Title: Enhancing Operational Efficiency and Revenue Analysis: A Data-Driven Study for DTDC Courier Franchise Services

A Final report for the BDM capstone Project

Submitted by

Name: **Avishi Prasad**

Roll number: **23f2003893**

IITM Online BS Degree Program

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

Contents

1 Executive Summary

This project explores data-driven operational improvements for a DTDC Courier Franchise in Sharda Nagar, Kanpur, Uttar Pradesh. Despite its long-standing presence, the franchise has recently faced challenges such as unclear region-wise revenue trends, increased parcel returns, and significant delivery delays. To address these, one month of transaction data (October 2024) was collected from physical delivery logs and Proof of Delivery (POD) records, resulting in over 225 structured entries in Excel. Fields included parcel weight, delivery status, transaction amount, service type, and region.

Microsoft Excel was used for data cleaning and analysis (COUNTIF, AVERAGEIF, PivotTables), while Power BI was used to visualize metrics like revenue by region, service preference, delivery delays, and return rates. Additionally, the OpenRouteService API was used to estimate travel time and distance for validating delivery delays.

Key findings include: Delhi and Bangalore contributed over 25% of revenue; express and heavy parcels, though fewer, generated disproportionately high revenue; return rate was low at 0.89%, possibly due to manual data entry issues; over one-third of deliveries were delayed, especially to Surat and Delhi. Recommendations include implementing address verification, prioritizing high-revenue parcels, optimizing routes using travel-time data, improving return logging, and adjusting staffing based on peak days. The project highlights how even small franchises can leverage accessible tools for data-driven decision-making.

2 Analysis Process and Methods

2.1 Data Collection

Data for this study was collected over a series of three in-person sessions with the franchise owner. Physical delivery records, primarily printed bills and Proof of Delivery (POD) sheets, were accessed and photographed. Each photograph was transcribed manually into a Microsoft Excel file. While this process was time-consuming, it ensured a high level of familiarity with the data and its structure.

Each record was entered with fields including parcel ID, date of booking, date of delivery, destination region, parcel weight, service type (standard or express), transaction amount, and final delivery status (delivered or returned). Data was collected only for outbound shipments from the Kanpur franchise during the month of October 2024, resulting in 225 usable entries after cleaning.

2.2 Data Cleaning and Preprocessing

Tools Used: Excel functions, filters, and validation tools.

Before proceeding with analysis, the data underwent several preprocessing steps to ensure consistency and reliability. In Excel, inconsistencies in text labels (e.g., “Express” vs “express”) were standardized using LOWER () and PROPER () functions. Duplicate rows, which were sometimes introduced during manual entry, were removed through the “Remove Duplicates” feature.

Weight formats were normalized (e.g., converting 5000g entries to 5kg), and blank entries were examined for context — most were discarded if critical fields were missing. Where feasible, null or incomplete cells were cross-referenced with physical documents. The final cleaned dataset maintained a uniform structure, which allowed for the reliable application of formulas and PivotTables.

In terms of structural integrity, the following fields were prioritized:

- Parcel ID – Unique identifier, used to trace deliveries.
- Region – Delivery destination grouped into standard zone names.
- Service Type – Binary classification as "Standard" or "Express".
- Parcel Weight – Grouped into <5kg and >5kg.
- Transaction Amount – The total revenue per parcel.
- Delivery Status – Marked as Delivered or Returned.

2.3 Method 1: Revenue Trend Analysis

Tool Used: Excel (PivotTables, Bar Charts)

The primary business goal was to understand how different regions contributed to revenue. For this, Excel PivotTables were used to segment the data by destination region and calculate total revenue, number of deliveries, and average transaction values.

Revenue was broken down by:

- **Service Type** (Standard vs Express)
- **Weight Class** (<5kg vs >5kg)
- **Day of Week**

This allowed us to create **multi-dimensional views** of the franchise's earnings and identify zones of underperformance. Charts like bar graphs and donut charts were generated using Excel's chart tools to visually compare revenue across various segments.

2.4 Method 2: Parcel Return Rate Analysis

Tools Used: Excel COUNTIF/COUNTIFS, Pie Charts

Parcel return analysis involved identifying patterns in failed deliveries. Using the COUNTIF function, the percentage of parcels returned out of total transactions were calculated. This metric was further grouped by region and service type to find out where and why returns occurred.

A surprising finding was the extremely low return rate (~0.89%), which led to discussions about data recording accuracy. Since returns are usually recorded manually by field staff, it was hypothesized that the observed rate might be lower than reality due to inconsistent logging practices.

Pie Charts were also used to visualize delivered vs. returned parcels and **highlight return hotspots** (e.g., Noida, Jhajjar).

2.5 Method 3: Delay Pattern and Route Analysis

Tools Used: Excel, OpenRouteService API via Python

Delivery delay analysis was done in two parts:

1. **Internal Analysis:**
    - Compared actual delivery date with expected timelines.
    - Used IF and DATEDIF formulas in Excel to find delayed entries.
    - Grouped delays by region to identify high-delay zones.
2. **External Validation (OpenRouteService API):**

- Wrote a Python script to send origin-destination pairs to OpenRouteService.
- API returned estimated travel durations and distances.
- Converted seconds to minutes and plotted travel time per region.

This method helped **validate delivery delays for distant cities** like Surat (14 hrs) and Delhi (6 hrs). By linking this delay data back to real-world travel metrics, we provided **evidence-based insights** to support route optimization recommendations.

3 Results and Findings

3.1 Region-wise Revenue

The franchise's revenue distribution showed significant disparities across different delivery destinations, highlighting the impact of geography and urbanization on courier demand. Metro cities such as **Delhi**, **Bangalore**, and **Lucknow** consistently emerged as top revenue contributors, indicating a strong correlation between high population density, business activity, and parcel volume. Delhi alone generated approximately ₹8,325, accounting for 15.11% of the total monthly revenue, making it the franchise's most profitable destination. Bangalore followed with ₹5,940 (10.78%), and Lucknow contributed ₹5,490 (9.96%), further emphasizing the dominant role of urban hubs in sustaining courier business revenue.

These trends can be attributed to several underlying factors. Metropolitan regions not only host a greater concentration of e-commerce activity and small-to-medium enterprises but also exhibit higher frequency in parcel dispatch and delivery needs. The presence of time-sensitive shipments and corporate clients in these areas may also explain the higher transaction values per parcel. Additionally, the availability of better infrastructure and more consistent delivery routes in metro areas contributes to smoother operations and greater customer satisfaction, fostering repeat usage of courier services.

In contrast, tier-3 towns and rural destinations such as Jhansi and Hardoi showed minimal contribution, each accounting for less than 1% of total revenue. These regions tend to have lower commercial demand, shorter shipping distances, and reduced shipment volume, resulting in fewer high-value transactions. Given this disparity, a strategic recommendation would be to prioritize marketing, customer engagement, and operational investments in metro and tier-1 cities where the return on effort and resources is demonstrably higher. While rural areas should not be neglected entirely, their limited contribution suggests they are better suited for maintaining baseline service levels rather than growth-focused initiatives. For region-wise revenue heatmap refer to **Appendix Figure 13**.

Visuals:

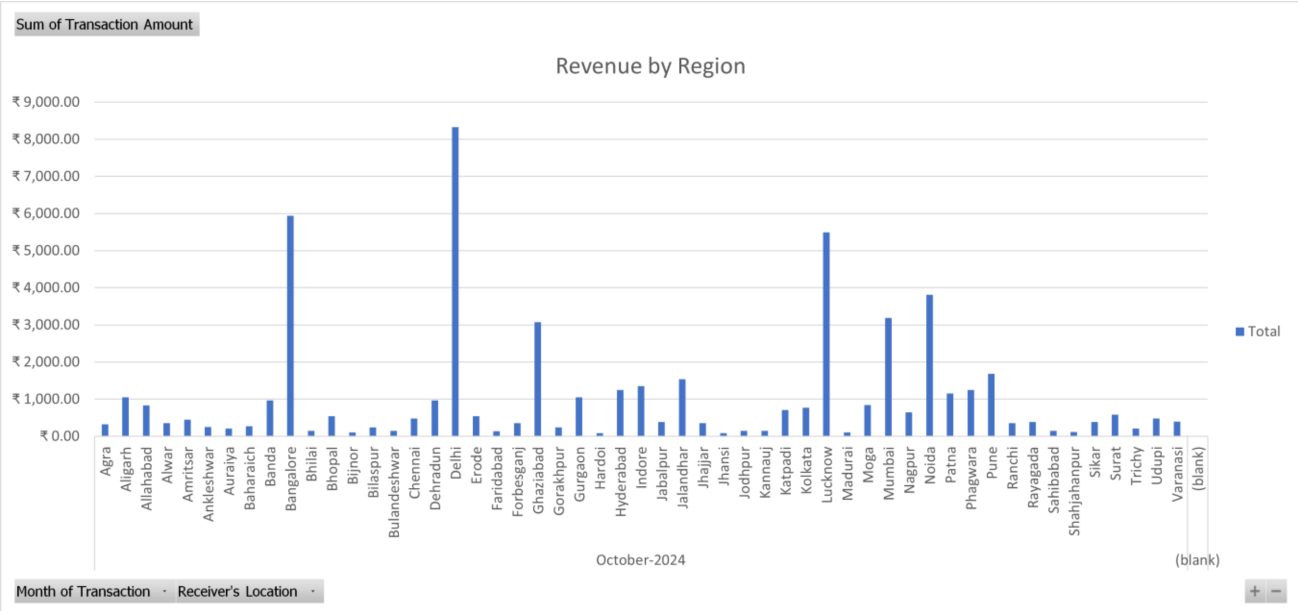

Figure 1: Bar Chart showing Region-wise Revenue Contribution

3.2 Parcel Return Insights

One of the initial concerns expressed by the franchise owner during preliminary discussions was the suspicion of a high parcel return rate, believed to be impacting customer satisfaction and overall operational efficiency. In courier businesses, a high rate of returned shipments often indicates issues such as incorrect address entry, failed delivery attempts, customer unavailability, or dissatisfaction with service timelines. Therefore, identifying and quantifying the scale of returns was considered an important part of this project's objective.

However, the data analysis painted a different picture. Out of the 225 parcels dispatched during the observed period (October 2024), only 2 parcels were marked as returned, resulting in a remarkably low return rate of approximately 0.89%. These two instances were associated with deliveries intended for the Jhajjar and Noida regions. At face value, this statistic might suggest a highly efficient delivery system with minimal errors and strong customer satisfaction. It would indicate that the franchise has robust address verification, efficient last-mile delivery practices, and reliable logistics partners.

Yet, a deeper evaluation raises important concerns about the completeness and reliability of return data. Since the franchise currently uses a manual system for recording parcel returns—often dependent on field staff to update logs accurately—there exists a possibility that not all failed deliveries are systematically captured. In fast-paced operational environments, especially during peak dispatch periods, delivery personnel may overlook or delay documentation of returned parcels, leading to underreporting. Furthermore, there may be ambiguity in defining what constitutes a "return" versus a "re-dispatch" or "delayed delivery," particularly if proper tracking protocols are not enforced.

This potential gap in data integrity suggests that the low return rate might be partially reflective of record-keeping limitations rather than flawless operational performance. It underlines the critical need for improving the accuracy and consistency of return data collection. The franchise could consider digitizing its return logging process, such as integrating barcoded Proof of Delivery (POD) slips or adopting a mobile app for field agents to record unsuccessful delivery attempts in real time. This would ensure a more transparent, accountable, and analysable return tracking mechanism.

Understanding the true nature and causes of parcel returns is essential for designing effective interventions. Whether due to incorrect addresses, unavailability of recipients, or service dissatisfaction, each return represents an opportunity to improve operational workflow and customer interaction. Without accurate data, however, such improvements remain speculative. Therefore, addressing the issue of underreporting is a critical step toward making data-driven decisions and achieving meaningful service optimization

Visuals:

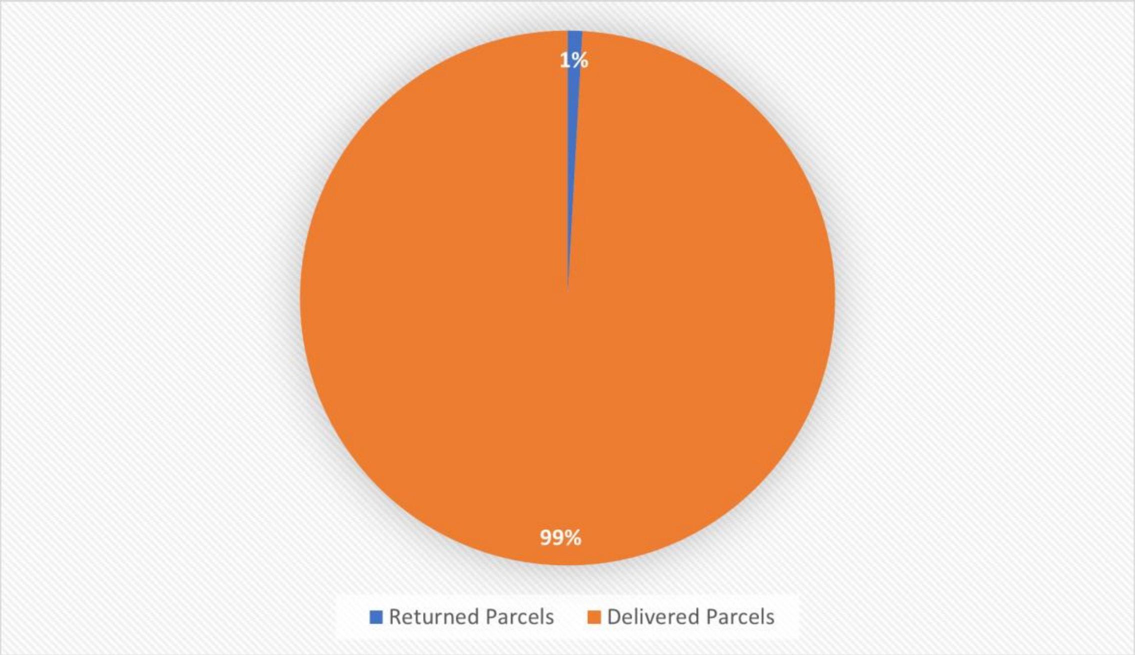

Figure 2: Pie Chart showing Delivered vs Returned Parcels

3.3 Delivery Delays and Travel Time

A critical area of operational concern uncovered during this study was the frequency and distribution of delivery delays. Timely delivery is a core component of customer satisfaction in the courier industry, and persistent delays can undermine service reputation, reduce repeat business, and increase logistical costs. The analysis revealed that approximately **34.53% of all transactions experienced delays**, a figure that signals substantial room for improvement in delivery efficiency.

Among the various regions analysed, **Surat and Delhi emerged as the most delay-prone destinations**, with delay rates of **50% and 63%**, respectively. This raised a key question: were these delays purely the result of inefficient operations, or were they influenced by external and unavoidable factors such as geographic distance and travel time?

To address this, an additional layer of analysis was performed using the **OpenRouteService API** (refer to **Appendix Figure 10** for python code), which enabled the estimation of actual road distances and expected travel durations between the origin point (Kanpur) and each major destination. The findings from the API provided essential context. **Surat**, located nearly **880 kilometres from Kanpur**, has an estimated travel time of **over 14 hours** by road under normal conditions. **Delhi**, while geographically closer at approximately **480 kilometres**, still demands **5.5 to 6 hours of road travel**. These insights suggest that a portion of the observed delays may be attributed not to internal inefficiencies, but to inherent logistical challenges driven by distance, route complexity, and possibly even real-time variables such as traffic or road conditions.

This correlation between delivery delays and geographic constraints **validates the importance of integrating real-world travel data into operational planning**. It also highlights the limitations of traditional expectation-based delivery timelines that do not account for variability in transit durations. Courier operations that fail to consider such factors often set unrealistic delivery windows, resulting in unmet expectations and potential dissatisfaction.

From a strategic perspective, these findings reinforce the **need for more dynamic and region-specific delivery planning**. For long-distance destinations like Surat and Delhi, the franchise could consider prioritizing **express service options**, optimizing **dispatch schedules**, or even **informing customers with realistic estimated delivery times** based on route analytics. Additionally, if demand from high-delay regions remains consistent, partnering with local logistics agents in those areas could reduce last-mile delivery times.

Ultimately, integrating tools like the OpenRouteService API into regular planning routines can offer actionable intelligence. Such data-driven planning not only improves delivery predictability but also enhances the franchise's ability to allocate resources efficiently, manage customer expectations better, and support continuous performance improvement in geographically challenging delivery corridors. For detailed table of parcel delay % per region, refer to **Appendix Figure 14**.

Visuals:

Reciever's Location	Delay Rate
Hyderabad	20%
Delhi	63%
Ghaziabad	29%
Noida	21%
Lucknow	46%
Mumbai	40%
Bangalore	44%
Pune	40%
Allahabad	25%
Dehradun	33%
Kolkata	40%
Gurgaon	17%
Surat	50%

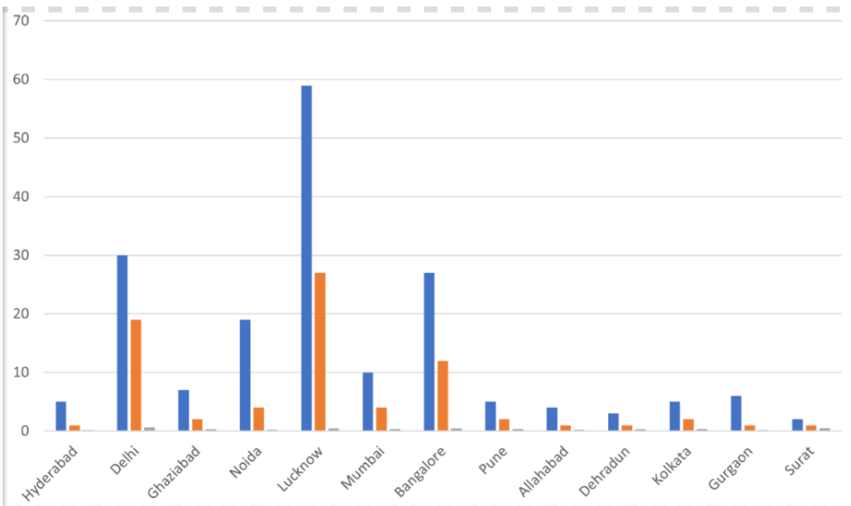

Figure 3: Bar Chart: Percentage of Delayed Parcels by Region

Open_Route_Service_API_Results		
Destination	**Distance (km)**	**Travel Time (hrs)**
Delhi	466.8	~6.0
Surat	1163.2	~14.0
Lucknow	86.8	~1.0
Banglore	1787	~22.0

Figure 4: Table: Distance and Travel Time Estimates from OpenRouteService

3.4 Service Type and Weight Category Analysis

Service preference and parcel weight significantly impacted revenue. Standard service was used in 88% of deliveries due to its affordability, yet express service, though only 12% in volume, generated 23% of the total revenue. This indicates a higher average revenue per express parcel (₹385) compared to standard parcels (₹225).

Similarly, while 92% of parcels weighed less than 5 kg and contributed 75% of revenue, the remaining 8% of heavier parcels contributed nearly 25% of revenue. This shows that heavier parcels, though fewer, are more lucrative. Delivery time also varied slightly with weight—heavier parcels averaged 2.99 days compared to 2.72 days for lighter parcels.

Visuals:

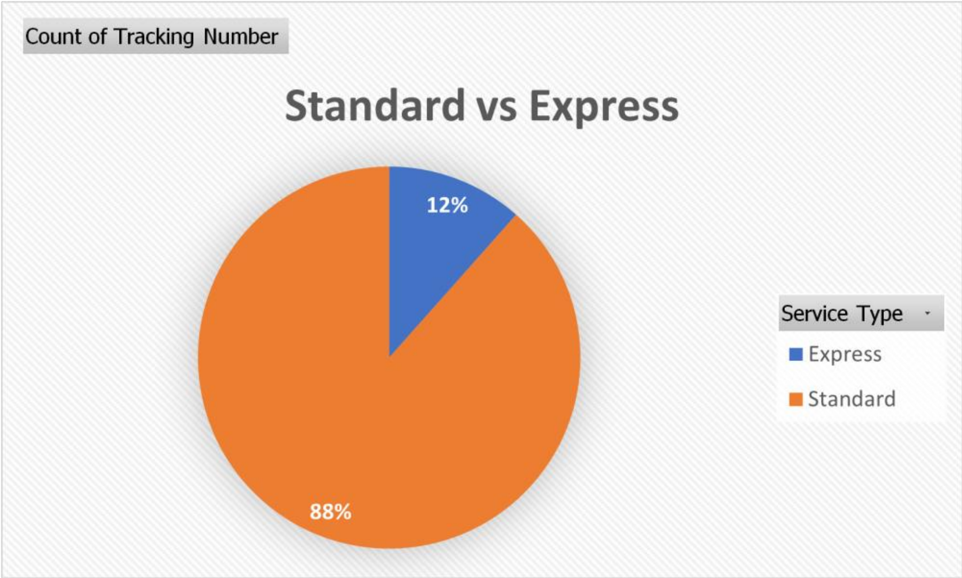

Figure 5: Pie Chart: Service Type Distribution

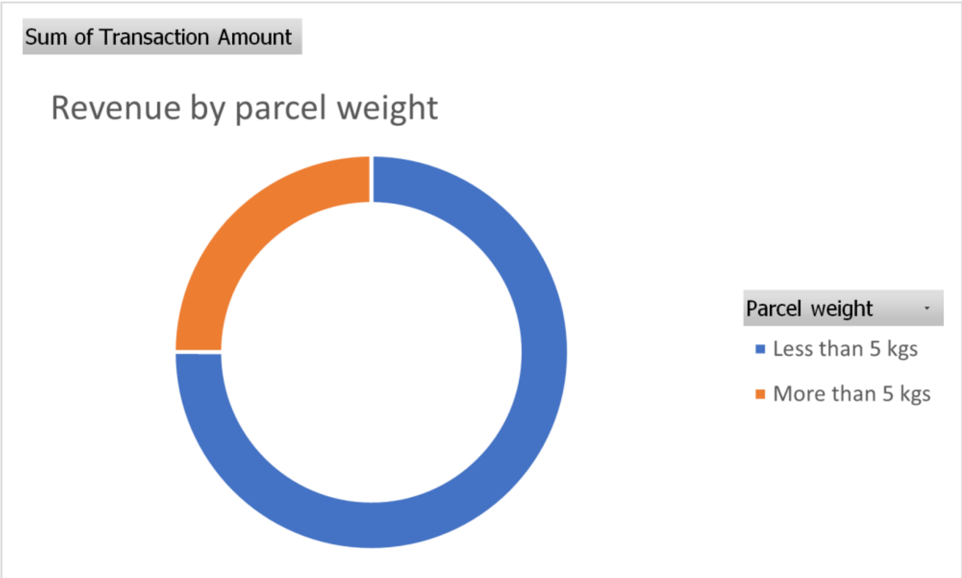

Figure 6: Donut Chart: Revenue by Weight Category

Row Labels	Count of Tracking Number	Sum of Average deivery Time(Days)	Average per Parcel
Less than 5 kgs	207	563.3000986	2.721256515
More than 5 kgs	18	53.75545691	2.986414273
Grand Total	**225**	**617.0555556**	5.707670788

Figure 7: Line Graph: Parcel Weight vs Average Delivery Time

3.5 Peak Activity Days

The franchise experienced maximum parcel activity on Mondays (28%) and Saturdays (20%), while Sundays (1.33%) and Tuesdays (8.44%) showed minimal activity. These patterns are useful for workforce planning and resource allocation. A detailed visualization of daily revenue trends further supporting these observations is provided in **Appendix Figure 12**.

Visual:

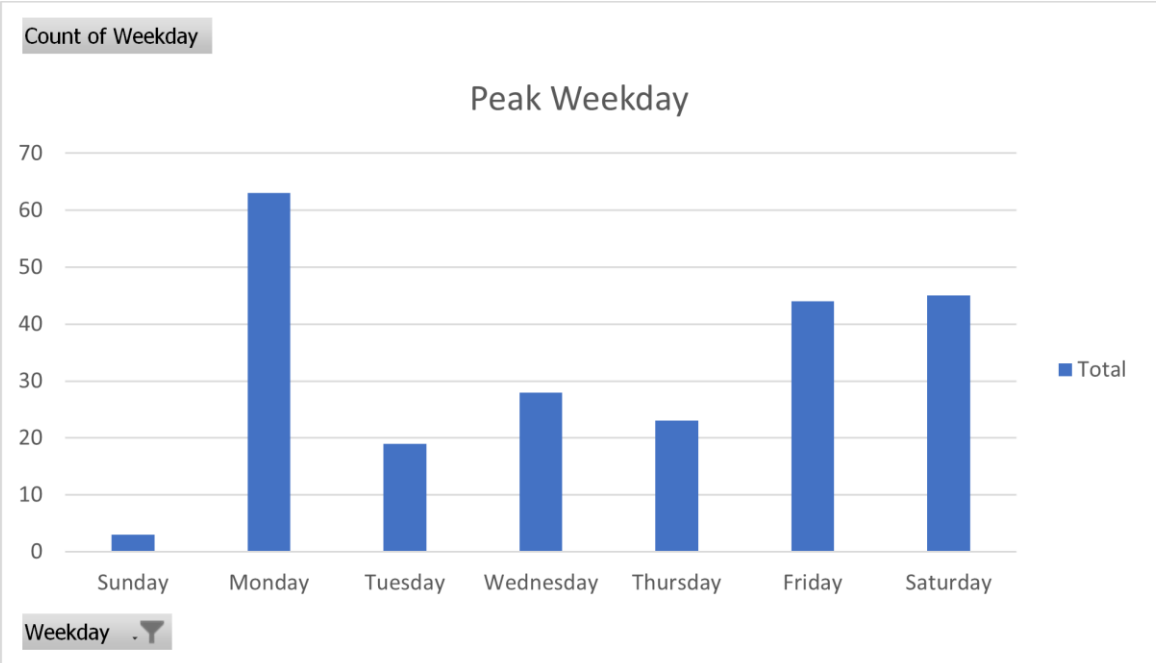

Figure 8: Bar Chart: Parcel Dispatch Volume by Weekday

4 Interpretation and Recommendations

4.1 Interpretation of Results

The analysis conducted in this project offers several critical insights into the operational dynamics of the DTDC Courier Franchise. A clear concentration of revenue in metro and tier-1 cities—particularly Delhi, Bangalore, and Lucknow—highlights these regions as strategic revenue drivers. This indicates a strong correlation between urbanization, business density, and courier demand. In contrast, smaller towns and rural areas contribute marginally to revenue, suggesting that business efforts in those areas may yield lower returns on operational investment.

The parcel return rate was surprisingly low at 0.89%, contradicting initial concerns. While this may indicate improved operational efficiency, it may also reflect incomplete data capture due to manual return logging. This discrepancy underscores the need for better data recording practices to make reliable inferences.

The delivery delay analysis revealed that logistical challenges—particularly for long-distance deliveries—are a significant contributor to service delays. The validation through travel-time data from OpenRouteService confirms that distance and travel duration directly influence delays. These findings align operational inefficiencies with geographic realities, suggesting a need for better route planning and prioritization mechanisms for distant zones.

A notable finding is the impact of service type and parcel weight on profitability. Although the majority of customers opt for standard service and lighter parcels, express deliveries and heavier parcels contribute a significantly larger share of revenue per transaction. This suggests that high-value segments, although smaller in volume, can be crucial to increasing overall profitability. Such insight reinforces the importance of not just volume-based metrics but also value-driven analysis when evaluating performance.

Lastly, the identification of peak operational days (Monday and Saturday) provides practical insight into customer demand patterns. These weekly trends, when considered alongside delivery bottlenecks, offer guidance for manpower allocation and resource scheduling.

Overall, the results paint a comprehensive picture of the franchise's performance, combining geographic, temporal, and operational dimensions to support evidence-based decision-making.

4.2 Recommendations

1. Focus on High-Value Parcels: Design targeted marketing campaigns promoting express and heavier parcel services. Discounts or loyalty points could encourage customers to use premium services.

2. Optimize Routes Using API Data: Use routing APIs to analyse traffic conditions and estimate travel times before dispatch. Prioritize express deliveries for distant cities like Surat and Delhi.

3. Digitize Return Logging: Switch from manual return logs to app-based entry or barcoded POD forms to track failed deliveries more accurately.

4. Implement Address Verification at Booking Stage: Introduce a PIN code validation system using Google Maps API or similar tools to ensure accurate address entry. This will help reduce failed deliveries and improve customer satisfaction.

5. Reallocate Manpower Based on Peak Days: Deploy more delivery agents on Mondays and Saturdays. Use Sundays and Tuesdays for admin work or rest days to balance labour efficiency.

6. Encourage Feedback Collection: Include QR codes or SMS links for customers to rate delivery service. This adds qualitative insight to the quantitative data.

7. Continue Using BI Tools: Maintain dashboards to track KPIs regularly and share visual reports with staff during review meetings. (Refer to Appendix Figure-11)

Visual:

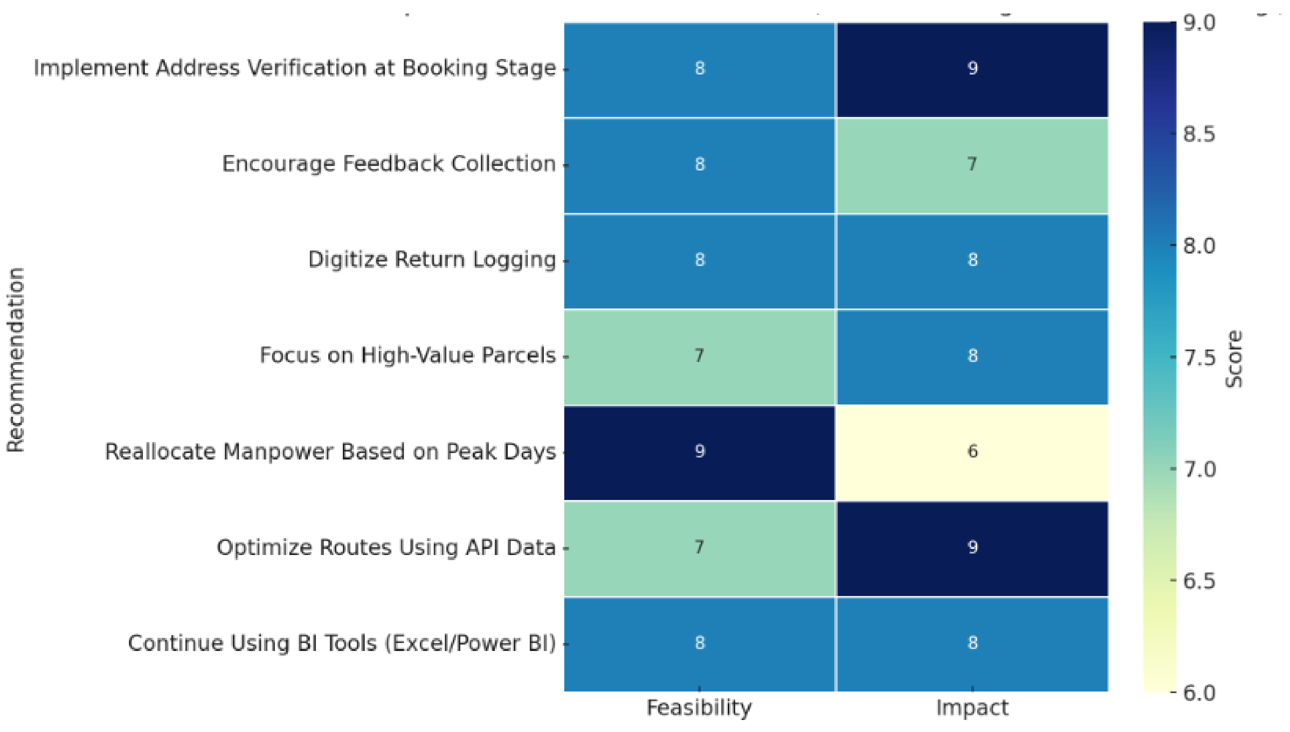

Figure 9: Heatmap Matrix of Recommendations (Impact vs Feasibility)

4.3 Scope for Future Work

This project lays the foundation for several advanced studies:

- **Seasonal Demand Analysis**: Study trends over festivals or holidays.
- **Predictive Analytics**: Use machine learning to predict delay risks or return probability.
- **Live Route Optimization**: Integrate GPS and traffic APIs for real-time route changes.
- **Customer Sentiment Analysis**: Analyse feedback using NLP tools.
- **Automated Reports**: Schedule Power BI dashboards to refresh weekly for timely decisions.

4.4 Limitations

Despite insightful findings, this project has certain limitations:

- **Time Constraint**: Data was collected for only one month, limiting trend visibility.
- **Manual Data Entry**: Prone to human errors.
- **Lack of GPS Tracking**: Delays could not be timestamped accurately.
- **No Inbound Parcel Data**: Only outbound deliveries were studied.
- **External Conditions Ignored**: Weather, traffic, and holidays weren't factored in.

5 Appendix

open_route_service.py > ...

```python
import openrouteservice

# Create client
client = openrouteservice.Client(key=[redacted])

# Coordinates (longitude, latitude)
locations = {
    "Delhi": (77.1025, 28.7041),
    "Surat": (72.8311, 21.1702),
    "Lucknow": (80.9462, 26.8467),
    "Bangalore": (77.5946, 12.9716)
}

# Kanpur coordinates
kanpur_coords = (80.3468, 26.4499)

# Loop through each destination
for city, dest_coords in locations.items():
    coords = (kanpur_coords, dest_coords)
    route = client.directions(coords)
    summary = route['routes'][0]['summary']

    # Convert distance to kilometers
    distance_km = summary['distance'] / 1000

    # Convert duration to hours and minutes
    duration_sec = summary['duration']
    hours = int(duration_sec // 3600)
    minutes = int((duration_sec % 3600) // 60)

    print(f"{city}: {distance_km:.1f} km, {hours} hours {minutes} minutes")
```

Figure 10: Python Code Snippet for OpenRouteService API

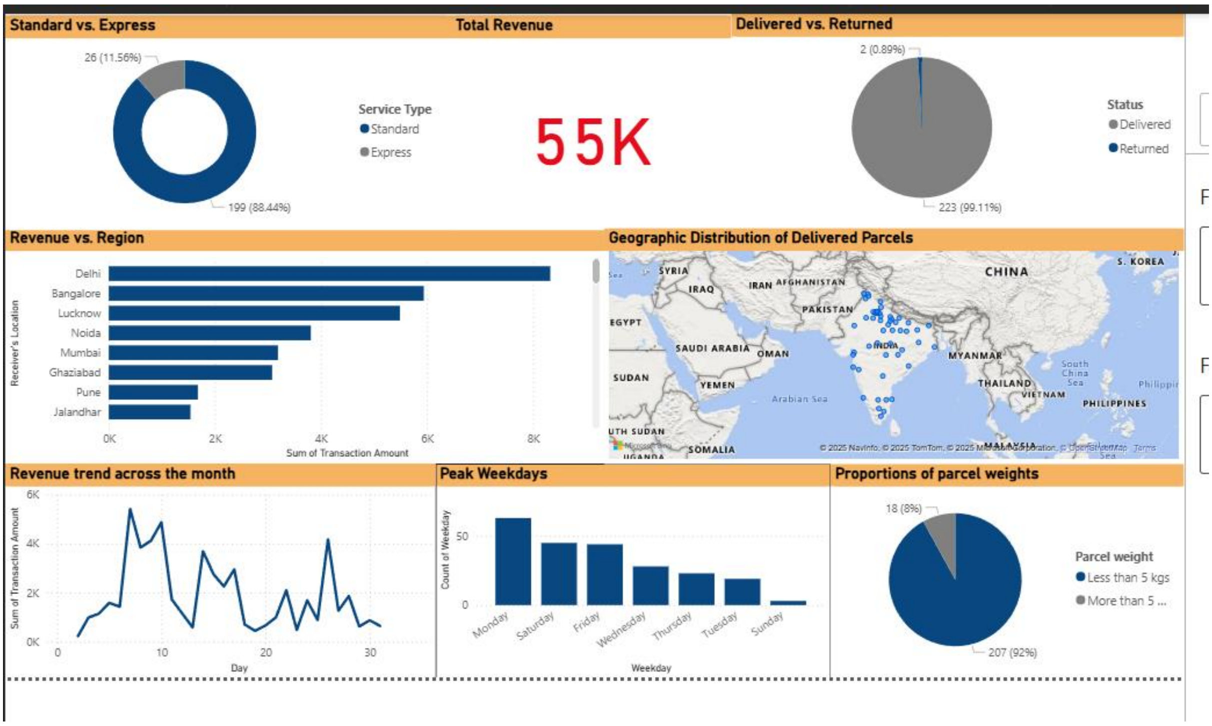

Figure 11: Screenshot of Power BI Dashboard

Figure 12: Bar Chart: Revenue Trend Across Days

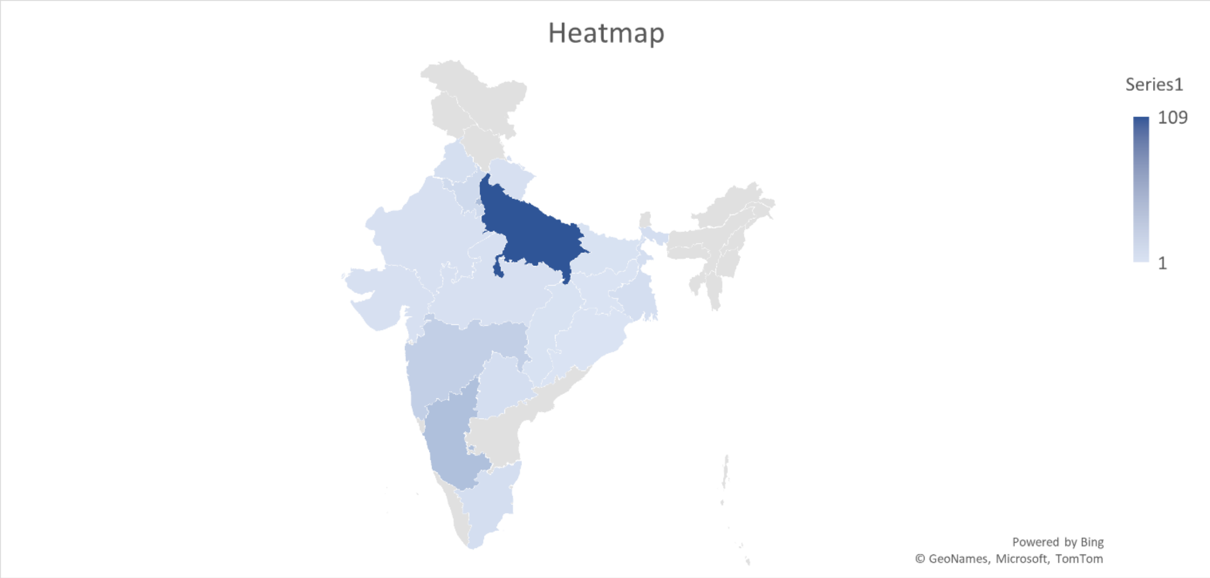

Figure 13: Map: Region-wise Revenue Contribution

Row Labels	Sum of Transaction Amount
October-2024	**100.00%**
Agra	0.58%
Aligarh	1.91%
Allahabad	1.51%
Alwar	0.64%
Amritsar	0.82%
Ankleshwar	0.45%
Auraiya	0.38%
Baharaich	0.49%
Banda	1.76%
Bangalore	10.78%
Bhilai	0.27%
Bhopal	0.98%
Bijnor	0.20%
Bilaspur	0.44%
Bulandeshwar	0.27%
Chennai	0.87%
Dehradun	1.76%
Delhi	15.11%
Erode	0.98%

Faridabad	0.25%
Forbesganj	0.64%
Ghaziabad	5.59%
Gorakhpur	0.44%
Gurgaon	1.91%
Hardoi	0.16%
Hyderabad	2.27%
Indore	2.45%
Jabalpur	0.71%
Jalandhar	2.80%
Jhajjar	0.65%
Jhansi	0.16%
Jodhpur	0.27%
Kannauj	0.27%
Katpadi	1.29%
Kolkata	1.40%
Lucknow	9.96%
Madurai	0.20%
Moga	1.52%
Mumbai	5.79%
Nagpur	1.18%
Noida	6.92%
Patna	2.09%
Phagwara	2.27%
Pune	3.05%
Ranchi	0.64%
Rayagada	0.71%
Sahibabad	0.27%
Shahjahanpur	0.22%
Sikar	0.71%
Surat	1.05%
Trichy	0.38%
Udupi	0.87%
Varanasi	0.73%
Grand Total	**100.00%**

Figure 14: Table: Parcel Delay % per region