Clustering Vehicles by Usage Patterns

Sevinch Pasilova (TA), Aman Ali, Younes Aref, Nathan Fecadu, Parshawn Haynes, Omar Ibrahim, Abhinav Kotamreddy, Chanh Ly, Badr Ramadan, Pragun Sethi, Kashyap Vallur, Victor Voznyuk

Allison Transmission

PURDUE UNIVERSITY | The Data Mine

Background and Mission

Introduction:
Allison Transmission's products find extensive use in various vehicle applications, including fire trucks, school buses, and garbage trucks. These vehicles exhibit different driving patterns.

Objective:
Categorize vehicles based on the usage metrics and the geographic areas where they operate.

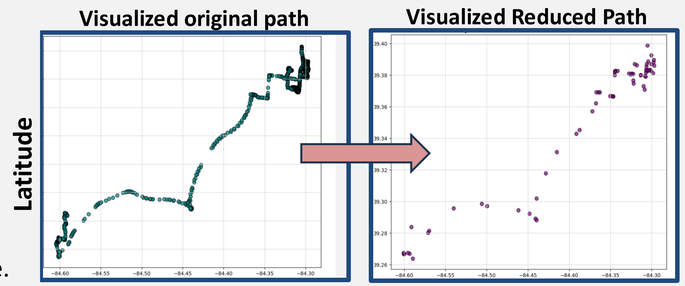
Methodology

- Reduced the GPS dataset using a variation of the RDP algorithm and extracted additional telemetry data through APIs and geographic datasets.
- Researched various clustering methods for both new and old datasets
- Experimented with multiple clustering techniques:

K-means | DBScan |Gaussian MM

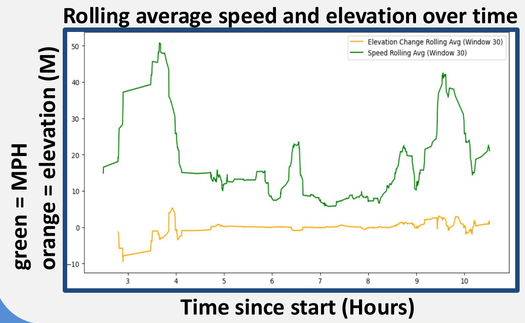
Reducing GPS Data Size

- **Grouping by Location and speed:** Data points were grouped by their location, giving us 3 sections:
  - **Highway** (not in city and 55+ mph)
  - **City** (within city limits and <55 mph)
  - **Traffic/Rural** area(neither of previous)
- **A Dynamic RDP Algorithm** was used to reduce redundant points while keeping data in the city dense, but highway sparse.
- **In Total Reduced** 2958 points to 489 data points



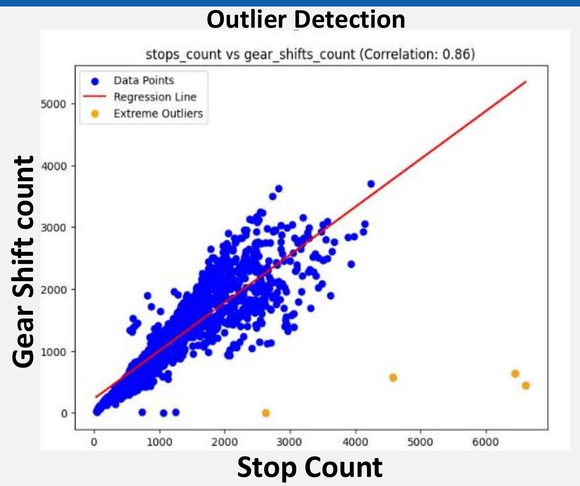
Extracting Elevation

- Using the reduced dataset and Google's elevation API we go accurate elevation data for each point
- Using the API data, we were able to get:
  - Elevation change
  - Road grade



Clustering Methodology

- **Applied logarithmic scaling** to normalize value distributions and reduce the impact of extreme values.
- **Handled outliers on a case-by-case basis**; removed data points only when an entire column contained unreasonable or non-representative values.
- **Performed correlation analysis** to find and remove features that were too similar, helping reduce redundancy and improve clustering accuracy.
- **Implemented Gaussian MM clustering** because the number of clusters is predetermined by the algorithm



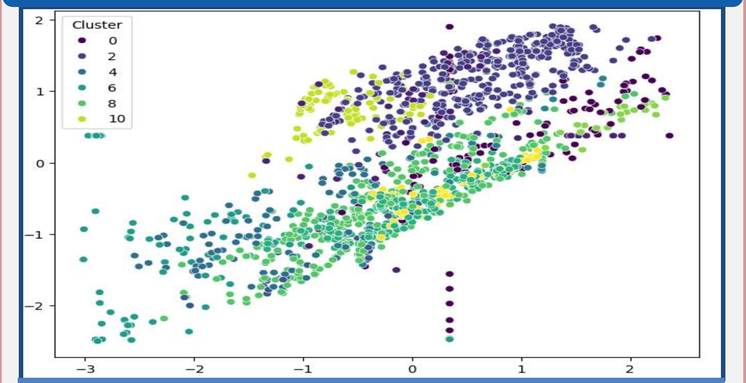
Telematics results

Telematics Datasets: From the GPS coordinates and time values we were able to extract:

	Data Type	Value
0	Total time within city	3.95 hours
1	Total time spent driving	8.27 hours
2	Total distance traveled (km)	191.71 km
3	Distance within bounds (city)	139.68 km
4	% of distance within city	72.86%
5	% of time within city	47.78%
6	Time on highway	0.64 hours
7	Distance on highway (km)	86.09 km
8	% of time on highway	7.72%
9	% of distance on highway	44.91%
10	Avg distance from highway (m)	1383.20 m

- Using these new data values, it is possible to increase the accuracy of the clustering algorithm, as more data points increases the accuracy of the clustering.

Clustering results



Cluster	AVG Trip Length	AVG Shift Frequency	Description
0	Long (5.8mi)	Few (1.4/mi)	Highway; Likely long-haul or highway fleets.
2	Short (3.0mi)	Moderate (2.0/mi)	Low; Possibly infrequent or limited-range use.
4	Moderate (5.0mi)	Frequent (3.2/mi)	Urban; Suggests stop-and-go traffic (e.g., delivery trucks).
6&8	Mid-range (4.5mi)	Balanced (2.5/mi)	Mixed; Likely regional or versatile fleets.
10	Moderate (4.3mi)	Fewer (2.0/mi)	Steady Service; Consistent driving (e.g., service vehicles).

Conclusion & Future Work

Conclusion: This analysis revealed key vehicle usage patterns, allowing us to effectively categorize vehicles by geography and application. These insights enabled the development of improved optimization, enhancing performance and efficiency.

Future work: will incorporate the newly developed GPS metrics into future clustering analysis which will result in more precise grouping of vehicles based on movement and regional patters. Using these findings, Allison Transmissions would be able to develop products that fit the needs of specific vehicles.

Acknowledgements: We would like to thank our Corporate Partner Mentors and The Data Mine Staff for their guidance and support of us.