

Supplementary Text

Defining place categories

To examine whether the same type of POIs shared similar visitation and contact patterns, we performed a clustering analysis of POIs based on three metrics: (1) crowdedness, (2) dwell time, and (3) variance of visitor numbers within a week. We first classified all POIs into subgroups using the first four digits of the NAICS code. For each subgroup, we calculated the mean values of all three metrics using data from January 2020 to February 2020 (prior to the pandemic). We normalized each metric by dividing the maximum value across all subgroups so that the three metrics are on the same scale. A Gaussian mixture clustering was performed. We identified three clusters (Fig. S1). Interestingly, subgroups of POIs within the same broader categories were generally clustered within the same cluster. Based on this preliminary analysis, we classified all POIs into five place categories using the first four digits of NAICS codes. The classification rule is provided in Table S1.

Retrospective forecasts

We used the EAKF to generate retrospective forecasts. The EAKF assumes a Gaussian distribution of both the prior and likelihood and adjusts the prior distribution to a posterior using Bayes' rule deterministically. To represent the state-space distribution, the EAKF maintains an ensemble of system state vectors acting as samples from the distribution. In particular, the EAKF assumes that both the prior distribution and likelihood are Gaussian, and thus can be fully characterized by their first two moments (mean and variance). The update scheme for ensemble members is computed using Bayes' rule (posterior \propto prior \times likelihood) via the convolution of the two Gaussian distributions. For observed state variables, the posterior of the i th ensemble member is updated through

$$y_{t,post}^i = \frac{\sigma_{t,obs}^2}{\sigma_{t,obs}^2 + \sigma_{t,prior}^2} \bar{y}_{t,prior} + \frac{\sigma_{t,prior}^2}{\sigma_{t,obs}^2 + \sigma_{t,prior}^2} y_t^o + \sqrt{\frac{\sigma_{t,obs}^2}{\sigma_{t,obs}^2 + \sigma_{t,prior}^2}} (y_{t,prior}^i - \bar{y}_{t,prior}). \quad [S1]$$

Here $y_{t,post}^i$ and $y_{t,prior}^i$ are the posterior and prior of the observed variable for the i th ensemble member at time t ; $\bar{y}_{t,prior}$ is the mean of the prior observed variable; $\sigma_{t,obs}^2$ and $\sigma_{t,prior}^2$ are the variances of the observation and the prior observed variable; and y_t^o is the observation at time t . Unobserved variables and parameters are updated through their covariability with the observed variable, which can be computed directly from the ensemble. In particular, the i th ensemble member of unobserved variable or parameter x^i is updated by

$$x_{t,post}^i = x_{t,prior}^i + \frac{\sigma(\{x_{t,prior}\}_n, \{y_{t,prior}\}_n)}{\sigma_{t,prior}^2} (y_{t,post}^i - y_{t,prior}^i). \quad [S2]$$

Here $x_{t,post}^i$ and $x_{t,prior}^i$ are the posterior and prior of the unobserved variable or parameter for the i th ensemble member at time t ; and $\sigma(\{x_{t,prior}\}_n, \{y_{t,prior}\}_n)$ is the covariance between the prior of the unobserved variable or parameter $\{x_{t,prior}\}_n$ and the prior of the observed variable

$\{y_{t,prior}\}_n$ at time t . In the EAKF, variables and parameters are updated deterministically such that the higher moments of the prior distribution are preserved in the posterior.

In the EAKF, we assumed a heuristic form of observation error variance (OEV): $\sigma_{lt}^2 = \sigma_{l0}^2 + (y_{lt}^o)^2/9$, where σ_{l0}^2 is the baseline OEV for neighborhood l and y_{lt}^o is the observed case in neighborhood l in week t . To account for heterogeneous disease burdens across neighborhoods, the baseline OEV σ_{l0}^2 was defined as 2500 (i.e., standard deviation 50). We chose a relatively large OEV to avoid ensemble degeneracy. Similar forms of OEV have been successfully used for inference and forecasting for a range of infectious diseases. During the weekly EAKF update, we looped through observations in all neighborhoods. The observation in neighborhood l was used to update model variables within the same neighborhood and global parameters (v_p for five place categories and β_h for household transmission). In implementation, we used 500 ensemble members.

A pseudo-code for the EAKF is provided below.

Algorithm 1. EAKF

Input: The transmission model \mathcal{M} , observations $\{y_{lt}^o\}$ in T weeks and L locations, the observational error variance (OEV) $\{\sigma_{lt}^2\}$, the initial ensemble of system state $\{z_i^0\}$ with 500 ensemble members.

for $t = 1$ to T **do**

Run model \mathcal{M} forward for one time step using system state $\{z_i^{t-1}\}$, obtain the ensemble of prior observed variables and system states at time t : $[\{y_{lt,prior}\}, \{z_{i,prior}^t\}] = \mathcal{M}(\{z_i^{t-1}\})$

for $l = 1$ to L **do**

Update observed variable in location l at time t using Eq. S1:

$$\{y_{lt,post}\} = EAKF(\{y_{lt,prior}\}, y_{lt}^o, \sigma_{lt}^2).$$

Update unobserved variables in location l at time t and parameters using Eq. S2:

$$\{z_{i,post}^t\} = EAKF(\{z_{i,prior}^t\}, \{y_{lt,prior}\}, \{y_{lt,post}\}).$$

end for

Update the ensemble of model state at time t using the posterior: $\{z_i^t\} \leftarrow \{z_{i,post}^t\}$, which will be used to initiate the model at the next time step.

Run the transmission model \mathcal{M} forward using current posterior state to generate forecasts.

end for

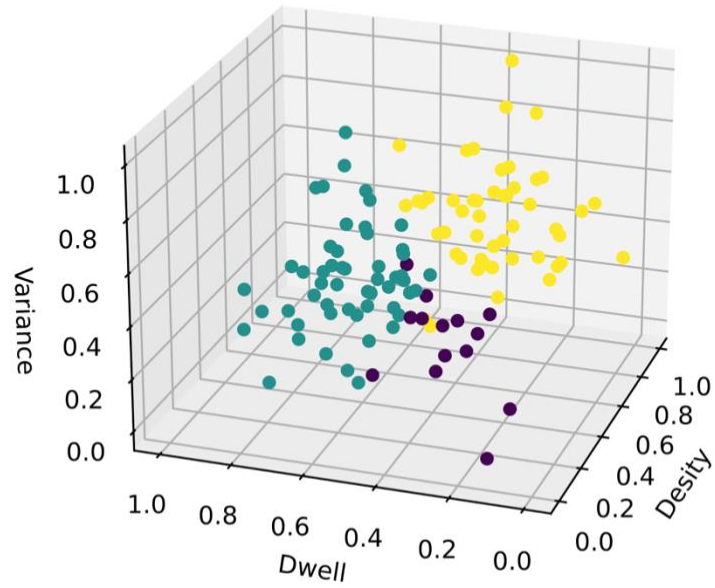


Fig. S1.

Clustering of POIs using crowdedness, dwell time, and variance of visitor numbers within a week. Each data point represents one subgroup of POIs with the same first four digits of the NAICS code. The three clusters identified by the Gaussian mixture model are highlighted in different colors.

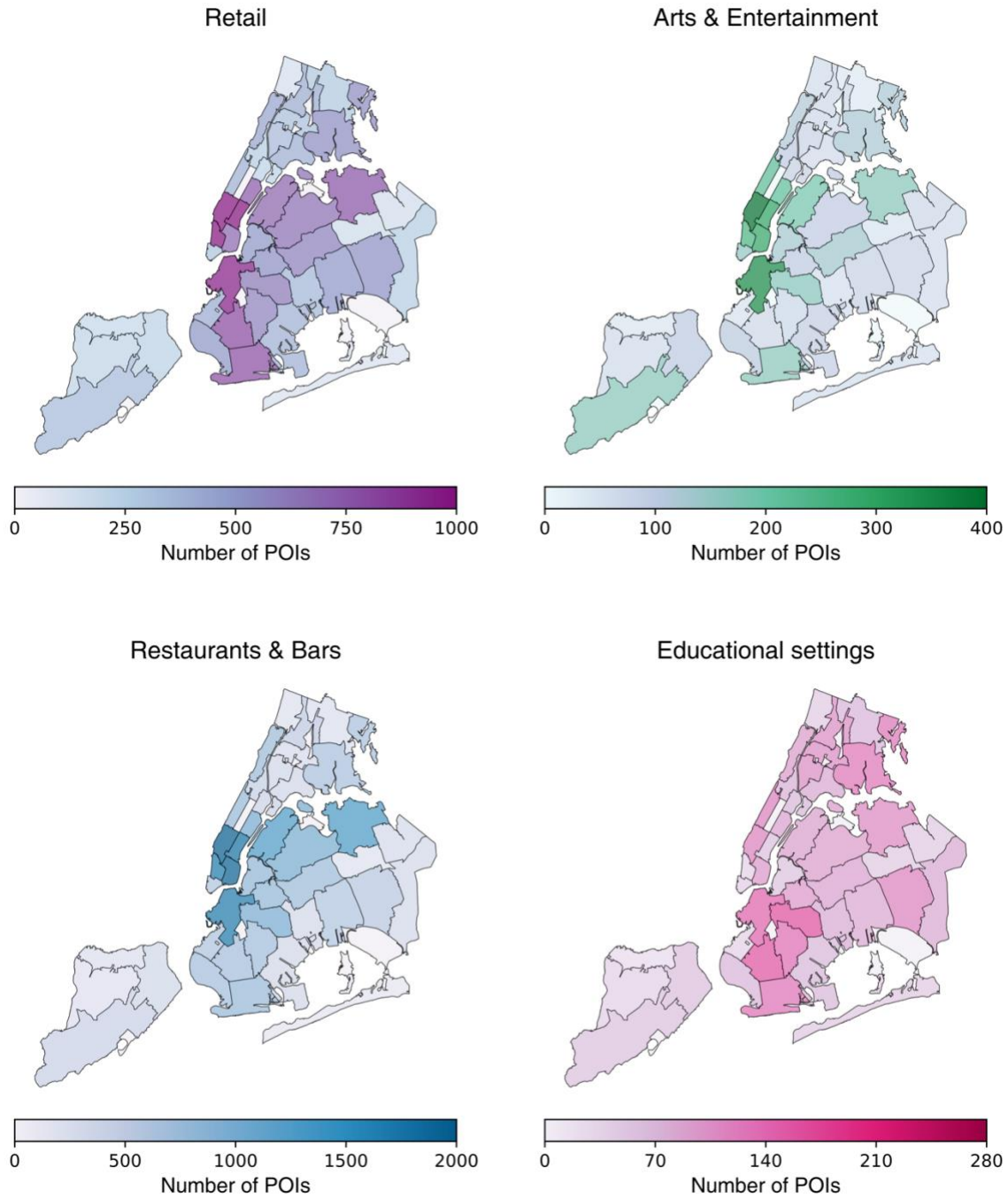


Fig. S2.

Geographical distribution of POIs. We visualize the number of POIs in each neighborhood for retail, arts & entertainment, restaurants & bars, and educational settings.

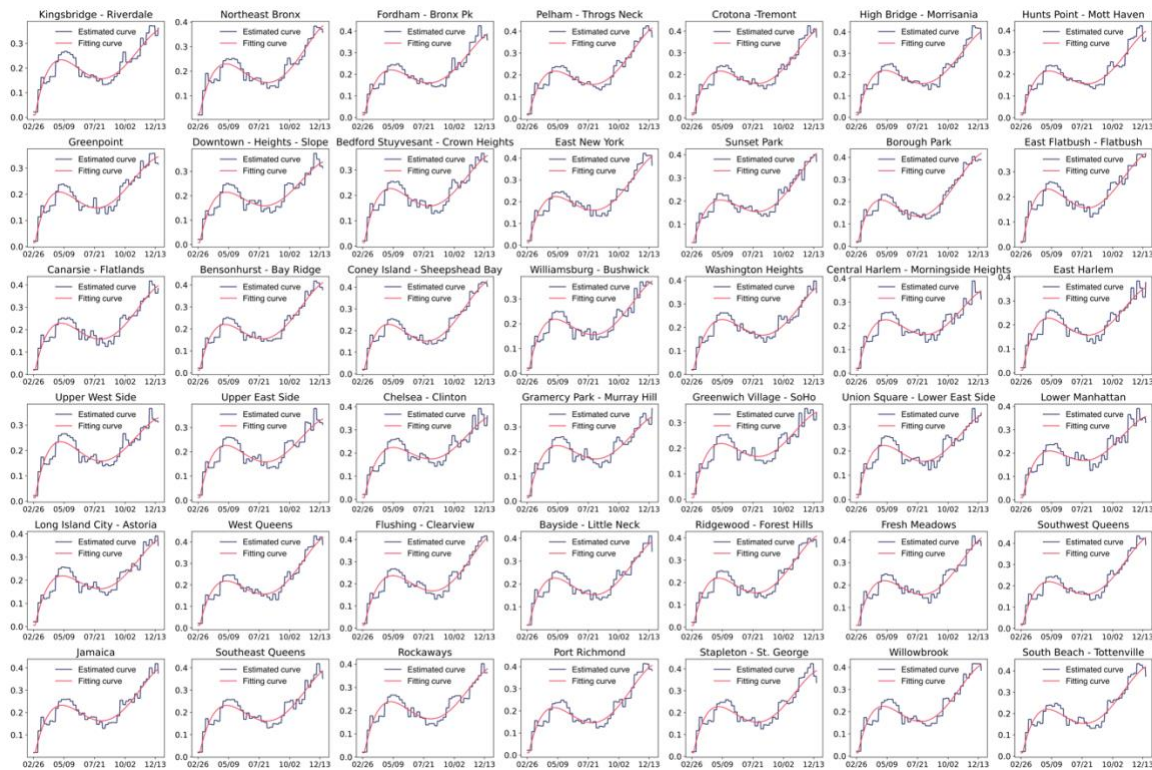


Fig. S3.

Weekly ascertainment rates in each neighborhood. The black lines are the estimated weekly ascertainment rate in Yang et al 2022 (36). We fit a smooth curve using a fourth-order polynomial function (red lines) for each neighborhood and used the smoothed curve to set ascertainment rates in our model.

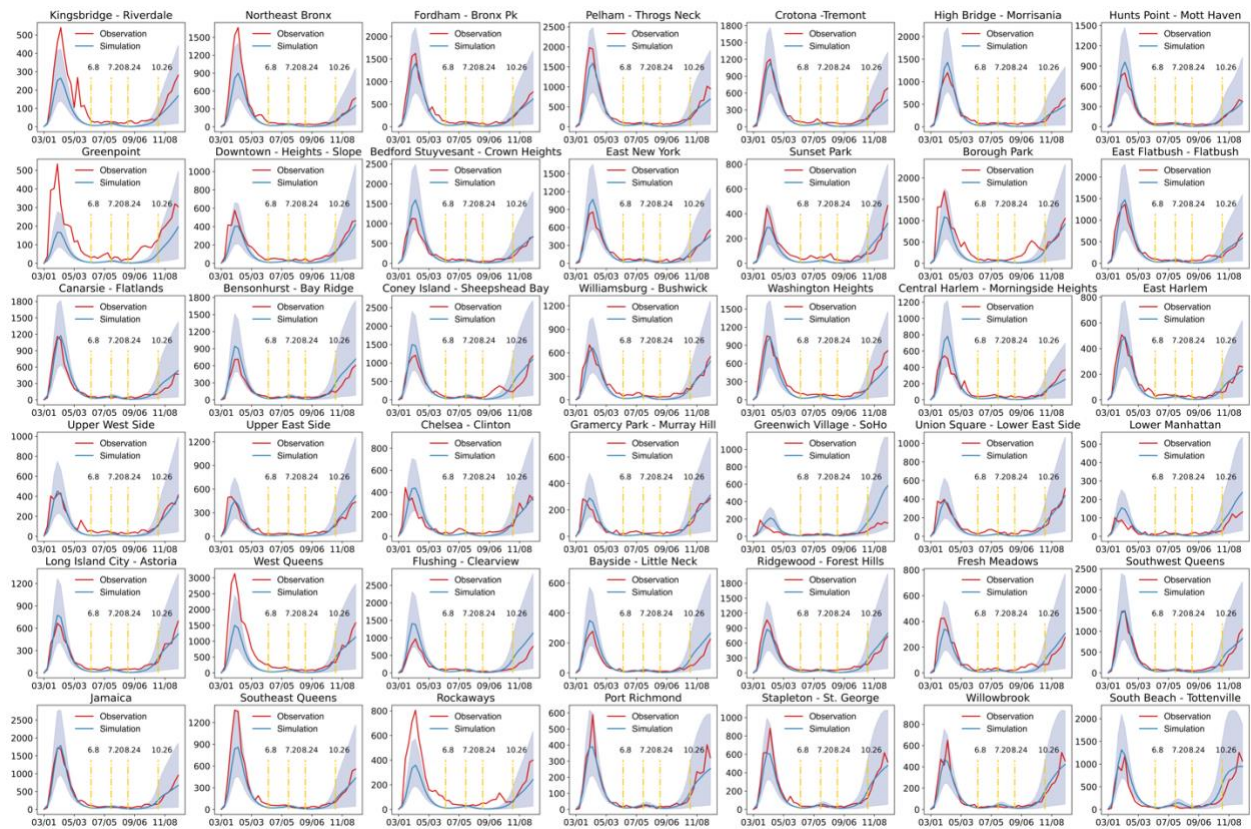


Fig. S4.

Model fitting in 42 NYC neighborhoods. Model simulations using estimated posterior parameters (blue) are compared with the reported weekly cases in each neighborhood. The blue shaded areas show 95% CIs, obtained from 500 simulations.

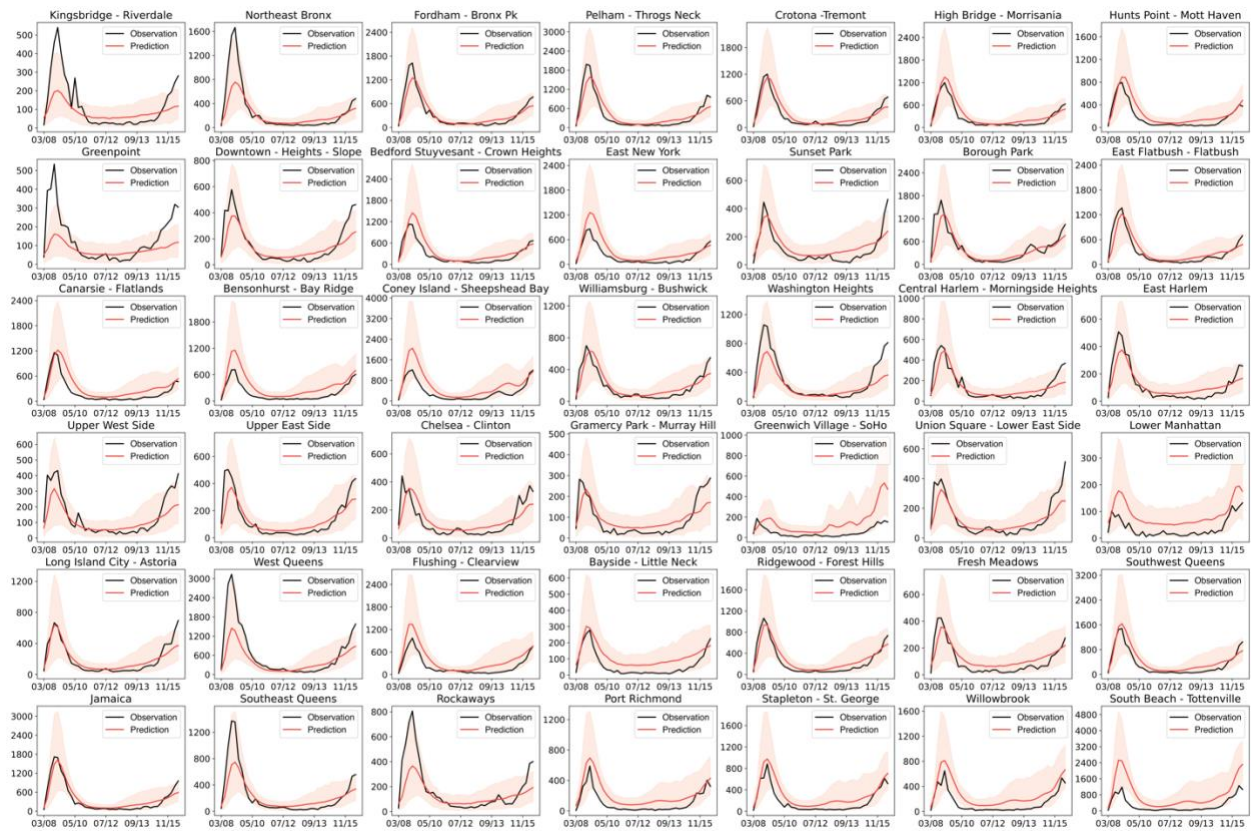


Fig. S5.

One-week ahead forecasts using the behavior-driven epidemic model. Shaded areas are 95% CIs obtained from 500 ensemble members.

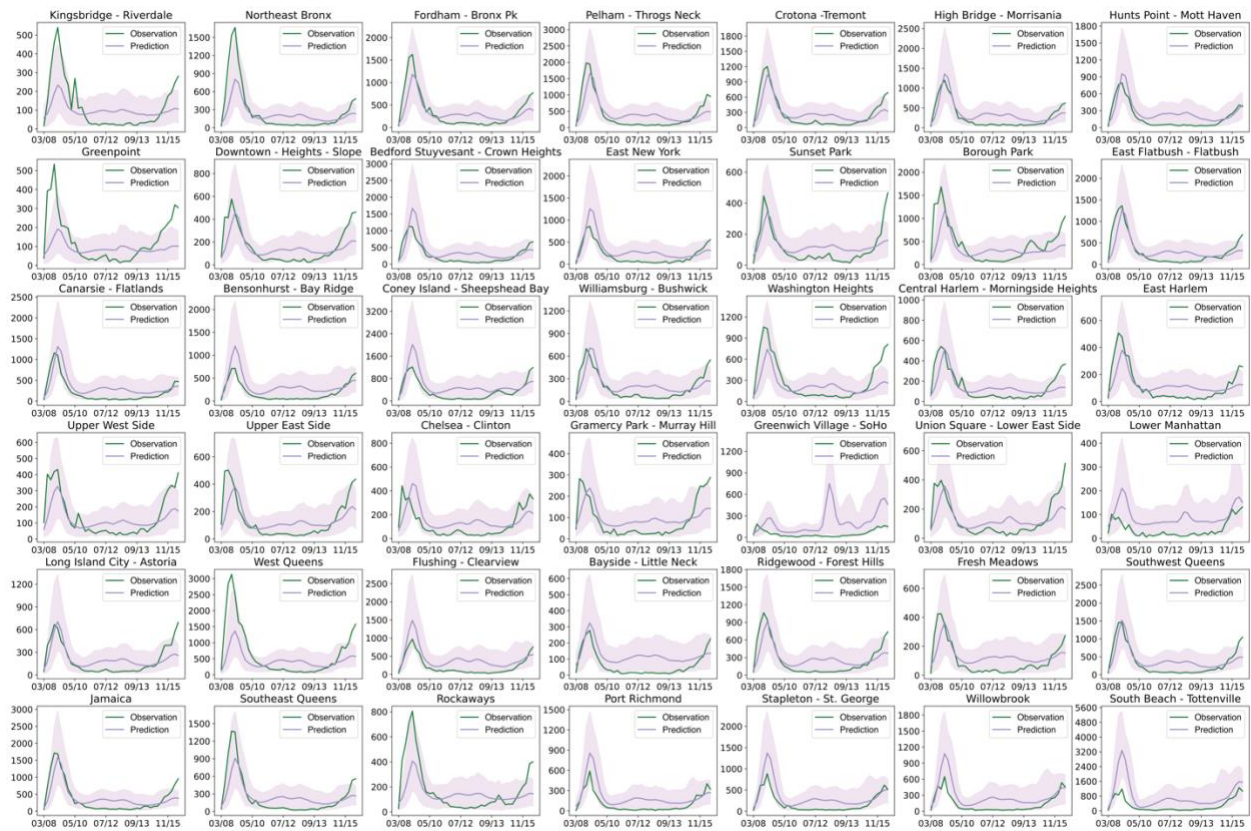


Fig. S6.

One-week ahead forecasts using the baseline model. Shaded areas are 95% CIs obtained from 500 ensemble members.

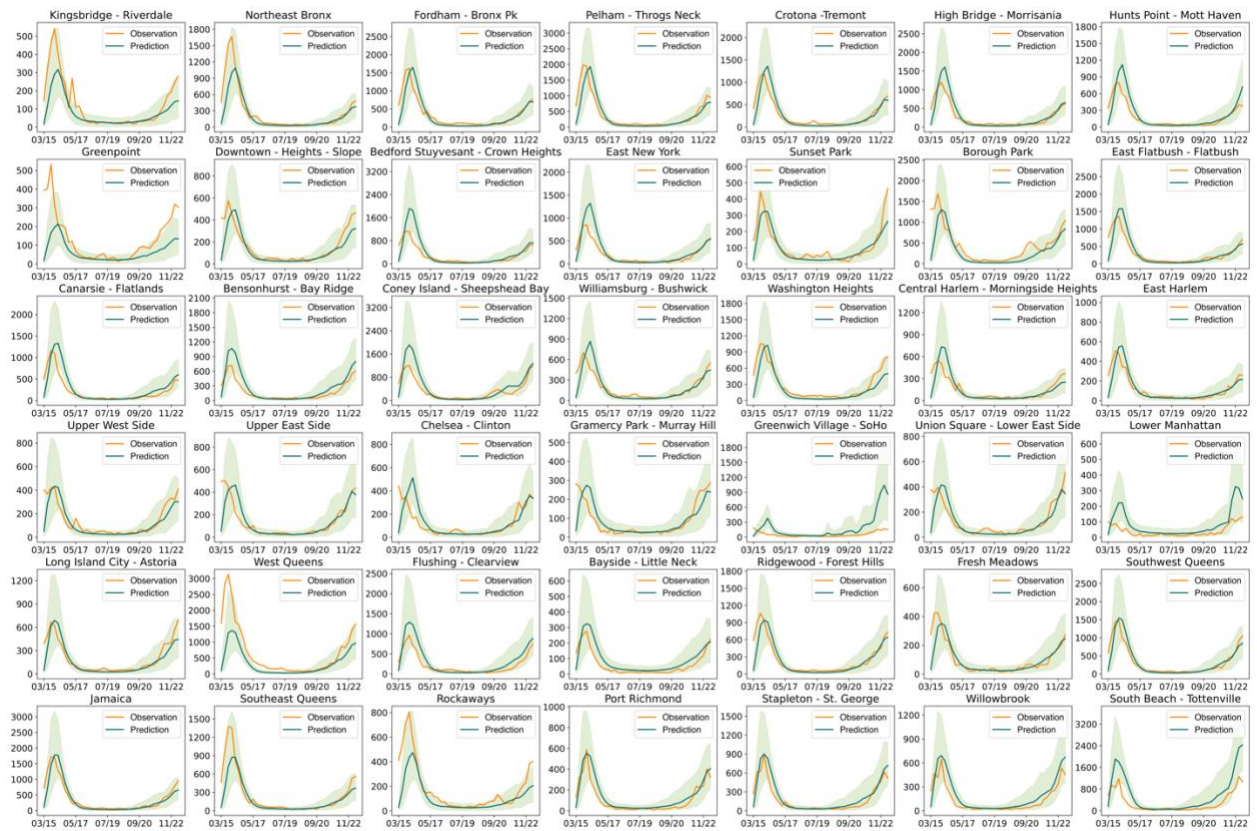


Fig. S7.

Two-week ahead forecasts using the behavior-driven epidemic model. Shaded areas are 95% CIs obtained from 500 ensemble members.

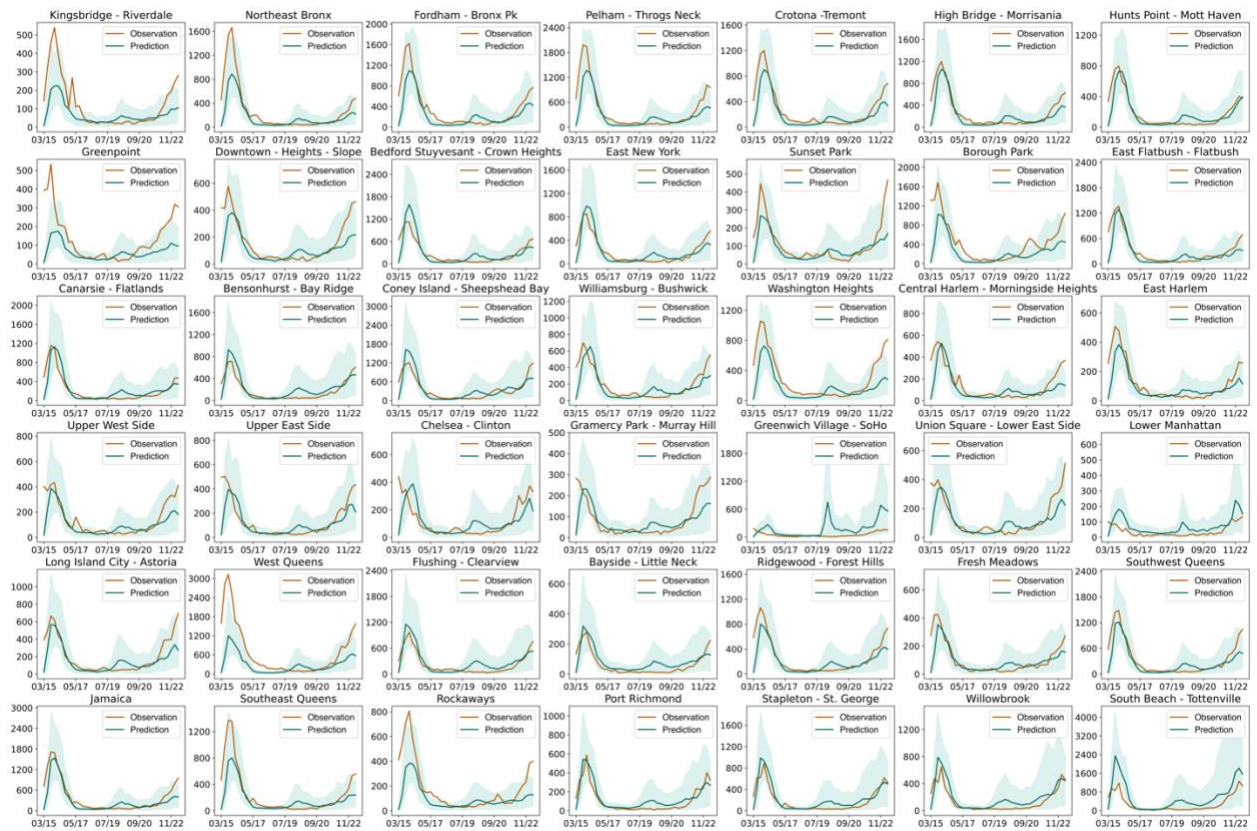


Fig. S8.

Two-week ahead forecasts using the baseline model. Shaded areas are 95% CIs obtained from 500 ensemble members.

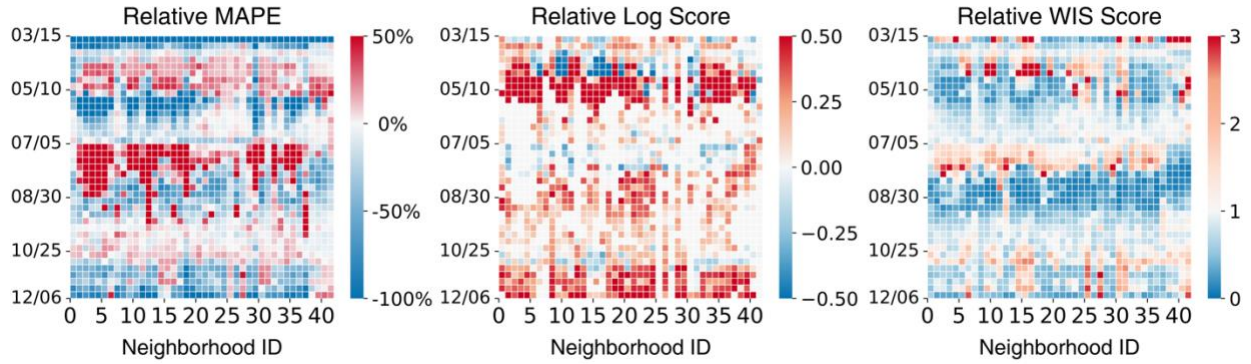


Fig. S9.

Comparison of two-week ahead forecasts. Comparison with a baseline model that does not incorporate mobility and contact patterns in place categories. From left to right, we present the relative mean absolute percentage error (MAPE) (MAPEs of the proposed model minus those of the baseline, with blue indicating better forecasts), relative log score (log scores of the proposed model minus those of the baseline, with red indicating better forecasts), and relative weighted interval score (WIS) (the ratio of the WIS scores of the proposed model to those of the baseline, with blue indicating better forecasts) for all 42 neighborhoods from March 8, 2020 to December 13, 2020.

Table S1.

Classification of POIs using the NAICS codes.

Place category	NAICS code and description	
Restaurants & bars	7223	Special Food Services
	7224	Drinking Places (Alcoholic Beverages)
	7225	Restaurants and Other Eating Places
Arts & entertainment	7111	Performing Arts Companies
	7113	Promoters of Performing Arts, Sports, and Similar Events
	7131	Amusement Parks and Arcades
	7139	Other Amusement and Recreation Industries
	7112	Spectator Sports
	7121	Museums, Historical Sites, and Similar Institutions
	7132	Gambling Industries
Educational settings	6111	Elementary and Secondary Schools
	6112	Junior Colleges
	6113	Colleges, Universities, and Professional Schools
	6115	Technical and Trade Schools
	6116	Other Schools and Instruction
Retail	4411	Automobile Dealers
	4412	Other Motor Vehicle Dealers
	4413	Automotive Parts, Accessories, and Tire Stores
	4421	Furniture Stores
	4422	Home Furnishings Stores
	4431	Electronics and Appliance Stores
	4441	Building Material and Supplies Dealers
	4442	Lawn and Garden Equipment and Supplies Stores
	4451	Grocery Stores
	4452	Specialty Food Stores
	4453	Beer, Wine, and Liquor Stores
	4461	Health and Personal Care Stores
	4471	Gasoline Stations
	4481	Clothing Stores
	4482	Shoe Stores
	4483	Jewelry, Luggage, and Leather Goods Stores
	4511	Sporting Goods, Hobby, and Musical Instrument Stores
	4512	Book Stores and News Dealers
	4522	Department Stores
	4523	General Merchandise Stores, including Warehouse Clubs and Supercenters
4531	Florists	
4532	Office Supplies, Stationery, and Gift Stores	
4533	Used Merchandise Stores	
4539	Other Miscellaneous Store Retailers	
Other places	Other NAICS codes	

Table S2.

Major policy changes in NYC in 2020.

Date	Policy	Source
March 13	Mayor Bill de Blasio issued a state of emergency. In partnership with the State, the City will place occupancy and spatial limits on New York City’s large venues and cancel all existing and future events exceeding a capacity of over 500 people until further notice, with restrictions put in place on establishments with a capacity under 500 people. The City has also set a target benchmark of getting 10 percent (roughly 35,000 people) of the City’s workforce to telecommute and 20 percent (roughly 71,000 people) on staggered work schedules	https://www.nyc.gov/office-of-the-mayor/news/138-20/mayor-de-blasio-issues-state-emergency
June 8	NYC begins Phase 1 reopening. New York City is now eligible to resume elective surgeries and ambulatory care. Restaurants and bars have been allowed to offer takeout and delivery throughout the pandemic. Outdoor businesses like drive-in movies, landscaping and gardening are allowed to resume. Nonessential retail stores, for delivery, curbside and in-store pickup are allowed to resume. Employees of construction jobs, wholesale, manufacturing, agriculture and retail companies can go back to work.	https://www.governor.ny.gov/news/new-york-city-enters-phase-one-reopening-today-governor-cuomo-announces-new-york-city-now
July 20	NYC begins Phase 4 reopening on July 19. Zoos, botanical gardens, historical sites, outdoor museums, and parks are allowed to reopen. Movie theaters, concert venues, amusement parks remain closed.	https://www.timeout.com/new-york/news/whats-next-after-phase-4-what-you-need-to-know-about-nycs-reopening-072220
August 24	Gyms and Fitness Centers Can Reopen Starting August 24. Museums, aquariums and other low-risk indoor cultural institutions were allowed to reopen.	https://www.governor.ny.gov/news/governor-cuomo-announces-gyms-and-fitness-centers-can-reopen-starting-august-24 https://www.nbcnewyork.com/news/local/cuomo-nyc-museums-can-reopen-later-this-month-bowling-alleys-on-monday/2568000
October 21	New York “Micro-Cluster” Strategy started.	https://www.governor.ny.gov/sites/default/files/atoms/files/MicroCluster_Metrics_10.21.20_FINAL.pdf

Table S3.

Posterior parameters estimated in the best-fitting model in the grid search and model selection of parameters a and b . We show the mean estimates and the 95% CIs from MCMC samples.

Category	Estimate	95% CI
Retail	1.85	(0.07-6.16)
Arts & Entertainment	0.63	(0.02-2.20)
Restaurants & Bars	27.88	(25.43-30.09)
Educational settings	17.48	(7.04-26.90)
Other places	21.21	(15.21-26.84)
β_h	0.0032	(0.0001-0.0104)

Table S4.

Geweke diagnostic for MCMC chain convergence. A z-score between $(-1.96, 1.96)$ or a p-value larger than 0.05 indicates the convergence of MCMC chains.

Category	Phase	Z-score	p-value
Retail	P1	-0.961	0.336
	P2	-0.243	0.808
	P3	-0.317	0.751
	P4	1.146	0.252
	P5	1.275	0.202
Arts & Entertainment	P1	0.068	0.946
	P2	0.123	0.902
	P3	-0.566	0.572
	P4	-0.396	0.692
	P5	-0.051	0.959
Restaurants & Bars	P1	2.107	0.035
	P2	-0.315	0.753
	P3	1.278	0.201
	P4	0.710	0.478
	P5	-1.705	0.088
Educational settings	P1	-0.863	0.388
	P2	-1.202	0.230
	P3	0.286	0.775
	P4	0.122	0.903
	P5	-1.328	0.184
Other places	P1	-0.547	0.584
	P2	0.976	0.329
	P3	-1.928	0.054
	P4	-0.037	0.971
	P5	-0.511	0.609
β_h	P1	-1.077	0.281
	P2	0.277	0.782
	P3	-0.832	0.405
	P4	-0.996	0.319
	P5	1.628	0.104