Traffic Data Analysis: A brief look into Georgia’s Roads

**By Jun Ming Li, Raheela Charania, Sharon Nancy**

**Background and Purpose**

Traffic congestion has been an ongoing issue within Gwinnett County. This impacts travel time, fuel consumption and overall productivity. To combat this, public transportation initiatives have been put in place, while multiple companies have decided to move to a hybrid or remote schedule. Despite this, congestion still exists in certain areas. This emphasizes the importance of collecting data to analyze traffic patterns. By doing so, we can understand the relationship between various traffic variables, predict congestion in future years and optimize route efficiency.

Through this analysis we aim to identify key factors that influence traffic flow and congestion. We reviewed data from the Georgia Department of Transportation’s (GDOT) Traffic Analysis and Data Application (TADA) site and incorporated the following variables in our analysis.

**Data Dictionary**

* D-factor: % of cars driving in the direction of traffic flow during peak times
* Single unit AADT: number of trucks in FHWA Categories 4 through 7
* ⁠Single unit AADT Peak: % of single unit trucks during peak hours of road use
* ⁠Combo unit AADT: number of trucks in FHWA Categories 8 through 13
* ⁠Combo unit AADT Peak: % of combo unit trucks during peak hours of road use
* ⁠Land use: Commercial (1); Residential (2)
* ⁠Number of lanes: Number of lanes the road had on either side

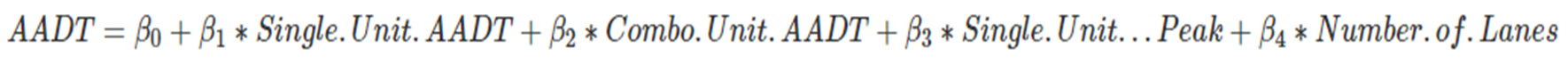
We then selected six high-traffic areas within Gwinnett County. We developed a predictive model for traffic conditions using multiple linear regression analysis on the data for those roads spanning the years of 2019 to 2022. This model was then applied to the 2023 data to determine its accuracy and predicting power. Overall, our model was about four times more accurate in predicting future AADT.

**Data and Methodology**

The traffic data was obtained from the Georgia Department of Transportation (GDOT). Using their Traffic Analysis and Data Application (TADA) tool, 6 highly used roads were selected around Gwinnett County, and their data was downloaded from their short-term active monitoring stations. The GDOT dataset contained information about the traffic flow over the year (AADT) and other variables relating to the vehicle type distribution and peak traffic flows. Additional information such as number of lanes and land use around the roads was gathered through scouting the areas around the roads.

To process our raw data, it was first filtered by year, applying a lookback period of 5 years from 2023. The information on the year 2024 was not available and was excluded from the model. Then each record was checked for outliers and missing values. It was found that some entries had incomplete records for Combo Unit Peak and/or D factor. To address these issues, the distribution of the variables was used to determine the best approach to replace the missing values and outlier. This was done to preserve the number of entries.



**Results**

To analyze our data, the whole data set was split into a training data set composed of entries for the year 2019 to 2022 and a testing data set of entries from 2023. An initial model was created using multiple linear regression. Then, VIF analysis and stepwise regression methods were then used to further refine the model and reduce the effects of multicollinearity and highlight the significance of variables. The final model is shown above.

The final model was tested against our testing data set, and it was found that compared to a baseline estimate using the mean AADT of the training dataset to predict future AADT, our model was nearly four times more accurate in predicting future AADT.





**Conclusions**

In conclusion, this analysis successfully highlights the ongoing issue of traffic congestion in Gwinnett County, emphasizing the need for data-driven solutions to optimize traffic flow. By utilizing data from the Georgia Department of Transportation's (GDOT) Traffic Analysis and Data Application (TADA), the study identified key factors influencing traffic patterns, including vehicle type distribution, peak traffic hours, and land use. The predictive model developed using multiple linear regression demonstrated the power of statistical methods in forecasting future traffic conditions. The model’s accuracy, being nearly four times more effective than a baseline estimate, proves the value of predictive analysis for better traffic management. The findings suggest that addressing factors like peak traffic flows and bottlenecks could significantly improve congestion, offering actionable insights for infrastructure planning and policy development. Furthermore, the study’s approach of filtering data and handling missing values ensures that reliable, robust predictions can be made, paving the way for smarter, data-driven decisions in urban traffic management.

We chose to study traffic congestion in Gwinnett County because it has a big impact on things like travel time, fuel usage, and productivity. Even though there have been efforts like public transportation improvements and more people working from home, traffic is still a problem in some areas. Understanding the reasons behind the congestion is important, as it can help us find ways to predict and improve traffic flow. By analyzing traffic data, we can figure out what factors are causing the issues and come up with solutions to make travel more efficient. This research is important not just for solving current problems but also for planning the future of transportation in the county. As Gwinnett County continues to grow, the insights from this study can help make decisions about policies and infrastructure that will improve traffic and transportation in the long run.

**Potential Predictions for Traffic Data / Future predictions.**

Based on our analysis, we predict that traffic congestion during morning peak hours (7 AM to 9 AM) will likely continue to increase unless we implement better traffic management solutions, such as optimized signal timings or additional lanes. Seasonal variations, especially during holidays or events, are also expected to worsen if not addressed with improved planning. External factors like weather-related delays, for example, a 20% reduction in speed during rainy conditions, and construction disruptions during peak hours will continue to impact traffic flow, making better planning crucial. We also anticipate that high-traffic areas, which are prone to bottlenecks, will experience persistent congestion unless infrastructure improvements, such as adding lanes or optimizing traffic signals, are made. Accident hotspots may see an increase in incidents unless targeted safety measures, such as improved signage and road repairs, are introduced. With population growth and urban development, traffic demand will rise, particularly in residential areas, unless we expand public transportation or improve road infrastructure. Lastly, policies like tolls or lane restrictions could help reduce congestion by limiting the number of vehicles on certain roads or improving flow during peak hours. These predictions highlight the need for ongoing investment in infrastructure, safety interventions, and policy changes to manage future traffic challenges.