WHITE PAPER

A vehicle's purpose: Providing vocation insights to transportation planners

GEOTAB ITS

About Geotab ITS

With access to one of the world's largest organically grown transportation datasets, Geotab ITS aggregates data from millions of connected vehicles to produce actionable transportation insights and urban analytics for transportation leaders across the United States and Canada.

These insights are driven from privacy-by-design principles and are provided through the Geotab ITS Altitude Platform, a secure, modular, open transportation analytics platform that enables partners and customers to quickly interact with the insights and to make informed decisions to improve safety, efficiency, sustainability and profitability of the infrastructure they are responsible for. To learn more, visit: **its.geotab.com**.

Acknowledgements

The author would like to thank Bob Bradley, Associate Vice President–Data Solutions; Kyle Liu, Data Science Team Lead–Geospatial; Natalie Smith, Senior Data Scientist; and Shweta Shah, Lead Product Manager–Data and Analytics for their valuable contributions and expertise provided during the development of this white paper.

Author: Nicki Schill, Marketing Manager, Public Sector Communications.

Transportation planners are digging deep into traffic data in order to make informed decisions for their region's infrastructure with the aim to create more efficient roadways.

That traffic data has traditionally been collected through either manual processes like mail-in surveys or increasingly through automated vehicle detection systems. The former suffers from low response rates and limited impact while the latter lacks a critical factor – context.

Without context, planners are left guessing as to why situations happened in the first place and cannot proactively plan for future incidents and infrastructure needs. More informed transportation planning decisions will require a deeper level of understanding context for traffic patterns and anomalies. One of those contextual factors that can play an integral part in aiding these decisions is vocation; understanding not just traffic volume but the vehicle's purpose on the road. Vocation and other contextual insights are a key part of the Geotab ITS Altitude platform. ITS data helps to provide critical answers for why vehicles are on the road in order to make informed decisions to support the overall flow of traffic and commercial freight activity.

Vocation sheds light on vehicle behavior

With over tens of millions of commercial vehicles in the US alone, it can be difficult to separate or classify them into useful groups or identifiers for meaningful analysis. Vocation classifications allow for deeper, useful insights into why a vehicle is on the road by honing in on the vehicle's purpose and function. The vocation of a vehicle can also then be combined with the industry classification, vehicle type and other vehicle features to segment the vehicle population even further. This is key information for a planner who can start to analyze vehicle behavior and make the necessary infrastructure adjustments.

For example, within the Altitude platform, a transportation planner can isolate last-mile delivery vehicles and understand exactly how frequently and how long they are using premium curbside space.

Vocation classifications

Using Machine Learning (ML) to automatically classify vehicles based on their driving pattern, Geotab ITS was able to identify several segments of vehicle behavior that we call vocations. The driving patterns for vehicles in Geotab ITS' ecosystem were closely monitored to develop the following five vocation classifications:

Vocation Name	Description
Local	The vehicle's range of activity is below 150-air-miles (or 241 km) thus qualifies for the **short-haul exemption** under Hours of Service Regulations. In addition, the vehicle does not exhibit behavior in line with other vocations such as hub-and-spoke and door-to-door. Metrics: ■ Geographic range < 241 km
Regional	The vehicle has a wide range of activity, over the 150-mile threshold, but tends to rest in the same location often. The vehicle is also neither hub-and-spoke nor door-to-door. Thresholds where vehicles are more likely to be classified as regional: ■ Geographic range **under** 670 km ■ Net movement **under** 90 km/day ■ Stays at domicile **over** 40% of the time
Long Distance	The vehicle has a very large range of activity and typically does not rest in the same location. The vehicle is also neither hub-and-spoke nor door-to-door. Thresholds where vehicles are more likely to be classified as long distance: ■ Geographic range **over** 670 km ■ Net movement **over** 90 km/day ■ Stays at domicile **under** 40% of the time
Hub-and-Spoke (On demand)	The vehicle spends many of its work days making multiple round trips from a singular location (a centralized hub). Typically the vehicle averages over one round trip per working day, with these round trips accounting for the majority of its total mileage. Thresholds **above** which vehicles are more likely to be hub-and-spoke: ■ 22% of work days consist of multiple round trips ■ 1.00 round trip per day ■ 90% of total mileage comes from round trips
Door-to-Door (Last mile)	The vehicle makes significantly more stops than most per work day but also tends to spend very little time per stop. This vocation will mostly include smaller vehicles traveling for last-mile delivery and rideshare purposes but may also include other larger vehicles that operate in a similar manner such as heavy-duty trucks working in waste collection. Thresholds where vehicles are more likely to be door-to-door: ■ Average stop time **under** 108 seconds ■ Average **over** 27 stops per day

Technical summary

The vehicle vocation system clusters every active vehicle into one of 5 vocations on a monthly basis based on activity in the previous 3 months. The system first analyzes raw telematics data (such as location, time, and speed) to create Contextualized Vehicle Movement Datasets (Section 1) of a vehicles domicile, the stops it makes, and its duty cycles. These movement data is then aggregated for each vehicle to generate Features (Section 2): These are a series of metrics that encapsulate the vehicles movement patterns. These Features served as inputs to three Clustering Models (Section 3), the outputs of all three models are then combined to determine vehicles' vocation.

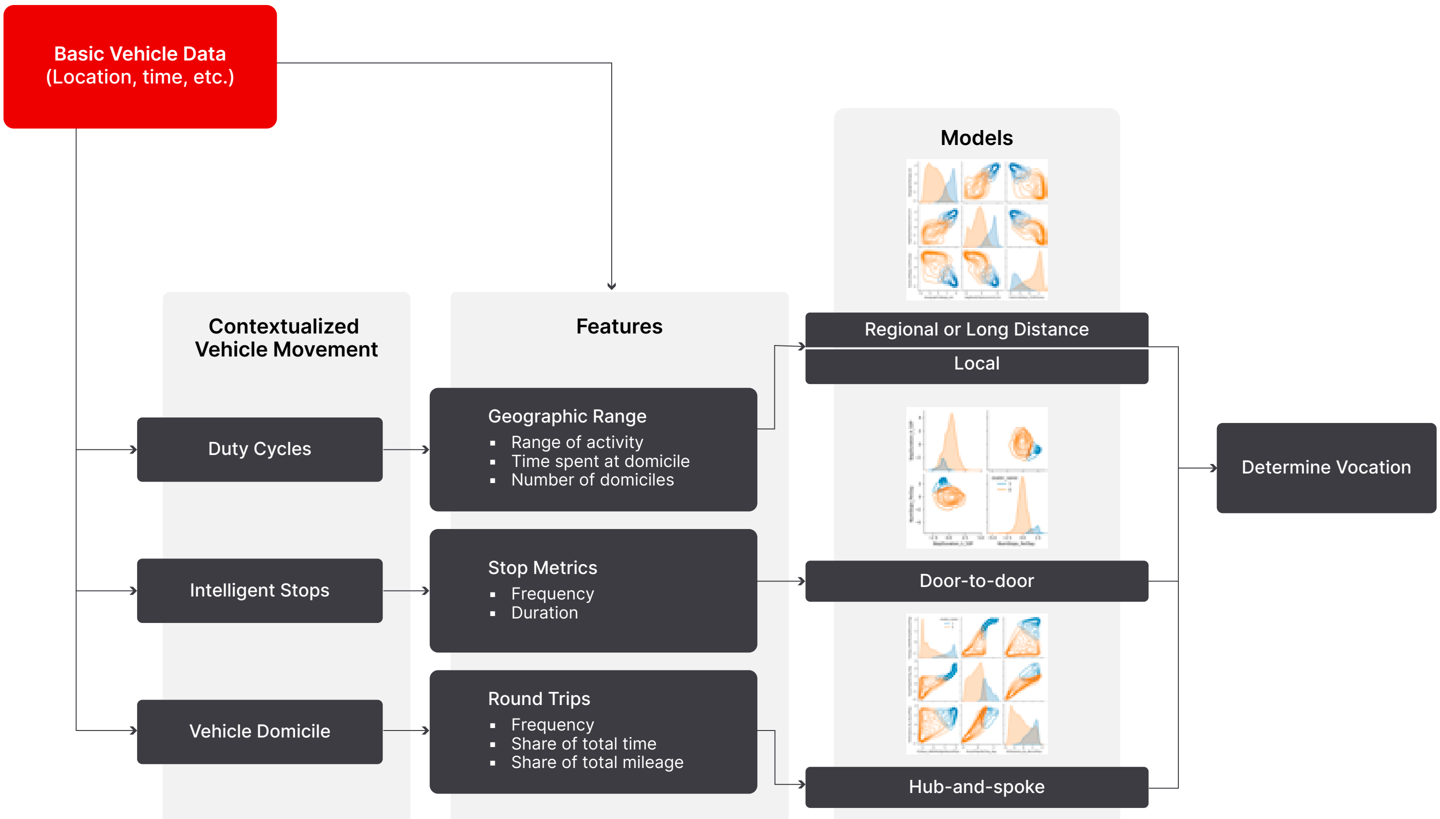


Figure 1

1. Contextualized vehicle movement

1.1 Duty cycles

Motivation: For the purpose of vehicle vocations, idling and stop events are especially important as many door-to-door vehicles make their deliveries without turning off the vehicle, resulting in short idling events.

Implementation: The duty cycle dataset can be thought of as less granular than raw GPS data, and more granular than trips. Every GPS point is given a status: “moving”, “idling”, “stopped”, “unknown”, or “invalid”. Any consecutive series of GPS points with the same status are grouped together into a “duty cycle” of the same status spanning the time window; i.e. moving cycle, idle cycle, stop cycle.

1.2 Intelligent stop locations

Motivation: While examining stop frequency and stop durations by vehicles, there were many cases where a truck makes multiple successive short stop or idle cycles when stopping at a warehouse or work site.

For instance, idling while waiting in line, stopping at the gate, stopping to unload, stopping again to re-load. Since this pattern looks very similar to a door-to-door or last-mile delivery vehicle making multiple deliveries the stop locations data set is used to group close-by successive stops into a single longer stop.

Implementation: To determine if multiple stops are part of the same location, a simple density based clustering algorithm is applied to stopping and idling events that meet defined minimum criteria, while also considering the road types at which the events occurred. The density based clustering considers all stops within a 30 meter radius as the same location.

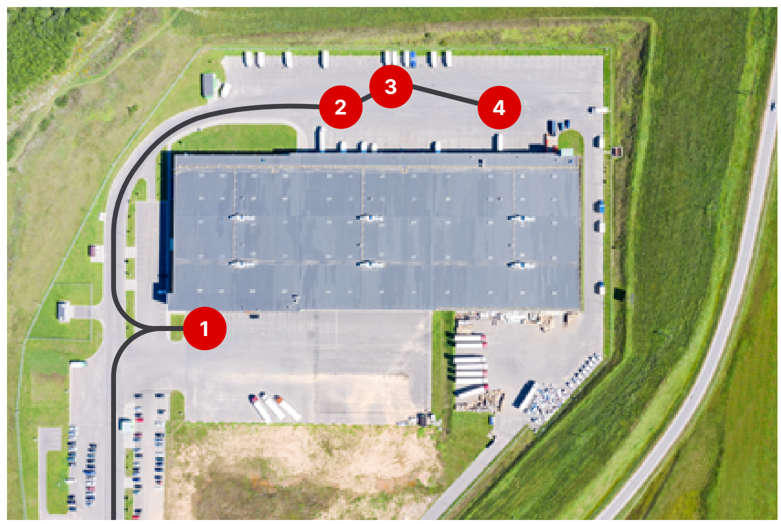


Figure 2: Example of multiple consecutive stops that should be grouped.

1.3 Domiciles

Motivation: A key metric that differentiates long distance vocation (typically long-haul or over-the-road trucking) from local and regional trucking is how often the trucker “sleeps at home”: A regional trucker may return home every weekend or after a few days aways, while long haul truckers may be away from weeks at a time.

To capture this pattern, we would need to determine where a vehicle “domiciles” every work day, and determine how often the vehicles domiciles away from “home”.

Implementation: Using stop locations from (1.2), a vehicles’ domicile can be defined as the longest stop of each day: For instance a city employee parking their issued vehicle in the municipality parking lot between 5pm and 8am would generate a 15 hour stop at their domicile every day.

2. Feature sets

In this section each feature set and corresponding metrics are explained. These describe an aspect of vehicle behavior which is indicative of a specific vocation.

2.1 Geographic range

The geographic range feature set contains metrics related to how far a vehicle tends to travel. This feature is used by the Truck Clustering model (3.1) to determine whether a vehicle is local, regional, or long distance.

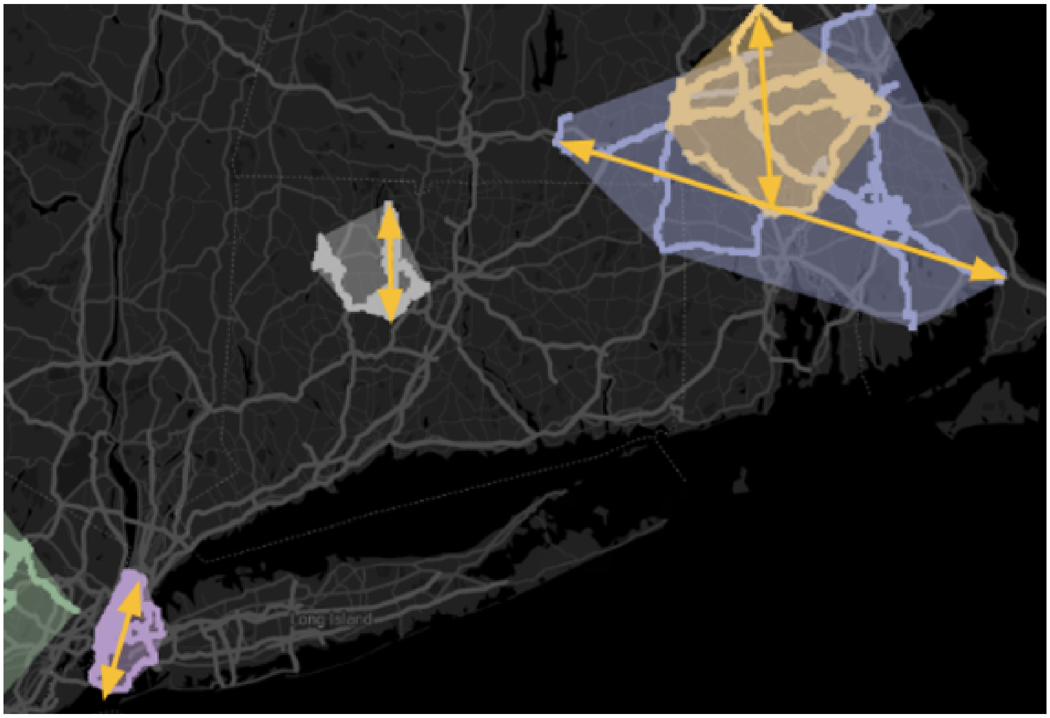


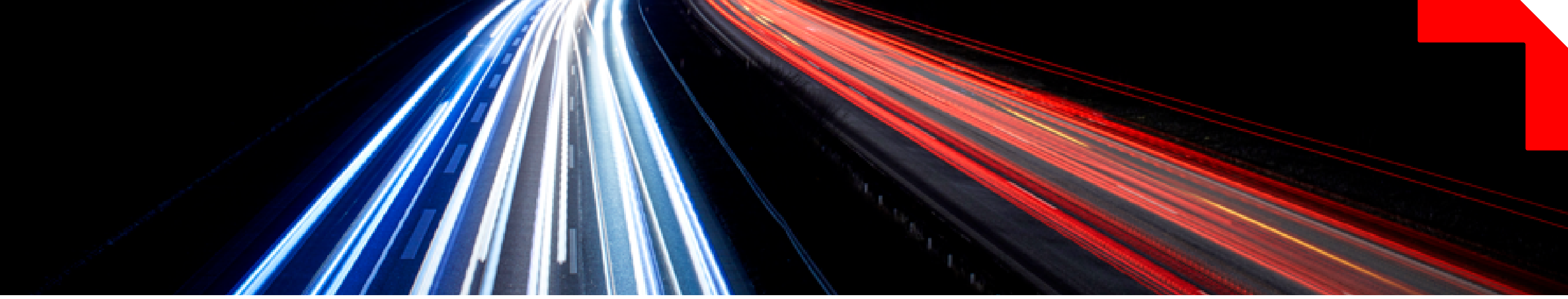
Figure 3: Illustration of geohash6 visited, convex-hull, and geographic range.

To measure how far vehicles travel, this feature set measures two basic elements: the actual geographic span of vehicles' activity and how often the vehicle domiciles at home.

Geographic range is determined by mapping the locations of all duty cycles at a geohash6 level. A convex hull function is then used to determine the smallest polygon that covers all the geohashes visited. The geographic range is then calculated by taking the longest cross-section of this polygon.

In addition to the geographic range, the daily displacement of each vehicle is calculated: Displacement is the crow fly distance between where the vehicle starts its working day and where it ends its working day. For example, a vehicle that returns home every day will have a displacement of 0. This is an important metric to track in addition to range because a local vehicle may take a single long trip, and a regional vehicle may happen to stay in a small region, but still take multi-day trips.

Where a vehicle domiciles is also an important feature related to the geographic range: as long distance vehicles tend to sleep away from home more than regional and local vehicles Considering domiciles allows us to have a geographic-agnostic way of determining what is "far". A long distance vehicle in the Northeastern US may have a smaller geographic footprint than a long distance vehicle in Kansas simply due to geography. If geographic range was the only factor being considered the vehicle in the Northeast may end up being considered regional. By tracking how often a vehicle sleeps at home, the model will recognize that the Northeastern vehicle is also long distance because it spent most of its nights away from home.



After normalization using power transform, the feature distribution across vehicles are as follows:

The three metrics shows two very clear clusters:

- One that tends to stay at home (primary domicile), tends to return to where they start the work day, and generally has lower geographic range: Consistent with local and regional vehicles.
- One that spends less nights at home, does not return to the same location in a day, and has a larger geographic range: Consistent with long distance vehicles.

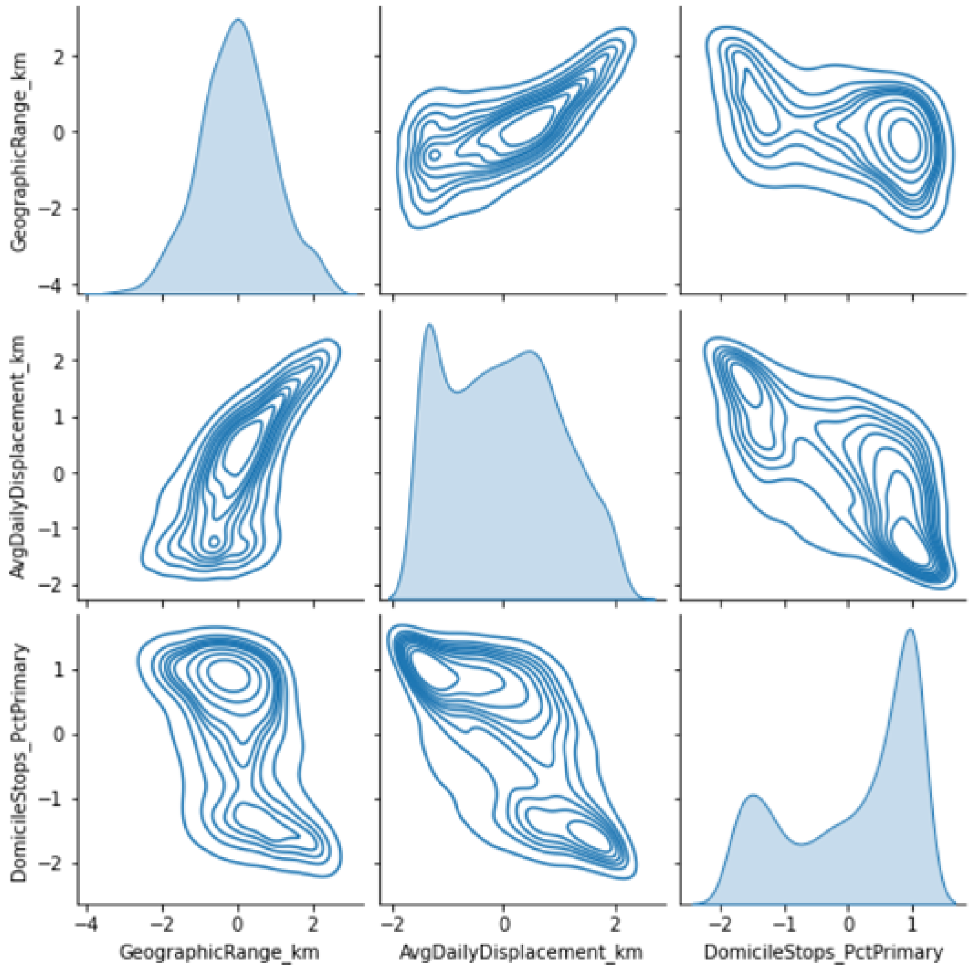


Figure 4: Distribution of range and domicile metrics for regional and long distance.

***NOTE**: To separate local and regional vehicles, a 150-mile (or 241 km) threshold is applied based on the **short-haul exemption** from FMCSA.

2.2 Stop metrics

As described in 1.2 Stop locations filters out stops due to traffic (traffic lights, stop signs, congestion) from stops due to fleet behavior (stop at customer, idling to make a delivery). As such, stop metrics are easily generated from aggregating every stop event.

Considering the number of stops per day and the average stop duration created a clear cluster of vehicles with higher stop frequency and lower stop duration. This is in-line with what we expect from door-to-door delivery vehicles.

This was confirmed by evaluating specific vehicles that identified as door-to-door and confirming that they indeed operate in functions such as, last-mile delivery, waste collection, and postal/parcel services.

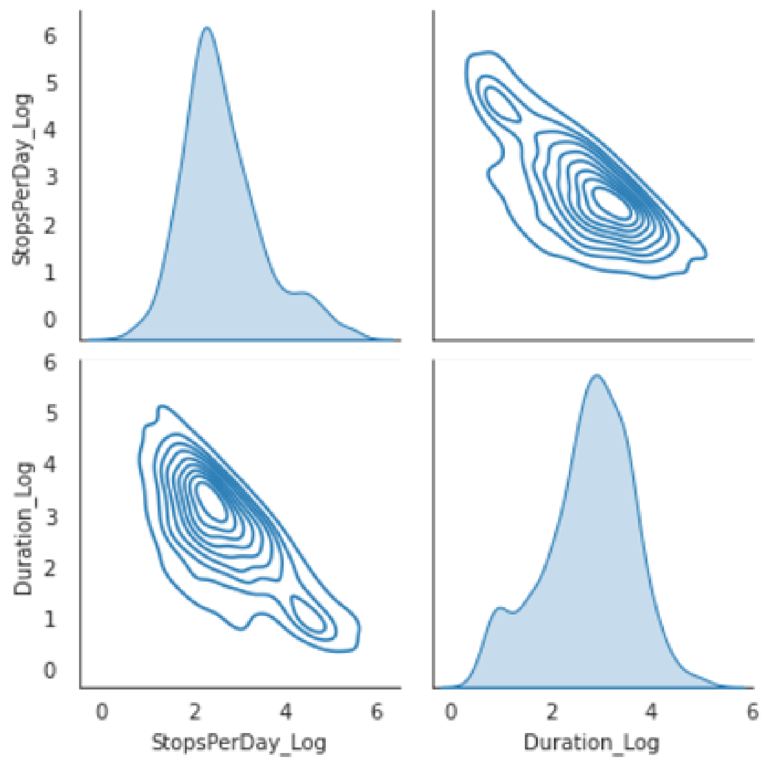


Figure 5: Distribution of vehicles by stop duration and frequency.

2.3 Round-trip-ness

The number of times a vehicle returns to its home base can be used to determine if the vehicle operates in a hub-and-spoke capacity.

With all stops clustered in (1.2) and daily domiciles identified in (1.3), we can now:

1. Define duty cycles by trips - where a stop of any duration defines the beginning/end of a trip.
2. By listing trips in chronological order, these trips can be grouped into larger "round trips" for when the vehicle has returned to domicile in the given day.
3. Using the round trips we count how many round trips were made on a given day, how many stops were made in each, and the total distance traveled in a given round trip.

After transforming the values, we can clearly see two close by clusters and one more distinct cluster:

- One, more distinct cluster, with many work days involving multiple round trips, where those round trips account for most of the total mileage, is very consistent with hub-and-spoke behavior.
- The two, more close by clusters, appear to be local fleets that tend to return home every night– thus having lots of its mileage from round trips but every few multi-round-trip work days. The other, with no round trips at all, are likely to be regional or long distance.

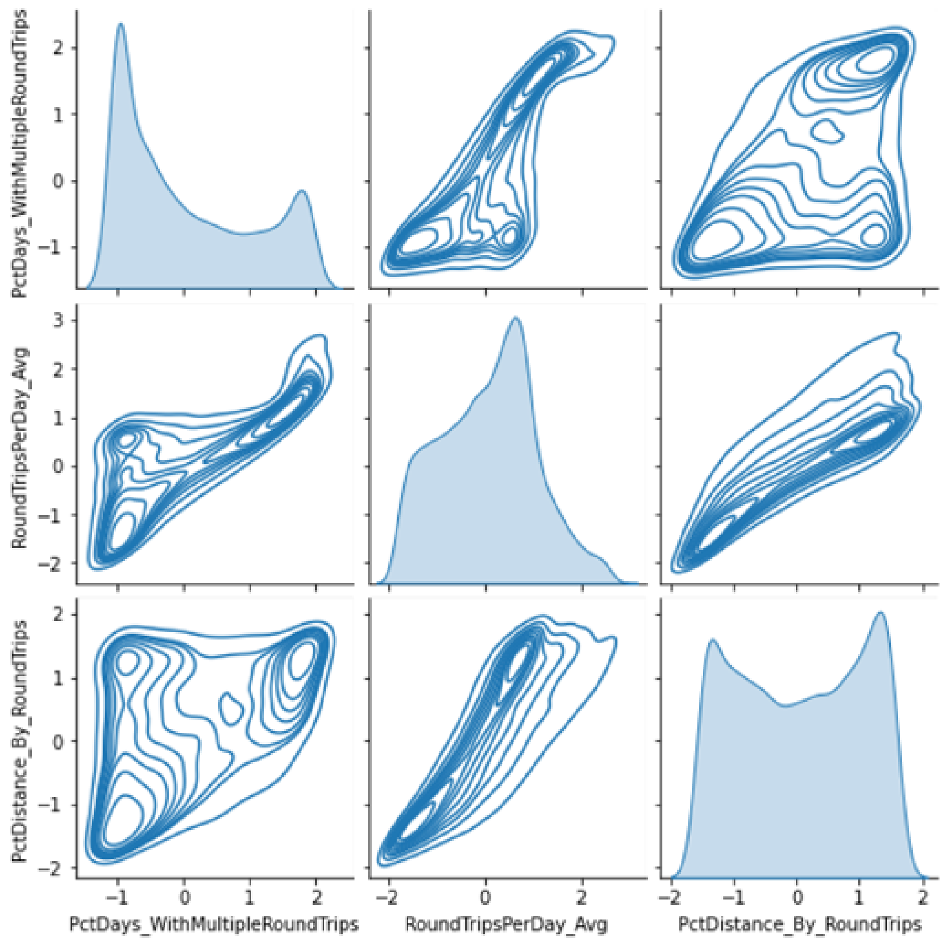


Figure 6: Distribution of round-trip and distance metrics for hub-and-spoke.

3. Models

Each of the three models are Gaussian mixture clustering models. Similar to other clustering models, they are unsupervised, and do not require labeled data, but rather seeks to find a grouping of data points such that differences between groups are maximized. Gaussian mixture models attempt to find a set of Guassian distributions (mean, standard deviation) that can explain the variation seen in the data. Gaussian mixture fits with our initial hypothesis that variations within each vocation are generally Guassian and the entire data set (all vehicles) is the sum of multiple Gaussian distributions.

All three models are similar and this pattern can easily be observed by looking at the resulting clusters of each model:

A) Long-distance detection model

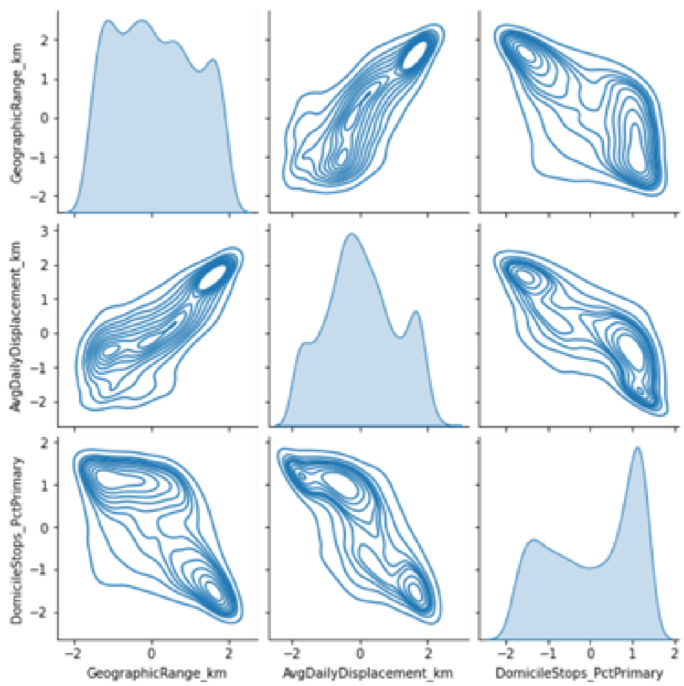


Figure 7: Distribution of vehicles.

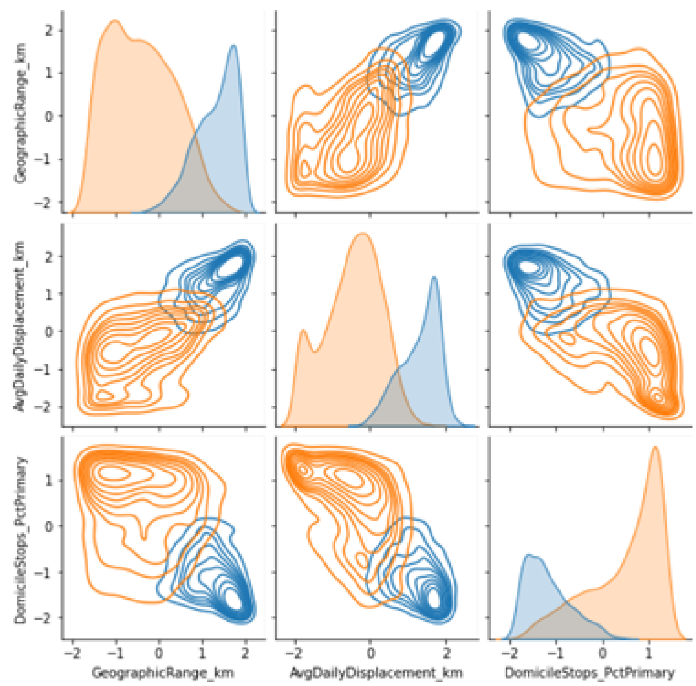


Figure 8: Classification.

B) Last-mile detection model

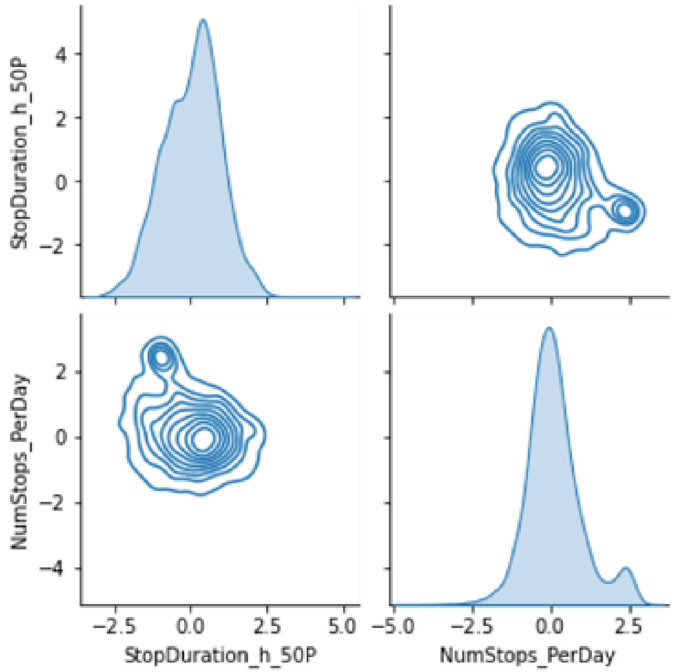


Figure 9: Distribution of vehicles.

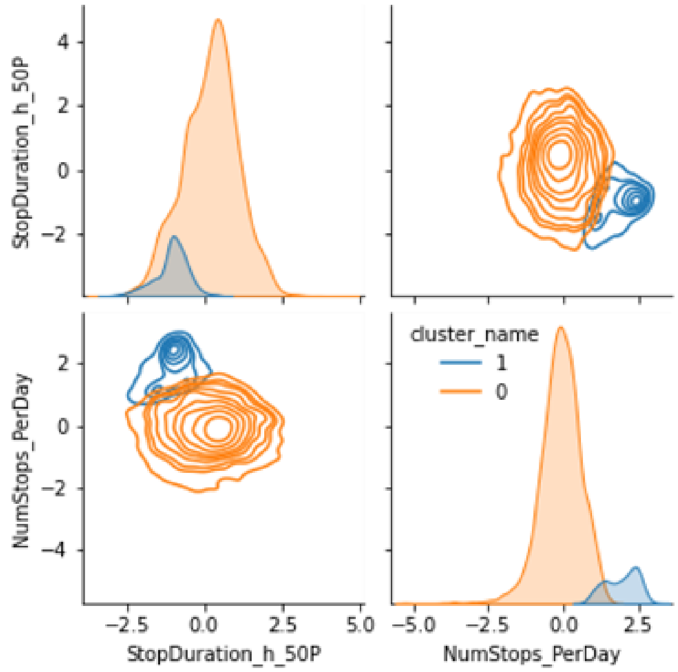


Figure 10: Classification.

C) Hub-and-spoke detection model

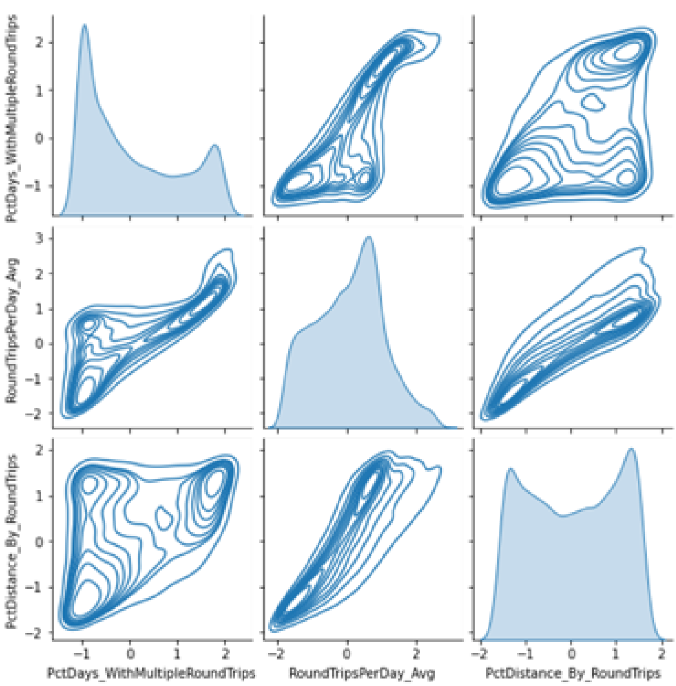


Figure 11: Distribution of vehicles.

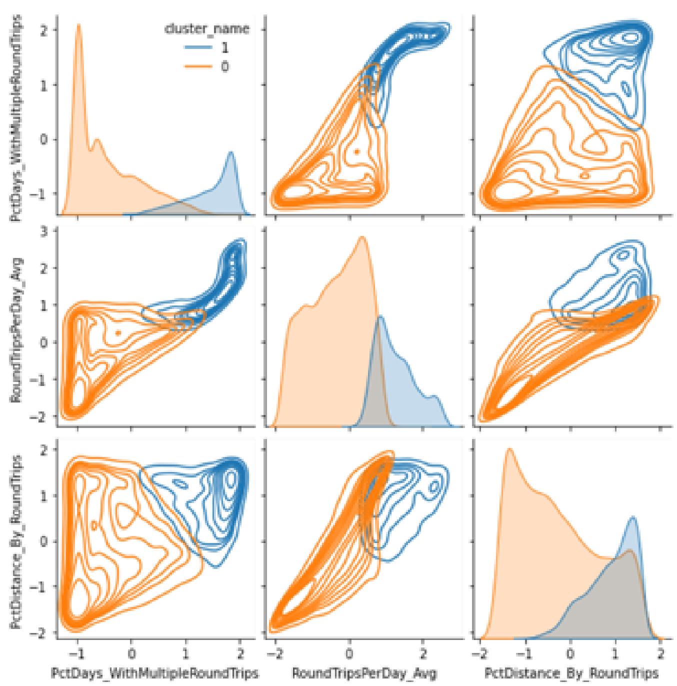


Figure 12: Classification.

How to use vocation data

There are many scenarios where having vocation data would allow transportation planners to make more informed and balanced decisions. For example, a Department of Transportation (DOT) might be interested in understanding how commercial goods movement is impacting congestion on state highways.

In order to accomplish this, the long distance vocation coupled with heavy-duty trucks could specifically be used in a targeted analysis to understand how these particular vehicles are moving through the highway network.

Another scenario might come as a result of the rapid rise in online shopping. Municipalities are increasingly interested in understanding the impact of last-mile delivery in city center areas. To run this analysis in Altitude, the door-to-door vocation coupled with the passenger or light-duty vehicles could be used for a targeted analysis into pinpointing where last-mile deliveries are frequently occurring and what impact they have on congestion and curbside parking.

An example analysis was conducted for the City of Columbus looking at door-to-door parking and delivery events to see the impact to curbside parking.

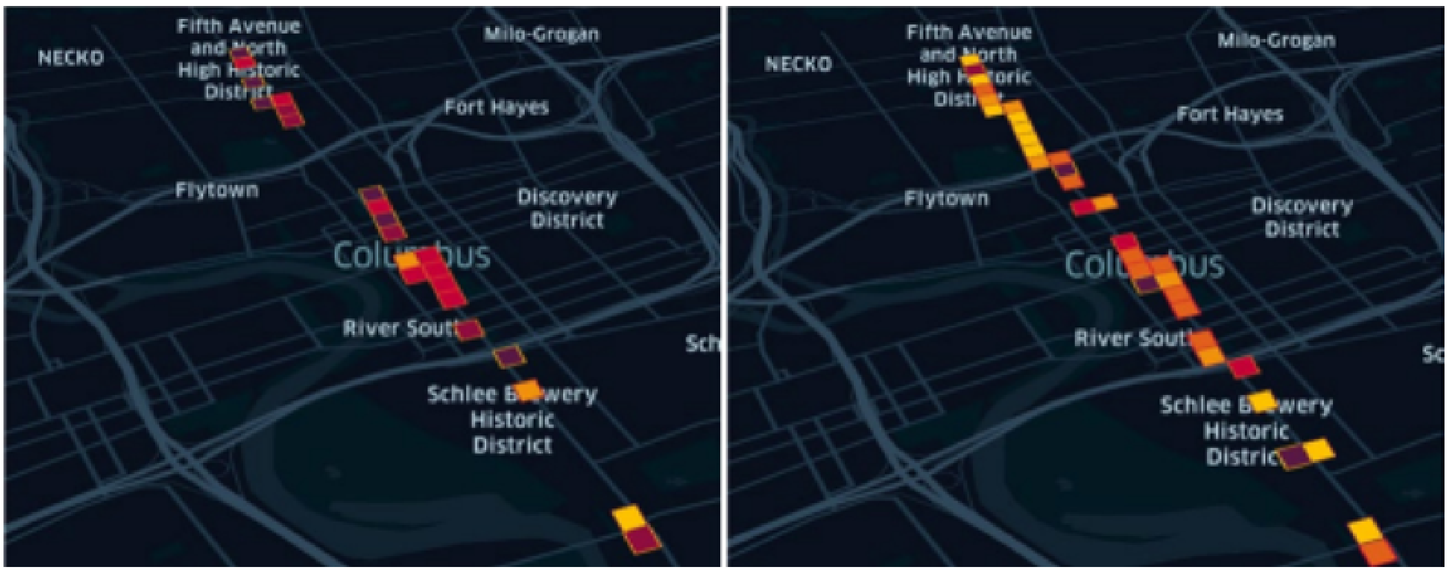


Figure 13: Columbus High Street door-to-door parking events increased 547% between October 2020 and October 2021.

Vocation data can also be used to contextualize a rise in curbside parking from an increase in parking events from door-to-door vehicles. A deeper understanding can be gleaned into how priority curb space is being utilized by understanding the vehicle classes as well as hours of day and days of week when most parking is occurring.

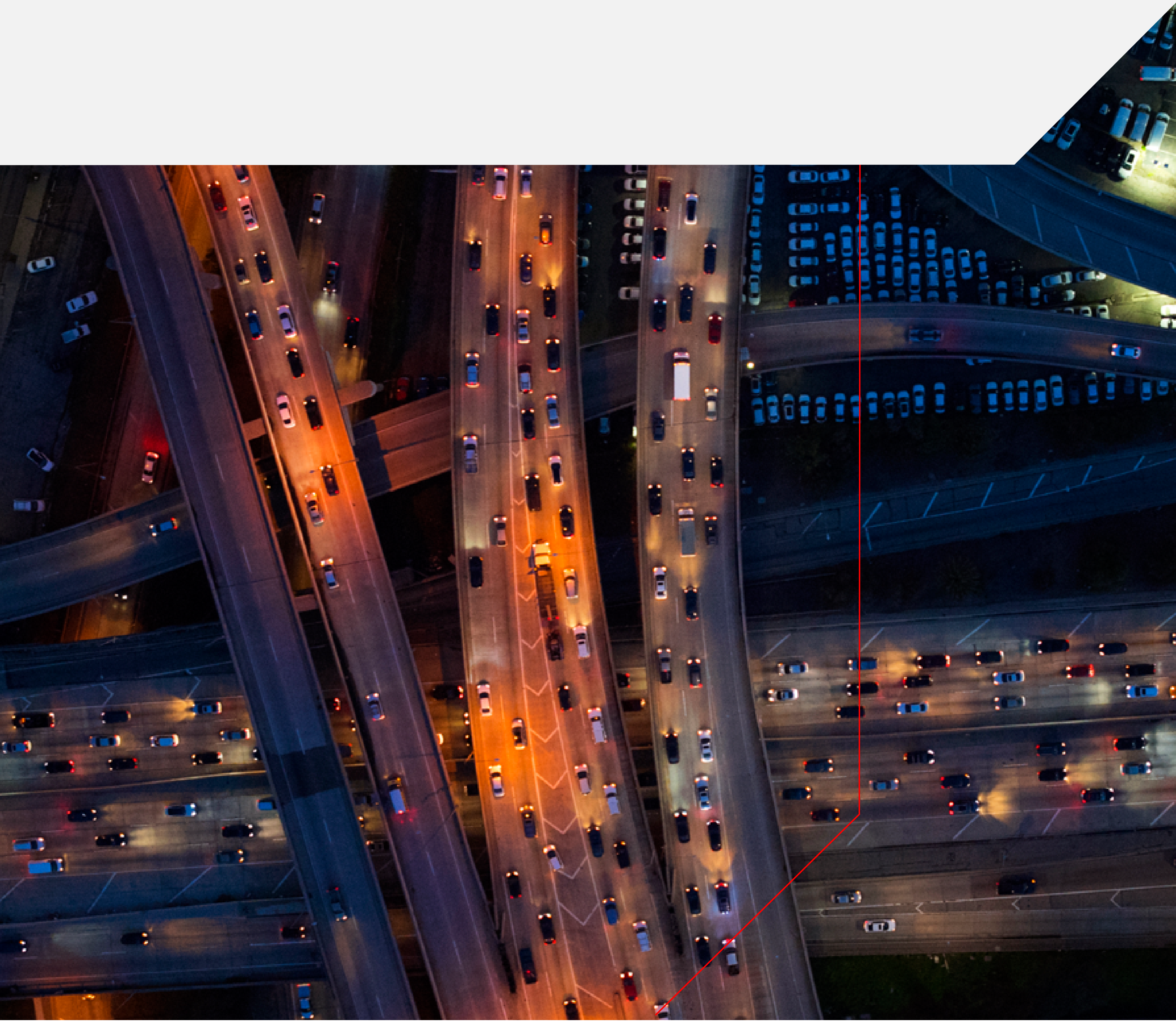
Combining vocation data with trip metrics and origin and destination analyses can lead to valuable insights into which routes are being used to move people versus goods or the types of vehicles being used for specific trips. This data allows for a better understanding into why patterns are emerging and the best course of action to take for required infrastructure changes.

Insights to inform

Vocation data arms transportation planners with deeper insight into how vehicles are being used on their roads. With access to this vehicle behavior data, patterns and anomalies start to become apparent and decisions for the necessary infrastructure to support these different types of traffic become much clearer as well.

Planners that combine trip context with traffic volume will start to transform the future of intelligent transportation, one that supports improved mobility and increases efficiency and productivity.

For more information on how the Altitude platform can provide comprehensive contextual insights for improved transportation planning, visit **its.geotab.com**.



Learn more at its.geotab.com
or email us at its@geotab.com

© 2023 Geotab ITS. All rights reserved.

GEOTAB ITS