The Price of Speed: Optimising Dark-Store Networks and Rider Assignment under Relaxed Delivery Promises in Indian Quick Commerce

Madhav Gupta[1], Aadee Pardasani[2], Aashi Khuthia[3], Jiya Jagawat[4], Prishi Doshi[5], Aarushi Rakesh[6]
[1,2,3,4,5,6]*NMIMS University, Mumbai, India*

***Abstract*—In January 2026 India's quick-commerce platforms withdrew their "10-minute delivery" claims after the Union Labour Ministry raised rider-safety concerns, and in August 2026 a Parliamentary Standing Committee recommended curbs on "unrealistic speed targets" without defining them. We combine a bibliometric review of 168 operations research articles on last-mile and on-demand delivery (2015–2026) with an integrated optimisation model of a Bengaluru dark-store network built entirely from public data. A coverage-constrained facility-location model opens stores and allocates demand at minimum store and rider cost while guaranteeing a promise T at a safe rider speed v; a transportation model plans replenishment, an assignment model dispatches riders at peak, and a Program Evaluation and Review Technique (PERT) model estimates on-time probability. At 15 km/h the cost-optimal network has about 240 stores and reaches 98.3% of residents for any promise of 15 minutes or more, whereas a 10-minute promise leaves 13.3% unreachable from available retail sites; serving everyone within 10 minutes requires 61% more stores and 8.4% higher cost per order. Rider speed, not the promise, drives cost: a 15 rather than 20 km/h cap raises cost per order by about 19%. We propose reachability within a safe ride radius as an operational definition of a "realistic" promise: promising 10 minutes only where reachable gives 86.7% of residents 10-minute service at 4.2% higher cost than a uniform 15-minute network.**

***Index Terms*—assignment problem, bibliometric analysis, dark stores, facility location, gig-worker safety, India, PERT, quick commerce, transportation problem.**

I. INTRODUCTION

Quick commerce (q-commerce) promises groceries and household essentials within minutes, fulfilled from small urban warehouses called dark stores. India has become its largest laboratory: in the April–June 2026 quarter Blinkit alone reported 2,443 dark stores and 331 million orders [1], and Swiggy Instamart a further 1,171 stores [2]. The model rests on a simple geometry: a store that serves only a 1.5–3 km radius can meet a 10-minute promise without riders travelling fast, which is how platform founders defended it [3], [4]. That defence has been tested. After a nationwide strike by gig workers on 31 December 2025, the Union Labour Ministry intervened and in January 2026 Blinkit, Zepto, Swiggy Instamart and Zomato dropped "10-minute delivery" from their branding [5].

In August 2026 the Parliamentary Standing Committee on Commerce recommended guidelines against "unrealistic speed targets" set by dispatch algorithms, but did not say what makes a target unrealistic [6]. The policy debate therefore lacks a quantitative answer to a classical OR question: what does each minute of delivery promise cost, and who pays for it the platform, through more stores, or the rider, through higher speed?

This paper answers that question for Bengaluru, India's most contested q-commerce market, and situates it in the OR literature. We pose three research questions:

RQ1: How has OR research on on-demand and last-mile delivery evolved, which methods dominate (location, assignment, routing), and where are the gaps for quick commerce and courier welfare?

RQ2: For Bengaluru, how do the number of dark stores, rider requirements and cost change as the delivery promise relaxes from 10 to 15, 20 and 30 minutes under a safety-capped rider speed?

RQ3: Which promise level and network design balance coverage, cost and rider safety, and what does this imply for defining a "realistic" speed target?

Our contributions are fourfold. First, a bibliometric map of OR research on last-mile and on-demand delivery from 2015 to 2026, which shows a fast-growing field (31.4% a year) centred on vehicle routing with new delivery modes, in which only 3 of 168 articles address quick commerce, none studies India and only 4 consider courier welfare. Second, an integrated and fully reproducible model chain coverage-constrained facility location, transportation, assignment and PERT built from public Indian data, together with an audit showing why widely used Kaggle delivery datasets are unsuitable for spatial modelling. Third, a "price of speed" curve for Bengaluru. Fourth, evidence for regulators on where a delivery promise stops being achievable through network design and starts depending on rider speed.
Section II reviews the literature; Section III describes the bibliometric protocol and modelling framework; Section IV presents the case and data; Section V formulates the models; Section VI reports results; Section VII discusses implications and limitations; Section VIII concludes.

II. LITERATURE REVIEW

A. Bibliometric overview

The search (Section III) yielded 168 journal articles on OR for last-mile and on-demand delivery published between 2015 and 2026. The field is young and accelerating: output grew at 31.4% a year between 2015 and 2025, 91% of the corpus appeared from 2021 onwards, and 2025 alone contributed 46 articles (Fig. 1). Core OR journals dominate: the European Journal of Operational Research published 20 articles and Computers & Operations Research 14, followed by Transportation Research Procedia and the Annals of Operations Research.

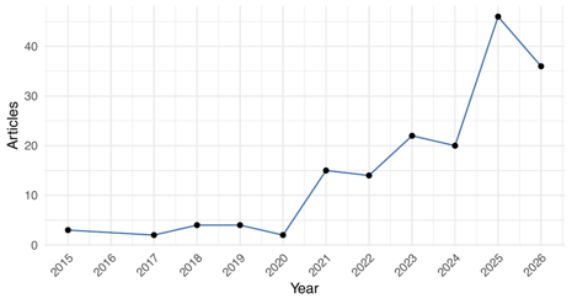


Fig. 1. Annual number of included articles (2026 is a partial year).

The most frequent keywords (Fig. 2) are last-mile delivery and vehicle routing, with mixed-integer programming (49 keyword occurrences) and adaptive large neighbourhood search (17) as the dominant solution technologies. Counting topics across titles, abstracts and keywords, drones and delivery robots appear in 54 of the 168 articles, assignment, matching or dispatch in 50, location models in 23 and on-demand, meal, grocery or instant delivery in 27. The thematic map (Fig. 3) places last-mile vehicle routing, delivery-cost reduction with order assignment, and facility location (with electric vehicles, crowdshipping and parcel lockers) among the basic themes: central to the field but internally diffuse. The motor themes, both central and well developed, are on-demand delivery routing with customer assignment, robot and drone delivery under deadlines solved by decomposition, branch-and-price for multimodal and drone routing, and emission-aware routing solved by genetic algorithms. Driver workload and social sustainability appear only in a niche theme with e-commerce well developed internally but peripheral to the field alongside food-delivery apps with reinforcement learning and depot location.

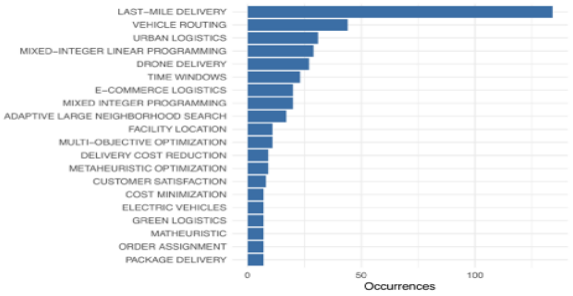


Fig. 2. Most frequent keywords in the 168 included articles (the keyword co-occurrence network is available in the project repository).

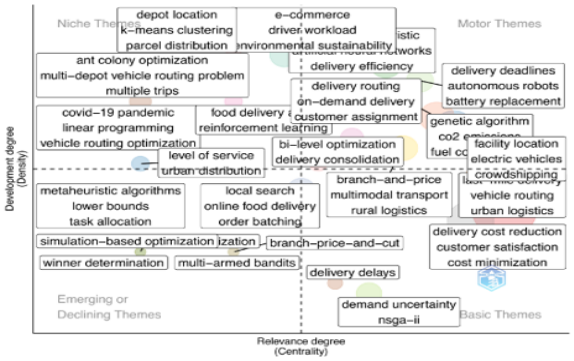


Fig. 3. Thematic map (centrality vs. density of keyword clusters).

The intellectual base, measured by co-citation, is anchored in routing. The most co-cited works are the drone-assisted travelling salesman problem of Murray and Chu [7] and its successors [8], [9], Solomon's time-window benchmarks [10], the adaptive large neighborhood search of Røpke and Pisinger [11], the survey of last-mile concepts by Boysen et al. [12], and two-echelon covering with cargo bikes and lockers [13].

B. Thematic synthesis

Table I codes 16 core studies selected for citations per year and for proximity to our problem.

Table I. Core studies (coded from abstracts).

Study	Problem	Method	Data
Deutsch and Golany [14]	parcel locker network location	0-1 ILP / uncapacitated facility location	real
Zhou et al. [15]	two-echelon VRP with pickup-facility delivery options	hybrid multi-population genetic algorithm	benchmark + real instance
Schermer et al. [16]	VRP with drones and en route launch/retrieval	MILP with valid inequalities; VNS/tabu search	synthetic/ benchmark
Liu [17]	online drone fleet dispatch for on-demand meal delivery	MIP with rolling-horizon progressive algorithm	simulation
Chen et al. [18]	VRPTW with vans carrying delivery robots	ALNS heuristic	synthetic/ benchmark
Alfandari et al. [19]	truck routing with autonomous delivery robots (tardiness)	MIP, branch-and-Benders-cut	synthetic/ benchmark
Li et al. [20]	review	literature review of two-echelon truck-UAV routing formulations	none/ review
Zhang et al. [21]	dynamic VRP with stochastic requests	knapsack-based linear approximations, multi-stage optimization	real
Simoni and Winkenbach [22]	crowdsourced food delivery order batching and assignment	graph-based heuristic, local search, agent-based simulation	real
Mao et al. [23]	meal delivery order dispatch	survival analysis, instrumental variables, simulation-based evaluation	real
Shuaibu et al. [24]	review of last-mile optimisation approaches	literature review	none/ review
Zamal et al. [25]	stochastic dynamic order assignment and two-echelon dispatching	MDP, cost function approximation + ALNS	real
Voigt et al. [26]	dynamic stochastic some-day delivery (multi-period PCVRPTW)	hybrid ALNS	synthetic/ benchmark
Bhattacharya and Kumar [27]	quick-commerce order allocation balancing fairness and efficiency	minimum-cost assignment, two-stage heuristic	real
Luo et al. [28]	micro-fulfillment center location-allocation for instant retailing	integer programming, branch-and-bound with Lagrangian relaxation	real
Carlsson et al. [29]	delivery zoning	VRP-based Voronoi partitioning, subgradient algorithm	real

Vehicle routing with new delivery modes.

The largest stream extends the vehicle routing problem to two-echelon systems, delivery options and new vehicles: multi-depot two-echelon routing with delivery options [15], trucks with drones [16], [17], and trucks launching delivery robots solved by adaptive large neighbourhood search [18] or Bender's decomposition [19]. Reviews classify these truck–drone formulations [20] and summarise the broader toolbox of exact, metaheuristic and learning-based approaches [24]. The stream is methodologically rich but mostly tested on synthetic benchmarks.

On-demand and dynamic delivery.

A second stream treats orders that arrive in real time: batching and assignment for crowdsourced food delivery [22], large-scale dynamic routing with stochastic requests [21], end-to-end order assignment and dispatch formulated as a Markov decision process [25], and dispatch policies evaluated on platform data, where maximising future orders outperforms minimising delay [23]. Delivery time is usually a performance measure here. Voigt et al. [26] are a notable exception: offering a "some-day" option yields large cost savings for small increases in delivery

time the closest analogue in the corpus to pricing a delivery promise.

Location of fulfilment infrastructure.
Location models appear in 23 studies, from parcel-locker networks reduced to uncapacitated facility location [14] to micro-fulfilment location–allocation for instant retail [28], building on covering models [30], [31]. Outside the screened corpus, Kewalramani and Khadilkar [32] site Indian dark stores heuristically under delivery-radius constraints.

Fairness and the workforce.
Only 4 studies consider couriers' or drivers' welfare. Bhattacharya and Kumar [27] show for quick-commerce order allocation that a fairness weight raises delivery time by 4–5% but cuts earnings inequality by a third, and Carlsson et al. [29] design equitable delivery zones that cut the maximum work span by about a fifth in a field implementation. Neither treats rider speed or safety as a constraint.

C. Research gaps
The bibliometric evidence supports three gaps. First, quick commerce is almost absent: only 3 of 168 articles mention quick commerce, dark stores or micro-fulfilment, and none (0) studies India, the largest q-commerce market. Second, the delivery promise is treated as given: only 4 articles treat delivery time or speed as something to be chosen or priced. Third, rider welfare is marginal (4 articles), and none models a safe-speed limit. Our study addresses all three by pricing the promise for an Indian city under an explicit rider-speed constraint.

III. RESEARCH METHODOLOGY

A. Bibliometric data collection
The course protocol prescribes an EBSCOhost search with the mandatory term's "operations" AND "research". Because EBSCO requires an institutional login, we ran the identical protocol on OpenAlex, an open index of scholarly works with abstracts, keywords, citation counts and reference lists, on 2026-10-06. The Boolean string was
"operations" AND "research" AND ("last-mile" OR "last mile" OR "on-demand delivery" OR "instant delivery" OR "quick commerce" OR "q-commerce" OR "meal delivery" OR "dark store" OR "micro-fulfil*" OR "crowdshipping") AND ("facility location" OR "covering" OR "assignment" OR "vehicle routing" OR "integer programming" OR "linear programming")
restricted to journal articles in English published 2015–2026. OpenAlex throttles long Boolean queries, so each topic term was queried separately on titles and abstracts and the union (7,576 records) was filtered locally: the method clause on title and abstract (715 records) and the mandatory terms on title, abstract, journal name and keywords approximating EBSCO's default search fields leaving 216 records with no duplicates (Fig. 4).
Titles and abstracts were screened against explicit criteria: a record was included if its primary contribution concerned the last-mile delivery of goods, food or parcels to end customers and it developed, applied or reviewed an OR method. Excluded records were coded as out of scope (37; e.g. "last mile" in disaster relief, health access or passenger transit), without an OR method (10; e.g. adoption surveys) or editorial (1). Screening was run in two passes: a first pass screened all records and flagged 24 as uncertain, and an independent second pass re-screened the flagged records, whose decision was adopted (it excluded nine of them, mostly generic routing studies without a last-mile focus). Every decision, its code and the second-pass rationale are recorded in the screening log. 168 articles were retained. Records and cited references were converted to the Bibliometrix format [33] for performance analysis and science mapping; citation counts are OpenAlex counts. Included records were exported to RIS for the Mendeley library.

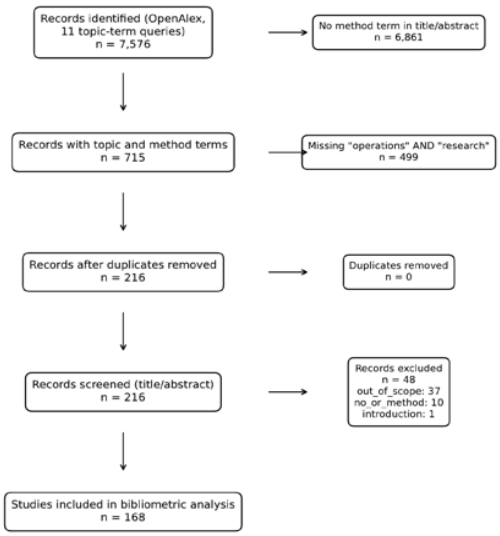


Fig. 4. Identification and screening of records.

B. Modelling framework

The case study links four classical OR models at three planning levels (Fig. 5). Strategically, a coverage-constrained facility-location model (M1) opens dark stores and allocates demand so that every served household is reachable within the promise T at a safe rider speed v, at minimum store and rider cost; rider requirements follow from cycle times (M2). Tactically, a transportation model (M3) plans daily replenishment from mother hubs to the chosen stores. Operationally, an assignment model (M4) dispatches riders to orders in a peak batch, and a PERT model (M5) converts the variability of picking, dispatch and riding into the probability that an order meets T. The promise T and speed v are swept as policy parameters; everything else is calibrated from public sources (Section IV).

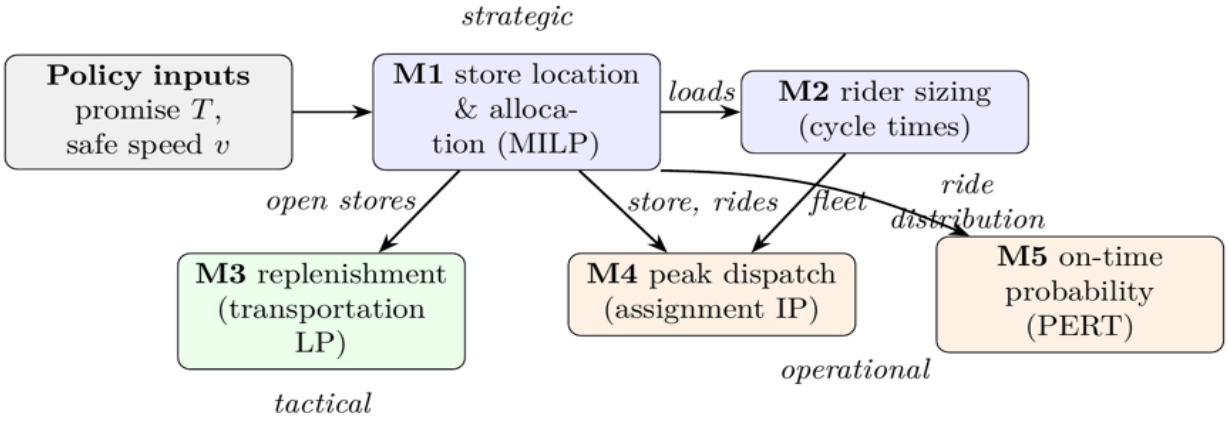


Fig. 5. Model chain linking strategic, tactical and operational decisions.

All models are implemented in Python with PuLP and solved with HiGHS; code, data provenance and a test suite verifying each model against hand-solvable or brute-force instances are available at https://github.com/MadhavGupta27/price-of-speed.

C. Use of generative AI

A generative AI assistant (Claude, Anthropic) was used for language polishing and for assistance with code and document formatting. The authors carried out the study design, literature screening, modelling, analysis and interpretation, reviewed all AI-assisted output, and take full responsibility for the content of the article.

IV. CASE AND DATA

A. Study area

The study area is the Bruhat Bengaluru Mahanagara Palike (BBMP) as delimited into 243 wards in 2022 [34]; the ward polygons cover 720 km^2, consistent with the official 712.5 km^2. Bengaluru hosts roughly 440 dark stores of the three largest platforms [35], making it the most contested q-commerce market in India.

B. Data sources and audit

Table II lists the inputs. Demand is spatialised with the WorldPop 2025 constrained population grid at 100 m [36], aggregated to 3,426 cells of about 0.46 km. WorldPop sums to 6.04 million inside BBMP below the 2011 Census count of 8.44 million [37] because it caps density in the dense core; we therefore keep its spatial pattern and scale it to the census total. Candidate store sites are the 1,480 supermarkets, grocery and convenience shops and warehouses mapped in OpenStreetMap within BBMP [38], merged into 1,133 candidate locations. Road distances are straight-line distances multiplied by a detour factor of 1.57 (interquartile range 1.41–1.81), the median road/straight-line ratio over a random sample of OSRM routes.

Table II. Data sources.

Input	Source	Use
Population	WorldPop R2025A, India 2025, 100 m constrained	Spatial demand weights
Boundary	BBMP ward map 2022 (OpenCity)	Study area
Candidate sites	OpenStreetMap, Geofabrik extract 2026-10-05	Possible dark stores

Input	Source	Use
Road detour	OSRM, OSM road network	Distance correction
Orders, stores, costs	Eternal and Swiggy shareholder letters Q1 FY27; press	Calibration (Table III)

Two popular Kaggle datasets on Indian food and grocery delivery were evaluated and rejected. In the food-delivery dataset, 98.4% of delivery points equal the restaurant coordinates plus an identical latitude–longitude offset drawn from 13 values, and road-traffic level is fully determined by the hour of day; the delivery geometry and timing are synthetic. The Blinkit sales dataset uses placeholder e-mail domains for all customers. Neither can support distance-based modelling, a caution for the many studies that use them.

C. Parameters

Table III summarises the calibration. Bengaluru demand is estimated as stores per platform times orders per store per day: Blinkit [1], Swiggy Instamart [2] and Zepto (mean of the two), giving about 600,000 orders per day. Rider cost uses the average of INR 102 per logged-in hour reported by Eternal for 2025 [39]. Speeds and pick-pack times follow statements by platform founders [3], [4]. As a validation, the implied rider cost per order in the base case falls within the reported per-order pay of INR 25–35.

Table III. Model parameters (base value and sensitivity range).

Parameter	Base (range)	Basis
Delivery promise $\boldsymbol{T}$	10, 15, 20, 30 min	Policy scenarios
Safe rider speed $\boldsymbol{v}$	15 (12–20) km/h	Global Indian [3]
Pick + pack (PERT $\boldsymbol{a, m, b}$)	1.3, 2, 4 min	India.com [4]; Zepto 76 s
Rider assignment; handover	(0, 0.5, 2); (0.5, 1, 2) min	Assumption
Daily orders, Bengaluru	600k (450k–750k)	Eternal Ltd [1] and Swiggy Ltd [2] and Quick Commerce Map [35]
Peak-hour share of daily orders	0.10 (0.08–0.12)	Assumption (evening peak)
Store fixed cost	INR 8 (6–11) lakh/month	Opex + capex/60 months
Rider cost	INR 125 (102–154)/h	Business Today [39]
Rider utilisation	0.6 (0.5–0.7)	Assumption

V. MODEL FORMULATION

Notation.

$\boldsymbol{I}$ is the set of demand cells with daily orders $\boldsymbol{q_i}$ and peak-hour orders $\boldsymbol{\lambda_i = s\, q_i}$ (peak share $\boldsymbol{s}$); $\boldsymbol{J}$ is the set of candidate sites; $\boldsymbol{d_{ij}}$ is road distance. Fulfilment comprises pick-and-pack, rider assignment and handover with PERT means $\boldsymbol{t^{pp}, t^{as}, t^{ho}}$, so the ride radius available under promise $\boldsymbol{T}$ (min) and safe speed $\boldsymbol{v}$ (km/h) is

$$\boldsymbol{r(T, v) = \frac{v}{60}\left(T - t^{pp} - t^{as} - t^{ho}\right).} \qquad \boldsymbol{(1)}$$

The reachable arcs are $\boldsymbol{A = \{(i, j): d_{ij} \le r(T, v)\}}$.

M1: coverage-constrained network design.

With binary $\boldsymbol{y_j}$ (open site $\boldsymbol{j}$), continuous $\boldsymbol{x_{ij}}$ (share of cell $\boldsymbol{i}$ served by $\boldsymbol{j}$) and $\boldsymbol{u_i}$ (unreachable demand),

$$\boldsymbol{\min \quad F \sum_{j \in J} y_j + \sum_{(i,j) \in A} c_{ij}\, x_{ij} + M \sum_{i \in I} u_i} \qquad \boldsymbol{(2)}$$

$$\boldsymbol{s.t. \sum_{j:(i,j) \in A} x_{ij} + u_i = 1 \quad \forall i \in I} \qquad \boldsymbol{(3)}$$

$$\boldsymbol{x_{ij} \le y_j \quad \forall (i, j) \in A} \qquad \boldsymbol{(4)}$$

$$\boldsymbol{x_{ij} \ge 0,\ u_i \ge 0,\ y_j \in \{0, 1\},}$$

where $\boldsymbol{F}$ is the monthly fixed cost of a store and

$$\boldsymbol{c_{ij} = 30\, q_i \left(\frac{2 d_{ij}}{v} + \frac{t^{ho}}{60}\right)\frac{w}{\rho}} \qquad \boldsymbol{(5)}$$

is the monthly rider cost of serving cell $\boldsymbol{i}$ from $\boldsymbol{j}$: a round trip at the safe speed plus handover, paid at $\boldsymbol{w}$ per logged-in hour with utilisation $\boldsymbol{\rho}$. The promise enters as a hard restriction on $\boldsymbol{A}$. Cells with no

reachable candidate are served by $\boldsymbol{u_i}$ at a large penalty $\boldsymbol{M}$ and reported as unreachable rather than making the model infeasible. Without capacity limits $\boldsymbol{x_{ij}}$ is integral at the optimum, so M1 is an uncapacitated facility-location problem on the reachable arcs; for $\boldsymbol{T = 10}$ it reduces to a set-covering structure [30], while for loose promises the cost trade-off dominates.

M2: rider sizing.
Each cell is served by its assigned store; peak riders at store $\boldsymbol{j}$ follow from cycle times (Little's law), $\boldsymbol{R_j} = [\sum_{i:\, j(i)=j} \quad \boldsymbol{\lambda_i}\,(\boldsymbol{2d_{ij}/v + t^{ho}/60})]$.

M3: replenishment (transportation problem).
Open stores are restocked daily from three mother hubs located at OpenStreetMap warehouse sites nearest the demand-weighted centroids of the store network. With hub supplies $\boldsymbol{a_h}$, store requirements $\boldsymbol{b_j}$ and distances $\boldsymbol{d_{hj}}$, $\boldsymbol{min} \sum_{h,j} \quad \boldsymbol{d_{hj} z_{hj}}$ subject to $\sum_j \quad \boldsymbol{z_{hj} = a_h}, \sum_h \quad \boldsymbol{z_{hj} = b_j}, \boldsymbol{z \geq 0}$. We compare the optimum with Vogel's approximation and the north-west corner rule.

M4: peak dispatch (assignment problem).
At a representative store we simulate a 10-minute peak batch of orders with arrival times $\boldsymbol{\alpha_o}$, ready times $\boldsymbol{\alpha_o + t^{pp} + t^{as}}$ and ride times $\boldsymbol{\tau_o}$, and rider slots $\boldsymbol{k}$ (a rider's successive trips) that become free at $\boldsymbol{\phi_k}$. The lateness of serving order $\boldsymbol{o}$ with slot $\boldsymbol{k}$ is $\boldsymbol{\ell_{ko} = max\{0, max(\phi_k, \alpha_o + t^{pp} + t^{as}) + \tau_o + t^{ho} - (\alpha_o + T)\}}$. The assignment model $\boldsymbol{min} \sum_{k,o} \quad \boldsymbol{\ell_{ko} x_{ko}}$ with $\sum_k \quad \boldsymbol{x_{ko} = 1}, \sum_o \quad \boldsymbol{x_{ko} \leq 1}, \quad \boldsymbol{x \in \{0, 1\}}$ is compared with first-in-first-out dispatch.

M5: fulfilment reliability (PERT).
An order passes four sequential activities: pick-and-pack, rider assignment, ride and handover. The first, second and fourth use the three-point estimates in Table III; the ride uses the 5th, 50th and 95th percentiles of demand-weighted ride times in the designed network. With PERT means $\boldsymbol{\mu_k} = (\boldsymbol{a_k} + \boldsymbol{4m_k + b_k})/\boldsymbol{6}$ and variances $\sigma_k^2 = ((b_k - a_k)/6)^2$, the probability of fulfilment within T is $\Phi\left((T - \sum_k \quad \mu_k)/\sqrt{\sum_k \quad \sigma_k^2}\right)$ [40]. We also report the rider speed needed to reach 95% on-time delivery for a given network.

VI. RESULTS

A. Network design across promises and speeds

Table IV and Fig. 6 report the cost-minimising networks (M1–M2) for every combination of promise T and safe speed v. Of the 29 mixed-integer programs solved for this paper, 27 were proven optimal within HiGHS's default gap tolerance (the longest took 154 s); the remaining 2 maximal covering with 50 and 100 stores at $T = 15$ (Section VI-B) stopped at the 600-second limit, so their reported coverage is a lower bound. Three results stand out.

Table IV. Cost-minimizing dark-store networks for Bengaluru (M1–M2). Cost per order is per served order and includes store and rider costs.

T (min)	v (km/h)	Radius (km)	Stores	Reachable pop. (%)	Riders (peak)	Orders/ store/day	Cost/order (INR)
10	12	1.21	390	79.3	6,705	1,220	50.4
10	15	1.51	320	86.7	6,902	1,626	43.4
10	20	2.01	226	94.2	6,872	2,501	35.6
15	12	2.21	270	95.9	10,295	2,130	49.3
15	15	2.76	240	98.3	9,275	2,458	43.2
15	20	3.68	203	99.8	7,929	2,950	36.3
20	12	3.21	261	99.4	11,341	2,286	50.8
20	15	4.01	239	99.9	9,700	2,509	43.9
20	20	5.34	204	100.0	7,968	2,941	36.4
30	12	5.21	260	100.0	11,553	2,308	51.2
30	15	6.51	238	100.0	9,733	2,521	43.9
30	20	8.68	202	100.0	7,987	2,970	36.4

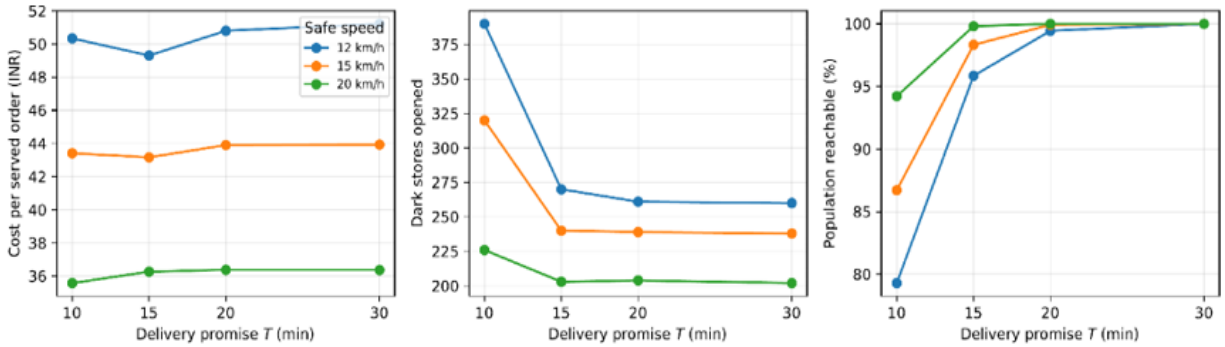


Fig. 6. The price of speed: cost per served order, stores opened and share of residents reachable, by promise T and safe rider speed v.

The promise decides who is served, not what it costs. At a safe speed of 15 km/h, a 15-minute promise reaches 98.3% of residents with 240 stores at INR 43.2 per order. Tightening to 10 minutes raises the store count to 320 but leaves 13.3% of residents mostly at the city's edge, where retail stock suitable for dark stores is sparse with no candidate site within reach (Fig. 7); the cost per served order barely moves (INR 43.4). Relaxing the promise beyond 15 minutes changes almost nothing: the cost-optimal network has about 239 stores whether the promise is 20 or 30 minutes, because the cost trade-off between stores and rider time, not the promise, then determines the radius.

The speed cap decides what it costs.: Because riders are paid per logged-in hour, cost per order falls steeply with rider speed: at T = 15 it is INR 49.3 at 12 km/h, 43.2 at 15 km/h and 36.3 at 20 km/h. Capping speed at 15 rather than 20 km/h costs about 19% more per order. This is the platform's economic incentive to push speed independent of the advertised promise and the quantity a regulator implicitly prices when it sets a speed norm.

The cost-optimal network is smaller than today's.: A single network serving all of Bengaluru's demand at minimum cost needs 202–390 stores across the scenarios, against roughly 440 operated by three competing platforms today. The gap is a measure of the duplication that competition creates.



Fig. 7. Optimal networks for T = 10 and T = 20 minutes at 15 km/h. grey cells are served; orange cells have no candidate site within reach.

The price of speed at equal coverage.: The coverage gap at T = 10 arises because mapped retail sites are scarce at the city's edge. Allowing a store anywhere on a ~0.9 km lattice of populated cells (860 candidate sites), a 10-minute promise reaches 100.0% of residents but needs 462 stores at INR 45.9

per order, against 287 stores at INR 42.4 for 15 minutes. Serving everyone within 10 rather than 15 minutes therefore needs 61% more stores and costs 8.4% more per order about INR 0.71 per order for each minute saved. The price of the last five minutes is paid mainly in real estate at the periphery, which the platforms' current networks largely avoid.

B. Fragmentation and a "realistic promise" policy

Fig. 8 solves the maximal covering problem for a given number of stores placed optimally. Within 15 minutes, 100 well-placed stores already reach at least 98.1% of residents. Within 10 minutes, 165 stores Blinkit's Bengaluru network reach at most 74.6%, and no number of stores at mapped retail sites lifts coverage above 86.7%. A 10-minute promise is therefore a promise to the dense core only, whichever platform makes it, and three platforms replicating 150-store networks buy little extra coverage.

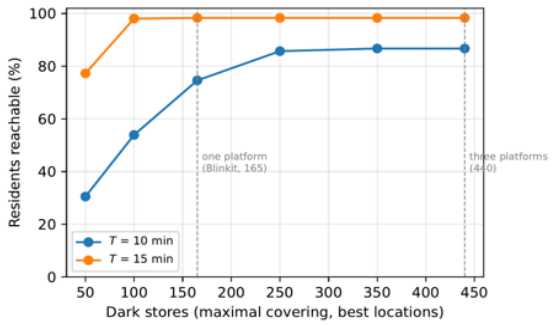


Fig. 8. Share of residents reachable with P optimally located stores (maximal covering, v = 15 km/h).

We also test the operational definition proposed in Section VII: promise 10 minutes wherever an address lies within the safe 10-minute radius of some candidate site, and 15 minutes elsewhere. This differentiated promise gives 86.7% of residents a 10-minute promise and still serves 98.3% overall the same coverage as a uniform 15-minute promise with 318 stores (78 more) at INR 45.0 per order, 4.2% above the uniform 15-minute network. A uniform 10-minute promise, by contrast, reaches only 86.7% of residents. Tying the promise to reachability thus delivers 10-minute service to most of the city at a small premium, without leaving anyone unserved and without requiring riders to exceed the safe speed.

C. Fulfilment reliability (PERT)

Table V gives the PERT estimate of fulfilment time in each network at 15 km/h. Inside the designed networks every served address is within the ride radius, so on-time probabilities are high: 99.7% at T = 10. A 10-minute promise met 95% of the time requires riders to average 12.1 km/h in the 10-minute network below the safe speed. Ten-minute delivery is therefore feasible without speeding, but only with a dense network and only for the 86.7% of residents that network can reach.

Table V. PERT fulfilment time by promise (network designed for that promise, v = 15 km/h); v_{95} is the rider speed needed for 95% on-time delivery.

T (min)	Mean (min)	SD (min)	P(on time)	v_{95} (km/h)
10	7.33	0.99	0.997	12.1
15	8.11	1.31	1.000	8.3
20	8.23	1.40	1.000	6.0
30	8.23	1.40	1.000	5.0

D. Peak dispatch (assignment)

The pressure appears when a network designed for one promise is asked to meet a tighter one. Table VI dispatches a 10-minute peak batch at a stressed store of the 15-minute network (90th percentile of mean ride distance, 1.3 km). With the full-sized fleet, a 15-minute promise is met for every order, but a 10-minute promise only for 64%; the late orders could be saved only if riders averaged about 20 km/h, above the safe speed. Under lean staffing the gap widens sharply. Optimal assignment reduces total lateness relative to first-in-first-out dispatch, although because it minimizes lateness rather than the count of late orders it occasionally delivers fewer orders strictly on time. The pressure is local: at the median-load store, where rides average 0.9 km, 60% of the sized fleet meets even a 10-minute promise for every order under optimal dispatch (first-in-first-out reaches 90%), and only at 40% staffing do orders run late. Tight promises thus fail first at the edges of store catchments, which is where riders would be pushed to speed.

Table VI. Peak 10-minute batch at a stressed store of the 15-minute network: optimal assignment vs FIFO.

Staffing	Riders	T (min)	On time optimal	On time FIFO	Mean lateness (min) optimal / FIFO
40%	32	10	17%	26%	4.24 / 4.47

40%	32	15	56%	59%	0.89 / 1.46
40%	32	20	100%	90%	0.00 / 0.25
60%	47	10	35%	49%	1.19 / 1.52
60%	47	15	100%	93%	0.00 / 0.10
60%	47	20	100%	100%	0.00 / 0.00
80%	63	10	64%	64%	0.82 / 0.83
80%	63	15	100%	100%	0.00 / 0.00
80%	63	20	100%	100%	0.00 / 0.00
100%	79	10	64%	64%	0.82 / 0.82
100%	79	15	100%	100%	0.00 / 0.00
100%	79	20	100%	100%	0.00 / 0.00

E. Replenishment (transportation)

For the base network ($T = 15$, $v = 15$; 240 stores), replenishment from three equal-capacity mother hubs costs 5.80 million order-km per day at the optimum. Vogel's approximation comes within 0.9% of the optimum, while the north-west corner rule is 2.3 times worse. Supplying each store from its nearest hub would cost 5.60 million order-km, so equal hub capacities carry a 3.5% premium. Because the store count hardly changes beyond 15 minutes, replenishment cost is insensitive to the promise in that range.

F. Sensitivity analysis

Fig. 9 varies each uncertain input over its range at the base case. The cost per order is most sensitive to rider cost (a swing of INR 13.4) and rider utilisation (INR 11.0): rider time dominates the cost structure, which is why speed is the platform's main lever. Higher rider cost or demand raises the optimal number of stores, shortening rides; higher store cost lowers it. Longer pick-and-pack times shrink the ride radius; cost per served order changes little because the hardest-to-serve cells drop out, which again shows that tighter time budgets are paid in coverage. The headline findings a store count near the base case, cost flat in T beyond 15 minutes, and speed as the cost driver hold across all ranges.

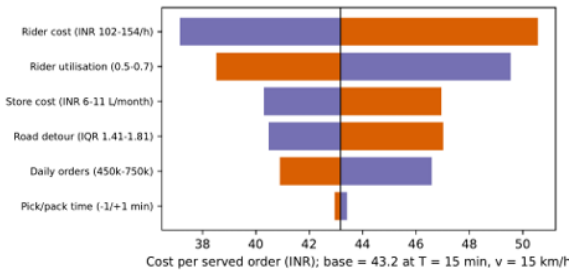


Fig. 9. One-at-a-time sensitivity of cost per served order ($T = 15$, $v = 15$).

VII. DISCUSSION

A. What makes a speed target "unrealistic"?

The results separate two quantities that the public debate conflates. The promise T is a network-design decision: it determines how many stores are needed and which neighbourhoods can be served. The speed riders must sustain is a cost decision: at a given network, faster riding is cheaper for the platform because rider time is paid per hour. A promise becomes unrealistic when the network cannot deliver it at a safe speed. Our results suggest a practical test for regulators: a promise is realistic for an address if the address lies within the safe ride radius of an open store after realistic pick-and-pack time exactly the reachability set A in M1. By that test, 15 minutes is realistic for 98.3% of Bengaluru residents with the cost-optimal network, while 10 minutes is realistic for 86.7% even with 320 stores. Promising 10 minutes beyond that footprint, or asking a 15-minute network to deliver in 10, can be met only by faster riding (Section VI).

B. Managerial implications

For platforms, the cost per order is nearly flat between 10 and 30 minutes once the network is designed for the promise; the expensive lever is rider speed. Platforms can therefore offer differentiated promises 10 minutes in dense cores, where mapped retail space allows it, and 15 elsewhere at modest cost (a uniform 10-minute promise for everyone costs about 8.4% more per order), and should resist meeting tight promises at peaks through speed. Dispatch quality matters most at stressed stores and lean staffing, where optimal assignment reduces lateness relative to FIFO.

C. Policy implications

A speed norm is more enforceable than a promise ban. A cap of 15 km/h raises cost per order by about 19%

relative to 20 km/h in our base case a quantifiable price of rider safety that can be weighed against accident risk. Platform disclosures of delivery radius and store density would allow regulators to check reachability directly.

D. Limitations

Demand is spatialised from residential population and calibrated from platform-level disclosures rather than order data; store counts come from a third-party tracker; candidate sites depend on OpenStreetMap completeness; road distances use a single detour factor and free-flow routing; rider cost is a per-hour average; and the model represents one network rather than three competing platforms. PERT assumes independent activities and a normal approximation. On the bibliometric side, the search used OpenAlex rather than the prescribed EBSCO database, citation counts are OpenAlex rather than Google Scholar counts, and country metadata were too sparse for a reliable country analysis.

E. Future research agenda

Each gap in Section II suggests an agenda. For quick commerce: competitive location models in which several platforms site stores in the same city, with capacity and inventory at dark stores. For promises as decisions: models that set differentiated promises jointly with network design, extending the some-day options of Voigt et al. [26] to minutes-scale delivery. For rider welfare: explicit accident-risk functions of speed in the objective, time-dependent travel times, and dynamic dispatch with fairness constraints building on Bhattacharya and Kumar [27] and Carlsson et al. [29]. For India: validation against platform order data, which would also remove our reliance on population as a demand proxy.

VIII. CONCLUSION

RQ1. OR research on last-mile delivery has grown by 31.4% a year since 2015, led by vehicle routing with drones, robots and two-echelon systems solved by mixed-integer programming and large neighbourhood search; quick commerce (3 of 168 articles), India (none) and rider welfare (4) remain almost unstudied, and the delivery promise is rarely a decision variable. RQ2. In Bengaluru the cost-optimal network has about 240 dark stores for any promise of 15 minutes or more; a 10-minute promise cannot reach 13.3% of residents from mapped retail sites, and serving everyone within 10 minutes needs 61% more stores at 8.4% higher cost per order. Rider speed, not the promise, drives cost: a 15 km/h cap costs about 19% more per order than 20 km/h. RQ3. A promise is realistic where an address lies within the safe ride radius of a store; on that basis 15 minutes is realistic for nearly all of Bengaluru and 10 minutes for 86.7% of residents, and a promise differentiated this way costs only 4.2% more per order than a uniform 15-minute network. Regulating speed norms and disclosure of delivery radii is more meaningful than regulating advertised minutes.

ACKNOWLEDGMENT

We thank Dr Vishal Pradhan, instructor of the Operations Research Management course at NMIMS University, Mumbai, for guidance throughout this project, including on the bibliometric research process. We also acknowledge the open data and software on which this study relies: map data © OpenStreetMap contributors (ODbL 1.0), obtained via Geofabrik; WorldPop, University of Southampton, for gridded population estimates; OpenCity.in and Devdatta Tengshe for the BBMP ward boundaries; OpenAlex for bibliographic metadata; and the developers of OSRM, PuLP, HiGHS and Bibliometrix.

DECLARATIONS

Funding: No funding was received for this work. Conflict of interest: The authors declare no conflict of interest. Data and code availability: All inputs are public (WorldPop 2025 population, CC BY 4.0; BBMP ward boundaries, OpenCity; OpenStreetMap, ODbL 1.0; OpenAlex metadata, CC0; cited company disclosures). Processed data, the screening log, provenance records and all code (MIT licence) are available at https://github.com/MadhavGupta27/price-of-speed. Use of generative AI: see Section III-C.

REFERENCES

[1]. Eternal Ltd., “Shareholders’ Letter, Q1 FY27,” filed with BSE and NSE, July 22, 2026; figures

as reported by MediaNama, "7 Key Takeaways from Eternal Q1 FY27 Earnings Call," 2026.
[2]. Swiggy Ltd., "Letter to Shareholders, Q1 FY27," NSE filing, July 30, 2026. 2026.
[3]. Global Indian, "India's Q-Commerce Boom: Stanford Dropouts' Startup Zepto Reaches Valuation of $900 Million," 2022.
[4]. India.com, "Grofers Founder Albinder Dhindsa Responds After Facing Criticism for 10-Minute Deliveries: Full Statement," August 2021.
[5]. Outlook Business, "Quick-Commerce Giants Drop '10-Minute Delivery' Promise After Govt Intervention," January 2026.
[6]. Outlook Business, "Quick Commerce Speed Algorithms May Face New Rules, Parl Panel Flags Rider Safety," August 2026.
[7]. C. C. Murray and A. Chu, "The Flying Sidekick Traveling Salesman Problem: Optimization of Drone-Assisted Parcel Delivery," Transportation Research Part C: Emerging Technologies, 2015, doi: 10.1016/j.trc.2015.03.005.
[8]. N. A. H. Agatz, P. Bouman, and M. Schmidt, "Optimization Approaches for the Traveling Salesman Problem with Drone," Transportation Science, 2018, doi: 10.1287/trsc.2017.0791.
[9]. D. Sacramento, D. Pisinger, and S. Røpke, "An Adaptive Large Neighborhood Search Metaheuristic for the Vehicle Routing Problem with Drones," Transportation Research Part C: Emerging Technologies, 2019, doi: 10.1016/j.trc.2019.02.018.
[10]. M. M. Solomon, "Algorithms for the Vehicle Routing and Scheduling Problems with Time Window Constraints," Operations Research, vol. 35, no. 2, pp. 254–265, 1987, doi: 10.1287/opre.35.2.254.
[11]. S. Røpke and D. Pisinger, "An Adaptive Large Neighborhood Search Heuristic for the Pickup and Delivery Problem with Time Windows," Transportation Science, 2006, doi: 10.1287/trsc.1050.0135.
[12]. N. Boysen, S. Fedtke, and S. Schwerdfeger, "Last-Mile Delivery Concepts: A Survey from an Operational Research Perspective," OR Spectrum, 2020, doi: 10.1007/s00291-020-00607-8.
[13]. D. L. J. U. Enthoven, B. Jargalsaikhan, K. J. Roodbergen, M. A. J. uit het Broek, and A. H. Schrotenboer, "The Two-Echelon Vehicle Routing Problem with Covering Options: City Logistics with Cargo Bikes and Parcel Lockers," Computers & Operations Research, 2020, doi: 10.1016/j.cor.2020.104919.
[14]. Y. Deutsch and B. Golany, "A Parcel Locker Network as a Solution to the Logistics Last Mile Problem," International Journal of Production Research, vol. 56, nos. 1–2, pp. 251–261, 2017, doi: 10.1080/00207543.2017.1395490.
[15]. L. Zhou, R. Baldacci, D. Vigo, and X. Wang, "A Multi-Depot Two-Echelon Vehicle Routing Problem with Delivery Options Arising in the Last Mile Distribution," European Journal of Operational Research, vol. 265, no. 2, pp. 765–778, 2017, doi: 10.1016/j.ejor.2017.08.011.
[16]. D. Schermer, M. Moeini, and O. Wendt, "A Hybrid VNS/Tabu Search Algorithm for Solving the Vehicle Routing Problem with Drones and En Route Operations," Computers & Operations Research, vol. 109, pp. 134–158, 2019, doi: 10.1016/j.cor.2019.04.021.
[17]. Y. Liu, "An Optimization-Driven Dynamic Vehicle Routing Algorithm for On-Demand Meal Delivery Using Drones," Computers & Operations Research, vol. 111, pp. 1–20, 2019, doi: 10.1016/j.cor.2019.05.024.
[18]. C. Chen, E. Demir, and Y. Huang, "An Adaptive Large Neighborhood Search Heuristic for the Vehicle Routing Problem with Time Windows and Delivery Robots," European Journal of Operational Research, vol. 294, no. 3, pp. 1164–1180, 2021, doi: 10.1016/j.ejor.2021.02.027.
[19]. L. Alfandari, I. Ljubić, and M. de Melo da Silva, "A Tailored Benders Decomposition Approach for Last-Mile Delivery with Autonomous Robots," European Journal of Operational Research, vol. 299, no. 2, pp. 510–525, 2021, doi: 10.1016/j.ejor.2021.06.048.
[20]. H. Li, J. Chen, F. Wang, and M. Bai, "Ground-Vehicle and Unmanned-Aerial-Vehicle Routing Problems from Two-Echelon Scheme Perspective: A Review," European Journal of Operational Research, vol. 294, no. 3, pp. 1078–1095, 2021, doi: 10.1016/j.ejor.2021.02.022.
[21]. J. Zhang, K. Luo, A. M. Florio, and T. V. Woensel, "Solving Large-Scale Dynamic Vehicle Routing Problems with Stochastic Requests," European Journal of Operational

Research, vol. 306, no. 2, pp. 596–614, 2022, doi: 10.1016/j.ejor.2022.07.015.

[22]. M. D. Simoni and M. Winkenbach, "Crowdsourced On-Demand Food Delivery: An Order Batching and Assignment Algorithm," Transportation Research Part C: Emerging Technologies, vol. 149, Art. no. 104055, 2023, doi: 10.1016/j.trc.2023.104055.

[23]. W. Mao, L. Ming, Y. Rong, C. S. Tang, and H. Zheng, "Faster Deliveries and Smarter Order Assignments for an On-Demand Meal Delivery Platform," Journal of Operations Management, vol. 71, no. 2, pp. 220–245, 2025, doi: 10.1002/joom.1354.

[24]. A. S. Shuaibu, A. S. H. Mahmoud, and T. Sheltami, "Last-Mile Delivery Optimization: Recent Approaches and Advances," Transportation Research Procedia, vol. 84, pp. 299–306, 2025, doi: 10.1016/j.trpro.2025.03.076.

[25]. M. A. Zamal, A. H. Schrotenboer, and T. V. Woensel, "End-to-End Logistics in Metropolitan Areas: A Stochastic Dynamic Order-Assignment and Dispatching Problem," Transportation Research Part B: Methodological, vol. 199, Art. no. 103249, 2025, doi: 10.1016/j.trb.2025.103249.

[26]. S. Voigt, M. Frank, and H. Kühn, "Last Mile Delivery Routing Problem with Some-Day Option," European Journal of Operational Research, vol. 324, no. 2, pp. 477–491, 2025, doi: 10.1016/j.ejor.2025.02.001.

[27]. M. Bhattacharya and M. Kumar, "Balancing Fairness and Efficiency in Order Allocation for Quick-Commerce Applications," Asia Pacific Journal of Operational Research, 2026, doi: 10.1142/S0217595926500363.

[28]. W. Luo, S. Deng, and X. Cheng, "Location-Allocation Problem Based on Order Distribution in Instant Retailing: An Efficient Branch-and-Bound Integrated Lagrangian Relaxation," Asia Pacific Journal of Operational Research, 2026, doi: 10.1142/S0217595926500120.

[29]. J. G. Carlsson, S. F. W. T. Lim, S. Liu, and H. Yu, "Equitable Delivery Zoning for Last-Mile Logistics: A Framework Validated with Implementation," Manufacturing & Service Operations Management, 2026, doi: 10.1287/msom.2024.1221.

[30]. C. Toregas, R. Swain, C. ReVelle, and L. Bergman, "The Location of Emergency Service Facilities," Operations Research, vol. 19, no. 6, pp. 1363–1373, 1971.

[31]. R. Church and C. ReVelle, "The Maximal Covering Location Problem," Papers of the Regional Science Association, vol. 32, no. 1, pp. 101–118, 1974.

[32]. P. Kewalramani and H. Khadilkar, "Heuristic for Optimisation of Dark Store Facility Locations for Quick Commerce Businesses," arXiv preprint arXiv:2312.11494, 2023.

[33]. M. Aria and C. Cuccurullo, "bibliometrix: An R-Tool for Comprehensive Science Mapping Analysis," Journal of Informetrics, vol. 11, no. 4, pp. 959–975, 2017.

[34]. OpenCity, "BBMP Ward Map – 2022 (243 Wards)," data.opencity.in, credited to Devdatta Tengshe, 2022.

[35]. WorldPop, "Global2 R2025A: Constrained Population Counts, India 2025, 100 m," University of Southampton, 2025.

[36]. Office of the Registrar General & Census Commissioner, India, "Census of India 2011: Provisional Population Totals, Cities – Bruhat Bengaluru Mahanagara Palike (8,443,675)," 2011.

[37]. OpenStreetMap contributors, "OpenStreetMap Data, India Southern Zone Extract, October 5, 2026," Geofabrik GmbH, Open Database License (ODbL) 1.0, 2026.

[38]. Business Today, "'Look at the Full Picture': Deepinder Goyal Breaks Down Delivery Partner Earnings with Hard Facts," January 2, 2026.

[39]. D. G. Malcolm, J. H. Roseboom, C. E. Clark, and W. Fazar, "Application of a Technique for Research and Development Program Evaluation," Operations Research, vol. 7, no. 5, pp. 646–669, 1959.