Trip Duration Prediction and CO_2 Optimization

A Data Science Approach to Urban Sustainability

1. Motivation & Problem

NYC taxis emit 450,000 tons CO_2 annually

Inefficient routes result in 15–25x more emissions per mile

Green Rides Initiative at 100% zero-emission by 2030

2. Research Questions

1. Can ML models accurately predict trip durations?
2. Which features best explain variability (time, weather, geography?
3. How much CO_2 can optimized routing save?
4. Can we operationalize via a real-time dashboard?

3. Data Sources

NYC Taxi Trip Data

1.4M records

Weather & Holidays

OSRM Routing

0.15 kg/km

4. Methodology

Feature Engineering

- time cycles
- distances (Haversine/Manhattan)
- weather
- clustering

CO_2 Framework

- 0.15 kg/km
- route reduction scenarios

Final RMSLE

stacked ensemble

.276

6. Ethical Considerations

- ✓ **Data privacy** no personal tracking
- ✓ **Algorithmic fairness** across boroughs
- ✓ **Advisory** optimization, not enforced

7. Future Directions

- Integrate real-time traffic APIs
- Pilot test with taxi flests
- Expansion to ride-sharing and last-mile delivery
- API integration for smart cities