A Data-Driven Approach for Capacity Planning and Enhancing Courier Efficiency for an Online Food Delivery Business.

Azhar Mansoori

A Thesis in
The Department
of
Supply Chain & Business Technology Management

Presented in Partial Fulfillment of the Requirements
For the Degree of
Master of Supply Chain Management
at Concordia University
Montreal, Quebec, Canada

July 2025

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Azhar Mansoori

Entitled: A Data-Driven Approach for Capacity planning and enhancing Courier Efficiency for an Online Food Delivery Business.

and submitted in partial fulfillment of the requirements for the degree of

Master of Supply Chain Management

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

________________________________ Chair
Dr. Ahmet Satir

________________________________ Examiner
Dr. Ahmet Satir

________________________________ Examiner
Dr. Mahesh C. Sharma

________________________________ Supervisor
Dr. Satyaveer S. Chauhan

Approved by

____
Dr. Xiaodan Pan, Graduate Program Director

July 16, 2025

____
Dr. Anne-Marie Croteau, Dean of Faculty

Abstract

A Data-Driven Approach for Capacity planning and enhancing Courier Efficiency for an Online Food Delivery Business.

Azhar Mansoori

The rapid growth of e-commerce and on-demand food delivery platforms such as Uber Eats, DoorDash, and Meituan has significantly influenced consumer purchasing behavior. To meet rising demand, platforms often rely on either in-house or crowdsourced couriers. While crowdsourcing helps reduce logistics costs, it also introduces operational challenges, particularly around courier performance and behavior, factors that can directly impact delivery efficiency and customer satisfaction. With order volumes increasing at a fast pace, there is a growing need for data-driven strategies that can support better planning and resource management.

This study uses real-world data from the Meituan food delivery platform to conduct an exploratory data analysis (EDA) of courier behavior and performance. Key performance indicators (KPIs) examined include delivery time, distance traveled, courier workload, order acceptance rate, courier activity and inactive time, spatial-temporal delivery patterns, fulfillment rates, and average delivery time during peak and off-peak periods. In addition, several machine learning models: Linear Regression, Random Forest, XGBoost, LightGBM, K-Nearest Neighbors, and Support Vector Machine, are implemented to predict order volumes across different times and regions. These models are evaluated using standard error metrics, including RMSE, MSE, MAE, MAPE and R-squared. By integrating insights from both the EDA and predictive modelling, this study proposes data-driven strategies to enhance operational planning and efficiency.

Keywords: Machine Learning, On-demand Food Delivery, Crowdsourced couriers, EDA, Courier Performance, Spatial-temporal analysis

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

The surge of e-commerce and on-demand delivery services in recent years, especially following the COVID-19 pandemic, has significantly altered consumer buying behavior patterns. This change is primarily due to the widespread use of smartphones, technological advancements, and the fast pace of modern life, accelerated during the COVID-19 (Sureeyatanapas & Damapong, 2024). Last-mile delivery services, a critical component of the supply chain, focus on the downstream process of delivering products directly to consumers. This has resulted in Last Mile delivery services getting increasing attention in logistics academic literature, particularly, from on on-demand food delivery sector.

In the food delivery sector, platforms like UberEATS, DoorDash, and Meituan provide end-to-end services for both consumers and restaurants (Chu et al., 2023). These platforms tend to follow a similar work process: order is placed by the customers through the apps, restaurants which are integrated in the apps receive, accept and prepare the orders, platform apps automatically assign the order to a nearby available courier to complete deliveries. Furthermore, these platforms also offer features such as real-time order tracking, customer reviews, and flexibility to choose between pickup and delivery (Sureeyatanapas & Damapong, 2024). The primary goal is to deliver orders as quickly as possible typically within 15 min – 30 mins after the order is placed by the customers.

On-demand food delivery services have been growing rapidly. According to Allied Market Research, the global market, valued at $3,120 million USD in 2016, is projected to reach $16,605 million USD by 2023, with a compound annual growth rate of 30% (Sureeyatanapas & Damapong, 2024). As demand increases, food delivery platforms are compelled to enhance their efficiency. However, the major challenge related to last mile delivery services is their expensive logistics operation costs which comprises approximately 53% of overall shipping costs (Elsokkary et al., 2023). Last mile delivery platforms are committed to minimize delivery duration while reducing operational costs, but the increasing order volumes, geographical dispersed delivery regions, and uncertain external environmental conditions often slow down timely deliveries (Elsokkary et al., 2023).

Market competition has intensified with increasing growth in the number of these delivery platforms, this is driven by the low switching cost of customers (Mao et al., 2025). The customers can make quick decision in switching to alternate platforms based on various factors which includes, price, delivery speed, reliability and promotion offers etc. For instance, customer anticipate quick food delivery typically within 30 min, and this time sensitivity can greatly influence their future orders decisions (Mao et al., 2025).

The complex last-mile delivery operation has received significant attention from researchers. One common strategy employed by on-demand platforms is to utilize the services of crowd-sourced couriers to reduce delivery costs and achieve fast deliveries. Crowdsourcing logistics replicates the gig economy model, for example, the ride-sharing Uber app or the housing rental Airbnb, where these platforms serve as a marketplace to match service providers with consumers in real time.

Instead of recruiting in-house couriers, crowdsourced couriers are independent drivers who have the flexibility to choose their work schedules and either accept or reject orders assigned to them by the platforms (Pourrahmani & Jaller, 2021; Castillo et al., 2022). Platforms like DoorDash, Uber eats, Deliveroo and Just Eat employ a crowdsource delivery system. Similarly, in 2015, Amazon also adopted this system by introducing Amazon flex which allows independent drivers to deliver same day packages to their customers (Alnaggar et al., 2021).

Although this approach helps reduce logistics costs, it poses challenges related to the inconsistency in courier performance and schedule flexibility, potentially impacting overall operational efficiency (Castillo et al., 2022). These challenges serve as a major driving force to study the behavior and performance of crowdsourced couriers and how it affects operational efficiency particularly in on-demand food delivery services.

Despite the advantages of crowdsourcing, the literature has yet to thoroughly examine the impact of courier performance and behavior on overall operational efficiency. Experienced couriers are more adept at navigating complex routes, managing multiple orders, and meeting delivery deadlines, thereby enhancing customer satisfaction. Conversely, new or less experienced couriers often require more time for deliveries and are inconsistent with meeting service quality. While on-time delivery metrics provide partial insights about courier capability, it fails to capture the impact of behavior pattern on operational efficiency, therefore a deeper, quantitative analysis across multiple performance and behavior metrics is needed to understand these differences comprehensively.

This study aims to fill this gap by conducting extensive exploratory data analysis on the performance and behavior attributes of couriers. Key behavior attributes examine in this study are activity status, inactive duration and acceptance rate along with operational performance metrics such as delivery time, distance travelled, delivery capacity, on time delivery rate etc. This provides holistic understanding of on demand delivery operational efficiency and these insights can help with inform data driven decision making related to performance improvement, planning and training of workforces.

Another underexplored area in the context of on-demand food delivery is spatial-temporal analysis, which examines how phenomena vary across both space and time. In logistics, spatial-temporal analysis has emerged as a key approach for predictive decision-making through understanding patterns in both space and time. The methodology has been extensively applied in domains such as traffic demand forecasting (Huang et al., 2022), urban growth planning (Aljoufie et al., 2013), air quality prediction (Wang & Song, 2018), CO_2 emissions analysis (Chen et al., 2023), and tourism demand forecasting (Yang & Zhang, 2019). Spatial-temporal predictive analysis can assist in improving resource allocation and operational planning across diverse sectors.

However, the application of spatial-temporal analysis in the on-demand food delivery sector remains relatively limited. Existing studies have focused majorly on route optimization and planning (Peng et al., 2023; Laynes-Fiascunari et al., 2024), with little emphasis on predicting spatial-temporal order density (the distribution of order volumes across geographic regions and time intervals). This information is critical, as delivery demand in urban areas is highly variable and sensitive to factors such as time of day, day of the week, weather, holidays, and promotional campaigns.

With the recent advancement in computational tools, availability of rich high dimensional datasets and technology such as Machine learning, it is now possible to perform complex spatial temporal analysis on future orders. Machine learning can be defined as an automated algorithm which uses statistical tools or models to learn patterns in rich datasets (Masini, R et al…, 2023). Machine learning models can be used for classification or regression purposes. Classification models involve predicting categorical outcomes such as identifying whether a transaction is fraudulent or non-fraudulent. On the contrary, regression models are used for predicting or forecasting continuous numerical values such as stock prices.

The academic literature on the application of machine learning models in last mile delivery has mostly focused on route optimization and courier assignment (Chu, H., Zhang, W., Bai, P., et al. 2023), (Dieter, P., Caron, M., and Schryen, G. 2023) and (Jahanshahi, H. et al. 2022) and (Bozanta, A. et al. 2021). However, the study on using these machine learning models for spatial temporal analysis of order density is sparse.

Accurately predicting spatial-temporal order density can enable platforms to proactively position couriers, reduce idle times, improve delivery speed, and optimize customer satisfaction. This thesis aims to bridge this research gap by developing and evaluating predictive ML models that forecast food delivery demand across regions and time intervals using real-world data.

1.1.Research Objectives

This thesis contributes to the literature on last-mile delivery by:

1. Conducting a comprehensive exploratory analysis of courier behavior and operational performance using a range of behavioral and performance-related metrics.
2. Developing and Implementing predictive machine learning models for spatial-temporal analysis of order density in the on-demand food delivery sector.

The main research questions are:

- How can machine learning models be employed to predict order density across regions and times?
- In what ways can insights from courier behavior and performance analysis, combined with spatial-temporal demand predictions, enhance order allocation strategies in last-mile delivery systems?

1.1.Thesis Structure

The thesis is organized as follows:
Chapter 2: Literature review on courier performance and behavior analysis, spatial-temporal analysis, and machine learning applications in last-mile food delivery.
Chapter 3: Methodology outlining the research design and data analysis approach.
Chapter 4: Results from the exploratory data analysis and predictive modeling.
Chapter 5: Discussion with Managerial Implication, Limitations and Future Research Direction.

Chapter 2. Literature Review

2.1 Last-Mile Delivery Performance and Behavior of Couriers

Assessing the performance and behavior of couriers is essential for improving operational efficiency in last-mile delivery systems, particularly in on-demand and crowdsourced logistics. A courier's reliability, route decision-making, and willingness to accept tasks can significantly influence key performance metrics such as delivery time, accuracy, service completion rate, and customer satisfaction.

Crowdsourced delivery outsources last-mile logistics to independent couriers coordinated via digital platforms, enabling scheduling flexibility and market scalability but also introducing variability and unpredictability in performance (Elsokkary et al., 2023). This decentralized nature of crowdsourced delivery complicates workforce management, giving rise to challenges in balancing fulfilment capacity with fluctuating demand and ensuring consistent service quality (Savelsbergh & Ulmer, 2022).

The assessment of courier performance typically relies on either objective performance metrics, subjective evaluations, or a combination of both. Multiple studies highlight the need to evaluate courier performance through both objective and subjective measures. Objective data such as delivery time, idle time, distance traveled, and order acceptance rates are increasingly used alongside customer ratings and textual feedback to understand individual and system-wide behaviors (Çelik et al., 2025; Rajendran, 2020). The following literature review outlines key findings related to courier behavior and performance in the last-mile and crowdsourced delivery context, with the aim of identifying relevant indicators and research gaps.

Driven by the growth of e-commerce, especially post-COVID-19, companies are increasingly prioritizing KPIs such as speed, reliability, and service quality. Timotius et al. (2023) and Sureeyatanapas and Damapong (2024) emphasize that timely deliveries, route knowledge, and careful order handling contribute significantly to customer satisfaction. However, due to the lack of quantitative performance data, logistics firms often rely on subjective measures like customer feedback (Şimşek et al., 2013). This study seeks to address that limitation by incorporating real-world courier performance and behavioral data.

Objective assessments based on GPS and platform quantitative data such as delivery time, idle time, and distance travelled, are gaining attraction. Elsokkary et al. (2023) findings revealed that reliability, travel distance, and historical performance significantly impact delivery outcomes, with less experienced couriers being most likely to cause delivery delays. Furthermore, operational behavior such as order acceptance or rejection plays a pivotal role in system efficiency (Savelsbergh & Ulmer, 2022; Wu et al., 2022).

Despite the growing prevalence of crowdsourced delivery models, measuring courier performance remains challenging due to variability in skill levels among independent contractors. Shen et al. (2022) emphasized the importance of aligning tasks with courier attributes like experience and reliability to maintain service quality. In addition, customer ratings have been shown to provide useful performance signals, Ravula (2023) observed a strong link between timely deliveries and customer satisfaction.

Quantitative performance determinants affecting individual and system-level efficiency can be broadly grouped into three categories: courier reliability, spatial behavior, and delivery capacity. The uncertain availability of non-contracted couriers introduces volatility in delivery capacity (Savelsbergh & Ulmer, 2022). This has been extensively noted in literature, particularly in Ermagun et al. (2020), who point out that non-professional couriers introduce liability and trust concerns in crowdsourced systems. Moreover, Savelsbergh & Ulmer (2022) further distinguish crowdsourced logistics by emphasizing behavioral unpredictability, whether couriers accept assigned tasks and how long they remain available is often unknown. Çelik et al. (2025), using user review text mining, confirmed that service levels for non-professionals are highly volatile and frequently unreliable.

Studies by Chu et al. (2023) and Dieter et al. (2023) highlight that behavioral and contextual factors such as traffic and route deviation are inherently difficult to predict and can adversely affect operational efficiency. Another critical aspect of research is the influence of spatial coordination on couriers' attributes. Ermagun et al. (2020) observed that delivery outcomes differ substantially between urban and suburban regions. Similarly, Savelsbergh & Ulmer (2022) note that couriers tend to reposition themselves toward dense, high-demand zones to increase earnings. This behavior reflects strategic, self-interested decision-making based on spatial familiarity.

Several studies have shown that couriers frequently deviate from platform-recommended routes based on local knowledge or logistical convenience. Savelsbergh & Ulmer (2022), Lorenzo-Espejo et al. (2024), and Dieter et al. (2023) found that such deviations affect delivery performance, time windows, and customer satisfaction. Although this could sometimes be beneficial due to better local navigation, however, frequent deviations from recommended routes could potentially reduce operational efficiency, due to increased delivery length or imbalanced courier workloads.

Delivery fulfilment capacity refers to the number of available couriers relative to order volume, is another critical factor influencing performance. The irregular supply of part-time or occasional couriers makes it difficult for platforms to align resources with fluctuating demand, resulting in inefficient order assignments (Savelsbergh & Ulmer, 2022). Additionally, workload management for individual couriers plays a pivotal role. Lorenzo-Espejo et al. (2024) identified three key operational factors that correlate with service completion rates: courier workload, the frequency of service attempts in an area, and the distance from the warehouse to the delivery zone.

Existing last-mile delivery literature has identified and measured courier performance across different operational metrics:

Delivery Service and Fulfilment Rates: Chen and Chankov (2017) used an agent-based simulation model to evaluate delivery service levels and crowd resource engagement under fluctuating supply-demand ratios and courier acceptable detour level. The results revealed that higher detour willingness improves delivery service footprints; however, this could also reduce courier efficiency due to increased travel time and competition.

Courier Workload and Service Density: Lorenzo-Espejo et al. (2024) identified moderate negative correlations between courier workload and service completion rate in a B2C logistics context. Their findings stress the need for efficient route planning along with courier workload balancing to enhance performance.

Spatial and Environmental Factors: Ermagun et al. (2020) used nested logit models to investigate shipment performance across urban and suburban regions. Their study revealed that environmental elements such as road density and socio-demographic attributes like income and ethnicity significantly influence courier behavior and delivery success rate.

On-Time Delivery vs. Delivery Volume: Castillo et al. (2018) compared traditional and crowdsourced fleets under variable demand conditions. Traditional couriers achieved higher on-time rates, whereas crowdsourced couriers completed more total deliveries. Notably, acceptance rates below 75% negatively affect overall system performance, which emphasizes the impact of courier participation on delivery success.

Deviations from Planned Routes: Dieter et al. (2023) and Chu et al. (2023) found that couriers frequently deviate from algorithmically suggested routes based on local knowledge or experience. Integrating such behavioral patterns into route optimization models improved delivery efficiency and compliance.

Performance Analysis: Brochado et al. (2024) adopted a hybrid approach combining Multi-directional Efficiency Analysis with explainable ML techniques like SHAP and LIME. They benchmarked providers on delivery time and service costs and linked poor performance to geographic and temporal constraints.

Courier Behavioral Characteristics

Courier Motivation and Decision Factors: Bathke & Münch (2024) used Choice-Based Conjoint analysis to investigate which delivery attributes influence couriers' acceptance rate. Among 193 German crowd shippers, delivery time and compensation were the most significant factors, while parcel weight and recipient familiarity were deemed less important. The study also found that personality traits such as thoroughness or sociability influenced delivery decisions, underscoring the value of personalized task offerings.

Participation Willingness and Acceptance Behavior: Ermagun & Stathopoulos (2021) analyzed courier engagement across stages (bidding, acceptance, pickup, delivery) using hazard models on a U.S. dataset. They found that task posting time, delivery deadlines, and sender-courier relationships influences behavioral decisions at different stages.

Volatility and Service Inconsistency: Text mining studies by Çelik et al. (2025) and Rajendran (2020) uncovered customer concerns regarding delivery delays, poor handling, and inaccurate delivery tracking in crowdsourced systems. Although crowdsource delivery offers delivery flexibility and reduces costs, satisfaction often declines over time due to inconsistent service delivery.

2.2 Spatial-Temporal Analysis in Last-Mile Delivery

The surge in e-commerce, particularly following the COVID-19 pandemic, has pushed companies to continuously enhance their last-mile delivery operations to meet growing consumer expectations for faster deliveries. This shift has driven both businesses and researchers to delve into spatial-temporal analysis within on-demand service sectors.

Spatial-temporal analysis examines both the spatial (location) and temporal (time) aspects of data, helping to answer questions about where and when events occur (An et al., 2015). In last-mile delivery, spatial-temporal analysis provides insights that can help optimize delivery routes and improve courier assignment. While this field is still evolving in last-mile delivery, previous studies have applied similar techniques in sectors like food delivery, urban planning, logistics, and on-demand services such as ride-sharing.

Peng et al. (2023) introduced a Spatial-Temporal Consolidation (STC) technique to restructure last-mile delivery. Traditional FIFO (First-In-First-Out) methods are often ineffective, as it involves multiple trips to the same customer which leads to increasing delivery costs. The STC approach uses a three-phase heuristic strategy that clusters parcels by delivery dates and locations. This not only helps in reducing travel time and costs but also offers flexible delivery options, thereby enhancing customer satisfaction.

Spatial-temporal analysis literature has also explored its application for enabling strategic business decisions, such as site selection for distribution centers. Fried and Goodchild (2023) studied trends in the placement of Amazon's distribution centers across the U.S. from 2013 to 2021, revealing a shift towards locating facilities closer to densely populated areas to better meet the demand for faster deliveries. During the COVID-19 pandemic, Feizizadeh et al. (2023) employed GIS tools to measure the impact of spatial and demographic factors on the online food delivery sector. Their analysis identified regions with higher order densities, providing valuable insights for urban restaurants considering new locations to capture increased demand.

Studies have also focused on underserved areas, like food deserts, where delivery services face challenges due to low population density. Haider et al. (2022) proposed establishing grocery delivery hubs at local convenience stores, consolidating orders based on location and time to make deliveries more feasible in these areas. In the context of on-demand food delivery services, Wang et al. (2019) conducted a review of trajectory analysis in ride-sharing apps, investigating the influence of space and time on supply and demand dynamics. Their findings, though centered on ride-sharing, can inform improvements in last-mile food delivery, such as demand forecasting, order allocation, and route optimization.

Spatial-temporal analysis can also help tackle logistics challenges like safety and route planning. Lin et al. (2022) examined motorcycle accident patterns in Taiwan for both food and non-food deliveries, finding that food delivery accidents often occurred in areas with high restaurant density, while non-food delivery accidents were linked to intersections and high population density. Laynes-Fiascunari et al. (2024) further enhanced traffic prediction by integrating real-time data from social media with deep learning models, using LSTM networks to optimize delivery routes and minimize delays. Despite this, the literature on predicting food order density across space and time is sparse, an area this thesis explores in depth.

2.3 Machine Learning Applications in Last-Mile Delivery

The surge in e-commerce and urbanization has led to a significant increase in the use of last-mile delivery services in recent years. However, this rapid growth in dynamic demand has motivated practitioners to adopt new technologies which can enhance operational efficiency. Machine learning, due to its advanced learning, predictive and decision-making capabilities, has become a powerful tool for last-mile delivery firms to be used for optimizing delivery routes, assigning orders, predicting disruptions, and approximating service completion times, thereby improving overall efficiency. This review explores data-driven machine learning approaches deployed to tackle challenges in last-mile deliveries.

Machine learning models have attracted considerable attention from researchers in recent years. In the context of last-mile deliveries, particularly on demand food delivery services, most studies focus on optimizing delivery routes and courier assignments. Chu, H., Zhang, W., Bai, P., et al. (2023) deployed hybrid approach of integrating machine learning with optimization systems using a Smart Predict then Optimize (SPO) framework. Their study incorporated influential features such as traffic, weather, driver data, and historical travel times to predict travel times. These accurate predictions were then used to optimize courier assignments and route planning.

Dieter, P., Caron, M., and Schryen, G. (2023) also proposed a hybrid approach, integrating machine learning with traditional optimization models for route planning. This study addressed a gap in the literature by considering driver behavior in route planning, a feature often neglected in previous research. The machine learning models predicted the actual routes taken by drivers, and these predictions were incorporated into traveling salesman problem models to minimize deviations between suggested and actual routes.

Similarly, Jahanshahi, H. et al. (2022) and Bozanta, A. et al. (2021) employed deep reinforcement learning algorithms for courier assignment and route planning, aiming to reduce delivery times by giving couriers the flexibility to prioritize higher-probability restaurants and reject distant orders. Jahanshahi et al. (2022) also examined courier utilization by determining the optimal number of couriers needed per day.

Focusing on environmental sustainability and efficiency in last-mile delivery, Ramírez-Villamil et al. (2023) conducted a study on a delivery company in Paris, France, applying the K-means clustering algorithm to cluster delivery areas based on proximity and then optimize routes with a routing algorithm. The results showed a 22.6% reduction in total travel time and distance covered by couriers.

Pegado-Bardayo et al. (2023) addressed the issue of incomplete deliveries using machine learning regression models and clustering techniques. Couriers often struggle to finish all assigned deliveries within their working hours. Regression models were used to predict high-risk orders likely to be incomplete, and clustering techniques helped optimize routes by placing these high-risk orders towards the end, ensuring efficient completion.

In the realm of clustering techniques and deep reinforcement learning for optimization in last-mile delivery, Arishi et al. (2022) explored the use of truck-drone combinations for parcel deliveries. K-means clustering grouped delivery areas based on the number of drones per truck and their flight range. Deep reinforcement learning algorithms predicted the optimal delivery routes for these clusters, providing a faster and more efficient method of delivery, particularly in congested urban areas where traditional approaches may be slow or impractical.

Another critical challenge in last-mile delivery is capacity planning under uncertainty. Bruni et al. (2023) tackled this issue by predicting the optimal number of bins (containers) needed with third-party logistics (3PL) providers during uncertain demand and supply periods. Classification machine learning models were used to forecast the number of bins required, helping companies avoid overbooking or under booking, thus reducing costs and enhancing efficiency.

The application of machine learning in last-mile delivery has extended to digital marketing within this sector. Yaiprasert, C., and Hidayanto, A. (2023) applied ensemble learning techniques combining decision trees, naïve Bayes, and K-nearest neighbor (KNN) algorithms for consumer behavior analysis and AI-driven recommendations. The study achieved high prediction accuracy, with decision trees and KNN reaching 100% accuracy, while naïve Bayes achieved 97%. This highlights the potential of machine learning in optimizing marketing strategies through consumer analytics and personalized recommendations.

2.4 Research Gap and Contribution to the Study

Although existing literature has identified several meaningful attributes influencing courier performance and behavior, notable research gaps remain. Firstly, prior studies rely on simulation-based or machine learning modeling approaches that highlight the importance of key variables. However, these methods often lack clear, interpretable insights into how spatial, temporal, and behavioral patterns visually affect courier and delivery performance.

Secondly, while the literature extensively covers machine learning applications in courier assignments and route optimizations, there remains a research gap in the area of spatiotemporal predictive analytics, particularly for food delivery services. Spatiotemporal forecasting using machine learning models could accurately predict order density in specific regions at given times. Such insights could assist on-demand delivery services like Uber Eats and DoorDash in efficiently allocating resources by directing couriers to high-demand areas at optimal times, ensuring timely deliveries and better service quality.

Additionally, literature still lacks comprehensive real-world analyses of courier skill variability. Ermagun et al. (2019) and Chu et al. (2023) stress the importance of addressing behavioral unpredictability, particularly in spatially diverse environments. Ermagun et al. (2020) point to a fundamental gap in understanding how courier behavior influences delivery outcomes due to limited access to platform data.

Echoing this, Ermagun & Stathopoulos (2021) emphasize the need to integrate behavioral data into performance evaluation frameworks. They propose developing systems to monitor, guide, and possibly train underperforming couriers while also communicating expected performance levels to customers. Çelik et al. (2025) further call for in-depth investigations into service volatility, while Lorenzo-Espejo et al. (2024) advocate for targeted performance modeling based on workload, delivery distances, and area typologies.

Lastly, many studies are geographically limited and use datasets that combine both traditional and crowdsourced couriers, potentially obscuring platform-specific dynamics. Important operational metrics such as delivery delays, acceptance rates, courier inactive status, and efficiency during peak versus off-peak periods, are insufficiently explored. While some studies acknowledge the interrelationships between performance indicators, they do not explicitly examine how these metrics vary across different spatial zones or time periods. Managerial recommendations around route optimization and courier scheduling are frequently proposed, effective implementation requires a deeper understanding of order distribution across regions and times, a dimension that remains underexplored in current research.

This thesis addresses the identified research gaps by conducting a comprehensive exploratory data analysis (EDA) on delivery operations from the Meituan platform, a prominent Chinese shopping service that employs crowdsourced couriers for food delivery. The EDA is structured across three key dimensions: order-level performance metrics, courier wave-level activity, and temporal dynamics during peak demand periods. The primary objective is to generate clear, interpretable insights into courier performance and behavioral patterns that directly influence logistics efficiency and operational effectiveness.

Furthermore, the findings from the EDA are integrated with a machine learning–based spatiotemporal prediction model of future order density. This combined approach aims to deliver actionable, real-world managerial insights to support improved courier allocation, resource utilization, and scheduling strategies in dynamic last-mile delivery environments.

Chapter 3. Methodology

3.1 Research Design

This study adopts a data-driven and exploratory research design to investigate courier efficiency enhancement and location-time-wise demand forecasting in on-demand delivery systems. Given the operational complexity and dynamic variability inherent in last-mile delivery networks, a data-driven approach leveraging exploratory data analysis (EDA) and machine learning predictive models was deemed essential to uncover actionable insights and support operational decision-making.

The methodology employed addresses two primary research objectives:

1. **Exploratory Data Analysis:** To examine courier performance metrics and extract operational insights to inform better decision-making.
2. **Machine Learning Predictive Modeling:** To conduct spatial-temporal analysis for forecasting future order densities across time and regions.

3.2 Data Description

Data Source

This study utilizes the food delivery dataset provided by Meituan as part of INFORM TSL (Transportation Science & Logistics) data driven research challenge 2024-25. INFORM is a society or community which aims to provide students, researchers and practitioners a platform to collaborate and address challenges in the TSL field.

Meituan is a Chinese shopping platform which provides consumer products and retail services such as dining, delivery, entertainment and travel etc. As part of TSL data driven research challenge, the company has provided real food delivery dataset in order to address some of the suggested sample questions mentioned in the data access page. The data and background page including data description and sample questions can be found using the link here: https://github.com/meituan/Meituan-INFORMS-TSL-Research-Challenge/blob/main/TSL-Meituan%20challenge_background%20and%20data_20240321.pdf

This study aims to address the two of the seven sample questions mentioned in the research challenge and in the previous section as follows:

This study specifically focuses on two sample research problems outlined in the challenge:

1. Enhancing understanding of skilled courier behavior to improve order assignment strategies.
2. Estimating spatiotemporal order structures to guide real-time courier relocation.

Data Description

The dataset name “all_waybill_info_meituan_0322.csv” was used for this study. The dataset captures the full lifecycle of orders, from creation to delivery, across multiple operational and spatial dimensions. The dataset comprises 654,344 rows, with 568,547 unique order id.

The original dataset included a date field ("dt") indicating that the orders were recorded over a seven-day period from October 16, 2022, to October 24, 2024. However, during the feature extraction process, inconsistencies were identified between the "dt" values and other time-related features (e.g., platform order time, dispatch time). Consequently, the "dt" field was deemed inaccurate for analytical purposes.

To ensure temporal accuracy, a new "Order Date" feature was derived directly from the **"platform_order_time"** field, as detailed in the feature extraction section.

Key features in the dataset include:

1. **Order Features:** *Date, Order ID, Waybill ID and is_prebook.*
2. **Courier Feature:** *Courier ID and is_courier_grabbed.*
3. **Time Feature:** *Platform order time, estimated arrival time, estimated meal preparation time, order push time, dispatch time, grab time, fetch time, arrival time and is_weekend.*
4. **Spatial Features:** *Sender Lat/Lng, Recipient Lat/Lng and Courier Lat/Lng, Area and Poi_id.*

Data Cleansing
To ensure the quality and consistency of the dataset:

Duplicate Removal: Orders initially assigned but rejected by couriers (identified via waybill IDs) were removed. Only records corresponding to accepted waybills were retained.
Final Dataset: 568,547 unique order IDs post-cleaning.
Missing Values: The dataset contained no null values.
Handling Rejections: Rejected orders had incomplete time-related information and were excluded from analysis to avoid introducing biases.

3.3 Data Preparation and Transformation

The dataset originally consisted of 23 features encompassing information related to orders, couriers, temporal aspects, and spatial dimensions. However, the spatial and temporal features were not initially in a format suitable for analysis or modeling. Consequently, feature transformation and additional feature extraction were conducted to standardize the data and enhance its suitability for machine learning applications, thereby improving the models' ability to capture complex patterns and make accurate predictions.

Geographical Coordinates
Original spatial coordinates were stored as large integers rather than standard micro degree format. Spatial fields for restaurants, customers, and couriers were divided by 10000000 to correctly represent latitude and longitude values as shown in Table 1 and Table 2 below:

Table 1. Current Format of Spatial Features.

Sender lng	Sender lat	Recipient lng	Recipient lat	Courier lng	Courier lat
1.75E+08	45905850	1.75E+08	45898250	1.75E+08	45906005

Table 2. Micro Degree Format of Spatial Features.

Sender lng	Sender lat	Recipient lng	Recipient lat	Courier lng	Courier lat
174.52993	45.90585	174.523	45.89825	174.5301	45.90601

Timestamps

The dataset contained several time-related features, including platform order time, estimated meal preparation time, dispatch time, grab time, fetch time, estimated arrival time, arrival time, and order push time. All of these features were recorded in Unix timestamp format, representing the number of seconds elapsed since January 1, 1970 (UTC).

To enable meaningful temporal analysis, it was necessary to convert these timestamps into human-readable datetime format. This transformation was performed using Python's datetime library. An example illustrating the conversion of the "platform order time" feature from Unix timestamp to standard datetime format is provided in Table 3 below.

Table 3. Date Time Format for Platform Order Time.

Feature	Before	After
Platform Order Time	1665935995	10/16/2022 3:59:00 PM

Feature Extraction

Although the dataset had 23 features related to orders, couriers, spatial and time, however, for machine learning modelling purposes particularly for spatial temporal analysis of order density and exploratory data analysis on courier performance, 11 new features were extracted from the current available features. This was done in order to enhance model training and exploratory analysis. The original dataset lacked key operational and behavior metric of couriers. These engineered features, summarized in Table 4, were essential to capture performance variability and behavioral patterns among couriers, thereby enhancing both model accuracy and the depth of exploratory insights.

Table 4. Description of Engineered Features

Features Extracted	Description	Example and Format
Order Date	Date when order was created	Date: 10/16/2022
Day of Week	0 = Monday... 6 = Sunday	Numerical
Delivery Time	Time in minutes taken to deliver the order	Minutes: 17.68 min
Distance	Delivery distance	Kilometer 3.49 km
Recipient Area	Customer's area ID	Numerical: 3, 4, 5….
Is Cross Area Delivery	Whether pickup and delivery areas differ	Binary: 0 (No) / 1 (Yes)
Ordered Areas	List of areas ordered from	List: [0 4 2 9]
Number of Unique Areas	Number of distinct ordered areas	Numerical: 3, 4, 5….
Delivered Areas	List of delivery areas	List: [0 4 2 9]
Number of Unique Delivered Areas	Number of distinct delivered areas	Numerical: 3, 4, 5….
Area - Focus	Indicator of specialized or multi-area operation	Binary: 0 (Specialized) / 1 (Multi Area)

3.4 Spatial Temporal Modelling

Objective: The objective of the spatial-temporal modeling was to forecast future order density across different time periods and geographic areas using supervised machine learning models. Reliable prediction enables dynamic courier allocation strategies, improving operational efficiency during peak and off-peak periods.

Machine Learning models used in this study is Linear Regression, Random Forest, XGBoost Regressor, Light GBM Regressor, K-nearest neighbor (KNN) and Support Vector Machine (SVM). Each of these models are described briefly below:

1. **Linear Regression**

Linear Regression is a parametric statistical method to construct the best fit line between the independent and dependent variables. This is normally done using linear least square method. This model is a simple and easy to interpret, however, the model assumes statistical linearity between the dependent and independent variables.

This model will act as the baseline, which will try to capture the best fit line between the between predictors and target variable, however, given the complexity of on demand delivery dataset, it is expected to be non-linear. Additionally, linear regression models are sensitive to the presence of outliers which means they can significantly distort the best fit line thereby affecting model performance and accuracy.

2. **Random Forest**

Random Forest is an machine learning algorithm, proposed by Breiman, with multiple decision trees (Yeşilkanat, C. M, 2020) and (Tongtian Zhu, 2020). A decision tree, also called a regression tree when used for predicting numbers, is a model that estimates a target variable by repeatedly splitting the data into smaller groups based on the values of the input features (Czajkowski & Kretowski, 2016).

The structure of a decision tree includes a root node (where the process starts), decision nodes (where the data is split based on specific features), and leaf nodes (where the final prediction is made). Each leaf node represents the average value of the target variable for that group of data (Czajkowski & Kretowski, 2016). At every step, the model evaluates how well each feature splits the data by calculating the prediction error (the difference between the predicted and actual values) (Pekel, E, 2019). The feature that minimizes this error is chosen for the split (Pekel, E, 2019). This process continues recursively until the tree reaches its stopping point (Pekel, E, 2019). In simple terms, a decision tree works like a flowchart: it starts by asking a question about one feature, then moves down different branches depending on the answer, eventually arriving at a final prediction.

Random Forest is an non-linear model which utilizes multiple decision tree to capture complex pattern between the independent and dependent variables (Gül, V. 2023). This algorithm can be used for both classification and regression tasks. The algorithm works by training each decision trees on randomly selected data points and features using bootstrap sampling, the final prediction is made by taking the average of individual tree prediction. (Yeşilkanat, C. M, 2020).

Random Forest algorithm demonstrate strong performance compared to other machine learning models, having higher accuracy (Tongtian Zhu, 2020). The model is computationally efficient, which means it is capable of handling high-dimensional dataset and training at a faster rate. Additionally, being a nonlinear model, the algorithm does not need to assume statistical linear assumptions between the dependent and independent features (Tongtian Zhu, 2020).

However, due to having multiple decision trees, the model gets complicated and challenging to interpret the relationship between the dependent and independent variables (Tongtian Zhu, 2020). Although, random forest is robust to outliers, but the algorithm can still be affected by it. A balanced dataset is recommended to use random forest regressor for prediction (Tongtian Zhu, 2020).

3. XGBoost Regressor

XGBoost or Extreme Gradient Boosting Regressor is an enhanced version of gradient boosting model, which offers faster solution and capable of handling high dimensional dataset (Gül, V. 2023). Gradient Boosting algorithm is an ensemble technique which uses the prediction of multiple decision tree models (weak learners), sequentially combine then and improve overall model prediction by optimizing the weights of individual models based on the errors made by the previous models (Sekeroglu et al, 2022). This process is repeated iteratively to minimize the loss or prediction errors and improve model's accuracy.

XGBoost incorporate both first order (gradient) and second order (Hessian) derivative of loss function compared to gradient boosting which only incorporate first order (Kang et al, 2021). The algorithm also tries to adjust the complexity of model to prevent overfitting (Kang et al, 2021). This means that XGBoost does not only look for the how much prediction deviates from actual value but also looks for the change in slope.

XGBoost are not only fast and accurate when compared with other machine learning models like Random forest and Linear Regression, but the model can process high dimensional dataset and avoid overfitting (Shehadeh et al, 2021). However, XGBoost achieve reliable prediction results with large dataset size and are not recommended to use with categorical features (Larsson et al., 2021). Manual encoding of categorical features is required when implementing XGBoost algorithm.

4. Light GBM Regressor

A highly efficient machine learning model, used for large dataset, aiming for faster training while using less memory (Gül, V. 2023). Light GBM is also another type of gradient boosting tree algorithm which is fast, efficient, improved accuracy and capable of dealing with large dataset multiple dimensions. It uses a histogram-based approach where the continuous parameters are stored in histogram bins which allows for increasing model training speed and less memory usage (Xu et al, 2023).

Unlike traditional tree models which uses level wise tree growth, Light GBM uses leaf wise tree growth. The leaf wise tree split allows for greatest gain of information which allow for reduce loss thereby improving prediction accuracy (Xu et al, 2023). Additionally, Light GBM models supports categorical input features directly without the need of pre-processing using one-hot encoding technique (Xu et al, 2023). However, one limitation of Light GBM model is the risk of overfitting (Ileri, K., 2025). Overfitting occurs when the model learns the training data too well, such that it affects its ability to make accurate predictions on unseen data.

5. **K-Nearest Neighbor:**

K-Nearest Neighbors (KNN) is a non-parametric, supervised learning algorithm used for both classification and regression (Luo et al., 2019). It works by calculating the distance (typically Euclidean) between a new input data point and all existing data points in the training set. The algorithm then selects the *k* closest data points and uses their target values to make a prediction (Halder et al, 2024). In regression, it returns the average of the neighbors' target values whereas in classification, it returns the most frequent class label among the neighbors (Halder et al, 2024).

This model can efficiently deal with complex dataset structure, have higher accuracy and are suitable for short term forecasting application e.g. traffic (Cai et al., 2016). However, selection of optimal *K* is crucial as it can affect model performance accuracy and cause overfitting (Halder et al, 2024). Additionally, with large dataset, the model can be computationally intensive (Halder et al, 2024).

6. **Support Vector Machine (SVR):**

A popular supervised learning machine learning model that can be used for both classification and regression purpose, the model has ability to deal efficiently with both linear and non-linear time series dataset (Liu et al., 2022). The fundamental idea behind SVR model is to solve regression problem by mapping the nonlinear original input features into a higher-dimensional feature space using Kernel function in order to construct a best fit straight line (Üstün et al, 2007).

SVR is suitable for solving problems with smaller sample sizes, can deal with high-dimensional datasets, and capture non-linear relationships. However, as the training sample size increases, its computational complexity also increases (Huang et al, 2022).

Table 5. Comparison of Machine Learning Models

Models	**LR**	**RF**	**XGBoost**	**Light GBM**	**KNN**	**SVM**
Strength	Easy to implement and interpret	Higher accuracy, computationally efficient and can handle higher-dimensional datasets (Tongtian Zhu, 2020).	Higher accuracy and faster than Random Forest, can also handle higher-dimensional and imbalanced datasets (Shehadeh et al, 2021)	Much faster, efficient and accurate compared to XGBoost and can also handle categorical variables in the dataset without the need for pre-processing (Xu et al, 2023).	Higher accuracy and suitable for space-time data (Cai et al., 2016)	Suitable for small to medium datasets. Just like RF and XGBoost, can handle high-dimensional datasets and capture non-linear patterns wells (Liu et al., 2022) and (Üstün et al, 2007).

Weakness	Sensitive to outliers (Yu & Yao, 2017) and assumes linear statistical assumptions	Difficult to interpret and requires balanced dataset (Tongtian Zhu, 2020).	Unlike LightGBM, manual encoding of categorical feature is required, (Larsson et al., 2021).	Overfitting risk (Ileri, K., 2025).	The increasing size of the dataset can make it computationally intensive. Optimal K value needs to be selected for higher accuracy and reduced overfitting overfitting (Halder et al, 2024).	Increasing computational complexity as the data size increases (Huang et al, 2022).

While the application of machine learning for spatiotemporal analysis in on-demand food delivery remains relatively underexplored in academic literature, similar predictive modeling approaches have been extensively studied in other domains. These include areas such as air quality monitoring, housing price forecasting, traffic flow prediction, and environmental analysis, each characterized by complex spatial and temporal dynamics that closely resemble the challenges in last-mile delivery systems.

Support Vector Machines (SVM) have shown robust predictive performance in air quality forecasting tasks (Liu et al., 2022). In the traffic flow prediction domain, K-Nearest Neighbor (KNN) has emerged as a particularly effective approach, as demonstrated by studies such as Luo et al. (2019), Cai et al. (2016), and Cheng et al. (2018). Likewise, tree-based models, including Random Forest, Gradient Boosting, and XGBoost, have been applied successfully in predicting environmental variables (Amato et al., 2020), traffic accidents (Al-Dogom et al., 2019), housing prices (Soltani et al., 2022), and electric vehicle charging behavior (Ge et al., 2020).

The widespread use of these models across diverse domains can be attributed to their effectiveness in capturing nonlinear relationships within high-dimensional datasets. This reinforces their methodological suitability for last-mile delivery applications, where data is not only large in volume but also inherently spatiotemporal in nature.

3.4.1 Input Features

The independent variables used for spatial-temporal prediction of order density encompassed both spatial and temporal dimensions associated with orders.

To ensure temporal consistency, a new Order Date feature was derived from the platform order time attribute. This adjustment was necessary because the original "dt" field in the dataset exhibited inconsistencies and did not align with other time-related features (e.g., platform order time, dispatch time, grab time) once converted from Unix timestamp to standard datetime format.

Following the creation of the Order Date feature, additional temporal attributes were extracted:

Day of the Week: Numerical representation of the weekday (0 = Monday, ..., 6 = Sunday).
Is Weekend: Binary indicator distinguishing weekends from weekdays.
Hour of Order Creation: Extracted directly from the platform order time, representing the hour at which each order was placed.
Day Part: Numerical representation of different periods of the day [Morning, Afternoon, Evening and Night].

For spatial analysis, while geographical coordinates for restaurants, customers, and couriers were available, the dataset also included an Area feature. This feature grouped restaurant locations into predefined regional clusters, similar to clustering techniques such as K-Means. Given that the Area feature sufficiently captured spatial segmentation, additional clustering using geo coordinates was deemed unnecessary. Moreover, since the primary modeling objective focused on predicting order density from the restaurant location perspective for courier allocation optimization, utilizing the Area variable was both sufficient and appropriate.

Summary of Input Features for Modeling:

Temporal Features: Hour of the day, Day of the week, is weekend indicator, Is Pre Book and Day part
Spatial Feature: Area

3.4.2 Target Feature

The target variable for the spatial-temporal modeling was order count, defined as the number of orders placed within specific time and spatial intervals. As the original dataset did not explicitly include an order density or order count feature, it was necessary to derive this variable.

The target feature was constructed by grouping the order IDs based on order date, area, and hour, resulting in a new dataset that captured the number of orders received in each area at each hourly interval for every day in the dataset.

3.4.3 Data Transformation

To ensure that the trained models do not suffer from underfitting, overfitting, or biased outcomes, visualizations were conducted to examine the distribution of both input features and the target variable. This helped assess whether the data was balanced and free from significant skewness. The input features exhibited a reasonably uniform distribution, making them suitable for modeling. However, the target variable, “Order count,” displayed a noticeable right skew, as illustrated in the Figure 1 below:

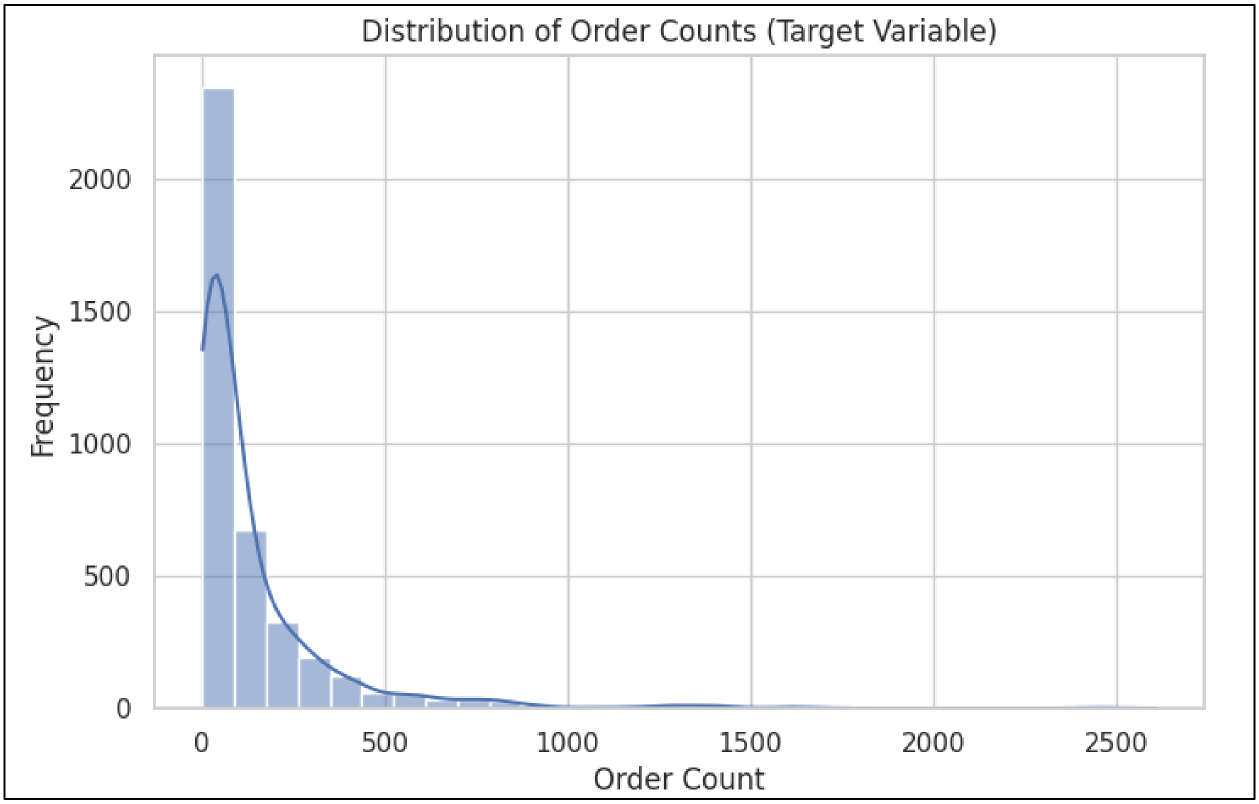

Figure 1. Distribution of Order Count

To reduce prediction bias and ensure a more balanced representation of the target variable, a log transformation was applied to the "Order Count." This transformation helped normalize the distribution, preventing the model from disproportionately favoring majority values. The post-transformation distribution of "Order Count" is shown in Figure 2 below:

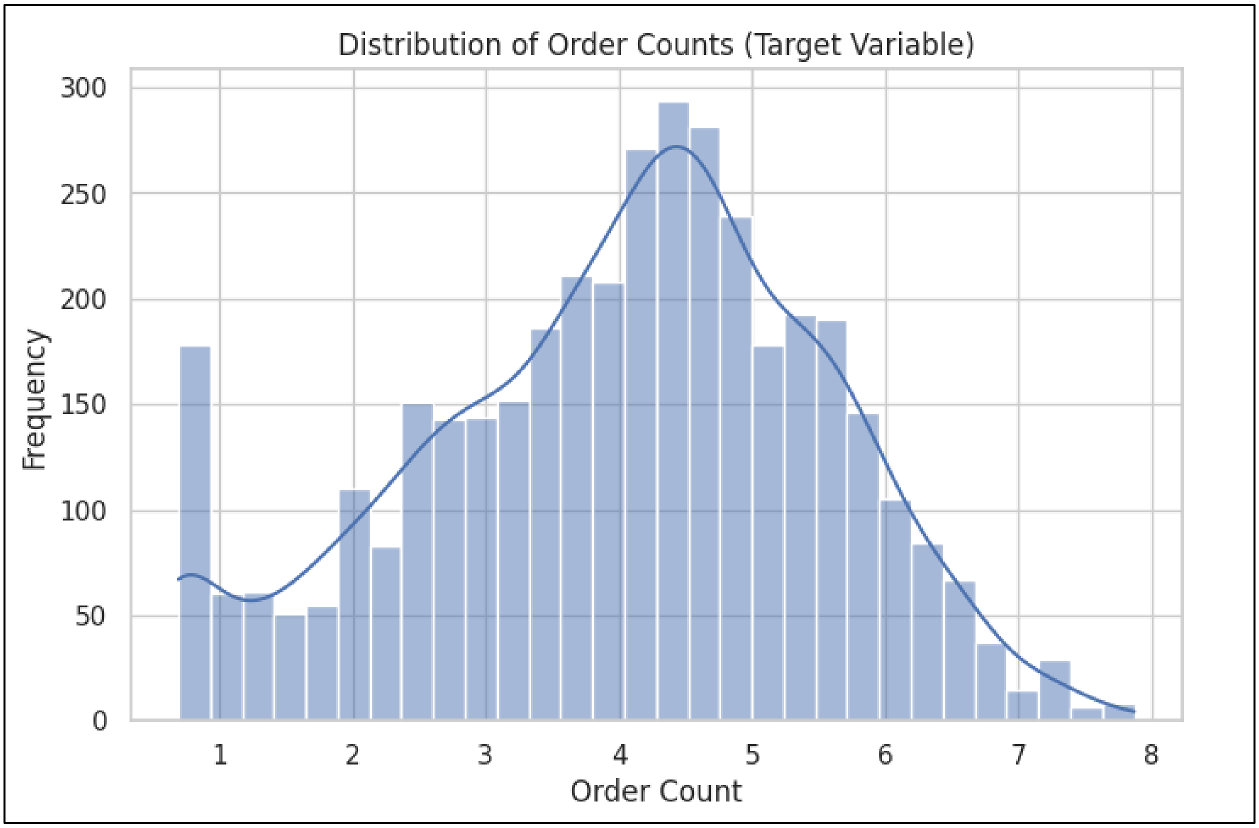

Figure 2. Order Count Distribution after Log transformation

3.4.4 Training / Test Split

Data Split: 80% for training, 20% for testing.
Training Groups: X_train, Y_train,
Testing Groups: X_test, Y_test.

The dataset was divided into training and test subsets using an 80:20 split. The training set (80%) is used to train the machine learning models by enabling them to learn the underlying patterns and relationships between input features and the target variable. The test set (20%), which remains unseen during training, is used to evaluate the model's generalization ability and assess how well it performs on new, real-world data.

This split strategy is crucial for identifying issues such as underfitting or overfitting. A model that performs well on the test data is considered reliable for deployment in real-world applications. Specifically, X_train contains 80% of the input features, while X_test contains the remaining 20%. The target variable, Y, is continuous in this study. Y_train provides the corresponding target values for training, helping the model learn from the patterns in X_train. Conversely, Y_test is used to assess how accurately the model can predict outcomes based on the unseen input features in X_test. The model's predictive performance is evaluated by calculating and comparing error metrics, as detailed in Section 3.4.5.

3.4.5 Evaluation Metrics

Root Mean Squared Error (RMSE): Measures the square root of the average squared differences between predicted and actual order counts.
Mean Absolute Error (MAE): Measures the average absolute differences between predicted and actual order counts.
Mean Squared Error (MSE): Measures the square difference between predicted and actual order counts.
Mean Absolute Percentage Error (MAPE): Measures the average of absolute percentage error between the predicted and actual order counts.
R Squared: It is a coefficient of determination used to determine the proportion of variance explained by the model. A higher R square value, closes to 100%, implies that prediction values are close to actual values.

Lower RMSE, MSE, MAE, MAPE and higher R squared values indicate better model predictive performance.

3.4.6 Cross Validation 5-Fold.

To ensure a robust evaluation of machine learning models performance used in this study, 5-fold cross-validation technique was employed. K-fold cross-validation technique is widely used in the data mining application for model evaluation and selection. This approach works by dividing the dataset into K equal subsets called folds. There are K iterations in total. In each iteration, one-fold is used as the validation or test set, while the remaining K-1 folds are used for training. This ensures that each data points are used for testing exactly once and for training in K-1 iterations. The model performance metrics are averaged across all K folds (Saud et al., 2020). In prediction or forecasting tasks, error metrics such as RMSE, MSE, MAE, MAPE and R-squared are used across all folds to provide more reliable and consistent estimates of model performance.

However, selecting an optimal K values plays a key role in reducing biases, variances and ensuring computational efficiency. Smaller K values typically, 2 or 3, are more likely to introduce bias, whereas larger K values, such as 10, can lead to increased computational cost (Zhang et al., 2023; Rodriguez et al., 2010). Academic literature in this domain has consistently used K values such as 5 or 10, which have been shown to provide reliable error estimate while being computationally efficient (Nti et al., 2021; Zhang et al., 2023).

Notably, 5-fold cross-validation has proven to provide a balance between computational efficiency and a reliable performance error estimate. For instance, a study by Sekeroglu et al. (2022) observed that Neural Networks, Logistic Regression, and Gradient Boosting achieved highest predictive performance using five-fold cross-validation. In this study we adopt 5-fold cross-validation approach for the evaluation and selection of best machine learning model in spatial temporal analysis of on demand food orders.

3.4.7 Model Comparison

Model performance was evaluated using five key metrics: Root Mean Squared Error (RMSE), Mean Squared Error (MSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and R-squared (R^2). The model that achieved the lowest error values and the highest R^2 score was deemed the most effective for accurately predicting future order distributions across various dates, time intervals, and geographic regions.

After identifying the best-performing model through 5-fold cross-validation, it was retrained on the full dataset to enhance learning and provide a more reliable estimate of its generalization capability. The model was then re-evaluated using the four key performance metrics outlined earlier.

Prior to training and dataset splitting, approximately 10 observations were intentionally set aside to serve as unseen data for final evaluation. In the final stage, this unseen data set was used to assess the model's ability to predict future order volumes and to evaluate how closely the predicted values matched the actual order counts.

3.5 Exploratory Data Analysis

Objective: The exploratory data analysis (EDA) was undertaken to develop a deeper understanding of courier performance, particularly regarding courier behaviors and operational bottlenecks in the last-mile delivery system. Additionally, the supply-demand balance was analyzed by examining the availability of couriers relative to the number of orders received, with a specific focus on peak demand periods.

3.5.1 Important Features for EDA

The EDA incorporated both original features from the dataset and additional derived features that were relevant to courier performance analysis. The key features include:

Courier ID: Unique identifier assigned to each courier.
Order Timestamp Features: Platform order time, estimated arrival time, estimated meal preparation time, order push time, dispatch time, grab time, fetch time, arrival time, and a weekend indicator (binary).
Spatial Features: Delivery area, delivered areas, and the count of unique ordered and delivered areas.
Temporal Features: Hour of the day, day of the week.
Performance Metrics: Total orders completed per courier, courier activity status, distance traveled per order, delivery time per order, and average delivery time per courier.

These features were used to evaluate courier behavior across different regions and time periods, providing a comprehensive view of operational dynamics.

3.5.2 Performance Metrics

Courier performance was assessed through several key performance indicators (KPIs):

Inactive status: Total time couriers were not active in the platform to accept orders.
Acceptance Rate: Percentage of assigned orders that were accepted by couriers.
Average Delivery Time: Average time taken to deliver an order.
Distance Traveled per Order (Distance_km): Distance traveled from order pickup to delivery completion.
Delivery Time: Total time elapsed from courier acceptance of an order to successful delivery.
Wave Order Status: Number of orders completed per wave (active status)

These metrics were analyzed both at the individual courier level and across different operational regions and time periods to generate both granular and system-wide insights.

3.5.3 Exploratory Data Analysis Tools

A range of statistical and visualization techniques were employed for the EDA:

Descriptive Statistics

Descriptive statistics play a fundamental role in summarizing large datasets into manageable and interpretable forms. As emphasized by Tukey (1977), these summaries facilitate the initial understanding of complex datasets by providing simplified views of variation. In this study, key performance indicators such as delivery time, distance traveled, inactive time, and acceptance rate were analyzed using measures such as the mean, median, mode, standard deviation, minimum, and maximum values. These metrics enabled the identification of distribution patterns, potential outliers, and skewness within the data.

Time Series Plots

Time series plots were employed to uncover temporal patterns in courier behavior and performance. These visualizations provide insight into how key operational metrics such as delivery time, distance traveled, inactive time, and activity status, fluctuate across different times of day and days of the week. This was particularly useful in identifying peak demand periods and understanding how courier efficiency changes in response to temporal variations in the on-demand food delivery context.

Univariate Analysis

Univariate analysis focuses on analyzing each variable independently to understand its distribution and characteristics. As Tukey (1977) describes, this approach offers descriptive insights into the central tendency, spread, and presence of extreme values in a single variable. In this study, box plots, histograms, and bar plots were utilized to examine variables such as delivery time, revealing skewed distributions and helping identify operational anomalies.

Bivariate Analysis

Bivariate analysis extends univariate techniques by exploring the relationship between two variables. The choice of visualization depends on whether the variables are quantitative, categorical, or a mix of both. Tools such as scatter plots and box plots were employed in this study to examine pairwise relationships between courier performance metrics, for instance, delivery time versus distance traveled, delivery time versus acceptance rate, and distance traveled versus acceptance rate. These visualizations helped reveal potential correlations and interdependencies between metrics.

Multivariate Analysis

Multivariate analysis enables the exploration of interactions among three or more variables. This approach is particularly valuable when analyzing spatial-temporal patterns in courier performance. For example, visualizations such as multi-axis line charts and multivariate scatter plots were used to explore how delivery time, order quantity, and fulfillment capacity interact across different time periods. These tools provided a more comprehensive view of complex dynamics within the dataset.

Correlation Analysis

Correlation analysis was used to statistically quantify the strength and direction of linear relationships between pairs of continuous variables. In this study, a correlation heatmap was utilized to visualize interdependencies among operational performance metrics, such as delivery time, distance traveled, and acceptance rate. This analysis was instrumental in identifying strong, moderate, or weak associations, informing both feature selection and deeper operational insights.

3.5.4 Peak Period Analysis

While performance metrics were analyzed throughout the entire operational day, particular attention was given to peak periods to assess courier supply and order demand imbalances.

For each order date, the operational day was divided into six four-hour time windows:

1. 00:00–04:00
2. 04:00–08:00
3. 08:00–12:00
4. 12:00–16:00
5. 16:00–20:00
6. 20:00–00:00

Within each four-hour window, the following were computed:

1. Total number of orders received
2. Number of available couriers
3. Average delivery time
4. Number of delayed and non-delayed orders

Multivariate scatter plots were used to examine the relationship between total orders, available couriers, average delivery time, and delay statuses across all time windows. This enabled the identification of peak periods for each order date and assessment of courier staffing adequacy during these periods.

3.5.5 Tools and Systems

The analysis was conducted using the following tools:

Data Preprocessing and Cleaning: Python (Pandas, NumPy) and Microsoft Excel
Data Visualization: Python libraries including Matplotlib, Seaborn, and Plotly
Statistical and Machine Learning Analysis: Python

All machine learning model implementations and data visualizations were performed in Google Jupyter Notebook environments. Data preprocessing tasks such as feature extraction and handling of missing values were executed using both Python scripts and Microsoft Excel utilities.

Exploratory Data Analysis Methodology Summary

Through extensive feature extraction, KPI computation, and multi-dimensional data visualization incorporating temporal and spatial dimensions, a comprehensive understanding of courier performance and behavior was achieved. The insights gained from EDA and spatial-temporal predictive modeling provides operational recommendations for dynamic courier allocation, particularly during peak demand periods, thus enhancing last-mile delivery efficiency.

Chapter 4. Results and Discussions

This section presents the findings from both the exploratory data analysis (EDA) of courier behavior and performance, as well as the spatiotemporal prediction of future order density using machine learning models. The outcomes offer actionable insights that can inform strategies for improving operational efficiency within on-demand delivery platforms.

4.1 Exploratory Data Analysis (EDA)

The EDA was structured around three analytical dimensions to uncover patterns in courier operations:

- Order Level Performance Metrics
- Wave Level Performance
- Peak Period Performance Analysis

4.1.1 Order Level Performance Metrics

Key performance indicators (KPIs) were computed for individual couriers based on available and derived features. These metrics were further visualized to highlight variations in courier behavior and performance. The KPIs include: *Acceptance Rate, Average Delivery Time per Courier, Average Distance Per Order, Total Orders completed by each courier, Courier Efficiency, Cross-Area rate, Courier Area Focus and Top 10 fastest and slowest couriers.*

For each of these indicators, descriptive statistics were supplemented with univariate, bivariate and multivariate visualizations to uncover operational trends and inefficiencies.

Acceptance Rate

Acceptance rate is defined as the proportion of delivery requests accepted by a courier out of the total number assigned. Figure 3 illustrates examples of couriers along with the number of orders assigned and the number accepted, offering a snapshot into courier responsiveness.

	courier_id	total_orders_assigned	total_orders_accepted	acceptance_rate
0	0	33	17	51.515152
1	1	21	17	80.952381
2	2	207	156	75.362319
3	3	126	91	72.222222
4	4	378	261	69.047619

Figure 3. Acceptance rate examples for each courier

Figure 4 below illustrates the distribution of order acceptance rates across all couriers included in the dataset.

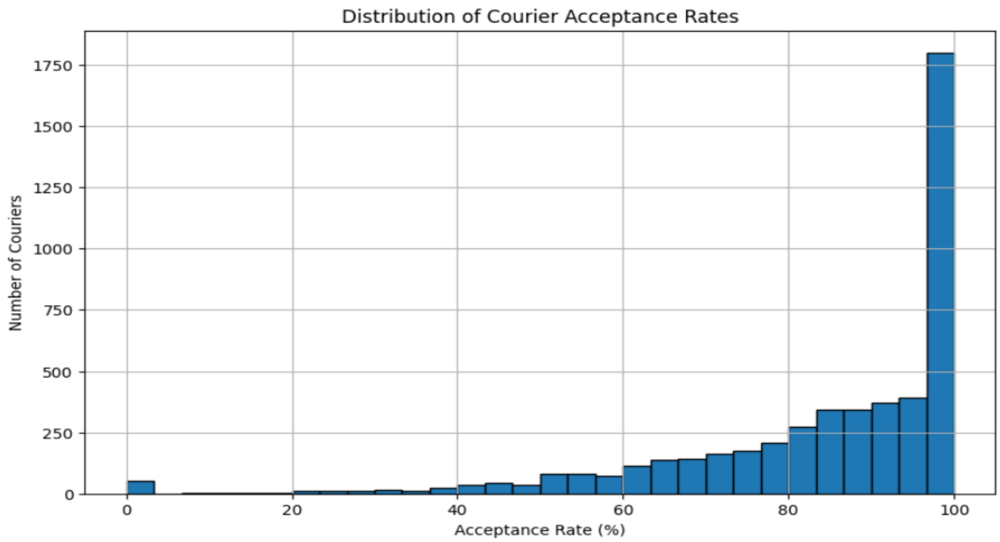

Figure 4. Distribution of Acceptance Rate

The distribution of courier acceptance rates indicates that the majority of couriers maintain an acceptance rate of at least 80% for assigned orders. Out of 4,955 unique couriers in the dataset, approximately 1,750 couriers exhibit an acceptance rate close to 100%. While a small subset of couriers displays lower acceptance rates, below 60%, however, these represent a minor fraction relative to the total courier population.

The descriptive statistics shown in Figure 5, further reinforce this trend: the average acceptance rate across all couriers is 84%, with a median of 90%. The interquartile range (IQR) spans from 76% to 100%, suggesting that a substantial proportion of couriers consistently accept the majority of orders assigned to them.

```
count    4955.000000
mean       84.743624
std        18.748192
min         0.000000
25%        76.923077
50%        90.769231
75%       100.000000
max       100.000000
Name: acceptance_rate, dtype: float64
```

Figure 5. Descriptive stats of acceptance rate

Average Delivery Time

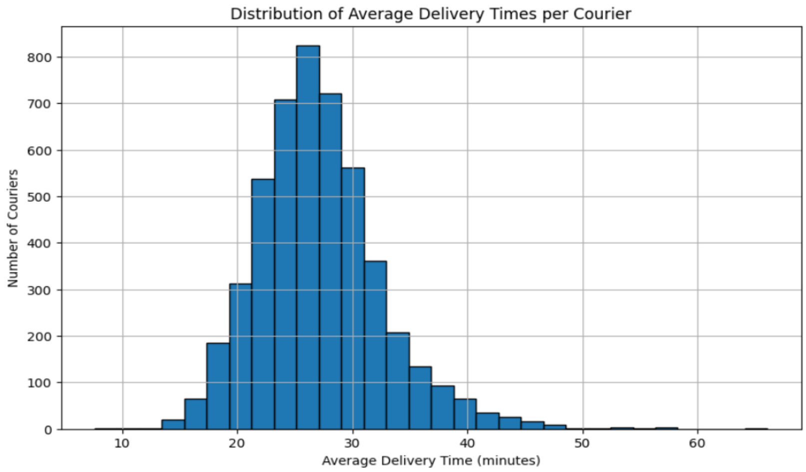

Figure 6. Distribution of Average Delivery Time

```
count    4901.000000
mean       27.098455
std         5.453358
min         7.616667
25%        23.523759
50%        26.646991
75%        29.949038
max        66.075000
Name: avg_delivery_time, dtype: float64
```

Figure 7. Descriptive statistics of Average Delivery Time

The histogram and descriptive statistics for the average delivery time per courier shown in Figure 6 and Figure 7 respectively, provides valuable insights into the variability of delivery performance across the workforce. The mean delivery time is approximately 27 minutes, with a standard deviation of 5.45 minutes, indicating moderate variability among couriers. The interquartile range spans from 23 to 29 minutes, suggesting that a large proportion of couriers' complete deliveries within this window.

The distribution, as illustrated by the histogram, is slightly right-skewed, reflecting the presence of a few outliers. While most couriers maintain consistent delivery times, a small number exhibit significantly lower (as little as 7.6 minutes) or higher (up to 66 minutes) average delivery durations per order.

Average Distance Travelled

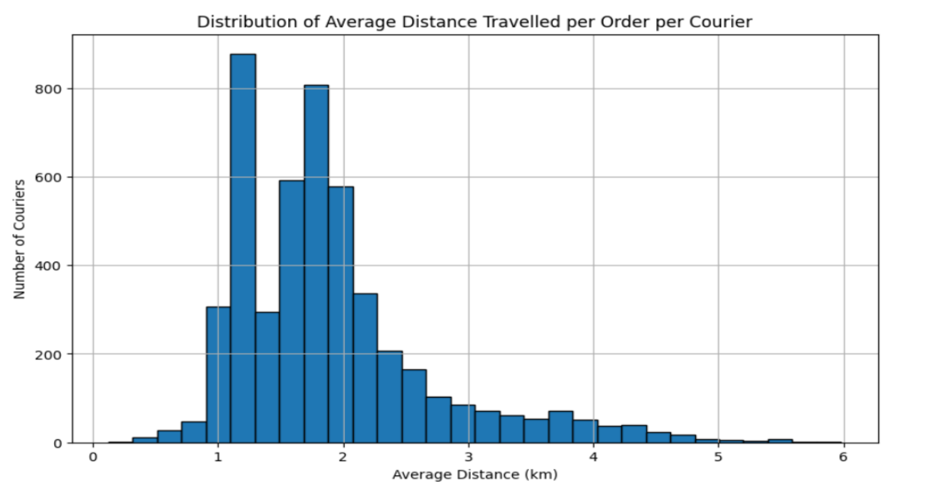

Figure 8. Distribution of Average Distance Travelled

```
count    4901.000000
mean        1.896525
std         0.799379
min         0.124708
25%         1.275356
50%         1.763263
75%         2.138371
max         5.980241
Name: avg_distance_km, dtype: float64
```

Figure 9. Descriptive Statistics of Average Distance Travelled

Figure 8 and Figure 9 shows histogram and descriptive statistics for the average distance traveled per order by 4,901 couriers respectively, reveal relatively low variability in delivery distances, with a standard deviation of 0.799 km. The mean distance per order is 1.86 km. The distribution is right-skewed, indicating that a large proportion of couriers typically travel between 1 and 2.5 km per delivery. However, a small number of couriers have experienced significantly longer delivery distances, with the maximum reaching 5.98 km. This suggests that certain couriers are assigned to more geographically dispersed areas.

Courier Performance- Acceptance Rate vs Average Delivery Time vs Average Distance Travelled

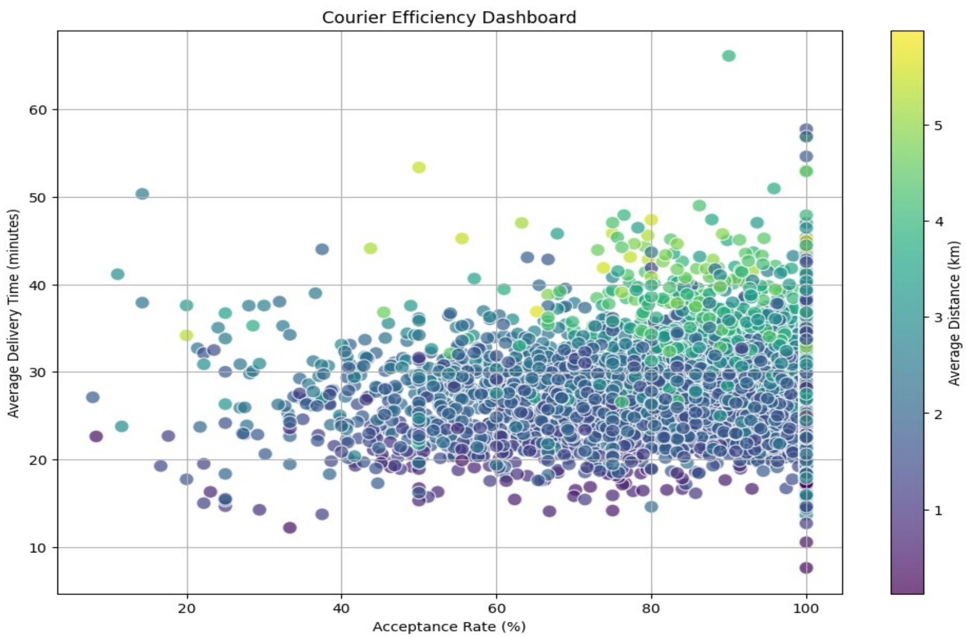

Figure 10. Courier Performance Multidimensional Scatter Plot

Courier performance, shown in Figure 10, was analyzed using a multidimensional scatter plot that captures the relationship between acceptance rate, average delivery time, and average distance traveled for each courier. In this visualization, each point represents a unique courier.

The distribution of points reveals that couriers with higher acceptance rates generally exhibit lower average distances traveled and correspondingly shorter delivery times per order. As expected, couriers covering longer distances tend to have longer average delivery durations, highlighting the direct impact of travel distance on delivery efficiency.

Cross-Area rate

The cross-area rate was computed for each courier to assess the geographical dispersion of their assigned deliveries. This metric indicates whether couriers are primarily fulfilling orders within the same area where the orders were placed or across different zones. Figure 11 below presents sample cross-area rates for selected couriers.

	courier_id	total_orders	cross_area_deliveries	cross_area_rate
0	0	17	12	70.588235
1	1	17	12	70.588235
2	2	156	139	89.102564
3	3	91	71	78.021978
4	4	261	222	85.057471

Figure 11. Cross Area Rate Courier Samples

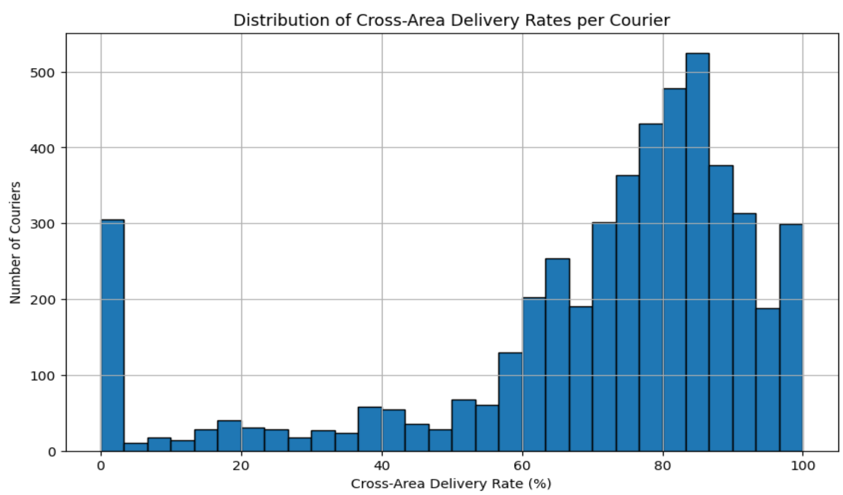

Figure 12. Distribution of Cross Area Delivery Rate

```
count    4901.000000
mean       70.342157
std        25.397756
min         0.000000
25%        64.150943
50%        77.966102
75%        86.290323
max       100.000000
Name: cross_area_rate, dtype: float64
```

Figure 13. Descriptive Statistic for Cross Area Rate

The proportion of cross-area deliveries was examined using the cross-area rate, which quantifies the extent to which couriers deliver orders outside the area in which they were received. Figure 12 and Figure 13 shows the distribution and descriptive statistics, which reveal considerable variability, with a standard deviation of 25.39%. The mean cross-area rate is 70%, with an interquartile range spanning from 64% to 86%, and a median of 77.96%. These figures suggest that a majority of couriers are engaged in cross-area deliveries.

The histogram exhibits a bimodal distribution, with notable peaks at 0% and between 60% to 100%. This pattern highlights operational diversity among the 4,901 couriers: approximately 300 couriers predominantly deliver within the same area as the order origin, while the remaining majority frequently operate across multiple geographic zones.

Courier Performance - Acceptance Rate vs Average Delivery Time vs Cross Area Delivery

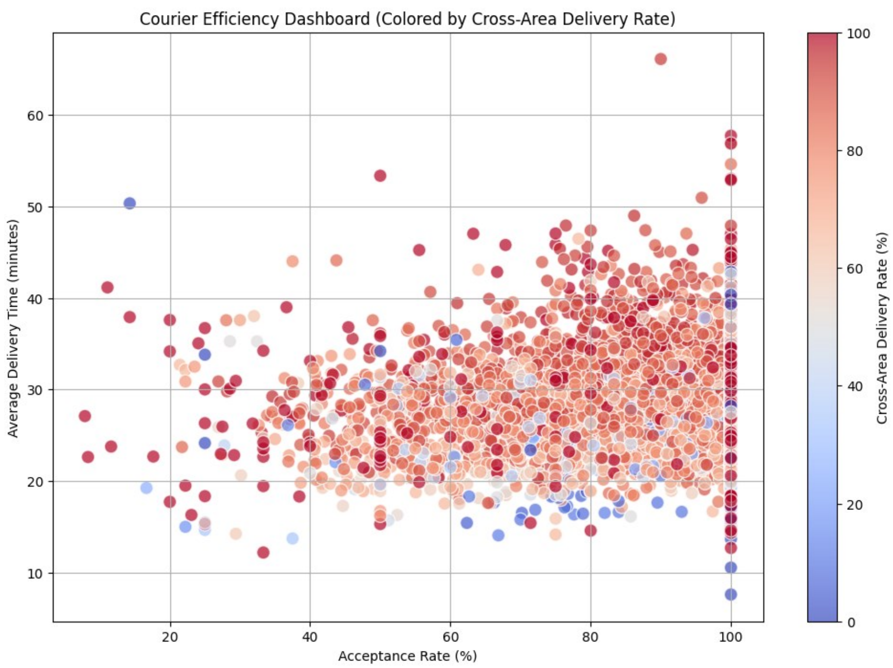

Figure 14. Courier Performance - Acceptance Rate vs Average Delivery Time vs Cross Area Delivery

To assess the influence of spatial factors on courier performance, multivariate scatterplot was developed to understand how acceptance rate and delivery time varies with increasing cross area engagements, as illustrated in Figure 14 above. In this plot, each point represents an individual courier, with the acceptance rate on the x-axis and average delivery time on the y-axis. The points are color-coded based on cross-area rate, ranging from light blue (indicating low or no cross-area engagement) to deep red (indicating high cross-area engagement).

The distribution reveals that a substantial number of couriers are highly mobile, frequently completing deliveries across different areas. These couriers tend to exhibit higher acceptance rates and typically complete deliveries within 20 to 40 minutes. In contrast, couriers with limited cross-area activity generally show lower average delivery times, though their acceptance rates vary. This spatial distribution underscores the impact of geographic flexibility on courier performance and operational behavior.

Area Focus - Multi Area vs Specialized

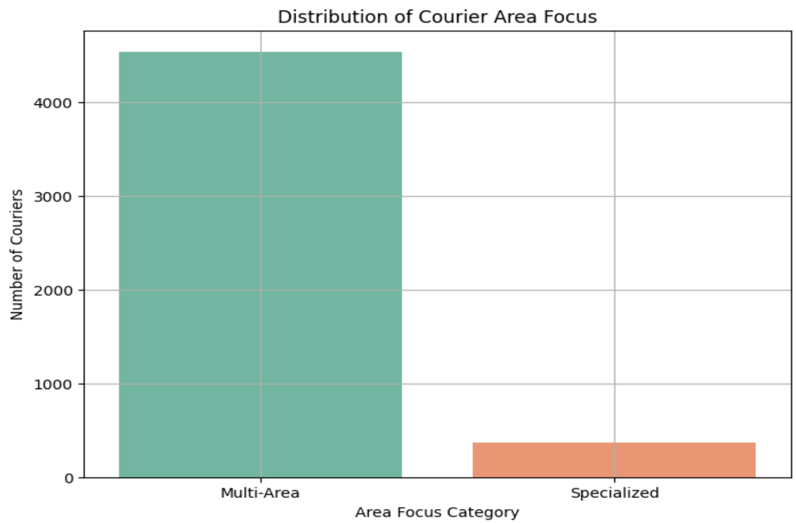

Figure 15. Distribution of Courier Area Focus

	acceptance_rate	avg_delivery_time	avg_distance_km	cross_area_rate
area_focus				
Multi-Area	84.897668	27.539712	1.935582	74.747696
Specialized	95.338061	21.630969	1.412583	15.754402

Figure 16. Comparison between Multi-Area and Specialized Couriers

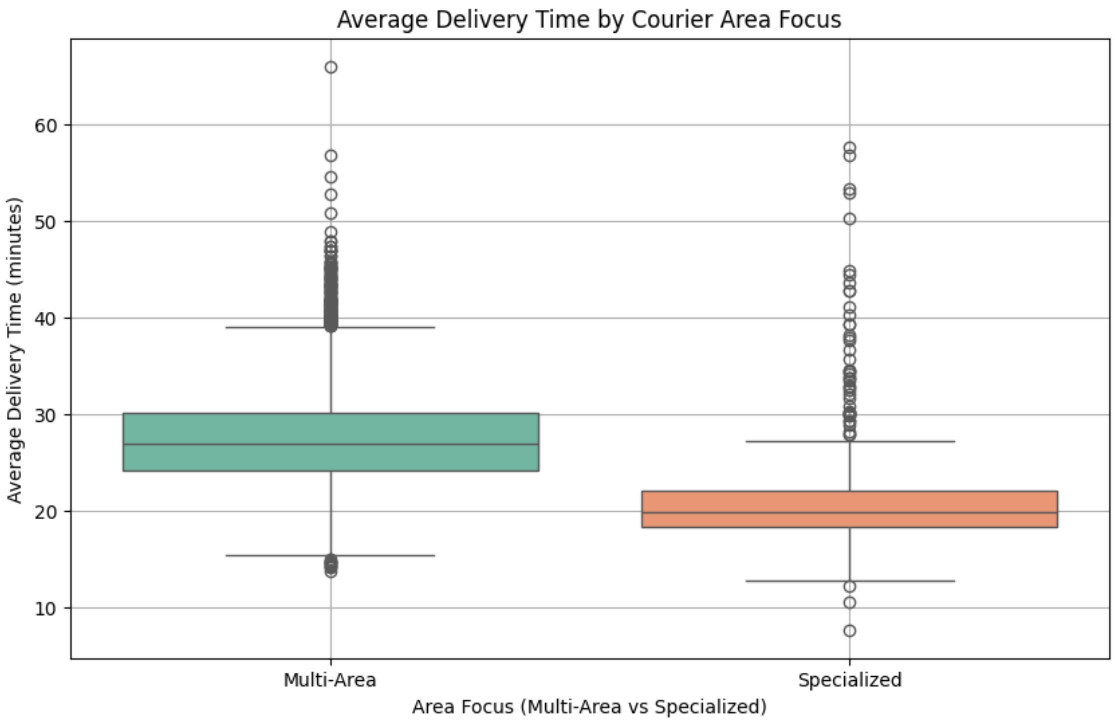

Figure 17. Average Delivery Time by Area Focus

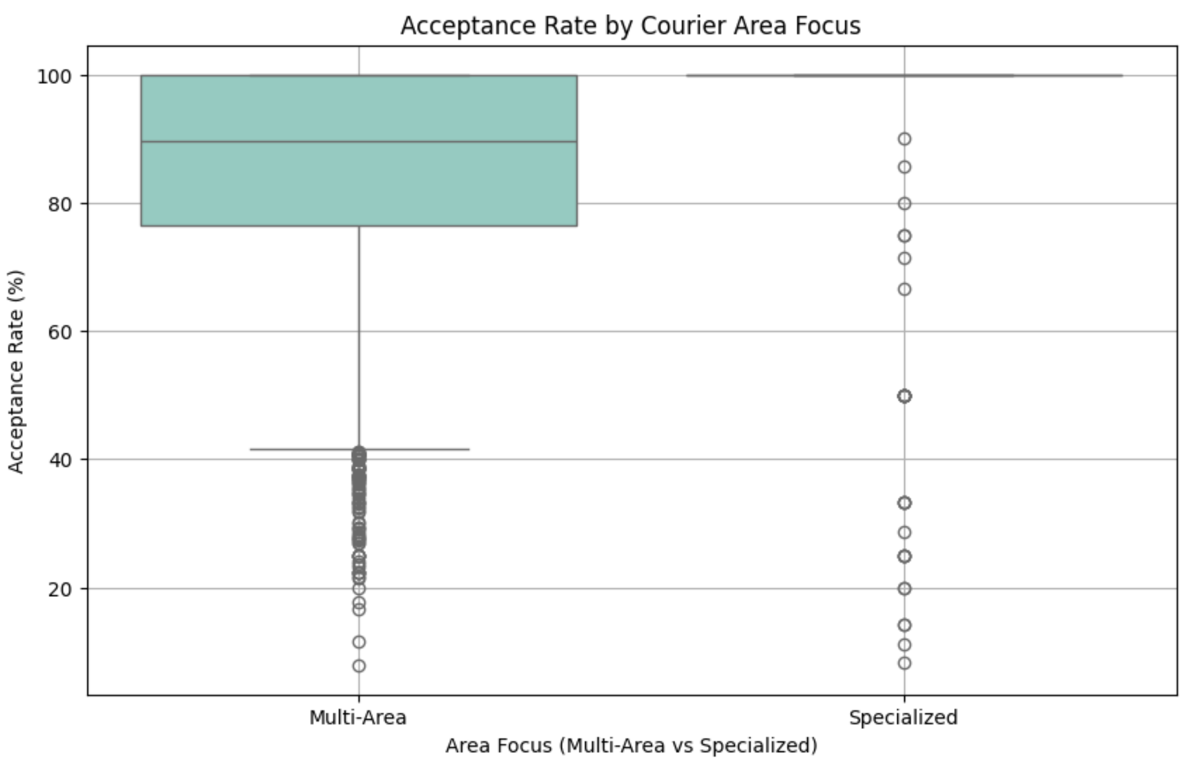

Figure 18. Acceptance Rate by Area Focus

Based on the spatial delivery patterns observed through cross-area delivery rates, couriers were classified into two distinct groups: Specialized and Multi-Area. Specialized couriers primarily deliver orders within the same geographic area where the order was received, while Multi-Area couriers frequently pick up orders in one area and deliver them to different zones.

Out of 4,901 couriers, 4,535 were categorized as Multi-Area and 366 as Specialized. Figure 15 shows this distribution for number of couriers who are engaged in Specialized and Multi-Area. Figure 16 shows, Specialized couriers demonstrated superior performance across several key metrics. On average, they achieved a higher acceptance rate of 95.3% and a lower average delivery time of 21.6 minutes. In comparison, Multi-Area couriers had an average acceptance rate of 84.9% and an average delivery time of 27.5 minutes. Additionally, Multi-Area couriers traveled longer distances per order on average, reflecting their broader spatial coverage.

These performance disparities between the two courier groups are also visually evident in the box plots shown in Figure 17 and Figure 18.

Fastest and Slowest Couriers based on Average Delivery Time

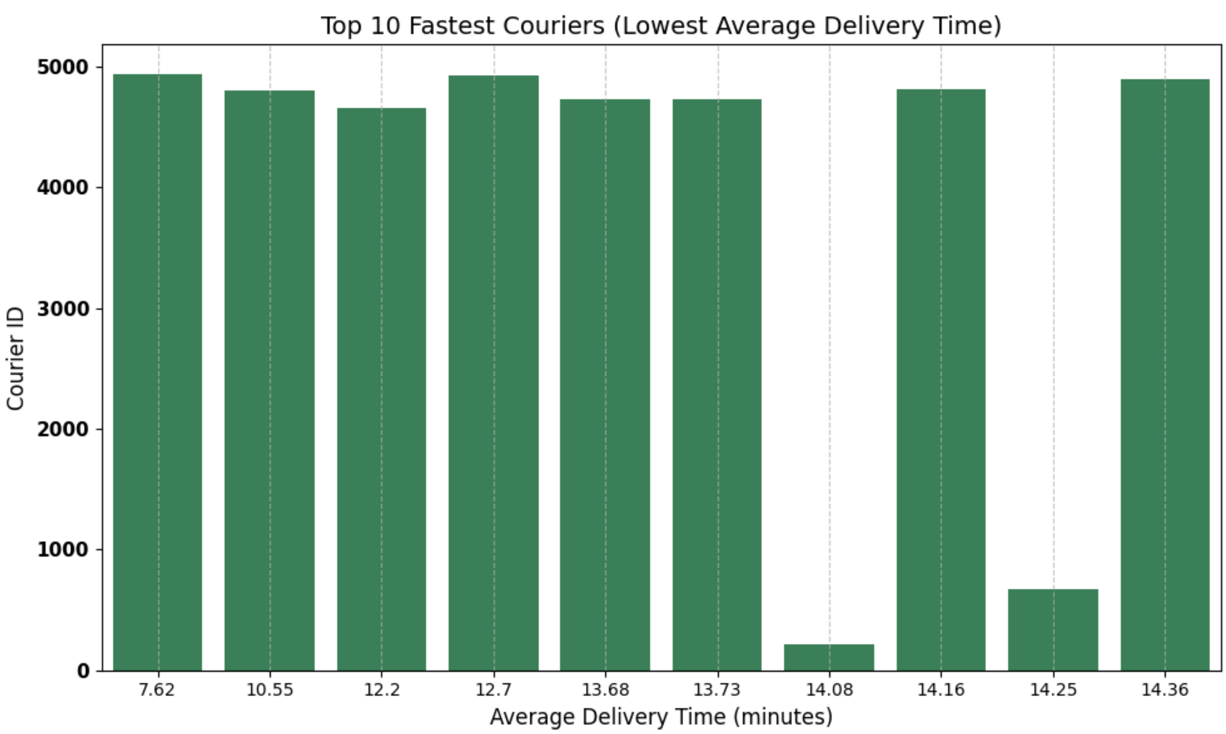

Figure 19. Top 10 Fastest Couriers

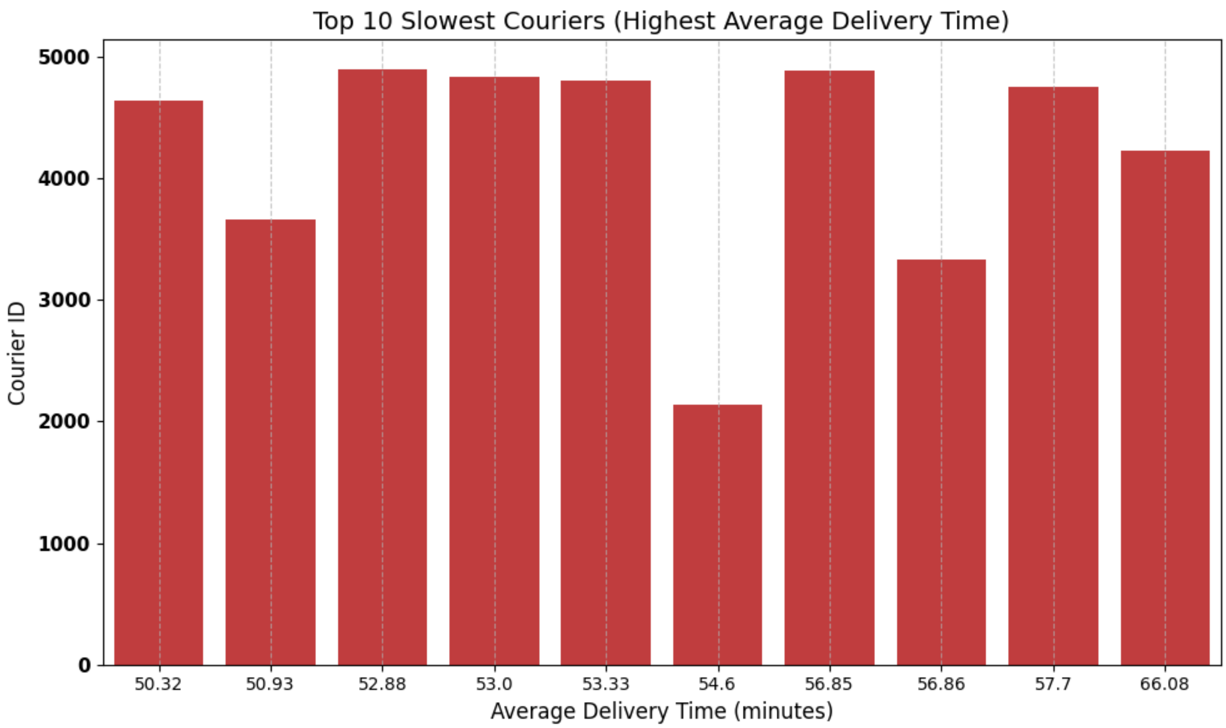

Figure 20. Top 10 Slowest Couriers

Table 6. Top 10 Fastest Couriers by Average Delivery Time

Courier ID	Average Delivery Time (min)
4937	7.62
4801	10.55
4662	12.2
4927	12.7
4731	13.68
4732	13.73
220	14.08
4810	14.16
668	14.25
4901	14.36

Table 7. Top 10 Slowest Couriers by Average Delivery Time

Courier ID	Average Delivery Time (min)
4223	66.08
4749	57.7
3329	56.86
4879	56.85
2136	54.6
4803	53.33
4831	53
4893	52.88
3658	50.93
4633	50.32

To examine the performance disparities between fast and slow couriers, the top 10 fastest and slowest couriers were identified based on their average delivery times. The fastest couriers as illustrated in Figure 19, recorded delivery times ranging from 7.62 to 14.36 minutes, whereas the slowest couriers had significantly higher averages, ranging from 50.32 to 66.08 minutes, shown in Figure 20. Additionally, Tables 6 and 7 clearly identify these couriers by their unique Courier IDs along with their corresponding delivery times.

This stark contrast underscores the substantial variability in last-mile delivery performance among couriers. While average delivery time is a primary and widely used metric for assessing courier performance, it alone does not provide a complete picture. As demonstrated in earlier analysis, spatial factors such as cross-area engagement and delivery zone distribution also play a critical role in influencing couriers' performance.

Pairwise Scatterplots - Average Delivery Time, Average Distance Travelled and Acceptance Rate

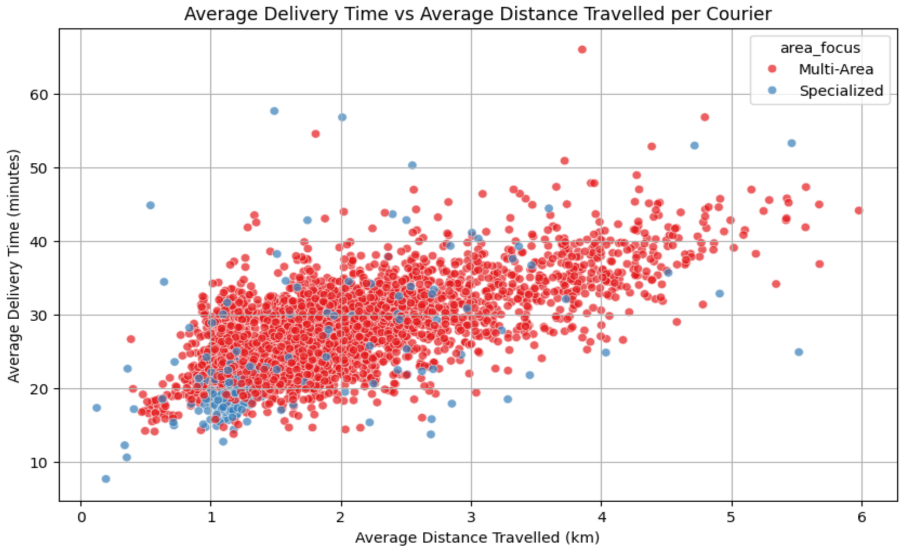

Figure 21. Scatterplot for Average Delivery Time vs Average Distance Travelled

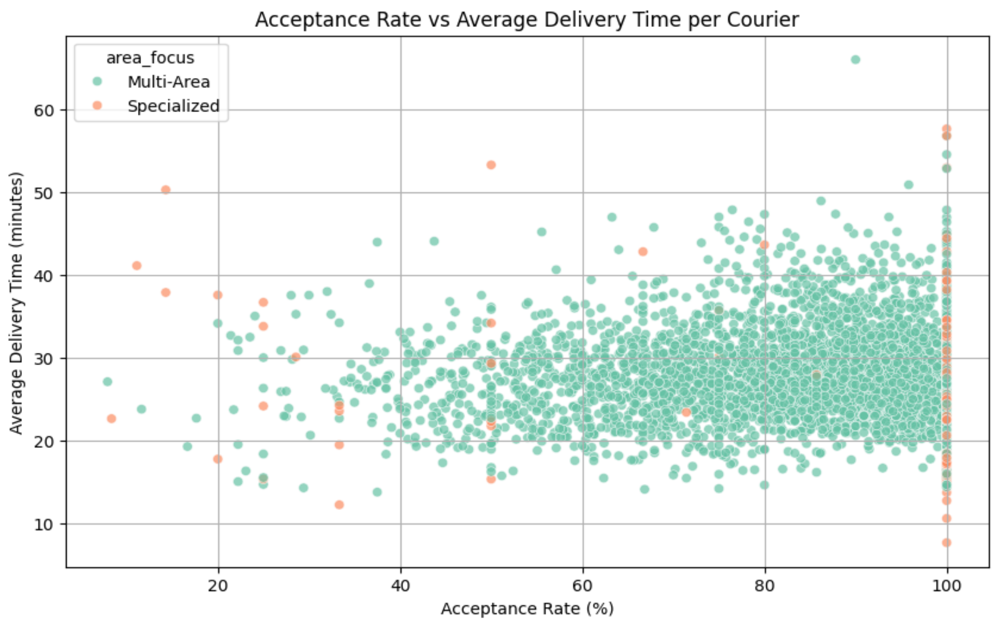

Figure 22. Scatterplot for Acceptance Rate vs Average Delivery Time

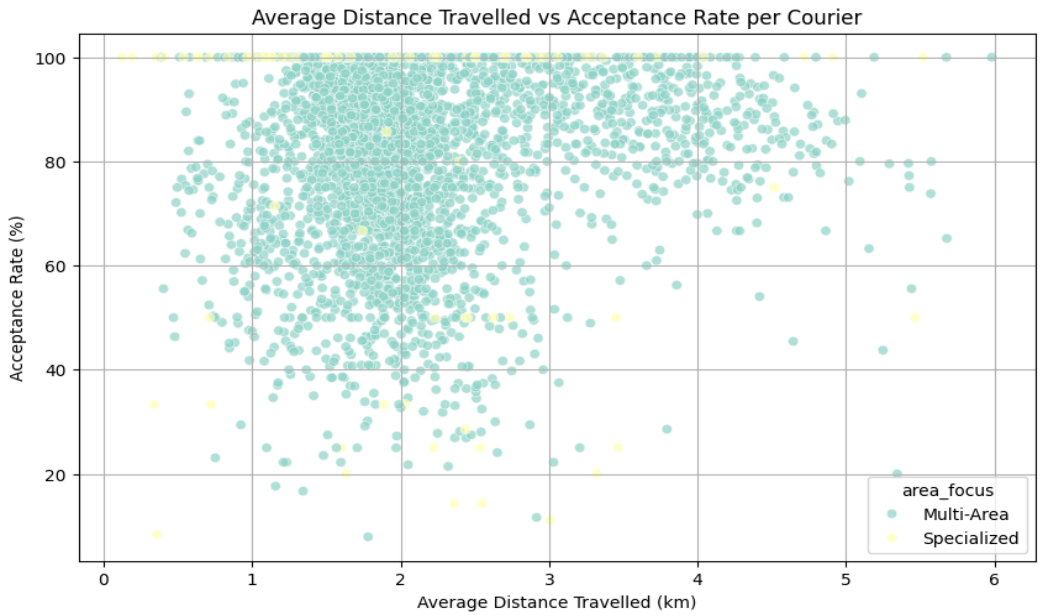

Figure 23. Scatterplot for Acceptance Rate vs Average Distance Travelled

To explore the interrelationships among key performance metrics for both Multi-Area and Specialized couriers, pairwise scatterplots were generated using average delivery time, average distance traveled, and acceptance rate. Figure 21 to Figure 23 illustrates how these operational metrics correlates with each other. Each data point represents an individual courier and is color-coded by courier type. As expected, Figure 21 shows a positive correlation is observed between average delivery time and average distance traveled. Multi-Area couriers tend to travel longer distances and exhibit higher delivery times compared to Specialized couriers, with few exceptions.

When examining acceptance rate against average delivery time in Figure 22, most couriers regardless of classification, demonstrate relatively high acceptance rates alongside moderate delivery times. However, couriers with lower acceptance rates show more variability in delivery performance. While no strong direct relationship is observed between acceptance rate and average distance traveled as shown in Figure 23, a dense cluster of data points appears at higher acceptance rates and delivery distances between 1 km and 3 km. Holistically, these pairwise scatterplots support earlier findings and reinforce the influence of spatial attributes on courier performance.

Correlation Matrix

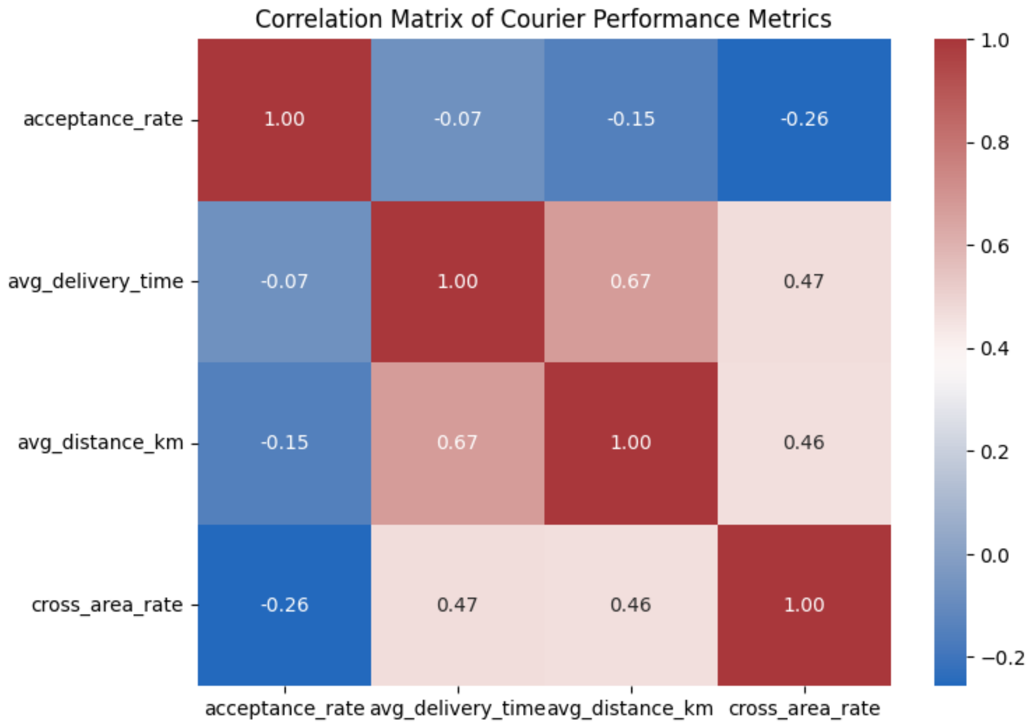

Figure 24. Correlation Matrix

To quantify the relationships among various courier performance metrics, a correlation matrix was constructed, as illustrated in Figure 24 above. As expected, the matrix reveals a strong positive correlation between average delivery time and distance traveled, indicating that longer delivery distances generally result in higher delivery durations.

Moderate positive correlations were observed between cross-area rate and both delivery time and distance traveled, suggesting that couriers frequently delivering across zones tend to cover more distance and take longer to complete deliveries. Conversely, acceptance rate demonstrated a weak negative correlation with delivery time, and a moderate negative correlation with both distances traveled and cross-area rate.

These findings suggest that couriers with higher acceptance rates are more likely to complete deliveries within the same geographic area, leading to improved efficiency in terms of both time and distance.

4.1.2 Wave Level Performance

A wave is defined as a continuous time sequence during which a courier transitions from an offline to online state, completes all assigned deliveries, and returns to an inactive status. Each sequence is assigned a unique wave ID for each courier. Analyzing performance at the wave level is crucial for understanding courier behavior in distinct operational cycles and examining how these behavioral patterns influence key performance metrics.

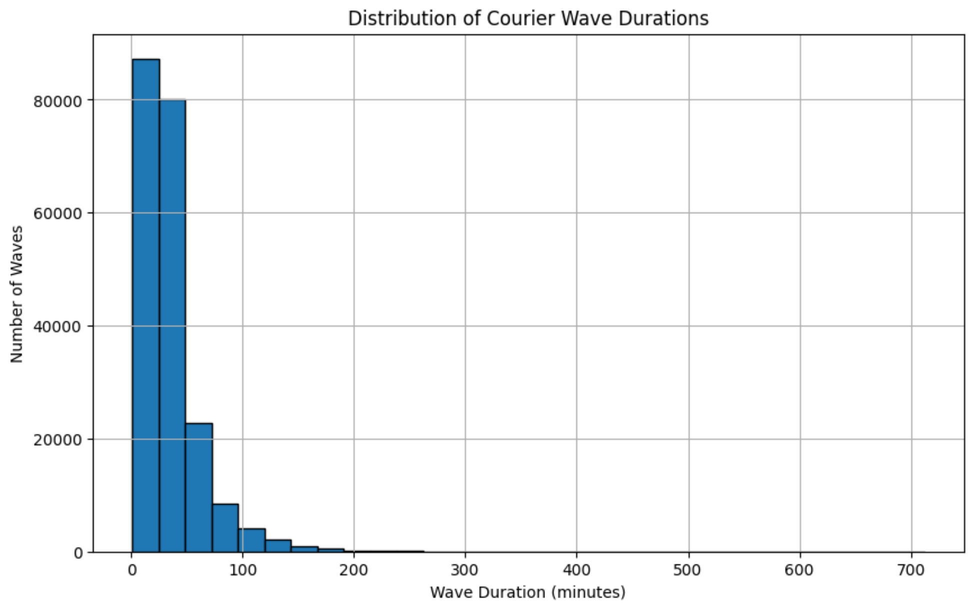

Figure 25. Distribution of Courier Wave Duration

```
count    206748.000000
mean         35.525091
std          26.914226
min           1.250000
25%          18.933333
50%          28.083333
75%          42.683333
max         712.366667
Name: wave_duration_minutes, dtype: float64
```

Figure 26. Descriptive Statistics for Courier Wave Duration

The descriptive statistics (Figure 26) and histogram (Figure 25) for wave duration across all couriers reveal that, on average, courier activity spans relatively short periods. The mean wave duration is approximately 35 minutes, suggesting that most couriers remain in active delivery status for brief intervals. The interquartile range extends from 18.9 minutes to 42.68 minutes.

However, the distribution is right-skewed with a long tail, indicating the presence of a small number of couriers with significantly longer active durations. The maximum observed wave duration is 712 minutes (approximately 11.8 hours), highlighting considerable variation in operational patterns among couriers.

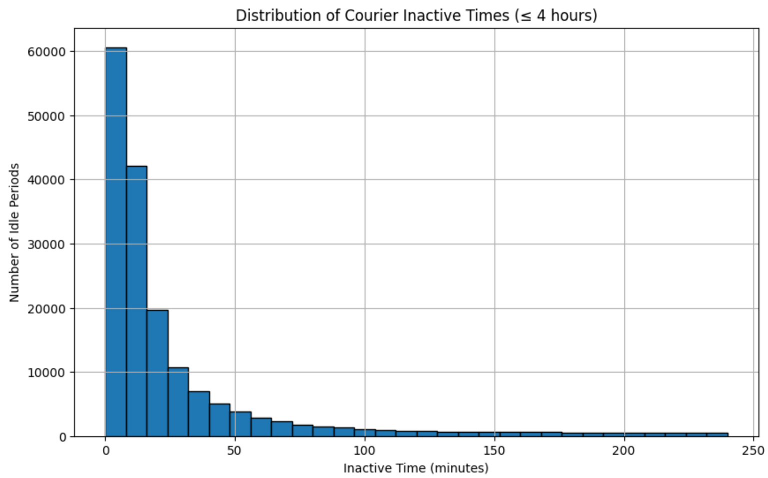

Figure 27. Temporal Distribution of Couriers Inactivity

```
count    170600.000000
mean         27.502231
std          42.093296
min           0.016667
25%           5.716667
50%          11.950000
75%          27.616667
max         239.983333
Name: inactive_time_minutes, dtype: float64
```

Figure 28. Descriptive Statistics for Couriers Inactive Time

To assess courier efficiency, inactive time is a critical performance metric, as it reflects the duration during a shift when a courier remains inactive or were not available to accept any orders. This metric was calculated using the unique wave ids assigned to couriers when they are active on the platform. The inactive time is the duration between the end of one wave and the start of another wave. Inactive time shows the duration for which couriers remain offline and were not available to complete deliveries. This metric shows the engagement time of each courier. Figure 27 above illustrates the distribution of inactive time across all available couriers.

The mean inactive period is approximately 27 minutes, with the majority of couriers exhibiting durations between 5 and 27.5 minutes as shown in Figure 28. However, the distribution in Figure 27 is right-skewed, indicating that a subset of courier's experiences significantly longer offline periods. Figure 28 shows maximum observed inactive time is 239 minutes, highlighting notable differences in engagement levels among couriers during their shifts.

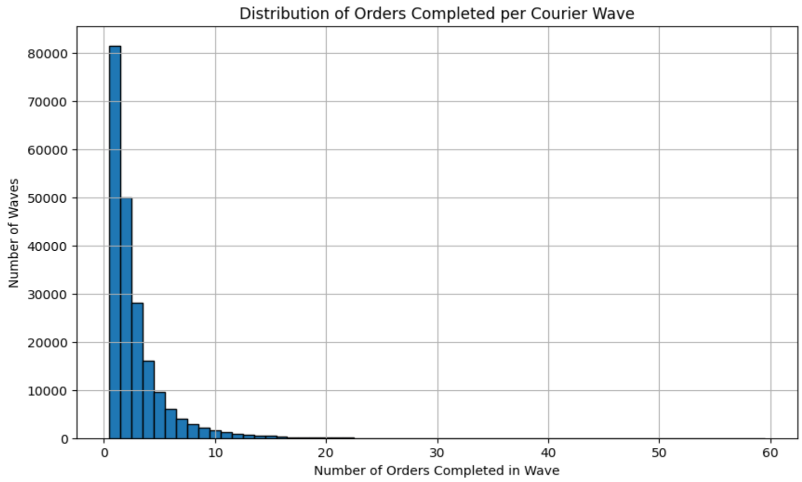

Figure 29. Distribution of Orders Completed per Courier Wave

Since each courier is associated with multiple wave ID (periods of active status), the number of deliveries completed per wave was calculated to understand workload distribution with respect to wave duration.

The distribution in Figure 29 indicates that most couriers typically complete between 1 to 3 deliveries per wave. However, a smaller subset of couriers completes 5 or more deliveries in a single wave, with rare instances showing delivery counts as high as 10 to 20.

Understanding the relationship between inactive time, wave duration, and the number of deliveries per wave provides valuable insight into individual courier workload. These metrics are essential for accurately evaluating courier performance and can be used to inform more efficient and balanced order assignments in future operations.

4.1.3 Peak Period Analysis

Peak period analysis of orders and courier activity using exploratory data analysis (EDA) is crucial for understanding the balance between supply and demand, specifically, the relationship between order volume and courier availability during high-demand time windows.

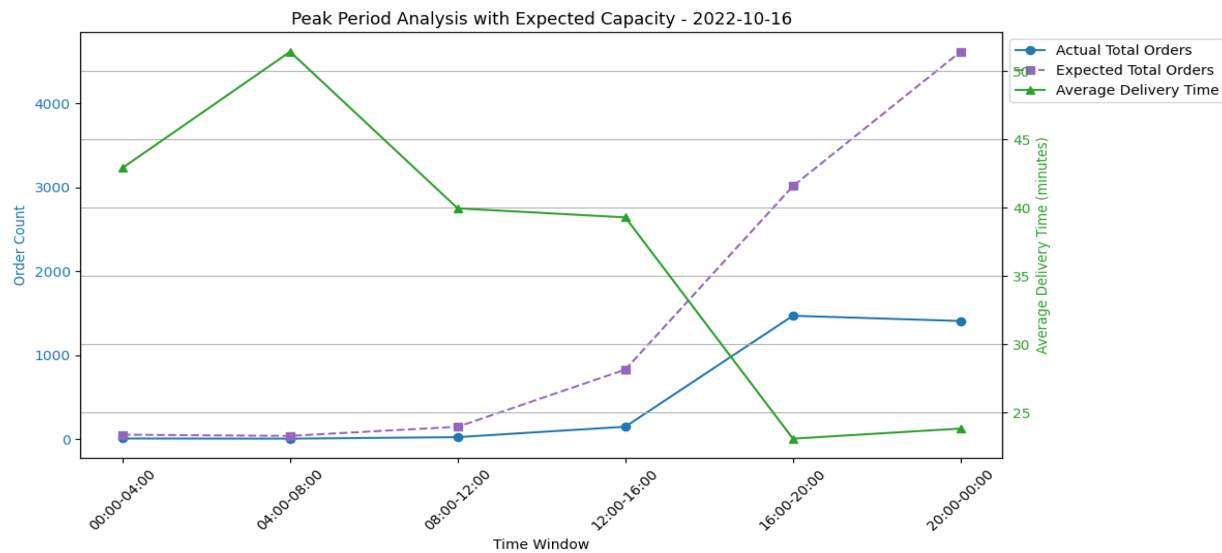

Figure 30. Line Chart for Peak Period 16th October 2022

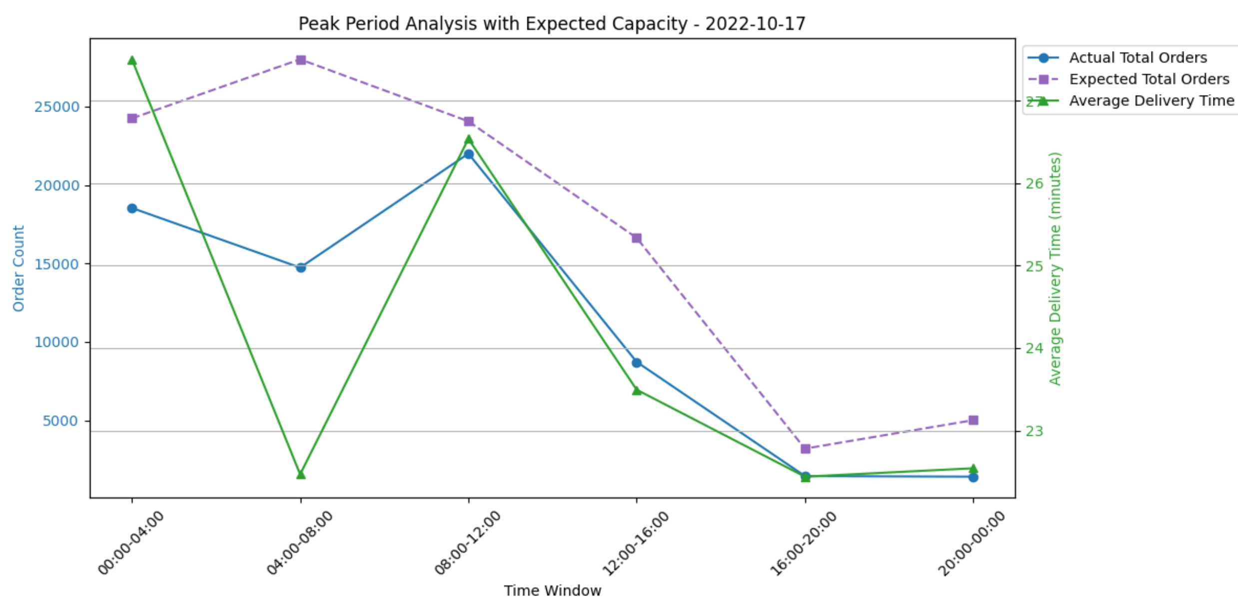

Figure 31. Line Chart for Peak Period 17th October 2022

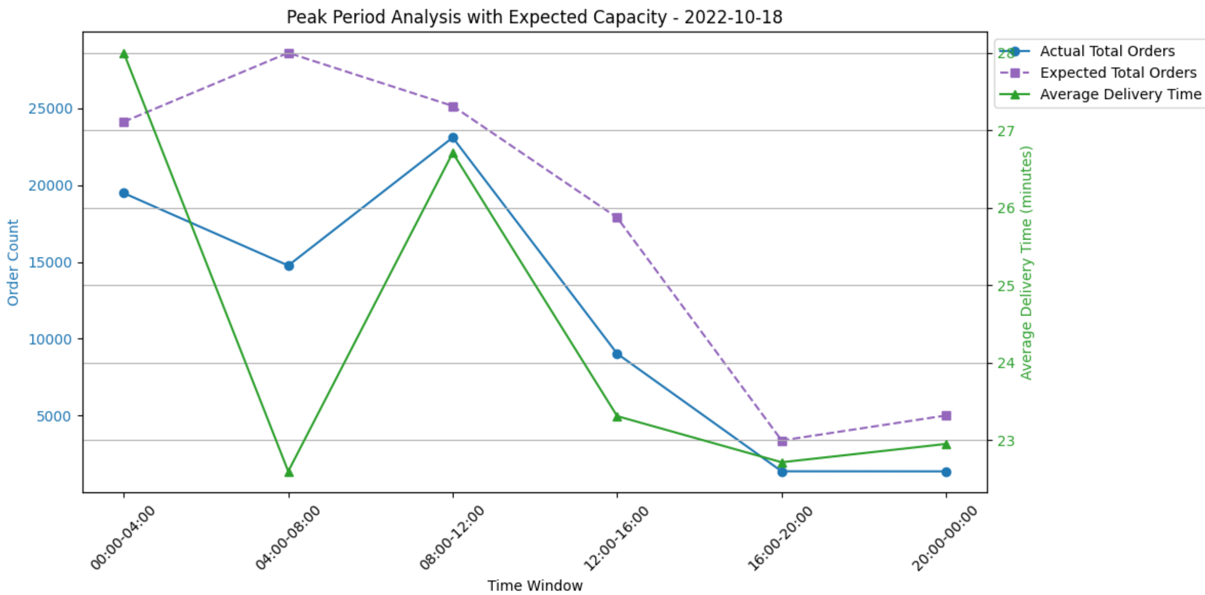

Figure 32. Line Chart for Peak Period 18th October 2022

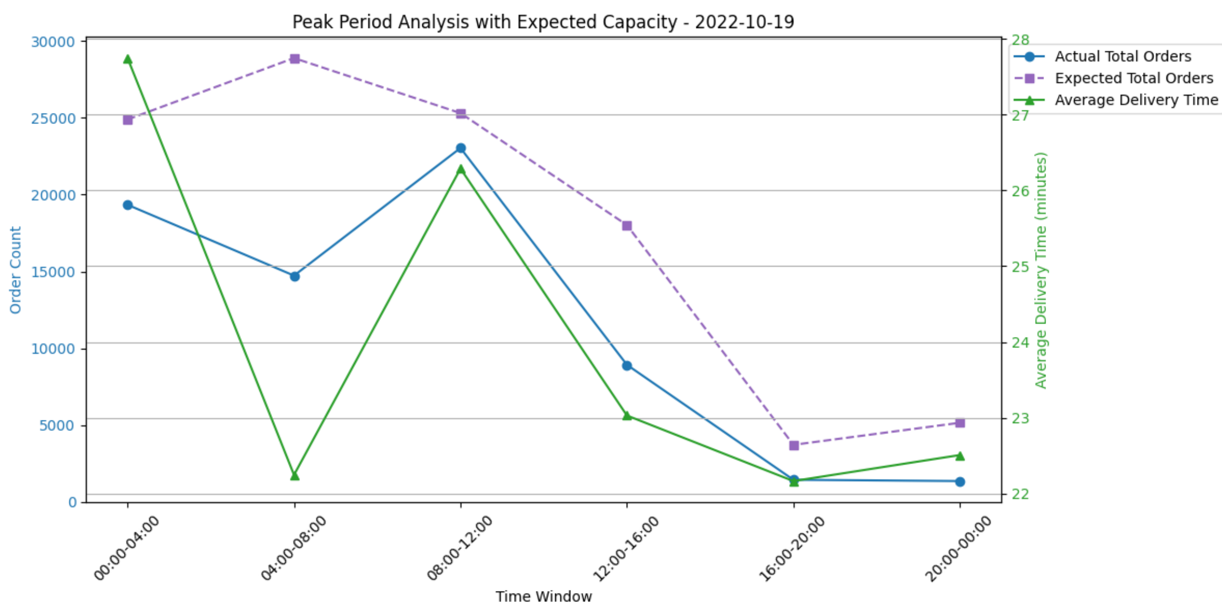

Figure 33. Line Chart for Peak Period 19th October 2022

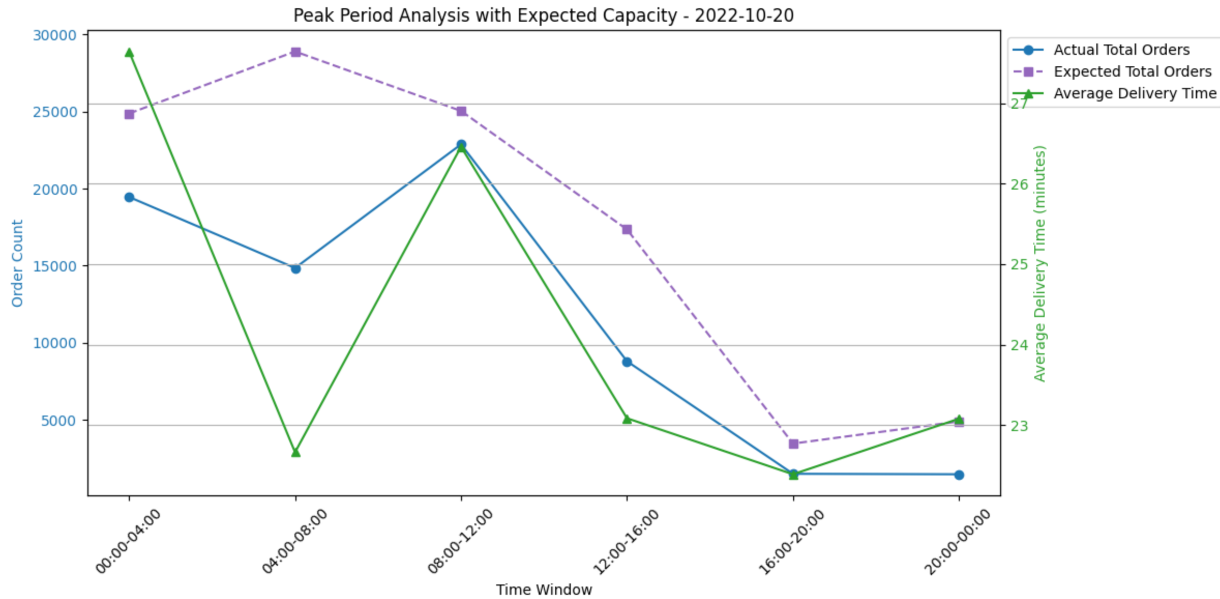

Figure 34. Line Chart for Peak Period 20th October 2022

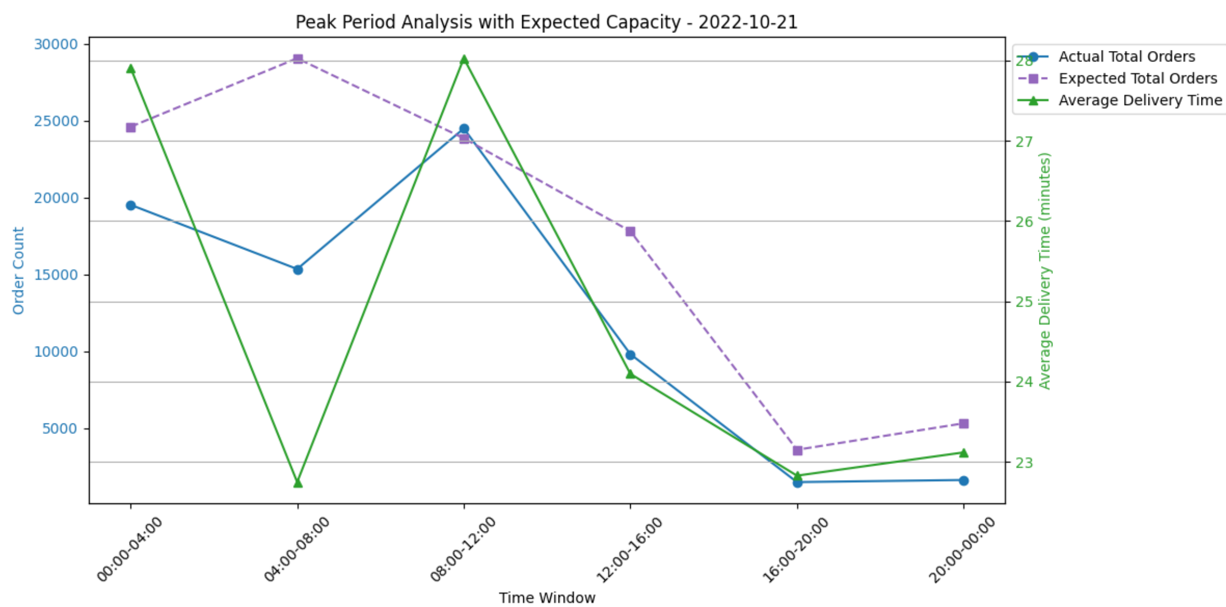

Figure 35. Line Chart for Peak Period 21st October 2022

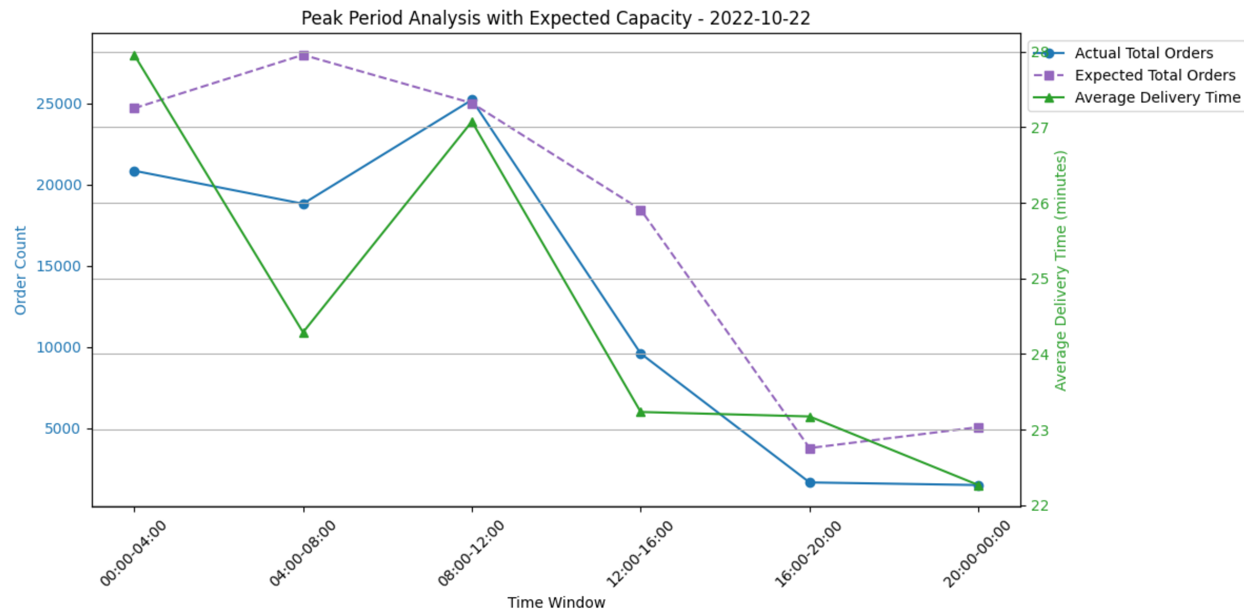

Figure 36. Line Chart for Peak Period 22nd October 2022

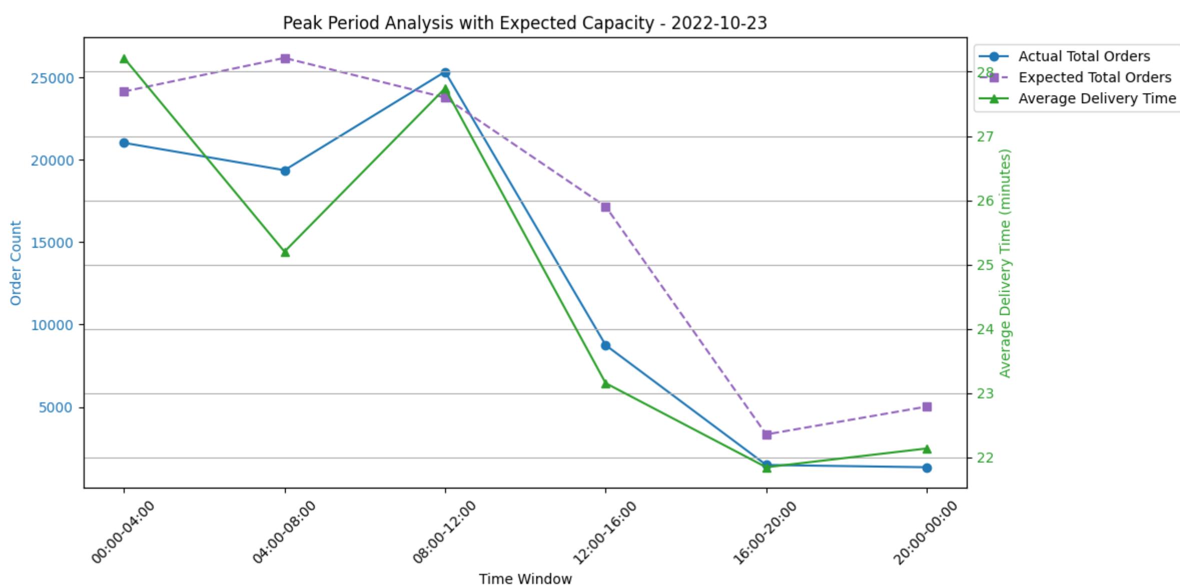

Figure 37. Line Chart for Peak Period 23rd October 2022

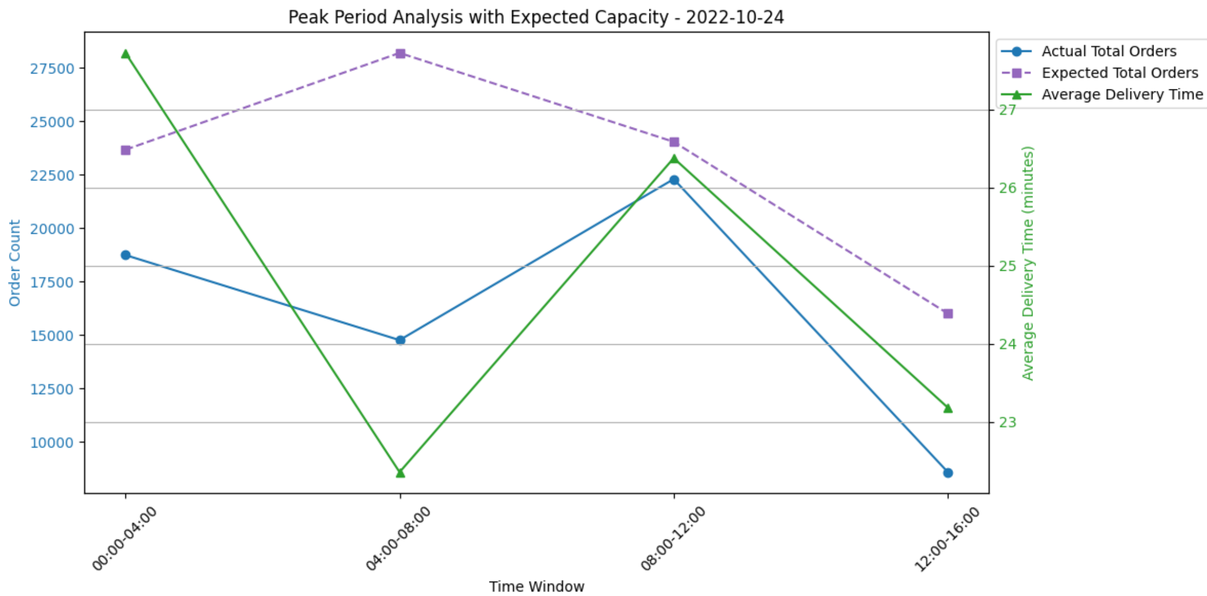

Figure 38. Line Chart for Peak Period 24th October 2022

Courier efficiency during peak periods was examined for the week spanning October 16 to October 24, 2022. To identify peak demand hours, order volumes were examined across four predefined time windows for each day, as shown from Figure 30 to Figure 38.

The expected total orders in these charts represent the estimated fulfilment capacity during each time window, calculated based on the number of available couriers and the average delivery time observed within those intervals. Across most days, the peak period was consistently observed between 8:00 AM and 12:00 PM, as illustrated from Figure 31 to Figure 38

Overall, the system demonstrated effective performance during peak periods, with courier availability and their corresponding fulfillment capacity exceeding the total number of orders received on most days. This indicates a healthy supply-demand balance. However, for the dates between October 21 and October 23, a noticeable decline in fulfillment rate was observed (Figure 35 to Figure 37), where order volumes either matched or exceeded the system's expected capacity during peak period.

This decline can be attributed to increased average delivery times during those peak intervals. Despite these temporary inefficiencies, the platform maintained strong overall performance, with average delivery times remaining below 27 minutes on most days. A notable exception occurred on October 16 (Figure 30), where limited courier availability during late night and early morning hours led to a rise in delivery times, ranging from 25 to 50 minutes.

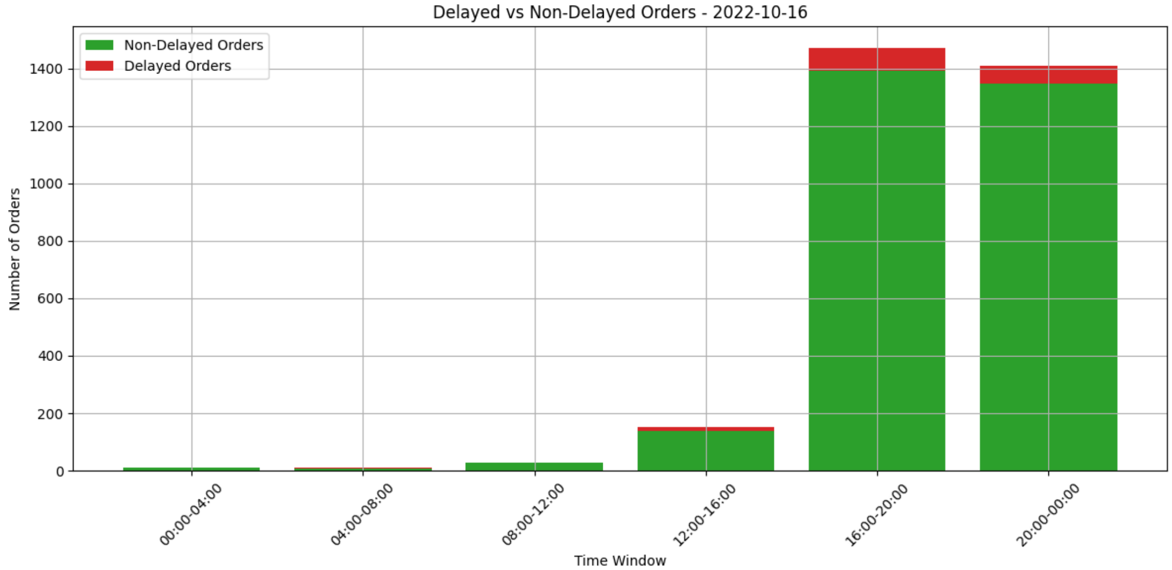

Figure 39. Delay vs Non-Delayed Orders 16th October 2022

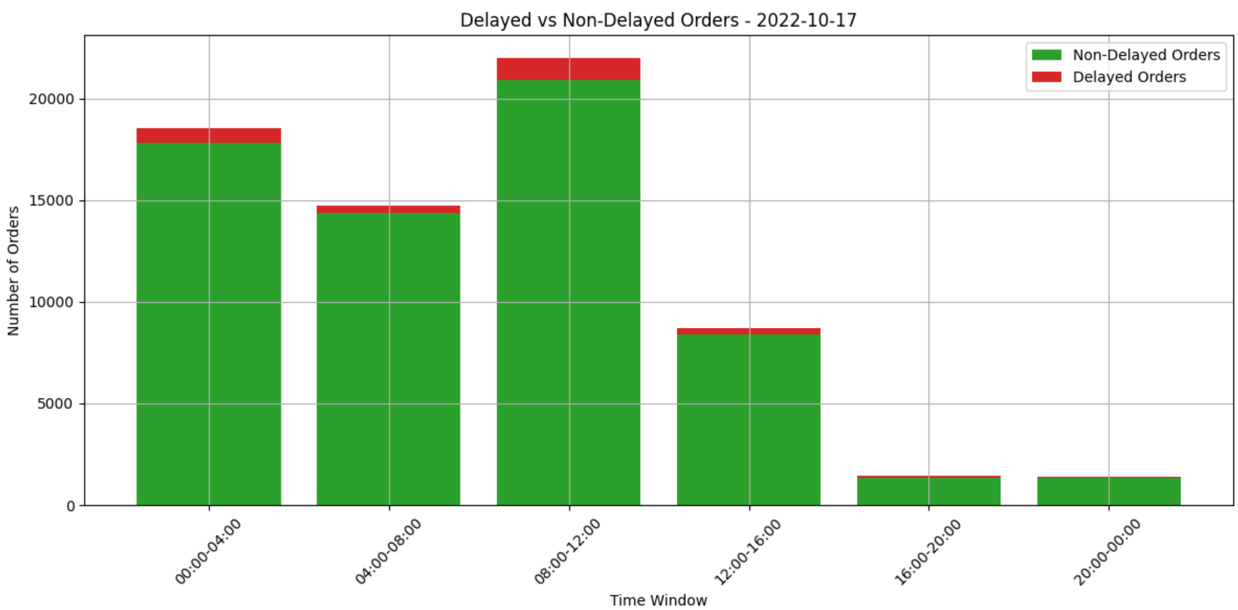

Figure 40. Delay vs Non-Delayed Orders 17th October 2022

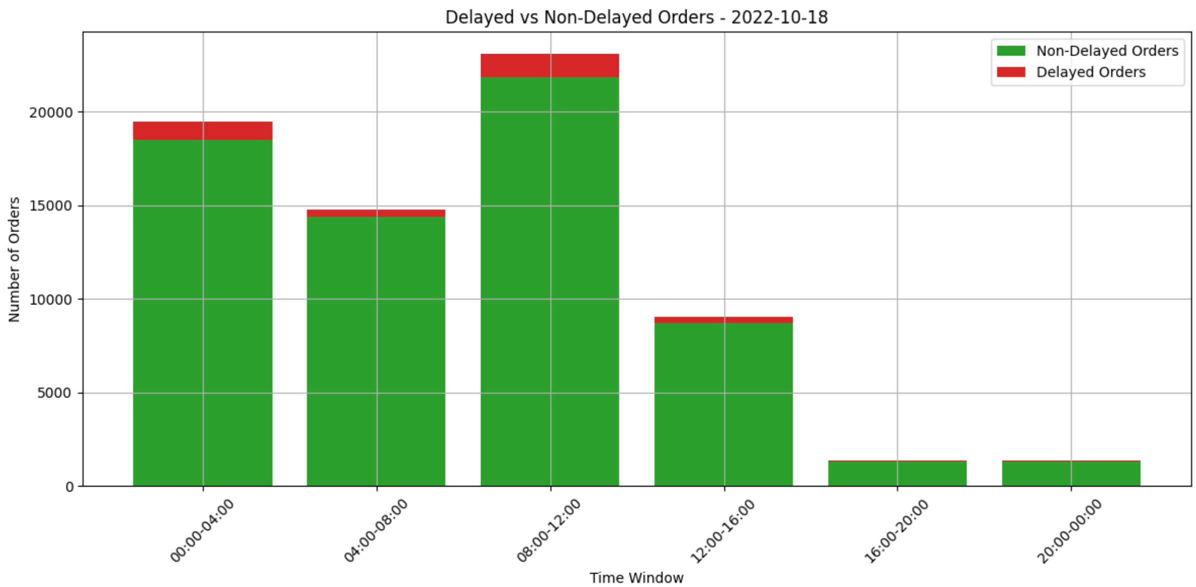

Figure 41. Delay vs Non-Delayed Orders 18th October 2022

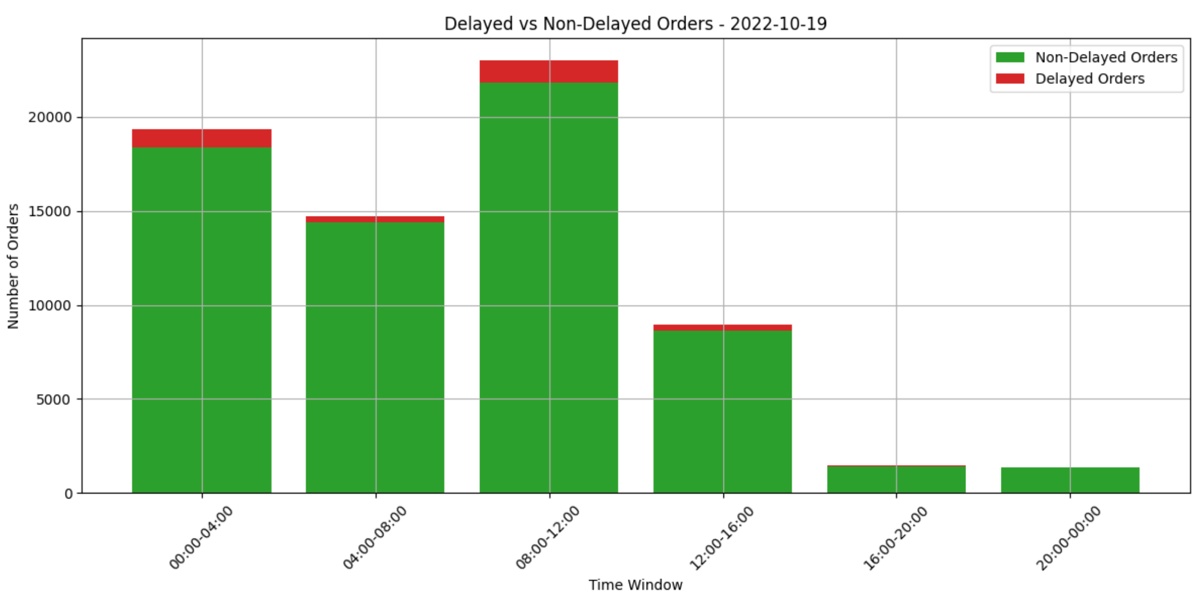

Figure 42. Delay vs Non-Delayed Orders 19th October 2022

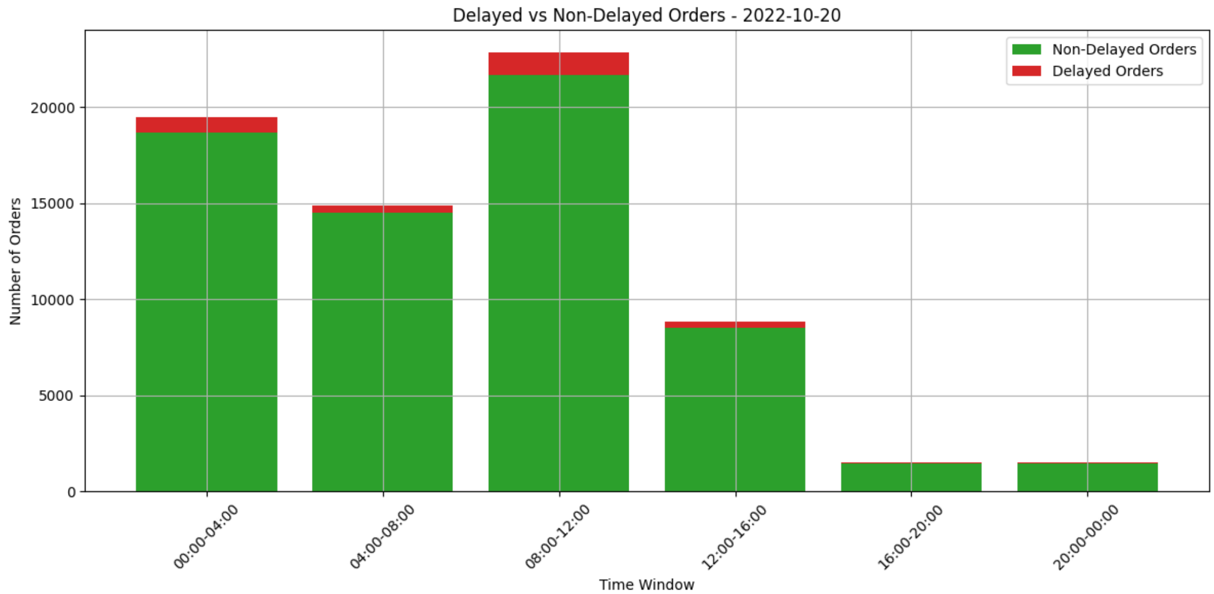

Figure 43. Delay vs Non-Delayed Orders 20th October 2022

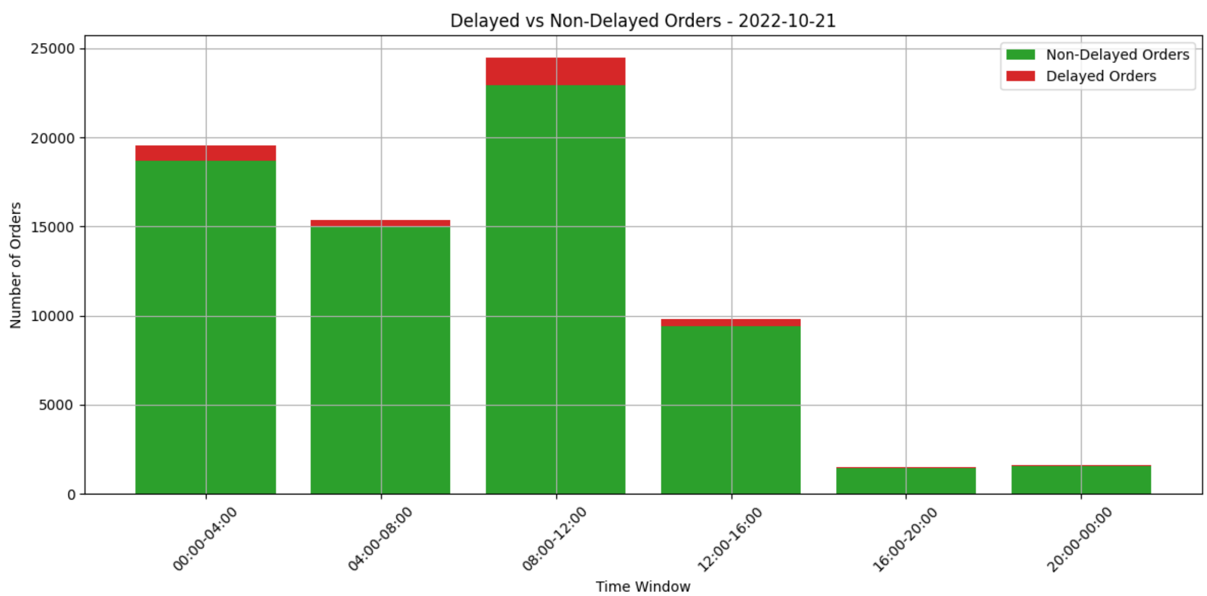

Figure 44. Delay vs Non-Delayed Orders 21st October 2022

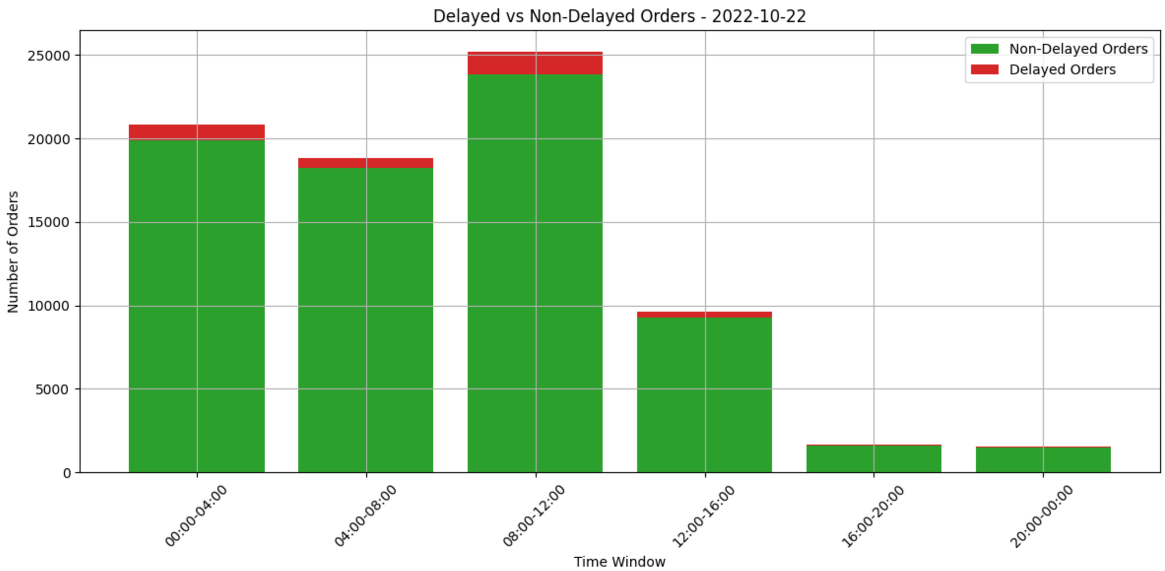

Figure 45. Delay vs Non-Delayed Orders 22nd October 2022

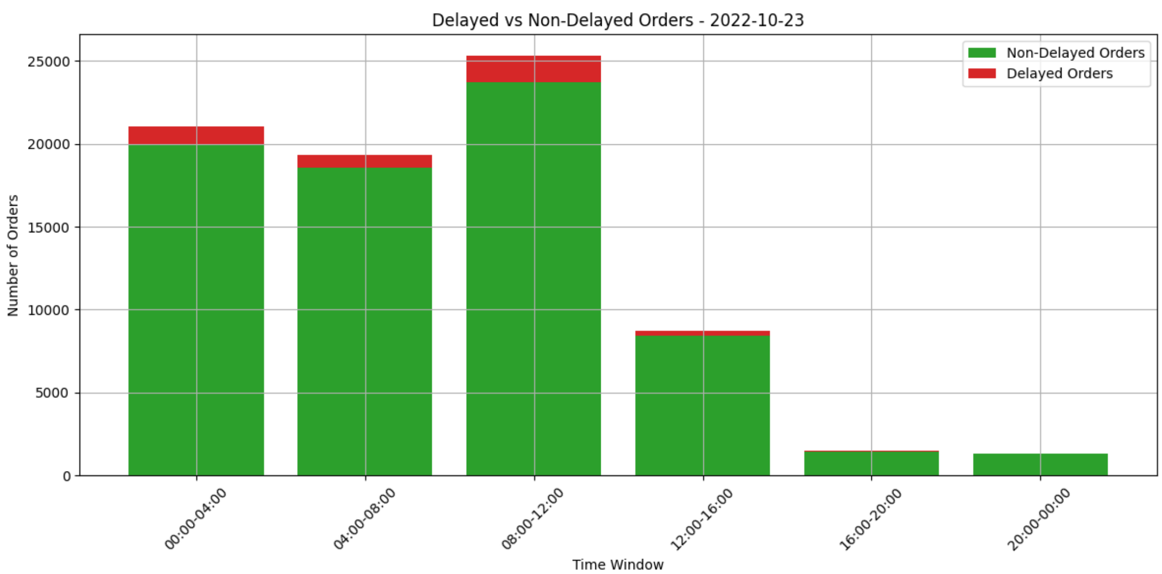

Figure 46. Delay vs Non-Delayed Orders 23rd October 2022

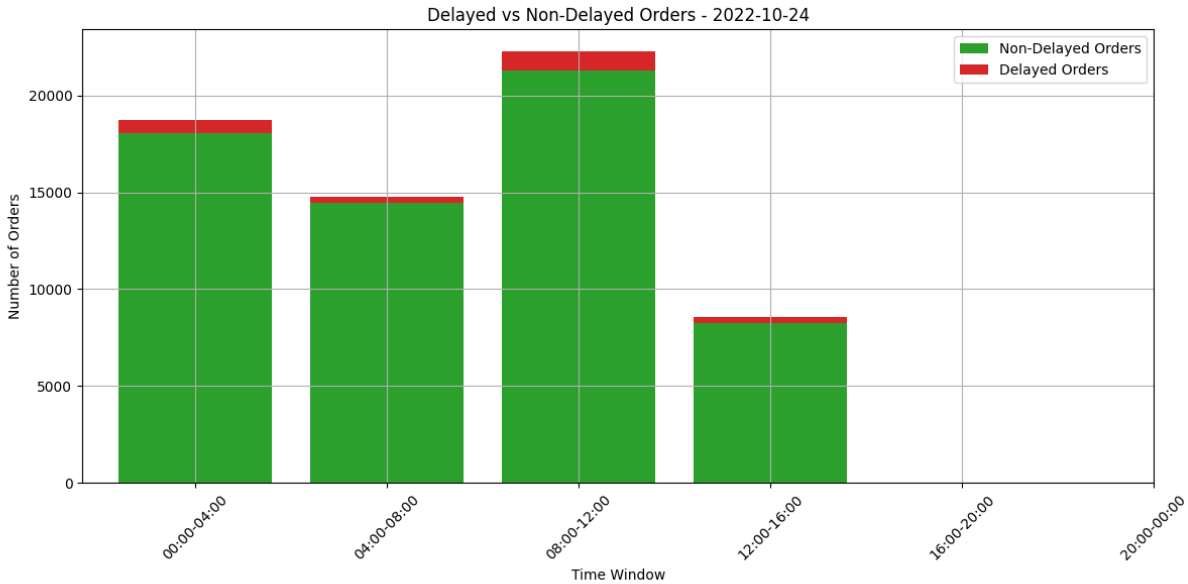

Figure 47. Delay vs Non-Delayed Orders 24th October 2022

To gain deeper insights into operational efficiency, a delivery delay analysis was conducted for the 9-day period from October 16 to October 24, 2022, segmented by the previously defined four-hour time windows (Figure 39 to Figure 47). Overall, delivery performance during this period remained stable and efficient. However, delays were most frequently observed during the late night to early morning (12:00 AM–8:00 AM) and early morning to afternoon (8:00 AM–12:00 PM) intervals (Figure 40 to Figure 47).

The time window from 8:00 AM to 12:00 PM emerged as particularly critical, with a noticeable increase in the delivery delay ratio across all order dates, indicating capacity strain during peak operational hours.

A consistent delay pattern was observed across the late night, early morning, and afternoon periods, where delay peaks were recurrent on each order date. In contrast, the late evening window generally exhibited minimal or no delays, likely due to reduced order volume during that time. An exception to this trend occurred on October 16 (Figure 39), where unusually high late evening order volumes contributed to an elevated delay ratio compared to other time windows on the same day.

4.2 Spatial Temporal Prediction of Future Orders

To forecast hourly order volumes across different geographic regions, various machine learning regression models were developed and evaluated. The objective of implementing these predictive models is to enhance operational efficiency in last-mile delivery logistics by enabling proactive courier allocation. By forecasting expected order volumes across different hours and locations, the platform can better align supply with demand, reducing delivery delays and improving operational efficiency and overall customer satisfaction level.

Models and Features used

The following regression models were trained and compared:

1. Linear Regression (Baseline Model)
2. Random Forest
3. XGBoost
4. LightGBM
5. K-Nearest Neighbors, KNN (K value ranging from 1 to 21)
6. Support Vector Machine (SVM)

Input Features:

1. Area: Geographic region
2. Hour: Hour of the day
3. Day_of_Week: Integer (0–6)
4. Is_Weekend: Binary indicator for weekends
5. Is_Prebook: Binary flag for pre-booked orders
6. Day_Part: Categorical variable (Morning, Afternoon, Evening, Night)

The categorical variable Day_Part was encoded using label encoding to convert it into a numerical format suitable for model input.

Target Variable

Order Count: The total number of orders received per area-hour instance, computed using unique order IDs.

4.2.1 Cross Validation K-Fold and Model Comparison

The dataset contained 3,948 records, with six input features and one target variable. An 80/20 split was applied to separate the data into training and testing sets. To ensure robustness and mitigate risks of bias, overfitting, or underfitting, all models were trained using 5-Fold Cross-Validation. All models were trained and evaluated using their default hyperparameter configurations.

Each model was evaluated using the following performance metrics: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Squared Error (MSE), Mean Absolute Percentage Error and R^2 score.

The K-Nearest Neighbors (KNN) model, which is sensitive to the choice of the hyperparameter k, was evaluated across a range of k values from 1 to 21 to identify the optimal setting. Based on the Mean Squared Error (MSE) from 5-fold cross-validation, the lowest error was observed at $k = 2$, as shown in Figure 48. Consequently, this optimal value was used for the KNN model in subsequent cross-validation comparisons with other models.

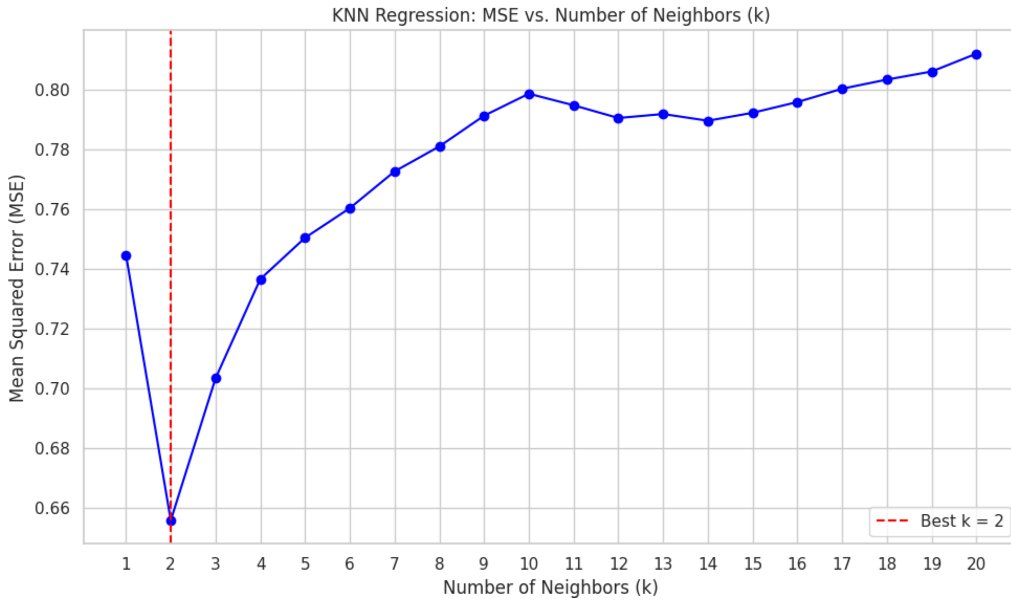

Figure 48. MSE for different k values in KNN

The average results for all the models across all folds are presented in Table 8 below, offering a quantitative assessment of model performance.

Table 8. Model Comparison Results

Models	RMSE	MSE	MAE	MAPE	R2
Linear Regression	1.136	1.291	0.910	0.351	0.433
Random Forest	0.491	0.245	0.247	0.126	0.892
XGBoost	0.511	0.269	0.244	0.121	0.882
LightGBM	0.468	0.226	0.258	0.127	0.901
KNN (K = 2)	0.805	0.653	0.568	0.243	0.714
SVM (RBF Kernel)	0.972	0.947	0.757	0.337	0.583

Among all the models evaluated, LightGBM demonstrated the strongest predictive performance, achieving the lowest RMSE (0.468) and MSE (0.225), along with a high R^2 score of 0.90, indicating excellent model fit and generalization capability. Random Forest and XGBoost followed closely as the second- and third-best performing models, with only marginal differences in error metrics compared to LightGBM.

In contrast, the Linear Regression model and distance-based algorithms such as K-Nearest Neighbors (KNN) and Support Vector Regressor (SVR) performed relatively poorly. These models were less effective at capturing the complex, non-linear relationships within the spatiotemporal features, leading to higher prediction errors and lower R^2 scores. A similar pattern is observed in the MAPE values, with LightGBM, XGBoost, and Random Forest achieving comparable results of 0.127, 0.121, and 0.126 respectively, followed by KNN at 0.243 and SVM at 0.337.

Table 9. Model Comparison Results with Original Order Volume Units

Models	RMSE	MSE	MAE	MAPE	R2
Linear Regression	223.89	50129.87	104.64	1.96	0.103
Random Forest	59.32	3519.41	18.89	0.87	0.94
XGBoost	58.93	3473.27	19.01	0.82	0.94
LightGBM	77.10	5944.55	25.79	0.79	0.89
KNN (k=2)	142.99	20447.69	59.27	3.34	0.63
SVM (RBF Kernal)	211.07	44549.46	93.11	2.74	0.20

The actual and predicted order volumes were reverse log-transformed to ensure that error metrics and R-squared values could be summarized and compared in the original order count units. Table 9 presents the results of the model comparison. A consistent pattern is observed among tree-based models, with Random Forest, XGBoost, and LightGBM producing lower error metrics and higher R-squared values.

While Random Forest and XGBoost showed slightly lower prediction errors than LightGBM when evaluated in the original order count units, LightGBM was chosen for final testing on the full training dataset for two main reasons:

- LightGBM delivered better performance when the target variable was log-transformed to address skewness. This outcome is considered more reliable, reducing the risk of bias, underfitting, and overfitting.
- LightGBM is more computationally efficient than the other models. Given the highly variable nature of order volumes and the complexity of real-world food delivery datasets, LightGBM's faster prediction capability is particularly valuable in dynamic environments with large and high-dimensional datasets.

4.2.2 Best Model Evaluation on Full training Dataset

After identifying LightGBM as the best-performing model through 5-Fold Cross-Validation, it was retrained on the entire training dataset and evaluated on the held-out test set to assess performance consistency and generalization capability. The test results on the full training set are shown in Table 10 below:

Table 10. LightGBM Results on Full Training Set.

LightGBM				
RMSE	**MSE**	**MAE**	**MAPE**	**R2**
0.433	0.187	0.246	0.127	0.925

The evaluation on the test data showed further improvements in model accuracy. The RMSE decreased to 0.43, MSE to 0.187, and MAE to 0.246, while the R^2 score increased slightly to 0.925. These results confirm that the LightGBM model not only performs well during cross-validation but also generalizes effectively to unseen data, making it a reliable choice for spatiotemporal order volume prediction.

Table 11. LightGBM Results in Original Order Volume Units.

LightGBM				
RMSE	**MSE**	**MAE**	**MAPE**	**R2**
86.29	7447.56	23.18	0.59	0.88

Table 11 presents the LightGBM results in the original order count units after training on the full training dataset. The results indicate that the model successfully explains 88% of the variation in order volumes across different regions and times. While the model's predictions show an average error of 86.29 orders, this is considered reasonable given the substantial variability in daily order volumes. For instance, as shown in Figure 50 (Section 5.1), the daily order counts range from 86 to 14,000, highlighting the high fluctuation in demand within the on-demand food delivery sector. In this context, although an RMSE of 86.29 may initially seem large, it represents a relatively small error when compared to the wide range of order volumes, indicating that the model generalizes well overall.

It is also important to note that the dataset contained only a limited number of relevant features. The model's predictive accuracy could potentially be improved with the inclusion of additional information such as weather conditions, road blockages, traffic congestion, vehicle types, and route directions, as these factors could help the model capture more nuanced patterns in order volume fluctuations.

4.2.3 Light GBM Prediction on Unseen Data

To further evaluate the generalization capability of the best-performing model, 10 observations were held out as unseen data prior to splitting the dataset into training and test sets. These samples were used exclusively for final validation after model development.

The LightGBM model was applied to predict order volumes on this unseen dataset. The resulting scatter plot of actual vs. predicted order counts (Figure 49) demonstrates strong alignment between predicted values and the best-fit line, confirming the model's reliability. The predictions are highly accurate for low to moderate order volumes, while a few deviations are observed in higher order counts. Nonetheless, the model exhibits excellent overall performance and robust generalization across diverse spatiotemporal contexts.

These findings reinforce the model's suitability for real-world deployment in predicting future order volumes and optimizing courier allocation strategies.

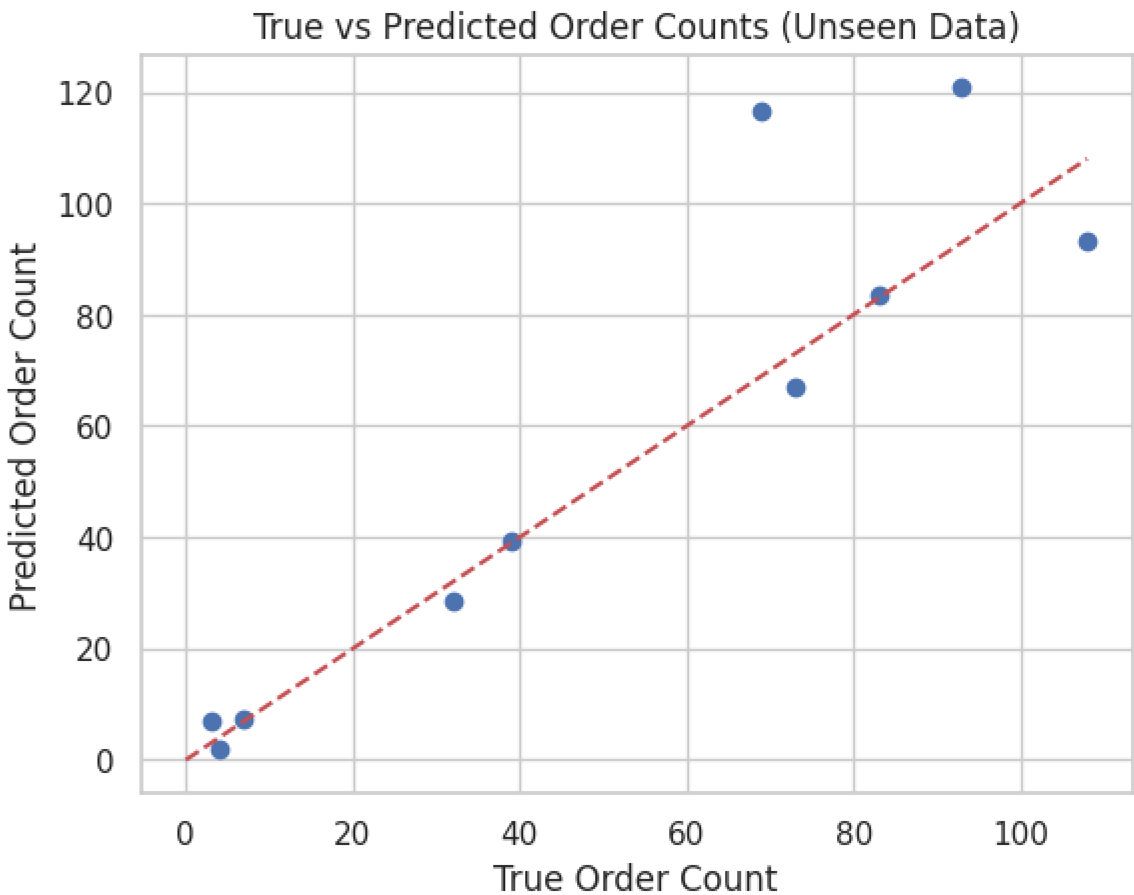

Figure 49. Actual vs True Order Count

Chapter 5. Conclusion

This study investigates the operational dynamics of last-mile delivery networks, with a particular focus on on-demand food delivery services. The analysis is conducted through two complementary approaches:

- Exploratory Data Analysis (EDA) to examine courier behavior and performance.
- Machine Learning-based prediction of future order density across spatial and temporal dimensions.

Insights derived from these analyses support data-driven decision-making for more efficient resource allocation, enabling improved delivery performance and better alignment with customer expectations.

5.1 Exploratory Data Analysis Insights

This study investigated the operational dynamics of last-mile food delivery systems by conducting an in-depth exploratory data analysis (EDA) of courier behavior and performance. The analysis was structured across three levels: order-level metrics, courier activity, and temporal dynamics during peak periods. These dimensions collectively offer valuable insights into efficiency bottlenecks, behavioral patterns, and strategic courier allocation potential.

Order-Level Courier Performance

The analysis included 4,955 unique couriers and focused on key performance indicators (KPIs), including average delivery time, distance traveled per order, order completion count, and acceptance rate. Descriptive statistics and visualizations highlighted high variability across these metrics. A majority of couriers maintained high engagement, as evidenced by an average acceptance rate of 84.7% (Figure 5). However, the presence of a long-tail distribution, with some couriers exhibiting acceptance rates between 40% and 70%, suggests the need to revisit order assignment policies (Figure 4).

Average delivery times generally ranged between 23 and 29 minutes, with a mean of approximately 27 minutes (Figure 7). While most couriers performed within this efficient bracket, a smaller subset averaged 30 to 50 minutes (Figure 6), indicating potential inefficiencies. The average distance traveled per delivery was 1.89 km, with outliers reaching up to nearly 6 km (Figure 9). Relying solely on average delivery time or distance for performance evaluation is insufficient; instead, a multidimensional analysis offers a more comprehensive understanding.

Using a multidimensional scatterplot, courier performance was examined based on average delivery time, acceptance rate, and distance travelled (Figure 10). The results revealed a direct correlation between distance and delivery time, as expected. Interestingly, couriers with higher acceptance rates tended to travel shorter distances and demonstrated quicker delivery times. This suggests a pattern where couriers frequently accepting orders are likely operating within localized, high-density zones, enabling faster turnaround.

These findings can support smarter and more effective management of order assignments. For instance, couriers who consistently accept a high number of orders and complete them on time could be rewarded with performance-based incentives, such as bonuses, to keep them motivated. On the other hand, couriers with lower acceptance rates and longer delivery times can be identified and offered targeted training to help improve their performance.

The analysis also uncovered valuable insights from the relationship between acceptance rate, delivery time, and distance traveled. Couriers with higher acceptance rates tend to travel shorter distances, which suggests they're often delivering within the same area or to nearby zones. This insight can be used to improve resource planning especially during peak and off-peak periods to boost operational efficiency across the system.

Spatial Analysis

To understand how spatial dynamics, influence performance, couriers were categorized as either Specialized (operating at the order origin region) or Multi-Area (delivering across different zones). The "cross-area rate" metric quantified the proportion of deliveries that occurred outside the pickup area. The histogram revealed a bimodal distribution, with peaks at 0% and between 60% to 100% (Figure 12), reflecting two dominants behaviors: strict area adherence versus extensive geographic mobility.

The mean cross-area delivery rate was 70% (Figure 13), highlighting that the majority of couriers are engaged in multi-area deliveries. However, comparative analysis showed that Specialized couriers consistently outperformed their Multi-Area counterparts (Figure 16). With just 366 couriers identified as Specialized (vs. 4,535 Multi-Area couriers), they consistently demonstrated better performance in terms of higher acceptance rates, shorter delivery times, and lower distances traveled. This suggests they are more efficient due to familiarity with their operating zones and faster turnaround times, whereas Multi-Area couriers may suffer from reduced availability due to longer travel times and reduced delivery frequency.

A scatterplot visualizing cross-area rate against acceptance and delivery time further emphasized these trends (Figure 14). Multi-Area couriers, while accepting a large number of orders, had longer delivery times, likely due to longer distances and unfamiliar routes. Specialized couriers, in contrast, benefited from consistent delivery zones, allowing faster, more frequent order completion.

These findings regarding specialized and multi-area couriers can support more effective, geographically informed order assignments during both peak and off-peak periods. Multi-area couriers, who travel across regions, can be tasked with more complex or critical deliveries as well as batch orders. In contrast, specialized couriers, with their experience in localized deliveries, can be strategically relocated to regions with higher demand. For instance, Figures 50 and 51 below present heatmaps illustrating the total number of orders received in various areas and the total number of orders delivered to recipient (customer) areas, respectively.

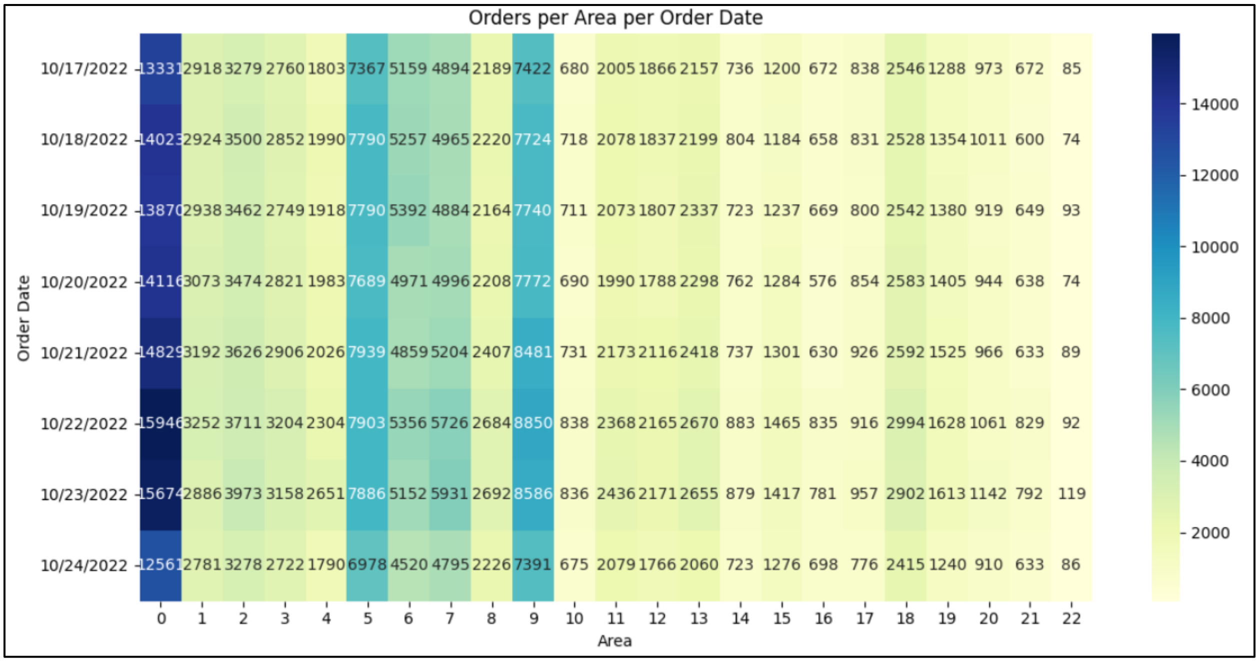

Figure 50. Total Number of Orders received in Different Areas each day

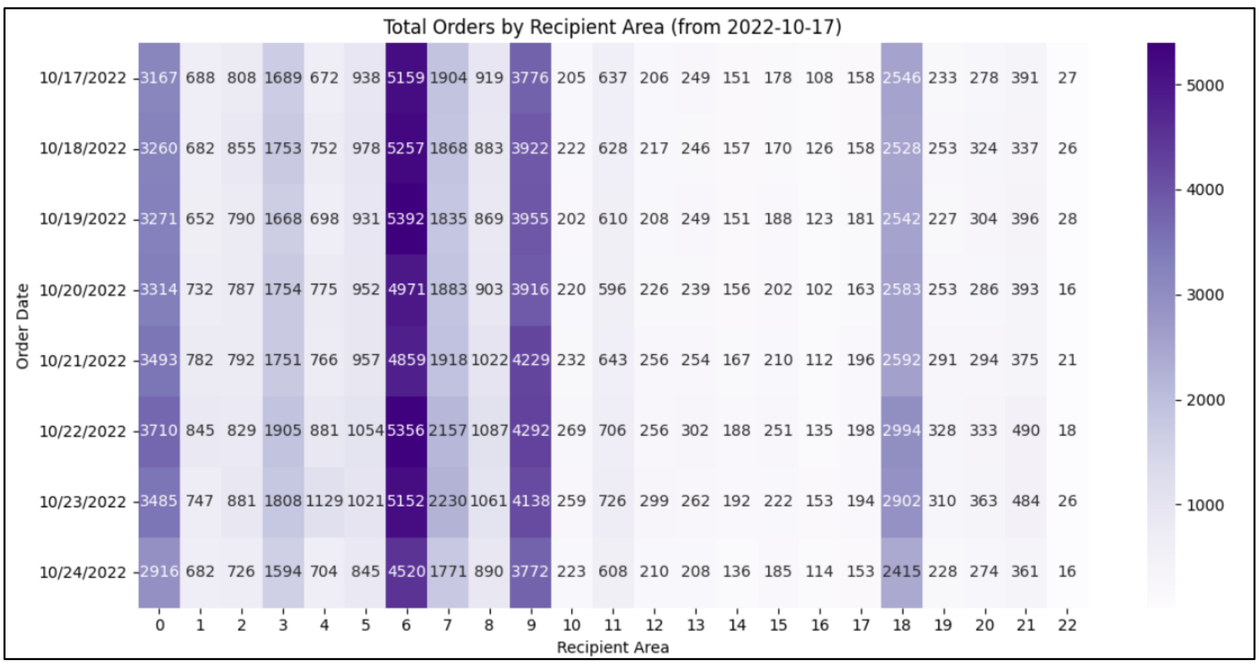

Figure 51. Total number of Orders delivered in recipient areas each day.

The heatmap reveals that, for example, on October 17, 2022, area 0 received a total of 13,331 orders (Figure 50), of which 3,167 (Figure 51), approximately 23.8% were delivered within the same region. These insights can help guide the strategic relocation of specialized couriers to high-demand regions where order origins and delivery locations coincide. Furthermore, since there are significantly more couriers handling cross-region deliveries compared to those operating solely within the order's origin region, encouraging a balanced distribution of order assignments is important. This can be achieved by monitoring each courier's order acceptance behavior. Specialized couriers typically accept orders when the origin and delivery locations are the same. However, these couriers could be trained to handle cross-region deliveries or incentivized with compensation or rewards to expand their service areas, especially during peak periods.

To explore interrelationships among performance variables, a correlation analysis was conducted (Figure 24). Strong positive correlations were found between delivery time and distance traveled, and moderate positive correlation between cross-area rate and delivery time. Moderate negative correlations were observed between acceptance rate and both cross-area rate and distance traveled. This implies that couriers with high acceptance rates typically perform localized deliveries, resulting in shorter distances and faster completion times.

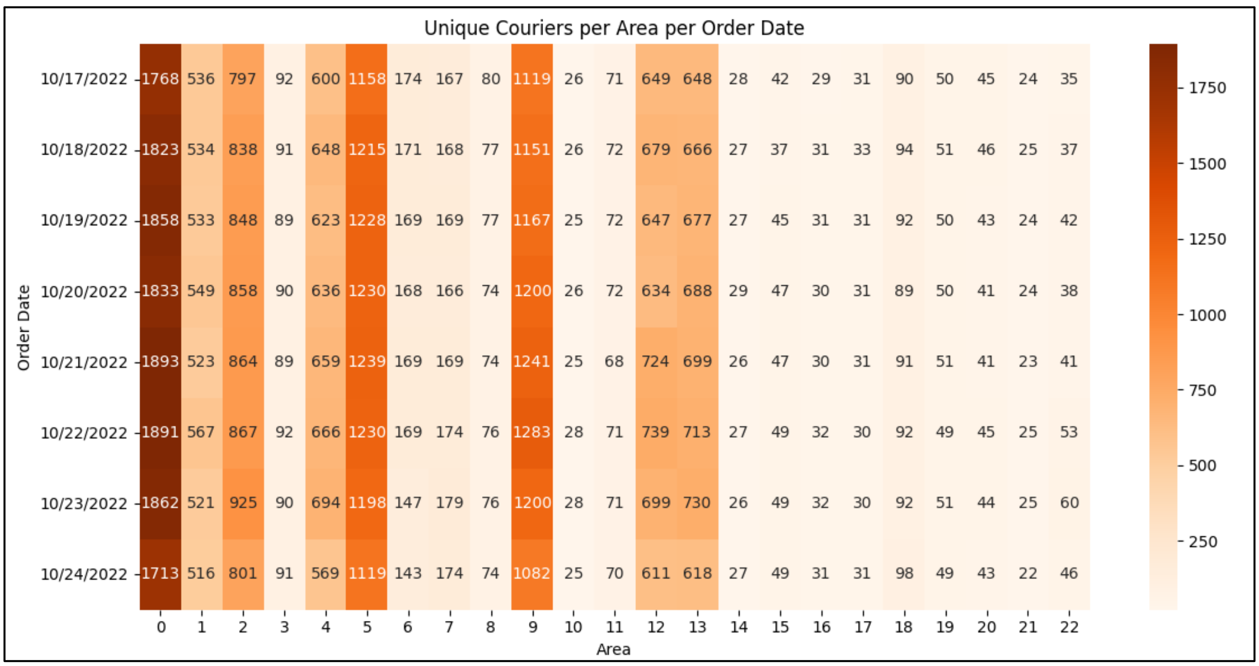

Figure 52. Number of Unique Couriers in each Order Area

Figure 52 highlights the number of distinct couriers operating in each order origin area on different order dates. While this study didn't include data on the exact delivery routes taken by couriers, due to its unavailability, combining insights from Figures 50, 51, and 52 with routing information in the future could support more efficient order batching and courier dispatch strategies. This could lead to reduced customer wait times and improved operational efficiency. Moreover, grouping order volumes by nearby or similar geographic regions, along with understanding where different couriers are active, can provide valuable insights into courier performance and behavior. This approach can help identify individual strengths and weaknesses, supporting more informed planning and performance improvements down the line.

Courier Incentivize and Capacity Planning

Wave analysis was used to investigate courier activity spans, with a wave defined as a continuous period during which a courier was actively completing deliveries. The average wave duration was 35 minutes (Figure 26), indicating short bursts of activity. While most couriers completed between 1 to 3 orders per wave, a small segment completed 5 or more, with rare instances reaching 10–20 orders (Figure 29), often accompanied by extended wave durations of up to 712 minutes (11.86 hours).

Inactive status of courier between waves averaged 27.5 minutes (Figure 28), but showed a rightly skewed distribution. However, some couriers experienced times exceeding 100 minutes (Figure 27), with extreme cases approaching four hours. These findings suggest an imbalance in supply-demand alignment and underutilization of courier resources. High inactive periods indicate lower courier engagement and inefficient platform operations.

These insights can help operational managers optimize resource capacity planning and strategically relocate couriers to high-demand areas. First, implementing a notification or alert system within the app can inform couriers about order frequencies, helping reduce their inactive duration and increase active engagement. Second, during peak periods such as shown in Figure 50, where area 0 consistently receives the highest order volume followed by areas 5 and 9, proactively relocating the maximum number of couriers to these regions can minimize downtime and boost efficiency. Third, price responsiveness strategies can promote a more balanced workload distribution during peak times by financially rewarding couriers who are willing to travel longer distances or handle batch orders. Conversely, during off-peak periods, the company can offer guaranteed minimum pay to couriers who remain active, thereby reducing downtime.

Temporal Analysis - Peak period

Temporal analysis divided each operational day into six four-hour windows. Across these periods, the morning-to-noon window (08:00–12:00) consistently emerged as the peak period (Figure 31 to Figure 38). For most days, courier availability exceeded order demand, indicating a stable supply-demand balance. However, between October 21st and 23rd, this balance was disrupted such that order volumes equaled or surpassed available capacity during peak times (Figure 35 to Figure 37).

Moreover, delay analysis revealed higher instances of late deliveries during early mornings, late nights, and afternoon periods. Notably, the 08:00–12:00 slot remained critical (Figure 40 to Figure 47), due to high order density and strained courier availability. The root cause appears to be insufficient courier availability during elevated demand, underscoring the need for more agile courier deployment strategies during known peak times.

The temporal analysis of peak and off-peak periods, combined with insights on delayed orders caused by insufficient courier capacity, supports dynamic courier relocation and shift scheduling. Figures 50 and 51 present heatmaps of total orders received and deliveries made across different areas. Integrating these findings with temporal patterns enables proactive adjustment of courier deployment to high-demand regions, especially focusing on those consistently achieving key performance indicators (KPIs).

Figures 19 and Figure 20 highlights the top 10 fastest and slowest couriers based on average delivery times. Categorizing couriers by performance can assist managers in assigning reliable performers to regions with high order volumes during peak hours. Additionally, these insights can guide decisions on hiring the optimal number of in-house or freelance couriers to effectively meet growing customer demand.

5.2 Spatial Temporal Machine Learning Prediction

The second major component of this study focused on predicting hourly order volumes across different geographic areas using supervised machine learning regression models. The objective was to develop a predictive framework that could support proactive courier allocation and workload management, particularly during periods of fluctuating demand in last-mile delivery operations.

A total of six regression models were trained and evaluated: Linear Regression, Random Forest, XGBoost, LightGBM, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM). Linear Regression served as a baseline model, providing a benchmark for detecting any linear patterns between the input features and the target variable. The remaining non-linear models were selected for their ability to capture complex relationships within high-dimensional datasets, a characteristic commonly observed in spatiotemporal delivery systems.

The input features used in model training included: Hour of the day, Day of the week, Area, Weekend indicator, Pre-booking status, and Day part (e.g., morning, afternoon, evening, night). The target variable, Order Count, was computed from the Order ID field by aggregating the number of orders received per area per hour.

The dataset was partitioned into an 80% training set and a 20% test set. To ensure a robust evaluation of model performance, all models were trained using 5-Fold Cross-Validation on the training set. This technique splits the training data into five equal parts, training the model iteratively on four folds while validating it on the fifth. Performance metrics (RMSE, MSE, MAE, MAPE and R^2 Score) were averaged across all folds. This approach was used to mitigate the risks of overfitting, underfitting, and sampling bias, ensuring a more stable and generalizable comparison of models.

Among the evaluated models, LightGBM emerged as the best performer. During cross-validation, it achieved the lowest RMSE (0.468) and a high R^2 score of 0.90 (Table 8). When retrained on the full training set and evaluated on the test set, its performance further improved, achieving an RMSE of 0.43 and R^2 of 0.925 (Table 10). These results reflect LightGBM's ability to efficiently model complex, non-linear relationships in structured tabular data, making it particularly effective for real-time demand forecasting in last-mile delivery environments.

In contrast, Linear Regression, KNN, and SVM performed comparatively poorly, likely due to their limited ability to model non-linear interactions and spatial dependencies in the data. While

Random Forest and XGBoost demonstrated stronger performance, LightGBM consistently delivered slightly better results across all metrics.

To further validate its generalization capability, the LightGBM model was tested on a small unseen dataset of 10 observations, which had been excluded from all previous training and validation phases. The predicted values closely matched the actual order counts, especially in low to moderate volume ranges (Figure 49). While slight deviations were observed for higher order volumes, overall predictive performance remained strong, reinforcing the model's robustness and practical applicability in real-world scenarios.

These findings underscore the feasibility of applying machine learning models, particularly gradient boosting algorithms, for spatiotemporal demand forecasting in delivery systems. Moreover, when integrated with the earlier exploratory analysis of courier behavior and performance, the predictive framework provides a comprehensive, data-driven approach for informed and adaptive operational planning in last-mile logistics.

Table 12. Number of Couriers Available against Predicted and Actual Order Count

Order Date	**Area**	**Hour**	**Day of week**	**Actual Order Count**	**Predicted Order Count**	**Number of available couriers**
23rd October 2022	18	1	6	83	83.39	40

Table 12 above, highlights a single observation from October 23, 2022, where the model predicted an hourly order volume of 83.39, remarkably close to the actual count of 83. However, only 40 couriers were active during that time. When this insight is combined with the temporal and courier wave-level analysis, it becomes clear how valuable such information can be for proactive resource planning.

From the wave analysis, we know that couriers are typically active for around 35 minutes per wave, during which they complete 1 to 3 orders. Based on this, having 40 couriers to handle 83 orders means each would need to complete about 2 deliveries within the hour. While that seems achievable on average, real-world challenges like traffic congestion, road closures, or bad weather, can easily reduce a courier's ability to meet that target. Since these external factors weren't included in the study, the actual courier capacity is likely slightly underestimated in ideal conditions. This suggests that the number of available couriers at that time was just a bit short. Managers can use these predictive models to anticipate order volumes in advance, especially during peak hours like 8:00 AM to 12:00 PM (as identified in the temporal analysis), and adjust courier deployment accordingly. This helps improve operational efficiency, reduce delivery delays, and enhance customer satisfaction.

Since this study forecasts demand at an hourly level, the model can be integrated into a real-time dashboard that managers can monitor regularly to make timely decisions about workforce allocation. These insights can also support dynamic pricing strategies to manage sudden surges in demand, like what happened from October 21 to 23, when orders outpaced available couriers and caused delays. On a strategic level, the same data-driven approach used in this study can be applied when expanding into new regions. Lessons learned from one area can inform courier hiring, resource planning, and operational strategy in other markets

5.3 Managerial Implications

The integration of machine learning–based spatiotemporal order prediction with behavioral and performance insights from exploratory data analysis (EDA) offers a robust foundation for data-driven decision-making in last-mile delivery operations. This integrated analytical framework enables platform managers to proactively plan courier allocation, optimize workload distribution, and respond more effectively to fluctuating demand. The following implications highlight key areas where this study's findings can be applied to enhance operational efficiency and transition to proactive logistics management.

1. Effective Resource Allocation through accurate forecasting

Accurate spatial-temporal forecasting using predictive models can significantly improve operational efficiency by enabling proactive resource allocation. For example, in the ride-hailing sector, real-time demand forecasting allows platforms to rebalance vehicle distribution across regions, improving fleet utilization and customer satisfaction (Ke et al., 2017) . Similarly, in the retail sector, accurate customer flow predictions help optimize inventory levels and staff scheduling across locations and time periods (Liu, 2024).

Given the high variability in the on-demand food delivery sector, similar forecasting approaches can be applied to support more informed decision-making. In this study, the LightGBM model outperformed other models in predicting hourly order volumes across different regions. Combined with insights from exploratory data analysis (EDA) which revealed a consistent peak period between 8:00 AM and 12:00 PM, these predictions can help operations managers schedule more couriers during high-demand hours or allocate them to regions with higher order density. This can ultimately improve order fulfillment rates and reduce delivery delays.

2. Geographically Informed Assignment

Geospatial attributes play a critical role in shaping the efficiency of on-demand services like ride-sharing and food delivery. For instance, in the on-demand transit sector, Liu et al. (2023) recommended allocating more private e-bike parking near educational institutions due to consistently higher demand in those areas. This kind of spatial analysis can similarly enhance operational strategies in last-mile food delivery.

In this study, exploratory data analysis (EDA) revealed notable performance differences between Specialized and Multi-Area couriers. Couriers who picked up and delivered within the same zone had higher acceptance rates and shorter delivery times, while those handling cross-zone deliveries often experienced lower performance. These insights suggest that operations managers can boost efficiency by aligning courier assignments with geographical familiarity and historical performance. High-demand zones can be consistently assigned to couriers who excel in those areas, while more experienced or flexible couriers can be designated for complex, multi-zone deliveries.

3. Recommendation System

Operational efficiency in on-demand services can be significantly enhanced through the use of data-driven recommendation systems. For example, Wang et al. (2019), in their study on trajectory analysis, highlighted how recommendation systems in on-demand transit help both drivers and passengers by suggesting faster travel routes, ultimately reducing idle time and improving service efficiency.

Building on this idea, the integration of exploratory data analysis (EDA) with spatial-temporal machine learning models, as demonstrated in this study, can support two forms of recommendation systems in the food delivery sector. Platforms such as Uber Eats, DoorDash, or Meituan could adopt similar systems to:

- Segment couriers based on performance, and
- Support targeted marketing or promotional campaigns.

From a marketing perspective, analysis of high-order-density regions can be used to tailor promotions or discounts to customers in those areas. If additional data, such as preferred cuisine types or frequently ordered restaurant categories is available, this information can further refine personalized offers to drive customer engagement and boost sales.

From an operations standpoint, courier segmentation based on order and wave-level performance metrics (e.g., acceptance rate, active status, delivery punctuality) can help identify top-performing couriers. These couriers can be prioritized for high-value or time-sensitive deliveries and rewarded through incentive programs to maintain their motivation and reliability. Meanwhile, couriers who show lower performance can receive targeted support such as training, performance coaching, or assignments within familiar local zones to help improve their efficiency over time.

4. Scheduling Workforce shifts

Ke et al. (2017) emphasized how spatial-temporal demand forecasting in the on-demand transit sector can help balance vehicle distribution and improve utilization rates. Similarly, Liu (2024) demonstrated that accurate predictions of customer flow in the retail sector allow managers to better align staffing schedules to meet service demand and enhance customer satisfaction. These insights are highly transferable to the food delivery sector. This study used real-world delivery data from Meituan, a platform that operates with a crowdsourced courier model. However, the findings are particularly relevant for companies that employ in-house couriers, such as Getir, where workforce scheduling is more centralized.

For example, wave-level analysis in this study showed that most couriers completed between 1 to 3 deliveries per wave, while a few handled up to 20. Additionally, inactive time distribution patterns revealed underutilization among certain couriers. Operations managers can use this type of insight to design better shift schedules that promote equitable workload distribution, reduce courier downtime, and maintain consistent service quality across time periods.

5. Strategic Resource Expansion and Capacity Planning

The integration of exploratory data analysis (EDA) with spatial-temporal forecasting using machine learning models can also play a vital role in strategic planning related to resource expansion and capacity management. For instance, Liu (2024) demonstrated how forecasting customer flow in the retail sector helps managers plan ahead for staffing and resource needs to meet future demand. Similarly, Sanaullah et al. (2021) provided policy recommendations in the on-demand transit sector to adjust fleet size in response to real-time passenger demand, aiming to reduce waiting times and improve service reliability.

In the context of food delivery, the integrated analytical approach used in this study can support long-term planning decisions across several dimensions. Predictive models offer data-driven insights for workforce scaling, courier fleet adjustments, and regional expansion. As demand grows, whether due to seasonal surges or promotional campaigns, more couriers and delivery resources can be proactively deployed. Additionally, when new delivery zones are introduced, historical patterns from EDA and forecasting outputs can serve as a benchmark to guide courier onboarding, set performance expectations, and design efficient routing strategies tailored to the new regions.

6. Disruption impact on Order Demand

Cottreau et al. (2025) demonstrated how analyzing service disruptions across space and time in public transit systems can support more effective operational management and resilience planning. A similar approach can be applied in the food delivery sector, where sudden disruptions such as adverse weather, heavy traffic, road accidents, or blockages can significantly impact delivery performance.

If historical or real-time disruption data is available, it can be integrated into existing predictive and spatial-temporal models to help managers anticipate and respond to such events more effectively. This enables the development of proactive risk management strategies, such as temporarily controlling order volumes, reallocating couriers to unaffected zones, or offering incentives to customers in case of expected delays.

5.4 Limitations and Future Research Direction

This research presented an analytical framework that integrates machine learning-based spatiotemporal prediction of future order density with exploratory data analysis (EDA) of courier behavior and performance. The goal was to support proactive and efficient courier allocation and resource planning in last-mile delivery operations. However, several limitations related to the dataset and modeling approach must be acknowledged.

Firstly, the machine learning models did not incorporate external contextual variables such as weather conditions, traffic congestion, and road closures, factors that can significantly impact delivery time and operational efficiency. Secondly, the EDA component did not include courier-specific attributes such as delivery experience, fleet availability, or customer ratings, which could offer deeper insights into behavioral patterns and performance variability.

Future research should consider integrating external contextual features, such as real-time traffic and weather data, into both the exploratory and predictive components of the analysis. Moreover, advanced modeling techniques such as Long Short-Term Memory (LSTM) networks, artificial neural networks, or reinforcement learning could be explored to enhance forecasting accuracy and adaptability. Lastly, future studies could bridge the gap between data analysis and operational execution by implementing simulation-based frameworks that use model predictions and courier behavior insights to support real-time courier allocation in response to dynamic order density patterns.

References

Aburas, M. M., Ahamad, M. S. S., & Omar, N. Q. (2019). Spatio-temporal simulation and prediction of land-use change using conventional and machine learning models: a review. Environmental Monitoring and Assessment : An International Journal Devoted to Progress in the Use of Monitoring Data in Assessing Environmental Risks to Man and the Environment, 191(4), 1–28.

Aljoufie, M., Zuidgeest, M., Brussel, M., & van Maarseveen, M. (2013). Spatial-temporal analysis of urban growth and transportation in Jeddah City, Saudi Arabia. Cities, 31, 57–68.

Alnaggar, A., Gzara, F., & Bookbinder, J. H. (2021). Crowdsourced delivery: A review of platforms and academic literature. Omega, 98.

Amato, F., Guignard, F., Robert, S., & Kanevski, M. (2020). A novel framework for spatio-temporal prediction of environmental data using deep learning. Scientific Reports, 10(1).

An, L., Tsou, M. H., Crook, S. E. S., Chun, Y., Spitzberg, B., Gawron, J. M., & Gupta, D. K. (2015). Space–Time Analysis: Concepts, Quantitative Methods, and Future Directions. Annals of the Association of American Geographers, 105(5), 891–914.

Arishi, A., Krishnan , K. and Arishi, M. (2022) Machine Learning Approach for truck-drones based last-mile delivery in the era of industry 4.0, Science Direct.

Bathke, H., & Munch, C. (2024). From Occasional to Active Crowdshippers: The Significance of Couriers' Characteristics. IEEE Transactions on Engineering Management, 71.

Bozanta, A. et al. (2021) Courier routing and assignment for food delivery service using reinforcement learning, Computers & Industrial Engineering.

Brochado, Â. F., Rocha, E. M., Addo, E., & Silva, S. (2024). Performance Evaluation and Explainability of Last-Mile Delivery. Procedia Computer Science, 232, 2478–2487.

Cai, P., Wang, Y., Lu, G., Chen, P., Ding, C., & Sun, J. (2016). A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting. Transportation Research Part C, 62, 21–34.

Castillo, V. E., Bell, J. E., Rose, W. J., & Rodrigues, A. M. (2018). Crowdsourcing Last Mile Delivery: Strategic Implications and Future Research Directions. Journal of Business Logistics, 39(1), 7–25.

Çelik, İ., Aydemir, E., Cebeci, H. İ., & Güner, S. (2025). Last-mile delivery performance of crowdsourced couriers: text mining of user comments. International Journal of Logistics Research and Applications, 1–22.

Chen P., Chankov S.M., & 2017 IEEE International Conference on Industrial Engineering and Engineering Management, IEEM 2017 2017 12 10 - 2017 12 13. (2018). Crowdsourced delivery for last-mile distribution: An agent-based modelling and simulation approach. IEEE International Conference on Industrial Engineering and Engineering Management, 2017-December, 1271–1275.

Chen, H., Yi, J., Chen, A., Peng, D., & Yang, J. (2023). Green technology innovation and CO2 emission in China: Evidence from a spatial-temporal analysis and a nonlinear spatial durbin model. Energy Policy, 172.

Cheng, S., Lu, F., Peng, P., & Wu, S. (2018). Short-term traffic forecasting: An adaptive ST-KNN model that considers spatial heterogeneity. Computers, Environment and Urban Systems, 71, 186–198.

Chu, H., Zhang, W., Bai, P. et al. (2023) Data-driven optimization for last-mile delivery. Complex Intell. Syst. 9, 2271–2284 (2023).

Cottreau, B., Manout, O., & Bouzouina, L. (2025). Spatio-temporal impacts of unplanned service disruptions on public transit demand. *Transportation Research Interdisciplinary Perspectives*, *30*.

Czajkowski, M., & Kretowski, M. (2016). The role of decision tree representation in regression problems - An evolutionary perspective. Applied Soft Computing Journal, 48, 458–475.

D. Al-Dogom, N. Aburaed, M. Al-Saad and S. Almansoori, "Spatio-temporal Analysis and Machine Learning for Traffic Accidents Prediction," 2019 2nd International Conference on Signal Processing and Information Security (ICSPIS), Dubai, United Arab Emirates, 2019, pp. 1-4, doi: 10.1109/ICSPIS48135.2019.9045892

dev.family (2023) On-Demand Food Delivery Platforms - Market, Trends & Opportunities, Medium.

Dieter, P., Caron, M. and Schryen, G. (2023) Integrating driver behavior into last-mile delivery routing: Combining Machine Learning and optimization in a hybrid decision support framework, European Journal of Operational Research.

E. Castillo, V. et al. (2022) Designing technology for on-demand delivery: The effect of customer tipping on crowdsourced driver behavior and last mile performance, Journal of Operations Management.

Elena Bruni, M., Fadda, E., Fedorov , S., & Perboli, G. (2023, April 28). A machine learning optimization approach for last-mile delivery and third-party logistics. Computers & Operations Research.

Elsokkary, N. et al. (2023) Crowdsourced last mile delivery: Collaborative workforce assignment, Internet of Things.

Ermagun, A., & Stathopoulos, A. (2021). Crowd-shipping delivery performance from bidding to delivering. Research in Transportation Business & Management, 41.

Ermagun, A., Shamshiripour, A. & Stathopoulos, A. Performance analysis of crowd-shipping in urban and suburban areas. Transportation 47, 1955–1985 (2020).

Ermagun, A., Shamshiripour, A. and Stathopoulos, A. (2019) Performance analysis of crowd-shipping in urban and suburban areas - transportation, SpringerLink.

Feizizadeh, B., Omrazadeh, D., Ghasemi, M., Bageri, S., Lakes, T., Kitzmann, R., … Blaschke, T. (2023). Urban restaurants and online food delivery during the COVID-19 pandemic: a spatial and socio-demographic analysis. International Journal of Digital Earth, 16(1), 1725–1751.

Fried, T. and Goodchild, A. (2023) E-commerce and logistics sprawl: A spatial exploration of last-mile logistics platforms, ScienceDirect.

Ge, X., Shi, L., Fu, Y., Muyeen, S. M., Zhang, Z., & He, H. (2020). Data-driven spatial-temporal prediction of electric vehicle load profile considering charging behavior. Electric Power Systems Research, 187.

Gül, V. (2023) *Machine learning part-4 (random forest-GBM-xgboost-lightgbm-catboost)*, *Medium*.

Haider, Z., Hu , Y., Charkhgard, H., Himmelgreen, D., & Kwon, C. (2022, April 7). Creating grocery delivery hubs for food deserts at local convenience stores via spatial and temporal consolidation. ScienceDirect.

Halder, R. K., Uddin, M. N., Uddin, M. A., Aryal, S., & Khraisat, A. (2024). Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications. Journal of Big Data, 11(1).

Huang, F., Yi, P., Wang, J., Li, M., Peng, J., & Xiong, X. (2022). A dynamical spatial-temporal graph neural network for traffic demand prediction. Information Sciences, 594, 286–304.

Huang, H., Wei, X., & Zhou, Y. (2022). An overview on twin support vector regression. Neurocomputing, 490, 80–92.

Ileri, K. (2025). Comparative analysis of CatBoost, LightGBM, XGBoost, RF, and DT methods optimised with PSO to estimate the number of k-barriers for intrusion detection in wireless sensor networks. International Journal of Machine Learning and Cybernetics, 1–20.

Jahanshahi, H. et al. (2022) A deep reinforcement learning approach for the meal delivery problem, Knowledge-Based Systems.

Kang, J., Guo, X., Fang, L., Wang, X., & Fan, Z. (2021). Integration of Internet search data to predict tourism trends using spatial-temporal XGBoost composite model. International Journal of Geographical Information Science, 36(2), 236–252.

Ke, J., Zheng, H., Yang, H., & Chen, X. (. (2017). Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. *Transportation Research Part C*, *85*, 591–608.

Larsson, A., Berg, J., Gellerfors, M., & Gerdin Wärnberg, M. (2021). The advanced machine learner XGBoost did not reduce prehospital trauma mistriage compared with logistic regression: a simulation study. BMC Medical Informatics and Decision Making, 21(1), 192.

Laynes-Fiascunari, V., Rabelo, L., Gutierrez-Franco, E. (2024). Enhancing Last-Mile Delivery: Social Media Insights and Deep Learning Applications. In: Garrido, A., Paternina-Arboleda, C.D., Voß, S. (eds) Computational Logistics. ICCL 2024. Lecture Notes in Computer Science, vol 15168. Springer, Cham.

Lin, P. C., Shen, C. W., Wang, J., & Yang, C. M. (2022). Spatial analysis of accidents involving food delivery motorcycles in Taiwan. Transportation Planning and Technology, 45(4), 335–357.

Liu, C.-C., Lin, T.-C., Yuan, K.-Y., & Chiueh, P.-T. (2022). Spatio-temporal prediction and factor identification of urban air quality using support vector machine. Urban Climate, 41.

Liu, H.-W. (2024). Mining spatial-temporal patterns from customer data to improve forecasting of customer flow across multiple sites. Journal of Retailing and Consumer Services, 79.

Liu, S., He , L. and Jun Max Shen, Z. (2020) On-Time Last-Mile Delivery: Order Assignment with Travel-Time Predictors, Management Science. Available at:

Liu, S., Zhang, F., Ji, Y., Ma, X., Liu, Y., Li, S., & Zhou, X. (2023). Understanding spatial-temporal travel demand of private and shared e-bikes as a feeder mode of metro stations. Journal of Cleaner Production, 398.

Lorenzo-Espejo, A., Muñuzuri, J., Onieva, L. et al. Exploring the correlation between courier workload, service density and distance with the success of last-mile and first-mile reverse logistics. Cent Eur J Oper Res (2024).

Mao, W., Ming, L., Rong, Y., Tang, C. S., & Zheng, H. (2025). Faster Deliveries and Smarter Order Assignments for an On-Demand Meal Delivery Platform. *Journal of Operations Management*, *71*(2), 220–245.

Masini, R. P., Medeiros, M. C., & Mendes, E. F. (2023). Machine learning advances for time series forecasting. Journal of Economic Surveys, 37(1), 76–111.

Nti, I. K., Nyarko-Boateng, O., & Aning, J. (2021). Performance of machine learning algorithms with different K values in K-fold cross-validation. International Journal of Information Technology and Computer Science, 13(6), 61-71.

Pegado-Bardayo, A., Lorenzo-Espejo , A., Muñuzuri, J., & Aparicio-Ruiz, P. (2023, August 27). A data-driven decision support system for service completion prediction in last Mile Logistics. Transportation Research Part A: Policy and Practice.

Pekel, E. (2019). Estimation of soil moisture using decision tree regression. Theoretical and Applied Climatology, 139(3-4), 1111–1119.

Peng , X. et al. (2023) A three-phase heuristic for last-mile delivery with spatial-temporal consolidation and delivery options, ScienceDirect.

Pourrahmani, E. and Jaller, M. (2021) Crowdshipping in last mile deliveries: Operational challenges and research opportunities, Socio-Economic Planning Sciences.

Rajendran, S. (2021). Improving the performance of global courier & delivery services industry by analyzing the voice of customers and employees using text analytics. International Journal of Logistics, 24(5), 473–493.

Ramírez-Villamil, A. et al. (2023) Reconfiguration of last-mile supply chain for parcel delivery using machine learning and routing optimization, Science Direct.

Ravula, P. (2023) Impact of delivery performance on online review ratings: the role of temporal distance of ratings. J Market Anal 11, 149–159 (2023).

Rodriguez, J. D., Perez, A., & Lozano, J. A. (2010). Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(3).

Sanaullah, I., Alsaleh, N., Djavadian, S., & Farooq, B. (2021). Spatio-temporal analysis of on-demand transit: A case study of Belleville, Canada. Transportation Research Part A, 145, 284–301.

Saud, S., Jamil, B., Upadhyay, Y., & Irshad, K. (2020). Performance improvement of empirical models for estimation of global solar radiation in India: A k-fold cross-validation approach. Sustainable Energy Technologies and Assessments, 40, 100768.

Savelsbergh, M.W., Ulmer, M.W. (2022) Challenges and opportunities in crowdsourced delivery planning and operations. 4OR-Q J Oper Res 20, 1–21 (2022).

Sekeroglu, B., Ever, Y. K., Dimililer, K., & Al-Turjman, F. (2022). Comparative Evaluation and Comprehensive Analysis of Machine Learning Models for Regression Problems. Data Intelligence, 4(3), 620–652.

Shehadeh, A., Alshboul, O., Al Mamlook, R. E., & Hamedat, O. (2021). Machine learning models for predicting the residual value of heavy construction equipment: An evaluation of modified decision tree, LightGBM, and XGBoost regression. Automation in Construction, 129.

Shen, S. et al. (2022) An optimization approach for worker selection in Crowdsourcing Systems, ScienceDirect.

Şimşek, B. et al. (2013) Driver performance appraisal using GPS Terminal Measurements: A conceptual framework, Science Direct.

Soltani, A., Heydari, M., Aghaei, F., & Pettit, C. J. (2022). Housing price prediction incorporating spatio-temporal dependency into machine learning algorithms. Cities, 131.

Sureeyatanapas, P. and Damapong, K. (2024) Performance assessment and comparison of online food delivery service providers based upon the aggregated perspectives of restaurant operators, Science Direct.

Timotius, E. et al. (2023) Buyers-sellers' value of Courier Services: Assessment in the Indonesian C2C e-commerce, International Journal of Retail & Distribution Management.

Tongtian Zhu. (2020). Analysis on the Applicability of the Random Forest. 1607(1).

Tukey, J. W. (1977). Exploratory data analysis (Vol. 2, pp. 131-160). Reading, MA: Addison-wesley.

Usman Miko, N., & Abbas, U. (2023, January 20). Determinants of efficient last-mile delivery: evidence from health facilities and Kaduna Health Supplies Management Agency | Emerald Insight. Emerald Insight.

Üstün, B., Melssen, W. J., Buydens, L. M. C., & CAC 2006 Campinas SP Brazil 20060910-20060914. (2007). Visualisation and interpretation of Support Vector Regression models. Analytica Chimica Acta, 595(1-2), 299–309.

Wang, J., & Song, G. (2018). A Deep Spatial-Temporal Ensemble Model for Air Quality Prediction. Neurocomputing, 314, 198–206.

Wang, S., Li , L., Ma, W., & Chen, X. (2019, September 23). Trajectory analysis for on-demand services: A survey focusing on spatial-temporal demand and supply patterns. ScienceDirect.

Wang, S., Li, L., Ma, W., & Chen, X. (2019). Trajectory analysis for on-demand services: A survey focusing on spatial-temporal demand and supply patterns. *Transportation Research Part C*, *108*, 74–99.

Wu, Z., Peng, L. and Xiang, C. (2022) Assuring quality and waiting time in real-time spatial crowdsourcing, ScienceDirect.

Xia, Y., & Leung, H. (2006). Nonlinear spatial-temporal prediction based on optimal fusion. IEEE Transactions on Neural Networks, 17(4), 975–988.

Xianglong Luo, Danyang Li, Yu Yang, & Shengrui Zhang. (2019). Spatiotemporal Traffic Flow Prediction with KNN and LSTM. Journal of Advanced Transportation, 2019.

Xu, K., Han, Z., Xu, H., & Bin, L. (2023). Rapid Prediction Model for Urban Floods Based on a Light Gradient Boosting Machine Approach and Hydrological–Hydraulic Model. International Journal of Disaster Risk Science, 14(1), 79–97.

Yaiprasert, C. and Nizar Hidayanto, A. (2023) AI-driven ensemble three machine learning to enhance digital marketing strategies in the food delivery business, Intelligent Systems with Applications.

Yang, Y., & Zhang, H. (2019). Spatial-temporal forecasting of tourism demand. Annals of Tourism Research, 75, 106–119.

Yeşilkanat, C. M. (2020). Spatio-temporal estimation of the daily cases of COVID-19 in worldwide using random forest machine learning algorithm. Chaos, Solitons and Fractals: The Interdisciplinary Journal of Nonlinear Science, and Nonequilibrium and Complex Phenomena, 140.

Yu, C., & Yao, W. (2017). Robust linear regression: A review and comparison. Communications in Statistics - Simulation and Computation, 46(8), 6261–6282.

Zhang, X., & Liu, C.-A. (2023). Model averaging prediction by K-fold cross-validation. Journal of Econometrics, 235(1), 280–301.

Appendices

Python codes for spatial temporal prediction and EDA analysis can be found on the following link mentioned below:

Machine Learning Spatial Temporal Order Volume Prediction

https://www.kaggle.com/code/azharmansoori/spatial-temporal-order-volume-prediction-ml

EDA analysis of Courier Performance and Behavior in On-Demand Food Delivery Platform

https://www.kaggle.com/code/azharmansoori/eda-of-courier-performance-and-behavior