Final Project Component	Plan Presented in Proposal	Progress to Date
Research Question	1) The initial task is to forecast the speed violations; can we predict the likelihood of speed camera violations at specific intersections based on historical violation data? 2) Are there spatial patterns or clusters of speed camera violations across different intersections? 3) How do speed camera violation rates vary over time, such as daily, weekly, or seasonal patterns?	Completed
Description of Related Work	Using data mining, Amiruzzaman (2018) predicts traffic infractions, pinpointing peak hours and days for infractions. For classification tasks, several algorithms work well, including SVMs and Naive Bayes. The E-LLM model, which outperforms AR and neural networks, is proposed by Mozaffari et al. (2015) for vehicle speed forecasting. Shawky et al. (2017) identify posted speed limits and traffic volume as important variables for the best location of speed cameras on rural highways. DeepAR is a probabilistic forecasting technique that Salinas et al. (2020) present and shows notable accuracy gains over conventional methods. In their proposal, Wang et al. (2022) emphasize the usefulness of the adaptive Kalman filtering approach for traffic management and safety improvement when utilizing the ARMA framework for high-speed vehicle speed prediction.	Completed
Data Collection (Selection)	I have selected the data from the Chicago Data Portal, and the number of infractions that have happened in Children's Safety Zones every day for each camera. The data gets updated daily and currently has 378,726 instances of speed violations. There are 9 attributes in the dataset. Dataset Link: https://data.cityofchicago.org/Transportation/Speed-Camera-Violations/hhkd-xvj4/data_preview	Completed
Description of Data	The dataset includes information on speed enforcement cameras, including their addresses, unique camera IDs, violation dates, number of violations, X and Y coordinates, latitude, longitude, and location coordinates. Each address may have multiple cameras, each with its own ID. The coordinates are geocoded using Illinois State Plane East for X and Y coordinates, and WGS84 for latitude and longitude.	Completed
Data Preprocessing	I will use Python libraries like scikit-learn, pandas, matplotlib (for EDA), and numpy. Upon first glance, certain rows appear to have information about the camera, the date of the violation, and counts but no geographic coordinates. I will not discard these data rows because they contain valuable information. I will change the date format from a string to a datetime format to simplify processing and ensure consistency. The camera location and violation counts are absent from certain rows. When data for a particular camera on a particular date is unavailable, I will use 0 for the violation count. If I come across any duplicate entries, I will eliminate them all while keeping the first instance observed.	In Progress. To be completed by Apr 22, 2024
Selection ML Algorithms	I will be experimenting the DeepAR algorithm from AWS SageMaker, especially for time series forecasting tasks, for model training. Because DeepAR can capture temporal patterns and data dependencies it is a good choice. I might also investigate alternative algorithms like LSTM networks or conventional time series models like ARIMA.	Completed
Testing Data	Metrics like mean absolute error (MAE), mean square error (MSE), and root mean square error (RMSE) will be taken into account in order to assess the models' performance and determine how accurate the predictions are. In addition, I intend to evaluate the models' robustness using the cross-validation technique in order to prevent overfitting and generalization.	To begin by Apr 22, 2024 and complete by Apr 29, 2024
Results	My analysis relies on the DeepAR algorithm and time series models to provide accurate forecasts of speed camera violations at specific intersections. DeepAR and time series models generate forecasts that not only capture the overall trend but also account for seasonality, holidays, and other temporal patterns, enhancing the precision of predictions. This enables authorities to anticipate when and where speeding incidents are likely to occur, allowing for targeted interventions and resource allocation. By incorporating the outputs of DeepAR and time series models into decision-making processes, stakeholders can implement proactive strategies that effectively address variations in speeding behaviours, ultimately contributing to improved road safety outcomes.	To begin by Apr 29, 2024 and complete by May 3, 2024
Interpretation and Discussion	Using historical data, my analysis aims to predict speed camera violations at particular intersections. It may be possible to reduce speeding incidents by taking proactive steps. In order to do this, I will look for spatial patterns or clusters of violations across intersections. This helps me implement targeted interventions. To further understand daily, weekly, and seasonal trends, I will investigate temporal variations in violation rates. This will make it easier to put time-sensitive plans into action.	To begin by Apr 29, 2024 and complete by May 3, 2024
References	Amiruzzaman, M. (2018). Prediction of traffic-violation using data mining techniques. *Proceedings of the Future Technologies Conference (FTC) 2018*, 283-297. https://doi.org/10.1007/978-3-030-02686-8_23 Mozaffari, L., Mozaffari, A., & Azad, N. L. (2015). Vehicle speed prediction via a sliding-window time series analysis and an evolutionary least learning machine: A case study on san francisco urban roads. *Engineering Science and Technology, an International Journal*, *18*(2), 150-162. https://doi.org/10.1016/j.jestch.2014.11.002 Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2020). Deepar: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, *36*(3), 1181â€“1191. https://doi.org/10.1016/j.ijforecast.2019.07.001 Shawky, M., Sahnoon, I., & Al-Zaidy, A. (2017). Predicting speed-related traffic violations on Rural Highways. *Proceedings of the 2nd World Congress on Civil, Structural, and Environmental Engineering*. https://doi.org/10.11159/icte17.117 Wang, Y., Yu, C., Hou, J., Chu, S., Zhang, Y., & Zhu, Y. (2022). Arima model and few-shot learning for vehicle speed time series analysis and prediction. *Computational Intelligence and Neuroscience*, *2022*, 1â€“9. https://doi.org/10.1155/2022/2526821 https://github.com/aws/amazon-sagemaker-examples/tree/main https://docs.aws.amazon.com/sagemaker/latest/dg/deepar.html https://www.geeksforgeeks.org/deepar-forecasting-algorithm/ https://sagemaker.readthedocs.io/en/v2.106.0/algorithms/time_series/deep_ar.html	In Progress. To be completed by May 3, 2024