

Forecasting How Busy a Venue Will Be, and Counting How Many People Are There

Jayden Bansal · Flock Social LLC, Pennsylvania · Research summary, October 2026
Full paper (30 pages, every data source credited): flockcorp.com/research · KIT and St. Gallen library data by permission

Research question

Can AI tell how crowded a venue will be before you go? And can a credit-card-sized computer model be designed to count the number of people inside a venue using a thermal camera without identifying individuals?

Answers. 1. One hour ahead, yes (54.9% exact against the chart's 30.2%), further ahead only slightly; the gain comes from fresh live readings, not the larger gradient-boosted model (30.8%). 2. Yes in simple rooms (94.3% exact against the heat rule's 40.7%), not yet in crowds or inside a working venue.

What I built

- **The forecast in Flock**, my app for planning nights out. It takes each venue's popular-times chart (how busy it usually is at each hour of the week) and shifts it by how far recent live readings ran above or below it, or uses a reading from the past hour directly (page 3, Equations 1 and 2). Readings were collected for 1,754 venues in Philadelphia and the Lehigh Valley (hourly) and Miami (twice daily); the data provider has sent almost none since 12 October. A gradient-boosted model with 106 inputs (listed on page 3) is tested against the live app and ships only if it wins.
- **Owl**, a 718,550-parameter neural network that counts people on a FLIR Lepton 3.5 thermal camera, its main model taking 14 to 16 ms a frame on one Raspberry Pi 5 core, trained on a physics-based generator of thermal scenes and six licensed real datasets. It runs inside **Flux**, a venue sensor I designed, which sends only numbers (Figure 1).

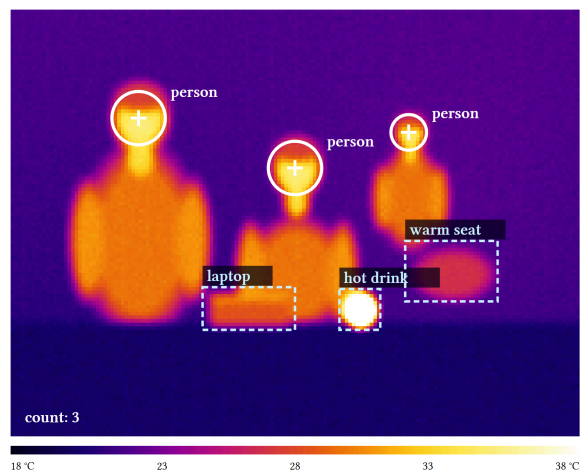


Figure 1. What Owl sees and what it reports. A 160 × 120 thermal frame in degrees Celsius, drawn the way Owl's training generator draws a scene, at the camera's true resolution. Owl puts a point on each person's head and a named box on every other warm thing, here a laptop, a hot drink and a seat someone just left, which a simple heat threshold would count as people. Only the points are counted: this frame counts as three.

How it was tested

The original forecast weights were fitted on earlier days and scored on **119,939 live readings** from 9 September to 8 October 2026 that they never saw, at 1,754 venues, with uncertainty intervals from 2,000 redraws of whole dates and repeat-the-last-reading baselines. The forecast target is the provider's live reading, a 0 to 100 score relative to the venue's own peak, so related patterns were then tested on public counts: **6.5 million hourly records from 31 sources in 15 countries**. The core 5.3 million counts (315 dataset-site series) were analysed under a plan whose pass rules were fixed before any data were downloaded; the rest add occupant counts in Singapore rooms, library seat estimates in Karlsruhe and Wi-Fi connections at campus access points in Barcelona.

Key findings

Result	Flock	Baseline
Question 1: the forecast		
Exact crowd level (of five, Quiet to Packed), 30-day test, one hour ahead	54.9%	30.2% popular times
Within one crowd level, same readings	78.4%	64.8% popular times; 75.2% a constant
Mean absolute error on the provider's 0 to 100 scale, same readings	18.00	26.40 (31.8% lower for Flock)
Exact level when read an hour earlier (48% of readings)	77.0%	

Result	Flock	Baseline
Exact level two to 24 hours ahead (current rule)	33.2–33.6%	30.2%
Public data: one-hour error, curve + latest departure (15 datasets)	6.41	9.58 curve alone
Wi-Fi users at 153 campus access points: hour-old value wins at	152 of 153	typical-week curve
People in 5 rooms (ROBOD): one-hour MAE (0 to 100), hour-old count	12.18	21.63 curve
Question 2: Owl		
Owl, frames counted exactly, five of six benchmark sets	82–96%	9–77% heat rule
Owl, sealed set of 300 frames no version had seen (independent AI counts)	94.3%	40.7% heat rule
Owl, 12 more public sets pooled (2,976 unseen frames)	40.4%	33.0% heat rule

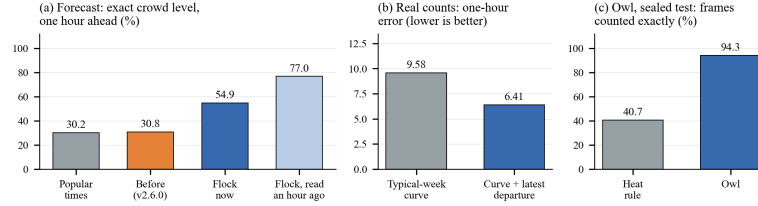


Figure 2. Headline results. (a) Exact crowd level of five, one hour ahead, on 119,939 held-out live readings; “Before” is the model Flock served until 26 September. (b) One-hour mean absolute error on the 15 pre-registered public real-count datasets (0 to 100 per site). (c) Owl against the heat-threshold rule on 300 sealed frames.

- **The freshest reading of a venue is worth more than any correction to its typical-week curve.** Where the venue was read in the hour before, those readings carry 92% of the exact-level gain over popular times and 96% of the error reduction, and production now serves that reading directly; otherwise the trailing offset is the correction.
- **The finding holds on occupancy data.** On occupant counts, library seat estimates and campus Wi-Fi connections, the hour-old value beats the typical-week curve, as it does on the provider’s readings. Across the 15 registered datasets the typical-week curve is close to calibrated (mean slope 0.92) and a departure from it lasts for hours, so carrying the latest departure onto the curve (Equation 4) cut pooled one-hour error by 33%.
- **The popular-times curve is a poorly calibrated forecast.** Against live readings, each point of curve is worth half a point of busyness (slope 0.50, correlation 0.36).
- **Owl generalises across its training cameras and, in generated stress tests, ignores the warm objects that fool a threshold** (hands, mugs, screens); in crowded classrooms, against face labels, it misses by one person on average, its predecessor by almost five. On 338 unseen lab-room frames of three or more people (BU-TIV) it counted 34.6% exactly and 79.3% within one person, where the heat rule managed 3.6% and 18.9%. It beat the rule 22.7% to 3.0% exact beside hot robot arms in a factory (WLRI-HRC) and 16.7% to 3.8% on a harbour front (LTD). The rule did better on 32×24 and 32×32 arrays, on pooled frames of three or more people, in an atrium of lone, distant people and with walkers filmed at head height (LITIV); large or distant crowds in unseen scenes (a hall, a café, a harbour) defeat both.

How Owl counts

Owl is a CenterNet-style network: it predicts maps of confidence scores, not temperatures, one for people ($\hat{P}$) and nine for warm objects (hand, pet, hot drink, food, laptop, screen, heater, lamp, warm seat). The count is the number of peaks in the person map above a confidence threshold τ , a peak being a pixel p no lower than any pixel in its neighbourhood $\mathcal{N}(p)$,

$$\hat{N} = \left| \left\{ p: \hat{P}(p) \geq \tau, \hat{P}(p) = \max_{q \in \mathcal{N}(p)} \hat{P}(q) \right\} \right| \quad (5)$$

each peak checked against the physics of skin temperature before it counts. If five or more are found, a crowd model of the same size reads the frame too, and the count is the largest of the two peak counts and its rounded sum of people per grid cell. Training penalises missed and false peaks (a penalty-reduced focal loss), misplaced boxes and wrong cell totals, $L = L_{\text{focal}} + \lambda_{\text{box}} L_{\text{box}} + \lambda_{\text{den}} L_{\text{den}}$ (weights λ_{box} , λ_{den}), on 40% generated frames and 60% from a licensed pool of 30,648 real and 8,000 rendered frames (70,402 labelled people).

Scope

- The forecast is scored against the provider’s 0 to 100 estimates, not headcounts; its large gains are one hour ahead.
- Owl’s 300-frame sealed set holds one or two people per frame (independent AI counts, not yet checked by people); its benchmark sets also guided its development, and it has not yet been tested in a working venue.

How the forecast is made

For venue v at hour h on weekday d : $b(v,d,h)$ is the venue's smoothed weekly curve; o_v is the median of (observed – curve) over its newest 20 live readings taken before h (within 28 days, at least two), clamped to ± 50 ; r_v is its latest earlier reading, a hours old.

$$\hat{y}_0 = b(v, d, h) + 0.85 \cdot o_v \quad (1)$$

$$\hat{y} = (1 - w_a) \cdot \hat{y}_0 + w_a \cdot r_v, \quad w_1 = 1, \quad w_2 = 0.85, \quad w_3 = 0.25, \quad w_a = 0 \quad \text{for } a \geq 4 \quad (2)$$

$$\hat{y} = \text{round}\left(\text{clip}\left(b + 0.5 o_v + \text{clamp}(\hat{\delta}, \pm 50), 0, 100\right)\right), \quad \hat{\delta} = f(x), \quad x \in \mathbb{R}^{106} \quad (3)$$

$$\hat{y}_{t+1} = \text{clip}\left(b_{t+1} + \text{clamp}(y_t - b_t, \pm 50), 0, 100\right) \quad (4)$$

The one-hour-ahead 54.9% result scores (1) and (2), each step rounded and clipped to 0 to 100, with weights fitted by mean absolute error on 1 to 5 September and fixed through the 30-day test. Since 13 October production serves a simpler rule ($w_2 = w_3 = 0$), chosen on that window; on the same readings it scores 54.7% one hour ahead and 33.2 to 33.6% two to 24 hours ahead, retrospective figures. (3) is the delta model, which adds half the offset and its predicted departure $\hat{\delta}$ to the curve: XGBoost, 800 trees at depth 8, learning rate 0.01, with the venue's own curve refused as an input. (4), the latest departure carried onto the next hour's curve, had the lowest one-hour error on 14 of 15 real-count datasets (exploratory).

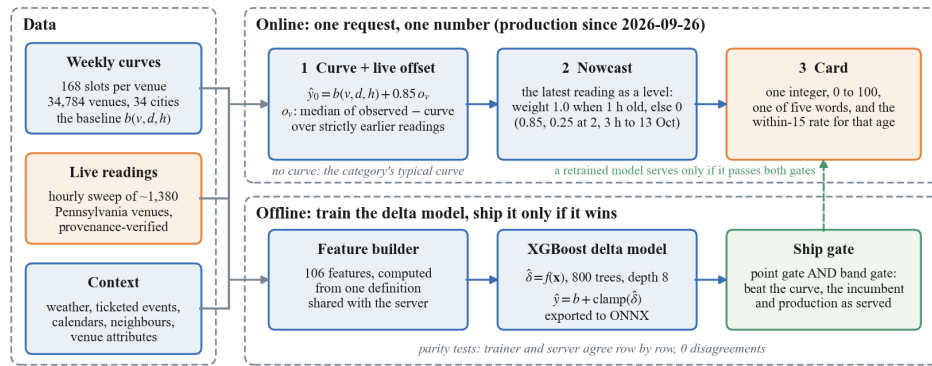


Figure 3. The system. The online path serves every request today. The offline delta model reads 106 inputs and can replace the online path only by passing the ship gate's two tests: point accuracy on held-out cities and crowd levels against production.

Every input the delta model reads (106)

Time (19)	hour, weekday, month and the sine and cosine of each; weekend; Friday or Saturday after 18:00; lunch, dinner, late night, morning; four season flags
Daylight (3)	daylight hours, hours after sunset, after-sunset flag (solar estimate from latitude and month)
Holidays (6)	holiday, school break, holiday eve, special night, special-night boost and suppress weights (31 national calendars, 1,351 dated special nights)
Weather (18)	temperature, humidity, wind, rain flag; nine condition groups; temperature anomaly against a climate table; four weather interactions
Events (16)	any ticketed event within 2 km; nearest event's attendance, its log and distance; its type as five flags; count, total attendance and its log within 2 km; large event; event $\times$ weekend, dinner, bar
Category curves (2)	typical value of the venue's category at this weekday and hour; the same refined by price and rating tier
Neighbours (2)	log count of venues within about 830 m; their mean curve value at this weekday and hour
Venue (5)	category (13 classes), price level, rating, review count and its log
Place types (30)	one flag for each of the 30 most common place types
Feedback (4)	mean crowd report, log report count, has-report flag, mean past prediction error (verified in-app reports, 28 days)
Row type (1)	realtime or typical-week row
Added for the retrain	live readings (5): the latest earlier reading's departure from the curve and its age, the trailing offset and its count, the curve one hour earlier; sports games (6): a tracked team plays today, number of games, evening game, home game, distance to the arena, arena within 10 km

What comes next

A retrain of both models in early November 2026, the forecast scored on 14 held-out days against thresholds fixed in code on 25 and 26 September (the production rule it must beat was chosen on 13 October with data up to 8 October, so the test is untouched by that choice only on later dates); Owl scored once, after the freeze, on Flux's own camera in rooms of one to fifteen people and on 215 untouched LITIV 2012 frames with independent AI counts not yet checked by people; and a venue study pairing hand counts with Owl and the provider's live scores.