

Citi-Bike Data Challenge

Jonathan Lennartz

July 1, 2025



IS THERE AN OPPORTUNITY?

If we reduce claims we

1. reduce temporary cost coverage, legal fees, administrative costs, etc and
2. get publicity and improve brand reputation

Scale of the problem:

- Roughly 5,000 crashes involving bikes and cars annually
- ~40% of these are likely Citi Bike users
→ 2000 cases per year

HOW CAN WE DO SOMETHING?

Nudging is promising:

- Effective: Up to 30% crash reduction demonstrated
- Versatile delivery channels:
 - Within Citi Bike app
 - AXA mobile app
 - 3rd party wearables
- Multiple intervention types:
 - Push notifications
 - Incentive programs
 - Route suggestions

Key considerations:

Nudging fatigue: Keep interventions rare to maintain effectiveness

Neighborhood discrimination: Avoid reinforcing existing spatial inequalities

Usage disincentive: Don't discourage overall bike usage

Solution: Precise identification of dangerous *area × time* combinations

DATA DRIVEN CITI BIKE OPTIMIZATION

Extensive data-driven optimization across multiple domains:

Operations: Rebalancing optimization, station relocation, dynamic corrals

User Experience: Interface redesign, app improvements, kiosk enhancements

Business Models: Bike Angels program, discounted memberships, incentive systems

Policy & Equity: Expansion to underserved areas, public accountability metrics

Why no data + crash data for safety interventions?

→ It's complicated ...

DATA IS COMPLICATED: SPACE AND TIME MATTERS

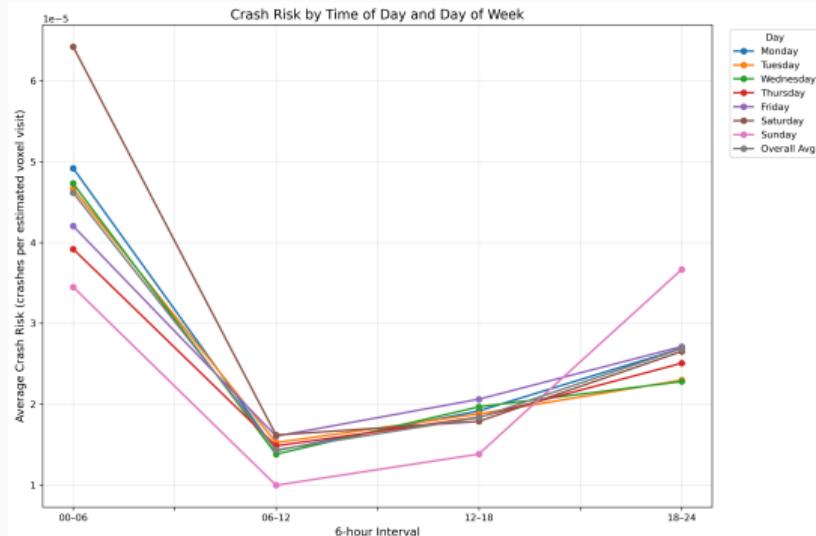


Figure 1: Risk varies by day and time

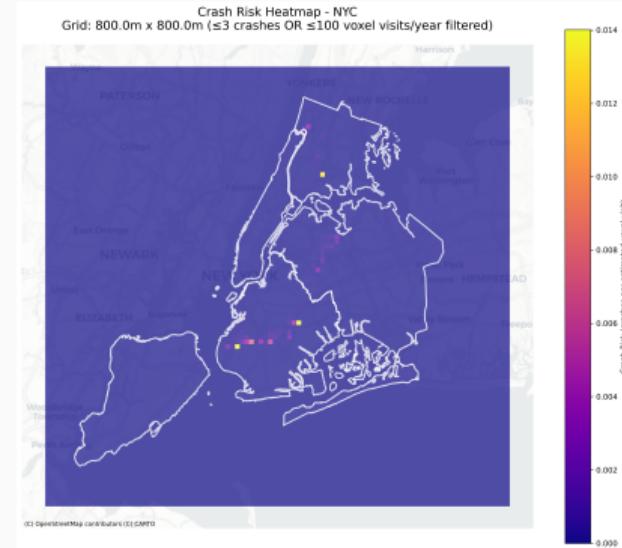


Figure 2: Risk varies by location

High spatial and temporal variability in crash risk

COMPLEX INTERACTING EFFECTS

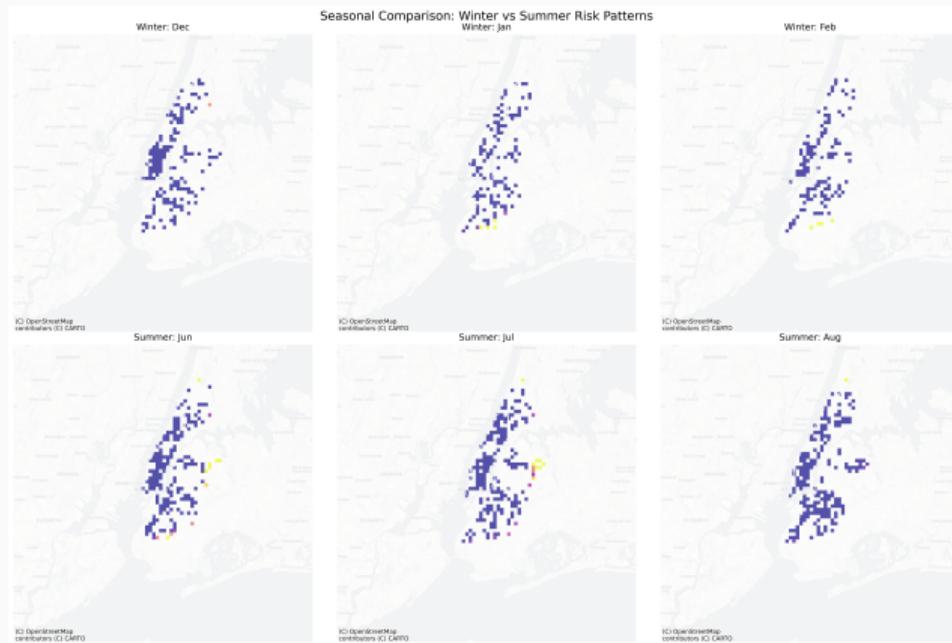


Figure 3: Seasonal risk patterns differ spatially

Problem: Despite rough rasterization

- spatial clusters are not homogeneous and inconsistent across time and
- crash cases per voxel approach outlier territory.

CURRENT BEST SOLUTION: SIMPLE, DIGESTIBLE RULES

Honest assessment: With current data alone, we cannot reliably identify specific danger zones. Instead, we should focus on easy rule.

Time-based warnings:

Be extra careful during night hours and weekends

Peak risk periods: Friday-Sunday evenings

Location-based guidance:

Exercise caution in areas with low bike traffic

Increased vigilance near highways and major intersections

Avoid isolated routes during off-peak hours

Future direction: Predict dangerous areas ahead of time by integrating additional data sources (weather, traffic, events) and conduct careful re-analysis.

IMPLEMENTATION DETAILS

For full details on data processing and analysis methods, please consult the documented source code:

`src/apps/rasterize_data.py`

`src/utils.py`

`src/jupyter/vis_results.ipynb`