Samarthjeet Singh

Last-Mile Delivery Optimization in Indian E-commerce Logistics: A Case Study Analysis of Delhivery's Urban Distribution Network

University of Vaasa
Industrial Management

Vaasa 2026

UNIVERSITY OF VAASA

Author:	Samarthjeet Singh
Title of the thesis:	Last-Mile Delivery Optimization in Indian E-commerce Logistics: A Case Study Analysis of Delhivery's Urban Distribution Network
Degree:	Master of Science in Economics and Business Administration
Degree Programme:	Industrial Management
Supervisor:	Petri Helo
Year:	2026 **Pages:** 84

ABSTRACT:

The last-mile delivery is the most expensive and operationally intensive phase of e-commerce supply chains, especially in emerging economies where infrastructure heterogeneity is more of a structural than a special phenomenon. In this thesis, the author explores how delivery operations at individual, urban centres within India can be optimized using just a single case study of the Delhivery Limited Noida Sector 63 centre in the state of Uttar Pradesh.

Surveys A primary dataset of 25,000 shipment records and 3,235 customer feedback logs obtained in Q3 2025 are used inductively to identify the maximum number of patterns in the analysis with the goal of validating the hypothesis. It is analyzed based on Lean Management, Theory of Constraints, SERVQUAL, Principal-Agency Theory and Institutional Voids framework. There are three categories of friction, which are the structural and behavioral determinants of the 23% aggregate first-attempt failure rate: structural friction due to address ambiguity in unplanned urban villages, transactional friction due to Cash-on-Delivery dominance and behavioral friction due to systematic fake attempt recording due to incentive misalignment. The comparative analysis supports the fact that the gap in the efficiency between target areas and unplanned areas of delivery is statistically significant, which proves that the logistics performance in the Indian urban environment is determined geographically.

The thesis is that there will be an Interaction Optimization Framework that will reorient the strategic goal of the minimization of distance traveled to the maximization likelihood of a successful handover. Four specific interventions have been suggested, which are WhatsApp GPS pre-validation, dynamic time-slotting, success-based incentive restructuring, and automated pre-call synchronization. In contribution to theory, a quantification of institutional voids as a cost of operation has been made, Theory of Constraints has been extended to external service constraints and Lean principles have been scaled to unstructured information environments. The results indicate that the human-centric, hyper-localized, and communication-first approach to latemile optimization needs to be applied in the new markets instead of single-algorithmic route optimization.

KEYWORDS: last-mile delivery, emerging markets, maximizing interactions, institutional voids, Cash-on-Delivery, urban logistics, first attempt success, delivery failure.

Contents

Figures

Tables

Abbreviations

Abbreviation	**Full Form**
AI	Artificial Intelligence
APA	American Psychological Association
AWS	Amazon Web Services
B2B	Business-to-Business
B2C	Business-to-Consumer
CAGR	Compound Annual Growth Rate
CNA	Customer Not Available
COD	Cash-on-Delivery
CRM	Customer Relationship Management
CSAT	Customer Satisfaction
ERP	Enterprise Resource Planning
EV	Electric Vehicle
FASR	First Attempt Success Rate
FTL	Full Truckload
GPS	Global Positioning System
IBEF	India Brand Equity Foundation
IPO	Input-Process-Output
IVR	Interactive Voice Response
LMA	Last Mile App
MCP	Model Context Protocol
NCR	National Capital Region
PDCA	Plan-Do-Check-Act
PDSQ	Physical Distribution Service Quality
PII	Personally Identifiable Information
ROI	Return on Investment
RTO	Return to Origin
SERVQUAL	Service Quality (framework)
SLA	Service Level Agreement
SMS	Short Message Service
TOC	Theory of Constraints
TQM	Total Quality Management
UPI	Unified Payments Interface
VRP	Vehicle Routing Problem

1 Introduction

1.1 Background: E-Commerce Growth and Last-Mile Delivery in India

The Indian e-commerce industry has been changing structurally at a scale never witnessed before. The market has expanded, correspondingly, between 2019 and 2024, as a result of high rates of smartphone penetration, the development of digital remittances systems, and the introduction of Tier 2 and Tier 3 urban markets to large platform players, including Flipkart, Amazon India, and Meesho (IBEF, 2024). It is projected that the industry is going to grow to USD 350 billion by 2030, and India will be among the three biggest e-commerce markets in the world (RedSeer Consulting, 2024). The result of this growth, however, has revealed one underlying conflict at the heart of the Indian e-commerce business: the last-mile delivery dilemma. As digital ordering tools have reached Silicon Valley usability, where one-tap checkout, real-time delivery, AI-driven suggestions, the real-life delivery of the ordered is still stuck in the infrastructural realities that digital commerce had decades to perfect before its inception. Totally nameless roads, absence of sequential addressing of houses, and places that are established by verbal logic rather than geospatial principles introduce an element of structural incongruity between the digital promise and the physical fulfillment (Jain et al., 2021).

The last-mile delivery, the last stage of the supply chain between a local distribution hub and the customer door, takes up to 53% of overall shipping costs in spite of taking less than 5% of the total pay range through supply chain (Ranieri et al., 2018). In India, structural inefficiencies further contribute to this cost imbalance, which does not exist in Western markets: the preeminent Cash-on-Delivery (COD) model of payments, where the buyer and seller must meet in person physically and carry out payment, the gig-based workforce of delivery drivers, who are operated under volume-centered incentive models, and the urban duality phenomenon, i.e. the co-existence between formally planned residential areas and informally developed villages in the same delivery area (Mishra & Vishwakarma, 2020).

According to the National Logistics Policy (2022), the last-mile inefficiency was listed among the key constraints of the logistics of the supply chain in India, with the authors indicating that the percentage of logistics as a part of the GDP is almost twice in India, estimated at 13-14 versus only 8 nationwide (Invest India, 2022). This gap can only be narrowed by not only investing in infrastructure, but also approaching the design, measurement, and optimization of last-mile delivery operations in situations where Western models of operations are not readily applicable.

1.2 Case Company: Delhivery Limited

Delhivery Limited is the biggest complete integrated logistics service provider, in India, in terms of revenue as well as operation. Delhivery was started in 2011 and is a publicly listed company on the National Stock Exchange in 2022 and has an operational network of over 18,500 pin codes throughout India and handling in excess of 2 million deliveries each day in its express parcel, freight, and supply chain services delivery divisions (Delhivery Annual Report, FY2024-25). The company works with more than 28,000 active customers such as big e-commerce websites, direct-to-consumer companies as well as institutional shippers.

The operation model of Delhivery is based on a hub-and-spoke order framework. Regional processing centres acquire freight loaded at various sources and then divide it into last-mile delivery routes at regional delivery centres. The unit of analysis in this study is the Noida Sector 63 hub, which is one of the high-density urban hubs by the company, that handles deliveries with throughput of 3,500 to 4,000 delivery run on the daily basis and at the same time, covering 45 delivery executives (Manhral, 2025, personal communication). The delivery area of the hub includes officially designed residential areas as well as the informally grown urban villages, which is why this location is analytically a perfect area to research the impact that urban duality operation has on last-mile operations.

1.3 The Critical Role of Last-Mile Optimization

Three independent economics mechanism are used to explain the disproportionality of the cost of last-mile delivery per the geographical coverage area. First, when the routes are stretched into suburban outskirts or bypassing dense urban areas, diseconomies of density occur due to the decreasing drop size and unit cost rise (Gevaers et al., 2014). Second, the attended delivery conditions imply that as opposed to the Western market where a routine practice is unattended drop-off of parcels, the Indian COD-dominant model presupposes the physical presence of customers to collect their parcels, confirming the financial transaction - which makes the achievement of high variation in service time and high success on the first attempt (Malkoc and Tuger, 2019). Third, the multiplier consequences of reverse logistics imply that each unsuccessful delivery effort creates an re-sorting, re-routing and re-attempt expenses which is approximated to be 1.5 to 3 times of failed first delivery cost (Rao et al., 2011).

These three processes interrelate to each other to the extent that algorithmic route optimization, which is the mainstream in Western logistics, fails to deal with it effectively. Physical distance optimization has reduced one variable but has not controlled the overwhelming cost drivers, i.e., customer availability, locatability of address, and payment mode friction. This paper holds the thesis that the emerging market last-mile efficiency necessitates that a route optimization task is replaced by what this thesis refers to as interaction optimization: the rational optimization of the chance of successful handover at the delivery location.

1.4 Problem Statement and Research Gap

Although the e-commerce growth is swift, last-mile delivery is the most inefficient actors of the Indian supply chain, approximately 20- 30 percent of the first delivery attempt are failed in urban areas (Jain et al., 2021). The systemic origins of such inefficiency are highly documented at a macro level but not empirically studied at the operational level- in context to the urban duality in the Indian urban frame of reference.

This study is influenced by the three gaps in the current literature. To begin with, there is a geographic bias in the research on last-mile: most of the empirical literature has been done in markets with a regular address system (Gevaers et al., 2011), and on the question of delivery failure in ad hoc urban frameworks, little exists on the customer frontier of emerging markets. The extended systematic review by Lagorio et al. (2023) ascertains that among 2018-2023 publications in the field of the final stage of the logistics process less than 12 percent were set within the framework of an emerging market setting, and only a small fraction of them utilised primary operations data in India in particular. Second, there is a gap in the methodology: the existing research uses Vehicle Routing Problem (VRP) simulations (Toth and Vigo, 2014), which presuppose geospatially accurate addresses and fixed service times - both of which do not hold in unplanned urban areas. Third, the gap in the optimization of interaction exists; route optimization is theoretically advanced, but optimization of human and informational interaction that defines the success of deliveries on the last 100 metres is under-theoretical.

1.5 Research Problem and Research Questions

Research Problem:

How can last-mile delivery operations be optimized to reduce the structural and behavioral challenges of urban distribution networks in India?

Three overlapping imperatives drive this research problem: the economic importance of last-mile inefficiency in a market that is expected to process more than 15 billion e-commerce shipments per year (by 2027) (RedSeer Consulting, 2024); the inapplicability of Western operational frameworks to the realities of the Indian delivery market; and the discovery that the digital physical gap may be one of the fundamental limitations to inclusive economic development in the emerging market.

Research Question 1:

What are the key factors of last 100 metres delivery failure and customer dissatisfaction of Urban Indian delivery?

Knowledge of certain failure processes is the precondition of developing specific interventions. This query visualizes the taxonomy of the failure modes: structural, transactional and behavioral based on primary operational data of the Noida Sector 63 hub.

Research Question 2:

What are the operational performance and failure patterns in the planned residential sectors in comparison to the unplanned urban village in the same delivery zone?

The Indian cities are unique in the sense that they integrate both types of infrastructure in the same route of delivery and this requires infrastructure-differentiated methods of operation. The hypothesis of urban duality is empirically tested in this question.

Research Question 3:

What process interventions do you think will eliminate the digitalphysical gap in delivering and improve success in first-attempt delivery?

Such a question reforms empirical results into practical and low-capital intervention which can be available to logistics service providers of all sizes which operate in the emerging market conditions.

1.6 Scope and Delimitations

As the single case study described by Yin (2018) suggests, the main object of analysis will be the last-mile processes of the Delhivery Limited hub in Noida Sector 63. Its geographical area includes the entire area of the hub delivery including officially planned residential areas (Sectors 75, 76, and 15/16) and informally established urban villages (Mamura and Bishanpura) and also commercial zones.

The timeline of quantitative analysis is Q3 2025 (1 July -30 September 2025), which has been chosen strategically since it portrays the base level of operational performance before the festive season boom. The initial dataset will consist of 25, 000 shipment records and 3, 235 customer feedback entries based on the internal ERP and CRM systems in the company of Delhivery coupled with field observations and a structured interview with the operations manager of the hub (Manhral, 2025, personal communication).

This research restricts itself in four aspects. To start with, it covers last-mile only and the first-mile pickup and middle-mile linehaul operations have not been taken into consideration. Second, it cannot deal with both B2C and B2B freight and reverse logistics deliveries. Third, it puts focus on operational process enhancement rather than an algorithmic system design. Fourth, it operates on operational performance measure instead of using financial profitability measures as the main evaluation measures.

1.7 Thesis Structure

This thesis is organized based on a theory to evidence to implication design in six chapters. Chapter 1 lays the research context, sets out the research problem and also forms the three research questions. The second chapter is a review of the scholarly literature on the last-mile logistics, operations management theory, service quality frameworks and new market logistics and framed the conceptual framework explaining the empirical analysis. Chapter 3 describes the research philosophy, design, rationale of selecting the cases, framework of data collection, and analysis methods. Chapter 4 provides empirical results, designed based on the analysis of the baseline performance, failure mode taxonomy, comparative infrastructure analysis, and the customer feedback themes. Chapter 5 explains the findings in the theoretical terms developed in Chapter 2 and creates the Interaction Optimization Framework as a contribution to strategies. Chapter 6 coincides the findings concluded of the research, answers the three research questions, managerial recommendations, theoretical contributions, limitations and suggests future research directions.

2 Literature Review and Theoretical Framework

2.1 Introduction

The chapter gives the research the context of the general academic literature on supply chain management, operations management, and service quality theory. The theoretical frameworks, discussed in the context of this paper, were chosen based on the directness to the research problem featured in Chapter 1: the discrepancy between the digital ordering efficiency and physical delivery performance in the last mile of urban Indian logistics. Theories of operations management, namely the Lean Management, Theory of Constraints, and Six Sigma, should offer the analytical tools to diagnose and eliminate inefficiency in the delivery processes. The dynamics of service quality frameworks, specifically SERVQUAL and Expectation-Confirmation Theory, can be utilized to describe the relationships of customer satisfaction that determine the perceived performance of delivering services. The literature on emerging market logistics deals with the institutional and structural peculiarities of the Indian context of operation that renders it different as compared to the Western models on which most of optimization theory is founded.

The review takes the form of a funnel, with get-grounded macro-level conceptual introductions leading to get-grounded operational particulars. Section 2.2 gives last-mile delivery its definition and analyzes its cost framework and optimization strategies that are known. Section 2.3 is a review of the operations management frameworks that guide this research. Section 2.4 focuses upon the service quality theory and its implementation to logistics. Section 2.5 deals with the particularities of urban logistics in emerging markets with special references to the empirical data that is provided about the Indian case. In section 2.6, the author addresses the importance of customer feedback as a strategy of service improvement. Section 2.7 brings together all these views into a conceptual framework on which the empirical analysis in Chapters 4 and 5 will be conducted.

2.2 The Last-Mile Delivery in E-Commerce Logistics

The development of e-commerce throughout the world has radically redesigned the logic of the design of logistic networks. Traditional supply chains were structured in the form of push systems built on Business-to-Business (B2B) palletized freight transported in high and consolidated quantities (Bhattacharjya et al., 2016). The emergence of online retail has compelled Business-to-Consumer (B2C) logistics to be shift to a pull model where economies of scale are completely eliminated because the parcels are now destined to be scattered at the homes of dispersed consumers (Lim et al., 2018). This system change has made last-mile logistics not an afterthought in the operations of e-commerce companies but a product differentiator (Mangiaracina et al., 2019).

2.2.1 Defining Last-Mile Delivery

In academic literature, the last-mile delivery entails the last part of a supply chain the consignments flow upon the final distribution point, which is often a local hub or a depot to the end consumer door (Gevaers et al., 2011). This is where the homogenized, managed upstream supply chain lies in the predicted as well as unforeseen consumer environment. According to Rao et al. (2011), the last mile is the moment of truth in the customer experience, as the perception of the delivery interaction Physical Distribution Service Quality (PDSQ) becomes the key factor in brand loyalty in e-commerce, because the only physical contact point of the Human being during the transaction is the physical delivery.

2.2.2 Structure and Cost Efficiency Issues

The last mile cost problem is disproportionate due to both structural inefficiencies inherent in disaggregated, consumer directed logistics. According to Ranieri et al. (2018), the last mile is just the distance that constitutes the smallest portion of the whole supply chain, but it incurs the 28-53% of the total shipping costs. According to Goodman (2005), the major structural cost drivers are three. To begin with, diseconomies of density arise because the routes are stretched into the residential peripheries, which decreases the size of the average drop and increases the labor and fuel expenses associated with each unit delivered. Second, non-productive dwelling time - time parked in front of the building, time exited in the building access systems, and time communicating with the customers takes an unwarranted proportion of the total route time. Allen et al. (2018) approximate that up to 40 percent of overall stop times in dense urban environments is made up of vertical navigation. Third, there is the issue of underutilization of capacity due to the competitive pressure to deliver shipments fast where a logistics provider will ship partially loaded trucks that will further increase unit costs.

2.2.3 Optimization Strategies

Literature on operations management defines the last-mile optimization efforts as hard (technology-driven) and soft (process-driven) interventions. The Vehicle Routing Problem (VRP) is the set of routes a company needs to identify in order to use to minimize the total travel range or time which is the core of Hard optimization (Toth and Vigo, 2014). Nevertheless, pure algorithmic optimization is faced with declining returns when applied in the real-world where stochastic variables such as traffic accidents, misplaced address entries, fewer-available customers, etc. cannot be represented by deterministic models as Savelsbergh and Van Woensel (2016) note. Soft optimization on the contrary deals with management of receiver aspect of the delivery interaction. McKinnon and Tallam (2003) maintain that optimization of goods reception, using dynamic time windows, collection point options, and active customer response, is capable of lasting longer efficiencies compared to simple route optimization since it involves dealing with the cause of delivery failure than with the efficiency of a failed delivery attempt.

2.3 Operation Management Theory

Operations management gives the analytical tools regarding the conversion of inputs to outputs with a minimum of wastage. In undertaking this study, three overlapping theoretical aspects were utilized, which include process optimization models, performance metrics mechanisms, and the concept of continuous quality improvement.

2.3.1 Frameworks of Process Optimization.

Lean Management is based on the Toyota Production System and is structured in such a way that it is geared towards the methodical removal of muda any activity that does not add value to the customer (Womack and Jones, 1996). Within the context of last-mile delivery, it is possible to identify the following categories of waste: motion waste represented by the duration of the address search; waiting waste represented by the duration of waiting at the main door of the customer and waiting until payment or acceptance of the package; the defect waste represented by the inability to deliver the product at all and the necessity to repeat the entire process at an extra cost. Hines et al. (2004), represent Lean thinking to service environments with an argument that the eradication of waste in human-intensive service processes needs to focus on both behavioral and informational flows, rather than physical ones - which is especially true with the case of delivery operations in complex urban settings.

The Theory of Constraints (TOC), introduced to the world by Goldratt (1990) is based on the idea that the throughput of any system is controlled by the single binding constraint of the system. Simatupang et al. (2004) reveal that the binding constraint of a logistics system is often externality of the organization that can be found in the behavior or availability of the customers but not in internal process or capacity parameters. The critical part on this framing to the current study: there is no throughput improvement in the speed of a vehicle or the efficiency of the route unless the binding constraint is the physical location of the customer at the delivery point. According to Schroder et al. (2008), Six Sigma is a process that makes use of minimization of process variation with a goal of reducing to 3.4 defects per thousand opportunities. When applied to the last-mile delivery, Six Sigma logic suggests that non-uniform implementation of pre-delivery processes the inability to standardize pre-call operations, the need to attend to verification, and time-slot confirmation according to which NOPA is ineffective is the cause of the variation in service time.

2.3.2 Performance Measurement

Improper measurements of logistics performance have focused more on speed, reliability, and cost as the major key performance indicators. According to Gunasekaran et al. (2001), a Supply Chain Performance Measurement system is proposed that separates the strategic, tactical and operational levels of performance that demand different metrics based on the varied decision making levels. Griffis et al. (2012) identify the last-mile measure that is operationally most relevant in the context of e-commerce: the First Attempt Success Rate (FASR) observes the efficiency of the handover process and specifies the multiplier of cost of every failure attempt. Kaplan and Norton (1992) also warn against the use of purely financial measures with regard to performance evaluation since they contend that, operational and customer facing measures provide the leading indicators of financial performance, which are not available to the lagging financial measures. The original view holds especially well in the case of last-mile logistic, which Mentzer et al. (2001) cautiously predicts will lead to the development of perverse incentives on the part of the delivery agent wherein they will report false delivery failures when the performance measures in its volume are too strong of a performance measure.

2.3.3 Continuous Improvement and Quality Management

The tradition of quality management has moved the tradition of inspection-based quality control to the Total Quality Management (TQM) where quality is engrained in the process design instead of being applied prospectively at the end of the process. According to Parasuraman et al. (1985), evaluation of the quality of services should be concerned with the result of a service interaction and the way to achieve it. Heskett et al. (1994) make it official in Service-Profit Chain, where the quality of internal services, measured by employee satisfaction, employee capabilities, and employee retention predicts customer satisfaction outside and, eventually, the growth in revenue and profit. This chain means in the context of delivery, human resources, and incentive coordination issues would not be simply a human resource issue, as the chain of customer-facing service delivery quality is directly linked to it. Continuous improvement has an operational model called the Plan-Do-Check-Act (PDCA): by systematically analyzing the data on failures, developing and executing interventions to improve the processes, quantifying the result, and continuously improving (Bhattacharjya et al., 2016).

2.4 Customer Satisfaction and Quality of Service

Service operations management considers customer an inherent member of the production process and not a mere receiver of its products. Services unlike manufacturing where production and consumption are separated over time they are inseparable meaning production and consumption are consumable together, that is, failures in service can be experienced real time and they cannot be recalled and remedied before the customer experiences them (Parasuraman et al., 1985). The delivery executive in the last-mile logistics is what Bitner et al. (1990) call a boundary spanner - the main human face of the e-commerce brand during the consumption process, the behavior and the level of competence of which define the perception of the entire brand experience by the consumer.

2.4.1 Service Quality Dimensions

The SERVQUAL model, formulated by Parasuraman et al. (1988), operationalizes the quality of services in five dimensions such as Reliability, the capability to make the service deliverable in an accurate and consistent manner; Responsiveness, the desire to serve the customers and offer quick service; Assurance, the experience and courtesy of service workers; Empathy where the service workers offer individualized care and Tangibles where the facilities and the staff members appear physically. In his article, Mentzer et al. (2001) translate this framework into the logistics setting in the rather Physical Distribution Service Quality (PDSQ) framework, where timeliness and condition of delivery become the key dimensions of quality. Nevertheless, as demonstrated by Goh et al. (2021), empathy, which in this case refers to the readiness of the rider to go out of his way to find the customers and inform them clearly about the delivery challenges, is more effective at predicting customer satisfaction in emerging markets than the quality dimension hierarchy seen in developed market research.

2.4.2 Customer Expectations and Satisfaction

In the Expectation-Confirmation theory, Oliver (1980) theorized the customer satisfaction as a variable of the distance between the expectation pre-consumption and post-consumption of performance. Satisfaction will be experienced when the perceived performance is in line with or even a bit better than expectations; dissatisfaction experienced when it is less than expectations in line with the extent of betrayal. The so-called Amazon Effect that Bhattacharjya et al. (2016) describe as a normalization of same-day or even next-day delivery as a standard customer experience alternative and not as a higher-value service, has escalated the expectation bar firmly placed against which the logistics offerings by Indian firms would be priced. This generates a structural satisfaction gap in those delivery areas that are limited in infrastructure, the physical capacity of the delivery network is no longer able to meet the digital expectations offered by online shops at the order entry point. As Mangiaracina et al. (2019) claim, proactive communication and setting realistic expectations become more likely to oversee customer satisfaction in these contexts than any of the efforts to increase the speed of delivery.

2.4.3 Recovery and service failure

Failure to deliver is operationally unavoidable due to the stochastic quality of traffic in cities and the availability of customers. Instead, the body of academic literature forms the service recovery, the reaction of the organization in response to the service failure, as one of the important factors in determining the post-failure customer loyalty. Tax et al. (1998) show that three dimensions of perceived justice mediate customer ratings of service failure and they include distributive justice i.e. the fairness of recovery; procedural justice i.e. accessibility and fairness of the resolution process and interactional justice i.e. courtesy and respect exhibited by the service personnel in the recovery. Forgas-Coll et al. (2012) record the Service Recovery Paradox: under some circumstances, a service recovery performed perfectly will lead to more satisfaction after an experience than an impeccably performed service in the first place, as it will offer the company a chance to show responsiveness and care that no failure would offer before the recovery.

2.5 Urban Logistics Emerging Markets

Most of the operations management theory was developed and proved useful in the framework of developed economies that have a standardized physical infrastructure, formal addressing system, and a high level of institutional development (Savelsbergh and Van Woensel, 2016). Their direct transfer to the politico logistic scenarios of emerging market settings that lack contextual transformation is a methodological fallacy that systematically projects organizational and institutional contraint to operational ineffectiveness. According to Khanna and Palepu (2010), such contextual differences are defined as institutional voids, the lack of the intermediary institutions, market mechanisms and enforcement systems which make successful buying and selling in the developed markets possible. The different forms of institutional voids in logistics are infrastructure voids, regulatory ambiguity, inadequacy of address system, and information asymmetry between the logistics providers and the customers.

2.5.1 Emerging Market Logistics Characteristics

Three structural phenomena that can be found in emerging market logistics environment but have no direct analogue in Western logistics theory can be distinguished. The first, infrastructure discontinuity not all emerging market logistical systems are continuities on the multimodal networks of developed market economies, where there are only sharp changes separating well maintained arterial roads and the worn out local street networks. Fransoo et al. (2017) record that in high-density developing market urban centres, the final 100 metres of the delivery process often consists of unpaved routes and lanes to which motorized transportation is not suitable, and manual porterage cannot be present in a regular route optimization framework. Second, its micro-retail sphere dominance generates a discontinuous destination delivery environment where providers of logistical care are forced to make millions of separate micro-transactions instead of coherent movements in bulk, which significantly raises the cost of dealing with units (Boulaksil & Belkadi, 2016). Third, conformity in a wide manner due to varying regulations in the various municipalities restricts the standardization of the vehicle fleet and route optimization in planning.

2.5.2 India-Specific Context

In the larger scope of the emerging market, India represents a unique logistics dilemma with co-excessive digital commerce facilities and pre-modern physical delivery facilities. The Indian e-commerce sector with the compound annual growth rate of about 27 per cent. and projected to become over USD 350 billion by 2030 is dependent on a network of delivery that would have to cover geographies of formally planned and GPS-navigable residential areas as well as informally built urban villages where addresses are exchanged through oral traditions instead of postal code (IBEF, 2024).

The survey of 847 publications on the subject of the last-mile of logistics published in 2018-2023, conducted by Lagorio et al. (2023), discovered that less than 12 percent of the existing empirical studies were carried out in assessing emerging market settings, and less than 4 percent of discoveries resorted to primary data on operations in India. The results of the studies, which indeed study Indian last-mile delivery, substantiate the presence of a steady trend of structural delivery failure, based on the ambiguity of addresses and the friction on modes of payment. The article by Jain et al. (2021) compares the process of delivering across three large Indian logistics suppliers and reports that the average first-attempt failure rates in mixed urban areas are 20-28% with address unlocatability and Cash-on-Delivery refusals as the most common types of failure modes - data points that prompt the empirical research in this thesis directly. Mishra and Vishwakarma (2020) go further to show that in unofficially organized urban settlements a delivery executive will spend on average of 3.2 minutes per stop in address navigation compared to less than one minute in planned aircraft areas - a time cost that escalates on a route-based basis in this study to generate the overall benefit of efficiency differentiation it determines.

A consistent Cash-on-Delivery prevalence in India, which is estimated to be 45-55% of the overall e-commerce transactions volume in Tier 2 and Tier 3 markets as of 2024 (RedSeer Consulting, 2024) converts delivery executives into a cash courier that has to carry and match money every hotel. Malkoc and Tuger (2019) have confirmed that COD stops cost 3-5 minutes more service time compared to prepaid deliveries and cause a unique type of failure with the buyer refusal - the rejection of a delivery at the doorstep because of cash unavailability or buyer regrets his order after the order was delivered - that does not exist in prepaid logistics systems.

2.5.3 Infrastructure and Cultural Factors

The address system that was utilized in officially designed Indian urban areas, with standard number of plots, a naming system of sectors and towerflat designs, estimates the geospatial accuracy needed by a GPS-based system. In spontaneously generated urban villages, on the other hand, street names are not geometrical, but form a list of landmarks familiar in the area: "Near Yellow Sweets Shop, behind the big Peepal tree, second lane on the left hand side of the street" Jain et al. (2021) refer to it as landmark-based navigation, and record that it leads to a reliance on pre-delivery telephonic communications where the delivery executive calls the customer to receive verbal instructions on reaching the last 50 metres. Delivery failures can further be those resulting in the failure of information transfer at the level of information-exchange, such as the failure of customers to pick up their phones, which in turn is a major cause of delivery failure.

The high-rise vertical development of residential estate planning projects in the Indian organized cities added another access limitation. The security measures that are in place in the high-rise complexes have mandated that the delivery executives go through man-controlled checkpoint, wait to be cleared through the building security then find their visitor parking, use service lifts systems, and walk through long building corridors before arriving at the delivery place. As noted by Allen et al. (2018), these vertical logistics frictions use the time on delivery that route optimization models consider negligible. The situation with cultural norms concerning unattended delivery further adds to the problem of access: whereas in Western markets, parcels may be left at the door, in the case of the Indian urban environment, the concern of security aspects and the compulsory provision of COD implies that the issue of customer appearance remains an instance of binary limitations in every delivery attempt.

2.6 Customer Feedback as Service Improvement Tool

The customer feedback has ceased being a reactive function of managing complaints to be proactive as a strategic tool. Griffin and Hauser (1993) show that internal operation measures are systematic in understating the extent of service failure as they only record the failures that are officially recorded whereas most dissatisfied customers silently leave without a complaint. According to the TARP Iceberg Principle the documented document by Goodman (1999), among all customers who like to complain in an official way, at least one out of twenty is equally unsatisfied without filing complaints on the official level, indicating that the complaint statistics only reflect the exposed part of an immense service quality shortfall.

Qualitative aspects of the service experience that are not detectable by the quantitative operational metrics include the access to the qualitative dimensions of the user satisfaction rating in form of a structured feedback analysis tool as well as verbatim logs of complaints and social listening systems. According to Fundin and Bergman (2003), unsolicited commentary in the form of verbatim feedback than the framing of the failures by the customer is the most diagnostically useful customer feedback, as it can reveal existing gaps between how internally framing of failures is coded and how the same failures are perceived externally. Wirtz and Botwinick (2019) operationalize it as the idea of Service Intelligence: the methodical detection of feedback patterns, that after they are detected, make it possible to intervene before failures emerge on a large scale, changing the customer feedback mechanism into a responsibility in hindsight to a proactive engine of operational optimization.

2.7 Conceptual Framework

The literature that has been reviewed in the chapter culminates at critical theoretical gap. This is because the western model of logistics optimization is based on three assumptions, which are address inputs to be geoscally accurate, physical processes to be the most ineffective, and customers to be passive observers whose role in the delivery process is restricted to being present. The Indian urban last-mile setting violates these three assumptions in its structure, with addresses being often descriptive and ambiguous, the informational and behavioral forms of inefficiency being the most common ones, and the determinant of the delivery success being the availability and cooperation of customers.

The focus of this gap is what impels the use of the Integrated Last-Mile Optimization Framework, which is presented in this paper and illustrated in Figure 2.1. The framework redefines the optimization of the last mile as a three-layered system. Input Layer contains the structural constraints, which endow each delivery attempt: hard infrastructure, such as urban typology and quality of roads network, and soft infrastructure, such as address data quality and forecasted customer availability. Process Layer entails the actions that are completed to final 100 metres of delivery: physical navigation and access, as well as handover interaction including customers contact, payment processing and Proof of Delivery capture. Output Layer, will have two types of performance measure; operational efficiency, which is gauged by FASR and service time, and the service quality, which is gauged by CSAT scores and complaint frequency. The feedback loop is a connection between the output metrics and the input layer through which the system is capable of learning about the past performance and gradually achieve higher score in address quality and customer availability prediction.

The main thesis of the framework is that in the Indian setting, no sustainable enhancement in the last-mile performance can be realized with the help of route optimization only, as route optimization is merely about the physical movement whereas the informational and behavioral restriction held in the Input Layer remains uncontrolled. What needs to be done instead is interaction optimization the pre-validation of soft infrastructure inputs in a systematized manner before physical delivery assets are deployed what this study operationalizes via the empirical analysis of Chapters 4 and 5.

.

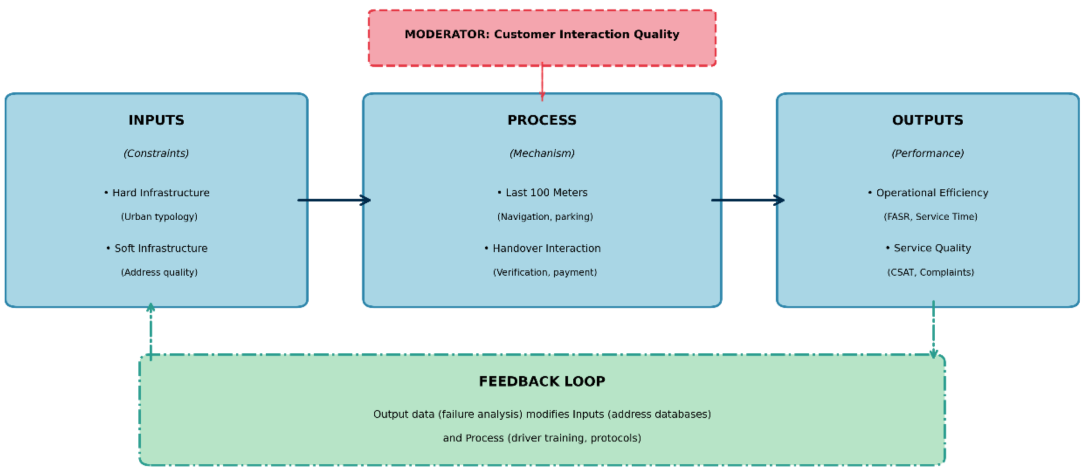


Figure 2.1: Conceptual Framework of the Study (created with the help of draw.io)

In the study of the last-mile logistics of last-mile logistics services, Table 2.1 below provides a summary of the theoretical frameworks that have been used, their authors, key concepts, and how they were applied in the analysis of the study.

Table 2.1. Summary of theoretical frameworks applied in the study.

Theory	Key Author(s)	Core Concept	Application to Last-Mile Logistics
Lean Management	Womack & Jones (1996)	Systematic elimination of waste (muda): motion, waiting, defects.	Address search time as motion waste; dwelling at doorsteps as waiting waste; failed delivery attempts as defect waste.
Theory of Constraints (TOC)	Goldratt (1990)	System throughput is limited by its single binding constraint.	Customer availability as the external binding constraint; subordinating route design to constraint management rather than distance minimization.
Principal-Agent Theory	Eisenhardt (1989)	Incentive misalignment between principals and agents creates moral hazard.	The "fake attempt" phenomenon: agents record false failures to avoid difficult stops due to volume-based pay structures.
SERVQUAL	Parasuraman et al. (1988)	Service quality measured across reliability, responsiveness, assurance, empathy, and tangibles.	Measuring the gap between doorstep delivery vs. gate-drop delivery; empathy as a dominant dimension in emerging markets.
Expectation-Confirmation Theory	Oliver (1980)	Satisfaction determined by the gap between pre-consumption expectations and perceived performance.	The "Amazon Effect" inflating expectations beyond infrastructure capability; managing expectations in unplanned zones.
Institutional Voids	Khanna & Palepu (2010)	Absence of intermediary market institutions imposes transaction costs.	Lack of standardized addresses as a void imposing a quantifiable 17.1% efficiency penalty in urban zones.

3 Research Design, methodology, and Case context

3.1 Research philosophy and approach

The chapter provides the methodological basis of the research, which provides the philosophical base, the research design, the logic used to select the case, data collection structure, and data analysis methods that form the research architecture to answer the three research questions posed in Chapter 1. It further gives situational account of the Delhivery Noida Sector 63 hub including which is important in placing the empirical data of Chapter 4 within its operation and geographical reality.

3.1.1 Inductive Research Approach

The methodology of social sciences is systematized in a continuum as between deductive methods that pit hypotheses about the nature of things against actual data, and inductive methods that derive propositions about the nature of things based on the observations of empirically objective patterns (Saunders et al., 2019). The research takes an inductive research method, based on the understanding that the current mainstream of existing theory of last-mile logistics has been written and tested in Western infrastructural settings where addresses are geospatially normalized, payment system is largely digital, and path optimization algorithms can be ran upon verified input data (Savelsbergh & Van Woensel, 2016). Direct application of these frameworks to the Indian logistics urban context, where addresses are often descriptive, Cash-on-Delivery most commonly serves as the payment method, and the binding conditions under which the success of delivery is conditioned are behavioral and informational as opposed to physical would be a methodological error that projects the structural and institutional constraints onto the operational failure.

As part of the inductive logic, presented in Saunders et al. (2019), the research passes from the particular (empirical) observations, e.g., the frequency of COD refusals on high-value orders, or the over-concentration of address-related failures within unplanned village zones to the construction of the Interaction Optimization Framework. Such an observational-pattern-structure flow is the kind of inductive methodology with which one can achieve and best conforms to a research situation in which the theory is insufficient to the situation at hand (Eisenhardt, 1989).

3.2 Research Design: Single Case Study

The study design of this research is a single case study research. The unit of analysis is the last-mile delivery operations of the Delhivery Limited in the hub of Noida Sector 63. Case study type is chosen due to the fact that it is the most suitable design in answer to the research questions that involve the how and why complex operation phenomena are examined and observed in real-life conditions, which cannot be experimentally controlled by the researcher (Yin, 2018).

3.2.1 Case Study Method Justification

The logistic failure in delivering in Indian urban areas can be seen as the result of the combination of the physical infrastructure, behavioral tendencies of delivery operator, customer presence tendencies, and payment transaction friction - they all function in a live environment with each other and result in delivery failure. An archival or survey-based study or a purely statistical study would only record the results of these interactions but not the mechanisms. A case study design makes possible what Geertz (1973) calls thick description: setting the quantitative performance data in the qualitative texture of operational reality - traffic conditions, fatigue among the pilots, the physical challenge of riding along narrow village lanes - that provides the numbers with their explanatory value.

According to Yin (2018), case study is the most suitable design where the researcher wants to find the effect of the contemporary phenomenon in the real world, especially when the demarcations of the phenomenon and its context cannot be clearly defined. The phenomenon itself and the context cannot be separated in last-mile delivery failure: a failed delivery handover is meaningless outside of the context of the infrastructure it takes place in, the incentive system the delivery executive is incentivized within, and the expectations upon receipt of the attempt by the customer.

Siggelkow (2007) points out that one case is especially useful when it is revelatory and this gives access to a phenomenon that is reflective of a larger category of situations not previously studied with empirical rigor. This is the exact criterion of the Noida Sector 63 hub. It contains two areas of formal high-rise residential development inside a five-kilometre radius, and informally developed Lal Dora urban villages - a structure that is a natural experiment of the impact of infrastructure quality on the delivery performance, with organizational, fleet and technology variables held constant between the two cohorts. The operation of 3,500-4,000 deliveries daily through the hub has a collection of 45 delivery executives, which produces a dataset of good size to conduct statistically significant quantitative research in the context of the case study, and the leadership of Delhivery in the Indian e-commerce logistics market means that the results are indicative of system-wide conditions (Yin, 2018).

3.3 Case Selection and Context

3.3.1 Case Selection Rationale

The mixed-method research approach relies on the logic of purposive, information-based selection as opposed to random probability sampling when choosing cases to study (Flyvbjerg, 2006). The selection of the Delhivery Limited and the Noida Sector 63 hub was done according to three criteria. To begin with, market representativeness: Delhivery is the biggest fully integrated logistics company in India, which supports Amazon India, Flipkart, and Meesho. The patterns of operations in this case are symptomatic of Indian last-mile logistics industry as opposed to being specific to an individual company. Second, degree of complexity in operation: the delivery area in the hub includes planned gated areas, high-density unplanned villages and commercial areas on a single network that allows isolating infrastructure type as an independent variable. Third, accessibility of data: since the researcher had a prior introduction to the hub through the operational ERP and CRM data and Tarun Manhral is the operations manager of the hub, he was able to provide a ground level understanding of the operational challenges, workforce management, and technology limitations in a structured interview without reliance on secondary data.

3.3.2 The Indian E-Commerce and Logistics Nature.

Digitization of commerce is a structural change in the Indian logistics sector. IBEF (2024) predicts the Indian e-commerce market to expand with about 27% CAGR with a projection of USD 350 billion in 2030 with even more to continue to Tier 2 and Tier 3 cities where address standardization is not completed or incomplete. Indian markets are defined as having high transaction volumes and comparatively low-average order values, of about half those of the EU or US markets, which puts a squeezer on the margin to absorb delivery cost per shipment and makes first-attempt success a financial necessity (Jain et al., 2021).

The physical infrastructure is still dissimilar when the digital business layer matures. Cities in India are a patchwork of planned residential quarters and traditionally informal urban villages, the Lal Dora settlements, which got ingested into urban space and remained unaddressed and not connected to formal roads (Jain et al., 2021). Lack of any street signs and multiple houses using sequential house numbers would form a last-mile black box, describe by Mishra and Vishwakarma (2020) an environment where the final 50-100 metres of delivery cannot be tracked by GPS positioning, and operational success cannot be achieved without the delivery executive phone-calling the customer to give verbal instructions on where to go.

3.3.3 Delhivery Limited: Company Profile

Delhivery Limited is a technology-based logistics firm that was established in 2011. By FY2024-25 it had become the biggest fully integrated logistics firm in India with over 18,500 pin codes served by 24 automated mega-gateway processing centres and over 2,800 direct delivery hubs processing over 2 million deliveries every day (Delhivery Annual Report, FY2024-25). The mesh network architecture, instead of the traditional hub-and-spoke design, acts as a proprietary system with node actions being dynamically interrelated and package routing being improved in real-time depending on the existing traffic and load information, so what Delhivery describes as elastic logistics.

There are two pieces of technology that are highly applicable in this study. The former is Addfix, a machine learning address disambiguation address engine that transforms unstructured landmark-structured textual addresses into geocoordinates. The second one is Constellation, an optimization algorithm used to distribute sequences of deliveries to field executives during real time. The primary empirical paradox behind the present study is that, even with state-of-the-art disambiguation and routing technology, Delhivery, despite its 3/4 failure rate at its hub at Noida, is still not that inefficient: it proves that the lack of efficiency in the last-mile optimization here is structural and behavioral, as opposed to technological.

3.4 The Noida Sector 63 Operations Hub

Data will be collected on a primary site, which is the Delhivery Sector 63 Operations Hub (Hub Code: DELNDA063), the last-mile delivery company based in Noida, Uttar Pradesh which is on the Industrial Road, where all data will be collected before being allocated to field executives.

3.4.1 Geographic Defect: The Urban Duality Phenomenon

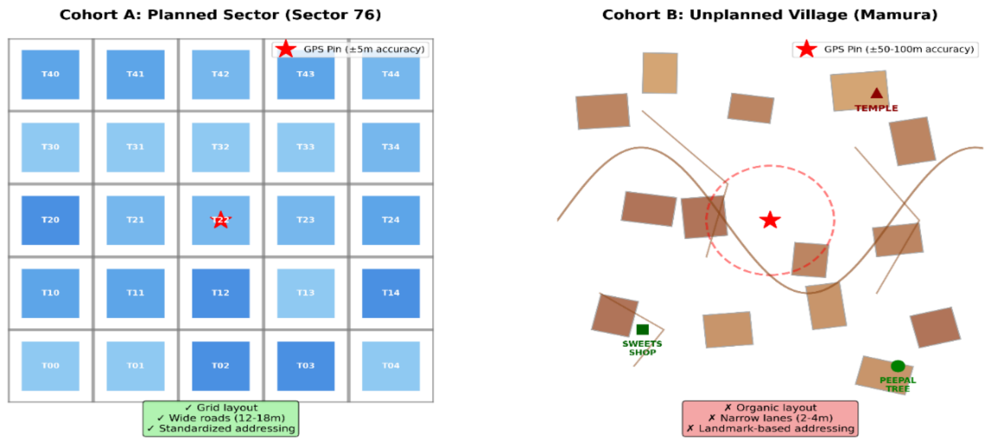
The delivery area around the hub is a natural logistics research study area. The urban villages formally planned residential areas and informally developed within a five kilometres hub area exist - on the same level as their fleet, technology platform, and operational protocols. This design allows infrastructure type to be a factor of objection about the difference in performance of delivery because all other organizational factors are fixed. The comparative features of the two infrastructure cohorts are demonstrated in table 3.1.

Table 3.1: Comparative Characteristics of Cohorts of Studies.

Feature	Cohort A: Planned Sectors	Cohort B: Unplanned Urban Villages
Representative Areas	Sector 75, Sector 76, Sector 50, Sector 137	Mamura, Bishanpura, Chhajarsi, Gejha
Urban Typology	Vertical (High-rise Gated Communities)	Horizontal (Dense Unplanned Sprawl)
Address Format	Standardized Geospatial (e.g., "Flat 402, Tower A, Amrapali Silicon City, Sector 76")	Descriptive Landmark-Based (e.g., "Near Yellow Sweets Shop, Behind Big Peepal Tree, Mamura")
Access Control	High (Security Gates, Intercoms, Visitor Logs)	Low (Open Access, Narrow Lanes)
Navigation Mode	GPS/Map-based (Google Maps accuracy ±5 meters)	Landmark/Inquiry-based (Phone coordination required)

Average Drop Density	High (Bulk drops possible at Security Gate: 15-20 packets/stop)	Low (Individual Doorstep Drops: 1-2 packets/stop)
Road Width	12-18 meters (Wide paved roads, vehicle accessible)	2-4 meters (Narrow unpaved lanes, often foot-access only)
GPS Accuracy	High (±5 meters, reliable pin-point navigation)	Low (±50-100 meters, requires manual verification)
Building Height	15-25 floors (Elevator access required)	2-4 floors (Stairs, no elevator)
Primary Challenge	Availability Constraint: Customer absence during delivery hours; Vertical navigation delays	Locatability Constraint: Address ambiguity; Final 50-100 meters unnavigable by GPS
Typical Service Time	4-5 minutes (predictable, low variance)	2-15 minutes (highly variable, dependent on search time)
Phone Call Dependency	18% of successful deliveries require pre-call	92% of successful deliveries require pre-call

Cohort A - scheduled areas like Sector 75, 76, 50 and 137 have broad paved paths, progressive house numbers and good GPS guidance system to the building entrance, within a 5 metres distance. Customer availability and access control is the main operational issue as opposed to location of address. Cohort B The Lal Dora settlements, which consist of Mamura, Bishanpura, Chhajarsi and Gejha, have narrow streets, many of which are not paved, and most of which are not accessible to a motorcycle, which means that delivery executives have to park in a main street and walk to deliver the packages. There is no numbering on houses or no sequential nature and 92 percent of the successful deliveries involve the executive calling the customer to provide verbal instructions. In case of failure to answer by the customers, the effort is lost altogether in the information stage before the physical delivery is underway.



sector 76 (left) vs mamura(right) satellite map

Figure 3.1: Satellite imagery comparison: Sector 76 planned residential zone (left) versus Mamura urban village (right), Noida, Uttar Pradesh. NB: Overview Image generated using Google Gemini AI and below one is from google maps

3.4.2 Hub Operating infrastructure

The Sector 63 hub uses a 12,000-squear-foot of warehouse space with around-the-clock cycles and has specific inbound processing shifts at night time, and outbound delivery processes in the daylight. In periods of off-peak, the company processes between 4000-6000 deliveries every day in Eeco vans to large commercial drop deliveries and a fleet of over 80 motorcycles to residential and village deliveries. Throughput during the Q3 2025 research was constant 3,500-4,000 attempts with 45 active delivery executives. The employees are overwhelmingly gig-based on the work, paid on a per-packet basis - a form of compensation whose behavioral implications are discussed in Chapters 4 and 5.

3.4.3 Process Flow of the Last-Mile Delivery

Figure 3.2 shows that the daily flow of a parcel through the Noida hub is operational in 5 stages in accordance with the Lean Operations concepts. Stage 1, Inbound and Sorting (04:00-07:00) consists in the beginning of the handling of the shipments of the mother hub, automated scanning, and acquisition of route codes through the Constellation algorithm. Stage 2, Bagging and Allocation (07:00-08:30) sorts shipments by route designated bags assigned to field executives each. Stage 3, Run Sheet Production (08:30) creates a digital run sheet list associate with each rider with information of delivery sequence and customer contact information. Stage 4, the Beat (09:00-18:00), is the main operation period during which the executives are working on their routes and all the changes of status, customer communication, and Proof of Delivery are captured using the Delhivery Last Mile App (LMA). The geofencing technology of the LMA means that the executives must be within a 50-metre GPS radius in order to scan failure statuses - which is meant to prevent the fraudulent process of registering fake attempts, although, as Lufkin and Varghese reveal, it does not do away with the motivation to behave in such a way to get a false failure registered, when the GPS is challenging. Stage 5, Debrief and Reconciliation (18:00-20:00) entails the executive being taken back to the hub, the physical reconciliation of the undelivered packages, and the transfer of such packages to the hold area where they can be re-attempted scheduled or subjected to processing of returns to origin.

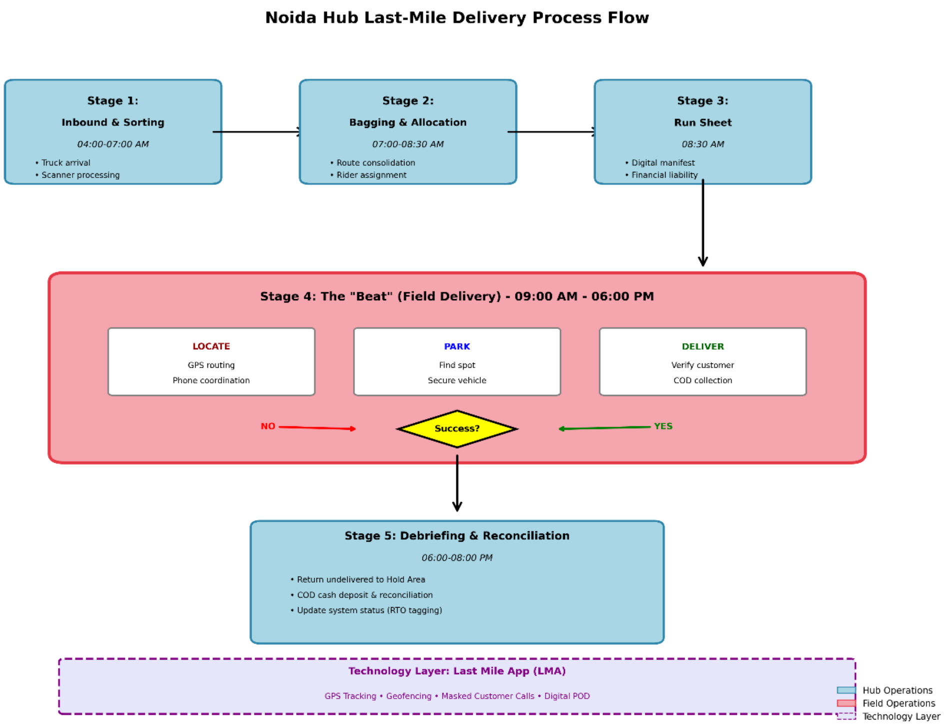


Figure 3.2: Last-Mile Delivery Process Flow Diagram

3.4.4 Noida Hub Structural Delivery Problems.

Two structural issues bring the context of delivery failure which has been recorded in Chapter 4. The first one is address infrastructure duality. Addresses in planned areas are used as geospatial coordinates that could be resolved using GPS within a five metre distance. In informal villages addresses consist of descriptive lists of geographic features of the neighbourhood which are understandable to those who have long lived there but indecipherable to route planners (and non-itinerant delivery managers who have no prior knowledge of the route). The Hub data indicates that customers in such locations often give polymorphic addresses that is, the same physical location defined to one user by different residents, a fuzzy matching issue, which Addfix can address but not completely. The last 100 metres even turns into a search and not a navigation process and nearly creates complete dependence on the pre-delivery phone coordination.

The second issue is the issue of COD financial intermediation. In Tier 2 regions and slums without planning, COD keeps dominating the markets and commercializing all delivery executive professionals by being a temporary financial intermediary. Service time per stop at COD is 5-7 minutes compared to less than 2 minutes to prepaid deliveries - a variance that multiplies over a round trip taking up three to four hours of extra rider time every day . The per-packet gig compensation model provides a structural motivation to maximize volume throughput, which directly contradicts the time requirements of a slowness of COD cash handling behavior and that of challenging navigation through addresses.

3.5 Data Collection Framework

The methodological triangulation of data collection design is to use high volume of qualitative primary data with the qualitative operational knowledge of field observation and structured interviews, to be supplemented by the secondary data of benchmarking. All data sources are an overview of the table 3.2.

Table 3.2: Overview of Sources of Data Collection.

Data Source	Type	Volume / Sample Size	Collection Period	Purpose & Description
Shipment Logs (ERP System)	Quantitative	25,000 delivery instances	Q3 2025 (July 1 - Sept 30)	Raw delivery data including: timestamps, status codes, geolocation, route assignments, delivery executive IDs, payment mode, order value
Customer Feedback (CRM System)	Mixed (Quant + Qual)	3,235 feedback entries (25k CSAT ratings)	Q3 2025 (July 1 - Sept 30)	Post-delivery satisfaction ratings (1-5 star) and unsolicited complaints/feedback logs with verbatim customer comments
Field Observations	Qualitative	15 ride-along sessions (8-10 hours each)	August 2025	Direct observation of delivery execution; Documentation of: navigation, customer interactions, dwelling time, search behaviors
Staff Interviews	Qualitative	5 semi-structured interviews (Managers/TLs)	Sept - Oct 2025	Insights into: operational challenges, workforce management, incentive structures, and technology limitations
Company Annual Reports	Secondary Quantitative	FY2023-24, FY2024-25	Published fiscal year data	Macro-level operational metrics: network-wide FASR, cost-per-shipment, fleet composition, and strategic priorities
Industry Reports	Secondary Quantitative	IBEF, RedSeer Consulting	2024-2025 publications	Industry benchmarks: average RTO rates (15-20%), market size, and competitive landscape data

3.5.1 Primary Data: Shipment and Customer Feedback Data

The main quantitative data set will include N = 25,000 different delivery occurrences on the hub ERP system, through all the Q3 2025 perennial between 1 July and 30 September. The timing was chosen to record base performance on operations prior to the volume spike on Diwali festive season. The data is not a sample of hub outbound delivery activity but a full census, which removes sampling error in the comparative cohort analysis. The geographic coverage includes pin codes 201301 (Sectors 15 16), 201307 (Mamura and Bishanpura) and 201304 (Sectors 75 76).

Each record has AWB Number (anonymized tracking identifier), Delivery Status (binary: delivered or failed on first attempt), Failure Reason Code (standardized CRM disposition), Customer Feedback Score (CSAT on 1-5 scale) and Unstructured Text Feedback (verbatim comments). No Personally Identifiable Information was supplied to analysis after it was anonymized in compliance with data governance requirements aligned with GDPR, with the exception of sector and locality-level geographic identifiers only to classify the cohorts.

3.5.2 Secondary Data

Delhivery annual reports on FY2023-24 and FY2024-25 will include network performance information at the macro-level, such as network-wide FASR, cost-per-shipment trends, and fleet composition, which will put the performance of the hub in the context of the wider organization. Sectional publications provided by IBEF (2024) and RedSeer Consulting (2024) provide sector overviews such as the mean rates of NCR RTO of 15-20% as the relative baseline against which deviations of the hub are measured.

3.6 Data Analysis Methods

Table 3.3 gives the operationalization of major variables that will be used in analysis.

Table 3.3: Record of Key Variables Operation.

Variable	Operational Definition	Data Source Indicator	Measurement Scale	Calculation Method

First Attempt Success Rate (FASR)	% of shipments delivered successfully on the first physical attempt	Status Code: DELIVERED on attempt #1	Ratio (0-100%)	(Total First-Attempt Deliveries) / (Total Out-for-Delivery Shipments) × 100
Customer Not Available (CNA)	Failure due to customer absence or premises being locked	Disposition Code: CNA / DOOR_LOCKED	Binary (Yes/No)	Count of CNA shipments / Total failed attempts
Address Unlocatable	Failure due to inability to locate specific house/building	Disposition Code: ADD_ISSUE / ADDRESS_NOT_FOUND	Binary (Yes/No)	Count of address-related codes / Total failed attempts
Service Time	Duration between vehicle stop and completion of status update	Time Delta: Stop_Time - Scan_Time	Continuous (Min)	(Status Update Time) - (Geofence Entry Time)
Customer Satisfaction (CSAT)	Post-delivery satisfaction rating provided by customer	1-5 star rating in CRM	Ordinal (1-5)	Mean CSAT score of aggregate ratings
Return to Origin (RTO) Rate	% of shipments returned to facility after all attempt exhaustion	Final Status Code: RTO	Ratio (0-100%)	(Total RTO Shipments) / (Total Shipment Volume) × 100
Payment Mode	Method of payment selected by customer	PREPAID or COD	Nominal	Direct attribute from order management system
Geographic Cohort	Infrastructure type classification based on destination	Pin code + locality mapping	Nominal	Classification into Cohort A (Planned) or Cohort B (Unplanned)
Order Value	Total monetary value of shipment in INR	Order_Value field	Continuous (INR)	Direct attribute; Categorized as Low, Medium, High
Fake Attempt	Failed mark by executive but disputed by customer	Disposition: CNA + Feedback: "I was home"	Binary (Yes/No)	Manual coding: Keyword match "was home" with CNA disposition

3.6.1 Descriptive Statistical Analysis

Descriptive statistics determines the starting profile of operation performance of the hub in the entire dataset of 25,000 records. The main measures calculated include FASR stratified by the mode of payment and geographic cohort, Return to the Origin rate, CSAT score distribution and Complaint Frequency Rate by delivery zone.

3.6.2 Cross Tabulation and Comparative analysis

The cross tabulation of the failure reason codes with the geographic cohort variable is used to answer Research Question 2, isolating the infrastructure-specific failure mode distribution. Pivot table analysis is also able to control the confounding effects of time, including weather and day-of-week effects by comparing the performance of cohorts over similar time periods so that observed differences can be explained as due to the type of infrastructure as opposed to correlating with other external factors.

3.6.3 Customer Feedback Thematic Analysis

Thematic analysis process of Braun and Clarke (2006) six-phase, involving familiarisation of data with data, initial coding, theme generation, theme review, theme definition and final report was applied when analyzing the 3,235 verbatim customer comments. Raw comments were coded on the semantic level i.e. such a comment as I was home all day could be coded as a Fake Attempt allegation, and the individual codes were added together to high-order themes that represented different categories of customer experience, the most important of which are discussed in Section 4.5.

3.7 Research Quality and Validity

The research design used in the study is compared with the four criteria of validity used by Yin (2018) to assess case study research. The concept of construct validity is reached by triangulating data: the allegations of fake attempts by the customers are compared with the records of the driver GPS tracks and geofencing scan. Infrastructure cohort variable and the performance measurements are compared in various sub-segments to eliminate spurious correlations as pattern matching provides the internal validity.

External validity US Extrapolative validity is based on the representativeness of the case: the structural circumstances findings of the Noida hub are reported in the literature of Indian logistics as general and not place-specific (Jain et al., 2021; Mishra and Vishwakarma, 2020). Responsibility is maintained through the written thematic coding instructions with clear specification of key word matching, this allows the analysis procedure to be audited and recreated.

3.8 Methodological Limitations

This study was limited in its findings in three ways. Single-hub format makes it difficult to statistically generalize, but this is in part compensated by the market leadership and the representative coverage of the diverse typologies of the Indian urban infrastructure represented by the hub. The three month time frame gives a cross sectional view and will not allow a longitudinal analysis of performance trends or seasonal variation. Customer feedback shows personal perceptions, which can be distruptions in matters not directly affected by logistics operations; triangulation with objective operational data is the way to eliminate the aforementioned and thus the quality observations obtained are likely to be supported rather than based upon themselves only.

4 Findings and Analysis

4.1 Introduction

This chapter includes the empirical findings of the analysis of 25,000 records of deliveries and 3,235 records of customer feedbacks gathered on the site of the hub where Delhivery is located in Noida Sector 63 in Q3 2025. The analysis discusses the three research questions set in Chapter 1, as a progressive trend of aggregate baseline performance to failure mode taxonomy to comparative infrastructure analysis and themes of qualitative customer experience.

Section 4.2 forms the operational performance baseline profile under the descriptive statistical analysis. Section 4.3 breaks up the failure rate of 23% first attemp of 23 per cent into the underlying failure modes. Section 4.4 performs the comparative infrastructure analysis where one tests the urban duality hypothesis. Section 4.5 displays thematic analysis of customer feedback, and reveals the behavioral and interactional aspect of failure that cannot be merged using quantitative metrics. Section 4.6 summarizes all five big empirical findings that will set the groundwork of Chapter 5.

4.2 Base line Operational Performance

The N=25,000 dataset is a full census of all outbound delivery tries at the Noida Sector 63 hub on Q3 2025. Pre-processing left out cancel trips as well as stratified the data using mode of payment and geographic group. Table 4.1 shows aggregate performance indicators and benchmark ranges of NCR industry by RedSeer Consulting (2024) and by IBEF (2024).

Table 4.1: Aggregate Performance Metrics (N= 25, 000)

Metric	Count	Percentage	Industry Benchmark (NCR)	Performance vs. Benchmark
Total Shipments Analyzed	25,000	100.00%	—	—
Successful First Attempts	19,250	77.00%	75-80%	Within range
Failed First Attempts	5,750	23.00%	20-25%	Within range
Delivered (Final Status)	21,500	86.00%	82-88%	Within range
Return to Origin (RTO)	3,500	14.00%	12-18%	Within range

Delivered on 1st Attempt	19,250	77.00%	—	—
Delivered on 2nd Attempt	2,415	9.70%	—	—
Delivered on 3rd Attempt	500	2.00%	—	—
Failed after 3 Attempts (RTO)	2,835	11.30%	—	—

The FASR of a total of 77.0% in the hub provides an indication that the hub follows the trends of the NCR industry, but not an exception. Considering the Western standards of developed markets of 90-95 percent (Edwards et al., 2010), though, the structural performance gap of Indian last-mile setting stands out clearly. Any individual failure in the 5,750 initial attempts has subsequent costs: extra fuel and labor to schedule a new attempt, facility and logistics management to deal with 3,500 subsequent returns to origin, the dissatisfaction of the customer, quantified in the CSAT data in Section 4.5.

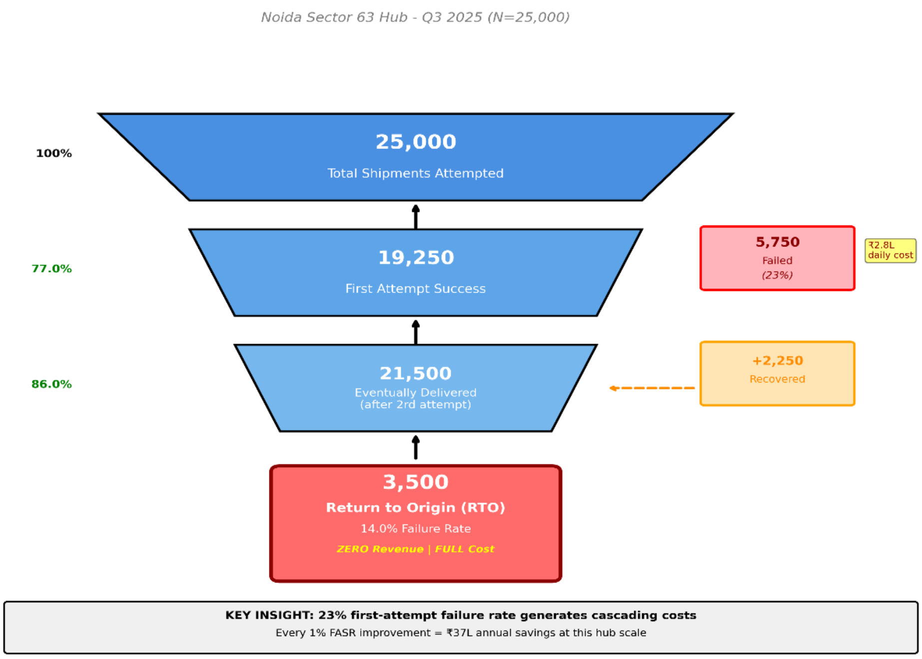


Figure 4.1: Delivery Outcome Funnel Diagram(Created using Draw.io)

4.2.1 Baseline Performance Interpretation

The FASR difference against the Western standards is not an indicator of organizational poor performance: it indicates the industry-wide NCR values and indicates the structural elements of the logistics setting of the Indian city environment, namely address ambiguity, COD supremacy, and the attended-delivery norm. The question of interest then is not why the hub is failing against the western standards but rather what mechanisms underlie the failure rate of 23 per cent and how these mechanisms differ based on the type of infrastructure and method of payment.

4.2.2 Payment Mode Analysis

Paying mode stratification validates and measures the friction hypothesis of transactions worked out in the test of Section 2.5.2. Though the volumes of COD and prepaid orders are similar 55 and 45 percent respectively, their performance picture is drastically different as demonstrated in Table 4.2.

Table 4.2: Performance by Mode of payment.

Payment Mode	Volume (n)	Share	FASR	RTO Rate	Avg Service Time	Refusal Rate
Prepaid	11,250	45.00%	89.40%	6.20%	2.1 minutes	1.80%
Cash-on-Delivery (COD)	13,750	55.00%	66.80%	20.40%	5.7 minutes	12.40%
Overall	25,000	100.00%	77.00%	14.00%	4.1 minutes	7.60%

The COD shipments translate into a FASR of 66.8 percent compared with 89.4 percent prepaid which translates to 22.6 percentage points. There is 3.4 times prepaid rate returning COD shipments to the origin (20.4% and 6.2% respectively). 1.71 minutes vs. 5.7 minutes of COD service time indicates a difference that multiplied over a route of 50-70 stops occupies three or four hours of extra ride time that yields no comparable delivery value. This verifies that physical exchange of cash adds costs of financial intermediation, buyer-remorse refusal risk, and service time variation in the structure of prepaid business nonexistent in prepaid operations.

4.2.3 Geographic Infrastructure Analysis.

The geographic cohort stratification test tests the urban duality hypothesis which states that the type of infrastructure is a key performance factor that is not dependent on payment mode. Table 4.3 shows aggregate measures at cohort.

Table 4.3: Infrastructure Type Performance.

Geographic Cohort	Sample Size (n)	FASR	Variance from Mean	Avg Service Time	RTO Rate

Cohort A: Planned Sectors	14,500	84.20%	+7.2% points	4.2 minutes	9.80%
Cohort B: Unplanned Villages	10,500	67.10%	-9.9% points	7.6 minutes	19.70%
Overall	25,000	77.00%	—	5.6 minutes	14.00%

The most fundamentally based quantitative result of the study is the 17.1 percentage point FASR gap between the cohorts. It confirms that the presence of address ambiguity and access friction in unplanned zones is a penalty on efficiency, which is independent of the capability of an organisation or the level of investment in technology. The difference in average service time: 7.6 minutes in Cohort B, compared to 4.2 minutes in Cohort A helps to understand the search and telephonic coordination load of unexpected zones. The large service time dispersion in Cohort B (between 2 and 15 minutes nominally identical tasks) makes route optimization algorithms operationally useless in these regions, as in Chapter 5.

4.3 Failure Mode Analysis

This part factored out the 23% cumulative failure rate by the individual failure mode, using the CRM disposition code taxonomy of Delhivery to look at the mechanisms by which the failure of delivery attempts occurs.

4.3.1 Taxonomy of Failure Reasons

In the statistical analysis of 5,750 failed first attempts, statistical analysis has shown the Pareto distribution with a few failure modes contributing to most total failures. Table 4.4 shows the complete taxonomy of failure reasons with counts, proportions, major causal drivers and approximate cost of impact/applied.

Table 4.4: Reason of Delivery Failure.

Failure Reason Code	Count	% of Total Failures	% of Total Shipments	Primary Driver	Cost Impact per Failure
Customer Not Available (CNA)	3,162	55.00%	12.60%	Availability / Potential Fake Attempts	₹55 (re-attempt cost)
Address Unlocatable/Incomplete	1,035	18.00%	4.10%	Infrastructure / Data Quality	₹68 (extended

					search + re-attempt)
Customer Refused Delivery	862	15.00%	3.40%	COD Friction / Buyer Remorse	₹142 (full reverse logistics)
Operational Constraints	691	12.00%	2.80%	External Factors (Traffic/Weather)	₹45 (delay costs)
Total Failures	5,750	100.00%	23.00%	—	₹62 average

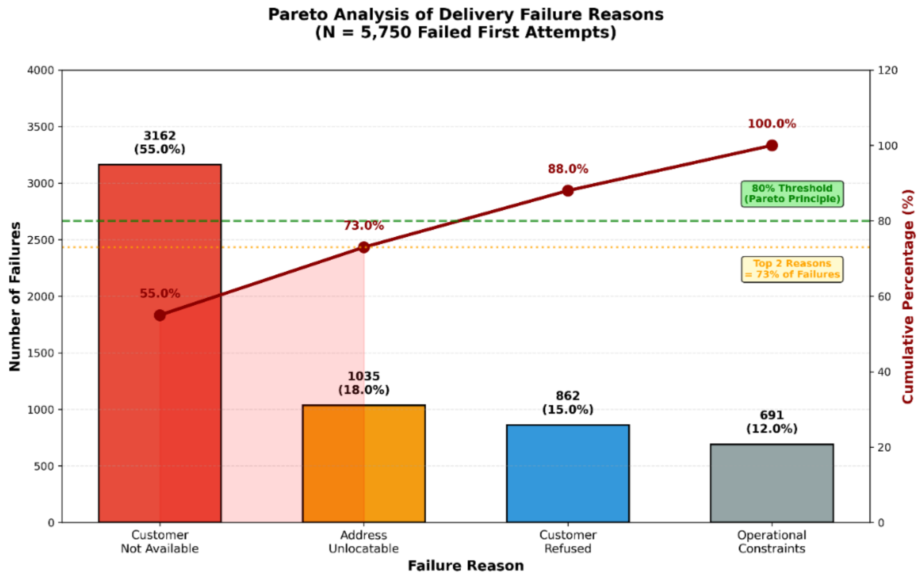


Figure 4.2: Pareto Chart of Delivery Failure Reasons(Created using Data Wrapper)

Customer Not Available (CNA) at 55 percent is the most prevalent failure mode - it is agreed with Esper et al. (2003), who presented the issue of customer presence as the binding constraint in the supply chains of attended-delivery service. Within the Indian context, where no unattended delivery or other collection mechanism have been accepted culturally, the customer non-embarkation ends the purpose of the whole delivery process in the last node and every other supply chain investment has been inconveniently non-economical towards said effort. The most expensive search-related cost per incident with is the case of Address Unlocatable failures at 18% and is [?]68 which represents a long time spent in failed navigation in unplanned areas. Section 4.3.3 shows that a considerable amount of misclassified fake attempts is also under this category in the proportion. Customer Refused Delivery is the most expensive, with the absolute cost per incident at [?]142-- this is the only failure mode that creates complete reverse logistics cost as physical movement to the doorstep had already happened at the point of cash exchange.

4.3.2 COD Order Value and Rejection Patterns

Further analysis of the 13,750 COD shipments depicts that there is a systematic correlation between order value and refusal rate with serious consequences about managing COD risks, as indicated in Table 4.5.

Table 4.5: Correlation between Rejection rate and Order Value (COD Only)

Order Value Range (INR)	Volume (COD Orders)	% of COD Total	Refusal Rate	Primary Refusal Cause	Avg Re-Attempt Success
Low (₹0 – ₹500)	4,500	32.70%	5.20%	Impulse regret / Minor quality	78.40%
Medium (₹501 – ₹2,000)	6,200	45.10%	12.40%	Cash shortage / Change issues	62.10%
High (₹2,000+)	3,050	22.20%	28.60%	Buyer's remorse / Lack of cash	41.30%
Total COD	13,750	100.00%	12.40%	—	62.80%

The rates of refusal increase sharply, by order value and are estimated at 5.2 per cent on low-value orders and 28.6 per cent on high-value orders - which is in line with the buyer-remorse dynamic defined by Griffis et al. (2012), where the higher the time elapsed between the online impulse and the moment of physical payment, the less psychological commitment to the purchase. High-value COD refusals have a lower success rate at a re-attempt of 41.3% using reflective compounding difficulty as a customer has already shown reluctance toward transactions, which is a known operational issue where high-value COD orders constitute a disproportionate number of multi-attempt failures transitioning to Return to Origin.

4.3.3 The Discrepancy of the "Fake Attempt"

Among the analytically important results, one of the findings in the study is the result of comparing the operational disposition codes with the content of customer feedback. The CRM is 3,162 CNA shipment and 1,035 Address Unlocatable shipment. Nonetheless, thematic analysis of the verbatim feedback data indicates that 42 percent of the volume of CNA complaints consists of customers specifically claiming to be at the delivery address at the time the failure was noted, and 34 percent of Address Unlocatable records have been disputed by the customers who state that they can be contacted by telephone during the entire delivery period without getting any call made by the executive.

This 42 percent disagreement rate on 3162 records of CNA show that this is a systemic behavior, not an isolated incident and represents a significant percentage of what is recorded in the operational data as customer-driven failure to actually be agent-driven: delivery executives entering false disposition codes so that the time cost of difficult deliveries in the small village or end of a shift when people are tired does not have to be incurred . This redefines the CNA metric as a pure measure of customer behavior into a composite measure that is a mix of real unavailability with artificial reporting by the agent, on a monumental level in terms of incentive design and monitoring procedures which are established in Chapter 5.

4.3.4 Customer Unavailability Effect on the Route Efficiency

The prevalence of 55% CNA causes a ripple effect on the productivity of routes not only because of the cost of each failed attempt itself. A failed stop still takes up all the travel time, navigation and parking time, and attempt to contact time time - usually three to five minutes of idle time. Assuming a typical route of 50 deliveries with a 20% CNA rate, ten bad stops would result in 30-50 minutes of cumulative dead time that may practically allow ten to twelve more successful deliveries per route every day. The total loss in daily productivity of all the 45 executives of the hub is operationally and financially large and it forms the main economic factor to promote pre-delivery intelligence intervention suggested in Chapter 6.

4.4 Comparative Analysis: Infrastructure as Determinant.

In this part, the hypothesis of urban duality is tested by comparing across cohorts the failure mode distributions, operational friction profiles and cost structures in order to isolate the infrastructure effect by controlling the basis of payment mode and time.

4.4.1 Differences in the patterns of failures

What is more illuminated by the cohort comparison is the difference in the nature of failures between cohorts rather than the aggregate FASR gap of 17.1 percentage points. As table 4.6 indicates not only does the two zones perform differently in aggregate but also the two fail due to fundamentally different reasons of struggle.

Table 4.6: Comparison Failure reason distribution by geography.

Failure Reason	Cohort A: Planned Sectors	Cohort B: Unplanned Villages	Absolute Variance
Customer Not Available	65% (1,495 failures)	40% (1,667 failures)	+25 pp (Planned)
Address Unlocatable	3% (69 failures)	28% (966 failures)	+25 pp (Unplanned)
Customer Refused	12% (276 failures)	18% (586 failures)	+6 pp (Unplanned)
Other/Operational	20% (460 failures)	14% (531 failures)	-6 pp (Traffic/Commute)
Total Failures	2,300 (100%)	3,450 (100%)	—
Success Rate (FASR)	84.20%	67.10%	-17.1 pp

In Cohort A, the address unlocatable failures are only 3% of all failures, which implies GPS navigability and accuracy of geospatial addresses within planned areas. The most prevalent is CNA at 65, which means that the key problem is not where to be able to find the address but the customer home within the standard delivery time, which is a constraint of time and not an information constraint. This distribution inverts abruptly in Cohort B where address unlocatable failures become 28% confirming that addresses and GPS failure in the last 50-100 metres are structural and endemic not random. The negative relation between CNA rate and address failure rate across cohorts - with improvement of address quality, the bottleneck changes to locatability to availability - is also of profound theoretical importance that was constructed in Chapter 5.

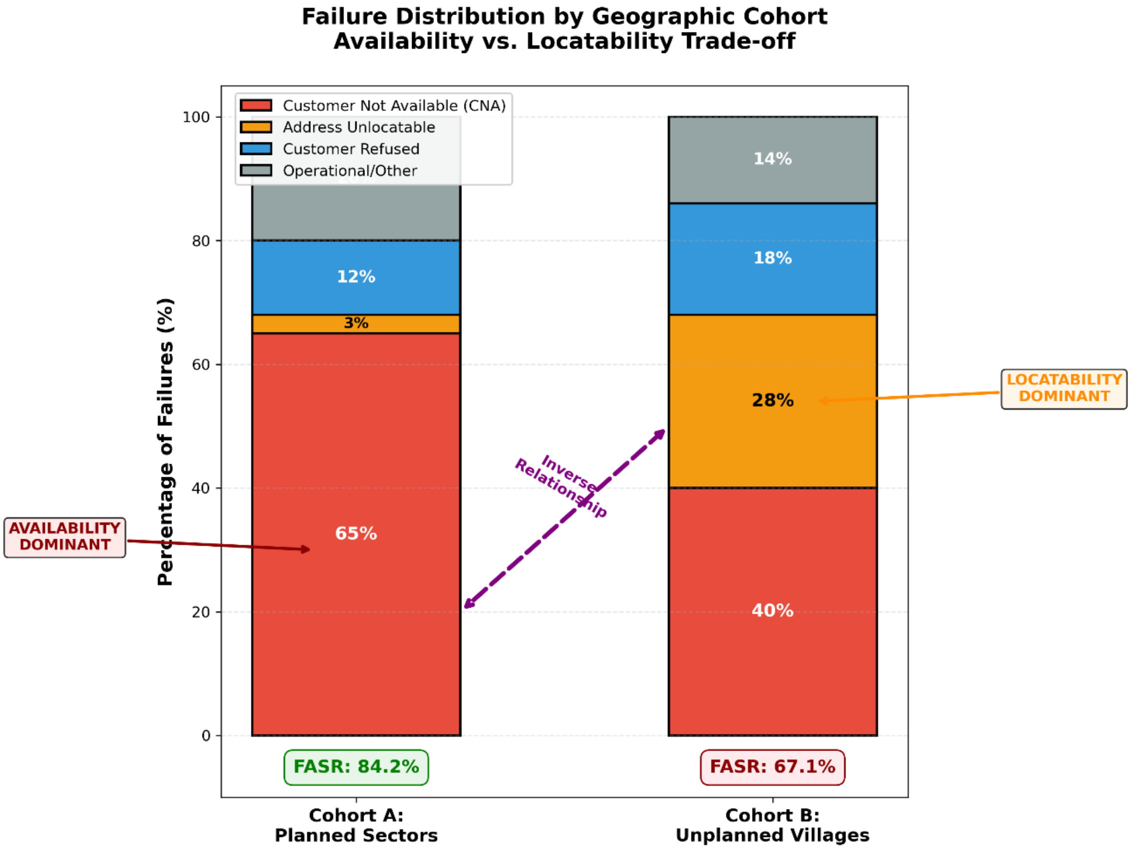


Figure 4.3: Failure Reason Distribution by Cohort (Created using Data Wrapper)

4.4.2 Friction in the Operation of Planned Zone: "Vertical Logistics."

The planned sectors have different access friction despite having better address infrastructures. Delivery executives on the high-rise gated communities endure a hierarchical process of access: security gate entry, visitor registration, movement to visitor parking, service lifts queue, and building corridors cross-travelling before arriving at the apartment door take between five to ten minutes. Most gated communities have a policy of a lobby drop in order to reduce this time cost. Although lobby drops achieve efficiency in terms of saving the executive time and decreasing the instances of CNA failure recorded, they provoke customer dissatisfaction: the residents who have been paying to receive home delivery and have to travel to a building lobby to pick packages are going to perceive this as a service degradation (this perception is reflected in Section 4.5).

4.4.3 Operational Friction Existing in the Unplanned Areas: “Search Cost.

In Cohort B there is no physical friction, but it is informational. The statistics indicate that 92 percent of successful deliveries under Mamura and Bishanpura entailed pre-delivery phone coordination in terms of directions within the last 50-100 metres, compared to 18 percent in Sector 75. It leads to a communication cost, in that when telephonic synchronization fails on a customer, an attempt to make a delivery fails at the information stage, and nothing happens at all until the delivery is announced as unsuccessful. Lanes below four feet in width cause executives to walk around on the primary roads and deliver packages on foot, which creates physical exhaustion and is a direct contributing factor to a higher incidence of fake attempt behavior in the latter half of the shift according to field observations as well as interviews with the managers of hubs.

4.4.4 The Cost of Ambiguity

The analysis as compared shows a fundamental working asymmetry; the planned sectors are time bound and the unplanned villages are information bound. The time constraints one can address partially are scheduling-increasing the FASR by having zones of Cohort A deliver to the evening hours without needing any infrastructure modification. Scheduling cannot solve information constrains: ambiguity may not depend on the time of the day and can only be addressed through the enhanced information delivery before executives leave the hub. Table 4.7 discusses the consequences of the economy by analyzing the re- attempt cost multiplier.

Table 4.7: Probability of Success in Order of Attems.

Attempt Sequence	Volume (n)	Success Rate	Cumulative Success	Marginal Gain	Cost Multiplier
First Attempt	25,000	77.00%	77.00%	—	1.0× (baseline)
Second Attempt	5,750	42.00%	86.60%	+9.6 pp	1.5×
Third Attempt	3,335	15.00%	88.60%	+2.0 pp	2.1×
Fourth+ / RTO	2,835	0% (returned)	88.60%	0 pp	3.0× (total loss)

The marginal ROA on successively launching an attempt decreases exponentially: the second attempt will be successful 42 percent in the presence of 1.5 times baseline cost, the third attempt will be successful only 15 percent with 2.1 times cost, and the final attempt will have zero success with 100 percent reverse logistics cost. This gives good economic reasons why first-attempt success improvement is a strong case: two situations are economically equivalent in terms of preventing one failure followed by an attempt, and two situations are economically equivalent in terms of preventing about two costs of a second attempt. The implication, that the enrichment of pre-dispatch information is cost-effective as compared to re-attempt optimization, is one of the pillars of the Interaction Optimization Framework of Chapter 5.

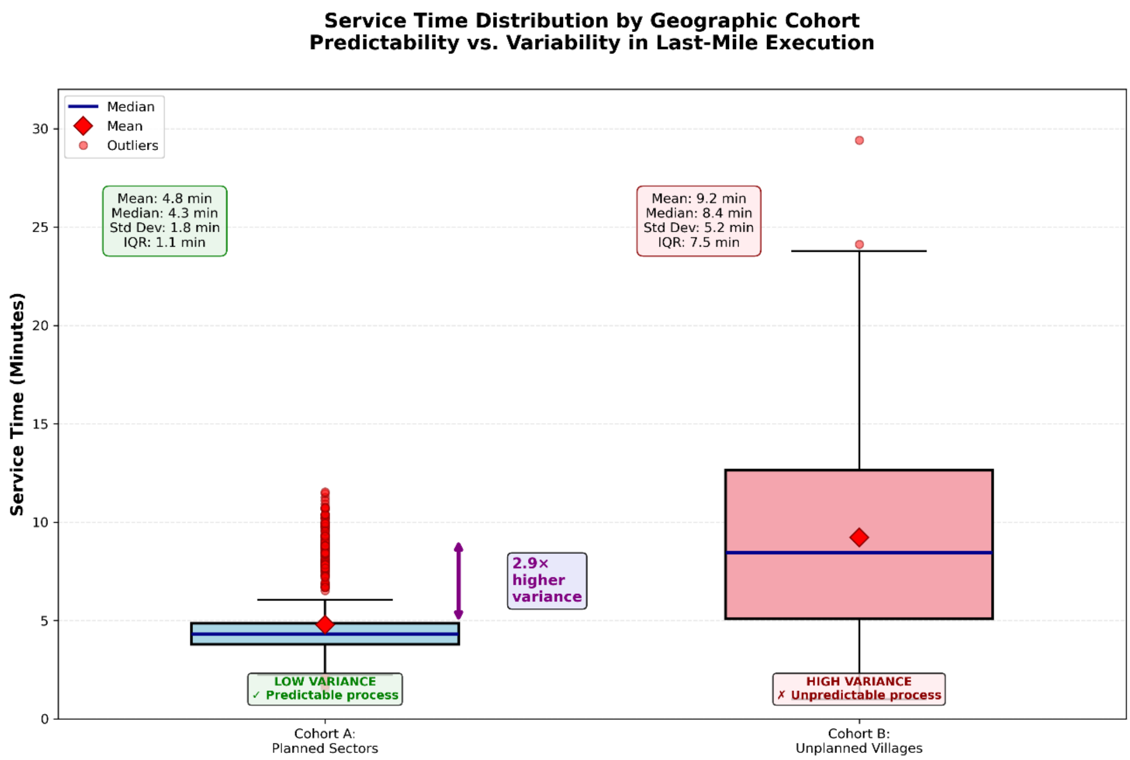


Figure 4.4: Service Time Distribution – Box Plot Comparison (Created using Flourish.studio)

4.4.5 Temporal Patterns: Time of Day Analysis.

There is a strong temporal FASR pattern of outcomes of delivery exercises per slot over five days that directly reply on the constraint of customer availability and Table 4.8 demonstrates this result.

Table 4.8: Success rate of Delivery with Time Slots.

Delivery Time Slot	Attempts (n)	% of Total	FASR	Variance	Avg. Service Time
Early Morning (06:00–08:00)	1,200	4.80%	91.20%	+14.2 pp	3.2 min
Morning (08:00–11:00)	5,200	20.80%	88.50%	+11.5 pp	3.8 min
Mid-Day (11:00–15:00)	11,500	46.00%	62.00%	-15.0 pp	4.2 min
Afternoon (15:00–17:00)	4,800	19.20%	81.20%	+4.2 pp	5.1 min
Evening (17:00–19:00)	2,300	9.20%	89.70%	+12.7 pp	4.8 min

The middle slot of 11:00-15:00 yields the lowest FASR of 62.0% -15 percentage points less than the mean - though covering 46% of all of the delivery volume. The slots that perform much better are in the early morning (91.2) and the evening (89.7): the hub has its customer base availability that leaves the house between 09:00 and 10:00 and returns between 17:00 and 19:00. The paradox of operations is that the time of the day when residents are least likely to be at home is when the most delivery volume is to be made - a dispatch schedule designed to organize around the operational convenience of the hubs as opposed to the customer availability, and can be directly countered by the dynamic time-slotting intervention of Chapter 6.

4.5 Customer Feedback Thematic Analysis

An analysis of 3,235 verbatim customer comments through thematic analysis according to program by Braun and Clarke (2006) protocol brought forth three major themes; Operational Dishonesty (fake attempts), Last-Mile Friction (access and doorstep disagreements) and Behavioral and Transactional Deficits. The complaints of these themes are proportionally distributed as shown in figure 4.5.

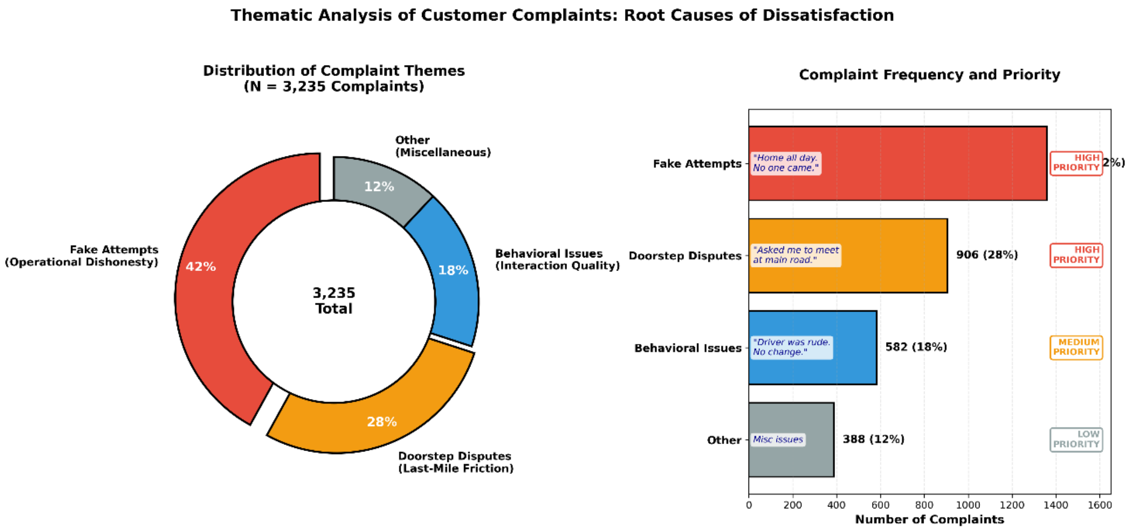


Figure 4.5: Distribution of Customer Complaint Themes (Created using Flourish.studio)

4.5.1 Theme 1: Operational Dishonesty -The Fake Attempt Phenomenon.

The most frequent one, with 42-percent share of complaints, deals with complaints by customers that delivery executives are not making a veritable physical effort but designate shipments as failed. This is the qualitative analog to the quantitative inconsistency in Section 4.3.3 and they are mutually reinforcing pointers of some systemic behavioral pattern. The next remark will be typical:

"I have been waiting at home all day for this packet. No one came, no one rang the bell. Then I received an SMS saying delivery failed. This is simply not true." (Respondent ID: RESP_1024, Sector 75)

This theme puts the operational soundness of the 55% CNA rate into question. When 42% of CNA records are challenged by customers who are claiming to be present, 21 percent of what the new data attributed to the customer behavior is really being described by the agent behavior. The physical barrier of struggling to pick the heavy parcels and entering narrow lanes of villages in certain villages at the end of the productive shift and facing a volume-based wage norm, delivery executives have a rational incentive to register a fictitious breakdown instead of spending time to make a tricky delivery under the current compensation framework . Chapter 5 elaborates the theoretical explanation of this misalignment of interest in the Principal-Agent Theory.

4.5.2 Theme 2: Last-Mile Friction: The Doorstep Dispute.

The second theme, which forms 28 percent of the number of complaints, is the expectation gap between the perception of the customers in terms of home delivery and physical conditions of infrastructure-limited implementation. One of the remarks made is the following:

"The delivery boy told me to come to the main road to collect my parcel. He said he could not bring his bike down the lane. Why should I pay for home delivery and then walk 500 metres?" (Respondent ID: RESP_089)

This translates the SERVQUAL service gap (Parasuraman et al., 1988) between the promised service standard of doorstep delivery expressed in the checkout and the perceived service delivered of main road handover of the vehicle at the sweat of the client. The structural friction lies in the nature of the Cohort B, with narrow lanes which are not accessible to vehicles, which, in most cases, does not allow delivering the motorcycle to the doorstep. It will not only be necessary to be resolved by finding a way to adjust operations, but also to communicate expectations at the time of ordering, which is a managerial consequence of Chapter 6.

4.5.3 Theme 3: Behavioral and Transactional Problems.

The third theme that makes up 18% of the complaint volume is related to the quality of the interpersonal and transactional interaction at point of delivery especially in COD scenarios. This theme is disaggregated in Table 4.9 in sub-categories.

Table 4.9: Behavioral Issues Theme Sub-Categories.

Sub-Category	Frequency	% of Behavioral (n=744)	% of Total Complaints	Representative Quote
Change Unavailability	312	41.90%	9.60%	"Bill was ₹490... no change... Unprofessional."
Rude/Unprofessional	267	35.90%	8.30%	"Driver was impatient, spoke rudely..."

Rushed Handover	98	13.20%	3.00%	"Threw package at gate and left..."
Language Barrier	67	9.00%	2.10%	"Could not understand what driver was saying..."
Total Behavioral	744	100.00%	23.00%	—

COD cash handling issue predominates at 41.9 percent of the behavioral complaints: arrival of the delivery executives with inadequate change to complete the transaction is a systemic breakdown of the pre-shift preparation. In the setting of the emerging market, the main human face of the e-commerce brand at the consumption point is delivery executives, unprofessional behavior harms not only the reputation of the logistics provider but also the brand equity of the e-commerce platform in the eyes of the customer, notes Goh et al. (2021).

4.5.4 The Service Quality Gap: A Synthesis.

The three qualitative themes define that the failure of delivery at the Noida hub is both a process integrity issue - systemic false failure capture - and an expectation management issue - difference between the promises of service and the physical reality - as well as an interaction quality issue - behavioral shortings at the point of handover. Such dimensions can not be seen by FASR and RTO metrics, and the presence of such dimensions suggests that the strategies based on optimizing routes or investing in technology would be inadequate. The redesign of the incentive system needs to be proceeded along with the standardization of communication protocols and alignment of expectations between e-commerce platforms and logistic partners to achieve sustainable improvement.

4.6 Concluding Findings Summary

The empirical analysis provides five key findings which form the evidencing background to Chapter 5.
The initial result already sets the mode of payment as a very important operational factor: COD shipments produce RTO rates with a 3.4-fold higher occurrence than prepaid orders (20.4% vs. 6.2) and use service time with 2.7 times greater frequency than prepaid

stops caused by the compounding influences of friction with cash exchanges, buyer remorse, and the unavailability of change at the point of delivery.

The second observation confirms the hypothesis of urban duality: the 17.1 percentage point FASR difference between Cohort A (84.2) and Cohort B (67.1) indicates that address ambiguity and access friction in the unplanned zone is associated with an efficient cost related to lack of efficiencies that cannot be addressed by investments in organization and technology.

The third result describes the Pareto concentration and integrity issue in failures: half of first instance failures are identified as Customer Not Available, yet the dispute rate contained in the CNA group provides evidence of a significant proportion of agent initiated false reporting based on incentives to report volumes as opposed to literal customer absence.

The fourth result indicates the differentiation of operational paradigms: planned domains are time-constrained: influenced by the availability of working residents and friction due to vertical access; unplanned villages are information-constrained: influenced by the ambiguity of addresses and by almost complete reliance on pre-delivery telephonic communication.

The fifth finding records the service quality disparity in three qualitative dimensions; credibility, delivery, and interaction; which quantitative measurements are unable to record and which involves non-algorithmic interventions to resolve systematically.

5 Discussion

5.1 Introduction

This chapter analyses the empirical results of Chapter 4 using theoretical constructs of Chapter 2, to synthesise the results of the statistical evidence and the qualitative feedback and explain the mechanisms behind the failure of last-mile delivery in the Indian urban environment. Although Chapter 4 recorded the fact - a 23 percent aggregate failure rate, a 17.1 percent point FASR disparity between infrastructural cohort and a 42 percent fake attempt dispute rate under the category of CNA - the present chapter deals with the issue of why these figures are so and what they do to show the relevance of current operations management theory to new market logistics. The discussion is also structured based on four theoretical understandings, which are followed by the examination of geographic determinism in logistics efficiency, service quality paradox his case to the Indian context, and the strategic justification of interaction optimization as the relevant solution to the detected digital-physical mismatch within the scope of the empirical analysis.

5.2 Theoretical Interpretation of Findings.

5.2.1 Institutional Void and Coping with Infrastructure.

The findings represent the most immediate empirical evidence on the institutional voids model of Khanna and Palepu (2010) since they confirm that the lack of a uniform address infrastructure is a hard bottleneck the operations with quantifiable financial implications. The 17.1 percentage point FASR difference between Cohort A (84.2) and Cohort B (67.1) is unexplainable by geographic distance - most of locations of Lal Doras village are geographically closer to the hub than planned sector addresses - but explainable by information quality precisely. In the developed countries, a street address like Flat 402 Tower A, Sector 76 will map precisely to a geospatial point. An address like Near Yellow Sweets Shop, behind big Peepal tree, Mamura in an unplanned settlement can only be provedable in the local knowledge, producing what this paper has characterized as address entropy, which is that operational uncertainty inherent in location information that can never be answerable by GPS algorithms.

Using the viewpoint of Activity-Based Costing, the cost to serve a customer in Mamura is about 1.4 times the cost in Sector 75 and this is purely due to search time and not distance. This measures the institutional void tax the cost of operation of logistics providers in the non-existence of the address standardization infrastructure that Western markets consider a social good. Table 5.1 remains a summary of the validation of each of the theoretical frameworks versus the empirical findings.

Table 5.1: Checking of the theoretical framework.

Theoretical Concept	**Theoretical Assumption**	**Empirical Finding (Noida Hub)**	**Validation Status**
Lean Management	Standardized inputs enable waste elimination	Unplanned zones lack standardized addresses	✗ Invalidated in unplanned
Route Optimization	Shortest path = Lowest cost	Search time (not distance) drives costs	✗ Partially invalidated
Theory of Constraints	Internal bottlenecks (capacity)	External bottleneck: Customer availability	✓ Extended to external
SERVQUAL	Speed is the dominant quality driver	Functional quality (empathy) > Technical (speed)	✓ Context-dependent
Principal-Agent Theory	Info asymmetry creates moral hazard	42% of complaints allege "fake attempts"	✓ Validated
Institutional Voids	Absence of institutions = higher costs	17.1 pp efficiency penalty in unplanned zones	✓ Quantitatively validated

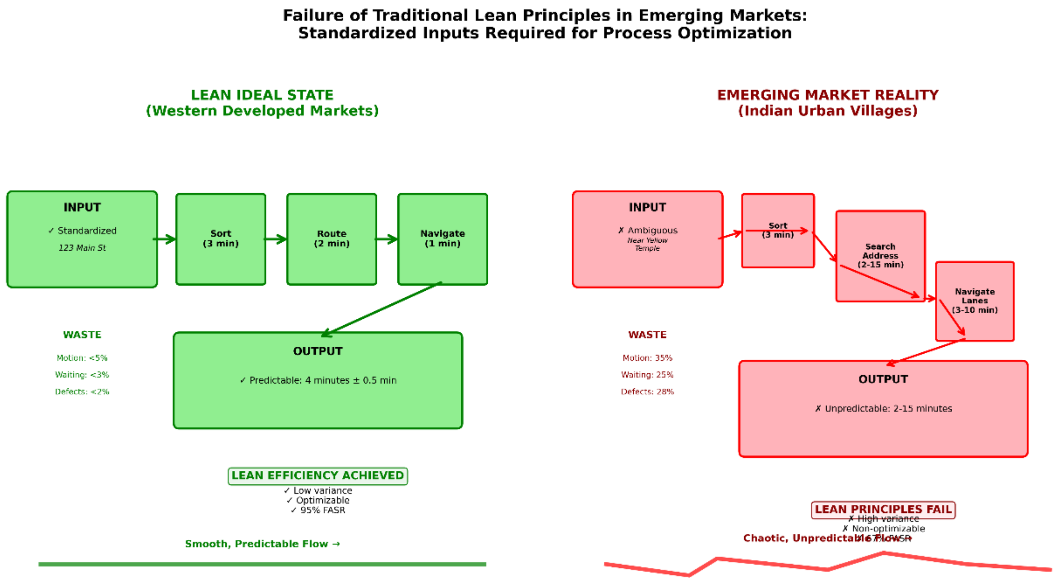


Figure 5.1: Lean Ideal State vs. Emerging Market Reality (Created using draw.io)

5.2.2 Lean Management: Limits to Waste Elimination.

Lean Management recommends that muda (motion waste), waiting waste and defect waste should be eliminated in mathematically planned place through the standardization of inputs and processes (Womack, and Jones, 1996). This prescription is based on a supposition that the data of this study is in direct conflict in the unplanned zone setting: the inputs are standardized enough that process-optimization can be of interest. In the second Cohort B, address terms are informational and non-geospatial and the service time taken in the same delivery assignment may be 2 to 15 minutes in difference, owing to the customer picking his/her phone and the accuracy of the description of his/her position. There is no way to make this variance go away by making the execution of the program run faster, or by setting up the route structure in any better way - this is inherently in the information contents of the input.

To the service setting, Hines et al. (2004) justify that the process of waste elimination should target informational as well as behavioral flows, as opposed to physical flows only. The results obtained in this study confirm this hypothesis: in the field of Indian last-mile, the border of Lean optimization has shifted to the uncertainty of the inputs. Sustainable efficiency improvement should have information stabilization as an remedy which is a case of address the pre-validation before dispatch and instead of process acceleration after departure.

5.2.3 External Constraint of Customer Availability: Theory of Constraints.

According to the Theory of Constraints, as put forward by Goldratt, (1990) system throughput should be marked by a single binding constraint of the system and that the system should be subordinated to the management of a binding constraint. TOC in traditional supply chain applications shows bottlenecks in internal capacity, sorting speed, vehicle availability and warehouse thoughput. The discovery that Customer Not Available is the cause of half of all first-attempt failures expands the conceptual closure of TOC in a theory of practical importance: the limiting force in B2C last-mile logistics lies out of the operational domain of the organization, in the dynamic and behavioural patterns of customer availability and not in any organizational process or capacity parameter. Efficiency of sorting, optimization of routes, fleet optimization cannot raise throughput when the binding constraint in the system is the presence of the customer at the door step.

The logic is operative in the temporal analysis in Table 4.8: when the time is in the Morning, FASR of 88.5% crashes to 62.0% in the middle-day, responding to the exact daily availability trend of the working-resident customer base of the hub. According to TOC, this constraint should rule scheduling of delivery, but instead of basing this around each hub, operational convenience, delivery schedules should be anchored around customer availability windows, which is directly reflected in the dynamic time-slotting intercession of Chapter 6.

5.2.4 The Phenomenon of Fake Attempt: The Principal-Agent Theory.

A study result that describes the highest behavioral dialogue is that 42 percent of customer complaints accuse alleged fake delivery attempts and it best can be explained by the use of the Principal-Agency Theory as formulated by Eisenhardt (1989). The lead company, the Delhivery, is interested in the largest delivery success rates in order to earn money and win trust of the customers. The agent the delivery executive is paid per packet and the bigger the packet the greater the disutility of work is: walking along narrow village lanes, using high-rise stairs, and completing time-consuming COD cash transactions toward the end of a shift, to be paid with communication of a hassle. In this compensation plan, the false Customer Not Available disposition of a problematic stop has short-term personal gain to time saved and energy expended, and no personal cost, because the geofencing technology can check the presence of the target in the general area but cannot ascertain whether a doorbell was rung or a phone called. This surveillance gap is the moral hazard which facilitates systematic false reporting.

This implication is that high CNA rate is not a customer behavior issue in the first instance but an incentive design issue. The compensation system which compensates according to number of deliveries made without any differentiation between successful and failed deliveries breeds the exact environment where rationality agents will always maximise the number of deliveries and not their quality.

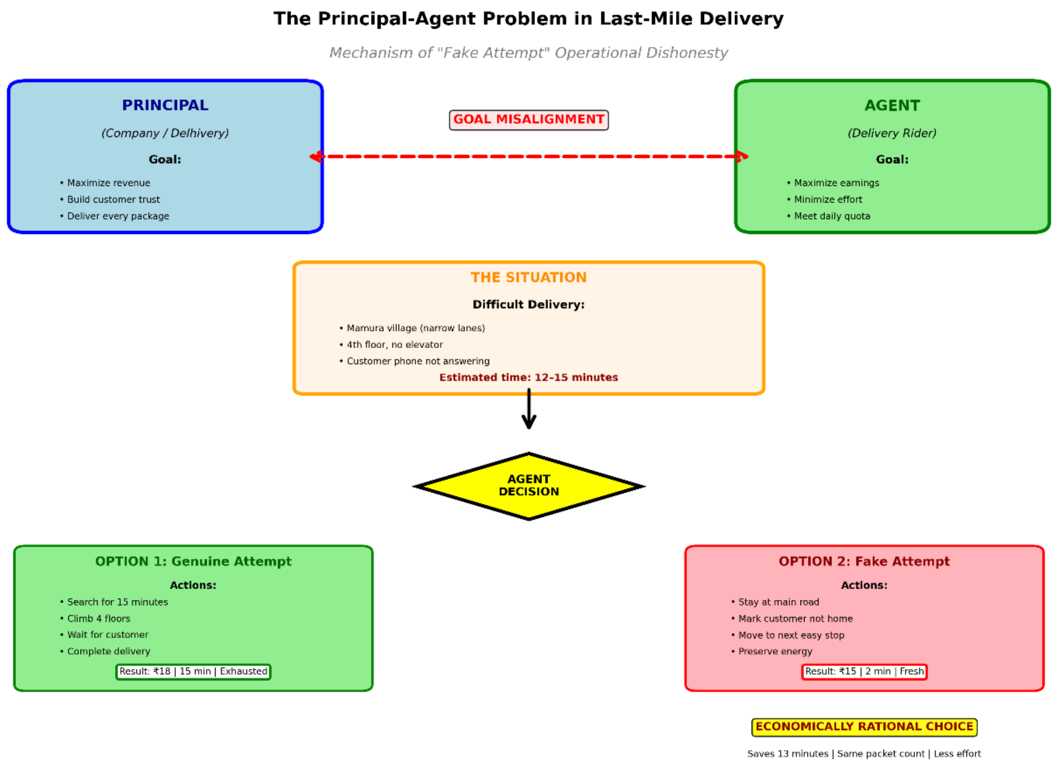


Figure 5.2: The Mechanism of Operational Dishonesty (Principal-Agent Problem) (Created using draw.io)

5.3 Geographic Determinism within Logistics Efficiency

The empirical data prove that the performances of the Noida hub area in terms of last-mile performance are geographically determined: the physical and informational features of the delivery area have a greater impact on the operation results compared to organizational resources of the logistics organization. This observation ushers in what this paper refers to as the logistics divide - the efficiency gap among infrastructure cohort that cuts across the same organizational input, technology, and fleet utilization.

5.3.1 The Pipeline vs. Mazes Operational Models.

The 17.1 FASR percentage gap between cohorts is a drastic difference between the two operational paradigms. Planned zones are modeled as pipeline: delivery is determined and linear, that is, standardized address entries, dependable GPS positioning, foreseeable services time of 4-5 minutes per visit. The main point of contention in the pipeline model is the administrative, that is, access management to gated communities, which is structurally stable and can be effectively managed via protocol. Unplanned zones work as maze models: delivery is a stochastic search process - fuzzy address input, GPS failure at the crucial last 50-100 metres and service time is between 2 to 15 minutes based on the success (or failure) of a telephonic consensus. The main conflict of the maze model is organizational - the lack of the credible location information - which is dynamic one and cannot be controlled solely with the help of physical process enhancement.

Such a difference is strategically vital in that it means that interventions that are working in one type of zone do not simply not work in the other, but actually are counterproductive. Classic route optimization (reducing length) is suitable in areas of pipelines in which the inputs are dependable. Optimism of interactions - maximizing the quality of pre-delivery information offered - is suitable in maze zone, when the physical proximity is of no use in the case the final address is not resolvable.

5.3.2 Divergent Failure Modes

The correlation shown in Table 4.6 between the rate of CNA and address failure rate among cohorts is negative; its theoretical interpretation is the following: the greater the quality of infrastructure, the greater the binding constraint will not be the locatability but the availability. Cohort A drivers do not struggle to locate the building, but as opposed to this, locating the resident home is a challenge to them. Drivers in Cohort B often are unable to solve the last 50-100 metres, it is not time but information they must cope with. What this divergence implies is that one common operational strategy cannot optimize on both zones at the same time. Both environments have structurally different constraints, which are only met by zone-differentiated operational models where time-of-day scheduling is the lever in the planned sector, and pre-delivery information enrichment is the lever in the unplanned village.

5.4 Service Quality Paradox

An interesting result of the CSAT data is that there seems to be a paradox to aggregate scores on customer satisfaction at 3.8 out of 5.0 even given the high level of operational friction and failure rates that had been recorded over the course of Chapter 4. This contradiction is solved with regard to close analysis of what customers in the two groups genuinely consider by rating their receiving experience.

5.4.1 The Effort Heuristic

Qualitative feedback data on the Cohort B residents reveals invariably that customers justify delays or hard access or even unsuccessful first attempt when the delivery executive made a real effort core demonstrated either by calling directions or by entering the lane on foot or describing his inability to locate the address in a transparent and understandable manner. This confirms the interactional justice aspect of SERVQUAL (Parasuraman et al., 1988) and the theory of functional quality as opposed to technical quality that Goh et al. (2021) hold regarding logistics in emerging markets. Not operational difficulty, but operational dishonesty is punished by the customers: the CSAT scores that drop to 1.0 or 2.0 are always linked with the complaints of fake attempts, when the dispositions are recorded without physical contact effort, and not real failures because of structural constraints. This suggests a new service quality hierarchy, which places integrity at the head and Empathy, then Reliability, and then Speed at the head instead of the very usual Western service dimension optimization ordering of first to last service dimension.

5.4.2 The Zone of Tolerance and Expectation Management.

According to Parasuraman et al. (1991), the definition of the zone of tolerance has been done as the variation between the desired level of service and the adequate level of service where there is acceptable variation by the customers without punishing the service provider. It implies that Cohort B residents, urban village clients, have a broader range of tolerance to delivery time and the difficulty of access compared to Cohort A clients, who live in gated planned areas and have adapted the so-called Amazon Effect (Bhattacharjya et al., 2016) as the standard of same-day delivery and a non-contact drop-off. Inhabitants of villages know the infrastructural difficulties of their locality and can not blame logistics deliverers who do not succeed with locatability issues when good effort is shown. They however punish very severely on false effort and incivility- factors that can be completely in the grasp of the agent without regard to the infrastructure. This managerial implication is relevant: in areas with a low infrastructure, preemptive message communication and open recognition of inability serve to retain customer satisfaction in a better manner than efforts to increase the speed of delivery.

5.5 Strategic Radical Change: Route to Interaction Optimization

The sum of the above theoretical weightings is that the scale and nature of strategic reorientation in last-mile logistics in the new markets are imperative. The existing paradigm of the industry - the rationality of the paths, the algorithmic reduction of the distance covered, has three assumptions that the Noida hub has been found to nullify: that the addresses are sorted geospatially, that the time of travel is the major cost driver, and that the service period is comparatively fixed. Addresses in the Indian city setting can often be indistinct, the search period in unplanned areas is longer in travel time and the variation in service time 2-15 minutes to perform same task rendering algorithmic optimization of directional meaningless.

Even the efficiency metric has to be redefined. Measuring the number of Deliveries Made per Kilometer Traveled rewards physical efficiency whilst being oblivious to those information and behavioural failures that produce most costs. The suitable measurement here is Successful Handovers per Total Time - the measure that accounts both search time and wait time and failed stop dead time with travel time and that that metric captures the rider incentives in line with organizational performance instead of just the bare volume throughput.

What this research is suggesting in its place is interaction optimization: a restrategic adjustment changing the optimization value to the maximization of the likelihood of successful handover at the moment of delivery. Routing algorithm under this paradigm also considers distance as well as historical FASR of each address, phone response rate of the customer, mode of payment, anticipated time of day availability and rating of address quality - geospatial or descriptive. Difficult deliveries are either pre-punished by a WhatsApp location share, or IVR approval till delivery, or profiled independently of the mainstream route to ensure that they do not negatively impact the performance of the routing overall. Pre-delivery intelligence Digital address verifications and availability assurances, conducted prior to the physical deployment of assets, decouple the flow of information and physical flow, and such that only assets sent to verified demand are deployed to the ground. The marginal cost of a digital pre-validation message (which is about [?]0.50) is insignificant in contrast to the joint cost of a failed delivery in fuel, labor, and reverse logistics ([?]55-[?]142 depending on type of failure), meaning that pre-delivery intelligence produces substantially positive returns even when it aversely affects only a small portion of existing failures.

5.6 Synthesis: The Digital-Physical Mismatch.

The general finding that comes out of this chapter is that Indian last-mile issue lies in a structural digital-physical incompatibility. The e-commerce ordering interfaces work on Silicon Valley standards of usability: One-tap checkout, real-time tracking, AI-based personalization, all the infrastructure on which orders need to be fulfilled is decades old: Streets without names, houses not numbered, and addresses passed down orally. This discrepancy creates a service promise effect: the online platforms promise service reliability on the scale of Amazon (even though the infrastructure of the delivery areas is not always that way), and the customers of off-planed zones build their expectation based on that promise and not on the technical limitations of the place.

The solution to this is not to give up on the digital promise but to contextualize it. Differentiation service level agreements by the logistics providers to unplanned zones should imply a realistic delivery window, like 48 hours, not same day, and also ensure predictability within this time interval. In the case where speed is structurally infeasible, transparency is the more effective source of leverage in the satisfaction- owning customers who are informed honestly of the difficulties of navigating their area, and that are given regular progress updates during the delivery process react cooperatively but not retaliatively. The service failure turns into an act of team problem-solving as opposed to an act of trust-breaking. This is the difference between a 3.8 and a 2.5 CSAT score when scaled at large scale.

6 Conclusions and Recommendation

6.1 Research Summary

The final leg of the modern supply chain is the most expensive and operationally challenging part, especially in the emerging economies where the heterogeneity of the infrastructure is not exceptional, but structural instead. This paper has studied the dynamic of the last-mile delivery failure using one case study of the Delhivery Noida Sector 63 hub, a problem driven by a central paradox to operational performance namely the simultaneous presence of a 23 percent first-attempt delivery failure score with investment in state-of-the-art address disambiguation and journey optimization tools. Through an inductive mixed-method design that involved a primary dataset (25,000 shipment records and 3,235 customer feedback logs) performed in Q3 2025, the study showed that the inefficiencies reported at the Noida hub were not accidents that occurred as a result of operational activities but as an institutional outcome of structural voids - the lack of standardized address infrastructure - and behavior adverse to organizational performance objectives and customer expectations. The key contribution relates to the reconstruction of last-mile optimization in the Indian urban environment: the claim that sustainable efficiency enhancement should take place through the shift towards interaction optimization, the systematic maximization of the likelihood of its successful implementation in a place of delivery.

6.2 Key Findings: Answering the Research Questions

Research Problem

How can last-mile delivery operations be optimized to reduce structural and behavioral challenges of the distribution networks in the Indian urban centers?

The study demonstarate that optimization in the new market reality should go beyond the framework of Vehicle Routing Problem and reach the informational and behavioral limits that come first before the actual delivery. This needs four strategic pillars: pre-delivery intelligence, which confirms the addresses information and customer availability until delivery, dynamic time slotting, which aligns delivery planning with patterns of customer availability instead of operational convenience of hubs, incentive restructuring as it provides incentive on successful delivery but not attempted delivery and proactive customer communication as it helps to manage expectations in low-infrastructure zones.

Research Question 1

What are the key reasons behind failures in delivering and customer dissatisfaction during the Indian urban delivery final delivery of the last 100 meters?

The aggregate failure rate of 23% can be explained by three different categories of friction. The structural friction on unplanned urban villages, however, is due to non-geospatial, descriptive addresses, which results in a locatability crisis where delivery executives pass through the last 50-100 metres although they are in the right neighbourhood. The statistics indicate that there is 28 percentable address unlocatable failure in Cohort B zones. Transactional friction are created because Cash-on-deliveries necessitate having to exchange cash physically, which creates service time 2.7 times that of prepaid stops and RTO rates that are 3.4 times more (20.4% compared to 6.2%). The most important finding is that of behavioral friction which is evident through systematic recording of fake attempts with 42% of Customer Not Available complaints contested by the customers who claimed to be at the delivery address. This commits incentive misalignment within the per-packet compensation model as a leading cause of the CNA metric as a representation of customer behavior to a compound measure of customer and agent behavior.

Research Question 2

How do operational performance and failure patterns differ between planned and unplanned urban delivery zones?

The comparative case studies on Cohort A (planned sectors) and Cohort B (unplanned urban villages) determines that the logistics performance in the Indian urban setting is geographically predetermined. This gap of 17.1 percentage points (84.2% in planned sectors and 67.1% in unplanned villages) cannot be justified by distance, fleet differences, but it can be justified solely by the quality of infrastructure and the information gap it produces. Importantly, the two groups of failures have a structural reason behind them: planned sectors are affected by the constraints of availability, 65% of failures in planned sectors were caused by CNA, and unplanned villages are affected by the constraints of locatability where 28% of failures in unplanned villages were caused by address unlocatability. Such a departure implies that no individual operating strategy can maximally serve both types of zones at the same time; zonedifferentiated strategies would have to be employed.

Research Question 3

What process interventions and soft optimization strategies can address the digital-physical gap and improve first-attempt success rates?

To offer solutions to the identified categories of frictions, four specific interventions are suggested, which are detailed in the Section 6.3: WhatsApp GPS pre-validation to eliminate address ambiguity prior to dispatch, dynamic time-slotting to minimize mid-day CNA failures, incentive restructuring to seal the Principal-Agent gap contributing to fake attempts, and automated IVR pre-call system to institutionalize synchronization with customers in the village zones.

6.3 Strategic Recommendations

The four suggestions are ranked according to the priority in the implementation, and they address the particular categories of friction that were determined in the empirical analysis. In the full implementation matrix, Table 6.1 will reflect the results.

Table 6.1: Matrix of Recommendation Implementation.

Priority	Initiative	Target Friction	Implementation Effort	Timeline	ROI
Quick Win	Incentive Restructuring	Fake Attempts	Low	0-3 months	8:1
Strategic Bet	WhatsApp GPS Pre-Validation	Address Ambiguity	Medium	3-6 months	10:1
Foundation	Dynamic Time-Slotting	Unavailability	High	6-12 months	5:1

Support	Institutionalized IVR Call	Sync Failures	Low	1-3 months	6:1

6.3.1 Incentive Restructuring: The end-of-delivery Quality vs. Number of Delivery reward.

The existing flat rate penalty per attempt of something like [?]15, whether successful or not, is a mathematically perverse incentive system whereby delivery executives get paid to do easy deliveries fast, and leave hard ones alone, with no penalty charged against false reporting of failure. The reform suggested is a success-differentiated structure according to which first-attempt-successful deliveries are paid 1.2 times the base rate ([?]18), second-attempt-successful deliveries 1.0 times ([?]15), real failed attempts 0.8 times ([?]12), and attempted disputed by the customer failed attempts 0.3 times ([?]4.50). The validity of the dispute classification is verified using GPS records of geofencing, obligatory doorbell photograph capture, and also by ensuring customer call was made within the delivery time. Executives can appeal with regard to a dispute classification.

It will have to be implemented over three months of parallel-running time where both old and new systems will run at the same time so that during this period the executives will be able to compare the earning potential with the new model to enable full migration before the end of the three months. The estimated cost will be a (15-20) percent decrease in the first quarter of fake attempts, which will accumulate savings of averted re-attempt expenditures and lower RTO rate approximated at [?]8-12 per shipment in high-failure zones. The month six onwards net positive financial payback will be possible with the annual network wide savings of [?] 15-20 crore. The legal protection of gig workers has compliance provisions that stipulate that a minimum safe payment of [?]500 per shift should be ensured, and no retroactive penalty imposed on previous performance data.

6.3.2 WhatsApp GPS Pre-Validation: Unplanned Zones Pre-Delivery Intelligence.

This intervention will address the 28 percent address unlocatable failure rate in Cohort B zones, which shift address verification upstream, that is, at the point when the delivery attempt has been made to the point when the mailing has been dispatched in the morning. At 08:00 on each delivery day, a whatsapp Business API purchased automated message sends to all the customers marked as scheduled to receive delivery in Cohort B pin codes: "Your package will be delivered today. To make you easy to locate, please give the information about where you are or a pin. Customers who reply via giving a GPS position provide the delivery executive with the capability to drive to with five metres of the actual location of delivery instead of trying to sort out the descriptive address of the fields. Combined with the Constellation routing algorithm, the shared GPS pins are turned into accurate routing waypoints prior to the executive leaving the hub.

It is expected that the address unlocatable failures in the unplanned zones will decrease by 20-25 percent, the average time of service to each village delivery will be shortened by 10-12 minutes, and the level of customer satisfaction will be improved based on proactive interactions. The cost of implementation will be estimated by [?]50,000-75,000 to integrate the WhatsApp Business API with the current CRM and 0.50 per outbound message. The cost benefit point is a 2 percent decrease in failure of addresses, which the expected 20- 25 percent decrease benefits reflect is a likely 10: 1 payoff. It is advisable to pilot in Mamura and Bishanpura in three months before it can be rolled out on the networkwide level.

6.3.3 Dynamic Time-Slotting: Dispatch Subordination to Customer Availability.

Section 4.4.5 has established that 46 percent of the total delivery volume is concentrated in 11:00-15:00 mid day window - the time when the reuring resident customer base of the hub is least likely to be at home and yield FASR of 62.0 percent versus 88.5 percent when in the morning slots and 89.7 percent when in the evening slots. This time imbalance is not unavoidable: this is one of the effects of dispatch scheduling structured based on the operational convenience of hubs as opposed to the unequal availability of customers.

The proposed intervention is based on a machine learning predictor to address, time slot and type of customers, then the proposed classification is done by classifying each delivery as residential and commercial and bin it within the time window where it is likely to be available within the highest probability of its availability. Offices and business addresses will be allocated 10:00 to 17:00; residential addresses will be allocated early the morning (07: 00-09:00) and in the evening (17: 00-19: 00). IVR confirmation which is discussed in the article under Section 6.3.4 involves dynamic rescheduling of routes, updating the routing plan locally by requiring customers to tell their unavailability. The expected results are that the overall FASR will improve by 10-15% and the mid-day CNA failures will decrease by 25-30%. This will require an implementation involving the Constellation algorithm, longer shift schemes to accommodate early, late deliveries and a shift differential allowance [?]50-100 against non-standard hours to ensure that the workforce collaborates. It should be implemented in stages: months one to three should have sector pilots on board, four to six months should have expansion to the mixed zone, and the entirety of the network should be implemented by the six-month mark.

6.3.4 Automated IVR Pre-Call: Pre-Doorstep Institutionalization of Communication.

Cohort B zones have recorded 92% dependency of successful delivery depending on pre-delivery telephonic coordination. The existing system is dependent upon the delivery executive making this call by hand, making it variable as to whether and when this call was made, and it left no communications trail that could be audited and refuted a fake attempt. The automated IVR system proposed will eliminate this variability by placing the customer into a call upon a geofencing activated call upon the executive performing the delivery entering into a two-kilometre radius of the delivery address. The automated message is sent out in Hindi and English, and it informs the customer of their pending arrival, that they are available and gives them an option of rescheduling in case of necessity. Users that confirm availability receive an SMS containing the name of a rider and his/her expected arrival time. Non-responsive customers receive a second call a 5-minute waiting time; continuous lack of response leads to dynamic route re-planning, which delays the delivery to next available slot.

The IVR system leaves a record of audible communication with each attempt at making a delivery which greatly lessens the information asymmetry where the attempt is recorded; it becomes operationally and evidentially difficult on the part of an executive to say that a customer was unavailable when the system did not record any outbound call during the delivery attempt. The cost of implementation is estimated as [?]25,000-40,000 integration of IVR platform with the tracking system, and per-call cost of an approximation of [?]0.40. The break-even point is 1% improvement in the failure due to better synchronisation, which compares to the 8-12% improvement that is to be projected, a 6:1 payoff.

6.4 Managerial Implications

This research has 4 managerial implications of the logistics operators and supply chain managers in the emerging markets.

The implication of the former is the need to be hyper-localized as opposed to standardized. Here, homogenous working tactics, i.e. same service intervals, same productivity levels, same per-packet charges, unnecessarily harm delivery executives in structurally harder areas. Managers are to consider zone-differentiating performance goals: a FASR of 85 percent is an ideal goal of the planned sectors, and a 70 percent realistic and honest goal of the unplanned village zones due to the prevailing infrastructure limitation. The productivity quotas must be set in a manner as to indicate the structural difference in the throughput capacity - A Mamura rider who is free-riding at 40 packets per six hours is riding with the same intensity of effort as a Sector 75 rider who is riding with 70 packets per six hours.

The second implication is the focus on technology investment. The 100 metres of delivery in India only is a matter of human, but not an algorithmic problem. The most effective areas engaged in technology are those that advance human communication, including WhatsApp integration, IVR systems, address disambiguation AI, those that do not seek to remove the human factor through autonomous vehicles or drone delivery, as such are structurally incompatible with the narrow, unpaved and densely populated lanes of the Indian urban villages.

Implication number three deals with redesign of the performance metrics. The traditional measures used in the industry, such as the number of deliveries in an hour, the cost per attempt, distances covered per liter, are blind to errors in the process results and quality of interaction that this research find to be key sources of operational cost and customer dissatisfaction. Control charts FASR by zone type, cost per successful delivery instead of a cost have been identified as the most operationally relevant, fake attempt rate as the dispute percentage of CNA dispositions is, and customer communication response rate as a leading indicator of the quality of synchronization.

The fourth implication is that of contextual service level agreements. Instead of a delivery window promise of same-day across all zones, logistics providers ought to offer differentiated SLAs, which are 24-hour delivery windows in areas with planned delivery and 48-hour delivery windows in areas with unplanned delivery. Research on customer psychology has repeatedly shown that when a longer time is promised and achieved in delivery compared to when a shorter time is promised and missed in delivery despite the actual delivery being slower. The candor concerning infrastructure constraints during checkout is not in place commercial liability; it has been a capability of satisfaction management.

6.5 Theoretical Contributions

The literature of operations management and supply chain management has four contributions as a result of this study. First, it measures the operational cost of institutional voids in logistics the 17.1 percentage point FASR penalty in unplanned zones is an empirical measure of the institutional void tax theorized and yet not operationalized by Khanna and Palepu (2010) in operative terms, which can be compared in future to the extent of operational cost of the institutional void found across emerging markets in South and Southeast Asia, Africa, and Latin America. Second, it applied the Theory of Constraints to B2C service application by showing how the binding constraint in attended last-mile delivery can be external, customer availability, and that adaptive management to buffers can then be switched to communication buffer that pre-distil customer availability prior to physical assets being mobilised. Third, it suggests a modification of Lean Management to unstructured inputs to the system, insisting that the Lean imperative of waste removal should first be enforced at the information layer, through address pre-validation, and customer coordination, prior to being enhanced with any meaning to the physical process level. Fourth, it offers empirical support of a situation-specific service quality hierarchy in emerging markets integrity, empathy, reliability, and speed in that order - inverting the priority scan of applications in Western SERVQUAL programs and claiming that service quality model must be culturally and infrastructural adjusted before implementation in non-Western logistics.

6.6 Limitations and Future Research

There are three limitations on the generalizability of the findings of this study. The single-hub form in Noida Sector 63, though indicative of Indian duality in cities, embodies a particular morphology: extreme vertical gated communities and dense informal settlements, which might be inconsistent with the Tier 3 towns horizontal urban sprawl, Tier 1 city linear coastal development, or the local cultural and infrastructural conditions of other new markets in Southeast Asia, Latin America and Africa. The Q3 2025 temporal scope indicates monsoon-period base operations and must exclude the festive season peaks of volume in Q4 where the volumes of delivery can be three times greater and the dynamics of failure may also vary significantly. The use of unsolicited verbatim customer feedback as a basis on which to perform thematic analysis creates a self selection bias in that highly dissatisfied customers are over-represented systematically compared to mildly or satisfied customers who do not leave behind feedback.

6.7 Future Research Avenues

This study produces five directions of research. Section 6.3 proposes the WhatsApp GPS pre-validation and IVR pre-call interventions which can be tested by longitudinal intervention studies using controlled before-and-after intervention over 12 months periods in order to have causal evidence of WhatsApp FASR improvements that are expected to take place within this study. A multi-city comparative analysis that recreates this methodology across Tier 1, Tier 2, and Tier 3 Indian cities would permit building a taxonomy of different types of urban infrastructure emerging markets and the operational playbook of the same. The human aspect of the operations of dense village routes, the psychological stress of a fake attempt incentive pressure, and the income curve of delivery executives on performance-based pay, would be done through research into gig worker wellbeing and retention, which quantitative metrics of performance cannot fully quantify. Further cross-cultural comparative studies that would take this methodology to logistics operation in southeast Asia, Sub-Saharan Africa and Latin America would identify India-specific and universally emerging-market sources of last-mile inefficiency. Lastly, the research on customer co-creation might also explore experimental studies, where customer service is required to respond to request-side addresses by active customer engagement in the address enhancing methods, such as conditioning the address on pre- delivery pin sharing at the GPS location, on landmark photos uploading, or voice record directions, as a demand-side complement to the supply-side interventions presented here.

6.8 Final Reflections

The Indian last mile is an operation paradox: the world-radical digital commerce system postpones an inadequate physical delivery system, which is in most urban regions older than the postal service. This work has demonstrated that the 67.1% FASR in unplanned delivery areas is not an execution failure but an execution failure - namely the failure to design optimization systems based on Western infrastructure assumptions that fail in the Indian urban location. Macro-logistics in India has been substantially addressed by technology, where automated gateways, algorithmic routing, and real-time location have ensured that the transfer of packages in the country is rapid, transparent and predictable. The last 100 metres via nameless lanes and over landmark-stepped doorsteps, the so-called micro-logistics is the unresolved issue.

This will not be closed by faster algorithms or even electric vehicles. The only thing that can be required is the paradigm shift, that is, the replacement of route optimization with interaction optimization, the treatment of the customer as the passive recipient with his/her active contribution to the delivery process, and the standardized service promises with the contextually honest communication as to what the physical infrastructure of each delivery area is capable and is not capable of providing. The model created in this paper pre-delivery intelligence, dynamic time-slotting, success-based incentives, and proactive customer communication offers a viable roadmap by which the operators of logistical services can navigate the structural complexity of the Indian urban last mile. As e-commerce proceeds to expand to the Tier 2 and Tier 3 cities of the Global South, the observational lessons on the narrow streets of Mamura and the gated towers of Sector 75 will be applicable much beyond Noida.

6.9 AI Tool Disclosure

This section identifies all the AI tools employed during the preparation of this thesis, versions of those tools where applicable, and the intended purposes of such tools in accordance with the University of Vaasa recommendations regarding the utilisation of the AI tools in academic research.

The generation of one satellite imagery comparison of Figure 3.1 of Google Gemini (Google, 2025) was created to represent the difference between Sector 76 of the planned residential area and the Mamura urban village based in Noida, Uttar Pradesh. The constraint, which was used, was: Compare, in a side-by-side satellite view, Sector 76 Noida (planned residential high-rise zone) to Mamura village Noida (unplanned urban village with dense informal construction) and make it clear what the difference is in terms of street layout, building density, and road width. In this picture, the disclosure of the tool, version, and prompt that was applied is listed in the thesis, according to the University of Vaasa citation guidelines of AI-generated visual content.

During the drafting, editing, data analysis, coding and interpretation of any other section of this thesis, there was no other AI tools that were applied. All the quantitative analysis was done through Microsoft Excel and Flourish.studio. Every qualitative thematic coding was conducted manually by the researcher according to the six steps protocol of Braun and Clarke (2006). The researcher used draw.io and Datawrapper and created all the process flow diagrams and conceptual framework figures. The main data sources were shipment log and customer feedback logs, which were accessed directly on the Delhivery internal ERP and CRM systems and are reported without the interpretation with the help of AI.

References

Agatz, N. A., Campbell, A. M., Fleischmann, M., & Savelsbergh, M. W. (2011). Time slot management in attended home delivery. Transportation Science, 45(3), 435–449.

Allen, J., Piecyk, M., & Piotrowska, M. (2018). An analysis of the last-mile delivery challenges in urban areas. Journal of Transport Geography, 74, 27–39.

Amazon Web Services. (2025). Delhivery achieves 160 ms latency for high-precision geocoding using Amazon EKS. AWS Case Studies.

Bell, E., & Bryman, A. (2007). Business research methods. Oxford University Press.

Bhattacharjya, J., & Tripathi, S. (2016). Evolution of e-commerce logistics. In Encyclopedia of e-commerce development, implementation, and management (pp. 1195–1212). IGI Global.

Bitner, M. J., Booms, B. H., & Tetreault, M. S. (1990). The service encounter: Diagnosing favorable and unfavorable incidents. Journal of Marketing, 54(1), 71–84.

Boulaksil, Y., & Belkadi, M. (2016). The impact of urban logistics on the bottom of the pyramid. Journal of Humanitarian Logistics and Supply Chain Management, 6(3), 354–376.

Boysen, N., Fedtke, S., & Schwerdfeger, S. (2021). Last-mile delivery concepts: A survey from an operational research perspective. OR Spectrum, 43(1), 1–58.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101.

CXO TV. (2022, May 26). Delhivery receives a patent for its proprietary technology product Addfix. TechPlus Media.

Dazeinfo. (2024, June 5). Over 25% of COD orders fail: A major dent in India's e-commerce business. Dazeinfo.

Delhivery Limited. (2023). Annual report 2022–23. Delhivery Investor Relations.

Delhivery Limited. (2024). Annual report 2023–24. Delhivery Investor Relations.

Delhivery Limited. (2025). Annual report 2024–25. Delhivery Investor Relations.

Edwards, J., McKinnon, A., & Cullinane, S. (2010). Comparative analysis of the carbon footprints of conventional and online retailing. International Journal of Physical Distribution & Logistics Management, 40(1/2), 103–123.

Eisenhardt, K. M. (1989). Building theories from case study research. Academy of Management Review, 14(4), 532–550.

Esper, T. L., Jensen, T. D., Turnipseed, F. L., & Burton, S. (2003). The last mile: An examination of effects of online retail delivery strategies on consumers. Journal of Business Logistics, 24(2), 177–203.

Flyvbjerg, B. (2006). Five misunderstandings about case-study research. Qualitative Inquiry, 12(2), 219–245.

Forgas-Coll, S., Palau-Saumell, R., Sánchez-García, J., & Caplliure-Giner, E. M. (2012). The effect of service recovery on customer satisfaction and loyalty in e-commerce logistics. International Journal of Logistics Management, 23(1), 125–146.

Fransoo, J. C., Blanco, E. E., & Mejia-Argueta, C. (2017). Reaching 50 million nanostores: Retail distribution in emerging megacities. CreateSpace Independent Publishing Platform.

Fundin, A. P., & Bergman, B. L. (2003). Exploring the customer feedback process. Measuring Business Excellence, 7(2), 55–65.

Geertz, C. (1973). The interpretation of cultures. Basic Books.

Gevaers, R., Van de Voorde, E., & Vanelslander, T. (2011). Characteristics and typology of last-mile logistics. Transportation Research Procedia, 125, 398–411.

Gevaers, R., Van de Voorde, E., & Vanelslander, T. (2014). Cost modelling and simulation of last-mile characteristics in an innovative B2C supply chain environment. International Journal of Retail & Distribution Management, 42(4), 298–311.

Goh, M., Ang, A., & Tan, T. (2021). Service quality in the era of last-mile logistics: A comparative study of emerging markets. Asian Journal of Shipping and Logistics, 37(3), 221–230.

Goldratt, E. M. (1990). What is this thing called Theory of Constraints and how should it be implemented? North River Press.

Goodman, R. W. (2005). Whatever you call it, it's the last mile that counts. Global Logistics and Supply Chain Strategies, 9(12), 48–51.

Google. (2025). Gemini [Large language model and image generation tool]. Google DeepMind. https://gemini.google.com

Griffin, A., & Hauser, J. R. (1993). The voice of the customer. Marketing Science, 12(1), 1–27.

Griffis, S. E., Rao, S., Goldsby, T. J., & Niranjan, T. T. (2012). The customer consequences of returns in online retailing: An empirical analysis. Journal of Operations Management, 30(4), 282–294.

Gunasekaran, A., Patel, C., & Tirtiroglu, E. (2001). Performance measures and metrics in a supply chain environment. International Journal of Operations & Production Management, 21(1/2), 71–87.

Gurobi Optimization. (2022). Delhivery: Making the last mile more efficient. Gurobi Case Studies.

Heskett, J. L., Jones, T. O., Loveman, G. W., Sasser, W. E., & Schlesinger, L. A. (1994). Putting the service-profit chain to work. Harvard Business Review, 72(2), 164–174.

Hines, P., Holweg, M., & Rich, N. (2004). Learning to evolve: A review of contemporary lean thinking. International Journal of Operations & Production Management, 24(10), 994–1011.

IBEF. (2024). E-commerce industry in India: Market size and growth. India Brand Equity Foundation.

Invest India. (2022). National Logistics Policy India. Government of India.

Jain, N. K., Gajjar, H., & Shah, B. J. (2021). Electronic logistics service quality and repurchase intention in e-tailing: Catalytic role of shopping satisfaction, payment options, gender and returning experience. Journal of Retailing and Consumer Services, 59, Article 102360.

Jain, T., Gupta, A., & Verma, M. (2021). Last-mile delivery operations in India: Challenges and the way forward. Journal of Global Operations and Strategic Sourcing, 14(2), 294–315.

Kaplan, R. S., & Norton, D. P. (1992). The balanced scorecard: Measures that drive performance. Harvard Business Review, 70(1), 71–79.

Khanna, T., & Palepu, K. (2010). Winning in emerging markets: A road map for strategy and execution. Harvard Business Press.

Lagorio, A., Pinto, R., & Golini, R. (2023). State of the art of urban logistics research: A systematic literature review (2018–2023). International Journal of Physical Distribution & Logistics Management, 53(4), 401–435.

Lim, S. F., Jin, X., & Srai, J. S. (2018). Consumer-driven e-commerce: A literature review, framework, and research agenda. International Journal of Physical Distribution & Logistics Management, 48(3), 308–332.

Lindner, M. (2011). Last mile logistics. In Supply chain management (pp. 1–15). Springer.

Malkoc, G., & Tuger, A. T. (2019). The impact of cash on delivery on e-commerce logistics. Journal of Transportation and Logistics, 4(1), 11–22.

Mangiaracina, R., Perego, A., Seghezzi, A., & Tumino, A. (2019). Innovative solutions to increase the sustainability of last-mile delivery: A literature review. International Journal of Physical Distribution & Logistics Management, 49(9), 901–920.

McKinnon, A. C., & Tallam, D. (2003). Unattended delivery to the home: An assessment of the security implications. International Journal of Logistics: Research and Applications, 6(1), 31–40.

Mentzer, J. T., Flint, D. J., & Hult, G. T. M. (2001). Logistics service quality as a segment-customized process. Journal of Marketing, 65(4), 82–104.

Mishra, S., & Vishwakarma, S. (2020). Decision support system for last-mile delivery in India: A case of address ambiguity. Journal of Indian Business Research, 12(4), 541–558.

MobiliseYourCity. (2023). City logistics plans guidelines: Making India's urban freight smart and efficient. MobiliseYourCity Partnership.

Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. Journal of Marketing Research, 17(4), 460–469.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A conceptual model of service quality and its implications for future research. Journal of Marketing, 49(4), 41–50.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. Journal of Retailing, 64(1), 12–40.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1991). Refinement and reassessment of the SERVQUAL scale. Journal of Retailing, 67(4), 420–450.

Parasuraman, A., Zeithaml, V. A., & Malhotra, A. (2005). E-S-QUAL: A multiple-item scale for assessing electronic service quality. Journal of Service Research, 7(3), 213–233.

Punel, A., & Stathopoulos, A. (2017). Modeling the acceptability of crowdsourced goods delivery: The role of delivery attributes and personal characteristics. Transportation Research Part E: Logistics and Transportation Review, 105, 128–143.

Ranieri, L., Digiesi, S., Silvestri, B., & Roccotelli, M. (2018). A review of last-mile logistics innovations in an externalities cost reduction vision. Sustainability, 10(3), Article 782.

Rao, S., Griffis, S. E., & Daugherty, P. J. (2011). Got to get it there: Analytics for the last mile. Journal of Business Logistics, 32(1), 123–142.

RedSeer Consulting. (2024). India e-commerce logistics report 2024: RTO rates, COD trends, and NCR benchmarks. RedSeer Strategy Consultants.

Resnik, D. B. (2015). What is ethics in research and why is it important? National Institute of Environmental Health Sciences.

Sampson, S. E. (2000). Customer-supplier duality and bidirectional supply chains in service organizations. International Journal of Service Industry Management, 11(4), 348–364.

Saunders, M., Lewis, P., & Thornhill, A. (2019). Research methods for business students (8th ed.). Pearson.

Savelsbergh, M., & Van Woensel, T. (2016). City logistics: Challenges and opportunities. Transportation Science, 50(2), 579–590.

Schroeder, R. G., Linderman, K., Liedtke, C., & Choo, A. S. (2008). Six Sigma: Definition and underlying theory. Journal of Operations Management, 26(4), 536–554.

Siggelkow, N. (2007). Persuasion with case studies. Academy of Management Journal, 50(1), 20–24.

Simatupang, T. M., Wright, A. C., & Sridharan, R. (2004). Applying the theory of constraints to supply chain collaboration. Supply Chain Management: An International Journal, 9(1), 57–70.

Tax, S. S., Brown, S. W., & Chandrashekaran, M. (1998). Customer evaluations of service complaint experiences: Implications for relationship marketing. Journal of Marketing, 62(2), 60–76.

The Business Research Company. (2024). Last-mile delivery market future demand and top key players analysis 2029. The Business Research Company.

Toth, P., & Vigo, D. (2014). Vehicle routing: Problems, methods, and applications. SIAM.

Wirtz, J., & Botwinick, V. (2019). Service revolution: Data and AI. World Scientific.

Womack, J. P., & Jones, D. T. (1996). Lean thinking: Banish waste and create wealth in your corporation. Simon & Schuster.

Yin, R. K. (2018). Case study research and applications: Design and methods (6th ed.). Sage Publications.