Initial TDM Validation Activity: Origin-Destination Core Data

TETC Transportation Data Marketplace Data Validation

Prepared by: TETC Validation Team

5/16/2024

Initial TDM Validation Activity: Origin-Destination Core Data

TETC Transportation Data Marketplace Data Validation

Inaugural Origin-Destination validation activity for TDM Validation Program

TDM-VAL-06

Data Categories*: Origin-Destination (OD)*

Publication Date: *May 2024 (Updated Nov 2024[1])*

5/16/2024

The Eastern Transportation Coalition is a partnership of 17 states and the District of Columbia focused on connecting public agencies across modes of travel to increase safety and efficiency. Additional information on the Coalition, including other project reports, can be found on the Coalition's website: www.tetcoalition.org

[1] Updated Table 1 to clarify vendor capabilities for passthrough/select link filtering.

Executive Summary

This report documents the first Origin-Destination (OD) data validation activity conducted by The Eastern Transportation Coalition (TETC) data validation team and focuses on evaluating OD data from four TDM vendors: AirsSage, Geotab, INRIX, and Streetlight. In contrast to Travel Time/Speed and Volume data, which have well-defined accuracy measures, OD datasets are new to the TDM and have not yet been evaluated for accuracy and quality. Validation of OD data is further complicated by lack of a qualified reference source (or ground truth) and major differences between vendor product offerings.

The general approach used to evaluate OD data is based on recommendations from a comprehensive literature review[2] and feedback from the TDM Validation Technical Advisory Committee (TAC). This strategy included requesting that vendors respond to a series of questions to help clarify their OD data products, investigating vendor OD data quality through both systematic comparisons and internal reasonableness checks, and, in the absence of a "reference data", conducting a side-by-side comparison of all vendors' OD datasets to test for agreement at different spatial scales.

This validation activity is composed of two parts: (1) a vendor questionnaire intended to identify the key aspects of each vendor's product offering, and (2) an analysis of sample data provided by each vendor in Richmond, VA. Given that this report marks the first time acquiring commercial OD data, the analysis focuses on the process of obtaining and interpreting OD data from each vendor, a basic descriptive analysis of the datasets, and preliminary analysis of comparison results. The sample datasets were chosen to coincide with other complementary datasets owned and managed by Viriginia Department of Transportation (VDOT) to support future follow-up analysis.

Several key takeaways emerged from the questionnaire and review of sample data:

- **TDM OD products are diverse, and each vendor reports trip patterns from different perspectives.** Key differences include data sources used (connected vehicle data vs smart phone location-based services (LBS) data), the types of vehicles captured (e.g., mixture of all vehicles, freight only), how trips are quantified (e.g., person-trips vs vehicle-trips), trip mode, and whether report trip counts reflect observed sample probes or population-level estimates. Furthermore, *each vendor uses slightly different logic to split GPS waypoints into separate trips, so the same GPS trajectory may result in different trip definitions*. These nuances should be taken into consideration when comparing vendor results to each other or to external data sources.

- **All vendors allow users to easily create custom OD queries, including support for user-defined geographic zones, customizable time periods, and other vendor-specific filters.** Although the query process differs across vendors, all vendors provide a straightforward way to produce meaningful customized OD datasets that meet necessary spatial, temporal, and other criteria common to most transportation analysis.

- **Understanding query options are critical, especially those that determine trip definition**. Specifying query parameters is one of the most challenging aspects of obtaining OD data, as these settings control query results, as well as how output data should be

[2] TDM-VAL-01: Traffic Volume Validation – Literature Review and Recommendations (link)

interpreted (e.g., sample counts vs population estimates, total counts for study period vs daily averages), how trips entering/exiting the study area are handled, and in some cases, how trips themselves should be defined. Parameters that determine trip definition, if exposed, are particularly critical.

- **Preliminary analysis of sample data from Richmond, VA at the county-level shows intuitive OD patterns for all 4 vendors**, including similarities in top Origins, Destinations, and OD pairs -- despite key differences in the types of trips captured by vendor products. Inspection of trip counts, average travel time and average trip distance for different scenarios (e.g., time of day, trip purpose) also yielded results that appeared reasonable based on local knowledge.

Table of Contents

Introduction

Transportation data sold through the Eastern Transportation Coalition (TETC) Transportation Data Marketplace (TDM) is procured from private industry based on contract specifications. The intent of the Coalition's validation program has evolved from the original Vehicle Probe Project validation which was primarily limited to ensuring that traffic data conforms to contractual standards for speed and travel time accuracy. Although the TDM still includes that essential function, it also has the flexibility to adjust to the needs of the Coalition members as the market evolves and data needs expand. The validation process is overseen by a technical advisory committee that sets general direction and reviews results. The TDM includes both quantitative and qualitative analysis of datasets available through the marketplace as appropriate for each data type. The marketplace currently contains six core data items: Travel Time/Speed, Volume, Waypoint, Origin-Destination, Freight, and Conflation, with all but one (Travel Time/Speed) being sold through the marketplace for the first time. As such, the validation team, under the guidance of the TETC Validation Technical Advisory Committee (TAC), is establishing benchmarks and methods for effectively evaluating data quality and value applicable to the different data sets.

This validation report evaluates one of the products sold in the TDM: Origin-Destination (OD) data. Based on the TAC's direction to focus on OD data, the validation team previously produced a report, TDM-VAL-3[3], containing a literature review and recommendations for validating OD data. The reader is encouraged to review this document, as the strategy outlined in TDM-VAL-3 guides the methodology used to evaluate vendor OD data in this report

This validation activity is composed of two parts: (i) a vendor questionnaire intended to expose important key aspects of each vendor's product, and (ii) an analysis of sample data provided by each vendor in Richmond, VA. The latter focuses on how to access and interpret vendor datasets, a descriptive analysis of the OD datasets, and preliminary comparative analysis of results across vendors. The study area was chosen to coincide with other complementary datasets owned and managed by Viriginia Department of Transportation (VDOT) to support follow-up analysis.

Given that this marks the Validation team's first look at commercial OD sold through the Traffic Data Marketplace, the focus of this report is to inform Coalition members to better understand OD data offerings, understand the extent to which commercial offerings show internal agreement, and introduce comparative methods and visualization to utilize in future OD studies.

Data Vendors

All four vendors that were selected through the TDM RFP process in the Origin Destination (OD) category participated in the study: AirSage, Geotab, INRIX, and Streetlight. Each vendor was highly engaged in the process, provided timely responses to a questionnaire and assisted with obtaining and interpreting the sample data for the analysis.

---

[3] Origin-Destination Validation - Literature Review and Recommendations (tetcoalition.org)

Part 1: Overview of Vendor Products

Although the four OD vendors are all selling products under the "Origin Destination" category, the products vary significantly in terms of underlying data sources, the types of trips that are captured, and other characteristics, reflecting the broad definition of OD outlined in the TDM RFP.

As such, a questionnaire was prepared by the validation team in coordination with the Technical Advisory Committee with the goal of uncovering the key aspects of each vendor's Origin-Destination product. The survey was divided into 4 sections, summarized below:

- **Methodology**: Questions about input data sources, travel modes captured, modeling processes, known biases
- **Data Characteristics**: Questions about spatial and temporal granularity, query options, data availability, and whether trip statistics can be provided.
- **Data Access / Privacy**: Questions about data access, formats, and privacy
- **Data Resiliency**: Questions about ability to handle disruptions to input data sources

Table 1 lists the responses from each vendor in a compact, comparable format, while the following subsections provide brief summaries of each vendor offering. *Note that these summaries are intended to orient the reader to each vendor's OD offering at a high level*. Table 1 contains details about specific OD data characteristics (e.g., specific spatial and temporal filtering options), and a more detailed discussion of data formats is reserved for Part 2.

AirSage

AirSage offers two types of OD products -- one based on Location Based Service (LBS) data acquired from smartphone apps and the other based on Connected Vehicle (CV) data. The former captures *person trips* across any mode and the latter represents *vehicle trips* – both of which reflect *population-level estimates* as opposed to reporting sample data. AirSage's definition of a trip end is when a device registers in the same location for 15 minutes, meaning that if a device dwells at a specific location for less than 15 minutes, it is determined to be only an intermediary stop, whereas if a device dwells for longer than 15 minutes, the current trip is terminated and a new trip is initiated once the device begins to move again. This setting can be adjusted per project requirements.

Spatial and temporal filtering options are vast; users can choose common zone geographies (e.g., MSA, county, TAZ, census track / block group / block) or define their own, and can query by day-of-week, time-of-day, and even define custom time periods down to 30-minute bins for the specified study period. The LBS-based OD product can be filtered for trip purpose (home / work / other), and trip statistics (e.g., trip length, distance) can be produced alongside OD data upon request (using a combination of LBS and CV data). Trip purpose is inferred by assigning 'home' and 'work' locations for each device based on GPS sightings over a period of time (via proprietary algorithm) and determining whether trips start or end at these locations (e.g., Home Based Work (HBW), Home Based Other (HBO), Non-Home Based (NHB)). Note that AirSage's LBS-based OD product continues to be available despite significant disruptions in the LBS data market dating from April 2022.

To access OD data, customers begin by meeting with an AirSage representative to discuss requirements and ensure that query parameters are chosen appropriately to align with study goals. Afterwards, OD data are made available for download in comma-separated value (CSV) format.

Table 1 - Summary of questionnaire and follow-up responses

	AirSage	Geotab	INRIX*	Streetlight
Methodology				
Data Sources	- LBS (measures person-trips) - CV (measures vehicle-trips)	CV: GPS data from commercial fleet vehicles' onboard telematics devices	- CV - LBS (Mobile devices)	- CV (personal vehicle) - CV (commercial fleet) - LBS (personal vehicle) - LBS (bike, ped, bus, rail)
Definition of trip end	Customizable. Default setting is when a device registers in the same location for 15 minutes.	Customizable. Default setting is "ignition off" or vehicle idles for 200 seconds.	Device does not move 200m within 10 min.	CV: "Ignition off" or vehicle moves less than 5m in 5min
Travel Modes	- LBS product encompasses all modes (cannot segregate) - CV product is vehicles-only (cars / trucks / buses)	Commercial vehicles only (can segregate by vehicle class, including detailed GVWR truck class)	Walk / vehicle / Unknown (Vehicles broken down by light/med/heavy duty)	Depends on data source and time period: - All vehicle, - Commercial vehicle (med/heavy duty) - Bike, Ped, Bus, Rail
Observed samples vs Population estimates	Population estimates	Observed samples (Expansion factors also provided at continuous count locations)	Observed samples	Population estimates (sample size also provided)
Known biases	Slight biases based on smartphone usage (income, age) and privacy laws. Efforts made to minimize bias.	100% commercial fleets (but diverse mixture of industries, fleet size, etc.)	None reported	None reported
Data Characteristics				
Average vs total trips	Either	Either	Total	Average
Trip Purpose	- LBS product: Home / Work / Other - CV product: N/A	Vocations: Door to door / Hub and Spoke / Local / Long Distance / Regional	N/A	- LBS product: Home / Work / Other - CV fleet: Geotab vocations designations
Standard zone geographies	Common zones include MSA, county, census tract / block group / block, TAZ.	Common zones include county, city, zip, TAZ, census tract, road segment	Lat/long of trip start/end provided (allows mapping to any standard zone)	Common zones include county, city, zip, TAZ, census tract/block group, road segment, transit stop/line
Custom zone geographies	Yes	Yes	Same as above	Yes
Passthrough / Select-link filtering	Yes	Yes	Yes**	Yes
Boundary handling	Only trips in study area are included. Trips entering/leaving boundary area can be measured by adding "external" zone/s (extending study area)	Trips entering/leaving boundary area must have origin, destination, or connector (passthrough) zone in the study area to be included. If using custom zone definitions, trips that	All trips that start, end, or pass through a region will be included. Lat/longs reflect actual trip origin or destination.	Trips entering/leaving the study area can be captured by define passthrough zones along boundaries or adding "external" zone

		start/end outside study area would need an additional "external" zone to be defined.		
Intersection-level OD	Not natively supported for TDM (can do on a project basis)	Yes, by specifying road segments as origin	No	Yes
Time aggregation	Time-of-day (hourly), day-of-week for specified date range	Time-of-day, day-of-week for specified date range	Individual trip start/end times provided. Can be aggregated as desired	Time-of day, day-of-week for specified date range
Custom time periods	Yes, down to 30-minute intervals	Yes	Yes (see above)	Yes, down to 15-min intervals
Data delivery frequency	Monthly (4 weeks after close of month)	API request executes in seconds, with data returned in minutes/hours. Data available for analysis 3-4 days after trip occurs	Data available with 72-hour lag after trips occur. Historical requests delivered in bulk.	Data available with 1–2-month lag (depending on mode). Historical requests go back to January 2016.
Trip statistics provided	Yes, upon request (requires combining LBS and CV data)	Yes	Yes. Basic stats available, and user has enough trip-level info to calculate.	Yes
Data Access / Privacy				
Data request	Consultation with vendor	Web UI or API	Consultation with vendor (Web UI if using via Trips Analytics)	Web UI
Data delivery format	Direct download link (csv file). See appendix for data schema.	- Web UI: CSV file download - API: Response as JSON format (see appendix for links to documentation)	CSV / Parquet files via S3	CSV / Shapefile download
Privacy considerations	- Device-level information not provided. - Data aggregated to level that is not PII	Data anonymized and aggregated such that individual trips are not reported (no PII)	- Trip start/end locations are tied to center of quad key - Comply with PII reqs (including Europe & CA)	- Data reflect groups of people, not individuals (no PII)
Resiliency				
Vulnerability to loss of data source	- Various data partners - Not overly reliant on single provider	Data comes from fleets that Geotab manages, so not reliant on external data sources	Not dependent on a single data source	- Many data providers, constantly being evaluated - No mode is reliant on a single provider
Level of disruption if primary source were unavailable	- Has weathered losses in the past and demonstrated ability to recover quickly (e.g., LBS data disruption in April 2022)	N/A (see above)	Minimal -- do not have one primary source and could still create OD info.	Minimal – have secondary sources under contract as contingency

***** INRIX's questionnaire responses are based on their flat file delivery format. Responses may be slightly different with their alternative method for accessing licensed OD data (Trips Analytics platform by CATT Lab).
** Requires purchase of Trip Paths dataset

Geotab

Geotab's OD product is based primarily on GPS data collected by its parent company, which deploys onboard telematics devices in commercial fleet vehicles. Their OD product represents actual commercial *vehicle-trips*; no modeling is done to estimate population-level OD patterns, although expansion factors to estimate commercial vehicle trips are provided at locations where reference traffic counts are available. Geotab's default definition of a trip end is when the ignition turns off or the vehicle idles for 200 seconds or more. However, Geotab allows the user to customize the trip definition by exposing a parameter in the web interface and API (e.g., to allow vehicles that idle for 15 minutes to be considered part of the same trip).

As summarized in Table 1, a full range of spatial and temporal filtering options are available; users can choose common zone geographies or define their own, and can query by day-of-week, time-of-day, or other custom temporal periods for analysis. Data can also be filtered by vehicle class (e.g., light / med / heavy duty trucks), vocation (e.g., long distance, regional, door-to-door) and industry (e.g., agriculture), which help provide insight into different types of commercial freight activity. Geotab is able to provide this level of granularity because their data are based on known fleet vehicles.

Geotab provides two primary ways to run queries: (1) a web interface (2) via API. The web interface, called Altitude, provides all necessary functionality to run queries, including creating custom O/D zones. Query results are presented visually through the UI and data can be downloaded for further analysis. The API provides programmatic access to the same query functionality, which may be particularly useful for large queries or automating data pulls. To make API queries easier, the user can set up a query using the web interface and then download the query settings as a JSON file to pass to the API.

INRIX

INRIX's OD product is based on GPS data sourced from connected vehicles and mobile devices, with a self-reported vehicle mixture of 89% passenger vehicles, 6% local fleet, and 5% long haul truck (note that this was prior to June 2023, when Wejo, a TDM vendor and connected vehicle data supplier, declared bankruptcy). The OD data represent *observed* trips and are not estimates of population level travel patterns. INRIX defines the end of a trip when a device does not move 200 meters within a 10-minute period.

Data can be filtered spatially and temporally to obtain desired subsets of the data; however, the specific filtering processes depend on how the data are acquired. To access data, the customer starts by communicating the desired study area – including possible custom zone geographies -- to the sales team, after which INRIX can share data via (1) direct file download or (2) through a web interface to the Trips Analytics platform via the University of Maryland Center for Advanced Transportation Technology Laboratory (CATT Lab).

If the web interface is used, users can apply spatial, temporal, and other filters (e.g., vehicle weight class) through the UI to generate and download OD matrices. However, if a direct download is used, INRIX provides a file that contains a record for each trip in the dataset. Each record contains start/end location and start/end timestamp as well as other attributes, enabling the user to map each trip to any spatial geography or temporal period. This level of granularity requires more data processing but enables a higher level of control by the user.

Streetlight

Streetlight Data offers OD products across several different modes (personal vehicles, commercial vehicles, bicycles, pedestrians, bus, passenger rail), which use different data sources. Broadly speaking, these products use either LBS data from mobile devices (LBS-based personal vehicles, bicycles, pedestrians, bus, passenger rail) or CV data (CV-based personal vehicles and commercial trucks). Spatial and temporal filtering options are highly customizable; users can choose common zone geographies or define their own, and can query by day-of-week, time-of-day, and define custom time periods down to 15-minute periods. For CV data, Streetlight defines the end of a trip when the ignition is turned off or the device does not move more than 5 meters within 5 minutes.

Certain query options depend on what mode and data source are used; for example, LBS-based OD product scan be filtered for trip purpose (home / work / other), while commercial trucks can be filtered by medium/heavy duty, or if Geotab data are included (an option for Streetlight data), Geotab's vehicle vocation designation. It is worth noting that the availability of LBS-based products has been impacted since April 2022, when the LBS market was significantly disrupted by changes in data policies at Apple and Google to better protect user privacy. As such, certain products such as LBS-based personal vehicle queries are only available prior to April of 2022.

A web interface, called Streetlight Insights, provides all the necessary functionality to run queries, including creating custom O/D zones. Query results are presented visually through the UI and data can be downloaded for further analysis (CSV format).

Part 2: Evaluation of Sample Data

Sample Data Request

Vendors were asked to provide data in Richmond, VA according to the following spatial and temporal filters. The specific details of how data were acquired varied by vendor.

- **Spatial Filter**: Custom Traffic Analysis Zones (TAZs) were chosen as the spatial geography for the sample data request and were selected to (1) be small enough to investigate granular trip patterns and (2) align with the zonal structure of Virginia DOT's travel demand model in Richmond. Figure 1 shows the 1203 TAZs near Richmond used for the query, which encompasses the following filters:

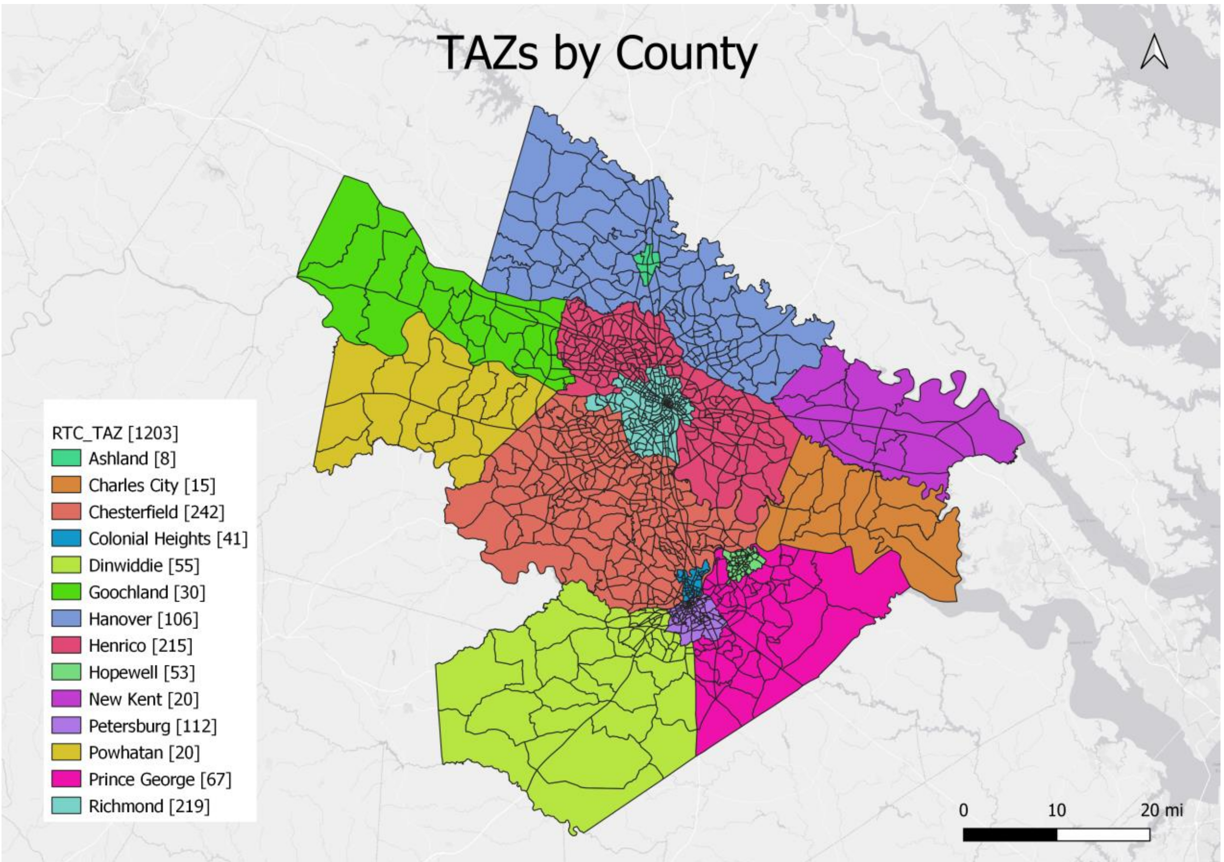

Figure 1 – Model TAZs near Richmond, VA, color coded by county.

- **Temporal Filters:**
  - *Date Range*: October 2022 was chosen to capture a full calendar month during the Fall year when schools are in session and there are minimal holidays in order to capture normal trip patterns

- *Day Type:* Results were segregated to Typical weekdays (Tues – Thurs), Weekends (Sat – Sun), and All days
- *Day Parts*: All day, AM Peak (6-9am), PM Peak (3:30-6:30pm), Off-peak (6:30pm-6:am)

- **Other filters** (varies by vendor). Each vendor provided slightly different filter options based on their product offerings. Examples included:
  - Trip purpose
  - Vehicle class

- **Trip statistics:** Additional information corresponding to the underlying trips was requested, including travel time and trip distance.

Data Delivery and Standardization

OD data was acquired through slightly different processes for each vendor, the details of which are described below. After querying, each vendor's dataset was standardized into a common format to support comparative plotting and analysis routines. This standardization process looked slightly different for each vendor; for some vendors it only required renaming columns, while others involved temporal aggregation or turning "total trip" counts into average daily values to make the datasets as comparable as possible. The final standardized data format includes the following fields:

- **vendor_name:** Name of vendor
- **origin_zone:** Origin zone name (corresponds to shapefile model TAZ id)
- **dest_zone:** Destination zone name
- **veh_class**: Vehicle classification (optional, vendor-specific)
- **trip_purpose:** Trip purpose (optional, vendor-specific e.g., HBW, HBO)
- **day_type:** Day type aggregation.
  - *Typical Weekday (Tue-Thu, Non-Holiday)*
  - *Weekend (Sat-Sun)*
  - *All (Mon-Sun)*
- **day_part**: Day part aggregation
  - *AM Peak (6-9am)*
  - *PM Peak (3:30-6:30pm)*
  - *All* day
  - *Off peak (XX am/pm – YY am/pm)*
- **numtrips_obs:** Number of daily **average observed trips** between zones for specified day_type and day_part. (not every vendor may report this)
- **numtrips_popest**: Number of daily **average estimated trips** between zones at the population level for specified day_type and day_part. Some vendors may not report this.
- **avg_tt:** Average travel time in minutes
- **avg_dist**: Average trip distance in miles

Key considerations when standardizing include the following factors:

- Two vendors (INRIX and Geotab) report the number of observed sample trips, while the two others (AirSage and Streetlight) report estimates extrapolated to the population level.

- Two vendors (INRIX and Geotab) report the total number of trips over the study period, while AirSage and Streetlight report average values per temporal period. To make count values comparable, average values are calculated and used for all vendors.

Even after standardizing, it is important to note that the resulting OD patterns and corresponding trip statistics are not directly comparable; each vendor is reporting different types of trips (e.g., person-trips vs vehicle-trips, passenger-vehicle vs freight-only). In particular, Geotab is distinct from the others in that it exclusively captures commercial freight trips.

AirSage

The validation team met with AirSage to discuss the parameters of the sample data request and shared a shapefile containing zone definitions. Based on this conversation about project goals and desired temporal and spatial scope, AirSage prepared a custom dataset to meet the requirements of the study. The delivery involved sharing two CSV files through an online link (weekday and weekend results).

The data provided for this exercise is based on AirSage's LBS-based product, which captures person-trips across any mode and represents population level estimates. Figure 2 shows a screenshot of the data format, which reports trip counts between origin and destination zones – broken out by weekday (WD) / weekend (WE), time of day, trip purposes, and home location. AirSage has several options for reporting trip purpose, but for this dataset, 3 are used: Home Based Work (HBW), Home Based Other (HBO), and Non-Home Based (NHB). Based on the chosen settings, the trip counts represent **averages** over the study period for the specified conditions.

Origin_Zone	Destination_Zone	Home_Zone	Start_Date	End_Date	Aggregation	Purpose	Time_of_Day	Count
584.0	582.0	51085320803.0	20221004	20221027	WD	HBO	H04:H05	8.8
257.0	156.0	51087200414.0	20221004	20221027	WD	HBO	H05:H06	8.5
877.0	1046.0	51760060700.0	20221004	20221027	WD	NHB	H02:H03	39.03
209.0	209.0	51760060700.0	20221004	20221027	WD	HBO	H09:H10	64.54
327.0	1252.0	51149850400.0	20221004	20221027	WD	HBO	H08:H09	16.61

Figure 2: AirSage raw data delivery format

There are a few points worth highlighting about the AirSage dataset. The first is that only trips that start *and* end in the study area (i.e., the Richmond model TAZs) are included in the dataset; trips that enter or leave the study area are not included. If these trip types are important to the user, the way to include them would be to define an additional zone outside the study area and label it as "external". Depending on the type of analysis of interest, this may or may not be important. The second point is that AirSage's LBS-based product does not come with trip length or travel time metrics by default (although it can be created upon request by integrating the CV dataset). The validation team did not consider this for inclusion at the time of the data request, so the data delivery does not include these metrics (although the validation team computed trip length as the distance between TAZ centroids). Finally, because the validation team's data request specified hourly temporal granularity, PM Peak and Off-Peak definitions were adjusted slightly (PM Peak: 3pm-6pm, Off Peak: 6pm-6am).

Overall, the process of acquiring data was streamlined. The up-front meeting with AirSage helped clarify issues right away, and the fact that AirSage runs queries on the user's behalf means that the user does not have to be concerned about understanding nuanced (and potentially impactful) parameter settings. However, the disadvantage is that it is harder to iteratively experiment with different query options, as any changes would require AirSage to create a new custom dataset.

Geotab

Geotab provides two methods for users to query and download data: a web application/analytics platform and an API. The validation team used the web interface for this exercise, as it provided a straightforward way to define custom spatial zones and specify query options. However, the ability to use an API offers significant flexibility and opportunities for efficiency if queries need to be made regularly. Figure 3 shows a screenshot of the Altitude user interface.

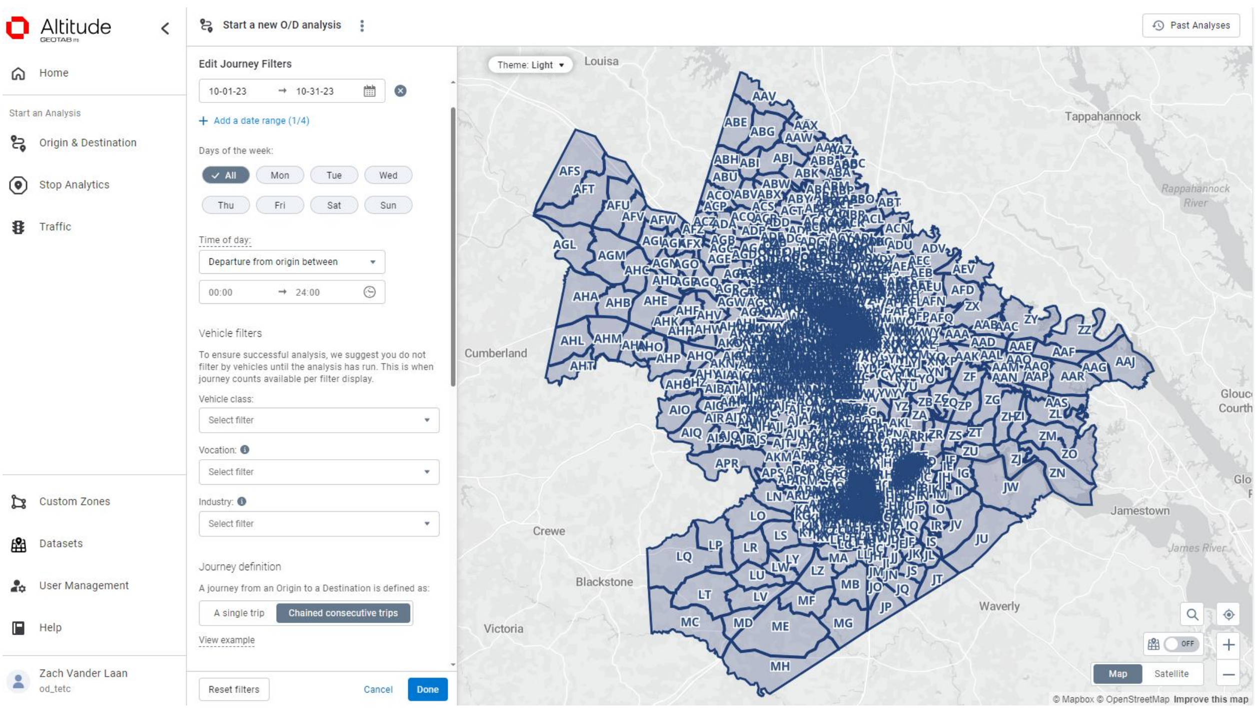

Figure 3 – Geotab web interface

Using the Altitude web platform, the first step was to define custom zone geographies for analysis. Geotab does provide standard geographies, but since the analysis focuses on VDOT's model TAZ zones, the zone definitions needed to be customized. Custom zones can be defined by drawing a bounding box on an online map or uploading a shapefile, GeoJSON file, or CSV file containing valid polygon geometries. After custom zone definitions are validated, they can be used as the geography for OD queries.

The query process is straightforward, but there are many available filters and parameter choices that can impact results. One of the most important parameters relates to how trips (or in Geotab's terminology, journeys) are defined. Geotab defines a trip when a vehicle starts moving until the ignition is turned off or the vehicle idles for 3 minutes 20 seconds. However, the user has the ability to chain trips together by ignoring stops smaller than a specified threshold, plus configure other aspects of trip chaining within origin and destination zones. The ability to define trips in this manner is unique; trip-making logic was not user-configurable in most other vendor query interfaces. However, care should be taken to define trips in a meaningful way because trip logic can significantly impact the results (see results section below).

Additionally, there are many types of freight-related filters available, indicating that Geotab has deep insights into the types of vehicles used, the types of trips they make, and the industries involved. Specific filter options include the following:

- **Vehicle Classification**, broken down into *Trucks* (light / medium / heavy-duty), *Non-Trucks* (passenger vehicles, multi-purpose vehicles, buses), and *Other Classes*.
- **Vocation** (descriptions below are taken from Geotab UI)
    - *Long distance* (e.g., Long Haul, Rental & company Vehicle) – Long travel distances, not often resting in the same locations
    - *Regional* (e.g., Building Supplies, Fuel Carrier) – Wide range of distances, often resting in the same locations
    - *Local* (e.g., HVAC, Beverage Distribution) – Operates within 150 mile radius of reporting location
    - *Door to Door* (e.g., Merchandise Deliver, Trash Collection) – High volume of short stops in a work day
    - *Hub and Spoke* (e.g., On-demand Goods, Auto-part Delivery) – Makes multiple round trips from central hub in a work day
- **Industry**: 20 different industry filters, examples of which include *Agriculture*, *Forestry*, *Fishing and hunting*, *Construction*, *Manufacturing*, and *Retail Trade*.

The validation team ran several different queries through the web interface and found the UI intuitive and easy to navigate. The queries ran successfully and produced summary statistics within the app; however, attempts to download the resulting OD data caused the software to hang without ever completing the task. Tests on small subsets of the data downloaded without issue, so it appears that this issue may be related to the size of the large query (1203 x 1203 OD matrix). Geotab's support team was very responsive and helped the validation team acquire the necessary results. However, it is likely that using the API directly would avoid this issue.

Figure 4 shows a screenshot of Geotab's data delivery in CSV format. Many more fields than can be shown visually in Figure 4 are included in the file, but the most important fields are highlighted below, including temporal, spatial, and vehicle class / vocation / industry filters. As mentioned above, Geotab reports **total trips** for each zone pair during the study period, so as part of the standardization process these values were averaged by temporal period (either by hour, day-part, or day – depending on the analysis). As an example, consider the trips for an OD pair during the Weekday AM Peak scenario. Geotab reports the total number of OD pair trips that take place between 6-9am for all typical weekdays (non-holiday Tuesdays-Thursdays) during the month. To create an average value for the Weekday AM Peak scenario that is comparable across vendors, this total count is divided by the number of non-holiday Tuesdays-Thursdays during the month, which in the case of October 2022 is 12.

DateFrom1	DateTo1	DOW	TimeFrom	TimeTo	VehClass	Vocations	Industries	ZonePairID	NumJourneys	AvgDuration(min)	AvgDistance(mi)
2022-10-01	2022-10-31	All days	00:00:00	23:59:59...	All	All	All	z-...	2891	3.6	1.46
2022-10-01	2022-10-31	All days	00:00:00	23:59:59...	All	All	All	z-3q7pmgjfsa...	1898	4.48	0.91
2022-10-01	2022-10-31	All days	00:00:00	23:59:59...	All	All	All	z-...	1674	2.35	0.74
2022-10-01	2022-10-31	All days	00:00:00	23:59:59...	All	All	All	z-...	1326	2.58	0.81
2022-10-01	2022-10-31	All days	00:00:00	23:59:59...	All	All	All	z-7e9ixmiwvl...	1168	1.79	0.45
2022-10-01	2022-10-31	All days	00:00:00	23:59:59...	All	All	All	z-59j3ufzn5x...	1134	1.55	0.3
2022-10-01	2022-10-31	All days	00:00:00	23:59:59...	All	All	All	z-...	1099	2.34	0.63
2022-10-01	2022-10-31	All days	00:00:00	23:59:59...	All	All	All	z-...	1090	2.12	0.48
2022-10-01	2022-10-31	All days	00:00:00	23:59:59...	All	All	All	z-8ypzvz65fiv...	990	1.74	0.39
2022-10-01	2022-10-31	All days	00:00:00	23:59:59...	All	All	All	z-63fgqzywtx...	958	1.75	0.39

Figure 4: Geotab raw data delivery format (csv)

INRIX

The validation team communicated the parameters of the data request to INRIX by email, which included sharing a Shapefile describing the study area zones. After creating the requested dataset, INRIX made their data available to the team through a dedicated AWS (Amazon Web Services) location, which could be accessed using provided credentials. Unlike other data vendors offering Origin-Destination (OD) level information, INRIX delivers the actual trips and their endpoints as latitude and longitude coordinates, which users can aggregate to any zone level of interest, provided they have the corresponding shapefile. This aggregation is achievable using spatial join tools, allowing for a highly flexible and customizable analysis based on geographical data.

Figure 5 shows a screenshot of INRIX data delivery in CSV format. Many more fields than can be shown visually are included in the file, but several fields are shown in the figure below, including temporal and spatial information and trip mode. Not shown in this figure is vehicle weight class, which is also available.

TripId	DeviceId	ProviderId	Mode	StartDate	StartWDay	EndDate	EndWDay	StartLocLat	StartLocLon	EndLocLat	EndLocLon
82fff64...	0015f226...	d07e70ef...	1	2022-09-14T11:47:21.000Z	3	2022-09-14T11:55:52.000Z	3	37.46014	-77.66166	37.45905	-77.65205
83718...	0038eba...	d07e70ef...	1	2022-09-14T22:02:09.000Z	3	2022-09-14T22:33:47.000Z	3	37.20353	-77.69737	37.0864	-77.53257
cf7c47...	003e98b...	5ef0b4eb...	1	2022-09-14T13:48:21.725Z	3	2022-09-14T13:56:09.315Z	3	37.43942	-77.593	37.43942	-77.56416
50502...	003f7121...	d07e70ef...	1	2022-09-15T18:39:48.000Z	4	2022-09-15T18:47:39.000Z	4	37.4307	-77.62184	37.42198	-77.6342
96984...	004a362...	26337353...	1	2022-09-12T12:54:15.000Z	1	2022-09-12T13:05:50.000Z	1	37.63979	-77.51335	37.6474	-77.48176
24a38f...	0086eff2...	d07e70ef...	1	2022-09-12T20:08:44.000Z	1	2022-09-12T20:36:45.000Z	1	37.44269	-77.65892	37.44269	-77.65892
bd54ef...	008f80b9...	d07e70ef...	1	2022-09-14T20:34:18.000Z	3	2022-09-14T20:42:17.000Z	3	37.50591	-77.61085	37.51027	-77.61085
721f75...	009cc482...	d07e70ef...	1	2022-09-14T16:34:05.000Z	3	2022-09-14T16:51:15.000Z	3	37.51027	-77.60398	37.46014	-77.54356
740e2...	00a6fb5b...	d07e70ef...	1	2022-09-15T09:24:47.000Z	4	2022-09-15T10:42:31.000Z	4	36.89335	-78.05168	37.29754	-77.27302
60e96...	00b7743...	d07e70ef...	1	2022-09-14T19:30:07.000Z	3	2022-09-14T19:32:16.000Z	3	37.25274	-77.32658	37.24836	-77.32933
661fa0...	00cf2712...	d07e70ef...	1	2022-09-15T11:18:56.000Z	4	2022-09-15T11:30:54.000Z	4	37.65175	-77.45429	37.68219	-77.59025

Figure 5: INRIX raw data delivery format (csv)

Streetlight

Streetlight provides a web platform enabling users to access a variety of datasets, including turning movement counts, Annual Average Daily Traffic (AADT), top routes between origins and destinations, and the number of trips in predefined Origin-Destination pairs. Figure 6 features a screenshot of the Streetlight user interface.

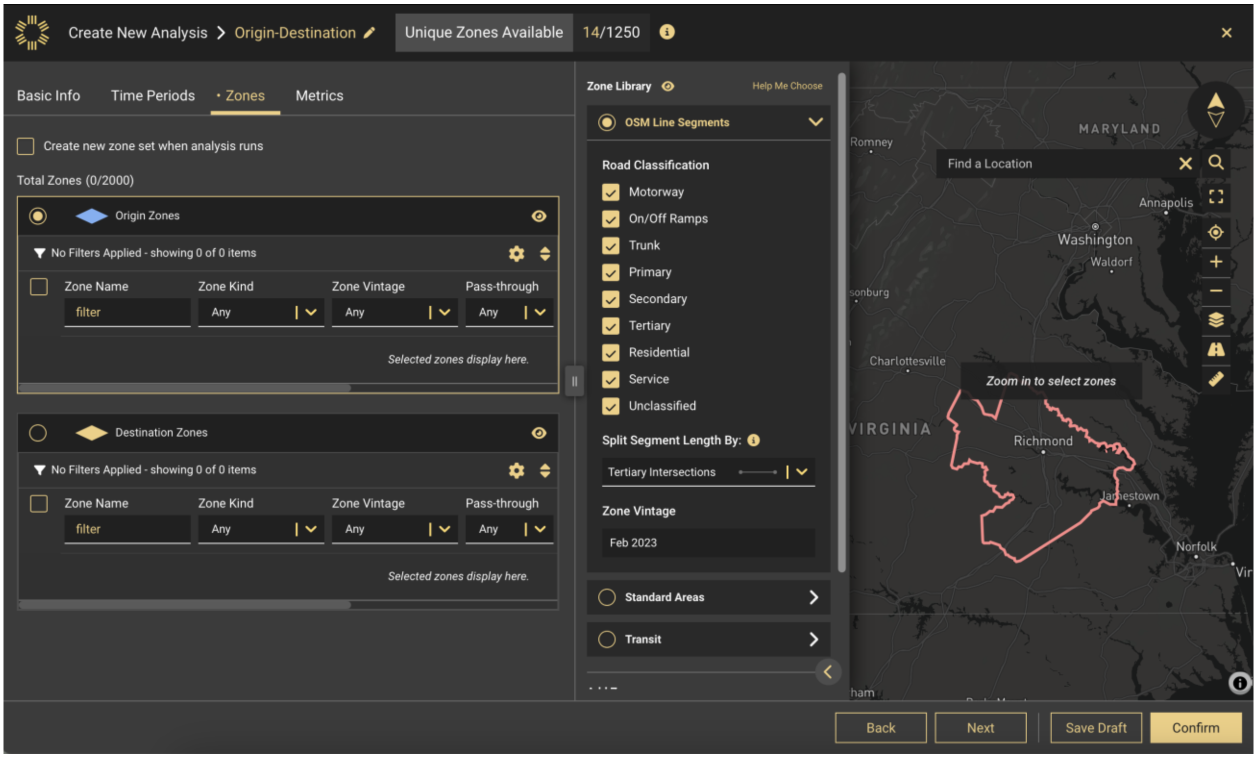

Figure 6: Streetlight web interface

After selecting the appropriate datasets, in this case, the Origin-Destination (OD) data, users can choose their preferred mode of transportation. This includes vehicles (LBS+ and CVD+), trucks (medium or heavy-duty), bicycles, pedestrians, buses, and rail. Additionally, the platform allows users to define time periods with user-customized day types and parts. Afterward, zones must then be input into the user interface. When selecting zones, users are presented with multiple options: they can choose from standard zones, such as those predefined by the US census, upload their own custom zones as a shapefile, or manually draw the zones of interest.

For this activity, the validation team uploaded a shapefile defining the custom TAZ zones in Richmond, VA. The process was straightforward, but the analysis could not initially be run due to a quota limiting the number of unique zones that can be used in analysis (500 Origin Zones + Destination Zones). This quota was increased to 2000 by the Streetlight support team, but even so, with about 1400 origins and 1400 destinations, the analysis was larger than could be supported in one query. This was easily handled by splitting the requests into smaller subsets that fit within the quota.

Figure 7 shows a screenshot of Streetlight data delivery in CSV format. Many more fields than can be shown visually are included in the file, but the most important fields are shown in Figure below, including temporal, spatial, and mode of travel.

DataPeriods	ModeofTravel	OriginZoneName	DestZoneName	DayType	DayPart	AvgODVolume
Oct 01, 2022 - Oct 31, 2022	All Vehicles CVD Plus - StL All ...	230	1199	0: All Days (M-Su)	0: All Day (12am-12am)	1
Oct 01, 2022 - Oct 31, 2022	All Vehicles CVD Plus - StL All ...	230	1199	0: All Days (M-Su)	2: Mid-Day (9am-3pm)	1
Oct 01, 2022 - Oct 31, 2022	All Vehicles CVD Plus - StL All ...	230	1199	1: Weekday (Tu-Th)	0: All Day (12am-12am)	2
Oct 01, 2022 - Oct 31, 2022	All Vehicles CVD Plus - StL All ...	230	1199	1: Weekday (Tu-Th)	2: Mid-Day (9am-3pm)	2
Oct 01, 2022 - Oct 31, 2022	All Vehicles CVD Plus - StL All ...	230	361	0: All Days (M-Su)	0: All Day (12am-12am)	2
Oct 01, 2022 - Oct 31, 2022	All Vehicles CVD Plus - StL All ...	230	361	0: All Days (M-Su)	2: Mid-Day (9am-3pm)	2
Oct 01, 2022 - Oct 31, 2022	All Vehicles CVD Plus - StL All ...	230	361	2: Weekend Day (Sa-Su)	0: All Day (12am-12am)	6
Oct 01, 2022 - Oct 31, 2022	All Vehicles CVD Plus - StL All ...	230	361	2: Weekend Day (Sa-Su)	2: Mid-Day (9am-3pm)	6

Figure 7: Streetlight raw data delivery format

Analysis Methods

In contrast to most TDM validation exercises, this validation activity does not involve statistical comparison to a reference data source. Instead, the intent is to describe and perform basic quality control checks on the data and compare/contrast results between vendors. It is important to re-emphasize that although comparing OD patterns and trip distributions between vendors can be instructive, all four vendors' OD products are slightly different (vehicle vs people movements, sample vs population estimates, personal vs commercial vehicle movements), and thus, differences are expected.

The results section below is broken down into the following subsections.

Descriptive Stats

- Number of Trips, Average Travel Time, Average Trip Distance
- Breakouts by day type/part and other filters

County-level OD Analysis

- OD Matrix Visualization
- Top OD Pairs
- Top Origins and Destinations

Results and Discussion

Descriptive Statistics

Tables 2-6 report descriptive statistics for each vendor, including trip count (average trips per temporal period) average trip distance, and average travel time. Results are broken down by scenario, including *Day Type* & *Day Part* and vendor-specific filters (e.g., trip purpose, vehicle weight class). Afterwards, Figures 8 and 9 show the distribution of trip length for each vendor side-by-side.

AirSage

Table 2 summarizes AirSage stats broken out by *Day Type* / *Day Part* and *Trip Purpos*e. It should be noted that trip distances for AirSage are calculated between the centroids of TAZ zones to which each origin and destination location are assigned. As a result, short trips that stay within a TAZ have distance values of zero assigned, which can impact overall average values.

The *Day Type* / *Day Part* results show that there are similar number of average daily trips for typical weekdays and weekends (2.69M vs 2.55M, respectively), with similar trips distances for each (averages of 5.8 miles and 5.01 miles, respectively. The AM Peak period shows a slightly higher average distance of about 7 miles. During typical weekday periods, the majority of trips (1.52M) are Home Based Other (HBO), while 844k are Non-Home Based (NHB), and just 319k are Home Based Work (HBW). HBW trips have the longest average trip distance (8.6 miles), followed by NHB and HBO (about 6.9 and 4.6 miles, respectively).

Table 2: AirSage descriptive stats

By Day Type & Day Part		
Scenario	**Count**	**Avg Dist. (mi)**
Typical Weekday (All hours)	2686530	5.80
Typical Weekday (AM Peak)	574514	7.03
Typical Weekday (PM Peak)	373830	5.56
Typical Weekday (Off peak)	1738186	5.45
Weekend (all hours)	2548717	5.01
By Trip Purpose*		
Scenario	**Count**	**Avg Dist. (mi)**
HBW	318530	8.62
HBO	1523507	4.58
NHB	844493	6.93
* Typical Weekday (All Hours)		

Geotab

Table 3 summarizes the trip count (average trips per temporal period) and average trip length/travel time for different scenarios, including breakdowns by *Day Type* & *Day Part* and Geotab-specific filters (vehicle class, vocation, industry). A unique feature of Geotab's OD product is the fine-grained control over how trips are defined, so Table 3 also reports results for two trip chaining parameter settings: (1) the default parameters and (2) a trip chain parameter of 15 minutes. Note that trip distance and travel time values are only reported for *Day Type* & *Day Part* scenarios. With many systematic queries this information could be acquired for all weight classes and vocations, but that exercise is left for possible future analysis.

Using the default trip chain parameters, the *Day Type / Day Part* results show that there are more than twice as many average daily trips for typical weekdays as compared to weekends (58.6k vs 27.4k, respectively), with slightly longer distance trips distances (averages of 1.3 miles and 0.75 miles, respectively) and travel times (3.9 min and 2.5 min, respectively). The AM Peak period shows higher travel times and distances than other time periods.

Table 3: Geotab descriptive stats

By Day Type & Day Part			
Scenario	**Count**	**Avg Dist. (mi)**	**Avg TT (min)**
Default Trip Chain Params			
Typical Weekday (All hours)	58609	1.31	3.91
Typical Weekday (AM Peak)	4242	1.62	5.10
Typical Weekday (PM Peak)	12823	0.69	2.55
Typical Weekday (Off peak)	32403	1.01	3.37
Weekend (all hours)	27001	0.75	2.49
Trip Chain Param = 15 min			
Typical Weekday (All hours)	8898	5.16	17.01

By Vehicle Class*			
Scenario	**Count Default**	**Count Trip Chain = 15 min**	**Percent of Default**
Light Duty Truck	25545	2687	11%
Med Duty Truck	4580	1459	32%
Heavy Duty Truck	6313	1919	30%
Non-Truck	20554	2351	11%

By Vocation*			
Scenario	**Count Default**	**Count Trip Chain = 15 min**	**Percent of Default**
Long Haul	2764	974	35%
Regional	5596	1819	32%
Local	9958	3419	34%
Door to Door	34374	1034	3%
Hub and Spoke	4701	1278	27%
None	1216	375	31%

* Typical Weekday (All Hours)

However, when 15-minute trip chaining parameters are used, the number of average daily trips drops significantly and the trip distance and travel time increases; there are about 1/7th of the trips during typical weekdays (8.9k vs 58.6k), with average distances increasing from 1.3 to 5.6

miles and average travel times increasing from 3.9 to 17.0 minutes. These results are intuitive because changing the trip chain parameter combines many short trips into longer ones and reinforces the fact that this parameter significantly impacts OD results.

Trip counts are also reported separately for vehicle weight class and vocation. As previously mentioned, the travel time and distances are not broken out for these scenarios. However, inspecting the number of average trips – and in particular, comparing these counts under default and 15-minute trip chain parameters – provides insight into the impact of these parameters. While all vehicle classes and vocations see a reduction in the number of trips, the impact is most profound for door-to-door vehicles (e.g., last mile delivery, garbage/recycling trucks); when the trip chain parameter is set to 15 minutes, only about 3% of the door-to-door trips are observed relative to the original value during default settings, as these very short trips are chained together. Given that more than half of the trips under default settings are assigned to the door-to-door vocation, users should determine whether this type of trip makes sense for their desired analysis. The rest of the results presented for Geotab will use the trip chain parameter of 15 minutes.

INRIX

Table 4 summarizes the trip count (average trips per temporal period) and average trip distance and travel time for different scenarios, including breakdowns by Day Type & Day Part and Vehicle Class. During weekdays, trips recorded were varying distances and travel times across different times of the day. The morning peak hours observed trips with the highest average distance of 10.35 miles and a travel time of 20.09 minutes, indicating potentially longer commutes. The evening peak showed a higher number of trips but with a slightly shorter distance and travel time compared to the morning. Off-peak hours witnessed the most trips, with the shortest average travel distance and time, suggesting more localized movement. Weekends showed a reduction in trips but with distances and travel times closely mirroring weekday off-peak patterns, implying a consistent pattern of movement across these periods.

When analyzing the data by vehicle class for a typical weekday, light-duty vehicles dominated the road with 121,831 trips, followed by medium-duty trucks at 28,711 trips, and heavy-duty trucks making up the smallest share with only 1,796 trips. Despite their lower numbers, heavy-duty trucks had the longest average distances and travel times, highlighting their role in longer-haul transportation. Conversely, light-duty vehicles, while more numerous, had shorter travel times and distances.

Table 4: INRIX descriptive stats

By Day Type & Day Part			
Scenario	**Count**	**Avg Dist. (mi)**	**Avg TT (min)**
Typical Weekday (All hours)	152338	8.17	16.60
Typical Weekday (AM Peak)	25355	10.35	20.09
Typical Weekday (PM Peak)	37533	8.10	16.91
Typical Weekday (Off-peak)	89450	7.58	15.48
Weekend (all hours)	97087	7.61	15.31
By Vehicle Class*			
Scenario	**Count**	**Avg Dist. (mi)**	**Avg TT (min)**
Light Duty Vehicles	121831	7.71	15.76
Med Duty Trucks	28711	9.86	19.85
Heavy Duty Trucks	1796	12.03	21.49
* Typical Weekday (All Hours)			

Streetlight

Table 5 indicates travel patterns across different *Day Types* and *Day Parts*, highlighting the volume and average distances traveled. On a typical weekday, across all hours, there were trips with an average distance of 8.23 miles. During the morning peak, trip average distance were longer, with an average distance of 10.56 miles, indicating more extensive commutes. The evening peak saw trips with an average distance matching the overall weekday average of 8.23 miles. Off-peak hours on weekdays accounted for the most significant number of trips with a slightly shorter average distance of 7.61 miles, suggesting closer destinations or more localized movement. Weekends showed a higher number of trips with an average distance of 7.90 miles, reflecting a mix of short and moderate distances typical of weekend travel patterns. This overview underlines the fluctuations in travel behavior, with peak times seeing longer distances traveled and weekends mirroring a blend of weekday off-peak movements.

Table 5: Streetlight descriptive stats

By Day Type & Day Part		
Scenario	**Count**	**Avg Dist. (mi)**
Typical Weekday (All hours)	3026760	8.23
Typical Weekday (AM Peak)	445339	10.56
Typical Weekday (PM Peak)	772372	8.23
Typical Weekday (Off-peak)	1745908	7.61
Weekend (all hours)	2289291	7.90

Comparison of Distributions

Figure 8 show the distribution of vendor trip distance for the *Typical Weekday (All Hours)* scenario. Whereas the previously reported tables summarized only the average values, these histograms provide additional insight into how distance is distributed.

These plots show that INRIX and Streetlight have nearly identical trip length distributions, with AirSage showing similar results aside from the 0 miles bin, which is not reflective of AirSage's data product, but rather an artifact of calculating distances centroid-to-centroid (internal trips have zero distance). In contrast, Geotab's trip distribution (using default query parameters) is markedly different from the others, with over 80% of trips falling in the 0-2-mile range and almost no trips reported above 10 miles. However, **this discrepancy does not imply that Geotab's data is wrong**; Geotab's dataset is different than the others in that it reports trips for commercial fleet vehicles – some of which make extremely short trips (e.g., "Door-to-Door").

Additionally, as was highlighted previously in Table 3, Geotab's trip chain parameter has a significant impact on how trips are defined. Figure 9 shows how the trip length distribution changes when the default value (200 seconds) is updated to 15 minutes. Although there are still differences with respect to the other vendors (as expected), the distribution is much more similar and less dominated by the last-mile and other fleet vehicles that make short trips. Subsequent tables and figures will use a 15 minute trip chain parameter for Geotab.

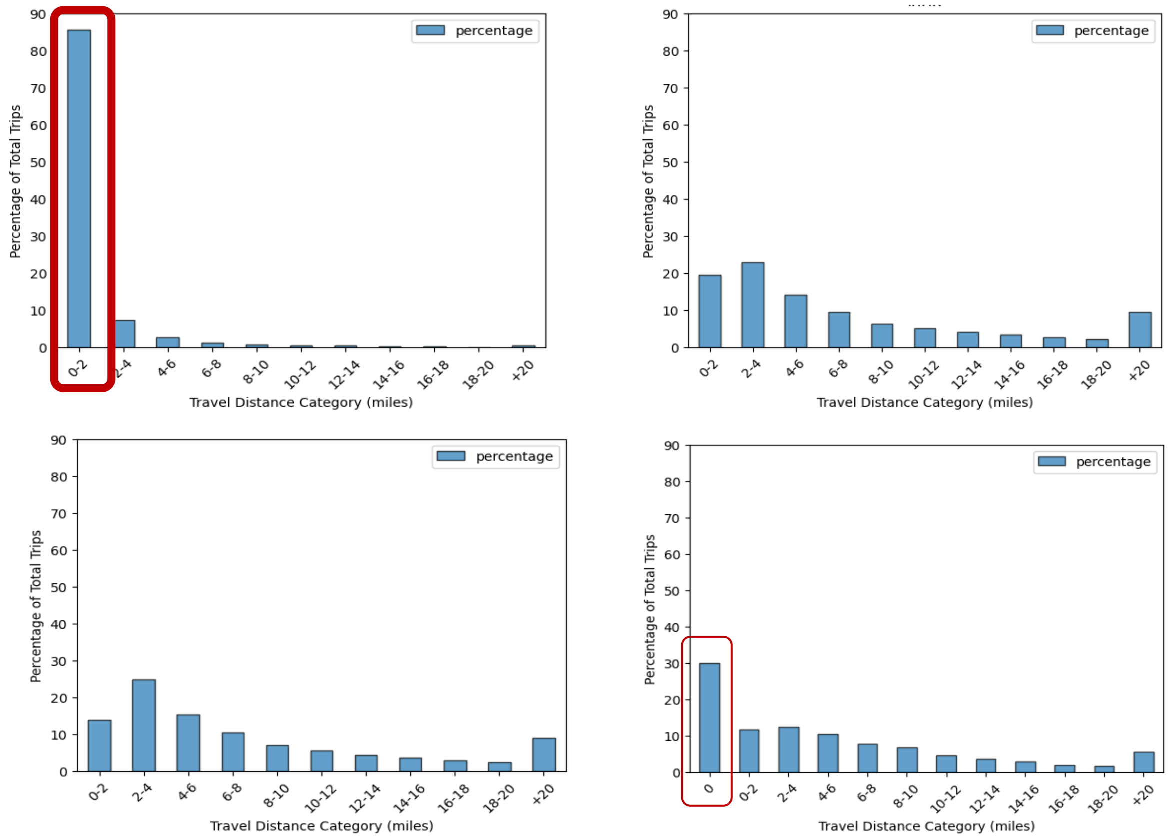

Figure 8: Travel distance distributions for all vendors

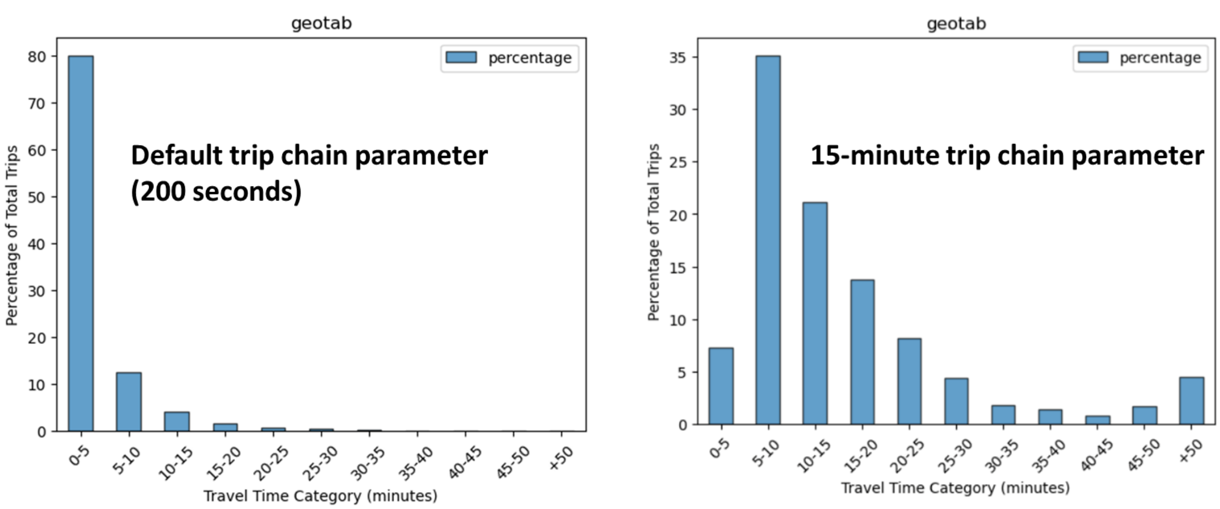

Figure 9: Trip distance distributions for default and 15 minute trip chain parameters

County Level OD Matrices

Figures 10-13 show visual representation of county-level OD patterns for typical weekdays, with an OD matrix color-coded to represent the intensity of trips between origin and destination counties. Cells with zero trips between O's and D's are colored white, while the non-zero cell values in the OD matrix are divided into quintiles; the smallest 20% of values are colored lightest shade, with the top 20% colored the darkest shade (and the middle 60% assigned to intermediate shades). This approach for coloring cells was chosen because it can be used regardless of the absolute values reported, making it applicable to vendors who are reporting sample trips or population estimates.

AirSage

Figure 10 highlights the fact that there are many internal trips (the diagonal values are the darkest color) and that many trips originate from and terminate at Chesterfield County. Additionally, it shows that the OD matrix is relatively symmetrical, indicating that there are a similar number of trips between both directions of OD pairs over the course of the study period. Follow-up analysis of row and column sums indicates that trips starting and ending in each county match within 15% (often within 2-3%).

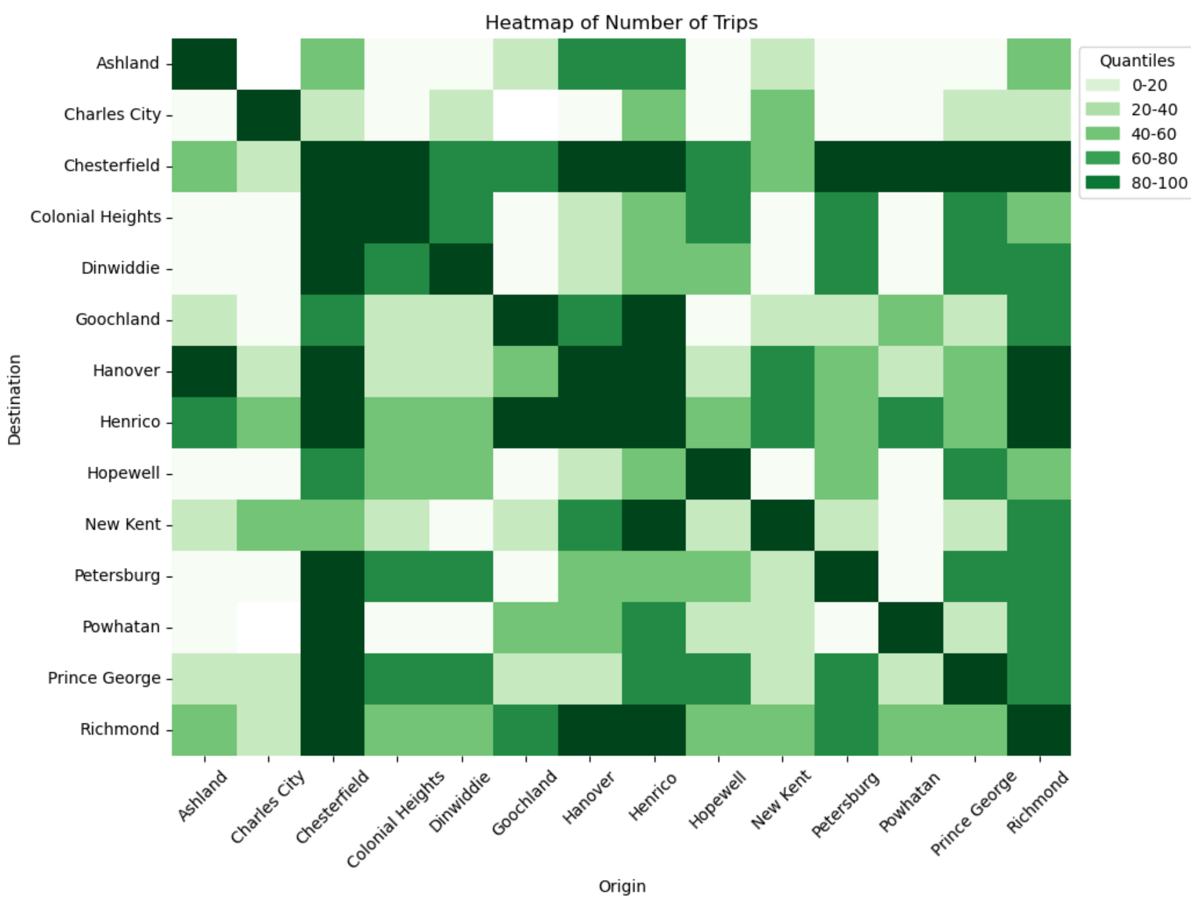

Figure 10: AirSage county OD matrix

Geotab

Figure 11 shows visual representation of county-level OD patterns with trip chain parameter a of 15 minutes, with an OD matrix color-coded to represent the intensity of trips between origin and destination counties. This OD matrix quickly highlights the fact that there are many internal trips (the diagonal values are the darkest color) and that many trips originate from and terminate at Chesterfield County. Additionally, it shows that the OD matrix is relatively symmetrical, indicating that there are a similar number of trips between both directions of OD pairs over the course of the study period. Follow-up analysis of row and column sums indicates that trips starting and ending in each county usually match within 1%.

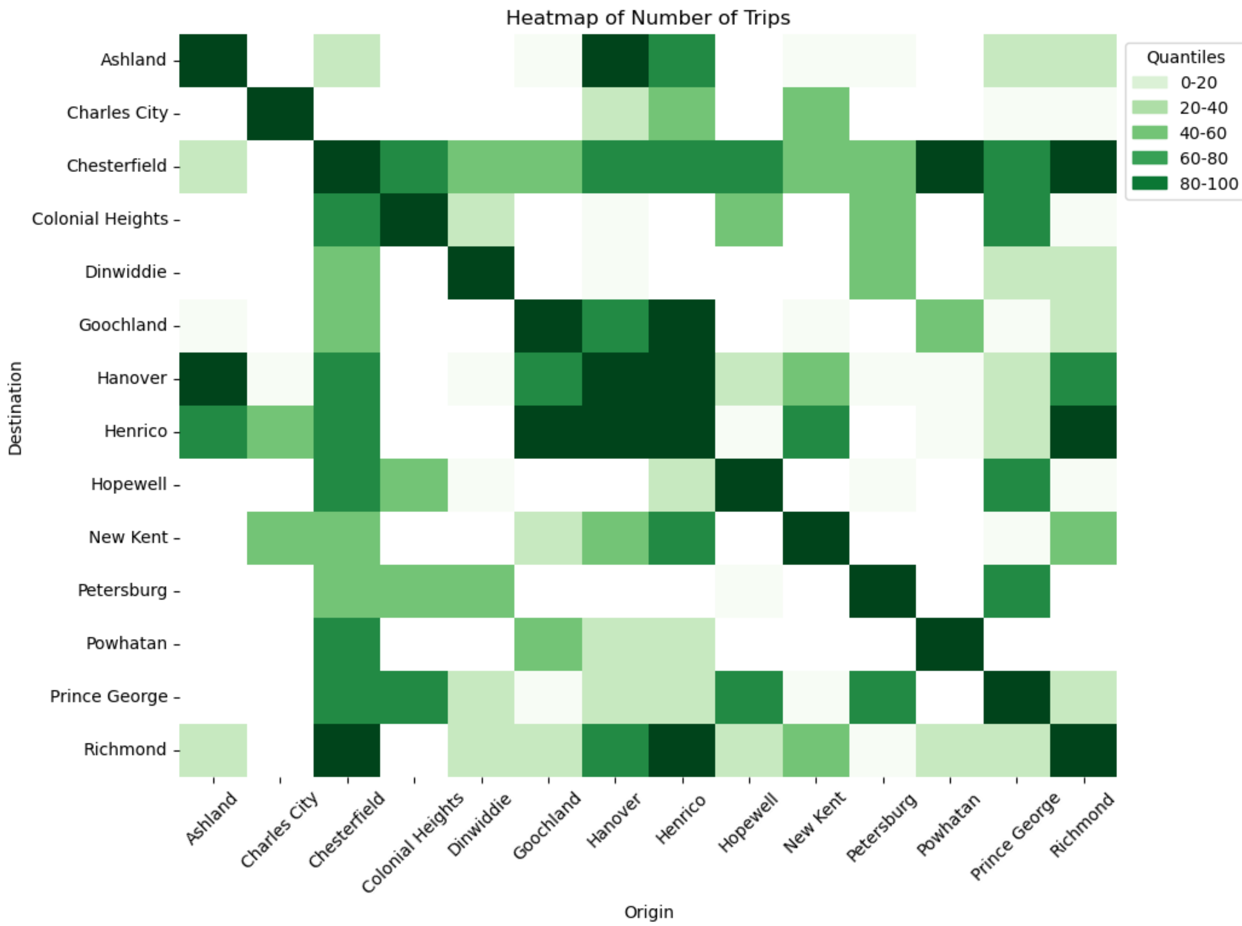

Figure 11: Geotab county OD matrix

INRIX

Figure 12 shows that INRIX's OD matrix is relatively symmetrical, meaning that there are a similar number of trips between both directions of OD pairs over the course of the study period. Additionally, it indicates that certain OD pairs, particularly those starting or ending in Chesterfield, Henrico, and Richmond, have a higher number of trips, as evidenced by the darker greens. Follow-up analysis of row and column sums indicates that trips starting and ending in each county usually match within 1%.

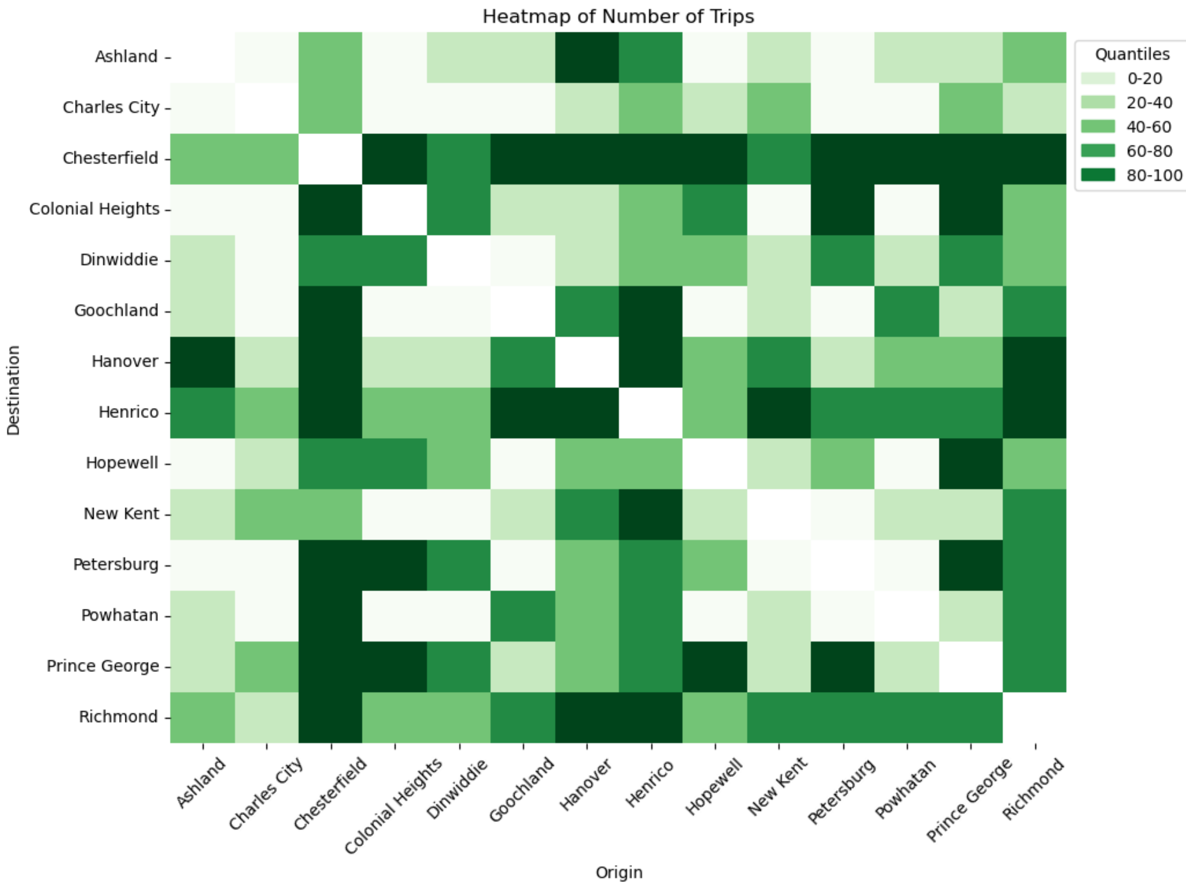

Figure 12: INRIX county OD matrix

Streetlight

Figure 13 shows a balanced OD matrix, with a similar number of trips in both directions between each origin and destination pair during the observed period. Moreover, the heatmap reveals that routes originating or concluding in Chesterfield, Henrico, and Richmond are amongst the most traveled, as indicated by the darker green shades. Follow-up analysis of row and column sums indicates that trips starting and ending in each county usually match within 1%.

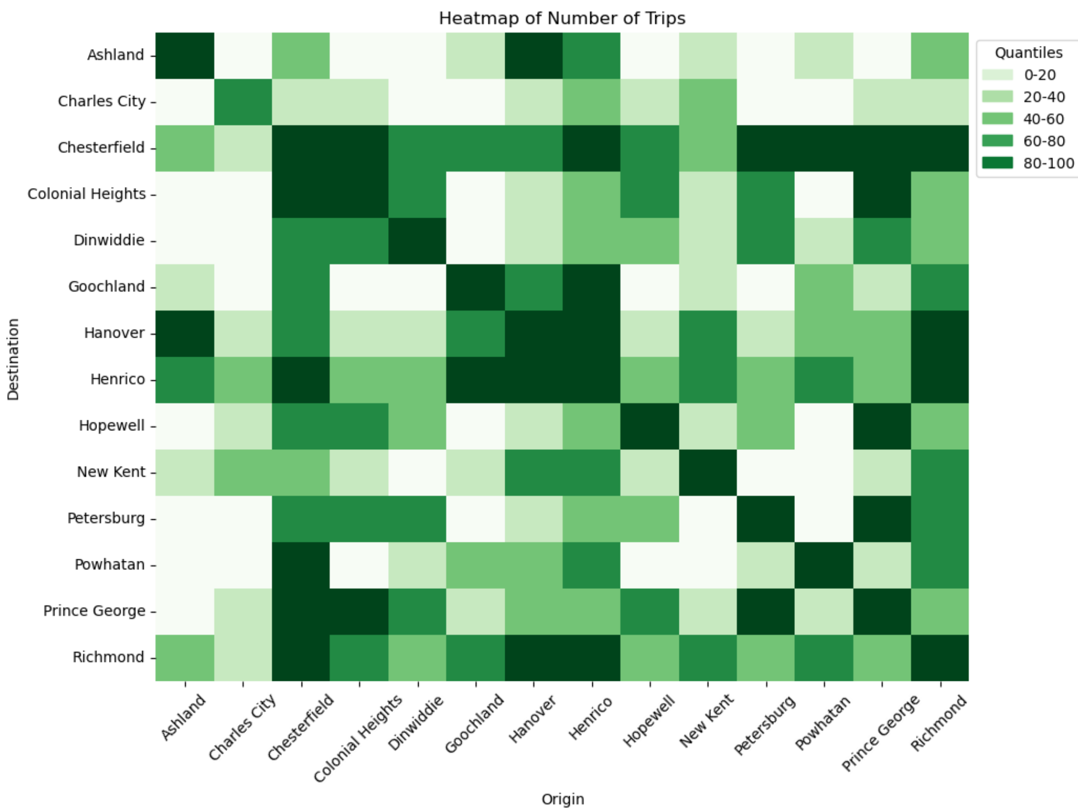

Figure 13: Streetlight county OD matrix

Comparison of OD Matrices

Figure 14 shows all of the OD matrices side-by-side. Although Geotab has more blank cells (indicating the absence of trips between certain OD pairs) than the others, in general the plots exhibit similar high-level patterns. All four show that there are many internal trips (diagonal values are the darkest) and identify Chesterfield as a key origin and destination. Additionally, they are all mostly symmetrical, meaning that a similar number of trips begin and end in a given county – a result that is expected when aggregating over the course of a month-long study period.

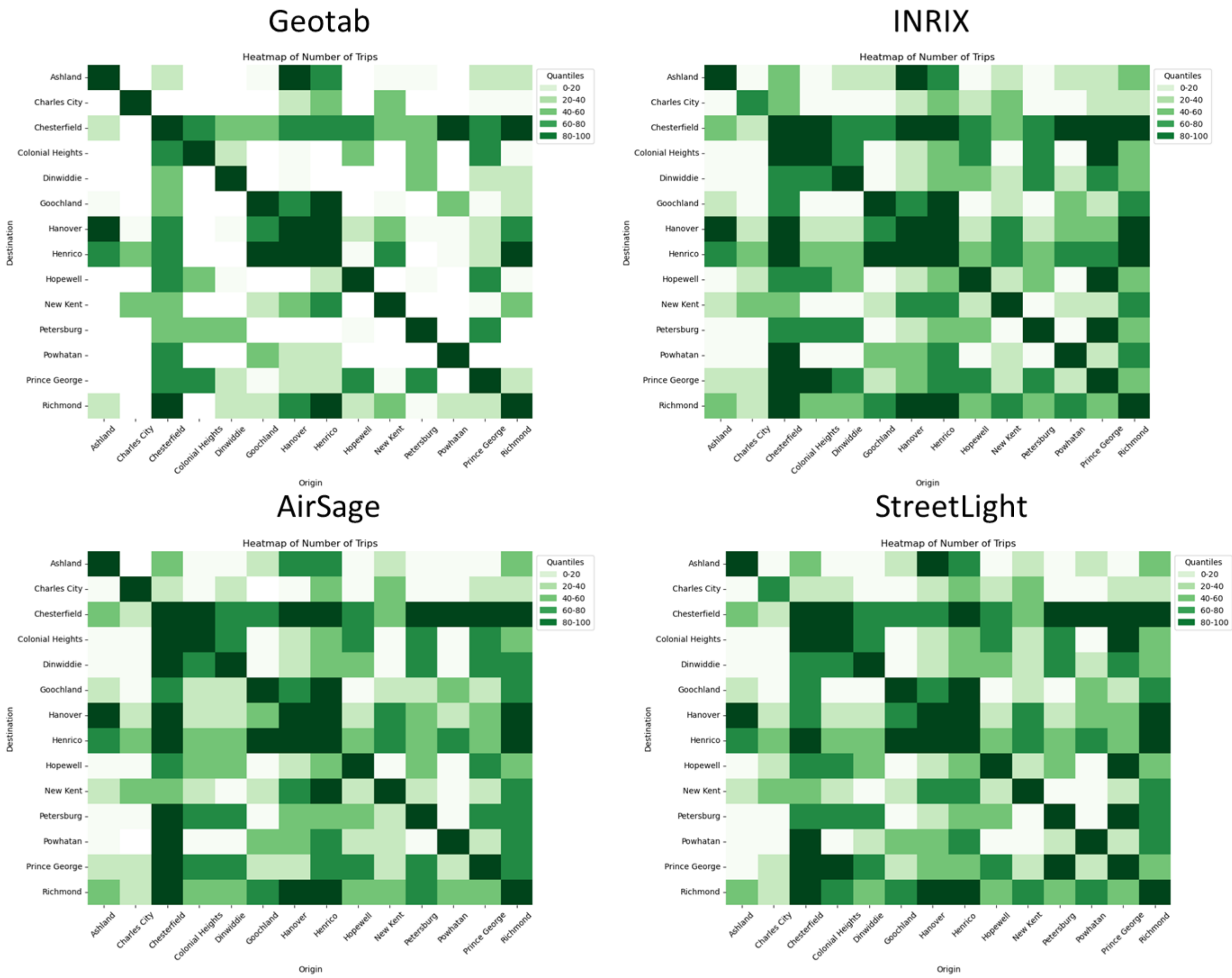

Figure 14: Comparison of county OD matrices

Top OD Pairs (County Level)

AirSage

Table 6 summarizes the top OD Pairs separately for (1) all trips, including intra-zonal trips, and (2) inter-zonal trips between counties. The top 4 OD pairs are within-county trips, with trips within Henrico and Chesterfield counties alone accounting for about 38% of all trips in the dataset. With intra-county trips excluded, trips between Henrico and Richmond (in both directions) and Chesterfield and Richmond (in both directions) are the dominant OD pairs. These pairs are shown visually below in Figure 15.

Table 6: AirSage top county OD pairs

Top OD Pairs: All Trips*				
Rank	**Origin**	**Destination**	**Percent**	**Avg Dist. (mi)**
# 1	Henrico	Henrico	19.6	2.42
# 2	Chesterfield	Chesterfield	18.6	3.21
# 3	Richmond	Richmond	11.7	1.21
# 4	Hanover	Hanover	4.6	2.79
# 5	Henrico	Richmond	3.7	6.80
Top OD Pairs: Inter-Zonal Trips*				
Rank	**Origin**	**Destination**	**Percent**	**Avg Dist. (mi)**
# 1	Henrico	Richmond	10.4	6.80
# 2	Richmond	Henrico	10.2	6.72
# 3	Chesterfield	Richmond	7.5	9.12
# 4	Richmond	Chesterfield	6.9	8.71
# 5	Chesterfield	Henrico	5.8	13.88
* Typical Weekday (All Hours)				

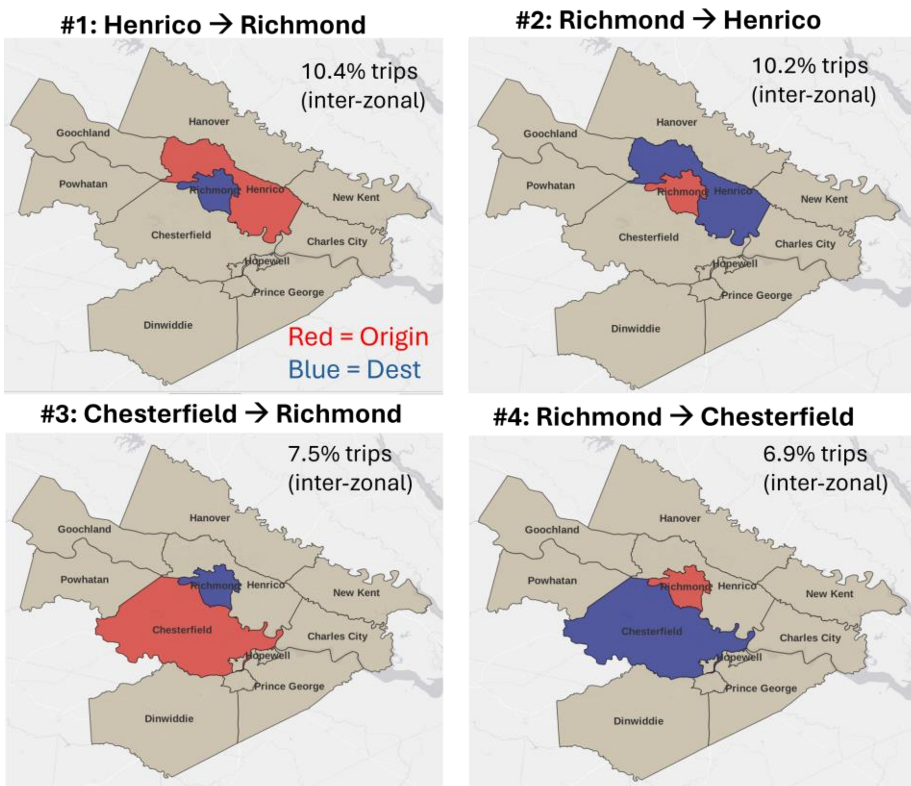

Figure 15: AirSage top county OD pair maps

Geotab

Table 7 summarizes the top OD Pairs separately for (1) all trips, including intra-zonal trips, and (2) inter-zonal trips between counties. The top 5 OD pairs are within-county trips, with trips within Henrico and Chesterfield counties alone accounting for about 50% of all trips in the dataset. With intra-county trips excluded, trips between Hanover and Henrico (in both directions) and Richmond and Henrico (in both directions) are the dominant OD pairs. These pairs are shown visually below in Figure 16.

Table 7: Geotab top county OD pairs

Top OD Pairs: All Trips*					
Rank	**Origin**	**Destination**	**Percent**	**Avg Dist. (mi)**	**Avg TT (min)**
# 1	Henrico	Henrico	27.8	3.94	14.46
# 2	Chesterfield	Chesterfield	22.8	5.16	19.31
# 3	Richmond	Richmond	12.8	4.91	17.84
# 4	Hanover	Hanover	9.0	4.09	12.72
# 5	Prince George	Prince George	6.2	4.26	15.58
Top OD Pairs: Inter-Zonal Trips*					
Rank	**Origin**	**Destination**	**Percent**	**Avg Dist. (mi)**	**Avg TT (min)**
# 1	Hanover	Henrico	9.8	13.45	30.25
# 2	Henrico	Hanover	9.6	11.58	26.73
# 3	Richmond	Henrico	7.6	6.16	18.29
# 4	Henrico	Richmond	6.6	6.96	22.08
# 5	Richmond	Chesterfield	6.0	10.54	26.52
* Typical Weekday (All Hours), Trip Chain = 15 min					

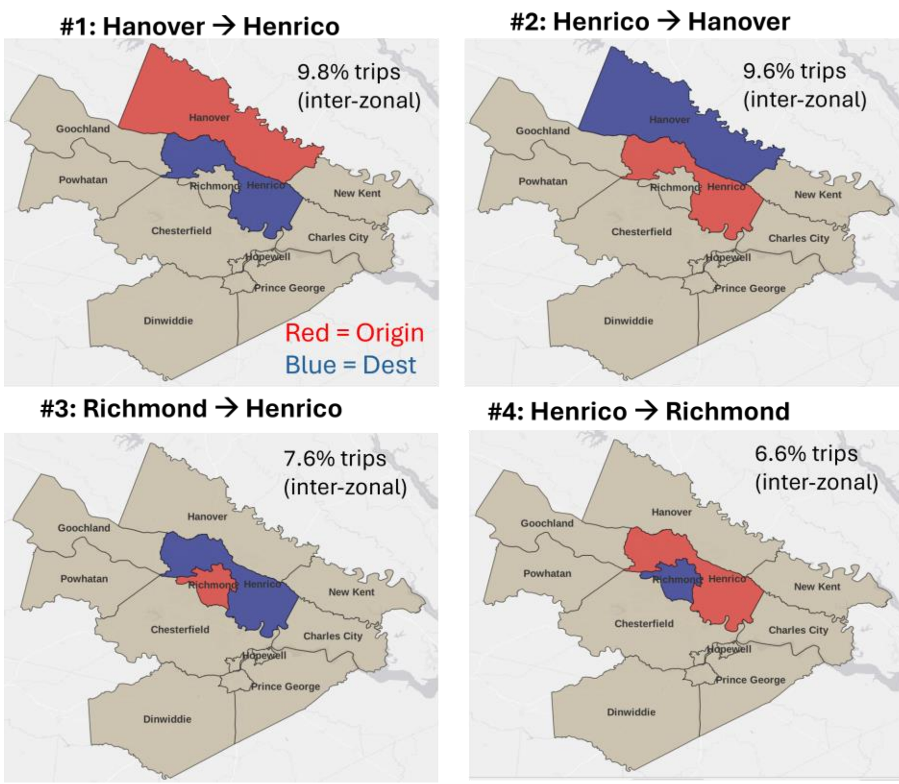

Figure 16: Geotab top county OD pair maps

INRIX

Table 8 summarizes the top OD Pairs separately for (1) all trips, including intra-zonal trips, and (2) inter-zonal trips between counties. The top 4 OD pairs are within-county trips, with trips within Henrico and Chesterfield counties alone accounting for about 42% of all trips in the dataset. With intra-county trips excluded, trips between Henrico and Richmond (in both directions) and Chesterfield and Richmond (in both directions) are the dominant OD pairs. These pairs are shown visually below in Figure 17.

Table 8: INRIX top county OD pairs

Rank	Origin	Destination	Percent	Avg Dist. (mi)	Avg TT (min)
Top OD Pairs: All Trips*					
# 1	Chesterfield	Chesterfield	22.6	5.48	13.21
# 2	Henrico	Henrico	19.2	4.50	12.04
# 3	Richmond	Richmond	7.6	3.38	11.60
# 4	Hanover	Hanover	6.5	5.69	13.77
# 5	Richmond	Henrico	3.3	8.31	19.07
Top OD Pairs: Inter-Zonal Trips*					
Rank	**Origin**	**Destination**	**Percent**	**Avg Dist. (mi)**	**Avg TT (min)**
# 1	Richmond	Henrico	9.6	8.31	19.07
# 2	Henrico	Richmond	9.5	8.30	19.38
# 3	Richmond	Chesterfield	6.7	11.96	23.11
# 4	Chesterfield	Richmond	6.5	12.11	23.48
# 5	Henrico	Hanover	5.2	12.88	22.30

* Typical Weekday (All Hours)

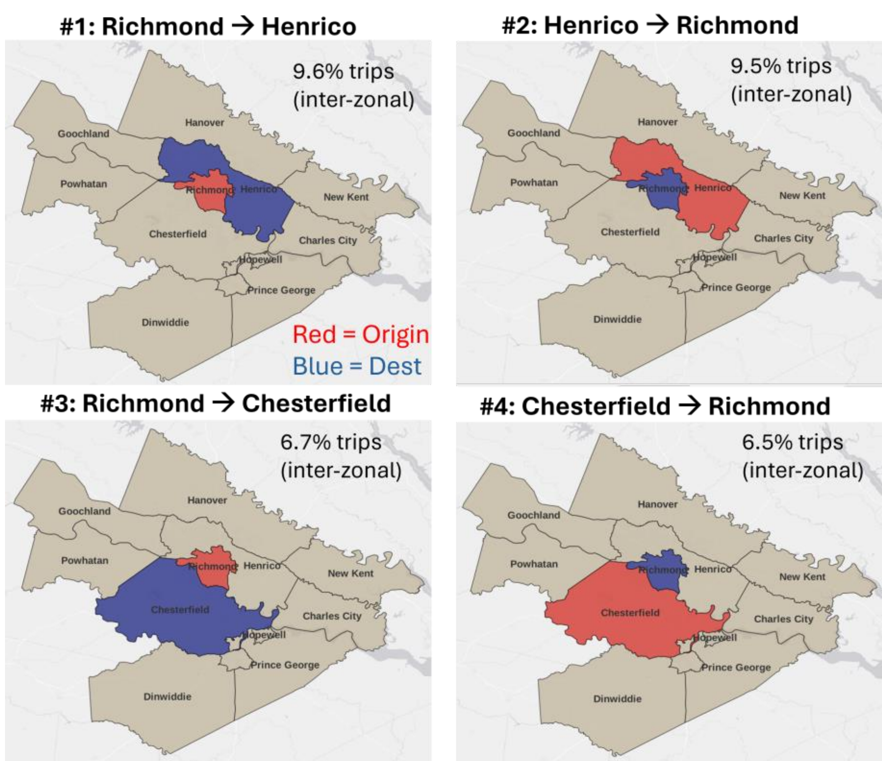

Figure 17: INRIX top county OD pair maps

Streetlight

Table 9 summarizes the top OD Pairs separately for (1) all trips, including intra-zonal trips, and (2) inter-zonal trips between counties. The top 4 OD pairs are intra-county trips, with trips within

Chesterfield and Henrico counties alone accounting for about 40% of all trips in the dataset. With intra-county trips excluded, trips between Richmond and Henrico (in both directions) and Richmond and Chesterfield (in both directions) are the dominant OD pairs. These pairs are shown visually below in Figure 18.

Table 9: Streetlight top county OD pairs

Top OD Pairs: All Trips*					
Rank	**Origin**	**Destination**	**Percent**	**Avg Dist. (mi)**	**Avg TT (min)**
# 1	Chesterfield	Chesterfield	22.4	5.92	
# 2	Henrico	Henrico	19.7	4.81	
# 3	Richmond	Richmond	7.8	3.55	
# 4	Hanover	Hanover	5.6	5.46	
# 5	Richmond	Henrico	3.7	8.32	
Top OD Pairs: Inter-Zonal Trips*					
Rank	**Origin**	**Destination**	**Percent**	**Avg Dist. (mi)**	**Avg TT (min)**
# 1	Richmond	Henrico	10.5	8.32	
# 2	Henrico	Richmond	10.5	8.31	
# 3	Richmond	Chesterfield	7.1	11.88	
# 4	Chesterfield	Richmond	7.0	12.15	
# 5	Henrico	Hanover	5.3	12.79	
* Typical Weekday (All Hours)					

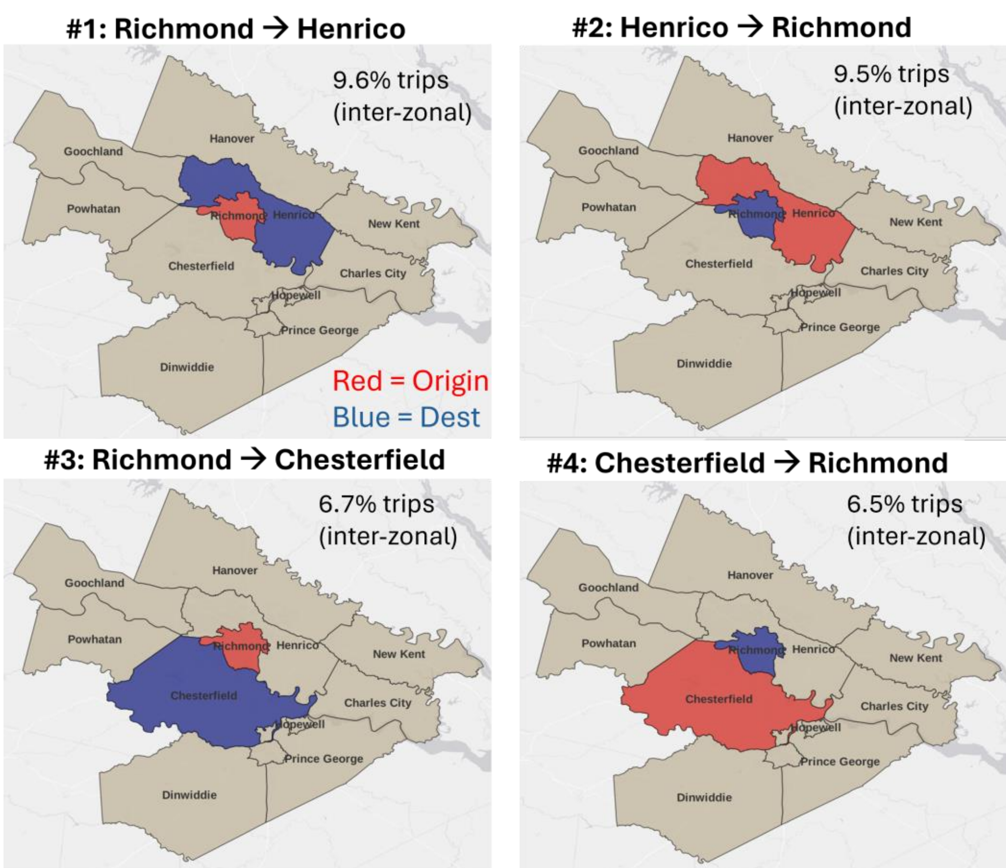

Figure 18: Streetlight top county OD pair maps

Top Origins and Destinations (County Level)

AirSage

Table 10 summarizes the top origins (to all destinations) and destinations (from all origins), with results broken out for (1) All trips and (2) Inter-Zonal trips that start and end in different counties. While the exact distribution of trips differs slightly between approaches, three counties produce and attract the most trips: Henrico, Chesterfield, and Richmond. These results are intuitive and generally in line with county populations.

Table 10: AirSage top county origins and destinations

Top Origins and Destinations: All Trips*				
Rank	**Origin**	**Percent**	**Destination**	**Percent**
# 1	Henrico	27.9	Henrico	28.6
# 2	Chesterfield	26.0	Chesterfield	26.0
# 3	Richmond	19.0	Richmond	19.7
# 4	Hanover	8.1	Hanover	7.7
# 5	Prince George	3.1	Petersburg	3.0
Top Origins and Destinations: Inter-Zonal Trips*				
Rank	**Origin**	**Percent**	**Destination**	**Percent**
# 1	Henrico	23.1	Henrico	25.2
# 2	Chesterfield	20.8	Richmond	22.5
# 3	Richmond	20.6	Chesterfield	20.7
# 4	Hanover	9.9	Hanover	8.5
# 5	Prince George	4.4	Prince George	4.0
* Typical Weekday (All Hours)				

Table 11 provides a similar breakdown but normalizes the number of trips by population (average daily trips per 1000 population). These normalized results change the top origin and destination rankings, with the top few counties producing about 8-15% of trips (rather than 20-25% unnormalized). On a per-person basis, the trips are distributed across counties more evenly and top counties include not just urban, but also rural locations (e.g., Dinwiddie).

Table 11: AirSage top county origins and destinations normalized by population

Top Origins and Destinations: All Trips*				
Rank	**Origin**	**Trips / 1k pop (%)**	**Destination**	**Trips / 1k pop (%)**
# 1	Colonial Heights	2965 (10.5%)	Colonial Heights	3001 (11%)
# 2	Dinwiddie	2499 (8.8%)	Petersburg	2403 (8.8%)
# 3	Petersburg	2389 (8.4%)	Richmond	2333 (8.6%)
# 4	Goochland	2314 (8.2%)	Henrico	2299 (8.4%)
# 5	Richmond	2252 (8%)	Hopewell	2204 (8.1%)
Top Origins and Destinations: Inter-Zonal Trips*				
Rank	**Origin**	**Trips / 1k pop (%)**	**Destination**	**Trips / 1k pop (%)**
# 1	Colonial Heights	1720 (13.9%)	Colonial Heights	1756 (15.5%)
# 2	Dinwiddie	1227 (9.9%)	Hopewell	1001 (8.8%)
# 3	Goochland	1198 (9.7%)	Petersburg	988 (8.7%)
# 4	Prince George	974 (7.9%)	Richmond	949 (8.4%)
# 5	Petersburg	974 (7.9%)	Goochland	912 (8%)
* Typical Weekday (All Hours)				

Geotab

Table 12 summarizes the top origins (to all destinations) and destinations (from all origins), with results broken out for (1) All trips and (2) Inter-Zonal trips that start and end in different counties. While the exact distribution of trips differs slightly between approaches, Henrico County produces and attracts the most trips, followed closely by Hanover, Richmond, and Chesterfield.

Table 12: Geotab top county origins and destinations

Top Origins and Destinations: All Trips*				
Rank	**Origin**	**Percent**	**Destination**	**Percent**
# 1	Henrico	29.5	Henrico	29.5
# 2	Chesterfield	23.8	Chesterfield	23.8
# 3	Richmond	13.8	Richmond	13.7
# 4	Hanover	10.2	Hanover	10.2
# 5	Prince George	6.4	Prince George	6.5
Top Origins and Destinations: Inter-Zonal Trips*				
Rank	**Origin**	**Percent**	**Destination**	**Percent**
# 1	Henrico	25.1	Henrico	25.6
# 2	Hanover	18.3	Hanover	18.0
# 3	Richmond	15.6	Chesterfield	14.7
# 4	Chesterfield	14.7	Richmond	14.0
# 5	Ashland	6.7	Ashland	6.4
* Typical Weekday (All Hours), Trip Chain = 15 min				

Table 13 provides a similar breakdown but normalizes the number of trips by population (average daily trips per 1000 population). Although these normalized results change the ranking of top origin and destinations, the average number of trips are more evenly distributed across counties. As a reminder, the trip values reported for Geotab represent samples (not population estimates), which is why these values are much smaller than vendors that are reporting values that are intended to represent the population.

Table 13: Geotab top county origins and destinations normalized by population

Top Origins and Destinations: All Trips*				
Rank	**Top Origin**	**Trips / 1k pop (%)**	**Destination**	**Trips / 1k pop (%)**
# 1	Prince George	13 (14.6%)	Prince George	13 (14.7%)
# 2	Powhatan	11 (12.4%)	Powhatan	12 (12.7%)
# 3	Hanover	11 (12.2%)	Hanover	11 (12.2%)
# 4	Henrico	8 (8.6%)	Henrico	8 (8.6%)
# 5	New Kent	7 (7.1%)	Goochland	7 (7.1%)
Top Origins and Destinations: Inter-Zonal Trips*				
Rank	**Origin**	**Trips / 1k pop (%)**	**Destination**	**Trips / 1k pop (%)**
# 1	Colonial Heights	1.1 (16.2%)	Goochland	1 (14%)
# 2	Hanover	0.9 (12.6%)	Colonial Heights	0.8 (12%)
# 3	Goochland	0.8 (11.7%)	Hanover	0.8 (11.9%)
# 4	New Kent	0.6 (8.6%)	Prince George	0.7 (10.3%)
# 5	Prince George	0.6 (8.6%)	Powhatan	0.6 (8.6%)
* Typical Weekday (All Hours), Trip Chain = 15 min				

INRIX

Table 14 summarizes the top origins (to all destinations) and destinations (from all origins), with results broken out for (1) All trips and (2) Inter-Zonal trips that start and end in different counties. Although there are minor variations in the trip distribution across these categories, Chesterfield, Henrico, and Richmond counties emerge as the leading three in terms of both producing and attracting trips. For all trips, Chesterfield County stands out as the foremost county, whereas Henrico takes precedence as the leading county for interzonal trips exclusively.

Table 14: INRIX top county origins and destinations

Top Origins and Destinations: All Trips*				
Rank	**Origin**	**Percent**	**Destination**	**Percent**
# 1	Chesterfield	29.1	Chesterfield	29.1
# 2	Henrico	27.6	Henrico	27.5
# 3	Richmond	14.2	Richmond	14.1
# 4	Hanover	9.9	Hanover	9.9
# 5	Prince George	3.8	Prince George	3.8
Top Origins and Destinations: Inter-Zonal Trips*				
Rank	**Origin**	**Percent**	**Destination**	**Percent**
# 1	Henrico	24.6	Henrico	24.5
# 2	Richmond	19.6	Chesterfield	19.4
# 3	Chesterfield	19.4	Richmond	19.4
# 4	Hanover	10.0	Hanover	10.2
# 5	Prince George	4.4	Prince George	4.4
* Typical Weekday (All Hours)				

Table 15 provides a similar breakdown but normalizes the number of trips by population (average daily trips per 1000 population). These normalized results change the top origin and destination rankings, with the top few counties producing about 10-12% of trips (rather than up to 29% unnormalized). On a per-person basis, the trips are distributed across counties more evenly and top counties include not just urban, but also rural locations (e.g., Goochland).

Table 15: INRIX top county origins and destinations normalized by population

Top Origins and Destinations: All Trips*				
Rank	**Top Origin**	**Trips / 1k pop (%)**	**Destination**	**Trips / 1k pop (%)**
# 1	Colonial Heights	202 (12.8%)	Colonial Heights	202 (12.8%)
# 2	Goochland	161 (10.2%)	Goochland	162 (10.3%)
# 3	Hanover	154 (9.8%)	Hanover	155 (9.8%)
# 4	Prince George	134 (8.5%)	Prince George	134 (8.5%)
# 5	Henrico	126 (8%)	Henrico	125 (7.9%)
Top Origins and Destinations: Inter-Zonal Trips*				
Rank	**Origin**	**Trips / 1k pop (%)**	**Destination**	**Trips / 1k pop (%)**
# 1	Colonial Heights	108 (16.3%)	Colonial Heights	108 (16.3%)
# 2	Goochland	89 (13.5%)	Goochland	90 (13.5%)
# 3	Prince George	53 (8%)	Prince George	52 (7.9%)
# 4	Hopewell	49 (7.3%)	Hopewell	49 (7.3%)
# 5	Petersburg	48 (7.3%)	Petersburg	48 (7.3%)
* Typical Weekday (All Hours)				

Streetlight

Table 16 summarizes the top origins (to all destinations) and destinations (from all origins), with results broken out for (1) All trips and (2) Inter-Zonal trips that start and end in different counties. Although there are minor variations in the trip distribution across these categories, Chesterfield, Henrico, and Richmond counties emerge as the leading three in terms of both producing and attracting trips. For all trips, Chesterfield County stands out as the foremost county, whereas Henrico takes precedence as the leading county for interzonal trips exclusively.

Table 16: Streetlight top county origins and destinations

Top Origins and Destinations: All Trips*				
Rank	**Origin**	**Percent**	**Destination**	**Percent**
# 1	Chesterfield	29.2	Chesterfield	29.2
# 2	Henrico	28.5	Henrico	28.5
# 3	Richmond	15.1	Richmond	15.1
# 4	Hanover	8.9	Hanover	8.9
# 5	Prince George	3.7	Prince George	3.6
Top Origins and Destinations: Inter-Zonal Trips*				
Rank	**Origin**	**Percent**	**Destination**	**Percent**
# 1	Henrico	24.9	Henrico	24.8
# 2	Richmond	20.6	Richmond	20.5
# 3	Chesterfield	19.2	Chesterfield	19.2
# 4	Hanover	9.3	Hanover	9.4
# 5	Colonial Heights	4.6	Colonial Heights	4.6
* Typical Weekday (All Hours)				

Table 17 provides a similar breakdown but normalizes the number of trips by population (average daily trips per 1000 population). These normalized results change the top origin and destination rankings, with the top few counties producing about 9-15% of trips (rather than 15-29% unnormalized). On a per-person basis, the trips are distributed across counties more evenly and top counties include not just urban, but also less populated locations.

Table 17: Streetlight top county origins and destinations normalized by population

Top Origins and Destinations: All Trips*				
Rank	**Top Origin**	**Avg trips / 1k pop (%)**	**Destination**	**Avg trips / 1k pop (%)**
# 1	Colonial Heights	4657 (15.4%)	Colonial Heights	4660 (15.4%)
# 2	Hanover	2736 (9.1%)	Hanover	2747 (9.1%)
# 3	Henrico	2578 (8.5%)	Henrico	2576 (8.5%)
# 4	Prince George	2574 (8.5%)	Prince George	2567 (8.5%)
# 5	Goochland	2441 (8.1%)	Goochland	2460 (8.1%)
Top Origins and Destinations: Inter-Zonal Trips*				
Rank	**Origin**	**Trips / 1k pop (%)**	**Destination**	**Avg trips / 1k pop (%)**
# 1	Colonial Heights	2706 (19.9%)	Colonial Heights	2709 (19.9%)
# 2	Goochland	1515 (11.2%)	Goochland	1533 (11.3%)
# 3	Prince George	1136 (8.4%)	Prince George	1129 (8.3%)
# 4	Petersburg	1081 (8%)	Petersburg	1082 (8%)
# 5	Hopewell	985 (7.3%)	Hopewell	996 (7.3%)
* Typical Weekday (All Hours)				

Conclusions

This OD validation report focuses on two components: (1) a vendor questionnaire intended to identify the key aspects of each vendor's product offering, and (2) an analysis of sample data provided by each vendor in Richmond, VA. Given that this report marks the first time acquiring commercial OD data, the analysis focuses on the process of obtaining and interpreting OD data from each vendor, a basic descriptive analysis of the datasets, and preliminary analysis of comparison results. The sample datasets were chosen to coincide with other complementary datasets owned and managed by Viriginia Department of Transportation (VDOT) to support future follow-up analysis.

Several key takeaways emerged from the questionnaire and review of sample data:

- **TDM OD products are diverse, and each vendor reports trip patterns from different perspectives.** Key differences include data sources used (connected vehicle data vs smart phone location-based services (LBS) data), the types of vehicles captured (e.g., mixture of all vehicles, freight only), how trips are quantified (e.g., person-trips vs vehicle-trips), trip mode, and whether report trip counts reflect observed sample probes or population-level estimates. Furthermore, *each vendor uses slightly different logic to split GPS waypoints into separate trips, so the same GPS trajectory may result in different trip definitions*. These nuances should be taken into consideration when comparing vendor results to each other or to external data sources.

- **All vendors allow users to easily create custom OD queries, including support for user-defined geographic zones, customizable time periods, and other vendor-specific filters.** Although the query process differs across vendors, all vendors provide a straightforward way to produce meaningful customized OD datasets that meet necessary spatial, temporal, and other criteria common to most transportation analysis.

- **Understanding query options are critical, especially those that determine trip definition**. Specifying query parameters is one of the most challenging aspects of obtaining OD data, as these settings control query results, as well as how output data should be interpreted (e.g., sample counts vs population estimates, total counts for study period vs daily averages), how trips entering/exiting the study area are handled, and in some cases, how trips themselves should be defined. Parameters that determine trip definition, if exposed, are particularly critical.

- **Preliminary analysis of sample data from Richmond, VA at the county-level shows intuitive OD patterns for all 4 vendors**, including similarities in top Origins, Destinations, and OD pairs -- despite key differences in the types of trips captured by vendor products. Inspection of trip counts, average travel time and average trip distance for different scenarios (e.g., time of day, trip purpose) also yielded results that appeared reasonable based on local knowledge.