NYU | Center for Urban Science + Progress

Network-Based Spatiotemporal Predictive Model for Car Crash Risk Assessment

Apr 26, 2024

Junyi Li, Qingyuan Feng, Jiuling Zhong, Linode Ye, Morris Gao

Background

Introduction, to the Issue

- Every year around 1.3 million lives are lost globally due to road traffic accidents making it the primary cause of death for children and young adults.

Impact on Progress

- These accidents not result in significant personal tragedies but also hinder sustainable development, particularly in low and middle income nations where a majority of these incidents take place.

Background

Advancements in Predictive Models

- Utilizing network-based spatiotemporal predictive models marks a significant advancement in urban road safety, allowing us to identify and predict potential crash hotspots more effectively.

Supporting Policy Making

- These sophisticated models provide essential insights into the interaction of demographic factors and road network characteristics, helping policymakers implement targeted safety measures more effectively.

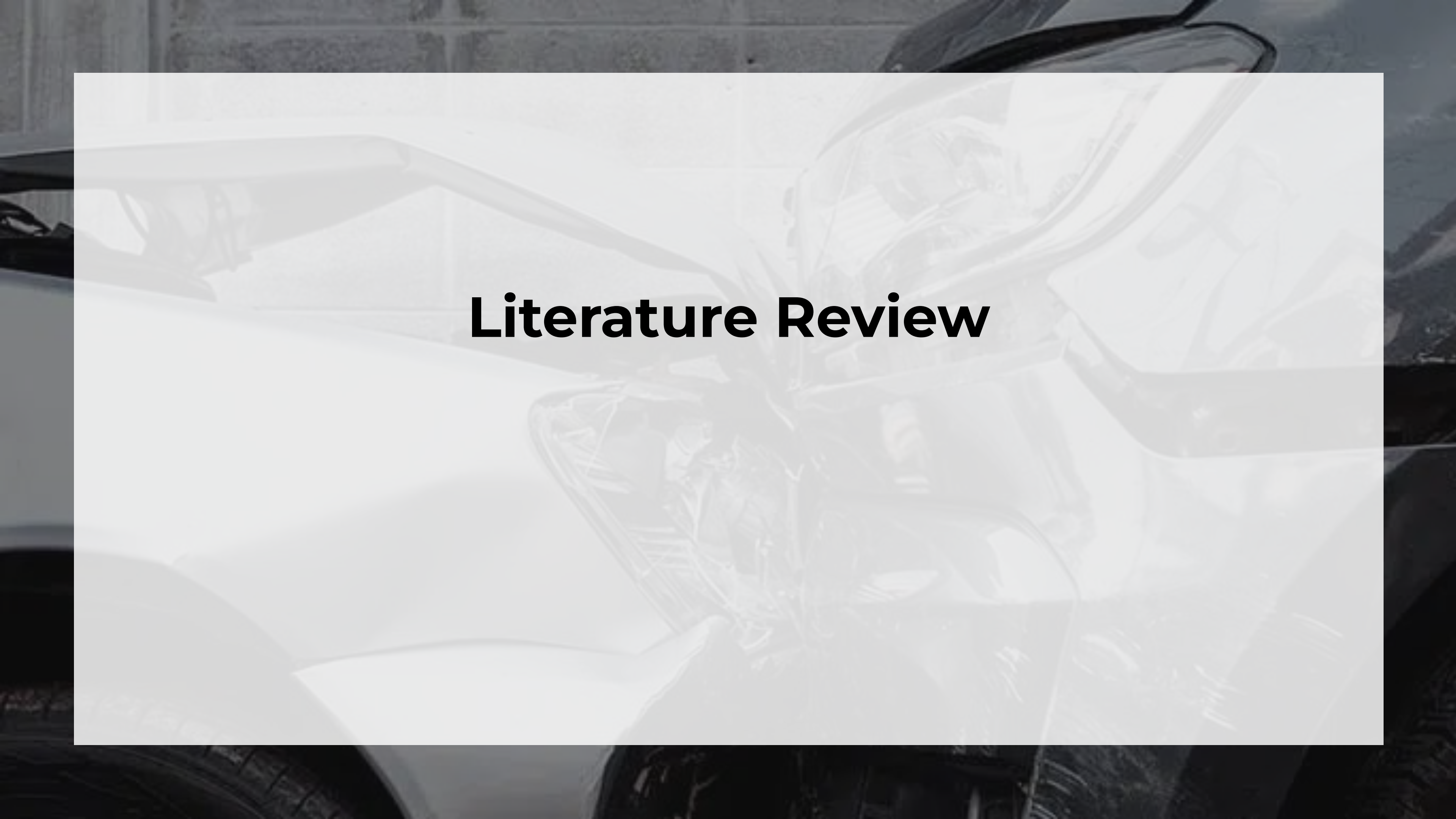
Long-term Benefits

- Continued research and improvement of these predictive models are vital. They serve as foundational tools in reducing traffic-related fatalities and injuries, ultimately leading to safer urban environments.

Problem Statement

The Study is aim to :

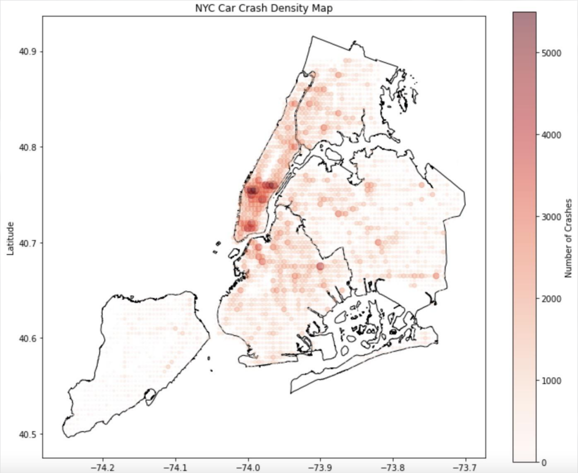
- **Explore the temporal and spatial distribution of car crashes**
- **Analyze the spatio-temporal autoregressive of traffic collisions**
- **Build a network-based predict model for Car Crash Risk**



Literature Review

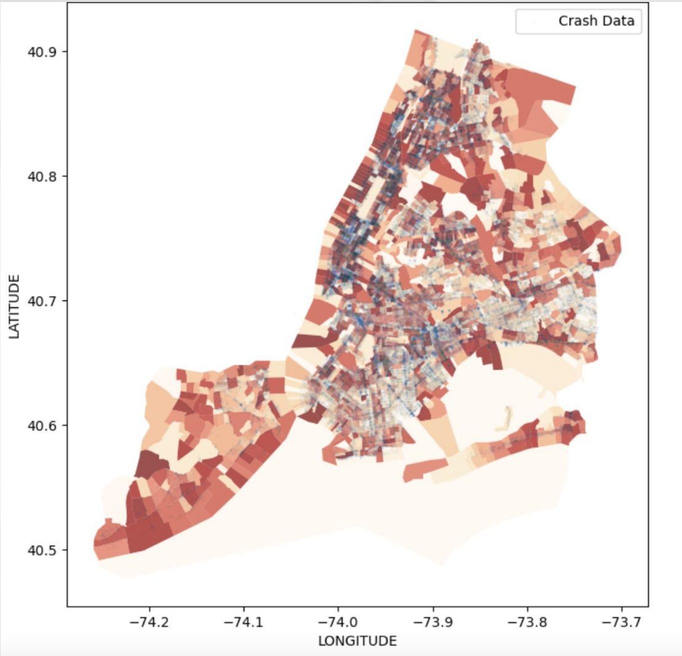
Data-Car Collision

```
collison22_23 = collison.loc[(collison['DATETIME'] > pd.Timestamp('2022-1-1 00:00'))]
collison22_23.shape
collison = collison22_23
```



NYC Open Data: Motor Vehicle Collisions - Crashes

Data-Demographic

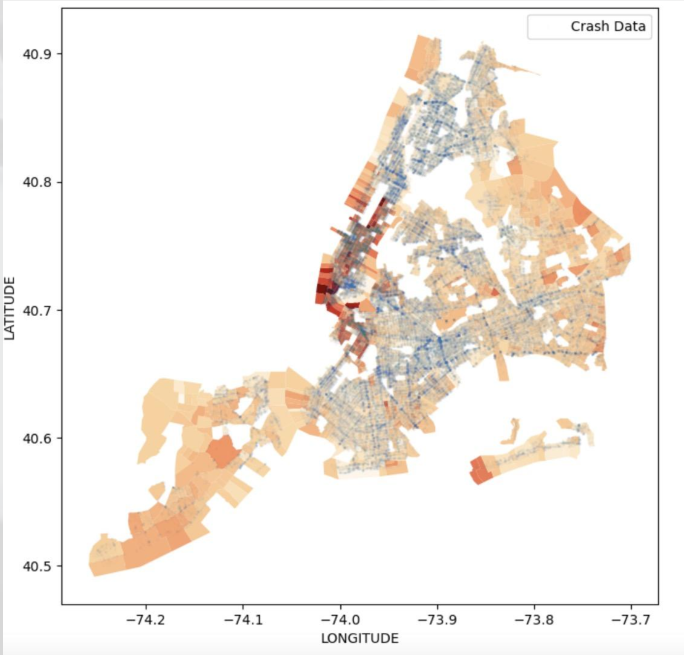


Crush Data vs Population by Census Tract
-NYC Planning

	GeoID	Pop1	geometry
0	36005000100	3,772	POLYGON ((-73.89772 40.79514, -73.89611 40.796...

https://www.nyc.gov/site/planning/planning-level/nyc-population/2020-census.page

Data-Demographic

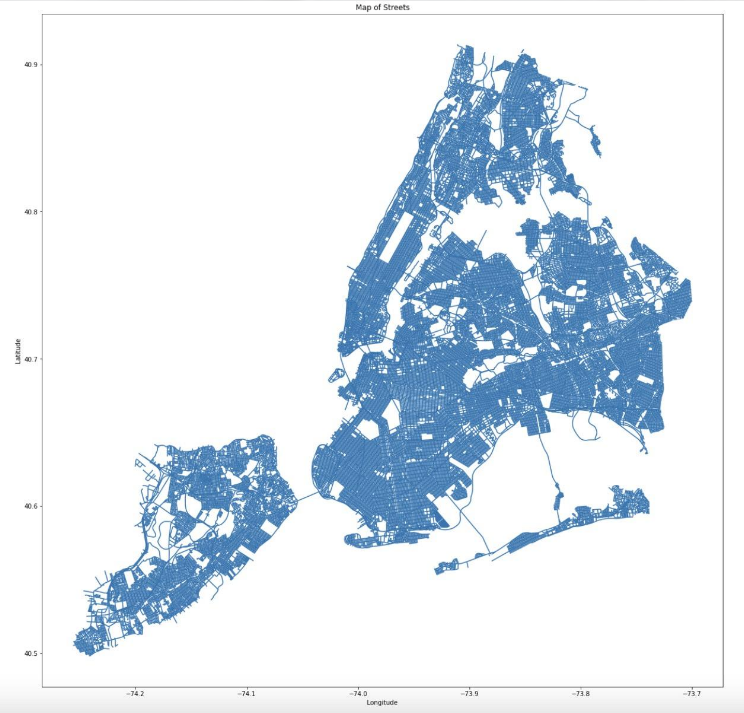


Crush Data vs Median Income by Census Tract
-**NYU Furman Center**

	short_name	long_name	Census Tract	2009-2013	2014-2018	GEOID	geometry
0	hh_inc_med_adj	Median household income (2022$)	36005041900	43887.524204	36984.666450	36005041900	POLYGON ((-73.88527 40.87923, -73.88466 40.879...

https://app.coredata.nyc/?mlb=false&ntr=&mz=16&vtl=https%3A%2F%2Fthefurmancenter.carto.com%2Fu%2Fnyufc%2Fapi%2Fv2%2Fviz%2F98d1f16e-95fd-4e52-a2b1-b7abaf634828%2Fviz.json&mln=false&mlp=true&mlat=40.709955&nty=&mb=roadmap&pf=%7B%22subsidies%22%3Atrue%7D&md=map&mlv=false&mlng=-74.009866&btl=Borough&atp=neighborhoods

Data-Street Network

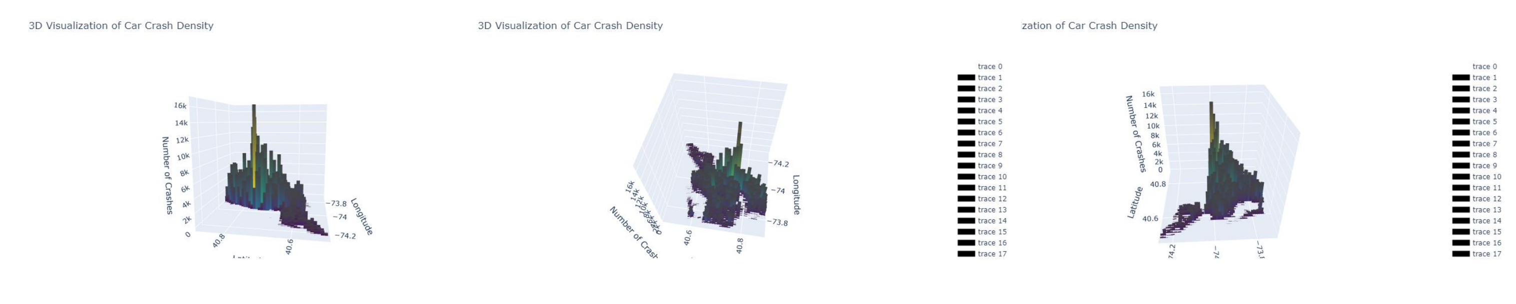


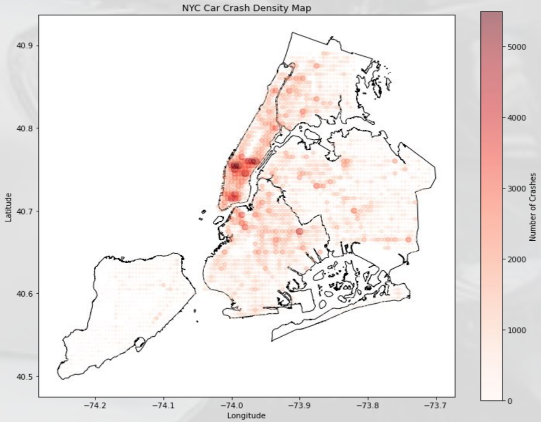
NYC Open Data

network

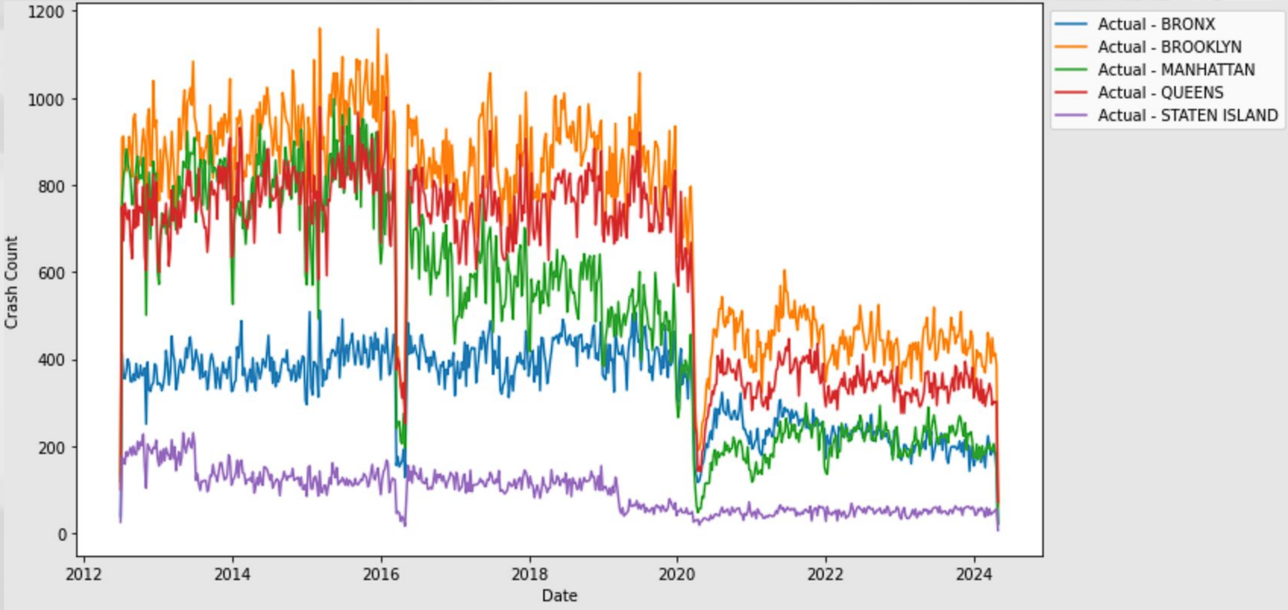
street_nm	honorarynm	old_st_nm	streetwidt	route_type	roadwaytyp	build_stat	record_st	paper_st	stair_st	cco_st	marg_wharf	edit_date	geometry
Arthur Kill Road	None	NaN	irreg>80	Mjr_st	Surface_ST	Improved	NaN	NaN	NaN	NaN	NaN	NaN	LINESTRING (-74.22128 40.55025, -74.22128 40.5...

Spatial Analysis





Time-series Analysis



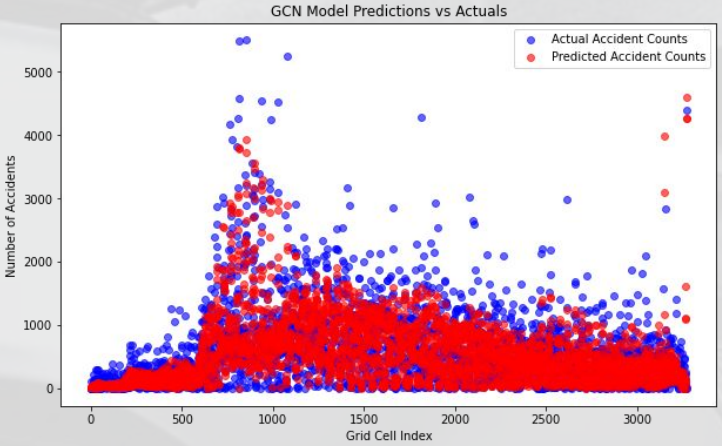
Significant drop after COVID-19

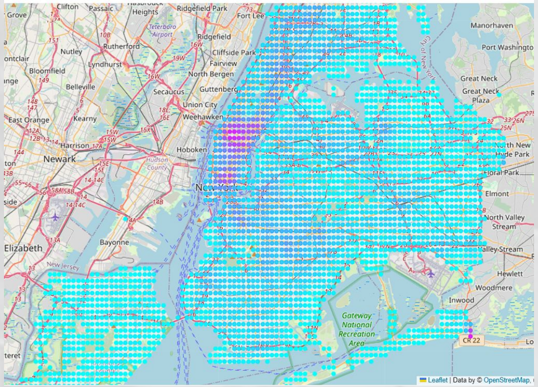
Strong seasonal pattern

Time series regression prediction is not very meaningful

Methodology

GCN (Graph Convolutional Network)





Methodology

GCN limitation:

Temporal autocorrelation (combine with LSTM or LSTnet, etc.)

Transportation network effect

Black box model

Methodology

The logistic regression function

$$\text{Logit}(p) = b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_k X_k$$

- Able to capture the many details relating to the crash occurrence;
- Have rigid assumptions about dependent with multiple categories;
- The logistic regression function is bounded by 0.0 and 1.0.

Dependent variable: Vulnerability of Crash

OF VEHICLE * 0.4 + # OF PERSONS INJURED * 0.6 + # OF PERSONS KILLED *1+ NO CRASH *0

Methodology

Variable: Spatiotemporal Index

$$Gravity[i]^r = \sum_{j \in G-\{i\}, d[i,j] \leq r}^{n} W[j]/e^{\beta * d[i,j] - plateau}$$

- W[j] is the weight of destination j, measured in hours of time interval;
- beta = 0.001 to control an exponential travel distance decay；
- Plateau radius = 0 ensure that distance decay is always included;
- For each i, the destinations j is the location where a crash happened within one month and within 3 kilometers travel distance;
- The index measures the attractive force from previous collisions around.

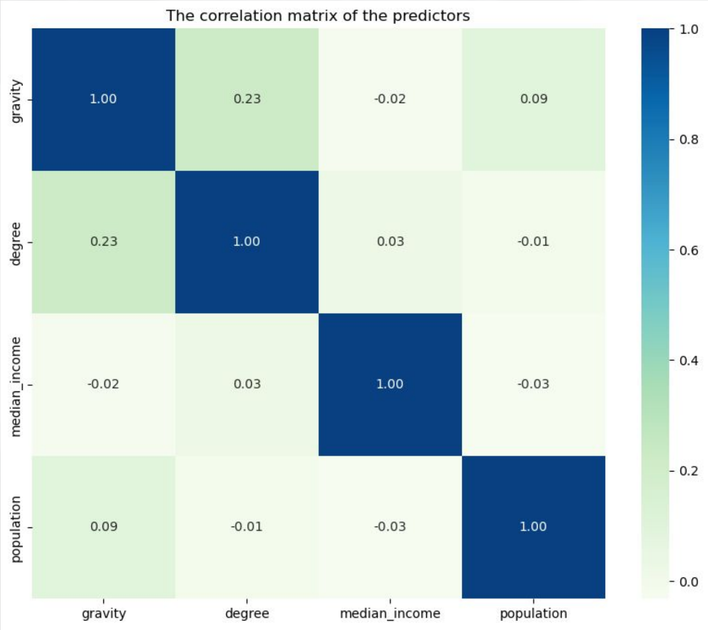
Model

R2: 0.774 In-sample MSE: 0.97 Out-of-sample MSE: 1.00

	coef	std err	t	P>|t|	[0.025	0.975]
gravity	5.241e-06	4.59e-07	11.412	0.000	4.34e-06	6.14e-06
median_income	1.149e-06	6.89e-08	16.665	0.000	1.01e-06	1.28e-06
population	4.17e-05	1.34e-06	31.012	0.000	3.91e-05	4.43e-05
degree	0.2792	0.002	158.626	0.000	0.276	0.283

Model

Correlation Matrix

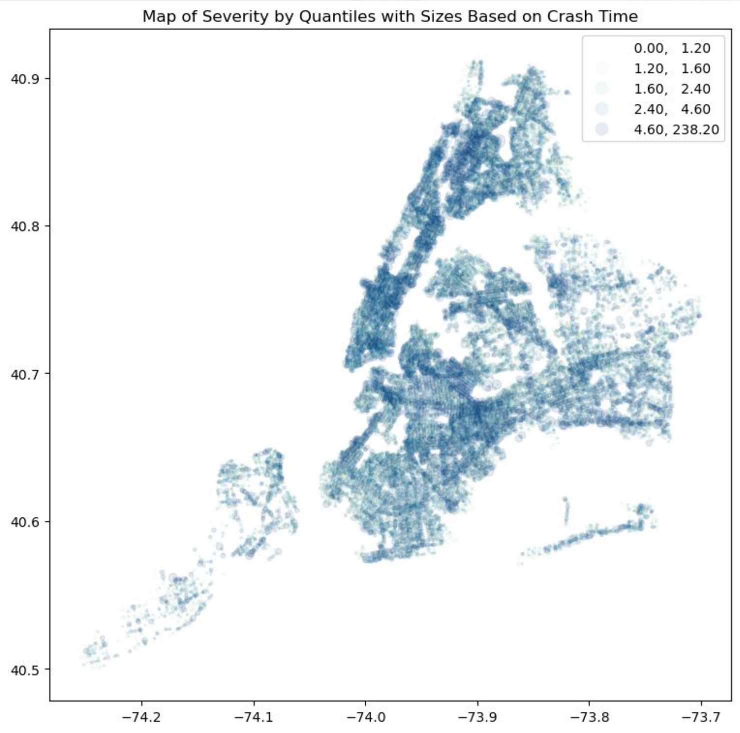


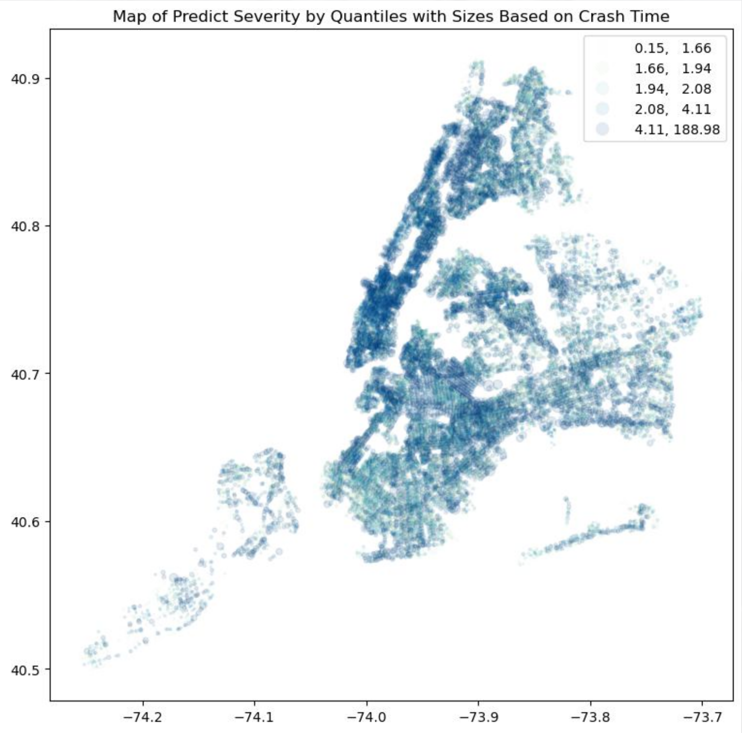
Variance Inflation Factor

Feature	VIF
Gravity	4.968909
Degree	10.182904
Median Income	3.965423
Population	4.942782

The VIF for 'degree'
is slightly above the threshold at 10.18, but its statistical significance and substantial contribution to the model justify retaining it.

Model


Map of Severity by Quantiles with Sizes Based on Crash Time
0.00, 1.20
1.20, 1.60
1.60, 2.40
2.40, 4.60
4.60, 238.20
40.9
40.8
40.7
40.6
40.5
−74.2
−74.1
−74.0
−73.9
−73.8
−73.7


Map of Predict Severity by Quantiles with Sizes Based on Crash Time
0.15, 1.66
1.66, 1.94
1.94, 2.08
2.08, 4.11
4.11, 188.98
40.9
40.8
40.7
40.6
40.5
−74.2
−74.1
−74.0
−73.9
−73.8
−73.7

Conclusion

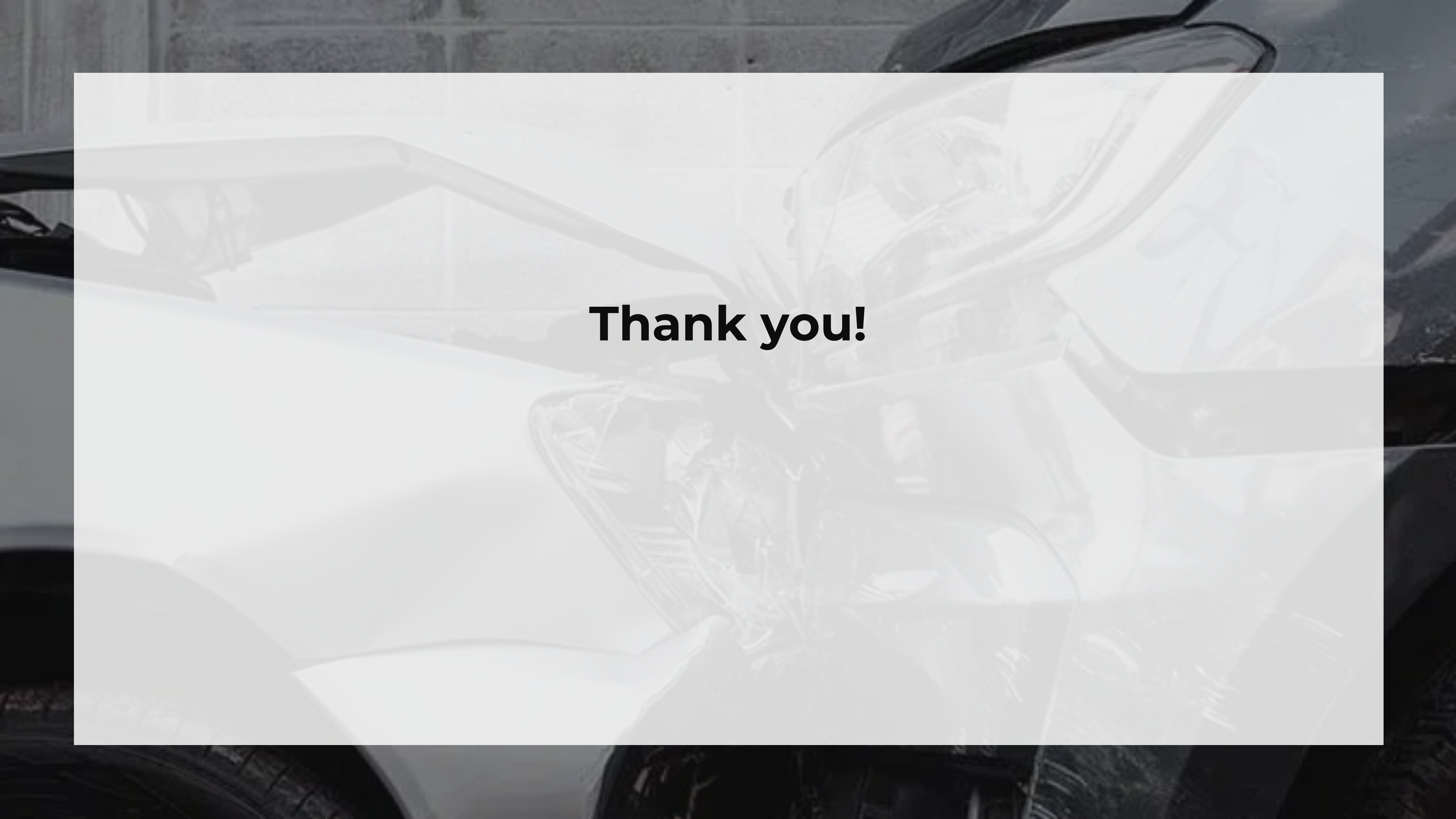
- **Downtown Manhattan has the highest car crash density**
- **After Covid-19 there has a significant drop of car crash occurrence, this pattern continue to nowadays**
- **Build a network-based predict model for Car Crash Risk with 0.774 R2**

Limitations

- Calculate of Gravity Index is highly time costs even with Dask
- Did not take other important network factors like betweenness centrality into account due to the limit computility

Recommendation

- Use betweenness centrality instead of degree to capture the structure of street network.
- Combine the idea of gravity index into the GNN structure to get a better model.



Thank you!