From this, we estimate that 6.9% of residential and open space land uses show a negative correlation with crash frequency. Through spatial analysis, potential risk hotspots including roads linking bridges and tunnels, and avenues leading with potential activity, are highlighted. The study also identified negative local spatial correlations in crash frequencies, suggesting significant safety variations within a relatively short distance. By mapping the differences between observed and predicted crash frequencies, we identified specific grid areas with unexpectedly high or low crash frequencies.

These findings highlight the crucial role of near-miss data in urban traffic safety policy and planning, particularly relevant with the imminent rise of autonomous and connected vehicles. By integrating near-miss data into safety estimations, we can develop a more comprehensive understanding of traffic safety and thus, more effectively address urban traffic risks.

### CONCLUSIONS
- Near misses have the highest correlation with crash frequency among all the variables.
- Other variables such as the number of intersections, number of bus stops, road length, residential land use rate, and open space land use rate significantly influence crash frequency in New York City.
- The near-miss-to-crash ratio, estimated to be 395:1 for the study urban area, can serve as a potential benchmark for other urban environments. The analysis of network and facility-related variables revealed that a higher number of intersections and bus stops, and longer road length, all contribute to an increased crash frequency.

### RESULTS AND DISCUSSION

**Spatial Distribution**
- First, grids encompassing large roads of bridges and commercial establishments are highlighted, due to the complexity of the road network at those locations, with more frequent merging and diverging points leading to more frequent lane changes by drivers.
- Second, high-risk grids predominantly face the eastern and southeastern parts of the city. These areas are characterized by high pedestrian activity due to the concentration of commercial establishments, public transit access points, and other urban amenities.

**Spatial Autocorrelation**
- Moran's I test is used to statistically analyze the crash count among those correlation coefficients for crash count. It identifies the spatial patterns of crash frequency in the study area. The road networks within these grids are often complex and multi-lane roads.

**Near-miss to crash ratio estimation**
- The inherent effect of near misses, such as vehicle or pedestrian crashes, is 0.8% on average. This implies that in our study area, near misses to crashes is approximately 135:1.
- To further refine the trend near misses to crash rate, we leverage two statistics: 10% of our observations had a high number of reported crashes, and Mobileye holds a 99% market share in ADS (ADAS).
- From this, we estimate that 9% of vehicles in New York City are equipped with Mobileye's Connected devices. This, in turn, allows us to calculate the near-miss-to-crash ratio in NYC as roughly 395:1.

**Near misses**
- The crash prediction model utilizes the previously selected variables to construct 2 models: the k-Nearest Neighbor (k-NN) and the Random Forest model.

**DATA PREPARATION**

### Methodology

**Grid-based method**
- Divides a geographical area into a grid of uniformly divisible by 300 ft.

**Empirical Bayes method**
- The empirical Bayes (EB) approach divides a world data and the expected crash frequency by Safety Performance Functions (SPFs).

**Empirical Bayes model**
- The empirical Bayes (EB) approach divides the observed crash frequency from the historical motor vehicle crashes of NYC into grid-based EB estimates.

**Spatial Analysis**
- The empirical Bayes (EB) model is utilized to analyze the spatial variation of the estimated crash frequency.

**RESULTS AND DISCUSSION**

**Near-miss to crash ratio estimation**
- The inherent effect of near misses, such as vehicle or pedestrian crashes, is 0.8% on average. This implies that in our study area, near misses to crashes is approximately 135:1.
- To further refine the trend near misses to crash rate, we leverage two statistics: 10% of our observations had a high number of reported crashes, and Mobileye holds a 99% market share in ADS (ADAS).
- From this, we estimate that 9% of vehicles in New York City are equipped with Mobileye's Connected devices. This, in turn, allows us to calculate the near-miss-to-crash ratio in NYC as roughly 395:1.

**Near misses**
- The crash prediction model utilizes the previously selected variables to construct 2 models: the k-Nearest Neighbor (k-NN) and the Random Forest model.