

MTA RIDERSHIP, ACCIDENTS, AND WEATHER PATTERNS IN NYC IN 2020: AN ANALYSIS

Sydney Mason, Meredith Rupp, Zain Khan,
Christian Dita, and Andrew Jensen

CONTENT

- Introduction of Research Question
- Data
- Results
- Conclusion & Future Implications

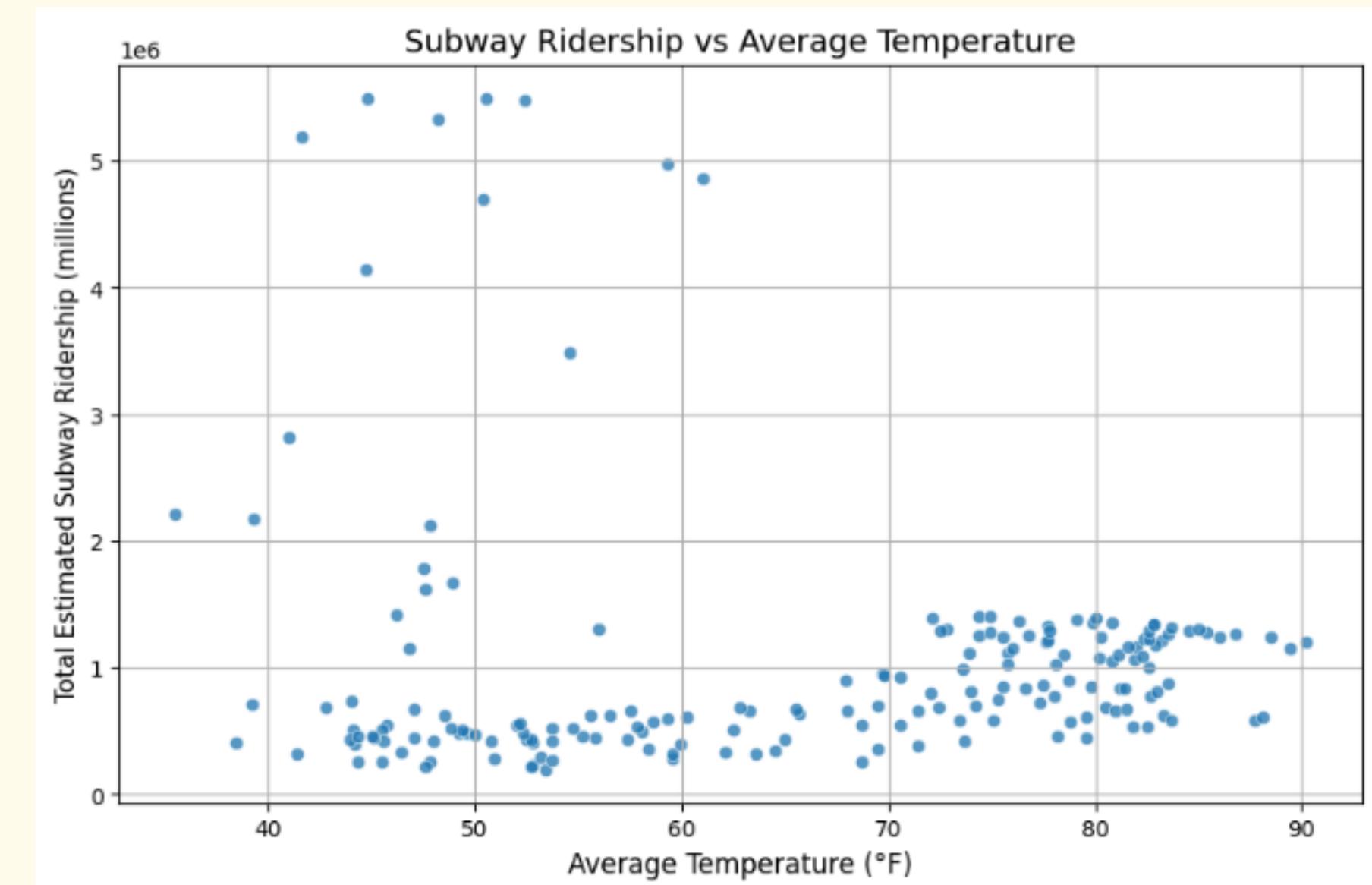
INTRODUCTION

Metro Transit Authority (MTA) is a fact of life in NYC

- Depending on the day, ridership varies considerably based on a myriad of factors, affecting expected revenue and needs for staffing
- Examined weather, ridership, and accident data to see if there were connections between the three to predict ridership as high or low above median ridership from March 01 - August 29 2020

OUR QUESTION

Can we predict if a day is high or low ridership on the MTA (above or below the median) based on daily accident counts and weather conditions?



MTA DATA

Origin of Data

- Collected by the Metropolitan Transportation Authority
- Covers daily ridership statistics across various transit modes
- March 1, 2020 - December 31, 2020

Variables

- Date
- Subways: Total Estimated Ridership

WEATHER DATA

Origin of Data

- Collected by the Laguardia Airport Weather Station, a branch of the National Weather Service
- Includes detailed daily weather reports from
- January 1, 2020 - December 31, 2020

Variables

- Temperature (°F) Avg
- Precipitation (in) Total
- Humidity (%) Avg
- Date

ACCIDENT DATA

Origin of Data

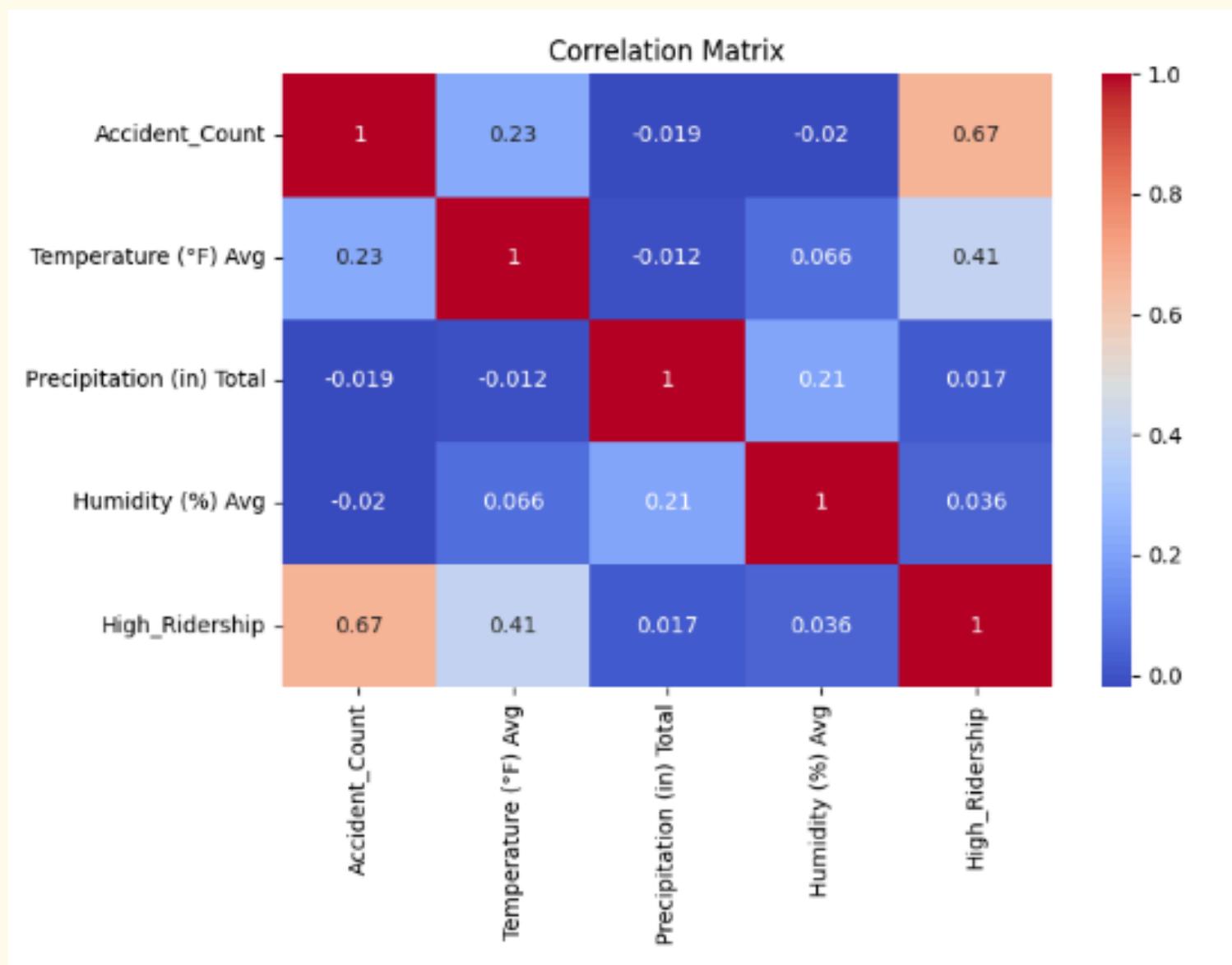
- Collected by the New York City Police Department and made available through NYC Open Data
- detailed record of all traffic collisions reported in the city
- January 1, 2020 - August 29, 2020

Variables

- Crash Date
- Crash Time
- Borough
- Number of Persons Injured
- Contributing Factor for each vehicle involved

RESULTS

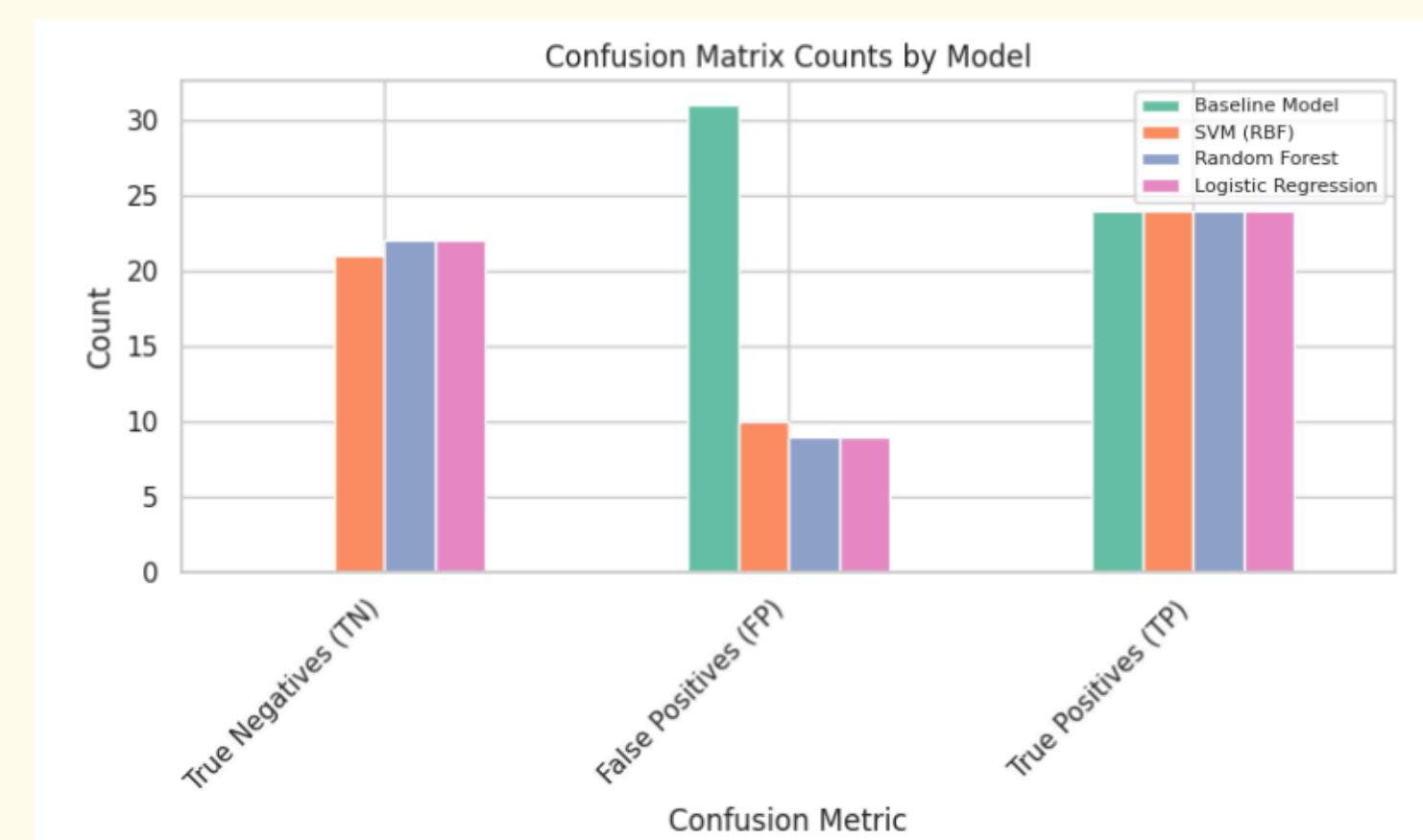
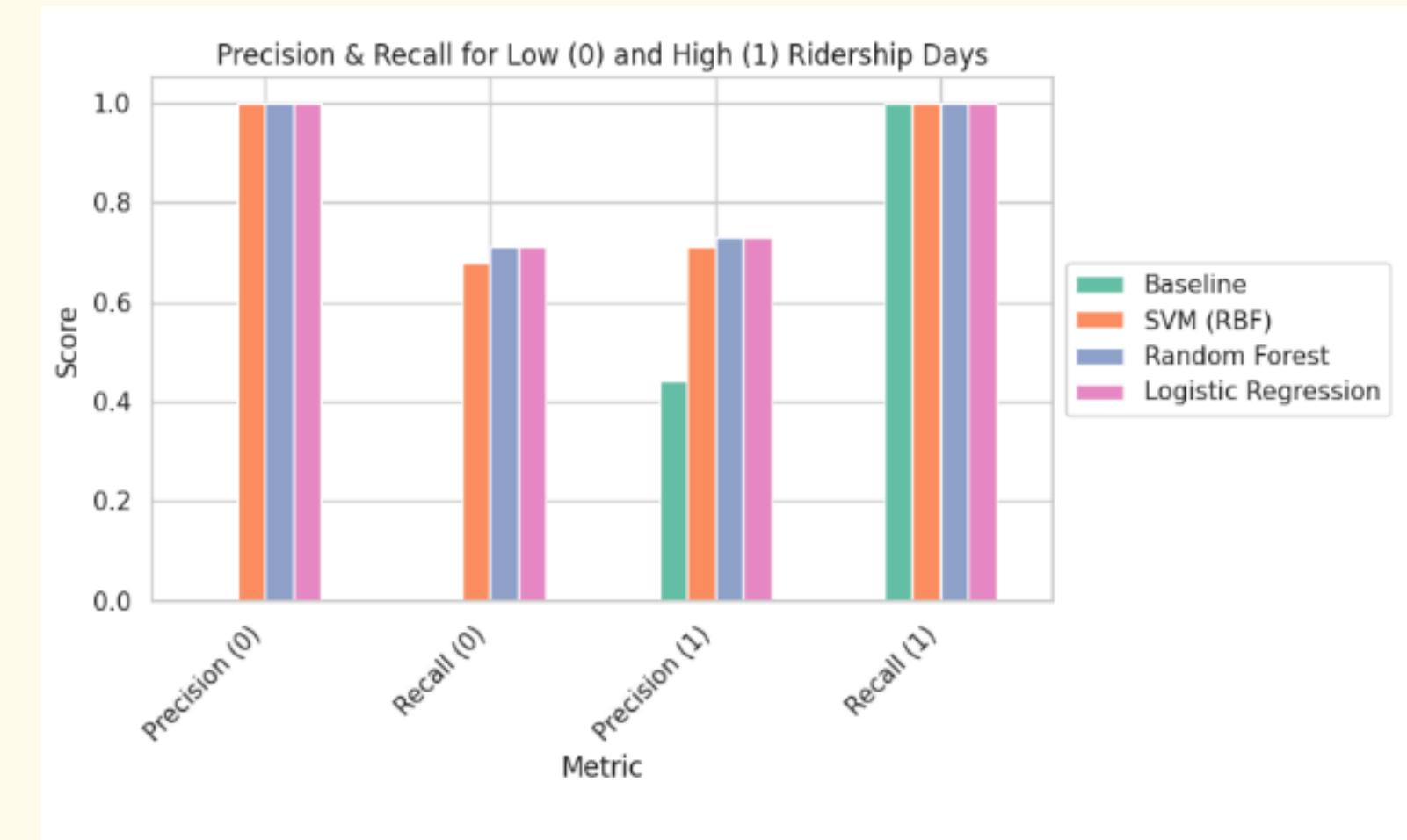
CAN WE PREDICT IF A DAY IS HIGH OR LOW RIDERSHIP ON THE MTA (ABOVE OR BELOW THE MEDIAN) BASED ON DAILY ACCIDENT COUNTS AND WEATHER CONDITIONS?



Feature	VIF
const	37.25
Accident_Count	1.05
Temperature (°F) Avg	1.06
Precipitation (in) Total	1.05
Humidity (%) Avg	1.05

RESULTS

- Baseline
 - accuracy score: 43.6%
- SVM
 - accuracy score: 82%
- Random Forest
 - accuracy score: 84%
- Logistic Regression
 - accuracy score: 84%



* Note that “Positives” are defined as days with High Ridership and “Negatives” are defined as days with Low Ridership.

BEST MODEL

- Either Logistic Regression or Random Forest
 - Logistic Regression final choice because it is a less complicated and therefore more accessible model.
 - Accuracy Score: 84%
 - Precision High: 73%
 - Precision Low: 100%
 - Recall High: 100%
 - Recall Low: 71%

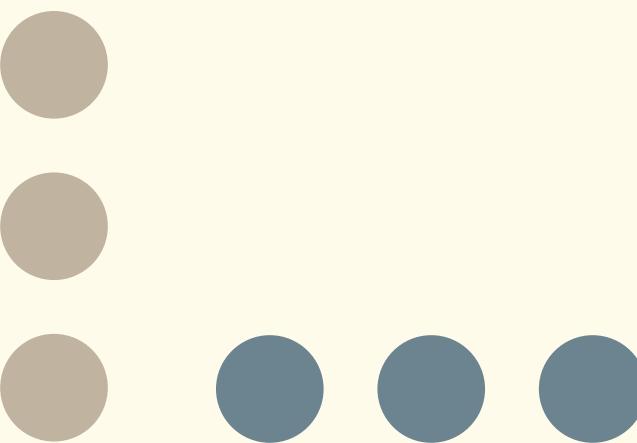
CONCLUSION

Applications:

- Allows MTA/ NYC government to predict revenue
- Staffing can be changed depending on ridership reducing the overall strain on budgeting

Future Study:

- Examining holidays/ special events to see if they affect ridership and non-Covid data



Questions?

Thank you

CITATIONS

American Meteorological Society. "Rain - Glossary of Meteorology." Glossary of Meteorology, Mar. 2024, glossary.ametsoc.org/wiki/Rain.

Balk, Tim. "78% Say NYC Subway Not Safe at Night, 24-Point Bump since 2017: Poll." New York Daily News, 19 Mar. 2024, www.nydailynews.com/2024/03/19/78-say-nycsubway-not-safe-at-night-24-point-bump-since-2017-poll.

DiMaggio, Charles, et al. "Spatial Analysis of the Association of Alcohol Outlets and Alcohol-Related Pedestrian/Bicyclist Injuries in New York City." *Injury Epidemiology*, vol. 3, no. 1, 4 Apr. 2016, <https://doi.org/10.1186/s40621-016-0076-5>. Accessed 24 Sept. 2019.

Encyclopedia Britannica, Encyclopedia Britannica, inc., www.britannica.com/dictionary/eb/qa/parts-of-the-day-early-morning-late-morning-etc. Accessed 28 Apr. 2025.

Masterson, Les, and Kelly Anne Smith. "Drunk Driving Statistics 2025." Edited by Patricia Louis, Forbes, Forbes Magazine, 3 Jan. 2025, www.forbes.com/advisor/car-insurance/drunk-driving-statistics/.

Shaik, Md. Ebrahim, et al. "A review on neural network techniques for the prediction of Road Traffic Accident severity." *Asian Transport Studies*, vol. 7, 2021, p. 100040, <https://doi.org/10.1016/j.eastsj.2021.100040>.

Sirikul, Wachiranun, et al. "Mortality-risk prediction model from road-traffic injury in Drunk Drivers: Machine Learning Approach." *International Journal of Environmental Research and Public Health*, vol. 18, no. 19, 8 Oct. 2021, p. 10540, <https://doi.org/10.3390/ijerph181910540>.