

BIG DATA FINAL PROJECT

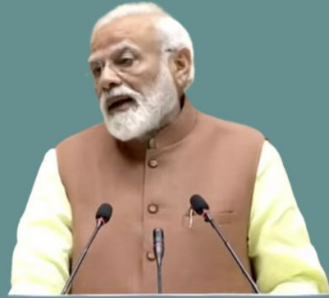
Team 5

Spatial, Temporal, and Weather Analysis of Accidents in Greater Nashville

Mark Raj, Daniel Zhang, Fateen Anam Rafid



Team 5



BIG DATA, Machine Learning, Artificial Intelligence

Mark Raj, Daniel Zhang, Fateen Anam Rafid

**Spatial, Temporal, and Meteorological Analysis
of Accidents in Greater Nashville**



Data & Methods

1. Spatial Analysis

- a. Geopandas to create spatial joins and create visualizations with plotly
- b. Used linear regression to see demographic correlations with response time/accidents

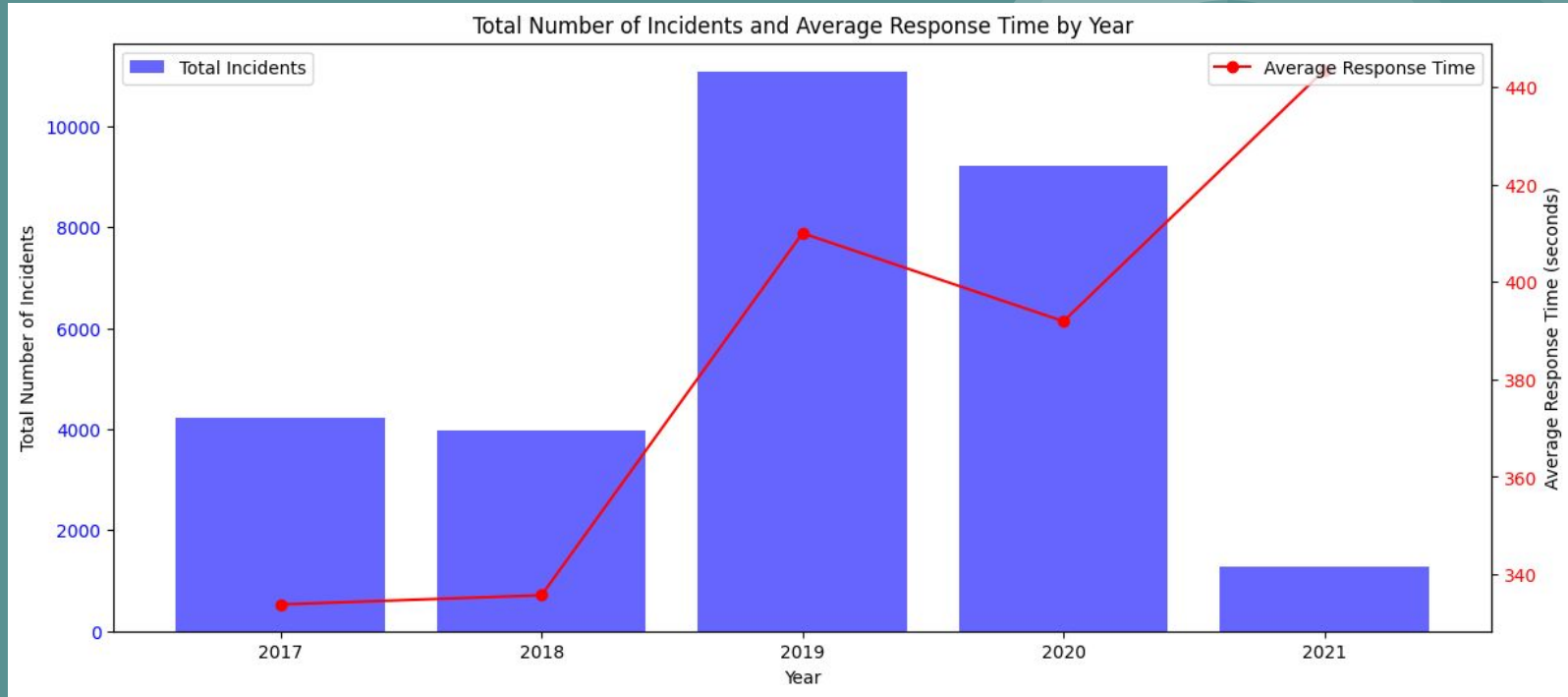
2. Temporal Analysis

- a. Conducted with Pandas and PlotLib to see year/month/time-of-day breakdowns

3. Meteorological Analysis

- a. Joined minute-by-minute weather data to accident data using Spark/EMR, filtering for the closest weather stations
- b. 2 machine learning models to predict response time based on weather factors

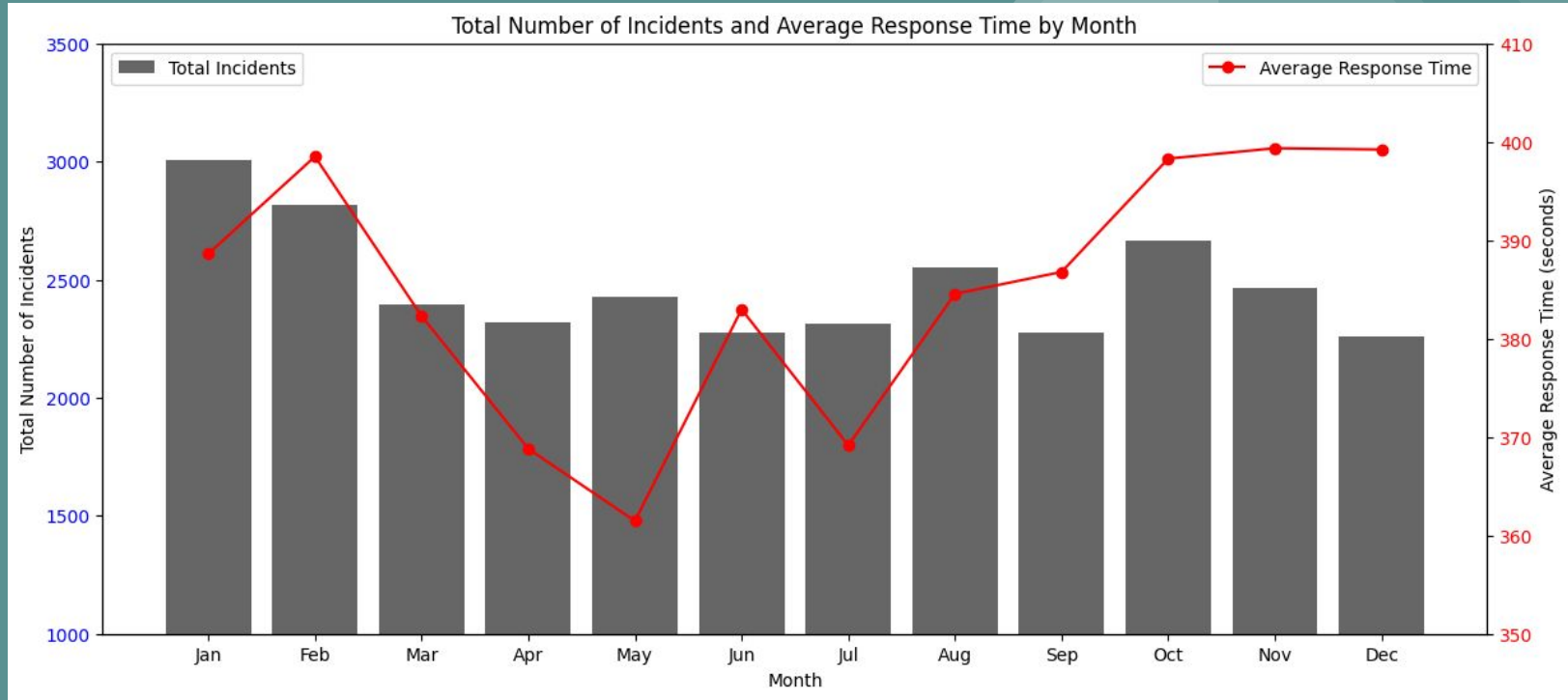
Temporal Analysis of Incidents



Key Takeaway:

- Rise of Total Accidents over the Years.
- Correlation can be seen with response time

Temporal Analysis of Incidents

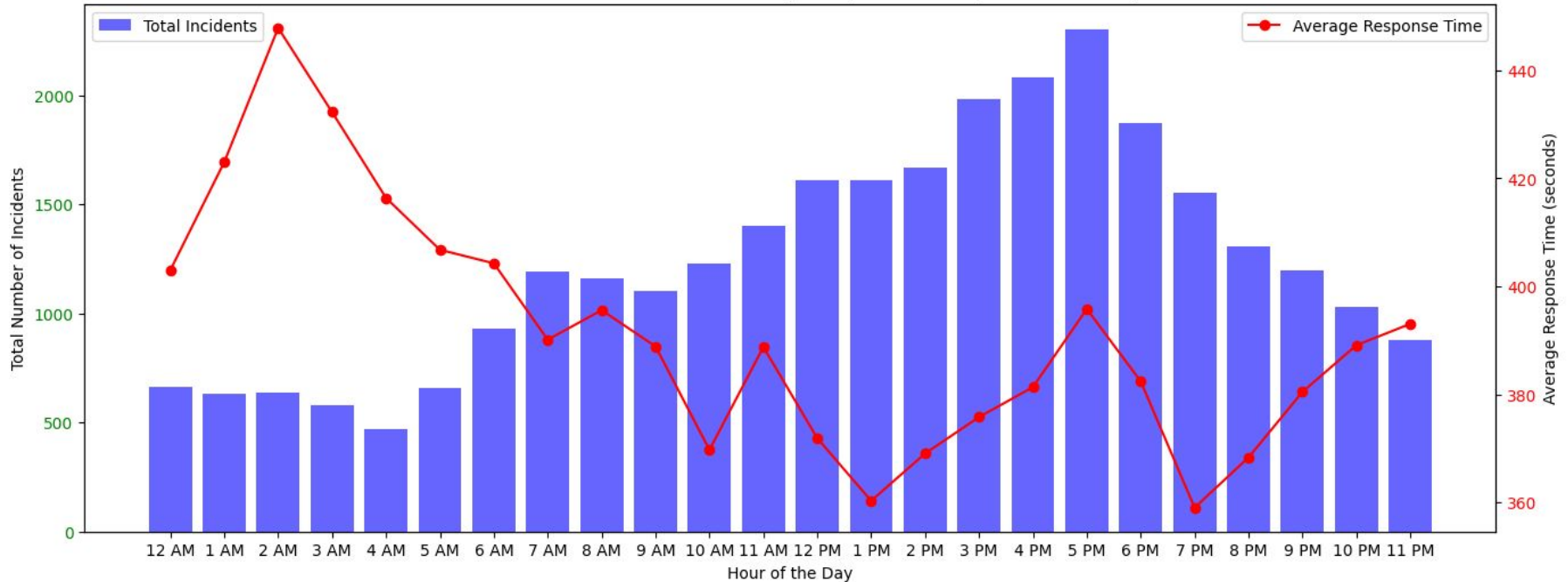


Key Takeaway:

- More Accidents happen during winter
- Correlation can be seen between total incidents response time

Temporal Analysis of Incidents

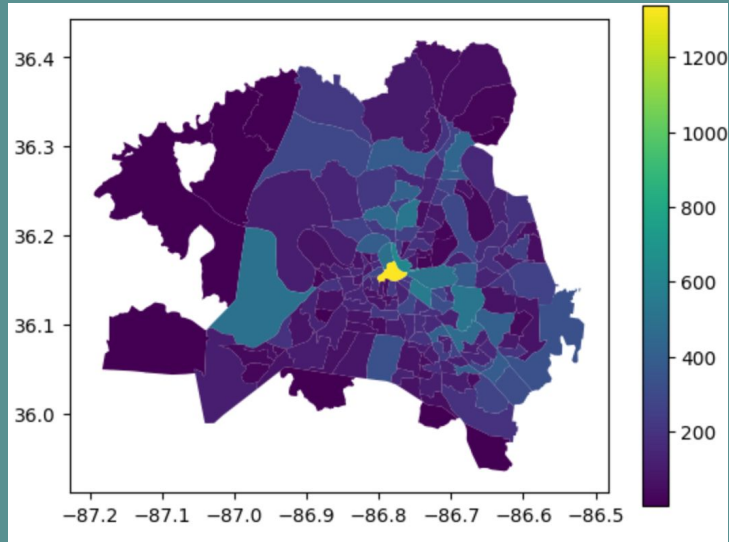
Total Number of Incidents and Average Response Time by Hour of the Day



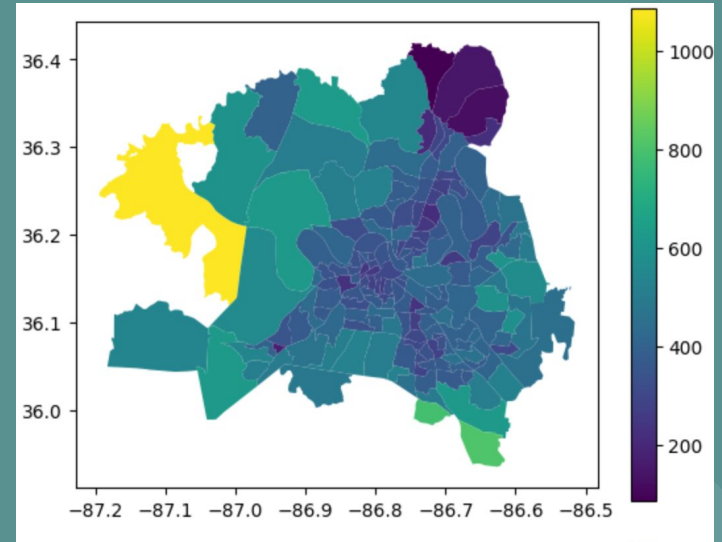
Key Takeaway:

- More accidents happen during the day
- Response time is higher after midnight. Inverse correlation.

Spatial Analysis of Incidents

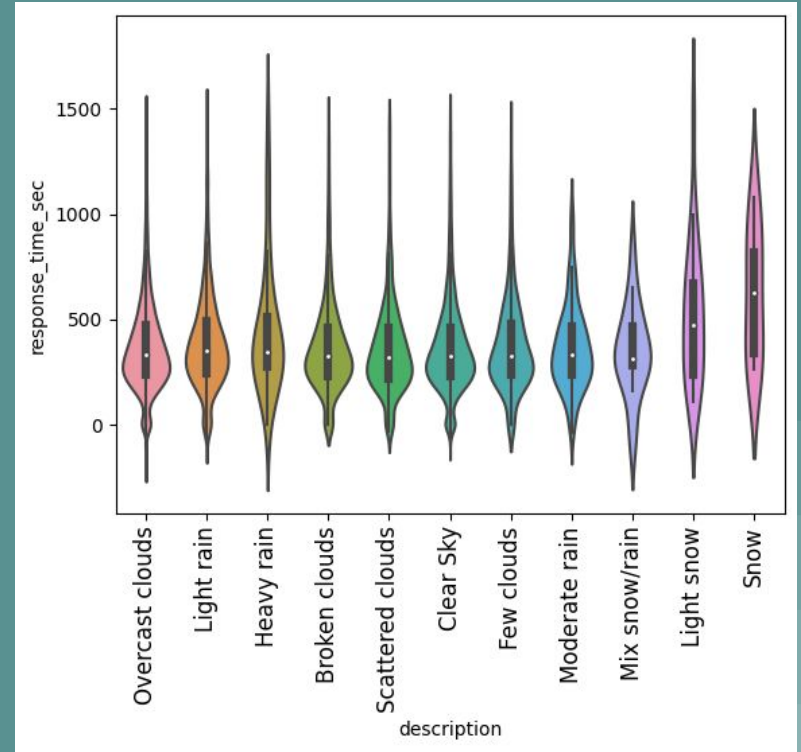
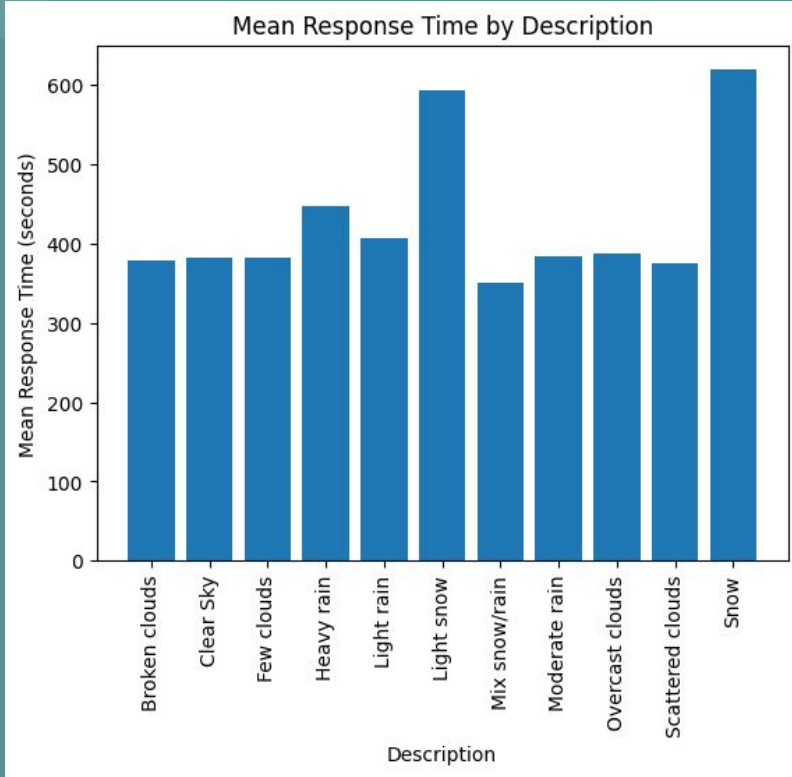


Number of Incidents



Average Response Time

Meteorological Analysis of Incidents



Meteorological Analysis of Incidents

Machine Learning

- Random Forest
 - RMSE: 156 seconds
 - $R^2 = 0.04$
- Gradient Boosting
 - RMSE: 54191 seconds
 - $R^2 = 0.09$
- Prioritized location and time over weather



Conclusions

- The primary focus for government should be on spatial disparities between areas of Nashville, contributing to longer response times
- Investments in weather-resistant systems would prove useful in lowering response times, including evenly distributing these systems throughout Nashville
- Incident frequency has a strong relationship with time-based metrics and local government should use time-based smart suggestions to keep ahead



Bibliography

- Pettet, G., Nannapaneni, S., Stadnick, B., Dubey, A., & Biswas, G. (2017, August). Incident analysis and prediction using clustering and bayesian network. In *2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)* (pp. 1-8). IEEE.
- Soni, A. (2018). Violin plots explained. Towards Data Science.
<https://towardsdatascience.com/violin-plots-explained-fb1d115e023d>