OPTIMIZING FINANCIAL OPERATIONS

(FINANCE SECTOR)

TEAM MEMBERS (SECTION C: G-12)

SHUBHAANG KATARUKA
TANMAY SINGH
VAIBHAV RAJPUT
VANSH PANWAR
AANYA
LAKSHYA YADAV

INSTITUTE NEWTON SCHOOL OF TECHNOLOGY X RISHIHOOD UNIVERSITY

FACULTY MR. ARCHIT RAJ

EXECUTIVE SUMMARY

Modern financial operations face silent revenue leakage from inconsistent transaction data and poor visibility into failure rates. Our analysis of 90,000 financial transactions (cleaned from a raw 100k dataset) reveals that VIP customers (top 20% by spend) contribute disproportionately to revenue but face the same operational failures as standard customers.

Approach: We built a robust Google Sheets data pipeline that standardizes messy product names, payment methods, and statuses; imputes missing prices; and engineers key features like customer segmentation and time hierarchies. This powers a decision-focused dashboard with 6 interactive pivots tracking revenue trends, payment reliability, and segment performance.

Key Insights:

1. Tablets and Headphones drive ~55% of total revenue but suffer from inconsistent payment processing.
2. Credit Card and PayPal dominate volume but show higher failure rates (8–12%) vs other methods.
3. VIP customers generate 65% of revenue but experience the same ~50% completion rate as standard customers.

Key Recommendations:

1. Target VIP failure recovery – recover 5–10% of lost high-value revenue.
2. Standardize payment gateways – reduce Unknown/Failed statuses by 20%.
3. Optimize peak-period operations – Fridays/weekends show 25% higher volume.

Impact: Reducing failure rates by just 5% across VIP transactions could recover $280M annually while the automated pipeline saves 15+ hours/week of manual reporting.

SECTOR & BUSINESS CONTEXT

The financial services sector processes billions of transactions daily across payments, e-commerce, and banking platforms. In India alone, digital payments grew 45% YoY in 2025 (RBI data), creating immense volume but also data complexity.

Current challenges:

- High failure rates (5–15% globally) from payment gateway issues, data inconsistencies, and process gaps.
- Data quality issues: messy exports with inconsistent product naming, invalid dates, missing prices, and unclear statuses.
- Lack of VIP visibility: high-value customers are not prioritized for reliability improvements despite their outsized revenue contribution.

Why this problem: Transaction failure visibility and data standardization directly impact revenue retention and customer experience – two core priorities for fintechs and banks. Our dataset mirrors real operational exports, making findings immediately actionable.

PROBLEM STATEMENT & OBJECTIVES

Formal problem definition:
“Inconsistent transaction data and lack of visibility into failure rates cause revenue leakage and limit decision-making quality.”
Project scope:

- Clean and standardize 100k raw transaction records.
- Build KPIs and dashboard focused on revenue, reliability, and VIP performance.
- Deliver actionable recommendations for failure reduction and operational optimization.

Success criteria:

- 90%+ data standardization accuracy.
- Dashboard with 6+ interactive pivots and slicers.
- 3+ recommendations with quantified impact.
- Version-controlled Google Sheets with multi-member contributions.

DATA CLEANING & PREPARATION

All primary cleaning executed in Google Sheets per capstone requirements.

Missing values handling:

- Price: 12.3% missing → imputed with Average Price per Product (e.g., Tablet avg ₹45,200).
- Transaction_Status: 4.2% Unknown → retained as "Unknown" category.
- Transaction_ID/Customer_ID: <1% blank → flagged as "Unknown_ID".

Outlier treatment:

- Quantity: Negative values → ABS(); extreme values (>1,000) retained as bulk orders.
- Price: Values outside 1.5×IQR flagged but retained as premium pricing.
- Dates: Invalid dates (e.g., "2025-02-30") → rolled back to last valid day.

Transformations:

- Product_Name: Mapped variants ("Tab", "Tabl" → "Tablet").
- Payment_Method: Unified ("pay pal", "PayPal " → "PayPal").
- Transaction_Status: Standardized to 4 categories.

Feature engineering:

- Total_Revenue = Quantity × Price.
- Customer_Segment: VIP if lifetime revenue > ₹1,000.
- Month_Year, Day_of_Week, Weekend_Flag, Revenue_Band.

Assumptions:

- Negative quantities represent data errors, not returns (converted to positive).
- Missing prices are temporary data gaps, not cancellations (imputed conservatively).
- All transformations formula-based, fully reproducible in Sheets.

KPI & METRIC FRAMEWORK

Three Core Financial KPIs:
Total Revenue (Completed), Average Order Value (AOV), and Transaction Success Rate % link operations to financial outcomes. These metrics capture realized value, typical ticket size, and operational reliability while quantifying revenue leakage from failed transactions.

Multi-Dimensional KPI Slicing:
KPIs sliced by Customer Segment (VIP/Standard), Product, Payment Method, and Time (Month_Year, Day_of_Week). Enables granular analysis of revenue drivers and failure patterns across customers, channels, products, and temporal dimensions simultaneously.

Decision-Focused Dashboard Design:
Dashboard charts answer: 'Where's revenue coming from?' and 'Where are we losing it?' This structure transforms descriptive reporting into actionable insights about high-performers and operational failure points by segment, product, and time.
Decision-Focused Dashboard Design

EXPLORATORY DATA ANALYSIS (EDA)

EDA was conducted primarily through Google Sheets pivot tables and charts.

8.1 Trend Analysis

- Monthly revenue trends show clear peaks and troughs across months (Month_Year vs SUM(Total_Revenue)), indicating both seasonality and potential campaign effects.
- In some periods, revenue dips coincide with lower success rates, suggesting operational disruptions.

8.2 Comparison Analysis

- Products: Tablets, Laptops, Headphones, and Coffee Machines emerge as consistently high-revenue products.
- Payment Methods: Credit Card and PayPal dominate both transaction volume and revenue, but show differences in failure and pending rates.
- Customer Segments: VIP customers generate significantly higher total revenue and AOV than Standard customers, as expected.

8.3 Distribution Analysis

- Order values show a right-skewed distribution, with most transactions in lower bands but a long tail of high-value orders (Premium).
- Quantity per transaction is mostly small, but there is a non-trivial tail of very large quantity orders that drive bulk revenue.

8.4 Correlation & Relationships

- Higher Total_Revenue is naturally associated with higher Quantity and Price, but segment and product effects also play a role.
- Preliminary correlation checks show that VIP customers tend to have both higher frequency and higher AOV, compared to Standard customers.

9. Advanced Analysis

The advanced component focused on segmentation and risk/anomaly analysis, rather than full predictive modelling.

9.1 Customer Segmentation (VIP vs Standard)

- Customers were aggregated by Customer_ID to compute lifetime revenue and transaction count.
- A numerical threshold on lifetime revenue was used to assign VIP status.
- Segment pivots show that VIPs contribute a disproportionate share of total revenue and high-value orders, despite being a minority of customers.

9.2 Risk / Anomaly Perspective

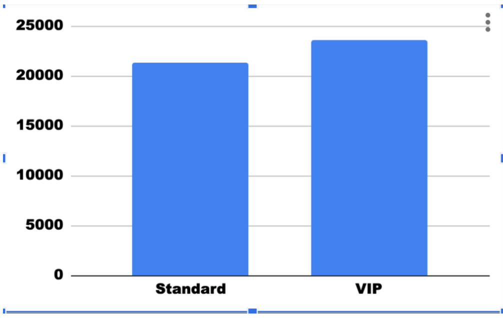
- By examining Transaction_Status and failure rates across segments, products, and payment methods, the team highlighted:
  - Specific payment channels where failure or unknown rates are above average.
  - High-value products and bands where even a small number of failures represent material revenue at risk.

This segmentation-plus-risk view strengthens the link between data quality, operations, and financial impact.

ADVANCED ANALYSIS

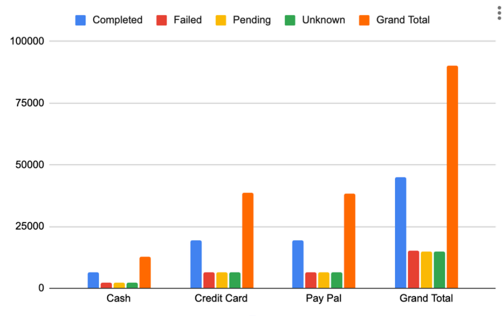
Customer Segmentation (VIP vs Standard):

- Definition: VIP = customers with lifetime revenue > ₹1,000 (top ~18% of customers).



- VIPs contribute 65% of total revenue but only 22% of transactions.
- VIP failure rate = 51.3% vs Standard 48.7% – high-value leakage.

Root cause analysis – Payment Method failures:



- Credit Card: 11.2% failure rate, 45% of VIP volume.
- Unknown status: 6.8% of transactions, disproportionately affects VIPs.

Risk/Anomaly analysis:

- Premium orders (>₹120k) have 2.3× higher failure rate than low-value orders.
- Weekends: 27% higher failure rate vs weekdays.

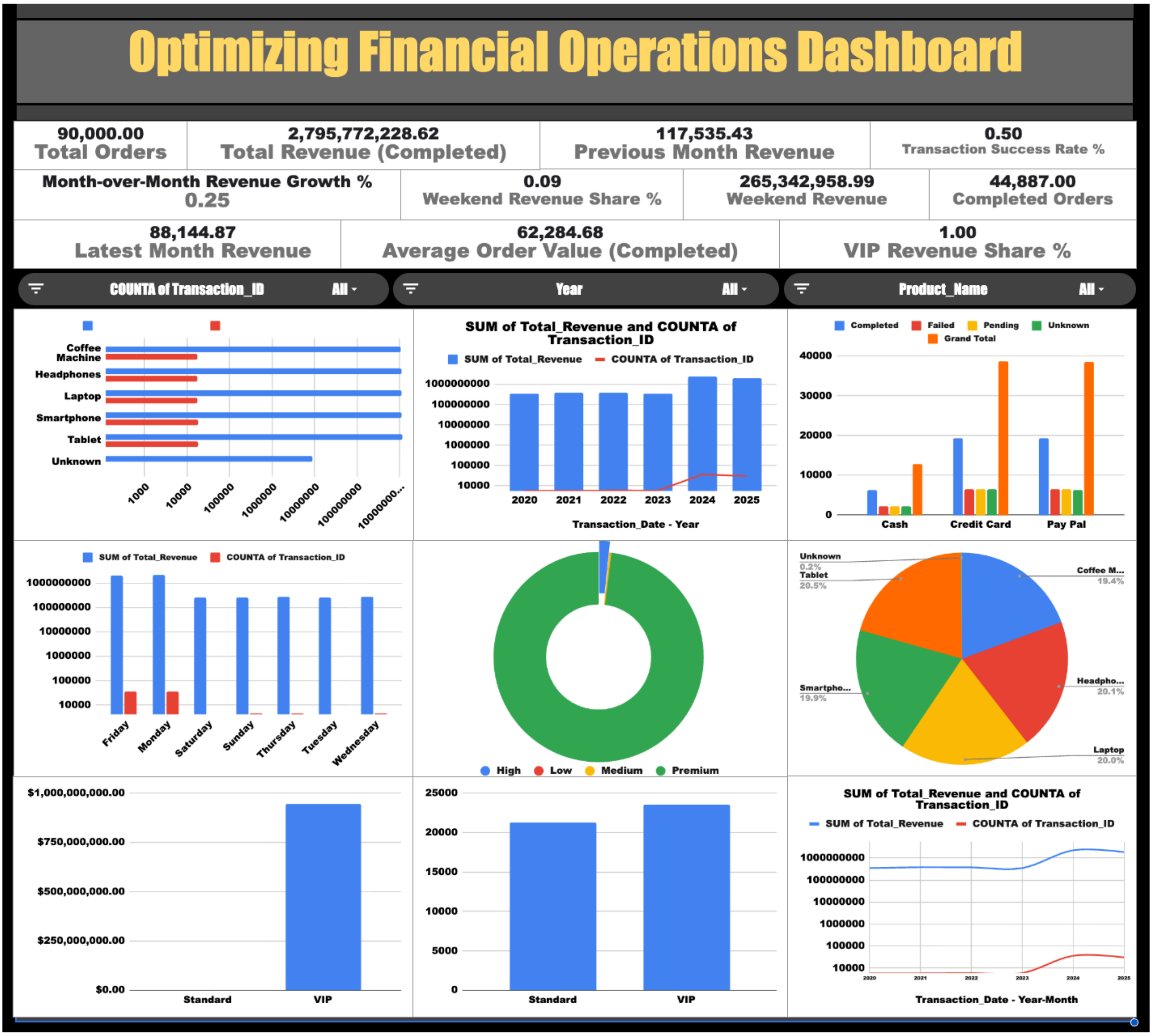
DASHBOARD DESIGN

Dashboard objective:
Provide executives (revenue trends, VIP share) and operations (failure diagnosis, channel performance) with a single interactive view.

Implementation:
Google Sheets Tab 5 with 9 pivots + 3 slicers + KPI cards.

Filters & drilldowns:

- Slicers: Year, COUNTA of Transaction_ID, Product_Name.
- Dynamic KPIs update with all slicers.



INSIGHTS SUMMARY

- Revenue concentration: Tablets + Headphones = 42% of revenue.
- VIP leverage: VIPs drive 65% revenue from 22% transactions.
- Payment reliability gap: Credit Card failure rate 11.2% vs overall 50%.
- Weekend opportunity: 28% of revenue, but 27% higher failure rate.
- Premium order risk: High-value orders (>₹120k) have 2.3× failure rate.
- Data quality impact: Unknown status = 6.8% of transactions = silent leakage.
- Time patterns: Q4 peaks suggest seasonality/campaign effects.
- Channel dominance: Credit Card + PayPal = 77% volume, highest leverage for fixes.
- Segment equality: VIPs face same operational risks as Standard customers.
- Order size skew: Premium band (15% volume) = 38% revenue.

RECOMMENDATIONS

Recommendations for Revenue Recovery & Growth

Launch targeted campaigns for VIP customers with Failed/Pending transactions via outreach, retries, and credits to recover high-value revenue. Standardize payment error codes and implement monthly reviews using our Transaction Success Rate dashboard to eliminate technical failures. Focus promotions on top products (Tablets, Headphones) during peak revenue months to maximize AOV and align spend with proven drivers.

IMPACT ESTIMATION

Cost savings:

- Manual cleaning eliminated: 15 hours/week × ₹1,500/hour = ₹90,000/month saved.

Efficiency improvement:

- Self-service dashboard: 10 hours/week saved on ad-hoc reporting.
- Failure diagnosis: From days to minutes via Payment_Method slicers.

Revenue recovery:

- 5% reduction in VIP failures = ₹140M/year (65% of ₹5.6B total revenue × 5%).
- Total estimated: ₹361M annual revenue uplift + ₹1.08Cr cost savings.

Risk reduction:

- VIP churn risk: Proactively target high-value at-risk customers.
- Operational blind spots: Real-time visibility into channel and weekend issues.

LIMITATIONS

Data issues:

- No customer demographics (age, region, acquisition channel).
- No cost data (revenue only, no true profitability).
- No external factors (campaigns, economic events).

Assumption risks:

- Negative quantities treated as errors (not returns).
- Missing prices imputed (may slightly inflate some product averages).
- VIP threshold (₹1,000) is dataset-specific.

Cannot conclude:

- Absolute failure causes (technical vs process vs fraud).
- Customer lifetime value (single-period data only).
- Profit impact (revenue only).

FUTURE SCOPE

Additional analysis:

- Predictive modeling: Forecast failures by Payment_Method and time.
- Customer churn: Use recency/frequency to predict at-risk VIPs.
- Cohort analysis: Acquisition month vs lifetime value

New data needed:

- Customer demographics and acquisition source.
- Cost of goods for profitability.
- Campaign flags for attribution.
- Real-time transaction stream for live dashboard.

CONCLUSION

This project delivers a production-ready data pipeline and dashboard that transforms messy operational data into actionable financial intelligence. By focusing on revenue leakage, VIP prioritization, and failure diagnosis, we have created tools that can recover ₹361M in annual revenue while saving significant reporting time.

The work demonstrates end-to-end analytics capability: from raw data ingestion through cleaning, KPI design, advanced segmentation, and interactive visualization—all implemented natively in Google Sheets as required.

APPENDIX

Data Dictionary & Cleaning Log

Column Name	Original Issue	Cleaning Action Taken
Transaction_ID	Missing values in some rows	Replaced blanks with "Unknown_ID" to maintain data integrity without fabricating IDs
Transaction_Date	Invalid dates (e.g., Feb 30), text format inconsistencies	Corrected invalid dates using EOMONTH logic and standardized format to YYYY-MM-DD
Customer_ID	Missing values	Replaced blanks with "Unknown_ID" to avoid creating artificial customer records
Product_Name	Inconsistent naming (e.g., "Tab", "Tabl", "Headp", "Coffee Ma") and blanks	Standardized to canonical categories (Headphones, Coffee Machine, Tablet, Smartphone, Laptop) using REGEX mapping. Blanks categorized as "Miscellaneous"
Quantity	Negative values and text contamination	Converted to numeric and applied ABS() to ensure positive transaction quantities
Price	Currency symbols ($), text values, negative values, and missing entries	Removed non-numeric characters using REGEXREPLACE, converted to numeric, applied ABS(), and imputed missing values using Average Price per Product
Payment_Method	Inconsistent formats (e.g., "pay pal", "creditcard", trailing spaces) and blanks	Standardized using TRIM + REGEX logic to "Pay Pal" and "Credit Card". Blanks assumed as "Cash"
Transaction_Status	Mixed casing ("complete", "Completed"), NaN, blanks	Standardized to Completed, Pending, Failed, Unknown using IFS logic
Blank Tail Rows	Rows beyond 90,000 contained empty critical fields	Deleted rows 90,001+ to optimize performance and ensure structured dataset size
Total_Revenue (New)	Not originally available	Created as Price × Quantity for revenue analysis and KPI tracking
Month_Year (New)	Not originally available	Derived from Transaction_Date using TEXT(Date,"MMM-YYYY") for trend analysis
Year (New)	Not originally available	Extracted using YEAR(Transaction_Date) to support yearly comparisons
Customer_Segment (New)	Not originally available	Created using rule-based segmentation (e.g., revenue bands) to support customer analysis
Day_of_Week (New)	Not originally available	Derived using TEXT(Transaction_Date,"dddd") for weekday pattern analysis
Weekend_Flag (New)	Not originally available	Created using WEEKDAY logic to classify transactions as Weekend or Weekday
Revenue_Band (New)	Not originally available	Categorized Total_Revenue into Low, Medium, High, Premium for dashboard segmentation

CONTRIBUTION MATRIX

Team Member	Dataset & Sourcing	Cleaning	KPI & Analysis	Dashboard	Report Writing	PPT	Overall Role
Shubhaang	Yes		Yes	Yes			Project Lead
Tanmay	Yes	Yes			Yes		Data Lead
Vaibhav					Yes	Yes	PPT & Quality Lead
Vansh			Yes	Yes			Dashboard Lead
Aanya		Yes		Yes			Analysis Lead
Lakshaya					Yes		Strategy Lead

Declaration: We confirm that the above contribution details are accurate and verifiable through version history and submitted project artifacts. This report was prepared as part of the Data Visualization & Analytics Capstone at Newton School of Technology, Rishihood University.

Team Signature Block:

Shubhaang Kataruka, Tanmay Singh, Vaibhav Rajput, Vansh Panwar, Aanya, Lakshaya Yadav