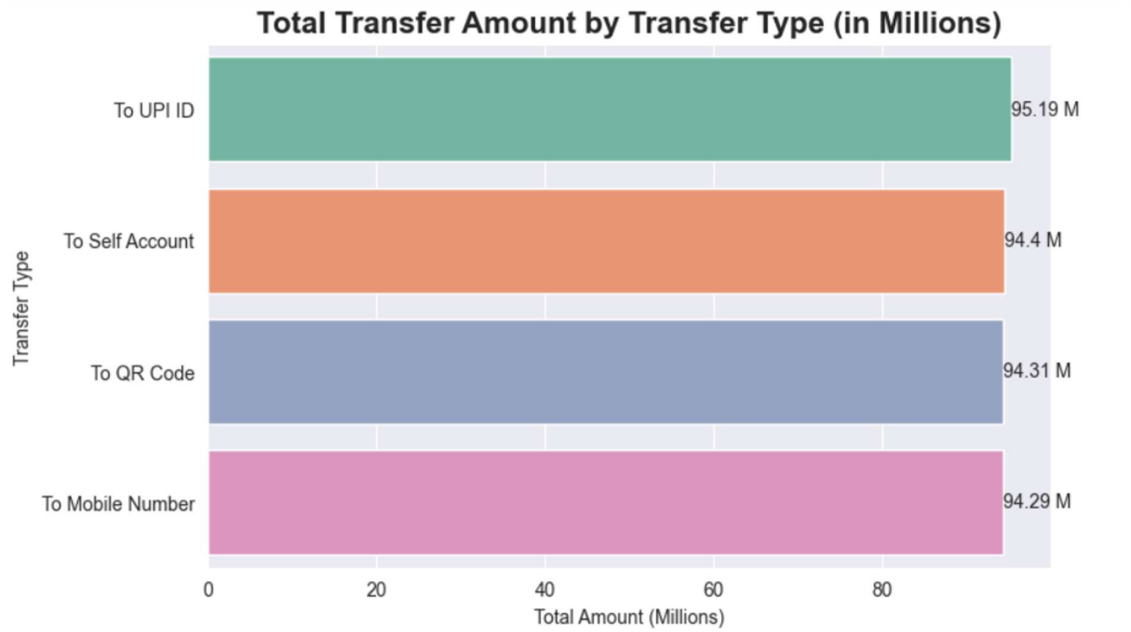
EDA was conducted to understand trends and distributions:

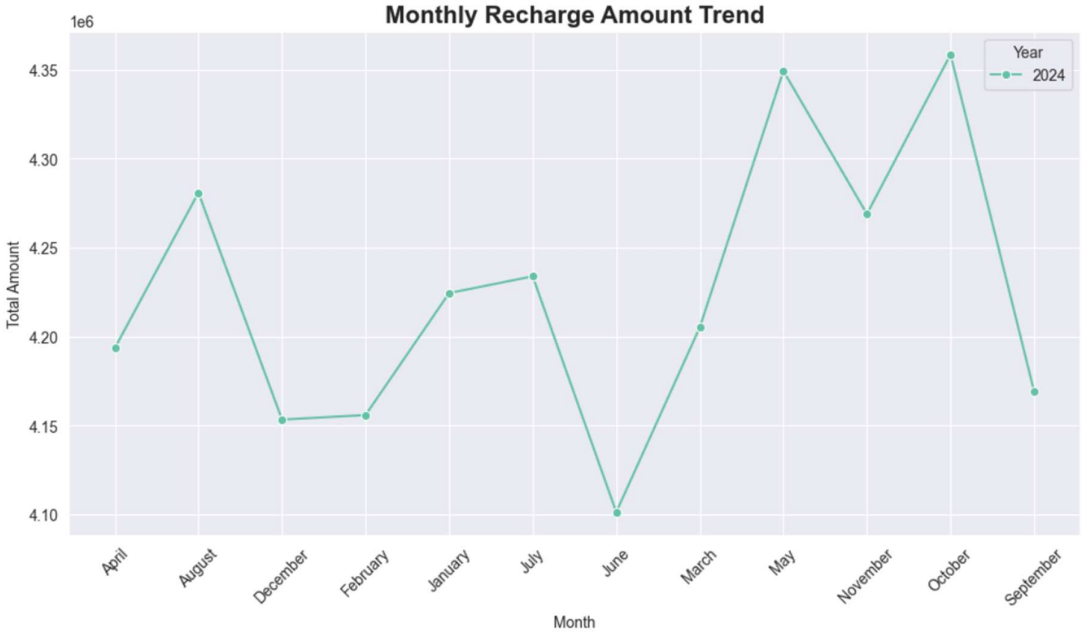
- Transaction volume trends across services
- Failure reason analysis for Recharge & Bills
- Customer activity segmentation
- Service-wise usage comparison
- Identification of high- and low-performing categories

Multiple visualizations such as bar charts, line charts, and donut charts were created to support findings.









---

7. SQL-Based Business Analysis

Data was loaded into a MySQL database with separate tables for each service. Business-focused SQL queries were written to:

- Identify top-performing services by transaction volume
- Analysis customer retention and repeat usage
- Detect major causes of transaction failures
- Segment customers based on service adoption

- Measure operational performance indicators

These queries simulate real-world analyst tasks in a production database environment.

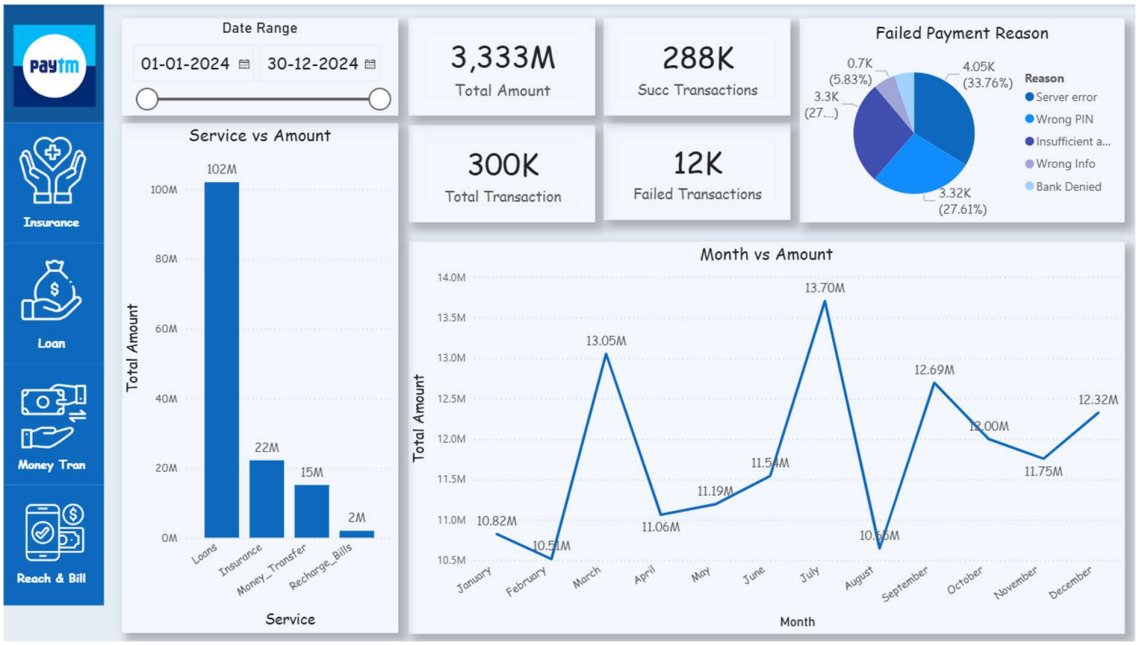
---

8. Power BI Dashboard & Visualization

An interactive Power BI dashboard was developed to present insights to stakeholders:

- Service-wise transaction distribution
- Failure reason breakdown
- Customer engagement overview
- Key KPIs and trend analysis

The dashboard enables non-technical stakeholders to explore insights and make data-driven decisions.



---

9. Key Business Insights

- Recharge & Bills contribute the highest transaction volume but also show significant failure rates.
- Certain failure reasons account for the majority of failed transactions, indicating operational improvement opportunities.
- A subset of customers actively uses multiple Paytm services, representing high-value users.
- Money Transfer shows stable performance but potential for increased adoption.
- Cross-selling opportunities exist between Recharge users and Financial services (Loans & Insurance).

---

10. Business Recommendations

Based on the analysis, the following actions are recommended:

- Improve payment gateway reliability to reduce recharge failures
- Introduce targeted loyalty programs for high-frequency users
- Promote cross-selling campaigns between Recharge and Financial services
- Optimize UX flows for failed transaction recovery
- Focus marketing efforts on high-performing customer segments

---

11. Conclusion

This project demonstrates a complete, industry-aligned data analytics workflow, from raw data to strategic insights. It highlights how Python, SQL, and Power BI can be effectively combined to solve real business problems in the fintech domain.

The project serves as a strong portfolio example for Data Analyst roles and reflects practical, job-ready analytical skills.

- GitHub Repository: *Provided in resume / portfolio*
- Power BI Dashboard: Shared via Google Drive link
- SQL Queries: Included as .sql files in the repository

End of Report