Chicago Traffic

Josiah Chung, Marc Edwards, Puneet Modi, Ankita Nambiar

Agenda

1. Business Problem Overview
2. Business Solution
3. Why Our Solution Works
4. Modeling
5. Assumptions and Limitations
6. Next Steps

Business Problem

- The 2022 Global Traffic Scorecard recently named Chicago as the worst traffic city in the US and the second worst in the world.
- Millions of Chicago drivers waste an estimated 155 hours in traffic every year.

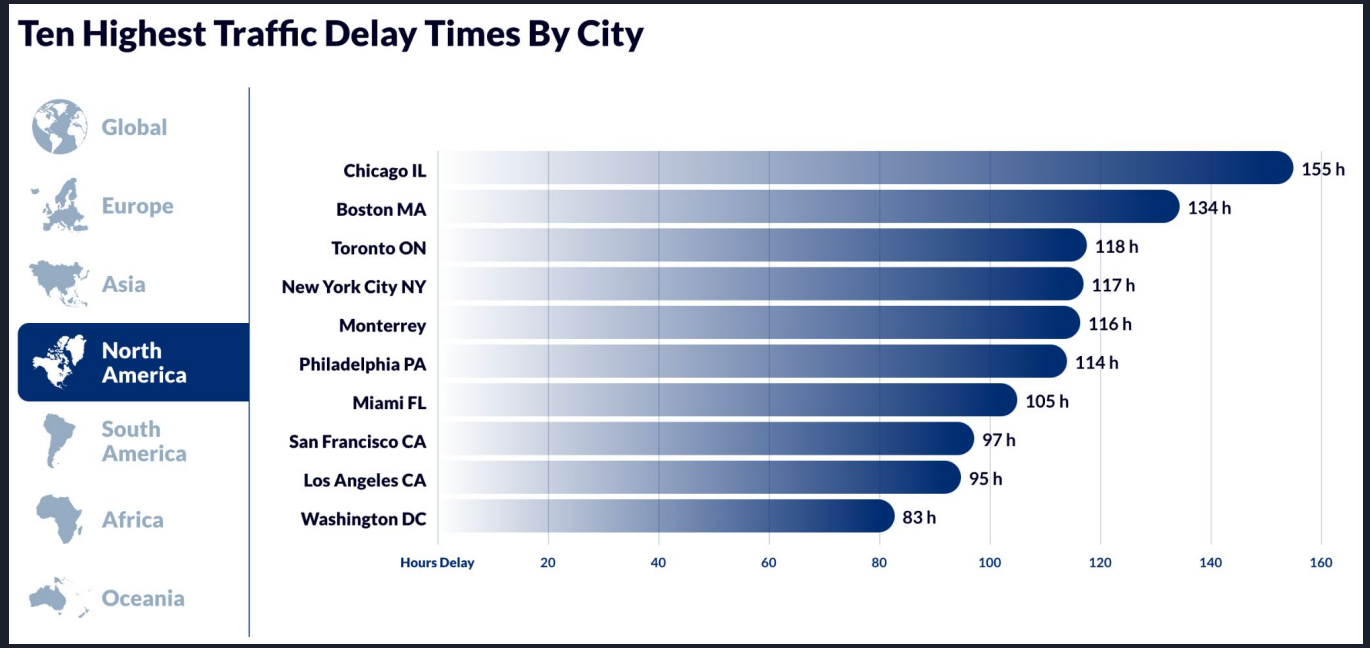
#1

US

#2

World

Business Problem



Business Solution

- **Objective**:

  The City of Chicago's objective is to decrease the time the average commuter spends in traffic. By uncovering factors that contribute the most to congestion and traffic, we can find a workaround to utilize the City's less congested and faster-moving roads. Thus, the average time spent in traffic can be reduced.

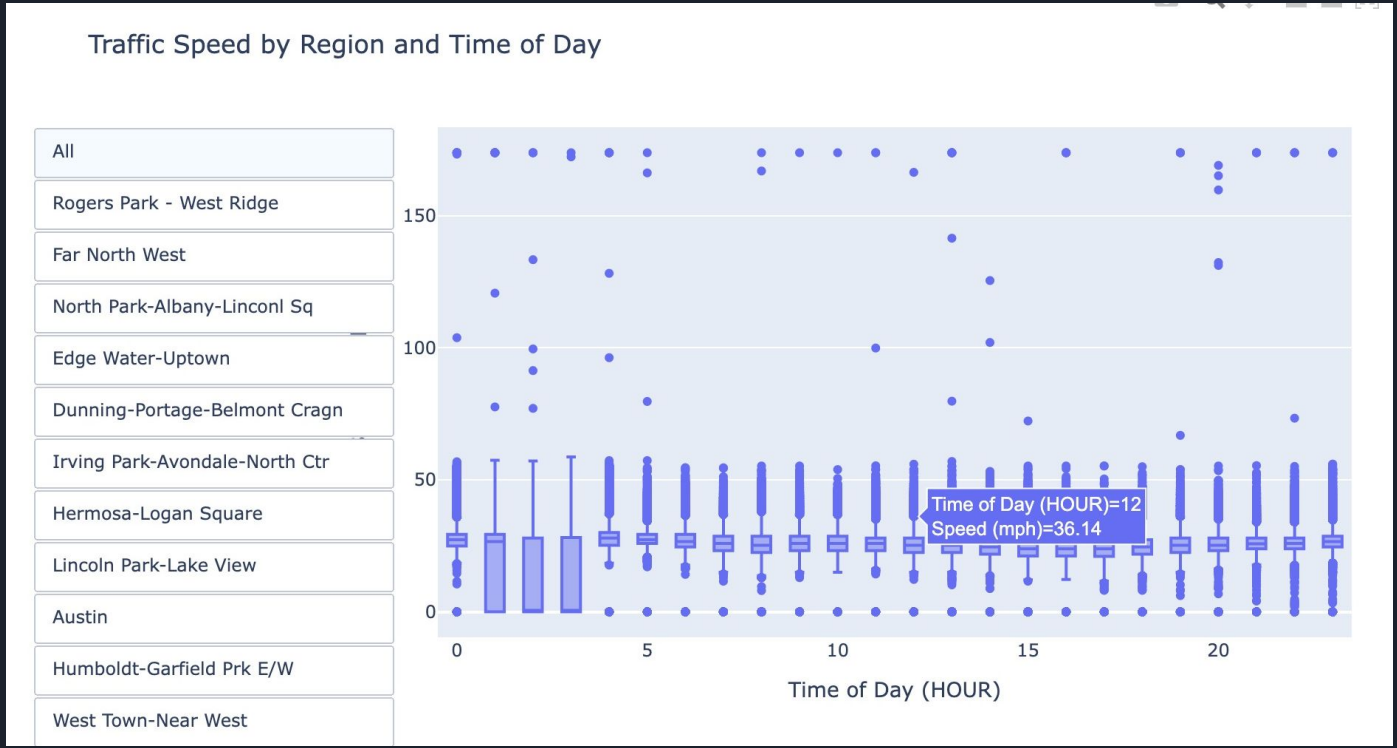
- **Potential Factors that Affect Time Spent in Traffic:**
  - number of vehicles on the road
  - the day and time of travel
  - the direction of travel, and much more

Why our Solution Works

- Data Source is Reliable
  - From the City of Chicago Data Portal

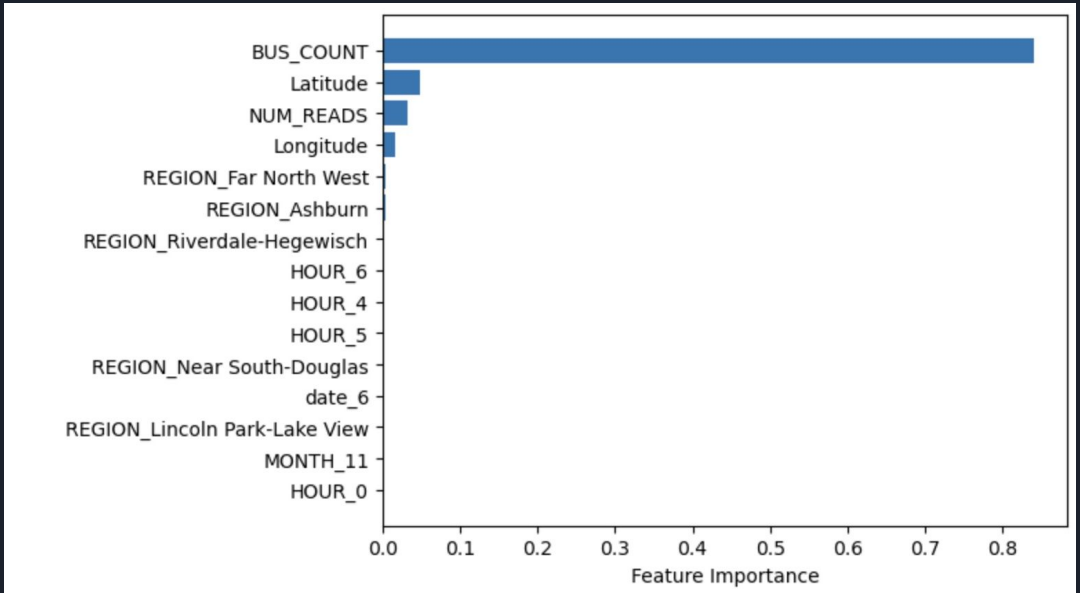
- Clear predictors that point to logical next steps to take
  - time, day, latitude, longitude, bus count, etc.

Speed vs Location and Time of Day



BUS_COUNT is Our Most Important Feature

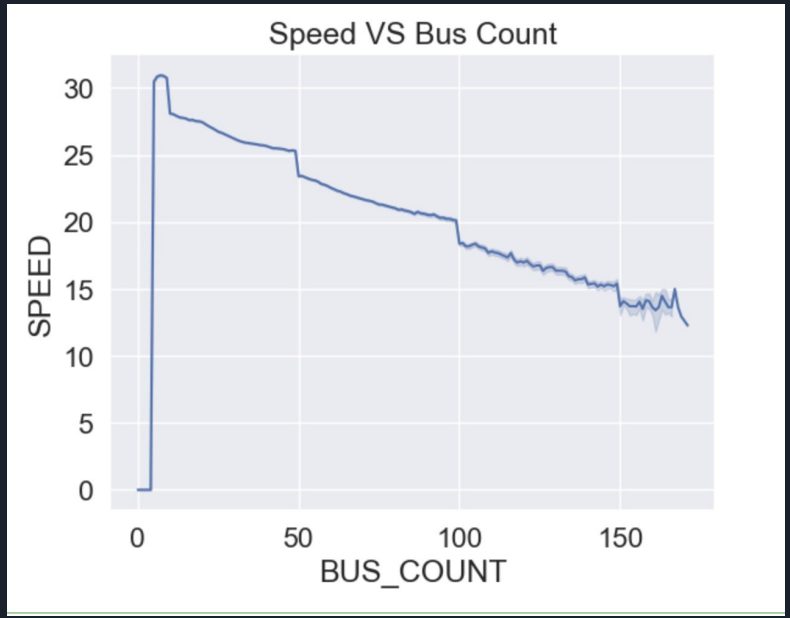
We ran random forest regression on our data and visualized the most important features that affect our target variable, Speed. BUS_COUNT is the most important feature with the score of ~ 0.85.



Since BUS_COUNT is a major predictor of speed, we must make sure to include it in our final dataset.

Evaluating Speed vs BUS_COUNT

- The average speed of all vehicles decreases considerably as the number of buses on the roads of Chicago increases.
- This shows an inverse relationship between Speed and Bus Count.
- As bus count approaches a large number, speed converges to about 15 mph.



Modeling

MODEL	RMSE
Linear Regression	0.0389
Decision Tree	0.0362

The low RMSE values means we can be more confident that the features we included in our model were useful in predicting our target variable, Speed.

The explanatory variables included: Bus Count, Vehicle Sensor Count (Num Reads), Latitude, Longitude, Hour, Month, and more.

What Worked and Didn’t Work

Worked:

- Ranking feature importance helped us parse through many categorical variables.
- We did not make any transformations to BUS_COUNT even though the distribution was skewed, and it turned out to be important to our evaluations.

Didn’t Work

- We tried adding minutes as a predictor of speed in our analyses, but we found that using minutes was unnecessary.
- We wanted to keep Zip Codes in our data for analysis, but we found that the zip codes provided in the data were erroneous.

Assumptions and Limitations

Assumptions

- We assumed that time of day will be one of the greatest predictors of traffic
  - Specifically, we assume that traffic will significantly increase during the mornings and afternoons, corresponding to rush hour traffic.
- We assumed that our dataset was representative of Chicago traffic; the data covers traffic estimates from 2022.
- We are measuring traffic congestion based on the number of GPS probes received from cars and the number of buses. We are assuming that these will provide accurate estimates of traffic.

Limitations

- Our devices could not handle more data than from just 2022, as 2022 alone came with over one million rows.

Next Steps

- The City of Chicago can create public service announcements (radio ads, TV ads, traffic alerts, etc.) to inform commuters of newly uncovered areas that have higher average speeds of travel and less congestion throughout the day, in particular the morning and evening rush hour.
- Provide recommendations to the City of Chicago for areas that may benefit from adding more street lanes or prioritizing streets to undergo repair. Our solution would be able to optimize travel by saving time and improving the travel experience.
- Since BUS_COUNT is a major predictor of speed and therefore congestion, it would be effective to prioritize properly managing bus scheduling and routing when trying to reduce areas of congestion.

THANK YOU