

PREDICTING THE SUCCESS OF RETAIL ESTABLISHMENTS IN NEW YORK

Authors:
Kateryna Peikova
Mykola Trokhymovych
Taras Sarvas
Maryanna Temnyk

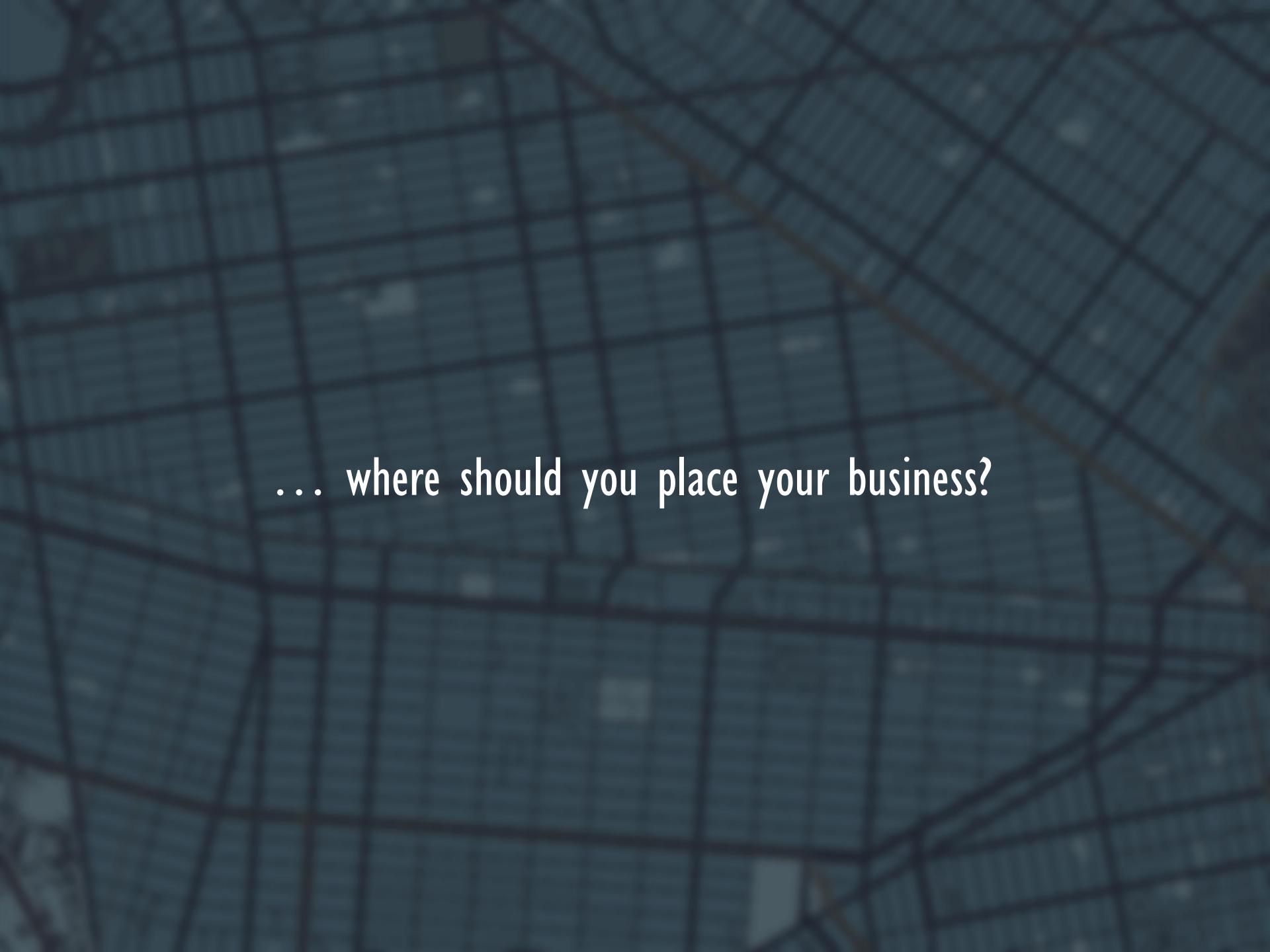
Supervisor:
Anastasios Noulas



A beautiful city of New York...



... and you are a happy business owner ...



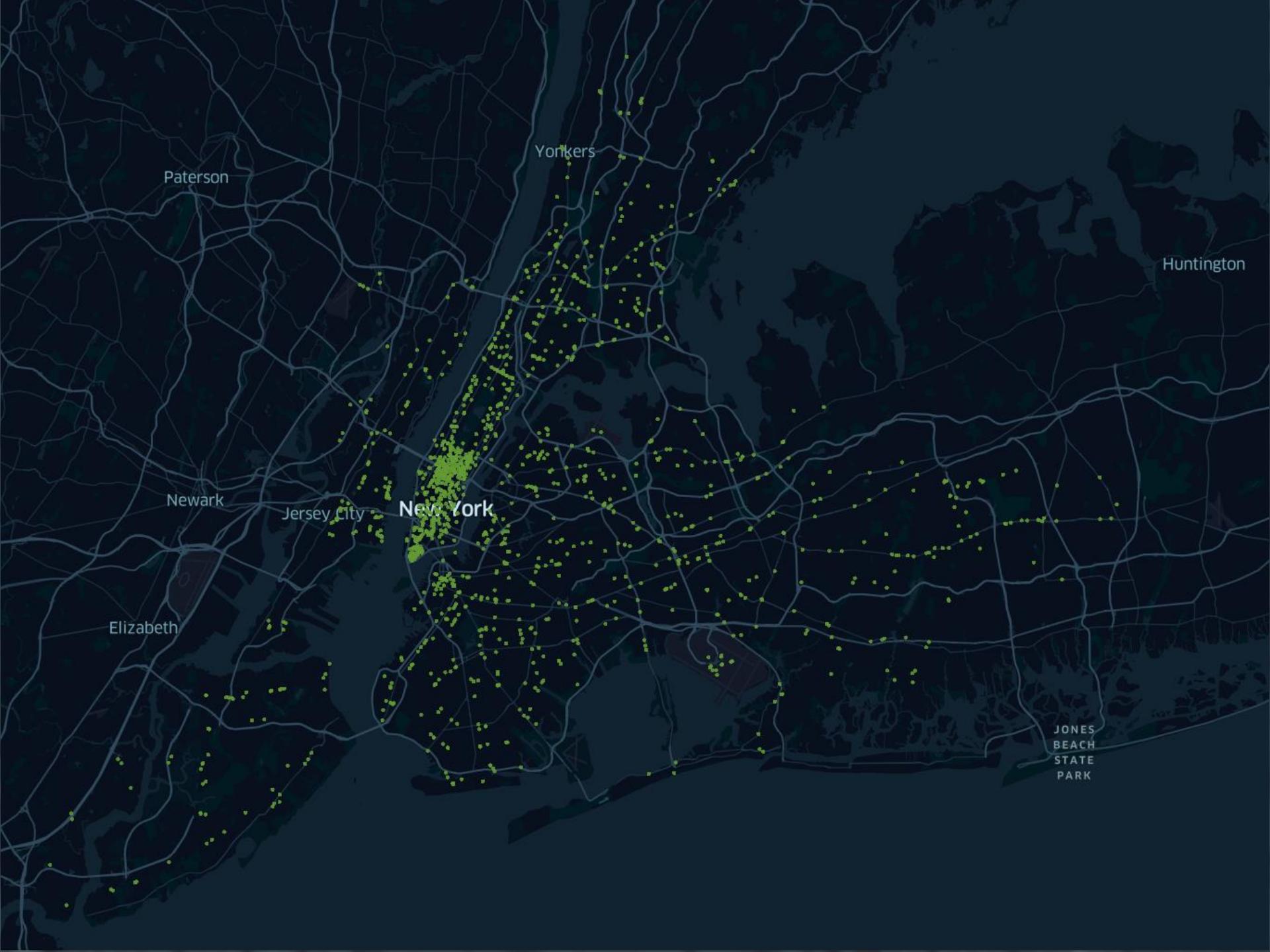
... where should you place your business?





wat







Data & Features

Data:

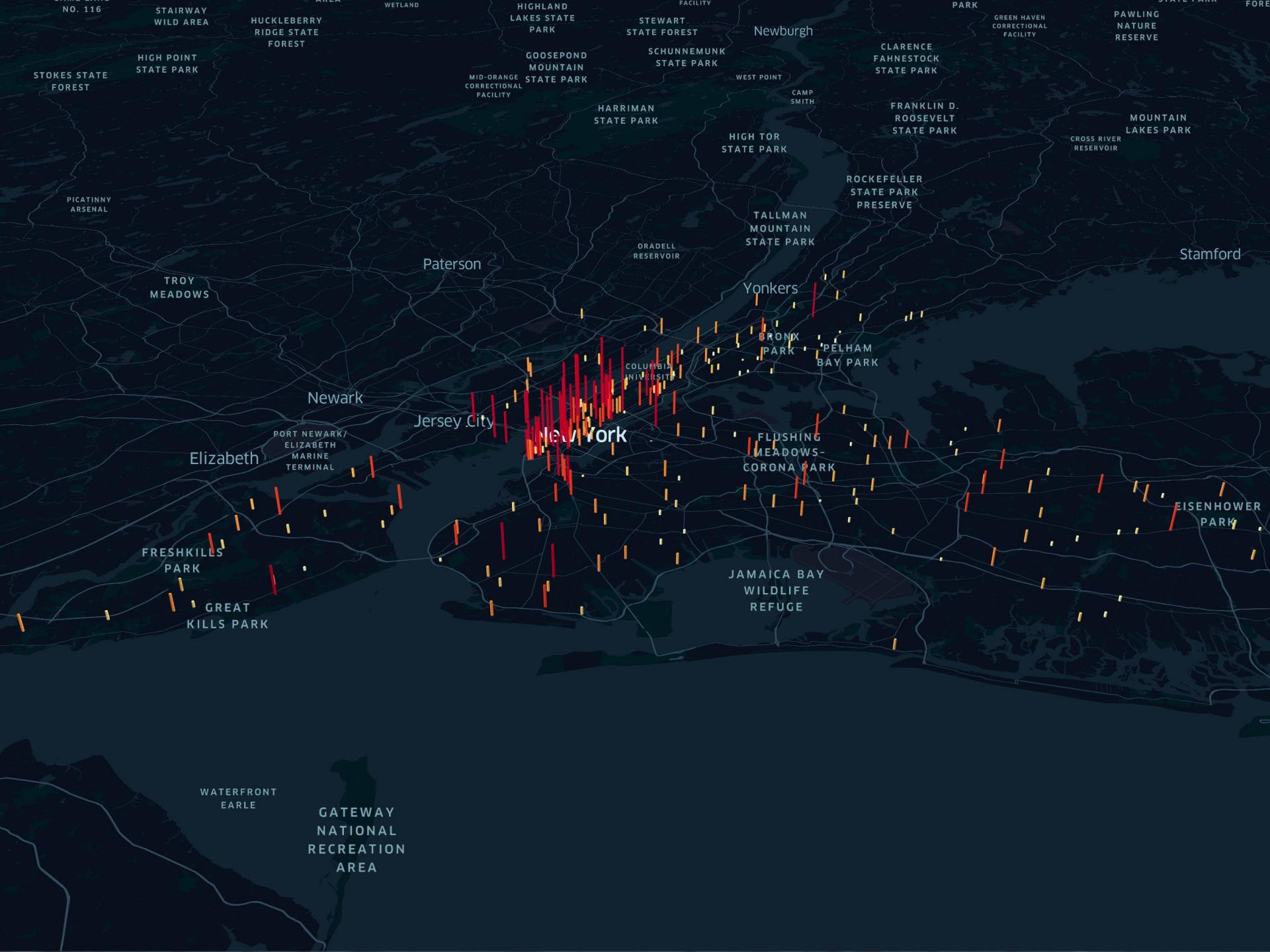
Foursquare check-ins
New York Yellow taxi

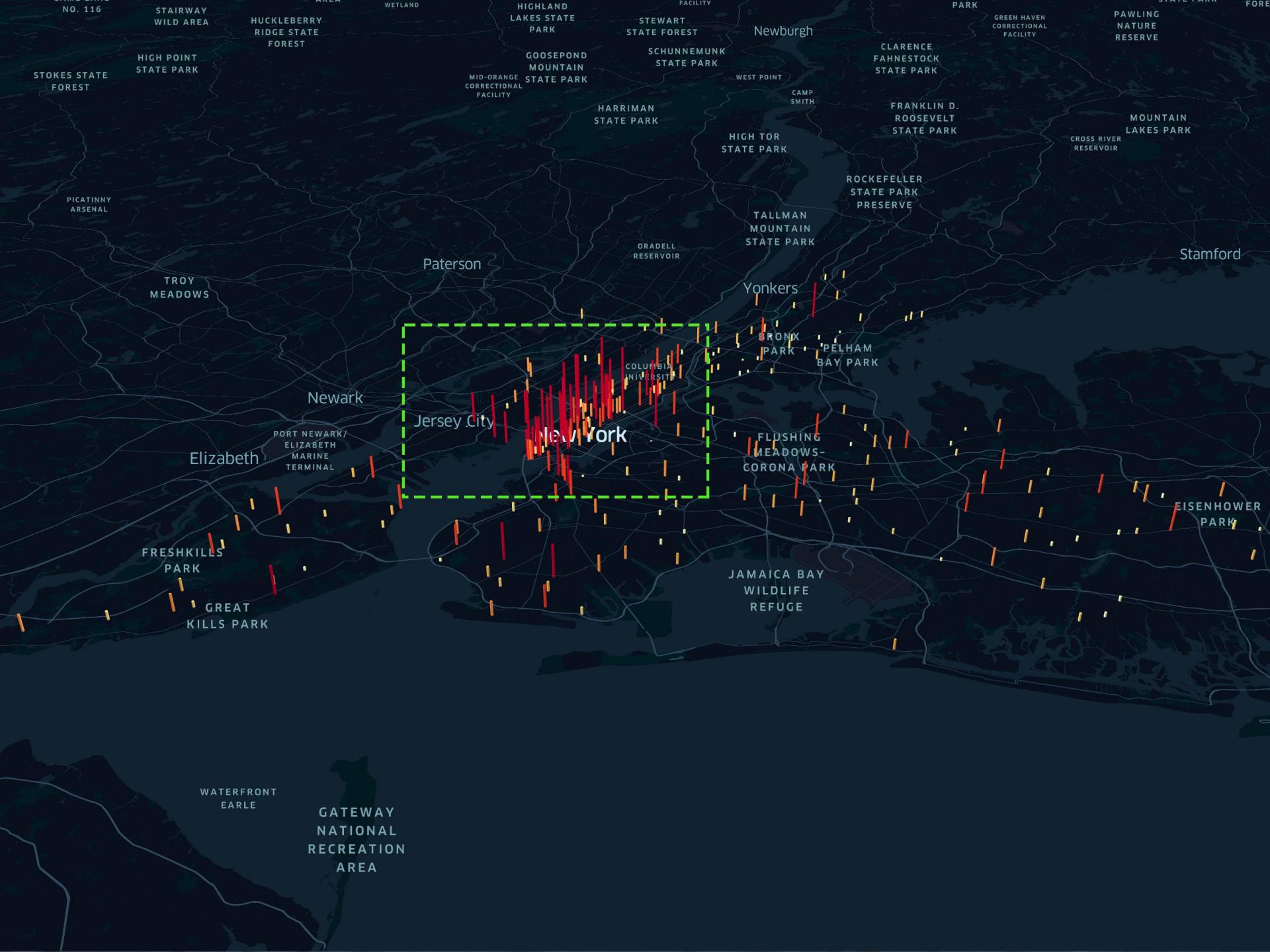
Geographic Features

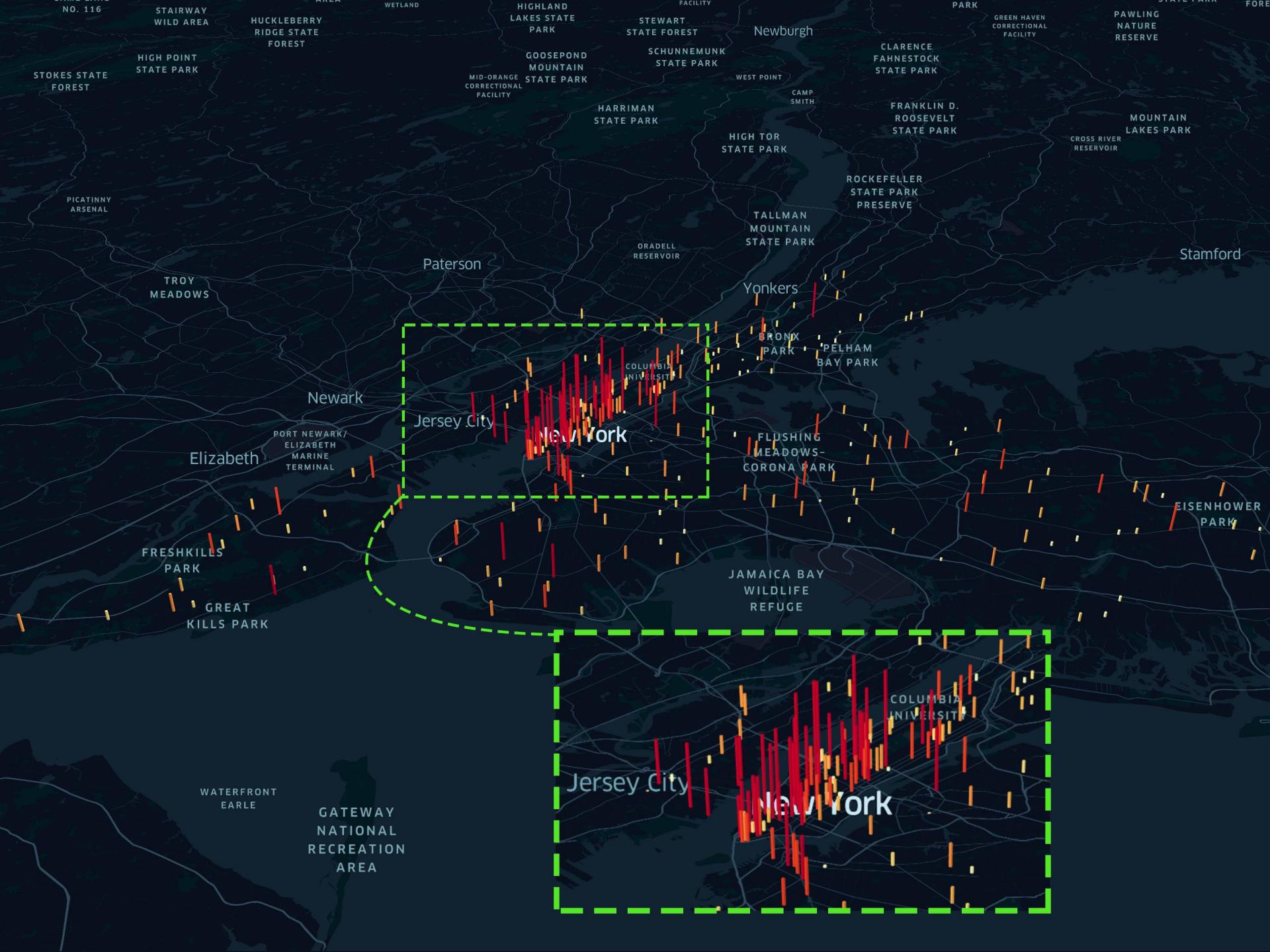
- ▶ Density
- ▶ Neighbors entropy
- ▶ Competitiveness

Mobility Features

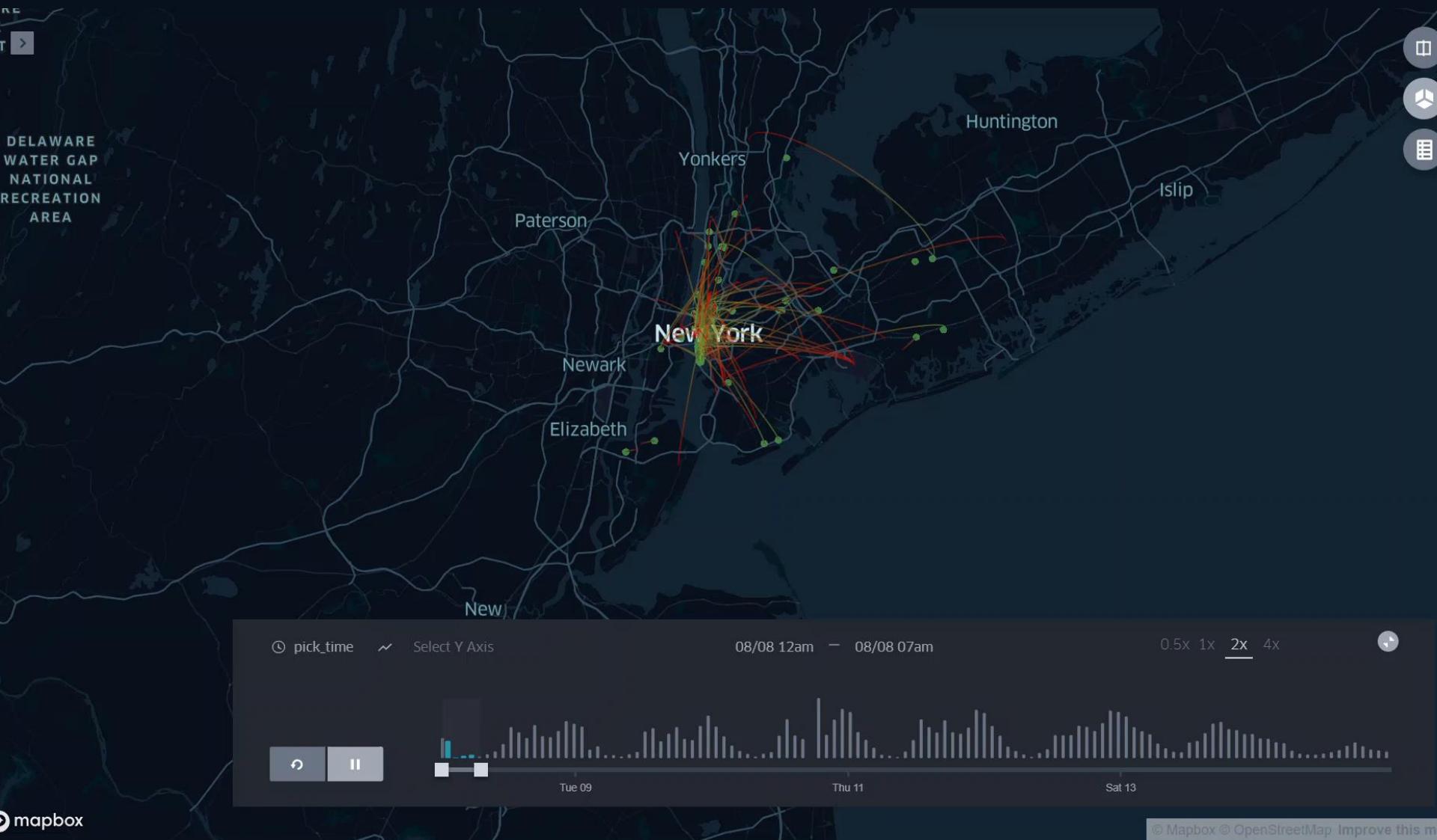
- ▶ Area popularity
- ▶ Transition density
- ▶ Incoming flow
- ▶ Transition quality
- ▶ Taxi transition density
- ▶ Taxi incoming flow





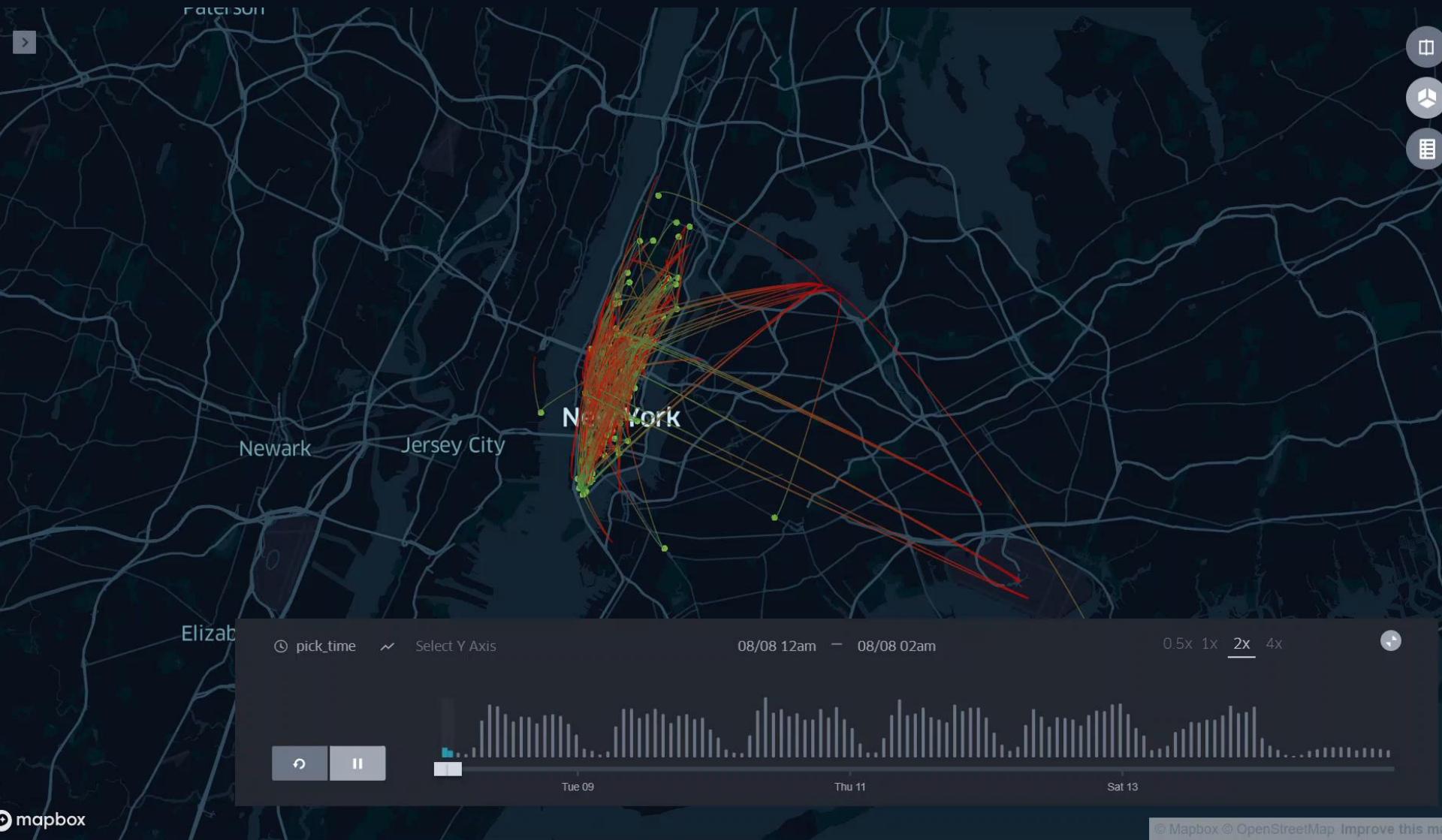


Foursquare income flow



<https://youtu.be/5IkWZJQKbxE>

Taxi income flow



https://youtu.be/GpQKUg_Zuvo

Model & Metrics

Learning Type:
Supervised learning

Models:

- ▶ XGBoost Regressor
- ▶ SGDRegressor
- ▶ DecisionTreeRegressor
- ▶ ElasticNet
- ▶ Lasso
- ▶ Ridge

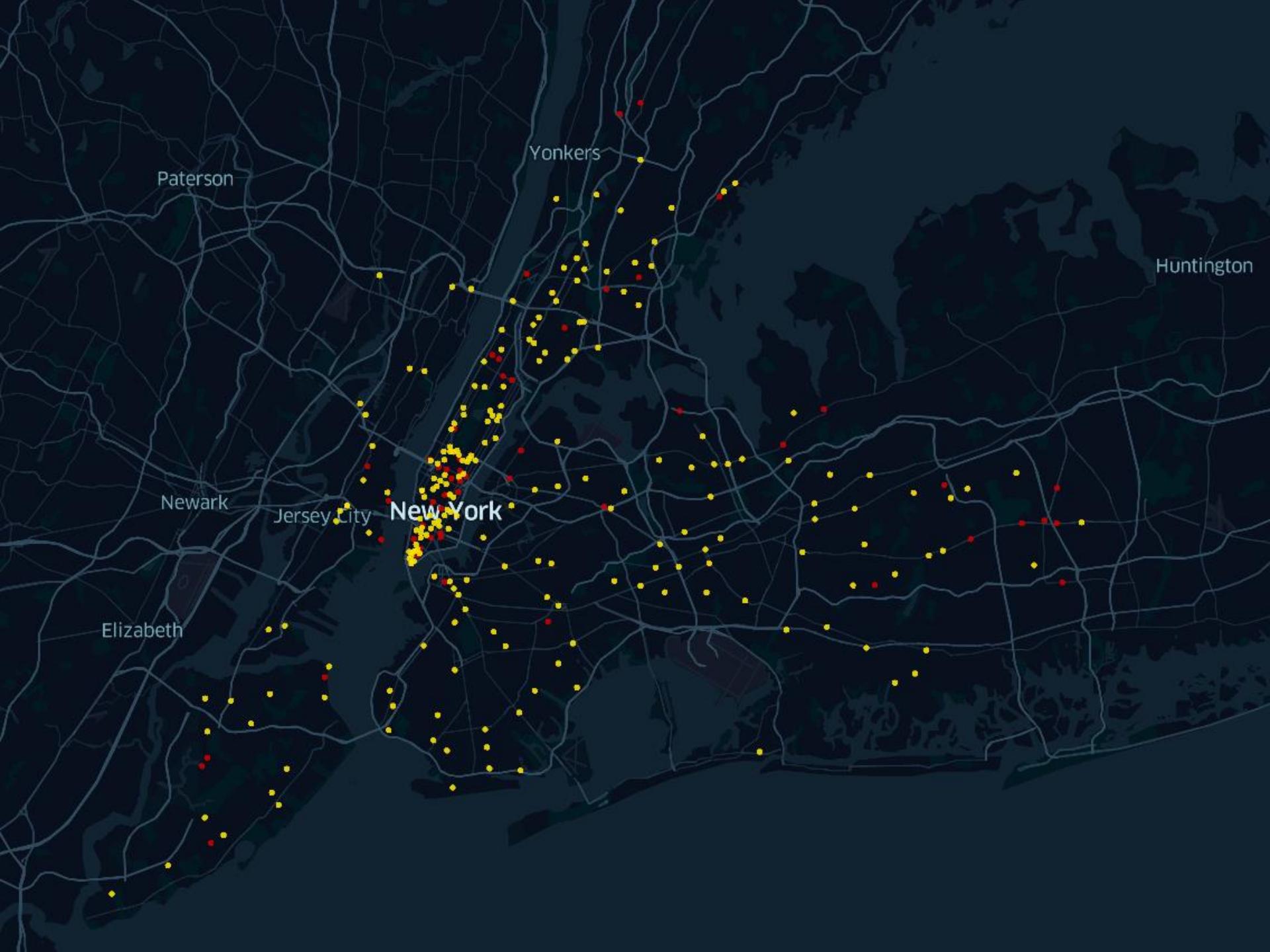
Hyperparameter
Optimization:

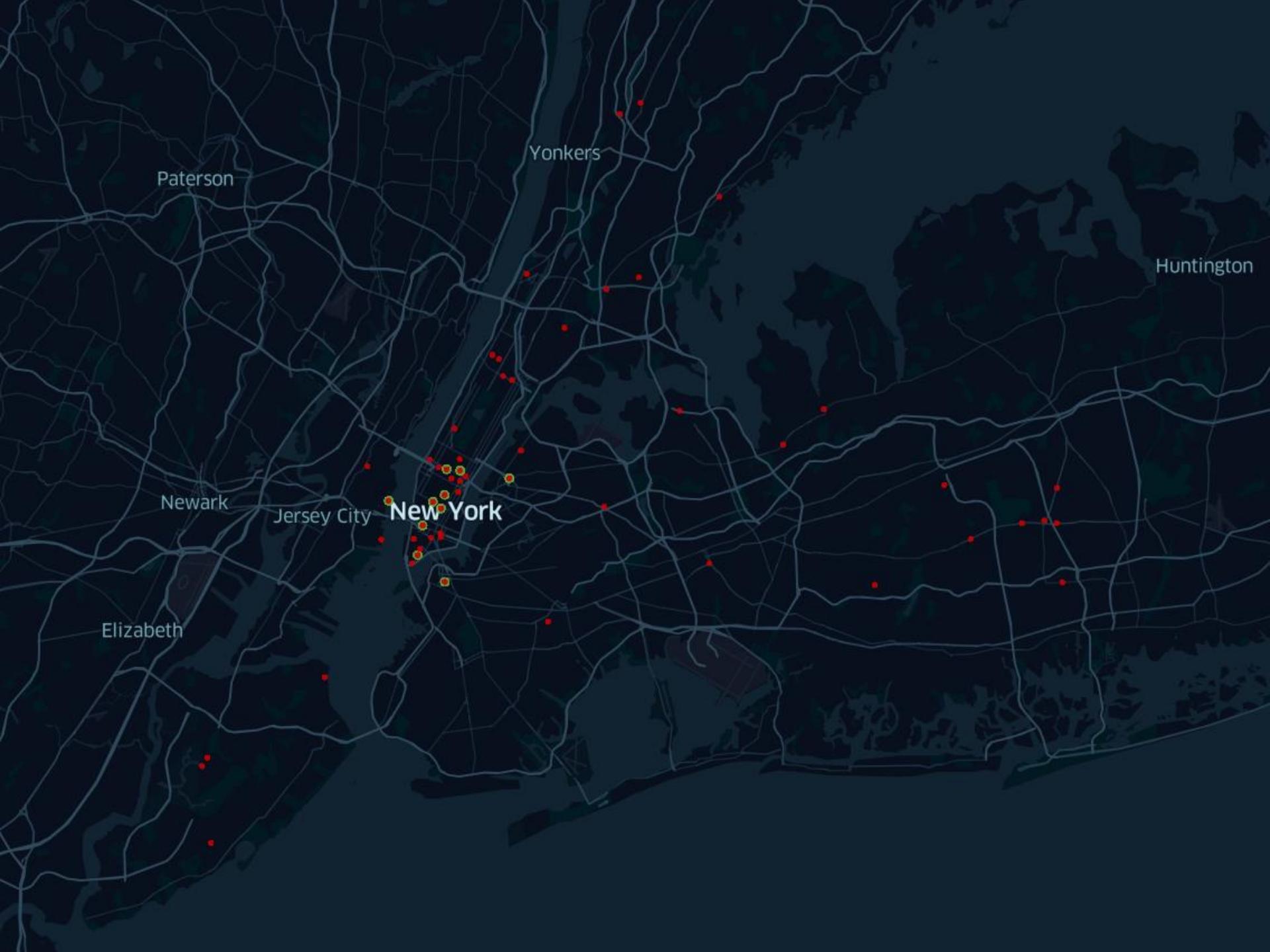
- ▶ Tree of Parzen
Estimators (TPE)
- ▶ RandomizedSearchCV

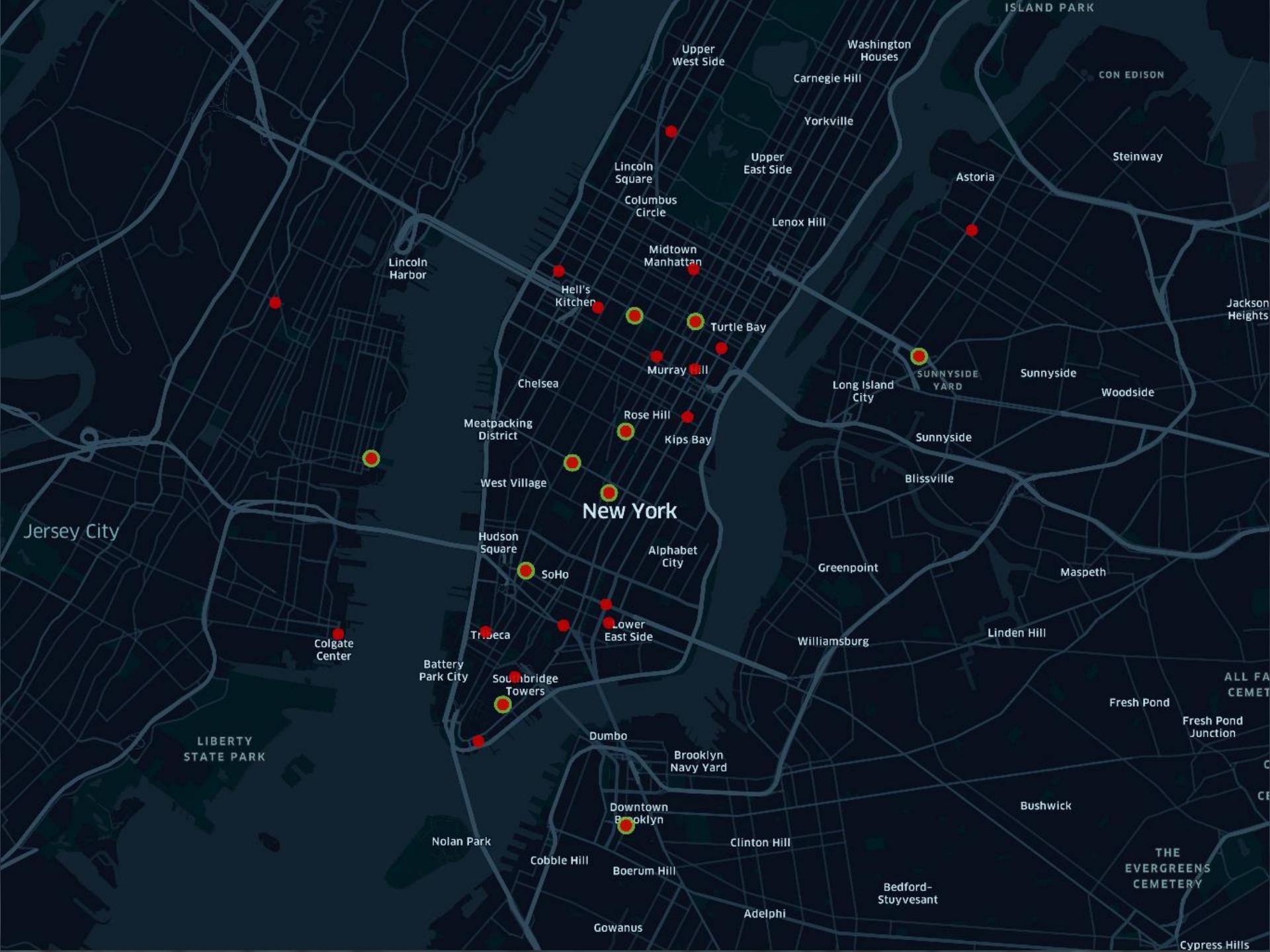
Metrics:

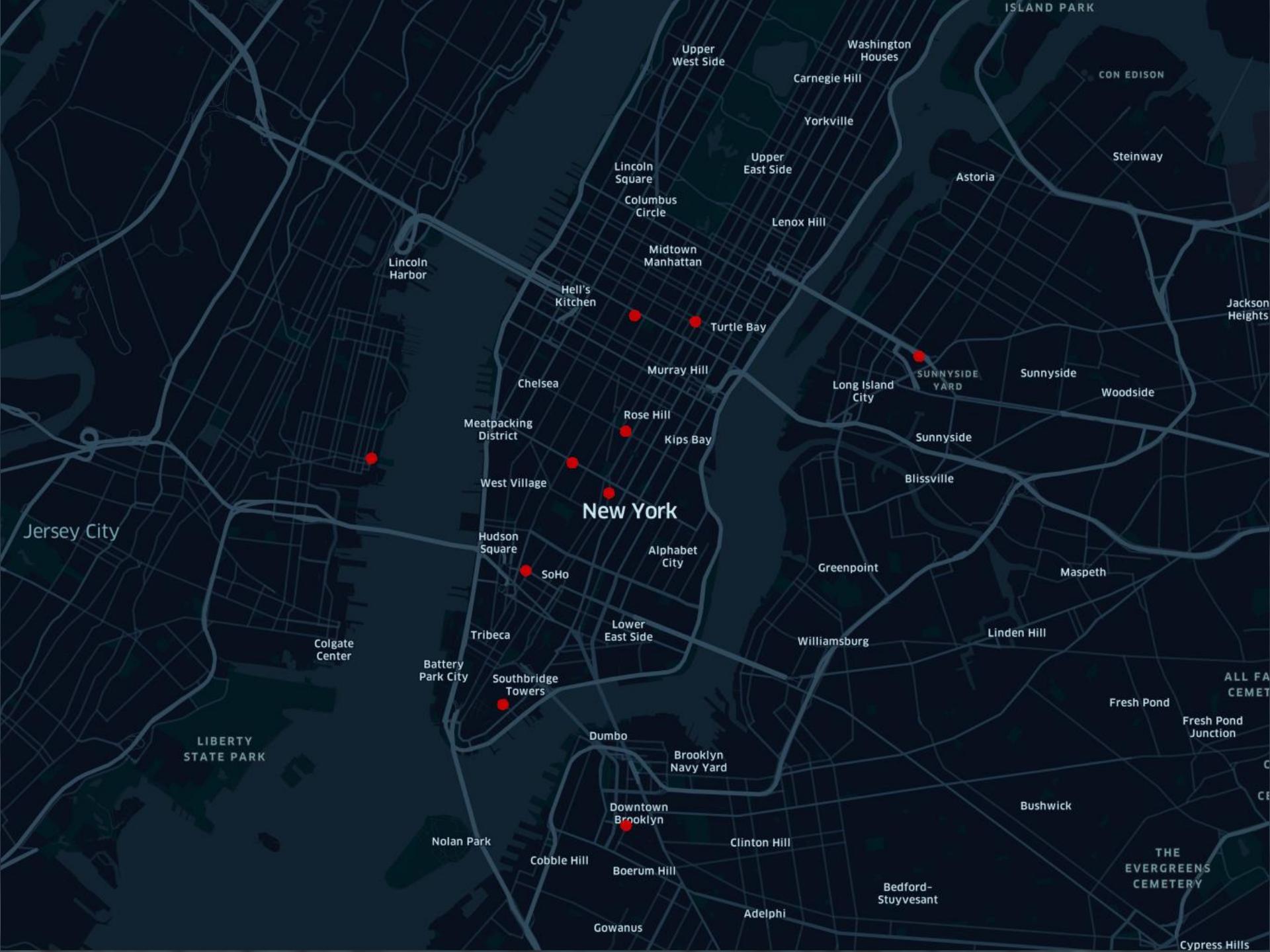
- ▶ NDCG@10











Correlation Heatmap

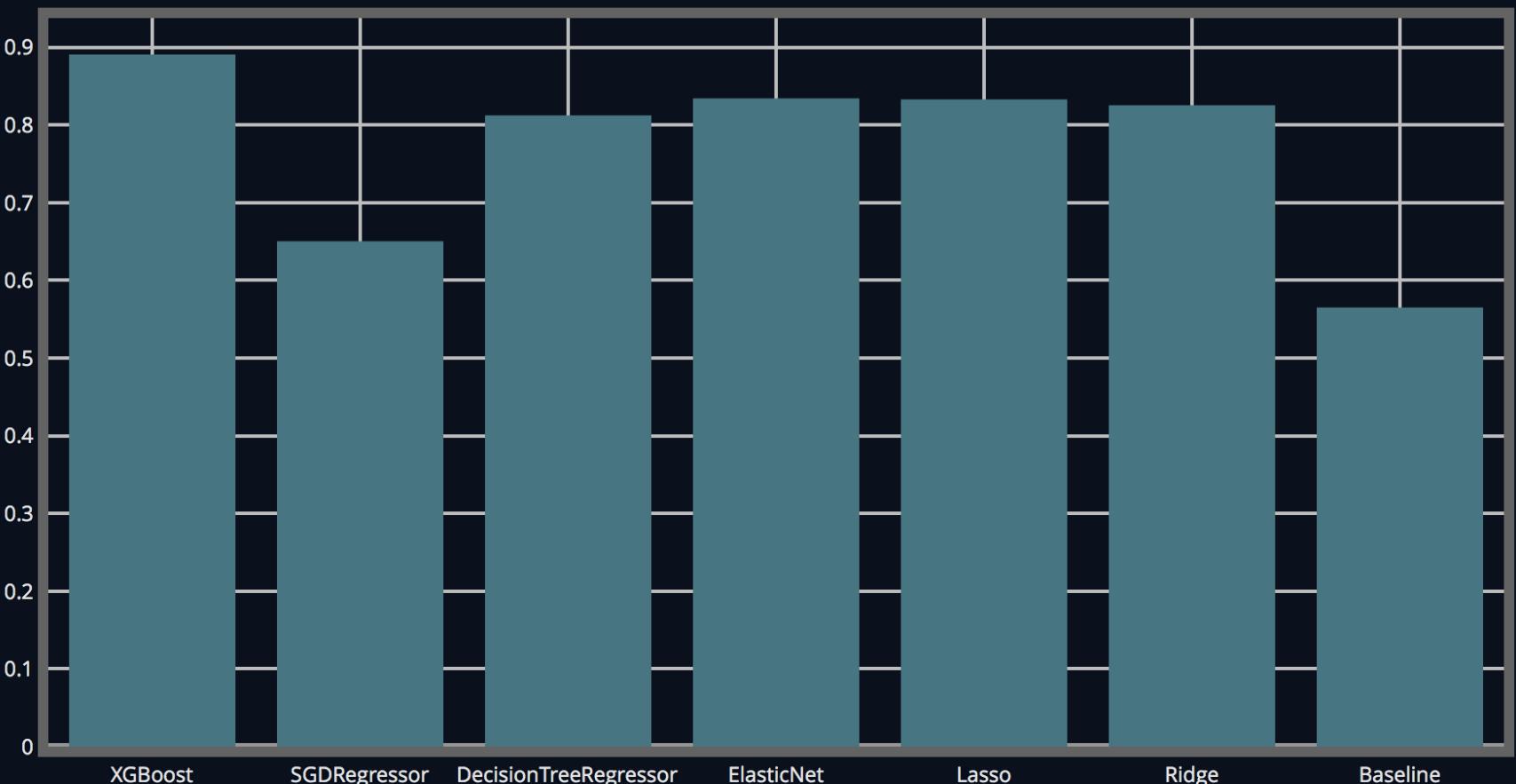
		Correlations between features and target									
		density	neighbors_entropy	competitiveness	area_popularity	transition_density	incoming_flow	transition_quality	taxi_area_popularity	taxi_incoming_flow	target_check_ins
density	density	1	0.61	0.42	0.84	0.79	0.86	0.9	0.88	0.87	0.59
	neighbors_entropy	0.61	1	0.59	0.52	0.49	0.53	0.54	0.52	0.52	0.38
competitiveness	competitiveness	0.42	0.59	1	0.35	0.33	0.35	0.36	0.36	0.35	0.29
	area_popularity	0.84	0.52	0.35	1	0.98	0.99	0.86	0.76	0.74	0.58
transition_density	transition_density	0.79	0.49	0.33	0.98	1	0.94	0.82	0.69	0.68	0.58
	incoming_flow	0.86	0.53	0.35	0.99	0.94	1	0.87	0.78	0.76	0.56
transition_quality	transition_quality	0.9	0.54	0.36	0.86	0.82	0.87	1	0.77	0.77	0.61
	taxi_area_popularity	0.88	0.52	0.36	0.76	0.69	0.78	0.77	1	0.99	0.48
taxi_incoming_flow	taxi_incoming_flow	0.87	0.52	0.35	0.74	0.68	0.76	0.77	0.99	1	0.45
	target_check_ins	0.59	0.38	0.29	0.58	0.58	0.56	0.61	0.48	0.45	1

NDCG Metrics

$$DCG_p = \sum_{i=1}^p \frac{rel_i}{\log_2(i + 1)}$$

$$nDCG_p = \frac{DCG_p}{IDCG_p}$$

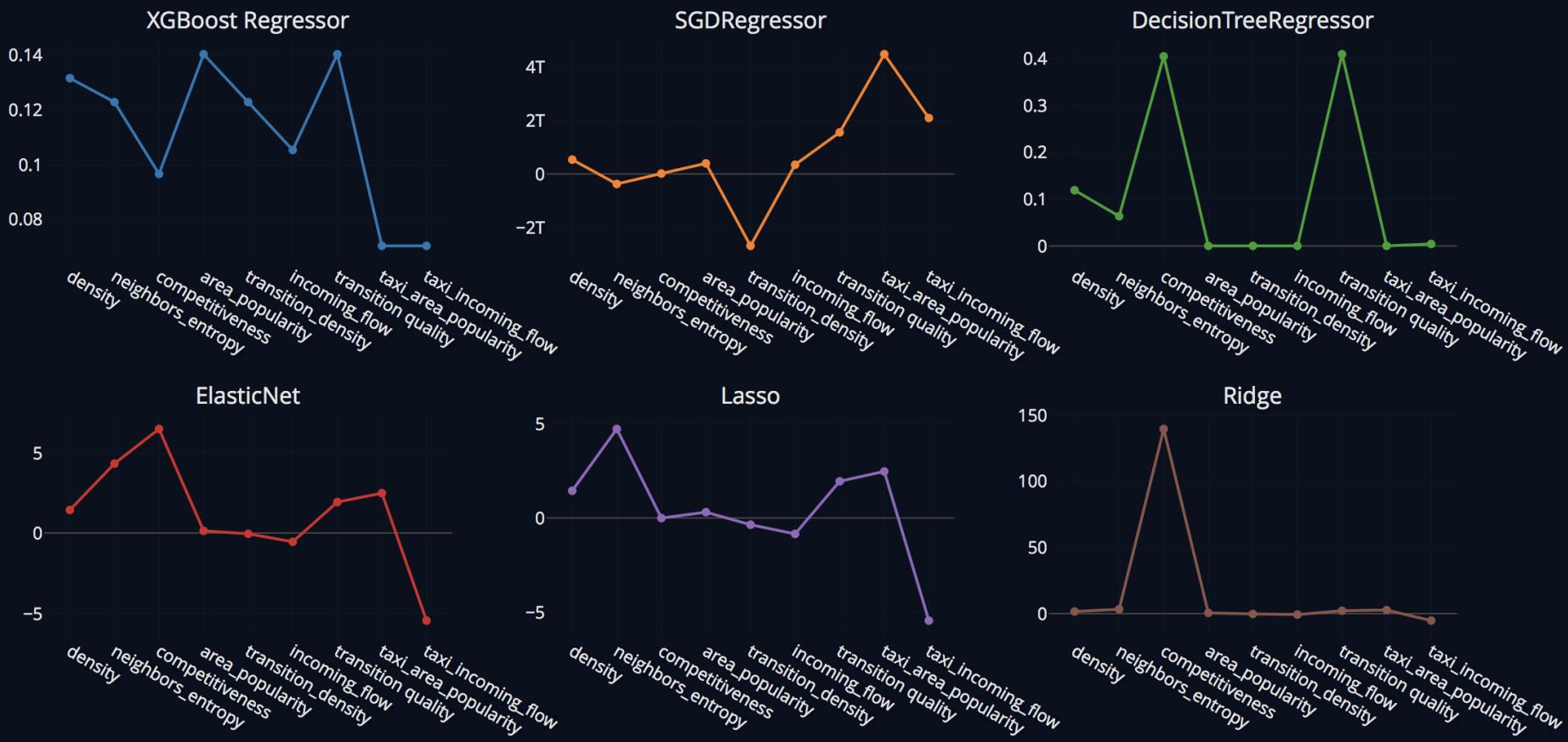
Model Performance

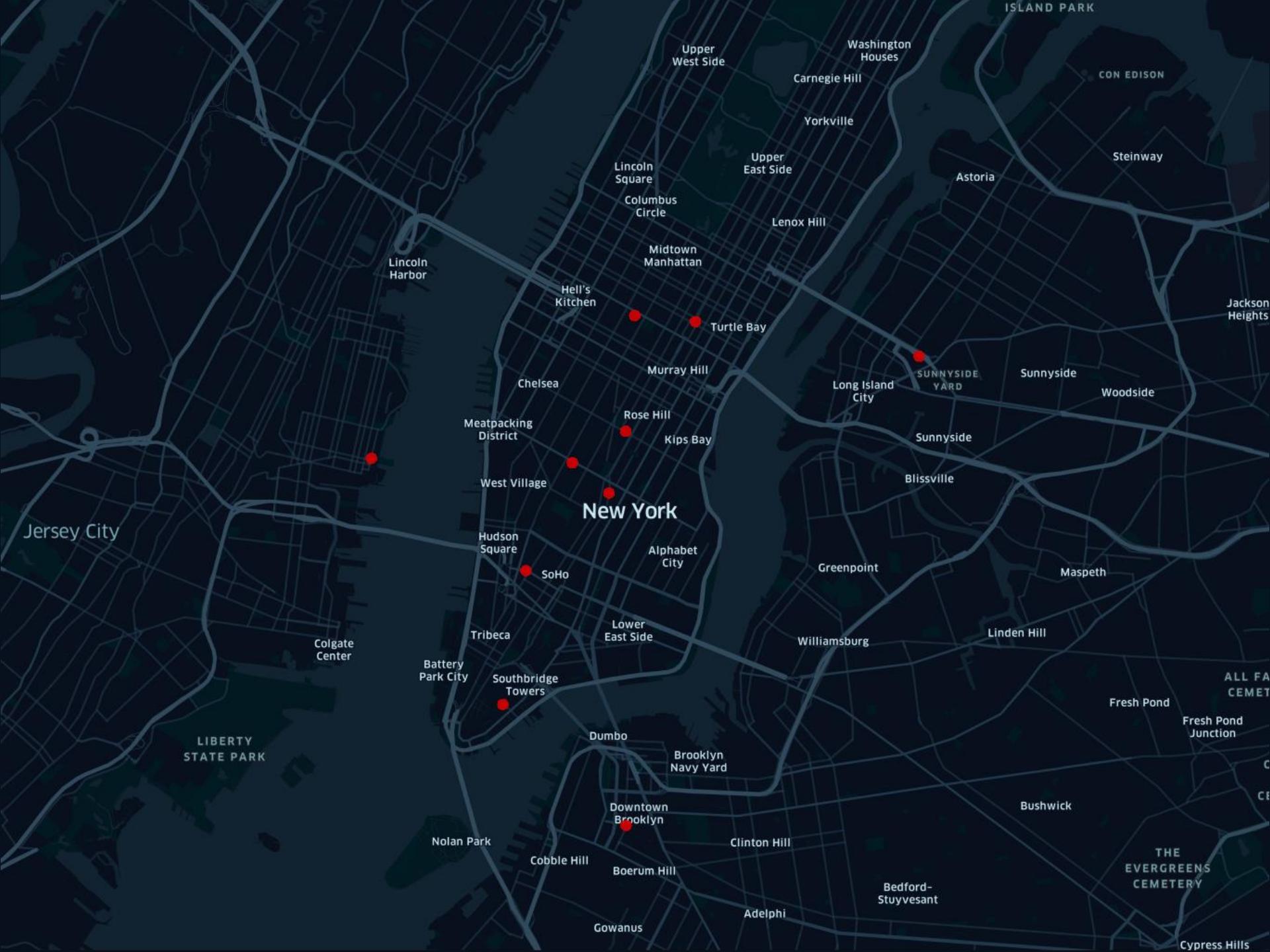


Model Performance

	cross_val_ndcg_mean	cross_val_ndcg_std
XGBoost	0.89053	0.0465816
ElasticNet	0.834106	0.0281904
Lasso	0.8328	0.0316315
Ridge	0.825261	0.0300194
DecisionTreeRegressor	0.812228	0.0673468
SGDRegressor	0.650265	0.0238567
Baseline	0.562394	0.0844208

Feature importance for each model





New York

Jersey City

Colgate
Center

LIBERTY
STATE PARK

Lincoln Harbor
Hell's Kitchen
Meatpacking District
West Village
Hudson Square
Tribeca
Battery Park City
Southbridge Towers
Dumbo
Nolan Park
Cobble Hill
Boerum Hill
Gowanus
Adelphi
Clinton Hill
Bedford-Stuyvesant
Bushwick
Fresh Pond Junction
C
CEMETERY
THE EVERGREENS CEMETERY
Cypress Hills

ISLAND PARK

CON EDISON

Steinway

Jackson
Heights

SUNNYSIDE YARD
Sunnyside
Woodside

Washington
Houses

Carnegie Hill

Yorkville

Upper
East Side

Lenox Hill

Lincoln
Square
Columbus
Circle

Midtown
Manhattan

Murray Hill

Turtle Bay

Rose Hill

Kips Bay

Alphabet
City

Lower
East Side

Long Island
City

Sunnyside
Blissville

Maspeth

Greenpoint

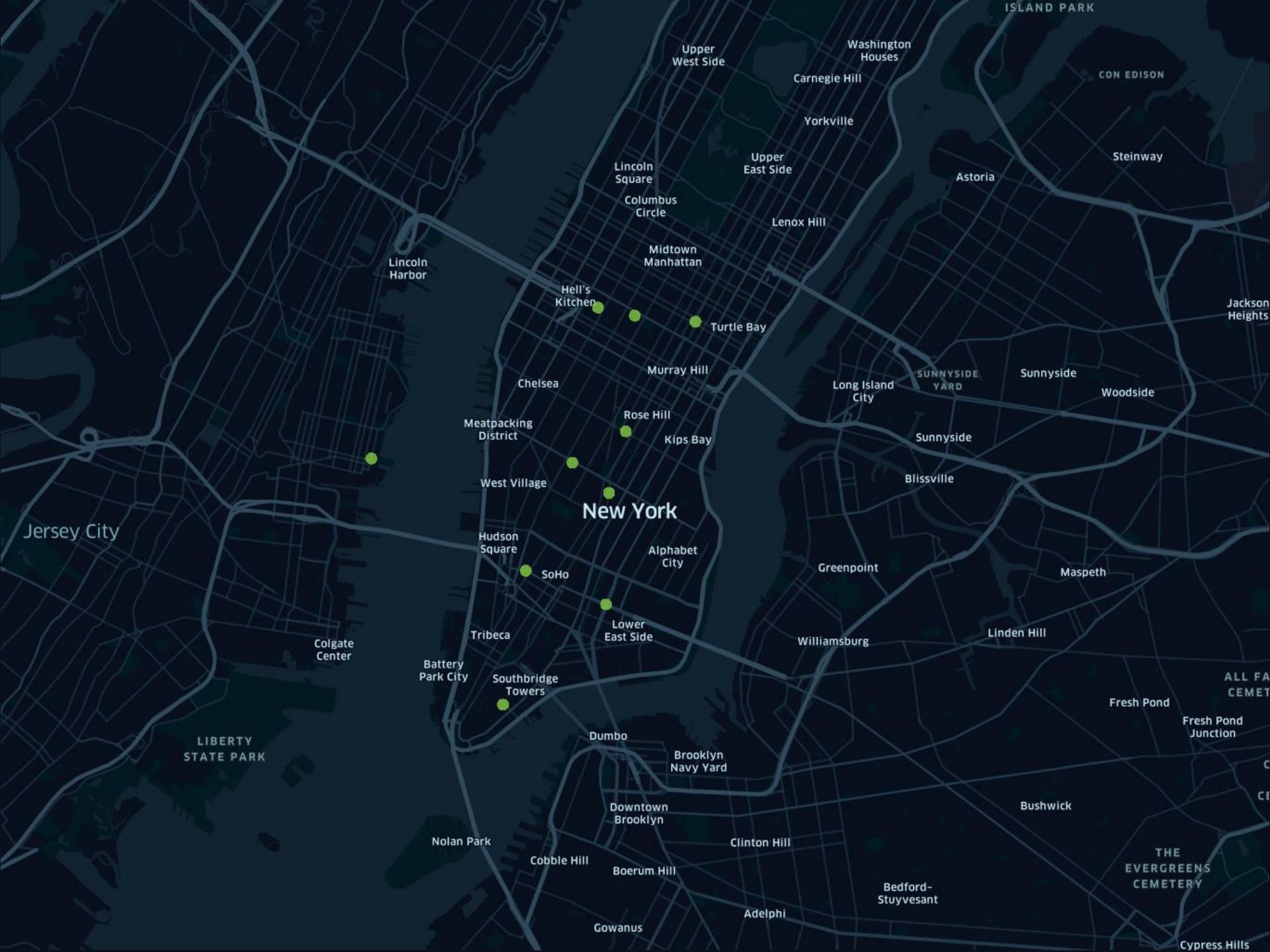
Williamsburg

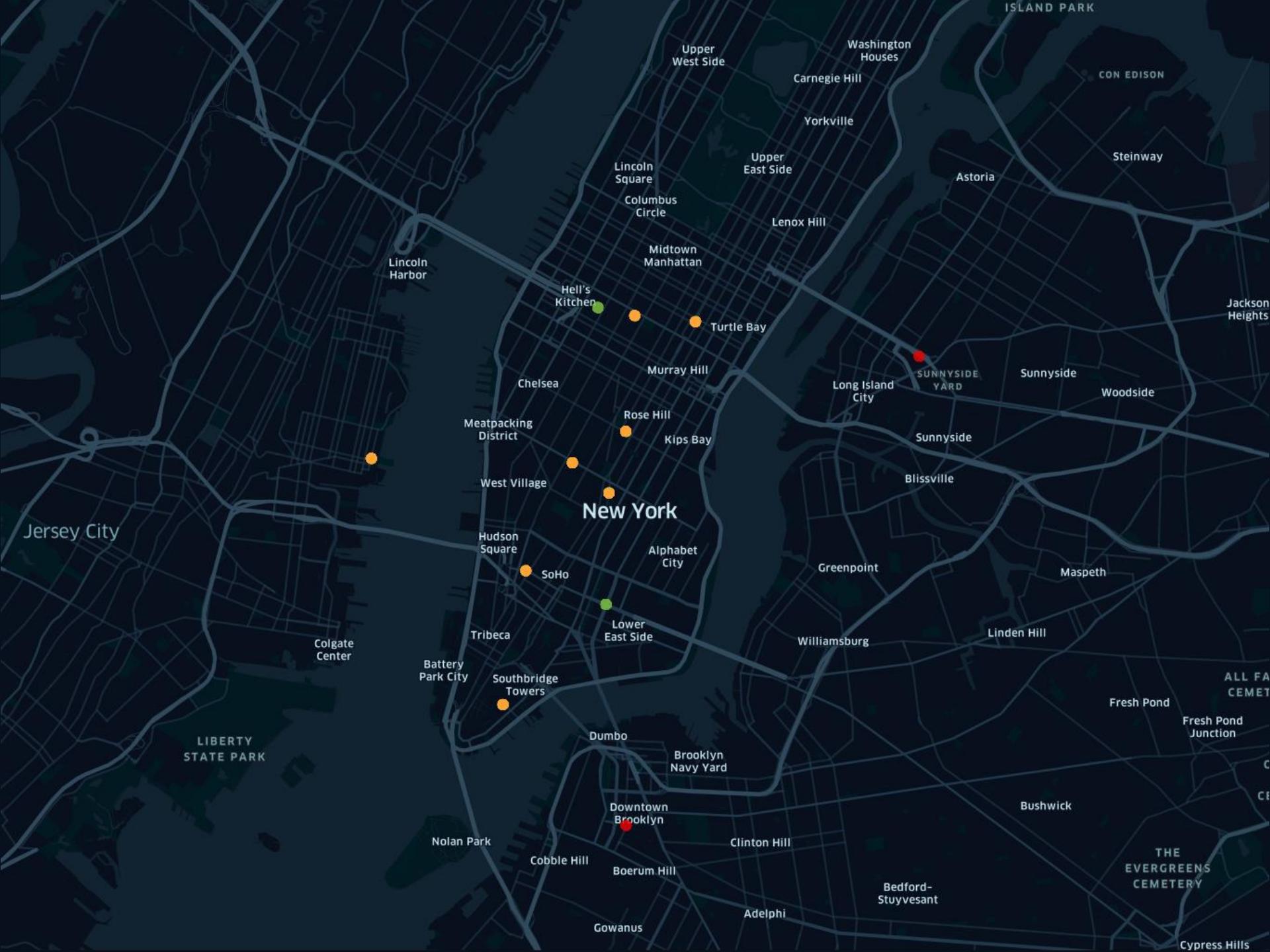
Linden Hill

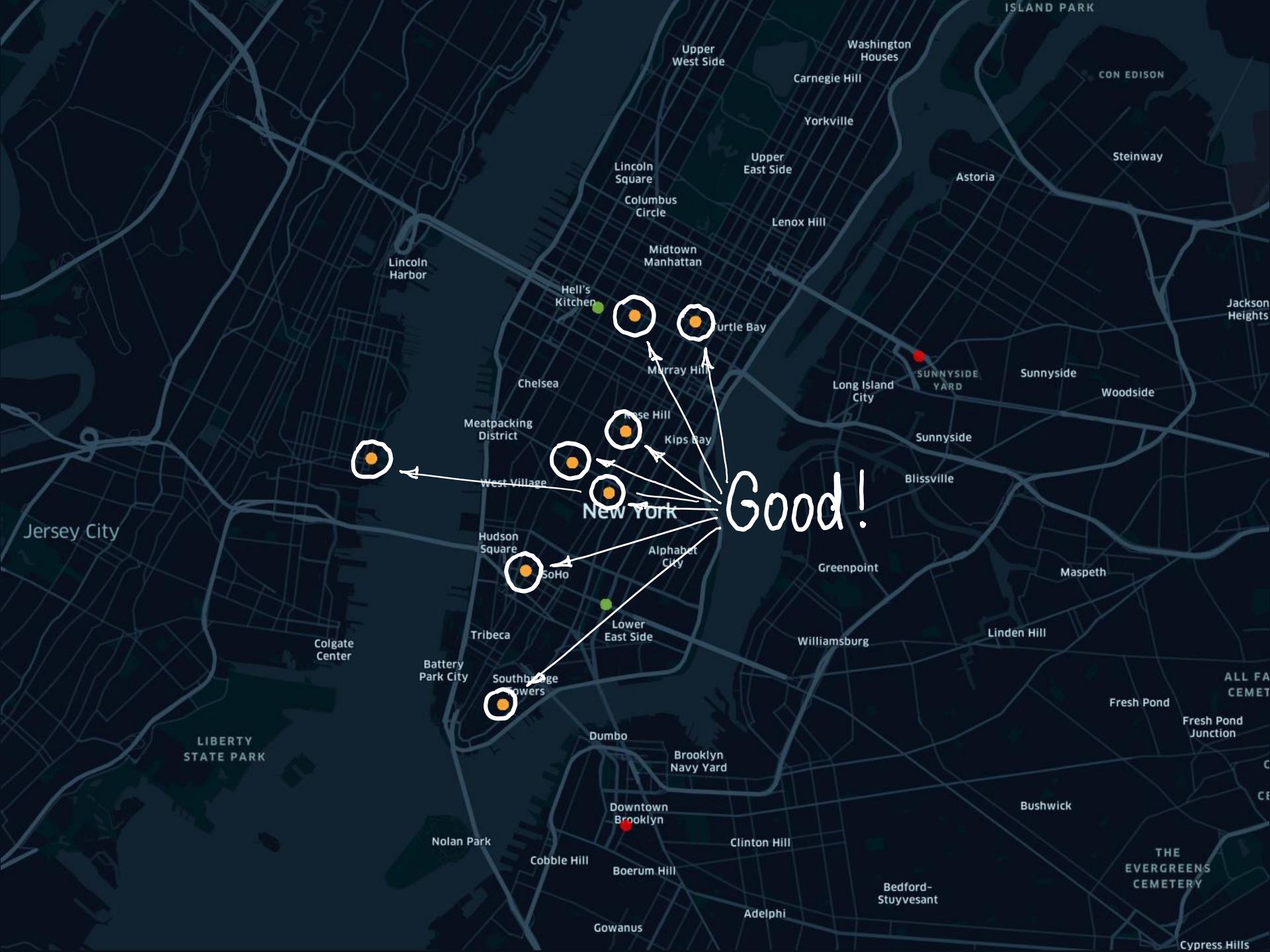
Fresh Pond
Junction

ALL FA
CEMET

THE
EVERGREENS
CEMETERY









And if you ever wonder ...



... can we do it better?

Sure!



#то_всё_граждаться

*everything can be programmed