Alexandra Veremeychik
Data 205: Capstone Experience in Data Science
CRN 22017
Dr. Jane Valentine,
Montgomery College
Maurice Valentine,
Montgomery County
Office of Consumer Protection

Trespass Towing Analysis

Trespass towing is the act of hauling a vehicle from one location to another without the consent of the vehicle's owner. In Montgomery County, Maryland, more than 100,000 vehicles have been towed in this manner since July 2021. The process is a stressful experience for vehicle owners. They are required to travel to a separate location and pay a significant fee. In many cases, lack of access to one's vehicle can put employment in jeopardy. Likewise, the process is taxing on the environment. Large tow trucks burn more fossil fuels, and owners are required to make additional potentially polluting trips to and from the location their car is stored.

By law, each invoice received for a trespass tow must include information prompting the vehicle owner to contact the Montgomery County Office of Consumer Protection with any complaints. In addition to investigating these complaints and ensuring merchants comply with regulations, the Office of Consumer Protection aims to reduce the number of trespass tows. The county enlisted Montgomery College to explore the data surrounding these cases, as well as some of the complaints filed against them. The goal of this partnership was to assess the trends in County tow data by analyzing County towing locations, towing agencies, and the towing populations while applying a racial equity and social justice lens.

The core of the data used to analyze trespass towing comes from a database maintained by the Montgomery County Police Department. An updated dataset is published to the dataMontgomery website monthly. Its current version includes over 100,000 entries detailing the the date each vehicle was towed, its location (address and coordinates), the reason it was towed, and information on the vehicle's make, model, and year of manufacture.

The Office of Consumer Protection also provided data on trespass towing complaints filed in the 2024 fiscal year. This data includes the location of the tow, the merchant in question, the practice being disputed, and the result of the case. This data was anonymized to preserve privacy.

Additional demographic, economic, and geographic data was merged with these two datasets in order to analyze patterns related to the characteristics of the towing locations. The Community Equity Index (CEI) is "a composite measure of five indicators of how the county's socioeconomic diversity is distributed, or which neighborhoods have concentrated advantage or disadvantage and which neighborhoods are representative of the county's diversity." Table 1 provides an overview of the key variables used in this analysis.

Measure	**Indicator**
Economic insecurity	Percentage of people who are under 200% of the poverty level
Earnings potential	Percentage of people without a bachelor's degree (or higher)
Housing stability and wealth building	Percentage of people who rent housing
Barriers to inclusivity and resource access	Percentage of people who speak English less than very well
Income	Per capita income

Table 1: Community Equity Index

The American Community Survey (ACS) is a yearly survey conducted by the U.S. Census Bureau that

collects data on the U.S. population's demographics, social, economic, and housing characteristics. Variables related to tract median household income, language spoken at home, foreign-born population, and means of travel to work were queried using Python and the Census API.

The Environmental Protection Agency (EPA) Smart Location Database (SLD) is a publicly available geographic data resource that measures location efficiency across the United States. The SLD contains over 90 attributes for each census block group. It was used to provide data on building and population density, as well as transit access.

Tools and Methods

Python played a pivotal role in these projects. The table below summarizes the Python libraries used and their specific purposes across all notebooks.

Package	**Role**	**Specific Purpose**
`pandas`	Data Manipulation	Aggregating, filtering, and preparing data for analysis and visualization.
`matplotlib.pyplot`	Visualization	Creating line charts, scatter plots, and bar charts for exploratory analysis.
`seaborn`	Statistical Visualization	Enhancing visualizations with heatmaps and statistical insights.
`numpy`	Numerical Computation	Supporting matrix operations, coordinate transformations, and calculations.
`folium`	Geospatial Visualization	Creating interactive maps to display geographic and temporal data.
`folium.plugins`	Geospatial Visualization Tools	Adding heatmaps and time-based visualizations to maps.
`squarify`	Visualization	Generating treemaps to display proportions, such as reasons for tows.
`plotly.express`	Interactive Visualization	Building interactive plots, including bubble and bar charts.
`plotly.io`	Input/Output Operations	Exporting visualizations as shareable HTML files.
`pygris`	Geospatial Data Retrieval	Obtaining coordinates from addresses and GeoIDs (census) from coordinates.

The following tools were utilized during the project for codebase organization, data analysis, website hosting, and deployment, and documentation:

Tool	**Purpose**
Google Colab	Running Python Code
Cursor AI	Organizing and managing the codebase efficiently with AI-assisted features.
GitHub	Hosting the codebase, dataset and enabling collaborative development.
Netlify & Vercel	Hosting and deploying interactive and dynamic web applications with ease, scalability, and smooth accessibility.
Overleaf	online LaTeX editor

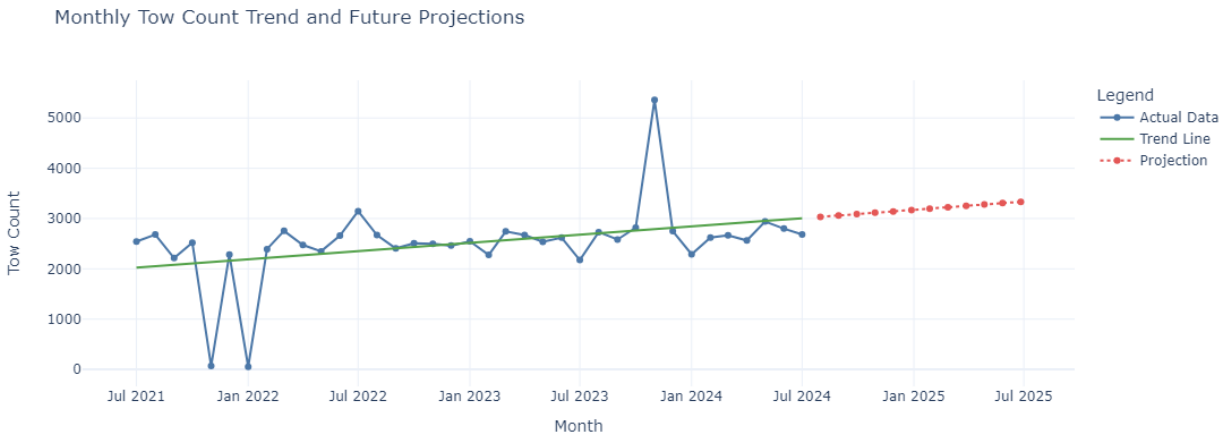

Figure 1: Linear regression projection

The data was wrangled using a custom Python script to correct various spelling mistakes. The pygris Python package was used to geocode the geoid from the towing coordinates and the coordinates from the complaint addresses. Repossessions were removed from the data, as they do not fall under the jurisdiction of the Office of Consumer Protection. Likewise, repossession tows are not required to be reported to the police, so those erroneously recorded in the database do not offer accurate parameters on the practice.

With corrections, exclusions, and filtered for the fiscal years 2022-2024 (July 1, 2021 - June 30, 2024), the dataset contains 93,059 cases of trespass towed vehicles in Montgomery County. Demographic, economic and geographic information from outside sources is delineated on Census tract boundaries. Montgomery County is divided into 250 Census tracts. The population of each tract ranges from 1,434 to 8,602 people and population density (residents/acre) ranges from 0.220 to 125.984.

Each Census tract has seen a range of 1 vehicle towed to 3053 vehicles towed in the last three years, with a mean of 373 vehicles. There is an upward trend of increasing tows over the past 3 years. The most common reason given for towing is the violation of a no overnight parking rule, with 53.6% of the cases labeled as such.

The tows were aggregated by census tract to compare the local characteristics of the tow locations. Correlations were measured to assess possible relationships that result in higher towing rates in a particular tract. Five are included in the table below and are further explored with the data product we created.

Variable	**Spearman Correlation**	**p-value**
Median Household Income	-0.6079	1.089e-09
Activity Density (jobs + residents/acre)	0.5580	4.238e-08
CEI	-0.5250	3.463e-07
Residential Density (housing units/acre)	0.5117	7.643e-07
Employment Density (jobs/acre)	0.3886	2.825e-04

Table 4: Spearman Correlation Between Local Characteristics and Towing Rates

Table 4 reveals that lower *Median Household Income* ($r = -0.6079$, $p < 0.001$) and lower *CEI* ($r = -0.5250$, $p < 0.001$) are significantly associated with higher towing rates. Conversely, higher *Activity Density* ($r = 0.5580$, $p < 0.001$), *Residential Density* ($r = 0.5117$, $p < 0.001$), and *Employment Density* ($r = 0.3886$, $p < 0.001$) are positively correlated with higher towing rates, indicating that areas with greater density and lower income experience more towing activity.

Data Product

The magnitude of the data and the issue required a robust manner to review relationships, trends, and patterns. A dashboard allows for the aggregation of various visualization methods. Due to the geographic nature of the data, one centered on the mapping of trespass towing cases and complaints was chosen. It is divided into two sections, with one consisting of a variety of graphs and the other a series of maps.

This map is accompanied by a bubble chart that illustrates this relationship, along with the relationship between the median household income of a Census tract and its population density (the number of people living in it). Together, the map and the bubble chart make clear that one is unlikely to get trespass towed from more advantaged areas. These do not provide enough information to suggest that strong or causal relationships exist between the CEI score of an area and the likelihood of being towed there. However, there is enough to warrant further attention and investigation.

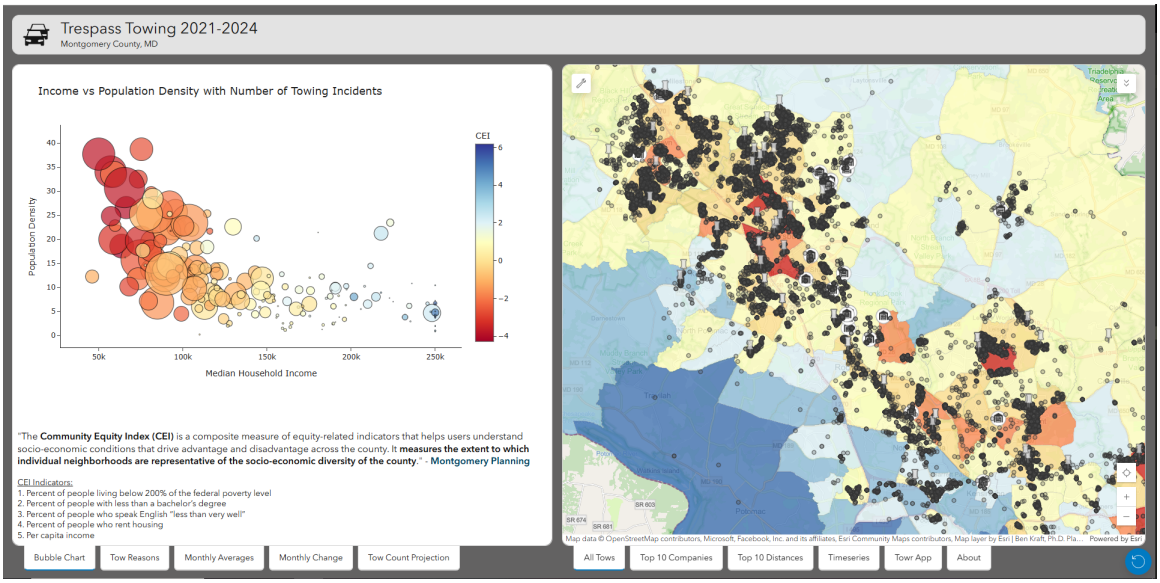

Figure 2: Interactive Dashboard

Among the visualizations included in the dashboard is a timeseries of the monthly average trespass tows for the top 5 towing companies, which make up 79.4 % of all trespass tows July 1, 2021 - June 30, 2024, as well as graph illustrating the rate of change in towing. The most active company, Henry's Wrecker Service, also exhibits dramatic volatility. Unethical business practices may or may not be responsible for rapid changes in towing rates, however if these metrics are correlated with certain behaviors than observing them may help identify possible issues.

One of the objectives of the project was to calculate and analyze the distances traveled by vehicles towed to registered storage lots. While the Montgomery County code Chapter 30C states that no vehicle may be trespass towed to a lot more than 15 miles from the parking location, county officials have not settled on a consensus as to whether this constitutes 15 driving miles or straight-line distance.

As calculating driving distances is costly in terms of resources and requires an API, finding the approximate driving distance of all 93,059 cases was impractical for the scope of this analysis. Driving distance has been correlated with straight distance in cases where significant shorelines or land barriers are present, as is true for Montgomery County. A sample of 10,000 cases was taken from the top 10 most active towing companies and their driving distances acquired through a Python script querying OpenStreetMap.

The Haversine formula is used to calculate the great-circle distance between two points on a sphere. It is given by:

$$d = 2R \cdot \arcsin\left(\sqrt{\sin^2\left(\frac{\Delta\phi}{2}\right) + \cos(\phi_1) \cdot \cos(\phi_2) \cdot \sin^2\left(\frac{\Delta\lambda}{2}\right)}\right)$$

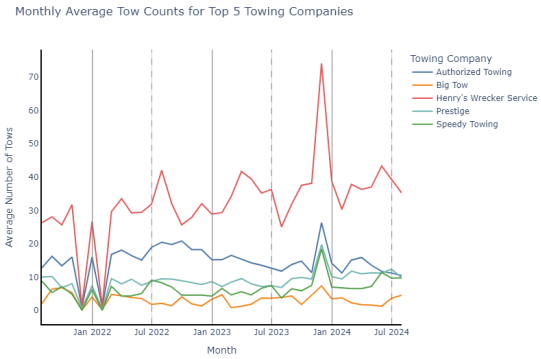

Figure 3: Monthly Average of Towing Data

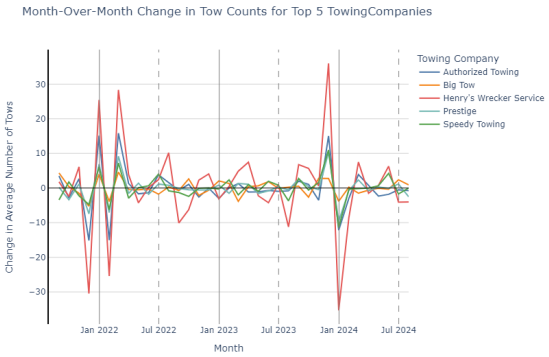

Figure 4: Monthly Rate of Change in Towing

where:

- $\Delta\phi = \phi_2 - \phi_1$: Difference in latitude (radians)
- $\Delta\lambda = \lambda_2 - \lambda_1$: Difference in longitude (radians)
- R: Earth's radius (3958.8 miles)
- ϕ_1, ϕ_2: Latitudes of the two points (radians)
- λ_1, λ_2: Longitudes of the two points (radians)

The Haversine distance was multiplied by a factor of 1.4 to estimate the effective towing distance, considering potential road conditions and detours. The actual towing distance was provided in meters and converted to miles using the conversion factor $1\,\text{km} = 0.621371\,\text{miles}$.

To evaluate accuracy, we calculated the Mean Absolute Error (MAE) between the estimated and actual distances:

$$\text{MAE} = \frac{1}{10000} \sum_{i=1}^{10000} |\text{Estimated Distance}_i - \text{Actual Distance}_i|$$

The results revealed a MAE of approximately 0.96 miles, indicating that the estimated distances deviated from the actual towing distances by less than 1 mile on average. This demonstrates a reasonable level of accuracy for the estimation method used. Various factors were tested, and the factor 1.4 resulted in the lowest Mean Absolute Error (MAE). This value of 1.4 is also supported by existing literature, as noted by Jorritsma 2, who demonstrated its effectiveness in providing accurate estimations. Distance estimations are used to create the third map in the dashboard, in which a deeper red represents vehicles that were towed further from their parking location to the storage lot. These estimations revealed that approximately 21.5% of tows in the sample violate the 15 mile rule.

The Haversine formula has the potential to be a resource-light method of flagging merchants for further inspection. To demonstrate its utility, we created two tools that incorporate it. The first is a pre-screening form that prompts a consumer to enter the name of the company that towed their vehicle, as well as the address it was towed from. The Haversine formula is then used to determine whether the Office of Consumer Protection should be alerted to inspect the distance the vehicle was transported. The second application is one that allows county employees to select any location on a map of Montgomery County to see the Haversine distance between it and all of the towing companies registered there, flagging those over an adjustable threshold. If such tools rely on an API, they become slow and costly. Our findings reinforce earlier research that deems straight-line distances adequate estimations of driving distances.

Application Overview

As part of the project, we developed four tools in collaboration with Dr. Rebin Muhammad of Montgomery College, each serving a specific purpose. The table below summarizes these tools, their purposes, and their URLs.

Website Name	Audience	Purpose	URL
Ranking Towing Company	County, Towing Companies	Provides rankings of towing companies based on performance in four categories, with formats for both general audiences and NOP.	`https://towrank.netlify.app/`
Towr	Local Residents	Displays historical towing data at a given address or current location to help avoid towing risks in Montgomery County.	`https://towr.netlify.app`
Towing Station Distance	County Officials	Analyzes distances between towing stations and user-specified locations for planning and operational purposes.	`https://towwatch.vercel.app`
Pre-Screening Complaint Form	Towed Vehicle Owners	Allows users to pre-screen towing complaints for accuracy and eligibility before filing an official complaint.	`https://towdist.vercel.app`

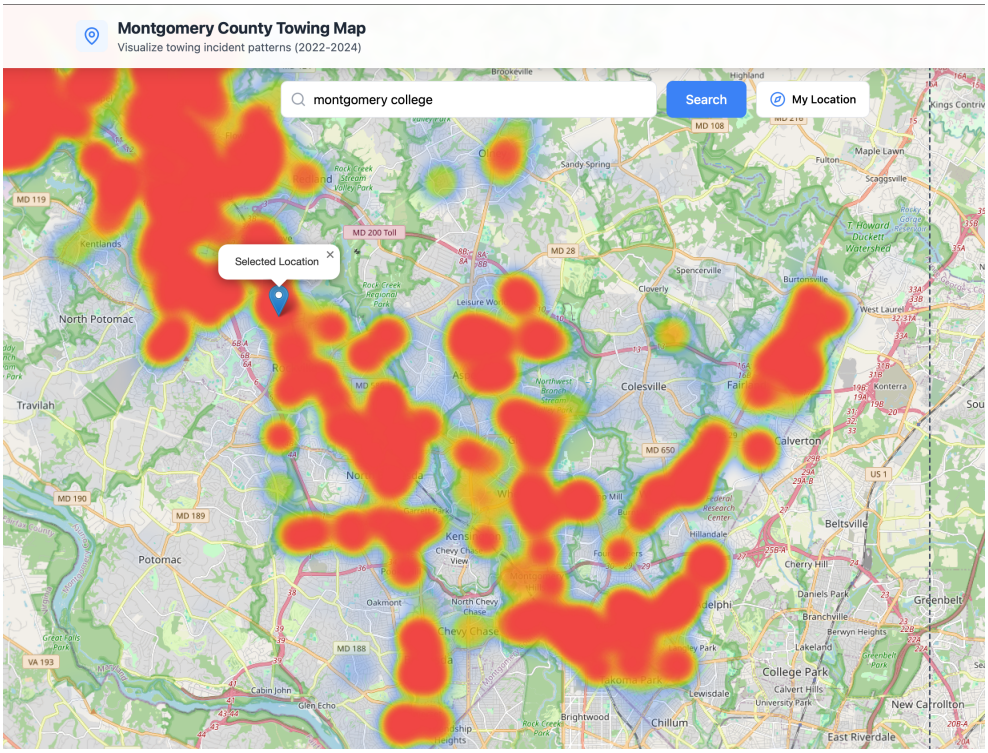

Figure 5: Interactive Tow Zone Hot Spot Map showing towing history by location.

These applications are meant to demonstrate the utility that data-driven solutions can offer the government and its citizens. In addition to the assistance they offer, similar applications can be developed that ethically collect data as well.

Acknowledgments

My endless gratitude goes out to everyone who supported this project. The expertise and example set forth by Dr. Jane Valentine was a beacon in times of challenge and uncertainty.

I am likewise grateful to all of the individuals involved in the partnership between Montgomery College and the Montgomery County government. Outreach and education manager of the Office of Consumer Protection Maurice Valentine's weekly guidance allowed the project to flourish, as did that of the Technology and Enterprise Business Solutions' Kathy Luh and Victoria Liu. Samuel Buo, Shaun Carew, and David Elswick supplied the resources and knowledge needed to ensure a well-rounded approach.

This partnership would not exist without Professor Rachel Saidi, who's excellence as a mentor elevates the work of all those that surround her.

My gratitude extends to the Mathematics, Statistics, and Data Science department at Montgomery College, and in particular the instructors who helped build the foundation of my data skills, Professor Mais Alraee and Professor Abdirisak Mohamed. Furthermore I would like to acknowledge the camaraderie of my fellow students as buoyant through a turbulent semester.

Words could never express the monumental role the support of my family played.

References

1. Wikipedia. "Haversine Formula." Available at: https://en.wikipedia.org/wiki/Haversine$_formula$.

2. Jorritsma, R. (2013). "A Nationwide Comparison of Driving Distance Versus Straight-Line Distance to Hospitals" *PLoS Medicine*, 10(12).
   Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC3835347/.