Metropolitan Washington Council of Governments

GEN3 DATA DEVELOPMENT

December 29, 2021

PREPARED FOR:

METROPOLITAN WASHINGTON COUNCIL OF GOVERNMENTS

SUBMITTED BY:

RSG

WITH BASELINE MOBILITY GROUP

55 Railroad Row
White River Junction, VT 05001
802.295.4999
www.rsginc.com

CONTENTS

1.0 INTRODUCTION 3

2.0 HOUSEHOLD TRAVEL SURVEY 5

2.1 INTRODUCTION 5

2.2 2015 HTS DATA FORMAT 5

2.3 PART TIME WORKER IMPUTATION 6

2.4 SURVEY PROCESSING APPLICATION(SPA) TOOL 11

PERSON TYPE CODING 11

TOUR PURPOSE CODING 12

JOINT TRAVEL EPISODES 13

3.0 TRANSIT ON-BOARD SURVEY PROCESSING 17

3.1 METRORAIL 2016 PASSENGER SURVEY DATA PROCESSING 18

3.2 MARC 2016 ON-BOARD SURVEY (OBS) DATA PROCESSING 31

3.3 VRE 2019 ON-BOARD SURVEY (OBS) DATA PROCESSING 43

3.4 ON-BOARD SURVEY CODING RESULTS 1

LIST OF FIGURES

FIGURE 2-1: RANDOM FOREST FEATURE IMPORTANCE 9
FIGURE 2-2: PART TIME WORKER CONFUSION MATRIX 10
FIGURE 3-1: HOME LOCATIONS OF METRORAIL TRIPS BY TIME PERIODS 21
FIGURE 3-2: HOME LOCATIONS OF METRORAIL TRIPS BY TIME PERIODS WITHIN OR CLOSE TO THE MODELED AREA 22
FIGURE 3-3: A 75MILE BUFFER AREA AROUND THE MODEL STUDY REGION TO SEPARATE VISITOR TRIPS FROM EXTERNAL TRIPS 23
FIGURE 3-4: TRIP CATEGORIZATION FLOW CHART 23
FIGURE 3-5: MARC SERVICE MAP 32
FIGURE 3-6: ORIGIN AND DESTINATION LOCATION OF MARC TRIPS WITH COORDINATES IN THE ORIGINAL DATA 33

FIGURE 3-7: ORIGINS AND DESTINATIONS OF MARC TRIPS WITH COORDINATES COMPUTED USING A GEOCODER 34
FIGURE 3-8: STATION GROUPS DEFINED FOR TRIP WEIGHTING 37
FIGURE 3-9: VRE SYSTEM MAP 44
FIGURE 3-10: ORIGINS OF VRE TRIPS USING INCORRECT HOME LATITUDE-LONGITUDE IN THE ORIGINAL DATA 45
FIGURE 3-11: ORIGINS OF VRE TRIPS USING CORRECTLY GEOCODED HOME ADDRESS 45

LIST OF TABLES

TABLE 2-1: MATCHING HTS VARIABLES 7
TABLE 2-2: DERIVED HTS VARIABLES 7
TABLE 2-3: PERSON TYPE DEFINITIONS 11
TABLE 2-4: TRIP PURPOSE SCORING 14
TABLE 2-5 SPA OUTPUTS 15
TABLE 3-1: ON-BOARD SURVEY TRAVEL MARKETS 17
TABLE 3-2: YEAR 2018 TRIP WEIGHTING FACTORS BY STATION AND TIME PERIOD .... ERROR! BOOKMARK NOT DEFINED.
TABLE 3-3: GEN 3 MODEL TIME PERIODS 29
TABLE 3-4: METRORAIL ACCESS/EGRESS MODE SEGMENTATION 30
TABLE 3-5: STEPWISE IMPUTATIONS FOR IDENTIFYING ORIGIN AND DESTINATION TAZS 35
TABLE 3-6: 2018 TRIP WEIGHTING FACTORS BY STATION GROUPS 39
TABLE 3-7: MARC ACCESS/EGRESS MODE SEGMENTATION 42
TABLE 3-8: PROCESS OF DESTINATION TAZ ASSIGNMENT 46
TABLE 3-9: VRE ACCESS/ EGRESS MODE SEGMENTATION 0

LIST OF ABBREVIATIONS

(If Applicable—Generate with PerfectIt or e-mail DocSupport@rsginc.com for assistance)

1.0 INTRODUCTION

The National Capital Region Transportation Planning Board (NCRTPB or TPB), staffed by the Metropolitan Washington Council of Governments (MWCOG or COG), is the federally designated metropolitan planning organization (MPO) for metropolitan Washington. COG/TPB staff develops, maintains, applies, and improves the TPB's family of regional travel demand forecasting models, which are used for regional, long-range transportation planning in the metropolitan Washington region. In 2018, COG/TPB set out to develop a next-generation travel demand model. The project team, consisting of RSG and Baseline Mobility Group, recommended that COG transition from its current aggregate, trip-based travel demand model (i.e., Gen2 Model) to a simplified activity-based model (ABM) implemented in the open-source ActivitySim software platform, to be known as the Generation-3, or Gen3, Model.

Survey and census data are needed to build a calibrated, estimated, and validated activity-based Gen3 Model for the metropolitan Washington region. This report describes the datasets used to implement the Phase 1 Gen3 model and processing required to transform the data into formats required by ActivitySim.

DATA SOURCES

There are three primary sources of data used for Phase 1 model development:

1. Household Travel Survey: The primary source for model calibration and estimation is the 2017-18 MWCOG Regional Travel Survey (RTS)[1] and the 2018-19 Maryland Travel Survey (MTS)[2] that gathered household characteristics, demographic information, and full day travel patterns for approximately 35,000 persons in 16,000 households.
2. Transit On-Board Surveys: A set of transit on-board surveys conducted by three transit operators, including a 2016 Metrorail Passenger Survey collected by the Washington Metropolitan Area Transit Authority,[3] a 2016 Maryland Area Regional Commuter (MARC)

---

[1] "2017/2018 Regional Travel Survey: Technical Documentation" (Washington, D.C.: National Capital Region Transportation Planning Board, Metropolitan Washington Council of Governments, January 2021), https://www.mwcog.org/transportation/data-and-tools/household-travel-survey/.
[2] Westat, "Maryland Statewide Household Travel Survey," Final Report (Baltimore Metropolitan Council, August 2020), https://baltometro.org/sites/default/files/bmc_documents/data%26maps/transportation/maryland-travel-survey/200902_MTS_FinalReport.pdf.
[3] WBA Research, "2016 WMATA Travel Trends Survey," Final Report (Washington, D.C.: Washington Metropolitan Area Transit Authority, January 5, 2017).

Train Survey collected by Maryland Transit Administration (MTA),[4] and a 2019 Virginia Railways Express (VRE) survey collected by VRE, were used to create internal-external, external-internal, and visitor transit trip tables. The transit on-board survey data will also be used to prepare ActivitySim calibration targets for transit modes.

3. Census/ACS Data: Census data was analyzed to provide supplemental information such as auto ownership data and journey-to-work summaries. For most purposes, the census data is taken from the American Community Survey (ACS) subsets and provides demographic information in the MWCOG modeled region. For work commute flows, 2011-2015 data was used since it is the latest census dataset for which commuting flows are available. For other summaries, 2014-2018 was used since it is more representative of 2018 base-year conditions to date and consistent with the seed sample used in MWCOG Population Synthesizer.[5]

Each of these data captures different aspects of travel in the region. This report describes each survey and how the data was processed. Section 2 describes the household travel survey. Section 3 describes the transit on-board survey data processing.

Please note that the 2018 traffic counts and transit boarding counts that are used for Phase 1 model validation were provided by MWCOG staff and were documented separately. Thus, the preparation of the count data is not covered in this report.

[4] WBA Research, Foursquare ITP, and Jacobs, "Systemwide Summary: 2015-2018 Transit Origin-Destination Survey" (Maryland Department of Transportation, Maryland Transit Administration, June 2020).
[5] MWCOG Population Synthesizer: https://github.com/MWCOG/Population_Synthesizer

2.0 HOUSEHOLD TRAVEL SURVEY (HTS)

2.1 INTRODUCTION

The 2017-18 MWCOG Regional Travel Survey (RTS) and the 2018-19 Maryland Travel Survey (MTS) were used to obtain travel behavior data from residents throughout the MWCOG modeled area. The surveys collect information about the household, all members of the household, and vehicles owned by household members. The surveys required all household members to record their travel-related details in a 24-hour period. Travel details such as activity location, trip purpose, travel mode, activity arrival and departure times were recorded. Approximately 16,000 metropolitan Washington households consisting of 35,000 individuals were surveyed.

The 2017-18 MWCOG RTS and the 2018-19 MTS are the primary data source used to calibrate and evaluate the MWCOG ActivitySim model. Household, person, tour, and trip-level characteristics are summarized from the RTS/MTS and are used to estimate and calibrate models as well as compare to final model outputs. Processing of the RTS/MTS is done primarily through a Python-based Survey Processing Application (SPA) tool[6] which converts input household, person, and place files into household, person, tour, and linked trip files. The RTS/MTS data is first processed into a format required by the SPA tool which includes deriving a place file from the RTS/MTS trip file.

As part of the person-level coding, the SPA tool categorizes individuals into eight distinct person types required by ActivitySim (see Table 2-3). Part-time workers is one of the required person types, but the RTS/MTS did not ask any questions about part time work status. Thus, part-time worker status had to be imputed. RSG performed this imputation by applying a model that RSG recently developed for Southeast Michigan Council of Governments (SEMCOG) since they also had the same issue with their household survey. This is described in more detail below.

2.2 HTS DATA FORMAT

The RTS/MTS data consists of four files – household, person, trip, and vehicle files[7]. The household file contains all the household-level variables such as household size and the number of vehicles. The person file contains the details of each member listed in the household file, and it includes details such as person age and employment status. The trip file lists the travel details for each person during the assigned 24-hour period. Each record in this file represents a trip of the associated person from an origin location to a destination location.

---

[6] MWCOG SPA GitHub repo: https://github.com/MWCOG/SPA

[7] Refer to the RTS/MTS data dictionary for a complete list of variables in each file: https://app.box.com/s/6yqe1m8mkx7m5m2yjl0xr7ig6sj0zyfl

Please note that the raw RTS/MTS trip file contains linked trips. The remainder of this report uses the term 'trip" to refer to linked trips unless specified otherwise. For each trip, details of both origin and destination locations are available such as activity purpose, arrival and departure time, travel mode, and person IDs of travel party members. The trip file in its raw form can be analyzed in many ways to draw inferences about travel patterns.

Household, person, tour, and linked trip weights were provided in the RTS/MTS tables. RSG performed several checks against census data to ensure the expansion weights result in a representative sample. For example, the number of workers in the MWCOG region from the weighted HTS is approximately 3.8 million, which matches closely to the 2011-2015, 5-year ACS commuting data that had ~3.5 million workers. Note that 2011-2015 ACS data was used for commuting comparisons as that is the most recent census dataset for which these flows are available. Please note that MWCOG staff performed the reweighting of the RTS/MTS data per guidance from RSG. The household, person, tour, and trip weights in the RTS/MTS were subsequently updated for the Gen3 Model, Phase 1 development. The reweighting work will be documented separately.

2.3 PART-TIME WORKER IMPUTATION

Part-time worker status is an important variable used in the identification of person type in ActivitySim. We define part-time worker status consistent with the Census: all people 16 years old and over who usually worked 1 to 34 hours per week or less than 50 weeks in the year. Part time worker status is often provided in household travel survey data either in the form of a question asking how many hours an individual worked or asking directly whether they are a full-time or part-time worker. The RTS/MTS, however, did not have any question from which part time status could be conclusively derived. RSG recently encountered this situation when deploying ActivitySim for SEMCOG. To address this issue, we estimated a machine learning model to impute part-time status using an older household travel survey (2005 SEMCOG HTS) for the SEMCOG region which collected full/part time worker status. We then applied this model to the RTS/MTS data to impute which workers are part-time workers[8].

First, we coded a common set of data field variables between the 2005 SEMCOG HTS and RTS/MTS dataset. These variables are listed in Table 2-1. We then coded additional derived explanatory variables that we expect might have an influence on whether a person is a part-time or full-time worker (

Table 2-2).

[8] SEMCOG Data Processing Technical Memorandum. November 15, 2020.

TABLE 2-1: MATCHING SEMCOG HTS VARIABLES

2005 HTS FIELD NAME	RTS/MTS FIELD NAME	DESCRIPTION
AGE	AGE	Person Age
AGE18	AGE18	Person Age > 18?
GENDER	GENDER	Person Gender
LDRV	LICENSE	Driver's License status
WRKR	EMPLOYMENT_STATUS	Employment Status
STYPE	SCHOOL_TYPE	Level of School
MJOBS	JOBS_COUNT	Number of jobs
HHNUMPPL	HHSIZE	Household size
HHNUMVEH	HHVEH	Household vehicles
INCOME	HH_INCOME_DETAILED	Household income
WRKRS	NUMWORKERS	Number of HH workers

TABLE 2-2: DERIVED VARIABLES (2005 SEMCOG HTS AND RTS/MTS)

DERIVED VARIBLE NAME	DESCRIPTION
HHCHILD	Number of HH Children
HHADULT	Number of HH Adults
CHILD_PER_ADULT	HHCHILD / HHADULT
NUMWRKTRIPS	Number of work trips
TOTWRKDUR	Duration of work across all work trips
AVGWRKDUR	TOTWRKDUR / NUMWRKTRIPS

DERIVED VARIBLE NAME	DESCRIPTION
AVGTRAVTIME_WRK	Average travel time to work (sum of reported travel times for work trips, based on reported departure and arrival times)
AVGTRAVTIME_WRK_R	AVGTRAVTIME_WRK / NUMWRKTRIPS

Next, we removed all non-workers from both datasets since the model will only be trained on and applied to workers. We randomly selected 75% of the 2005 SEMCOG HTS workers to train the model and used the remaining 25% as a validation set to quantify model performance.[9]

We chose a random forest classification model to perform the part-time worker imputation.[10] Random forest models work by constructing a set of decision trees. Each decision tree consists of sequential binary cuts (i.e., divisions of the data into two parts) on a parameter of the input data. For example, one decision tree may cut on whether an individual is older or younger than 50 years old. Those older may then be cut on the number of work trips being less than 2 while the younger may be then cut on whether the number of household children is greater than 3. The parameters to divide the data are determined randomly in each branch of each decision tree, but the cut values are often selected such that the purest split between part-time and full-time workers is constructed. In other words, the procedure attempts to find the branches which minimize the within-group variance and maximize the between-group variance for the selected branch variable. Eventually, the decision tree makes enough cuts to break out part-time vs full-time workers in the final branches. A prediction of whether an individual is a part-time or full-time worker then comes from which branch they land in, based on the parameters and cuts of the decision tree. Many different decision trees using different parameters and cut values are constructed to form the random forest. The random forest final output is a weighted average of the predictions from all the decision trees included.

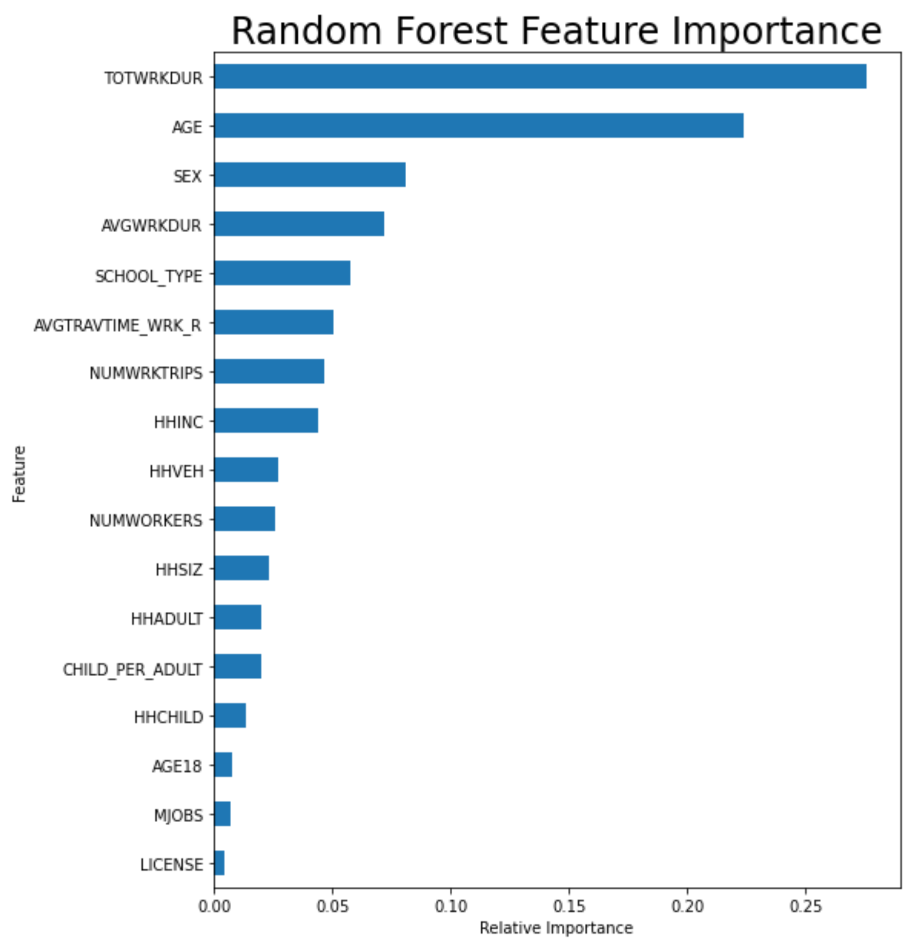
Training a random forest model involves the creation of these trees with parameters and cut values that produce the best possible predictions on the training data. Because so many different decision trees are included in the forest, this model is able to capture non-linear effects that more traditional logistic regression models fail to account for. Figure 2-1 shows the relative importance of the variables included in the worker status model. The relative importance of the variable is calculated based on how well the decision tree branch separates part-time or full-

[9] The split between training and validation datasets for machine learning models is usually either 66/33 or 75/25. We decided to use more of our data for training since we have a binary output prediction with many explanatory variables,

[10] We provide a brief overview of random forest classification here but encourage those looking for more detail to browse the many online resources available, e.g., Donges, Niklas. “A Complete Guide to the Random Forest Algorithm.” Built In, 2021. https://builtin.com/data-science/random-forest-algorithm

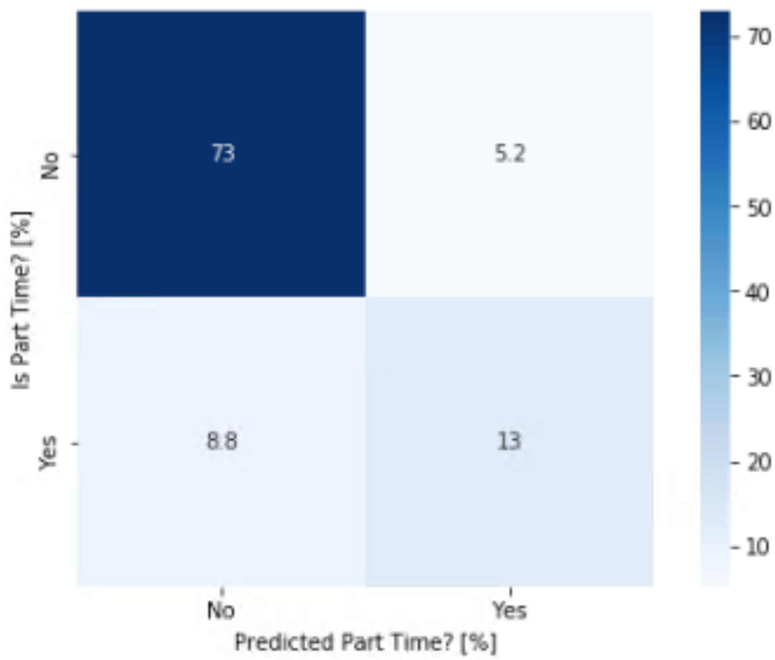
time workers. As expected, the total number of hours worked across all trips reported by a worker is the most important explanatory variable for the model to distinguish part time from full time workers. Age and gender are the next most important factors. Travel-related variables such as average travel time and work duration are also important factors. The relatively high importance of the *SEX* variable indicates that women are generally more likely to be part-time workers. Employment categories were investigated to see if there was additional bias toward a specific employment type.

FIGURE 2-1: RANDOM FOREST FEATURE IMPORTANCE



After training the model on 75% of the workers in the 2005 SEMCOG HTS, we used the model to predict the work status of the 25% holdout to determine model performance. The model predicted full versus part-time status with 86% accuracy. However, since part time workers constitute only 20% of the workforce, even a model that predicts full time for every worker would still achieve an accuracy of 80%. It is therefore important to report model precision and recall.[11] If the random forest model predicted a part-time worker, that prediction was correct 65% of the time (precision) and the model also correctly identified 56% of all part time workers (recall). Figure 2-2 shows the "confusion matrix" for part-time workers, which cross-tabulates the share of validation data by whether part-time status is observed and whether the model accurately predicted their status. When applying the model to the RTS/MTS, the same performance would be expected, but this cannot be explicitly verified because there is no part-time worker information in the RTS/MTS.

FIGURE 2-2: PART TIME WORKER CONFUSION MATRIX



After applying the model to the RTS/MTS to predict part-time work status, we checked the predictions to see if they make sense. For example, the number of part-time workers in the MWCOG RTS/MTS was imputed to be 17.4% of all workers. Additionally, women constituted about 53% of the imputed part time work force. The percentage of part-time workers in the 2013-2017 5-year ACS data for MWCOG is 16.5%, 64% of which are women. By construction, no individual under the age of 16 was imputed to be a part time worker since the model was trained on and applied to only employed persons which are required to be age 16 and older.

[11] In machine learning, model precision answers the question "What proportion of positive identifications was actually correct?", while model recall answers the question "What proportion of actual positives was identified correctly?"

We used the random forest model to impute the RTS work status before we ran the SPA tool which uses the prediction in its determination of ActivitySim person type.

2.4 SURVEY PROCESSING APPLICATION (SPA) TOOL

The SPA processes the raw survey files into data structures consistent with ActivitySim[12] and primarily involves the grouping of trips into tours. Input to the SPA tool consists of a household, person, and place file. Each of these input files need specific variables names to match the variable definitions in the SPA. A basic R script was used to rename variables in the raw RTS/MTS survey files, recode trip mode and purpose in the required format, and produce sequential household and person ID's. This R script also derives a place file from the RTS/MTS trip file in the format required for SPA. The SPA tools work off the place file to generate household and person files. Only the persons present in the place file are included in the output person file. Therefore, the R script adds a dummy home place record for persons with no reported travel in RTS/MTS. This ensures that all persons from the raw person file make it to the output person file.

Person Type Coding

The SPA first computes person type according to Table 2-3. It also checks for and resolves any inconsistencies between the reported purpose and location for work and work-related activities. Then the process loops through place records to identify tours and linked trips.

TABLE 2-3: PERSON TYPE DEFINITIONS

PERSON TYPE	AGE	EMPLOYMENT CATEGORY	STUDENT CATEGORY
1: Full-time worker	>=16	Full-time	Any
2. Part-time workers	>=16	Part-time	Not attending
3: University student	>=17	Not full-time	College+
4: Non-worker	>=16 and <=64	Non-worker	Not attending

[12] For more information on the variables and structures (tours, intermediate stops, etc.) used in ActivitySim, see RSG, and Baseline Mobility Group, Inc. "Gen3 Model Design Plan." Metropolitan Washington Council of Governments, National Capital Region Transportation Planning Board, July 2, 2020.

PERSON TYPE	AGE	EMPLOYMENT CATEGORY	STUDENT CATEGORY
5: Retired	>=65	Non-worker	Not attending
6: Student of driving age	>=16 and <=19	Not full-time	K-12
7: Student of non-driving age	>=6 and <=15	Not in labor force	K-12
8: Child too young for school	>=0 and <=5	Not in labor force	K-12, Not attending

A home-based tour is created for each sequence of place records starting and ending with "Home" as the activity purpose.[13] Work-based tours are created for each sequence of place records starting and ending with "Work" as the activity purpose. Within each tour, a trip is created for each pair of place records corresponding to an origin and a destination. Normally, a "trip linking" procedure links change-mode activities with previous and subsequent places. However, the trips in RTS/MTS trip file were linked by COG staff before SPA processing, thereby, leaving no trips with a "change mode" purpose. Therefore, the SPA tool skipped the "trip linking" procedure for all persons.

A trip is tagged as a joint trip if any of its constituting legs involves the person travelling with other household member(s).

Once all place records are processed and corresponding trips and tours are created, the SPA checks for tours that start before or end after the designated survey period (3 am to 3 am). Such tours are tagged as partial tours because some of their attributes either cannot be determined or need to be computed in a manner different from non-partial tours. Additional attributes are then derived from their constituent trips.

Tour Purpose Coding

To determine the tour purpose, a scoring system is used based on the activities within the tour, the duration of each activity, and distance of each activity from home. Table 2-4 shows the

[13] Out of almost 44,200 tours in the MWCOG household travel survey, around 2,850 tours did not start or end at home. These are retained in the dataset for purposes of tour generation if the primary activity purpose can be identified but would not be used in tour scheduling targets. The tours are tagged as 'partial tours' by SPA and are also logged in the error log file.

scores that are applied to each trip purpose by duration. The final score is the sum of the score from the table plus an additional distance term:

$$score_{final} = score_{table} + max\left(2 - \frac{Dist_{home}}{30}, 1\right)$$

where the distance from home is given in miles. The purpose of the trip with the lowest score is selected as the tour purpose. The idea behind this logic is to choose the purpose of the tour based on a hierarchy of activities, with mandatory (work and school) activities at the top of the hierarchy, followed by maintenance (escorting, shopping, and other maintenance) and discretionary (eating out. social/recreation, and other discretionary) activities. However, the procedure recognizes that some travelers might construct their activity schedules according to their own hierarchy of needs, so we provide fuzzy logic rules that take into account activity duration and distance as additional purpose ranking variables, making the assumption that longer and more distant activities are more important than shorter and closer activities. The rules treat shopping, maintenance, and other discretionary activities as equivalent in terms of hierarchy, such that the duration of the activity and the distance from home will control the identification of primary activity on the tour.

Joint Travel Episodes

Next, the SPA identifies joint travel episodes across household members. Due to the inconsistency often found in the data reported by different travel participants of a joint travel episode, rule-based intelligence is implemented in the SPA to identify and remove errors. The SPA further derives escort-related attributes (e.g., who is the chauffeur and whether the escorted individuals are being dropped off or picked up) for each travel group and tags the information to the associated trips.

Lastly, tours of different household members that contain identical series of joint travel are tagged as fully joint tours. Tours that contain both joint and non-joint trips are referred to as partially joint tours. As the last computational step of the SPA, escort related attributes at the trip level are aggregated to the tour level for partially joint tours.

TABLE 2-4: TRIP PURPOSE SCORING

PURPOSE		SCORE FUNCTION								
Code	Description	TD*=0	TD=60	TD=120	TD=180	TD=240	TD=300	TD=360	TD=420	TD=480
1-3, 10	Work/ University/ School/ Work-related	8	4	2	1.5	1.4	1.3	1.2	1.1	1
4	Escorting	8	6	4	3.5	3.4	3.3	3.2	3.1	3
5,6,9	Shopping/ Maintenance/ Discretionary	10	6	4	3.5	3.4	3.3	3.2	3.1	3
7	Eat Out	12	7	5	4.5	4.4	4.3	4.2	4.1	4
8	Social/Visit	14	8	6	5.5	5.4	5.3	5.2	5.1	5

Note: * "TD" indicates Tour Duration in minutes

The SPA produces seven data tables and two log files for error investigation purposes (see Table 2-5).

TABLE 2-5 SPA OUTPUTS

FILE NAME	DESCRIPTION
Households.csv	Each row contains selected attributes of an observed household including household income, workers, autos, size, children, and weight.
Persons.csv	Each row contains selected attributes of an observed person including age, gender, person type, employment status, student status, work and school location, and person weight.
Tours.csv	Each row contains attributes derived for an observed person tour, including tour purpose, mode, participants, origin and destination locations and times, and transit information if applicable.
Trips.csv	Each row contains attributes derived for an observed person linked trip.
Unique_joint_tours.csv	Each row contains attributes describing a fully joint tour made by multiple household members.
Unique_joint_trips.csv	Each row contains attributes describing a joint trip (unlinked) made by multiple household members.
Joint_ultrips.csv	Optional output used for analyzing data errors relating to joint travel episodes. Each row contains selected attributes of an unlinked person trip that is part of a joint trip.
Error_log.txt	Text file containing error and warning messages about unresolved data anomalies.
Recode_log.txt	Text file describing assumptions and recoding made to resolve known data anomalies and inconsistencies.

There are some input attributes, like person and household weights, that are not written to the SPA output but are required for downstream models. Since the SPA tool preserves person, household, and place ID's, these attributes can be merged in later from the SPA input household, person, and place files.

RTS data processed with the SPA tool serves as the primary data source for comparing model outputs and creating calibration targets discussed in the subsequent chapters.

3.0 TRANSIT ON-BOARD SURVEY PROCESSING

Transit on-board survey data was primarily used to develop Traffic Analysis Zone (TAZ)-level External-Internal (EI), Internal-External (IE), External-External (EE), and Visitor trip tables. There were three on-board transit surveys used for this purpose:

- A 2016 Metrorail Passenger Survey collected by the Washington Metropolitan Area Transit Authority (WMATA) in 2016 (April 7 through June 4, 2016) at all of the 91 stations during all time periods of weekdays and weekends.
- A Maryland Area Regional Commuter (MARC) Train Service 2016 On-Board Survey.
- A 2019 Virginia Railways Express (VRE) survey administered to riders on northbound trains in the AM peak period only.

The definition of each market is shown in Table 3-1. We include resident internal-internal (II) trips as a specific market for purposes of survey coding, even though the process of coding the data was not specifically focused on this market. In the table, II trips are for residents, while visitor trips include non-resident internal trips. EI, IE, and EE trips can be made by both residents and non-residents. Please note that inter-regional transit trips made by Amtrak and inter-city buses to/from this region are not considered in the Gen3 Model due to the lack of transit on-board survey data and their relatively small market share.

TABLE 3-1: ON-BOARD SURVEY TRAVEL MARKETS

Market	Definition	Potential data use
Internal-Internal (II)	Resident with both origin and destination trip end within MWCOG region	Check calibration targets for ActivitySim
External-Internal (EI)	Resident or non-resident with external trip origin and internal trip destination.	Trip table for assignment to transit network
Internal-External (IE)	Resident or non-resident with internal trip origin and external trip destination.	Trip table for assignment to transit network
External-External (EE)	Resident or non-resident (likely non-resident) with both trip ends external.	Trip table for assignment to transit network
Visitor trip	Non-resident overnight visitor with both origin and destination trip end within MWCOG region	Trip table for assignment to transit network

Survey data processing for each survey is described below, followed by summaries of the resulting processed data.

3.1 METRORAIL 2016 PASSENGER SURVEY DATA PROCESSING

During weekdays, the survey data was collected for the following five time periods:

- AM Peak (opening to 9:29 am), coded in the data as Time Period 1
- Midday (9:30 am to 2:59 am), coded in the data as Time Period 2
- PM Peak (3:00 pm to 6:59 pm), coded in the data as Time Period 3
- Evening (7:00 pm to 12:00 am), coded in the data as Time Period 4
- Late Night (12:00 am to close), coded in the data as Time Period 5

It should be noted that the AM Peak period as defined in the Metrorail survey was slightly longer than the Gen3 Model's definition of the AM Peak period (6:00 AM to 8:59 AM), but it was assumed to be the same for this analysis. The Evening and Late Night time-periods were combined to match to Gen3 Model's Night time period (7:00 PM to 5:59 AM). The survey Midday time period was assumed to be the same as Gen3 Model's Midday time period (9:00 AM to 2:59 PM) although the start time is one half an hour off.

The Metrorail survey collected information on passengers' home address location, passengers' residence jurisdiction within the WMATA service area, boarding and alighting station locations, trip purpose, mode of access and egress to and from the stations, and demographic information.

The 2016 Metrorail Passenger Survey data file received from the MWCOG-TPB (2016RailPassengerSurveyData-WMATA-Final-ZipCode-TAZ-LatLong.xlsx) had a total of 72,956 survey records of which 62,329 were collected on weekdays and 10,627 were collected on weekends. A breakdown of these survey records is provided below:

- Weekday AM, Midday, PM, Evening, and Late Night – 62,329 survey records
    - Weekday AM Peak – 25,172 survey records
    - Weekday Midday – 10,612 survey records
    - Weekday PM Peak – 21,089 survey records
    - Weekday Evening – 5,168 survey records
    - Weekday Late Night – 288 survey records
- Weekend AM, Midday, PM, Evening, and Late Night – 10,627 survey records (Note: These records were not analyzed for this data processing)

The 62,329 weekday survey records were extracted for further data analysis. The database had a total of 33 columns of attribute data, including home location geocodes, boarding and alighting station names, access and egress modes, and the estimated survey weights by boarding station and time period.

The data processing task involved the following main steps:

1. Origin and Destination TAZ Identification
2. Visitor Trip Identification
3. Transfer Trip Identification (to/from MARC and VRE)
4. EI/IE Trip Identification
5. Trip Weights Computation
6. Origin-Destination (O-D) Table Preparation

Data processing was conducted using both automated and manual processes. The steps are discussed next.

Step 1 - Origin and Destination TAZ Identification

In this step, the origin and destination TAZs were assigned to each trip record. The Metrorail survey does not report the origin and destination location for each trip record. However, the home location and the boarding and alighting stations are reported for each respondent. Therefore, the origin and destination TAZ were coded as the TAZs of the reported boarding and alighting stations. Out of the 62,329 survey records, 927 records (1.5%) had alighting stations recorded as “UNKNOWN”. These 927 records were excluded from any further processing. The remaining 61,402 records had complete information on boarding and alighting stations, and as such, origin and destination TAZs were filled in by joining them using a station-to-TAZ lookup table.

Step 2 - Visitor Trip Identification

As noted in Table 3-1, visitors are defined as non-residents who pay a multiple-day visit to and travel within the MWCOG/TPB Modeled Area (also referred to as “this region” or “the MWCOG region” in this report) (i.e., internal-internal trips by non-residents). For this analysis, visitors are separated from non-resident non-visitors who live in proximity to this region and travel to this region for work or other purposes, which usually does not involve an overnight stay (i.e., internal-external or external-internal trips by non-residents). To identify visitors, we use the home location of the respondents. The Metrorail survey contained home locations for 46,126 (out of 61,402) records. A review of these 46,126 records of geocoded home locations was found to reasonably match the underlying MWCOG TAZs, zip codes, cities, and counties. The home locations of these 46,126 records are depicted in Figure 3-1 and Figure 3-2, where color denotes time period of travel.

The map shows that the majority of the home locations are within or close to the MWCOG region, but some home locations are spread all across the US. The respondents with a home location well outside the MWCOG region are more likely overnight visitors who use Metrorail for an internal trip rather than a commute to or from home. Therefore, we developed a process based on the distance of the home location from the MWCOG Modeled Area boundary to separate visitors from non-resident, non-visitors. A buffer area of 75 miles around the study area was considered to separate visitor trips from all of the external trips. The trips made by respondents who live outside the 75-mile buffer area with access or egress mode other than MARC/VRE/Amtrak were categorized as Visitors Trips. There were 877 visitor trip records. Figure 3-3 illustrates how a 75-mile buffer reasonably captures external commuters' trips while filtering the visitors out.

FIGURE 3-1: HOME LOCATIONS OF METRORAIL TRIPS BY TIME PERIODS

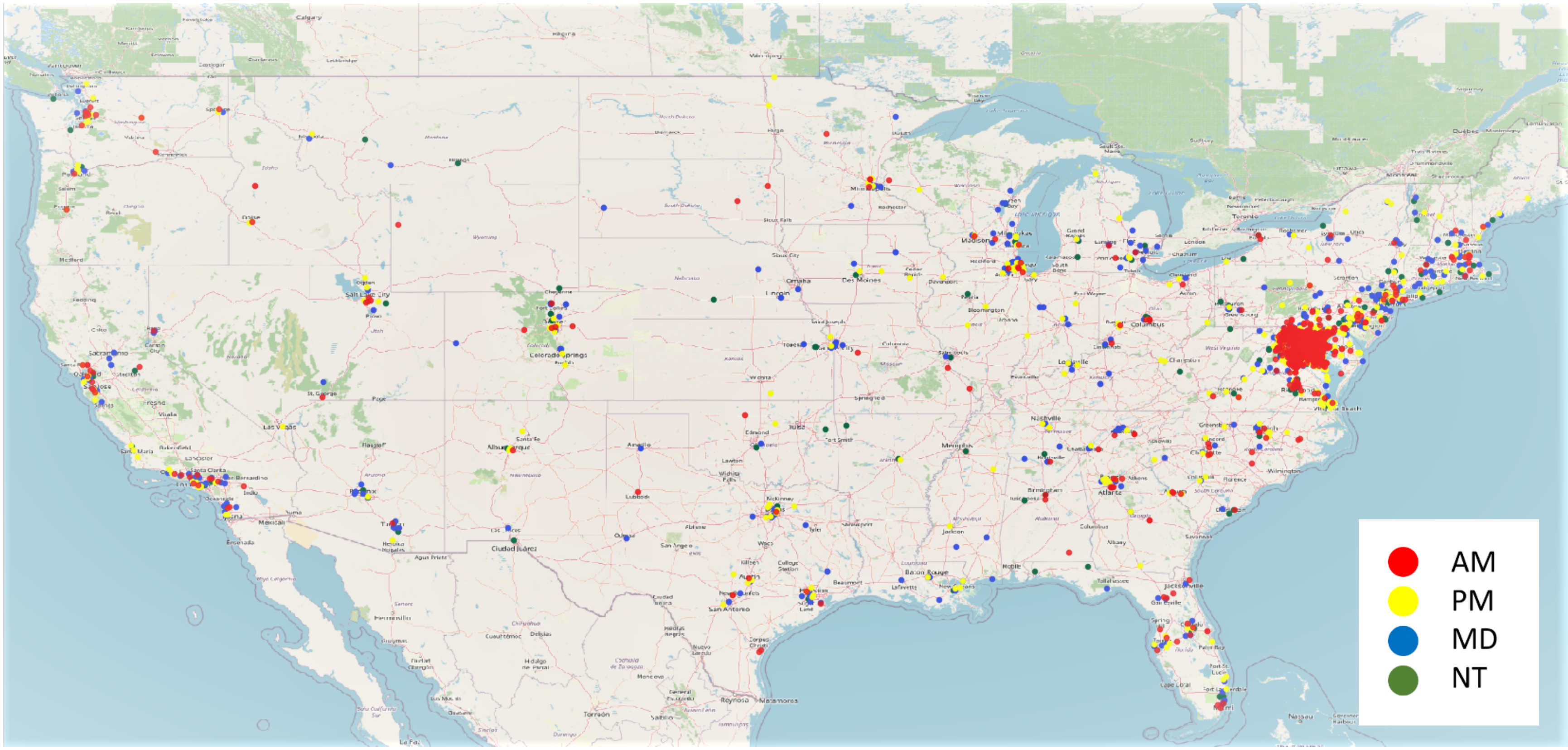


FIGURE 3-2: HOME LOCATIONS OF METRORAIL TRIPS BY TIME PERIODS WITHIN OR CLOSE TO THE MODELED AREA

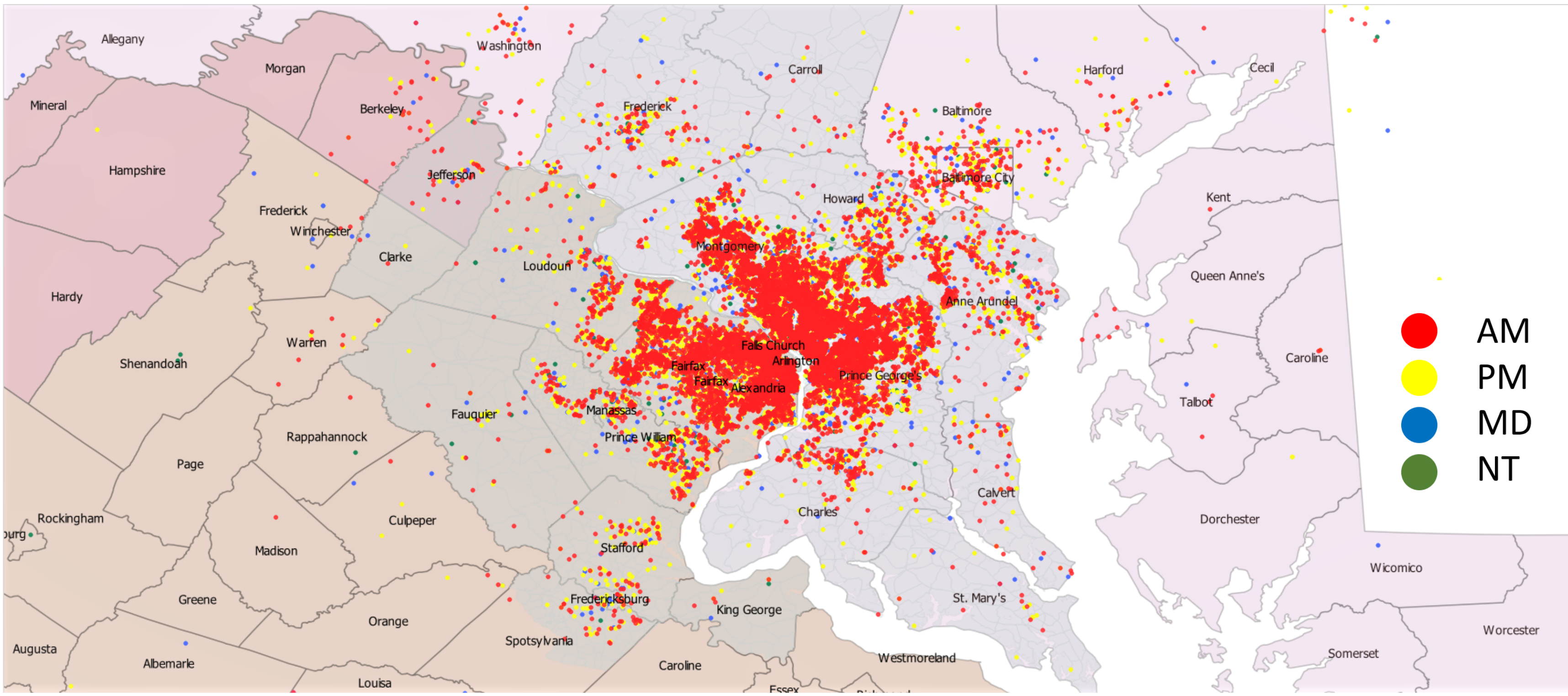


FIGURE 3-3: A 75-MILE BUFFER AREA AROUND THE MODEL STUDY REGION TO SEPARATE VISITOR TRIPS FROM EXTERNAL TRIPS

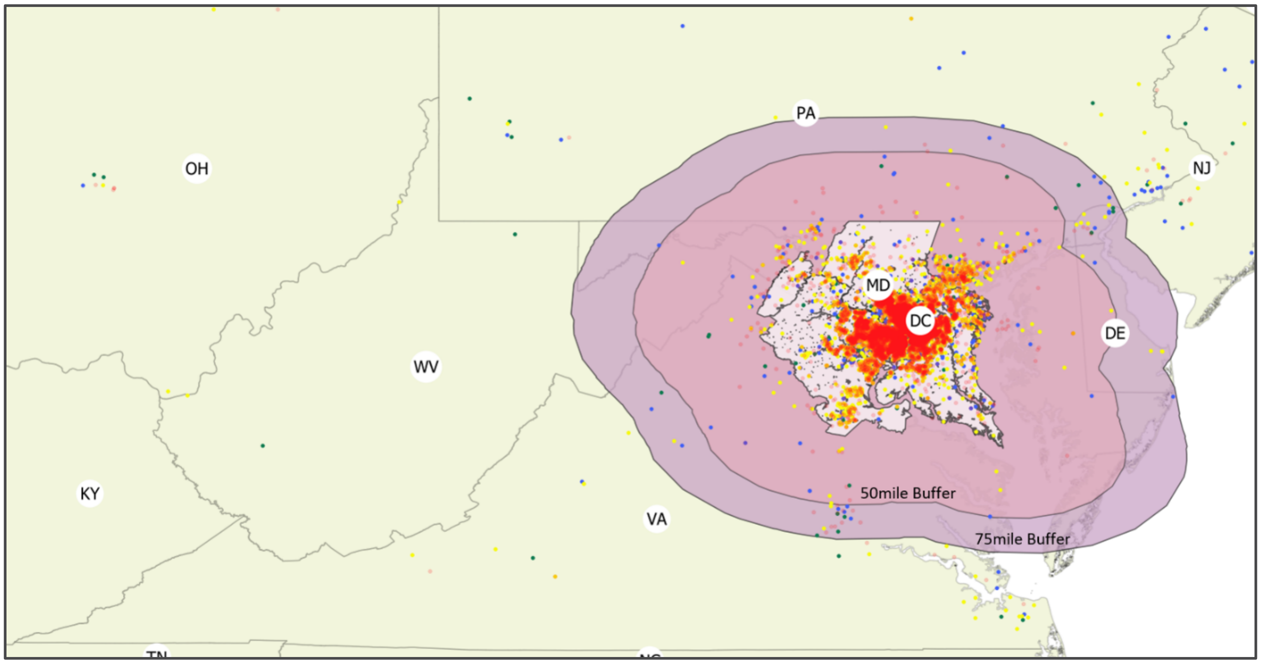
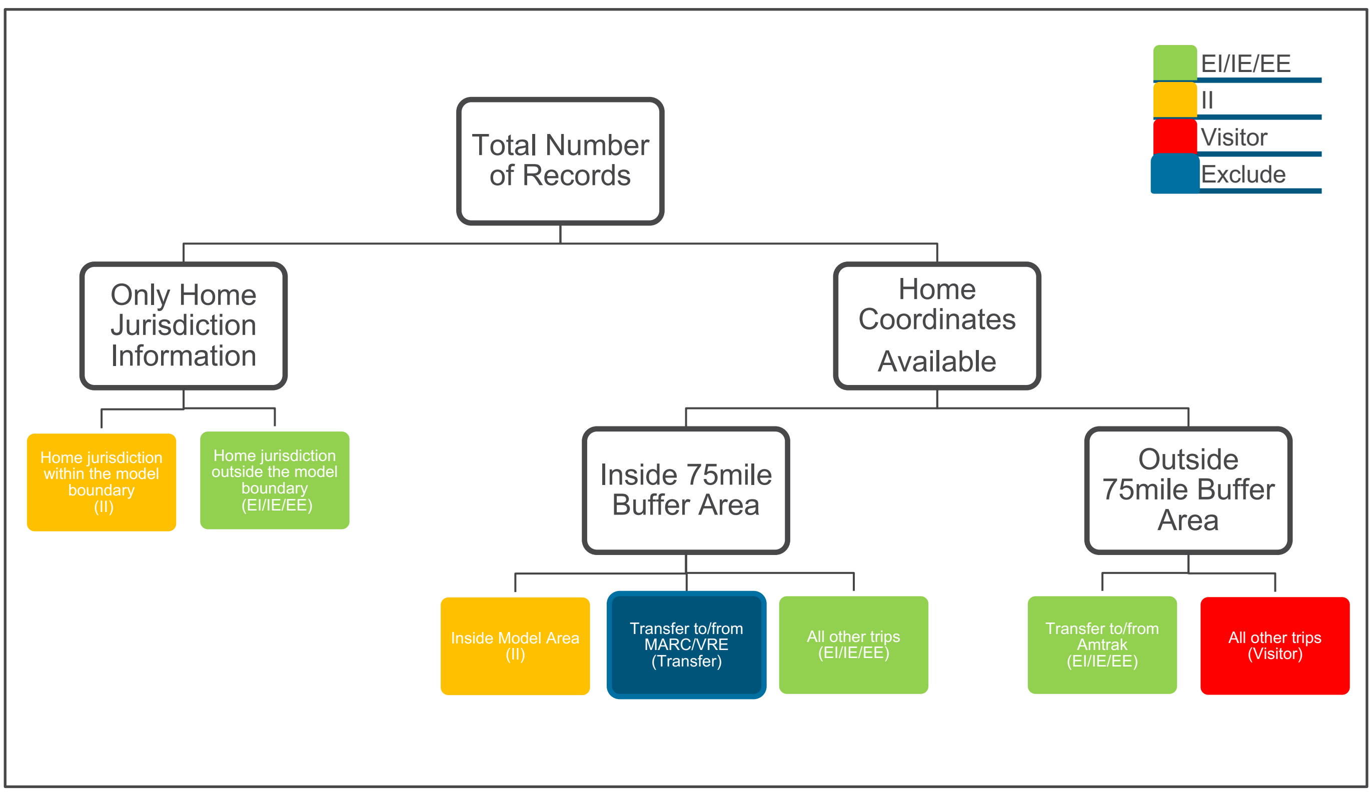


FIGURE 3-4: TRIP CATEGORIZATION FLOW CHART



Step 3 - Transfer Trip Identification

The Metrorail survey data includes many trips that transfer between heavy rail services such as Amtrak, MARC, VRE, and Metrorail. Additional steps were required to separate these trips and categorize them accordingly:

- Transfer Trips to/from MARC and VRE (Excluded from external trips)

  The Metrorail external trips which are made by non-residents living within a 75-mile buffer include trips that transfer to/from MARC and VRE. Including these transfer trips to/from MARC/VRE in Metrorail data will lead to double counting since these are accounted for in the VRE and MARC OBS surveys. Thus, these trips were not included in Metrorail trip tables.

  The external trips inside the 75-mile buffer area with a stated access or egress mode of Amtrak were also categorized as MARC/VRE transfer trips. It is a reasonable assumption that these trips have their access/egress mode as MARC/VRE instead of Amtrak. There were 418 transfer trip records in this category. These were excluded from the Metrorail external trips.

- Transfer Trips to/from Amtrak (Included in external trips)

  The external/visitor trips from the Metrorail survey also contained trips that transferred to/from Amtrak. Since no Amtrak OBS data was collected, all these trips that transferred to/from Amtrak were considered as external trips.

- The external trips made by non-residents living outside the 75-mile buffer area (visitors) with a stated access or egress mode of MARC/VRE were also categorized as Amtrak Transfer Trips. It is a reasonable assumption that these trips have their access/egress mode as Amtrak instead of MARC/VRE. There were 35 such records which were considered as Metrorail external trips.

Step 4 - EI/IE Trip Identification

Figure 3-4 shows how trips are classified in the Metrorail survey by market. Any trip with a home location within the modeled area is coded as an II trip. If the home location is coded as outside the model boundary, and only jurisdiction is known, it is assumed to be an EI, IE, or EE trip. The trip would be retained in the EI/IE/EE trip table though since the external trip end is unknown in the Metrorail survey, only the internal portion is retained. Residents who report a transfer to/from MARC or VRE are excluded from the output trip tables to avoid double-counting. Metrorail trips made by non-residents (those who live outside the modeled area, including visitors and non-resident, non-visitors) are coded either as EE/EI/IE trips if reporting a transfer to/from Amtrak or as visitor II trips. There were 1,222 trip records for EI/IE/EE in total.

There were 15,276 survey records where home geocodes were missing but had a valid response on residence jurisdiction. These 15,276 survey records were coded either as II trips or as IE/EI/EE trips based on whether the residence jurisdiction is inside or outside the model boundary. For example, survey records with residence jurisdictions in Baltimore County or in the City of Baltimore were coded as EI/IE/EE trips.

Two counties in Maryland, namely Carrol and Howard counties, are located inside the model boundary but were grouped as “Other Maryland” in the survey question related to residence jurisdiction (i.e., Question A). Similarly, one county in West Virginia, namely Jefferson County, is located inside the model boundary but was grouped as “Elsewhere” in the survey residence question. There are 651 survey records related to these two residence locations, namely Other Maryland and Elsewhere (with missing home geocodes), which were tagged as EI/IE/EE trips. It is possible that a small portion of these trips are II trips if the home locations are within the three counties inside the model boundary, namely Carrol (MD), Howard (MD), and Jefferson (WV) counties.

Step 5 - Trip Weights Computation

The Metrorail Passenger Survey data utilized in this analysis was from the year 2016. This data needed to be reweighted to match the 2018 WMATA ridership data, consistent with the model base-year.

The Metrorail 2016 survey data included weekday ridership weights by boarding station and time period. The Metrorail 2016 survey data was compared with the available WMATA 2018 Boarding data to compute 2018 factors by boarding station and by time period. The WMATA 2018 Boarding data included an attribute titled “HOUROFDAY” which was used to aggregate the data for four time periods assuming the noted hour as the entry/beginning time. The computation of the 2018 adjustment factors by boarding station and by time period are summarized in Table 3-2Figure 3-2.

TABLE 3-2: YEAR 2018 TRIP WEIGHTING FACTORS BY STATION AND TIME PERIOD

	Daily Boardings from 2016 Metrorail Survey By Gen3 Model Peak Period					Daily Boardings from 2018 Ridership Data by Station and By Hour					Weight Adjustment Factors			
BOARDING STATION	AM	MD	PM	NT	Total	AM	MD	PM	NT	Total	AM	MD	PM	NT
ADDISON ROAD	1,783.2	560.0	322.0	145.0	2,810.2	1,423.1	686.5	258.2	355.9	2,723.7	0.7981	1.2257	0.802	2.4546
ANACOSTIA	2,707.6	1,687.8	1,583.9	442.0	6,421.3	2,087.6	1,686.3	1,496.7	636.3	5,906.9	0.771	0.9991	0.945	1.4396
ARCHIVES-NAVY MEMORIAL-PENN QUARTER	354.4	1,285.0	5,901.6	1,336.0	8,876.9	241.7	1,303.3	5,415.0	1,062.2	8,022.2	0.6821	1.0142	0.9176	0.7951
ARLINGTON CEMETERY	12.0	422.0	349.0	109.0	892.0	14.9	627.9	482.7	6.4	1,131.9	1.2433	1.488	1.383	0.0583
BALLSTON-MU	4,738.6	1,790.6	3,469.6	963.0	10,961.8	3,622.7	2,218.7	2,906.8	789.3	9,537.5	0.7645	1.2391	0.8378	0.8196
BENNING ROAD	1,335.6	568.4	365.7	137.5	2,407.3	1,220.4	805.6	418.1	312.0	2,756.2	0.9137	1.4174	1.1433	2.2687
BETHESDA	3,238.2	1,859.4	3,693.5	1,326.5	10,117.6	2,538.6	2,295.0	3,502.2	1,023.8	9,359.6	0.784	1.2343	0.9482	0.7718
BRADDOCK ROAD	2,900.7	688.4	870.0	239.1	4,698.2	2,449.8	954.2	845.9	309.2	4,559.0	0.8446	1.3862	0.9723	1.2928
BRANCH AVENUE	4,579.4	734.2	423.0	148.0	5,884.6	3,590.9	891.1	337.4	619.5	5,438.9	0.7841	1.2137	0.7977	4.1859
BROOKLAND-CUA	2,300.3	1,286.9	1,979.5	624.0	6,190.7	1,748.6	1,617.2	1,905.3	619.0	5,890.1	0.7601	1.2567	0.9625	0.9921
CAPITOL HEIGHTS	1,154.3	375.3	194.0	74.2	1,797.8	1,001.1	482.6	234.4	229.8	1,947.9	0.8672	1.2861	1.2083	3.0978
CAPITOL SOUTH	865.0	1,563.0	4,095.4	911.0	7,434.4	782.0	1,464.9	3,708.2	581.5	6,536.6	0.9041	0.9372	0.9054	0.6383
CHEVERLY	809.2	200.9	135.5	43.8	1,189.3	577.1	257.5	104.0	98.8	1,037.4	0.7132	1.2817	0.7672	2.2583
CLARENDON	1,606.0	649.4	836.3	776.4	3,868.1	1,749.5	1,090.5	1,101.7	637.5	4,579.2	1.0894	1.6793	1.3173	0.8211
CLEVELAND PARK	2,131.0	799.0	625.6	294.0	3,849.5	1,618.1	1,219.1	608.8	389.7	3,835.7	0.7593	1.5259	0.9732	1.3255
COLLEGE PARK - U OF MD	1,841.8	953.0	1,048.0	306.0	4,148.8	1,356.4	946.1	1,082.6	452.6	3,837.7	0.7365	0.9928	1.033	1.4791
COLUMBIA HEIGHTS	4,455.1	2,410.3	3,136.2	1,545.1	11,546.8	3,134.6	2,898.0	3,279.7	1,451.0	10,763.2	0.7036	1.2023	1.0457	0.9391
CONGRESS HEIGHTS	1,256.5	661.1	346.5	151.4	2,415.4	1,022.4	703.3	476.7	273.4	2,475.8	0.8137	1.0638	1.3757	1.8064
COURT HOUSE	3,580.1	1,187.6	1,943.0	590.0	7,300.7	2,714.5	1,631.1	1,820.7	484.0	6,650.3	0.7582	1.3735	0.9371	0.8203
CRYSTAL CITY	4,284.5	2,416.1	5,221.8	882.9	12,805.3	3,106.1	2,642.1	4,804.2	827.7	11,380.1	0.725	1.0935	0.92	0.9376
DEANWOOD	692.0	267.2	172.0	72.0	1,203.2	588.6	398.6	253.7	154.1	1,394.9	0.8506	1.4917	1.4749	2.1396
DUNN LORING-MERRIFIELD	2,408.3	740.6	648.1	161.0	3,957.9	1,954.9	886.2	722.4	295.6	3,859.1	0.8118	1.1967	1.1146	1.8359
DUPONT CIRCLE	3,676.2	3,118.6	8,188.6	4,069.1	19,052.4	2,588.4	3,799.6	7,631.7	3,306.8	17,326.5	0.7041	1.2183	0.932	0.8127
EAST FALLS CHURCH	2,675.1	672.6	559.5	142.5	4,049.8	2,279.6	905.6	580.6	253.2	4,019.1	0.8522	1.3463	1.0378	1.7771
EASTERN MARKET	2,002.0	1,207.0	1,456.0	1,066.0	5,731.0	1,679.1	1,595.4	1,462.4	755.0	5,491.9	0.8387	1.3218	1.0044	0.7083
EISENHOWER AVENUE	864.6	336.3	373.0	95.0	1,668.9	749.6	452.1	623.2	209.2	2,034.2	0.867	1.3444	1.6709	2.2021
FARRAGUT NORTH	1,218.7	3,507.9	15,960.0	4,286.5	24,973.2	852.3	3,470.2	15,480.6	3,288.4	23,091.6	0.6994	0.9892	0.97	0.7672
FARRAGUT WEST	1,444.3	3,325.2	13,795.5	3,584.4	22,149.4	1,059.5	3,197.0	13,158.8	2,822.5	20,237.8	0.7335	0.9614	0.9538	0.7875
FEDERAL CENTER SW	421.0	1,037.9	3,923.6	424.0	5,806.5	288.9	1,059.9	4,102.1	365.5	5,816.4	0.6862	1.0213	1.0455	0.862
FEDERAL TRIANGLE	181.0	1,289.5	5,960.7	1,024.4	8,455.6	111.3	1,215.1	5,280.1	683.3	7,289.7	0.615	0.9423	0.8858	0.667
FOGGY BOTTOM-GWU	1,895.2	3,220.9	10,053.1	4,238.9	19,408.3	1,628.7	3,898.8	10,194.4	3,819.0	19,540.9	0.8594	1.2104	1.0141	0.9009
FOREST GLEN	1,357.2	311.4	313.0	92.0	2,073.6	1,267.9	430.8	225.6	140.9	2,065.2	0.9342	1.3833	0.7209	1.5314
FORT TOTTEN	3,670.5	1,581.3	1,747.4	625.0	7,624.1	3,296.4	2,020.2	1,812.5	926.6	8,055.6	0.8981	1.2776	1.0373	1.4825

BOARDING STATION	Daily Boardings from 2016 Metrorail Survey By Gen3 Model Peak Period					Daily Boardings from 2018 Ridership Data by Station and By Hour					Weight Adjustment Factors			
	AM	MD	PM	NT	Total	AM	MD	PM	NT	Total	AM	MD	PM	NT
FRANCONIA-SPRINGFIELD	5,112.1	1,147.7	705.8	211.0	7,176.6	3,361.4	1,425.1	738.2	619.4	6,144.1	0.6575	1.2417	1.0459	2.9357
FRIENDSHIP HEIGHTS	2,585.8	1,895.3	2,500.0	941.4	7,922.6	2,280.3	2,323.9	2,537.4	848.8	7,990.3	0.8818	1.2261	1.0149	0.9017
GALLERY PLACE-CHINATOWN	1,734.6	3,957.2	12,147.5	6,803.4	24,642.8	1,314.9	3,717.5	11,650.5	6,393.2	23,076.2	0.758	0.9394	0.9591	0.9397
GEORGIA AVE-PETWORTH	2,590.4	1,375.4	1,231.6	613.8	5,811.3	2,411.2	1,712.4	1,455.4	620.3	6,199.3	0.9308	1.245	1.1817	1.0106
GLENMONT	3,666.2	865.4	585.8	255.0	5,372.4	3,349.3	1,195.9	480.8	568.7	5,594.7	0.9135	1.3819	0.8209	2.2302
GREENBELT	4,058.3	1,014.9	623.7	251.0	5,947.9	2,973.0	1,241.0	696.3	642.3	5,552.5	0.7326	1.2227	1.1164	2.559
GREENSBORO	346.0	231.0	518.0	125.0	1,220.0	276.8	256.1	776.1	164.2	1,473.1	0.7999	1.1086	1.4982	1.3135
GROSVENOR-STRATHMORE	3,734.6	872.0	679.0	135.0	5,420.6	2,894.6	1,125.0	751.1	277.5	5,048.2	0.7751	1.2901	1.1062	2.0553
HUNTINGTON	5,448.0	1,037.3	555.4	182.0	7,222.7	4,261.8	1,391.0	656.2	784.6	7,093.6	0.7823	1.341	1.1816	4.311
JUDICIARY SQUARE	459.9	1,722.0	5,696.9	712.0	8,590.9	311.8	1,669.3	5,257.5	677.3	7,916.0	0.6779	0.9694	0.9229	0.9513
KING STREET-OLD TOWN	2,977.5	1,412.4	2,596.8	900.6	7,887.2	2,218.4	1,724.0	2,142.7	857.1	6,942.1	0.7451	1.2206	0.8251	0.9517
LANDOVER	1,137.7	259.5	135.7	70.0	1,602.9	932.2	363.7	129.4	197.2	1,622.5	0.8194	1.4014	0.9536	2.8176
LARGO TOWN CENTER	3,276.5	576.3	417.0	176.0	4,445.8	2,803.8	695.9	338.8	509.2	4,347.7	0.8557	1.2076	0.8124	2.8933
L'ENFANT PLAZA	2,752.2	2,990.9	14,413.6	1,488.0	21,644.7	2,273.7	3,462.7	13,053.5	1,647.8	20,437.8	0.8261	1.1578	0.9056	1.1074
MCLEAN	625.5	277.8	535.8	85.0	1,524.2	676.5	390.3	940.7	129.5	2,137.0	1.0816	1.4048	1.7556	1.5233
MCPHERSON SQUARE	1,695.9	2,030.5	9,119.0	2,291.5	15,136.8	1,319.7	2,260.8	8,780.0	1,733.0	14,093.6	0.7782	1.1134	0.9628	0.7563
MEDICAL CENTER	811.7	861.9	3,221.9	616.0	5,511.5	687.6	934.2	3,142.7	548.8	5,313.3	0.8471	1.0839	0.9754	0.8909
METRO CENTER	1,448.5	4,586.8	13,560.7	4,893.3	24,489.4	1,155.0	4,025.5	15,045.0	4,501.8	24,727.2	0.7973	0.8776	1.1095	0.92
MINNESOTA AVENUE	953.4	581.2	521.3	158.6	2,214.4	778.4	724.6	682.3	276.2	2,461.5	0.8165	1.2468	1.3089	1.7415
MORGAN BLVD	1,316.7	297.1	170.0	70.0	1,853.8	1,146.4	374.6	176.0	207.5	1,904.5	0.8706	1.2609	1.0355	2.9644
MT VERNON SQUARE 7TH ST-CONVENTION CENTER	1,273.5	723.8	1,766.6	627.0	4,390.9	975.8	1,108.0	1,929.0	749.7	4,762.5	0.7662	1.5308	1.0919	1.1958
NAVY YARD-BALLPARK	1,147.2	997.4	4,046.7	597.1	6,788.4	1,496.9	1,626.2	4,527.2	949.3	8,599.6	1.3048	1.6304	1.1187	1.5899
NAYLOR ROAD	1,574.0	456.9	264.4	175.0	2,470.4	1,222.5	595.1	296.6	351.1	2,465.3	0.7767	1.3023	1.1217	2.0063
NEW CARROLLTON	5,114.8	1,274.7	911.0	247.3	7,547.8	3,854.8	1,443.1	804.9	781.8	6,884.6	0.7537	1.1321	0.8836	3.161
NOMA-GALLAUDET U	2,812.0	1,623.0	4,309.2	715.2	9,459.4	2,521.0	2,343.7	4,293.6	767.9	9,926.2	0.8965	1.4441	0.9964	1.0736
PENTAGON	5,852.6	2,694.3	5,844.7	533.9	14,925.6	4,578.4	2,968.9	5,922.1	721.9	14,191.2	0.7823	1.1019	1.0132	1.3519
PENTAGON CITY	4,922.4	2,642.8	3,021.1	1,291.7	11,878.1	4,615.0	3,506.5	3,155.4	1,463.3	12,740.2	0.9376	1.3268	1.0444	1.1328
POTOMAC AVENUE	1,894.9	725.3	509.8	220.0	3,350.0	1,613.7	1,035.4	585.8	328.1	3,563.1	0.8516	1.4276	1.1491	1.4915
PRINCE GEORGE'S PLAZA	2,190.3	945.1	810.0	386.0	4,331.4	1,774.0	1,139.0	842.9	562.4	4,318.4	0.81	1.2052	1.0406	1.4571
RHODE ISLAND AVENUE-BRENTWOOD	2,754.4	1,475.9	1,206.1	483.0	5,919.5	1,950.2	1,535.3	1,034.8	534.3	5,054.6	0.708	1.0402	0.858	1.1063
ROCKVILLE	2,289.7	940.7	933.0	306.0	4,469.3	1,757.6	999.1	874.1	401.6	4,032.4	0.7676	1.0621	0.9369	1.3125
RONALD REAGAN WASHINGTON NATIONAL AIRPORT	797.7	2,031.0	1,881.8	1,479.2	6,189.7	565.4	2,192.9	1,901.3	1,335.0	5,994.6	0.7088	1.0797	1.0104	0.9025
ROSSLYN	3,730.4	2,433.2	5,641.4	1,122.1	12,927.1	3,123.1	3,263.5	6,103.8	1,262.7	13,753.1	0.8372	1.3412	1.082	1.1253

BOARDING STATION	Daily Boardings from 2016 Metrorail Survey By Gen3 Model Peak Period					Daily Boardings from 2018 Ridership Data by Station and By Hour					Weight Adjustment Factors			
	AM	MD	PM	NT	Total	AM	MD	PM	NT	Total	AM	MD	PM	NT
SHADY GROVE	8,532.0	1,804.7	1,076.1	279.1	11,691.8	6,633.4	2,335.4	1,197.2	1,258.6	11,424.6	0.7775	1.2941	1.1126	4.5095
SHAW-HOWARD UNIVERSITY	1,485.4	770.6	1,270.9	712.0	4,238.9	1,291.9	1,287.5	1,636.4	754.3	4,970.1	0.8697	1.6708	1.2876	1.0594
SILVER SPRING	5,353.8	2,497.9	2,825.7	1,083.9	11,761.3	4,776.3	2,987.0	2,785.1	1,099.4	11,647.8	0.8921	1.1958	0.9857	1.0143
SMITHSONIAN	251.9	1,322.1	5,895.0	585.0	8,053.9	213.4	1,645.4	6,004.4	709.7	8,572.9	0.8471	1.2446	1.0186	1.2132
SOUTHERN AVENUE	2,960.0	983.3	556.1	210.7	4,710.1	2,595.9	1,164.1	484.1	625.6	4,869.6	0.877	1.1839	0.8706	2.9693
SPRING HILL	398.4	329.0	408.0	171.0	1,306.4	364.6	258.1	438.4	129.3	1,190.3	0.9154	0.7843	1.0744	0.7559
STADIUM-ARMORY	1,018.6	560.9	541.5	163.2	2,284.1	961.1	752.6	648.8	210.4	2,572.9	0.9436	1.3417	1.1981	1.2896
SUITLAND	2,987.5	860.3	950.6	180.0	4,978.3	2,170.1	938.9	875.7	600.9	4,585.5	0.7264	1.0914	0.9212	3.3381
TAKOMA	3,224.3	958.6	640.0	233.0	5,055.9	2,370.4	1,436.5	891.1	376.2	5,074.2	0.7352	1.4985	1.3924	1.6144
TENLEYTOWN-AU	1,472.2	1,692.8	2,231.1	855.0	6,251.1	1,394.5	1,732.3	2,435.3	1,067.8	6,629.9	0.9472	1.0233	1.0915	1.2488
TWINBROOK	2,207.5	762.8	981.3	236.0	4,187.6	1,735.5	962.0	908.9	353.1	3,959.4	0.7862	1.2611	0.9262	1.496
TYSONS CORNER	538.0	569.9	1,389.6	880.9	3,378.4	509.5	608.0	1,834.9	675.8	3,628.2	0.947	1.0669	1.3204	0.7671
U STREET/AFRICAN-AMERICAN CIVIL WAR MEMORIAL/CARDOZO	1,767.7	1,286.4	1,523.6	1,950.0	6,527.8	1,398.4	1,248.7	1,615.8	1,244.0	5,506.8	0.791	0.9707	1.0605	0.6379
UNION STATION	9,786.5	5,424.4	12,226.5	3,770.1	31,207.5	8,273.3	6,076.8	10,899.4	3,286.6	28,536.0	0.8454	1.1203	0.8915	0.8718
VAN DORN STREET	2,315.0	395.0	325.0	167.0	3,202.0	1,660.0	541.1	312.7	275.7	2,789.5	0.7171	1.3698	0.9621	1.651
VAN NESS-UDC	2,403.3	1,154.5	1,201.6	388.3	5,147.6	1,977.7	1,844.4	1,481.6	628.4	5,932.1	0.8229	1.5976	1.233	1.6183
VIENNA/FAIRFAX-GMU	7,121.9	1,481.9	946.7	350.0	9,900.5	5,352.8	2,012.2	974.2	923.0	9,262.2	0.7516	1.3579	1.029	2.6371
VIRGINIA SQUARE-GMU	1,851.6	566.0	855.7	239.0	3,512.3	1,724.9	956.1	1,016.4	318.2	4,015.6	0.9315	1.6892	1.1878	1.3314
WATERFRONT	1,420.6	899.3	1,498.4	499.0	4,317.3	1,141.6	1,141.7	1,421.7	566.2	4,271.2	0.8036	1.2695	0.9488	1.1347
WEST FALLS CHURCH-VT/UVA	1,932.0	446.0	249.9	96.0	2,723.9	1,583.4	655.1	208.6	155.4	2,602.4	0.8196	1.4688	0.8345	1.6182
WEST HYATTSVILLE	1,985.0	681.4	533.9	250.0	3,450.4	1,568.0	793.6	499.8	432.9	3,294.3	0.7899	1.1646	0.9361	1.7316
WHEATON	2,205.4	654.2	490.0	232.0	3,581.6	1,645.5	926.5	570.5	422.9	3,565.4	0.7461	1.4163	1.1642	1.823
WHITE FLINT	1,541.2	659.4	1,119.0	282.6	3,602.1	1,322.5	901.0	1,051.4	298.6	3,573.4	0.8581	1.3664	0.9396	1.0568
WIEHLE-RESTON EAST	4,073.7	1,308.7	1,396.7	370.4	7,149.5	4,217.0	1,481.2	1,561.5	777.9	8,037.5	1.0352	1.1318	1.118	2.1003
WOODLEY PARK-ZOO	2,393.8	1,400.9	1,833.9	761.0	6,389.6	2,013.2	1,807.9	1,359.3	579.5	5,759.9	0.841	1.2905	0.7412	0.7616

Step 6 – Origin-Destination (O-D) Table Preparation

The Gen3 Model assigns transit trips in Origin-Destination (O-D) format. Thus, O-D trip tables, rather than P-A trip tables, were developed in this analysis. The trip tables are prepared at the TAZ level and are segmented by time period and access mode.

The trip data were converted to 2018 daily trips by multiplying average trip weight for every boarding station with an adjustment factor by time period. Time period was assigned to each trip record in the survey, which were slightly different from the Gen3 Model Time Periods shown in Table 3-3. O-D trip tables were generated for each of the time period of AM, MD, PM and NT.

TABLE 3-3: GEN 3 MODEL TIME PERIODS

TIME PERIOD	DESCRIPTION	BEGIN TIME	END TIME
1	AM Peak (AM)	6:00 AM	8:59 AM
2	Midday (MD)	9:00 AM	2:59 PM
3	PM Peak (PM)	3:00 PM	6:59 PM
4	Night (NT)	7:00 PM	5:59 AM

An access/egress mode segmentation was created for three core groups: WALK, PNR (park and ride), and KNR (kiss and ride). Table 3-4 shows how different access/egress modes were aggregated into these three core groups. Each trip mode is coded as ACCESS-TRANSIT-EGRESS to trace both access and egress modes. While a person may use auto mode on both access and egress end of the transit trip, typically, travel models allow drive access (either PNR or KNR) only at one end of the transit trip. Consequently, each trip is coded into the following mode combinations:

- WK-TR-WK
- PR-TR-WK
- KR-TR-WK
- WK-TR-PR
- WK-TR-KR

In this analysis, the drive access mode (PNR or KNR) was allowed at either home end of the trip or at access end of the trip for AM and MD trips (for trips with neither end at home). The drive access mode (PNR or KNR) was also allowed at the egress end of the trip for PM and NT trips (for trips with neither end at home).

TABLE 3-4: METRORAIL ACCESS/EGRESS MODE SEGMENTATION

	ACCESS			EGRESS		
GROUP	**WALK**	**PNR**	**KNR**	**WALK**	**PNR**	**KNR**
All trips starting from HOME	All remaining Modes SEE NOTE 1	• Drove alone and parked • Drove/rode with others and parked • Scooter/moped	• Dropped off by someone • Taxi/rideshare (Lyft, Uber, etc.) • Slug	Assigned to WALK regardless of egress mode for simplification		
All trips ending at HOME	Assigned to WALK regardless of access mode for simplification			All remaining Modes SEE NOTE 1	• Drove alone and parked • Drove/rode with others and parked • Scooter/moped	• Dropped off by someone • Taxi/rideshare (Lyft, Uber, etc.) • Slug
Remaining NON-HOME trips during AM & MD time periods	All remaining Modes SEE NOTE 1	• Drove alone and parked • Drove/rode with others and parked • Scooter/moped	• Dropped off by someone • Taxi/rideshare (Lyft, Uber, etc.) • Slug	Assigned to WALK regardless of egress mode for simplification		
Remaining NON-HOME Trips during PM & NT time periods	Assigned to WALK regardless of access mode for simplification			All remaining Modes SEE NOTE 1	• Drove alone and parked • Drove/rode with others and parked • Scooter/moped	• Dropped off by someone • Taxi/rideshare (Lyft, Uber, etc.) • Slug

NOTE 1: Personal bicycle, Capital bikeshare, Walked, Wheelchair, Skateboard, Airplane, Metrobus, Other bus (not specific), Shuttle, Amtrak, MARC, or VRE, ART/Arlington regional transit, Circulator (not specific), CUE/City University Energy Saver, DASH bus, Fairfax Connector, Loudoun County Transit, Prince George's County Transit/The Bus, PRTC/Potomac and Rappahannock Transportation Commission/OMNI Link, Montgomery County Ride-On, TAGS bus, Trolley, Bethesda Circulator, MTA commuter bus , MTA (not specific), commuter bus (not specific), train (not specific), DC street car, Eyre commuter bus, Bolt bus, Washington Flyer, Greyhound, DC Circulator, Megabus, other, selected multiple modes.

3.2 MARC 2016 ON-BOARD SURVEY (OBS) DATA PROCESSING

The OBS data file received from MWCOG had survey records from four separate surveys conducted for the different modes of public transit available in the region, namely light rail, local bus, MARC, and Metrorail. These four public transit surveys were conducted between 2015 and 2018 and collected systemwide passenger movement data for weekdays and weekends.[14] A breakdown of the 25,300 survey samples collected in these four surveys is provided below:

- Light Rail – 1,919 survey records
- Local Bus – 17,763 survey records
- MARC – 3,345 survey records
- Metrorail – 2,273 survey records

These 25,300 survey records had a total of 139 columns of attribute data, including home location geocodes, origin location geocodes, destination location geocodes, boarding stop/station name, alighting stop/station name, access mode, and egress mode.

Under this task, only the 3,345 MARC OBS trip records were processed to prepare average weekday EI/IE/EE trip matrices for the 2018 Base Year MWCOG model. The MARC sample included 278 records for weekend travel. The weekend records were removed from the MARC sample, since the Gen3 Model is intended to estimate travel on an average weekday. The rest of the data processing focused only on the 3,067-weekday records that captured travel on three MARC lines, namely Brunswick Line, Penn Line, and Camden Line.

The data processing task involved the following main steps:

1. Origin and Destination TAZ Identification
2. Visitor Trip Identification
3. EI/IE Trip Identification
4. Trip Weights Computation
5. Origin-Destination (O-D) Table Preparation

These four steps are discussed next.

Step 1 - Origin and Destination TAZ Identification

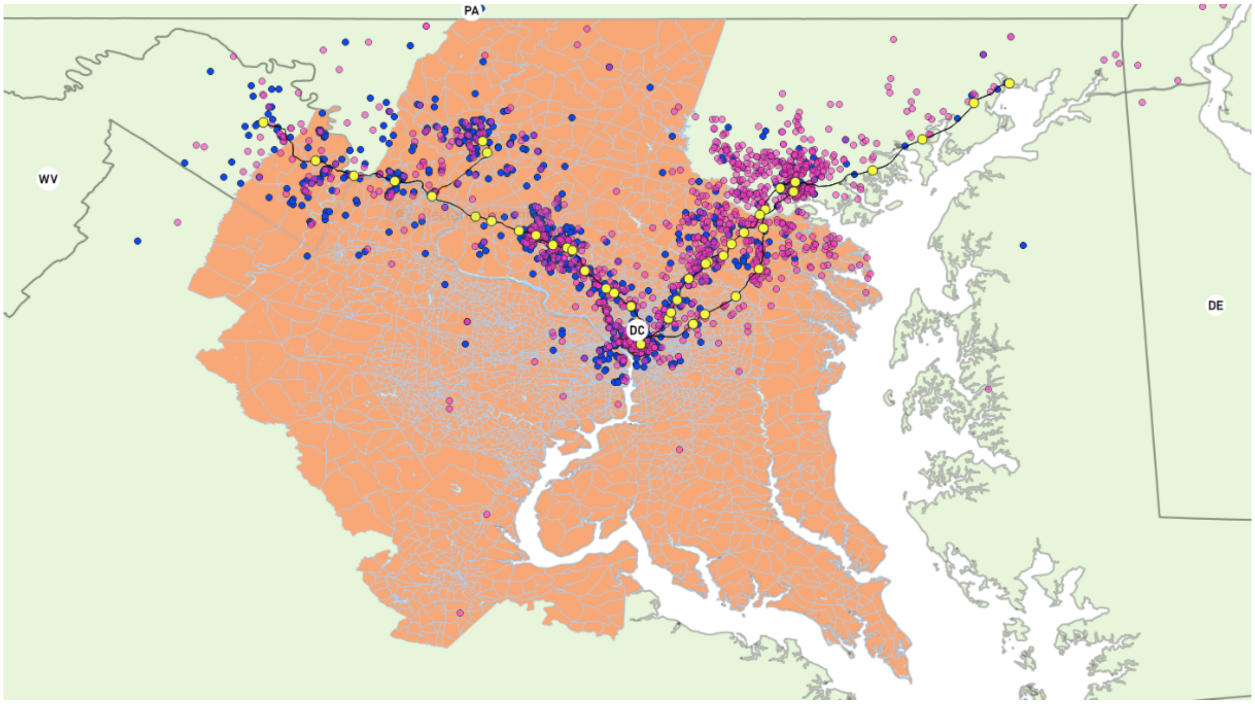
In this step, the origin and destination TAZs of each survey record were identified. The raw data included geocodes (latitude-longitude) for 1,855 trip records out of 3,067-weekday trip records. Figure 3-5 shows the MARC service map. We plotted survey origin and destination locations on

[14] WBA Research, Foursquare ITP, and Jacobs, “Systemwide Summary: 2015-2018 Transit Origin-Destination Survey.”

a map (Figure 3-6) along with the MWCOG TAZ polygon layer and MARC station locations. It can be observed that origin and destination locations are within a reasonable distance from the MARC stations.

FIGURE 3-5: MARC SERVICE MAP

FIGURE 3-6: ORIGIN AND DESTINATION LOCATION OF MARC TRIPS WITH COORDINATES IN THE ORIGINAL DATA

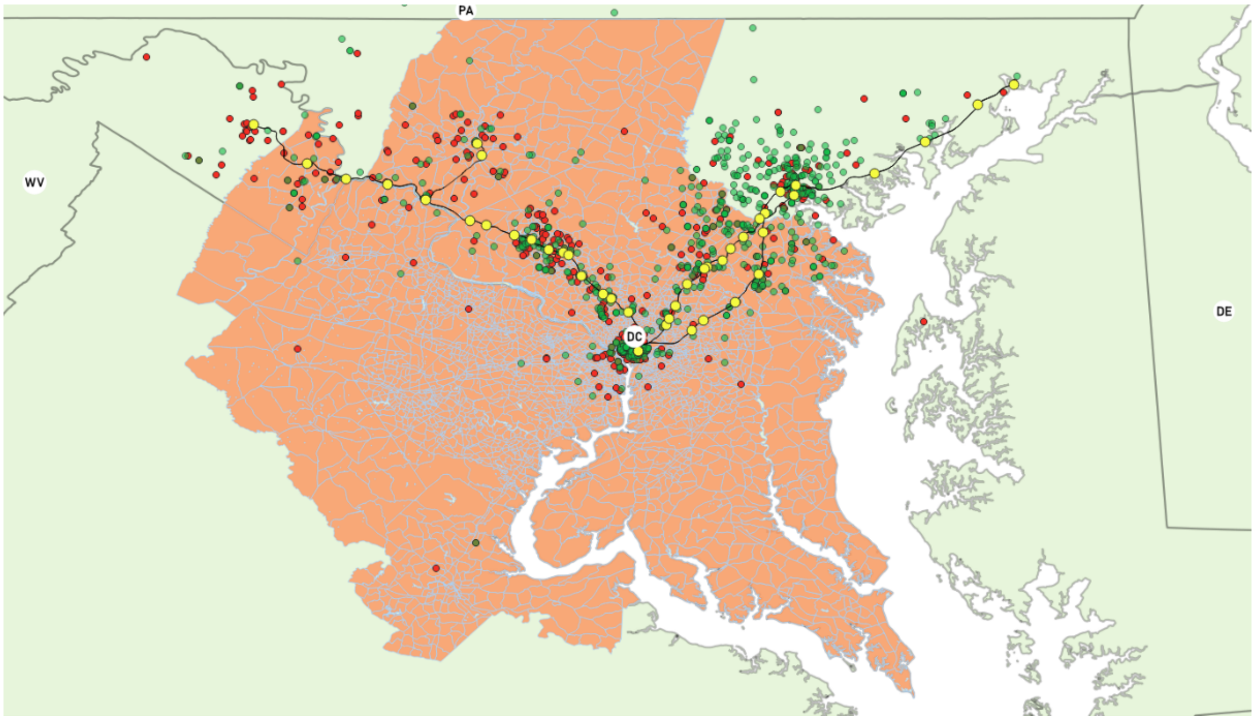


NOTE: The blue dots represent the origin of MARC riders in the original data. The pink dots represent the destination of MARC riders in the original data. The yellow dots represent the location of MARC stations. The model study area is shown in orange.

For the remaining trip records, we used all available origin and destination location information for geocoding. We deployed a Google API-based geocoder (Geocode by Awesome Table[15]) to geocode the home locations and performed similar QA/QC checks on these locations as before (see Figure 3-7).

[15] https://workspace.google.com/marketplace/app/geocode_by_awesome_table/904124517349

FIGURE 3-7: ORIGINS AND DESTINATIONS OF MARC TRIPS WITH COORDINATES COMPUTED USING A GEOCODER



NOTE: The red dots represent the origin of MARC riders as geocoded using the Google API-based Geocoder. The green dots represent the destination of MARC riders as geocoded using the Google API-based Geocoder. The yellow dots represent the location of MARC stations. The model study area is shown in orange.

Next, origin and destination TAZs were assigned to each trip record with valid origin and destination geocodes. The tagging of TAZs was performed using the point-in-polygon feature in QGIS. This feature requires a polygon layer and a point layer. The following files were used for this exercise:

- COGTPB3722.shp (polygon layer)
- Geocoded origin points and destination points (point layers)

The origin and destination TAZs were imputed for trip records with missing location information. Table 3-5 summarizes the methodology to systematically tag/impute the origin and destination TAZs. Using this methodology, the origin and destination TAZs were assigned to 2,761 trip records.

TABLE 3-5: STEPWISE IMPUTATIONS FOR IDENTIFYING ORIGIN AND DESTINATION TAZS

Sub-step	Number of Records	Notes
Review Origin Lat/Long and Destination Lat/Long in the original data	1,855	• 2,341 records had Origin coordinates • 2,278 records had Destination coordinates • 1,855 records had both
Geocode for missing Origin and Destination coordinates using a Google API based Geocoder	1,187	• Geocoded with available address information related to trip origins and destinations
Impute Origin and Destination Lat/Long from Home Lat/Long	29	• 14 Origin Lat/Long were filled with Home Lat/Long as the trips started from home • 15 Destination Lat/Long were filled with Home Lat/Long as the trip ended in home
Tag Origin TAZ based on Origin Lat/Long	2,823	• Using QGIS
Tag Destination TAZ based on Destination Lat/Long	2,775	• Using QGIS
Tag Origin TAZ based on Boarding Station TAZ	52	• For trips with missing Origin TAZ, Walk Access mode, and reported Boarding Station
Tag Destination TAZ based on Alighting Station TAZ	79	• For trips with missing Destination TAZ, Walk Egress mode, and reported Alighting Station
Tag Boarding Station by MARC Line	2,012	• Using reported Boarding Station and station shapefile
Export subsets of Penn Line, Camden Line, and Brunswick Line data for imputing Boarding Station based on Home location	90	• For 43 missing Penn Line Boarding Station for Home-based trips • For 25 missing Camden Line Boarding Station for Home-based trips • For 22 missing Brunswick Line Boarding Station for Home-based trips
Export subsets of Penn Line, Camden Line, and Brunswick Line data for imputing Boarding Station based on Origin location	965	• For 647 missing Penn Line Boarding Station for Non-Home-based trips • For 130 missing Camden Line Boarding Station for Non-Home-based trips • For 188 missing Brunswick Line Boarding Station for Non-Home-based trips
Impute Boarding Station, first from Home Location, and second, from Origin Location when Home location is missing	85	• Out of 90 records from all three MARC lines, 5 records were discarded
Impute Boarding Station from Origin Location	868	• Out of 965 records from all three MARC lines, 97 records were discarded
Tag Origin TAZ based on selected access modes of shuttle services from known origin locations	20	• For missing Origin TAZs and imputed Boarding Stations
Tag Origin TAZ based on external boarding stations	12	• For missing Origin TAZs and imputed Boarding Stations
Tag Destination TAZ based on selected egress modes of shuttle services from known destination locations	5	• For missing Destination TAZs and imputed Alighting Stations
Tag Destination TAZ based on external alighting stations	20	• For missing Destination TAZs and imputed Alighting Stations
Delete records with the same TAZ for both Origin and Destination location	17	• Not meaningful for modeling

Step 2 – Visitor Trip Identification

Visitor trips are internal trips made by non-residents. For the MARC survey, visitor trips were identified based on reported home locations and (if outside the MWCOG modeled area) how far they are from the MWCOG modeled area. We assumed that the respondents with a home location well outside the MWCOG region are probably visitors and not commuters. A buffer area of 75 miles around the study area was drawn to separate the visitor trips from the external trips. The trips with home locations outside the 75-mile buffer area were categorized as visitor trips. Using this methodology, a total of 411 weighted visitor trips were identified, and most of these visitor trips are EI/IE trips. Since visitor trips amounted to approximately 1% of total MARC daily trips, however, a separate visitor trip table was not created. All these EI/IE visitor trips were included as part of the EI/IE/EE trip table.

Step 3 - EI/IE Trip Identification

The EI/IE/EE trip identification process tags any trip with either its origin or destination outside the MWCOG modeled area as an EI/IE/EE trip. The MWCOG model has 3,669 internal TAZs. The locations which are outside the modeling region are not assigned a TAZ by the point-in-polygon process described in the previous section. The trip records with missing location information are not assigned a TAZ, either. The TAZ IDs of valid external trip locations were set to zero while those with missing locations were set to NA. Next, the EI/IE trips were identified. For example, a trip record with origin tagged as 0 and destination tagged as 3333 has been labeled as an EI trip. Similarly, a trip record with origin 3318 and destination tagged as 0 has been labeled as an IE trip. There was only one trip with both origin and destination outside the MWCOG modeled area, which was labeled as an EE trip.

After external trips were identified, a TAZ was assigned to the external end using the following rules:

- The external end on the Penn line was assigned to the nearest external traffic station (TAZID 3714).
- The external end on the Camden line was assigned to the nearest external traffic station (TAZID 3713).
- The external end on the Brunswick line was assigned to the nearest external traffic station (TAZID 3693)

Step 4 - Trip Weights Computation

The MARC OBS data utilized in this analysis were from the year 2016. Since the base year for the Gen3 Model is 2018, this data needed to be reweighted to match the 2018 MARC ridership data. Before computing trip weights, trip records with missing information were removed.

The MARC 2016 OBS data included monthly ridership weights. The monthly ridership weights were converted to daily ridership weights by dividing the monthly ridership weights by 22 (average number of weekdays in a month). Daily boardings were computed for each line and boarding station. This was compared against the observed 2018 MARC ridership data and a weighting factor was computed for each line and station pair. However, we noted that for some stations there were very few trip records or the observed boardings themselves were very low. Therefore, the stations on each line were grouped for weighting to avoid the lumpiness of trip weights.

A total of six station groups were defined for the Penn Line, three groups were defined for the Camden Line, and four station groups were defined for the Brunswick Line based on ridership volume and geographic proximity. Figure 3-8 shows these groups using the same color coding for stations belonging to the same group along the three MARC lines. For example, the red color dots represent the second group of stations along the Brunswick line, and the blue color dots represent the fifth group of stations along the Penn line. Larger stations such as the Union Station, BWI Airport station, and the Penn station were not grouped with other stations. The computed weighting factors by station groups are summarized in Table 3-6.

FIGURE 3-8: STATION GROUPS DEFINED FOR TRIP WEIGHTING

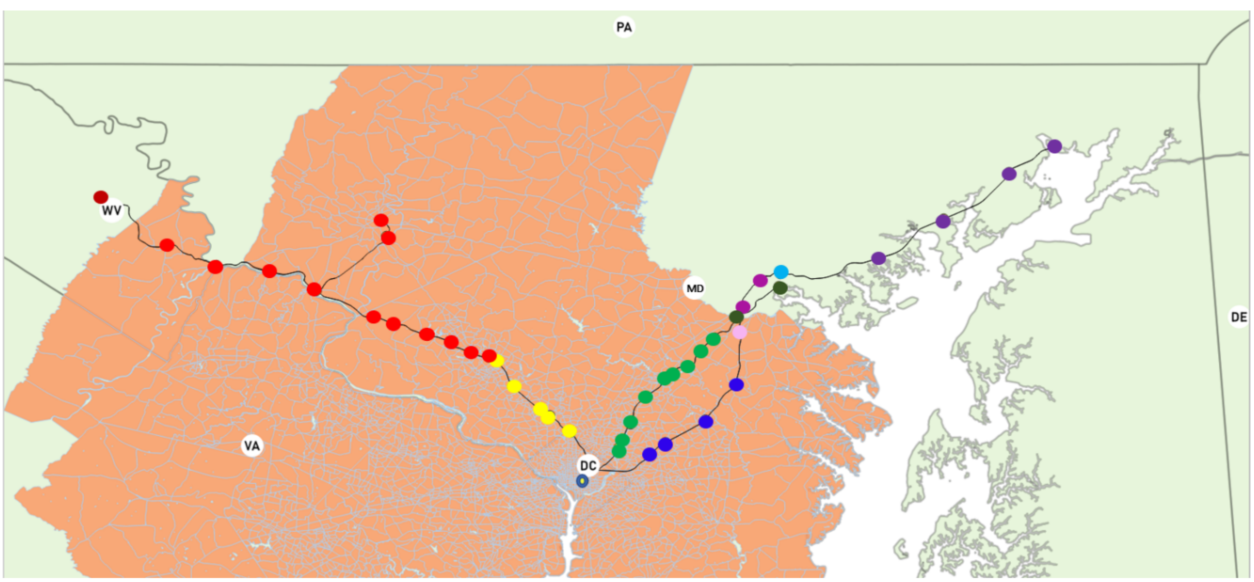


TABLE 3-6: 2018 TRIP WEIGHTING FACTORS BY STATION GROUPS

MARC Line	MARC Station	MARC Station Group	Sum of 2016 Daily Weights from OBS Data (2016 Nov)	2018 Weekday Boarding (2018 Nov)	Factor by Station Group
Penn	Perryville	P-1	113	128	2.2941
Penn	Aberdeen	P-1	81	214	2.2941
Penn	Edgewood	P-1	78	271	2.2941
Penn	Martin Airport	P-1	134	320	2.2941
Penn	Pennsylvania Sta	P-2	6,082	3,947	0.649
Penn	West Baltimore	P-3	488	823	2.4251
Penn	Halethorpe	P-3	477	1,517	2.4251
Penn	BWI Airport	P-4	1,999	2,472	1.2367
Penn	Odenton	P-5	1,071	2,984	3.2667
Penn	Bowie State Univ	P-5	81	819	3.2667
Penn	Seabrook	P-5	0	517	3.2667
Penn	New Carrollton	P-5	540	1,205	3.2667
Penn	Union Station	P-6	11,373	11,484	1.0097
Camden	Camden	C-1	311	429	1.3165
Camden	St. Denis	C-1	23	11	1.3165
Camden	Dorsey	C-2	476	585	1.1977
Camden	Jessup	C-2	8	1	1.1977
Camden	Savage	C-2	385	454	1.1977
Camden	Laurel Race Trac	C-2	0	2	1.1977
Camden	Laurel	C-2	550	680	1.1977
Camden	Muirkirk	C-2	371	440	1.1977
Camden	Greenbelt	C-2	59	63	1.1977
Camden	College Park	C-2	178	238	1.1977
Camden	Riverdale	C-2	91	73	1.1977
Camden	Union Station	C-3	1,642	2,436	1.4836
Brunswick	Martinsburg	B-1	99	87	0.8819
Brunswick	Duffields	B-2	103	85	0.948
Brunswick	Harpers Ferry	B-2	73	52	0.948
Brunswick	Brunswick	B-2	387	412	0.948
Brunswick	Frederick	B-2	128	89	0.948
Brunswick	Monocacy	B-2	205	316	0.948

MARC Line	MARC Station	MARC Station Group	Sum of 2016 Daily Weights from OBS Data (2016 Nov)	2018 Weekday Boarding (2018 Nov)	Factor by Station Group
Brunswick	Point of Rocks	B-2	352	319	0.948
Brunswick	Dickerson	B-2	30	18	0.948
Brunswick	Barnesville	B-2	95	83	0.948
Brunswick	Boyds	B-2	33	20	0.948
Brunswick	Germantown	B-2	798	842	0.948
Brunswick	Metro Grove	B-2	377	235	0.948
Brunswick	Gaithersburg	B-2	579	525	0.948
Brunswick	Washington Grove	B-3	138	43	0.6359
Brunswick	Rockville	B-3	529	493	0.6359
Brunswick	Garrett Park	B-3	353	50	0.6359
Brunswick	Kensington	B-3	302	210	0.6359
Brunswick	Silver Spring	B-3	710	496	0.6359
Brunswick	Union Station	B-4	3,937	3,013	0.7652
All 3 Lines	Union Station		16,952	16,933	
SYSTEMWIDE			35,838	39,501	

Step 5 – Origin-Destination (O-D) Table Preparation

Trip tables in origin-destination format are prepared at the TAZ level and are segmented by time period and access mode. The processed trip records were filtered to use only EI/IE/EE trips to generate the IE/EI/EE trip tables. A time period was assigned to each trip record according to the reported start time following the Gen3 Model time period definitions as illustrated in Table 3-3. The O-D trip tables were generated for each of the time periods.

An access/egress mode segmentation was created, namely WALK (walking), PNR (park and ride), and KNR (kiss and ride) as required for Gen3 Model. Table 3-7 shows how different modes were aggregated into these three broad access mode groups. Each trip mode is coded as ACCESS-TRANSIT-EGRESS to trace both access and egress modes. While a person may use auto mode on both access and egress end of the transit trip, typically, travel models allow drive access (either PNR or KNR) only at one end of the transit trip. Consequently, each trip is coded into the following mode combinations:

- WK-TR-WK
- PR-TR-WK
- KR-TR-WK

- WK-TR-PR
- WK-TR-KR

In this analysis, the drive access mode (PNR or KNR) was allowed at either the home end of the trip or at the access end of the trip for AM and MD trips (for trips with neither end at home). The drive access mode (PNR or KNR) was also allowed at the egress end of the trip for PM and NT trips (for trips with neither end at home).

TABLE 3-7: MARC ACCESS/EGRESS MODE SEGMENTATION

	ACCESS			EGRESS		
GROUP	**WALK**	**PNR**	**KNR**	**WALK**	**PNR**	**KNR**
Internal-Internal, and internal end of EI/IE trips	SEE NOTE 1	SEE NOTE 2	SEE NOTE 3	SEE NOTE 1	SEE NOTE 2	SEE NOTE 3
Internal-Internal, and internal end of EI/IE trips	SEE NOTE 1	SEE NOTE 2	SEE NOTE 3	Assigned to WALK regardless of egress mode for simplification		
Internal-Internal, and internal end of EI/IE trips	Assigned to WALK regardless of access mode for simplification			SEE NOTE 1	SEE NOTE 2	SEE NOTE 3
External end of IE/EI trips and boarding/alighting station outside model boundary	Assigned to WALK regardless of access mode as the boarding station is outside the model boundary			Assigned to WALK regardless of egress mode as the alighting station is outside the model boundary		
External end of IE/EI trips and boarding/alighting station inside model boundary and MARC was the first transit service used	SEE NOTE 1	SEE NOTE 2	SEE NOTE 3	SEE NOTE 1	SEE NOTE 2	SEE NOTE 3
External end of IE/EI trips and boarding/alighting station inside model boundary and MARC was not the first transit service used	Assigned to WALK regardless of access mode as the MARC boarding station is not the first stop			Assigned to WALK regardless of egress mode as the MARC alighting station is not the last stop		
EE trips and boarding station inside model boundary	SEE NOTE 1	SEE NOTE 2	SEE NOTE 3	Assigned to WALK regardless of egress mode for simplification		
EE trips and boarding station outside model boundary or MARC is not the first transit service	Assigned to WALK regardless of access mode as the boarding station is outside the model boundary			Assigned to WALK regardless of egress mode for simplification		
All trips starting from home	Accept the Reported Access as WALK, PNR or KNR			Assigned to WALK if the access is PNR or KNR		
All trips ending at home	Assigned to WALK if the egress is PNR or KNR			Accept the Reported Egress as WALK, PNR or KNR		
All non-home trips during AM and MD time periods	Accept the Reported Access as WALK, PNR or KNR			Assigned to WALK if access is PNR or KNR		

	ACCESS			EGRESS		
GROUP	WALK	PNR	KNR	WALK	PNR	KNR
All non-home trips during PM and NT time periods	Assigned to WALK if the egress is PNR or KNR			Accept the Reported Egress as WALK, PNR or KNR		

NOTE 1: WALK includes: a) Walked/wheelchair only, b) Bicycle, c) Bus/subway/Citylink/any public transportation other than plane or Amtrak, d) Other
NOTE 2: PNR includes a) Drove a car, b) Drove a scooter
NOTE 3: KNR includes: a) Rode with someone, b) Taxi or App based transportation service such as Uber, c) Shuttle bus (not specific), d) School shuttle bus (not specific), e) Work shuttle bus (not specific), f) Airport shuttle bus, g) Johns Hopkins shuttle bus, h) UMD shuttle bus, i) Gallaudet shuttle bus. j) Hospital shuttle

3.3 VRE 2019 ON-BOARD SURVEY (OBS) DATA PROCESSING

The 2019 VRE survey instrument had 9 questions. The responses were collected from riders on northbound trains in the AM peak period only. The raw survey data has 6,299 records with 30 attributes. The data processing dropped records that were missing key information such as the geographic location of the trip origin or destination and added new variables to aid in the generation of trip tables. The final processed dataset has 6,291 records and 52 attributes for the AM peak period. For the PM peak period, 6,291 synthetic records were created by transposing the AM peak period records. The combined dataset including the synthetic trip records for both AM and PM periods was used to generate the IE/EI/EE trip tables.

The data processing included the following main steps:

- Origin and Destination TAZ Identification
- Visitor Trip Identification
- EI/IE Trip Identification
- Trip Weights Computation
- OD Table Preparation

These steps are discussed in detail below.

Origin and Destination TAZ Identification

In this step, origin and destination TAZs were assigned to each trip record. Latitude-longitude coordinates of home locations were available in the data. Figure 3-9 shows the VRE system map. We plotted the home locations on a map (Figure 3-10) and found that home locations in some cases were mapped in states far beyond the MWCOG region. We compared this against other address information in the survey data and found the coordinates to be erroneous. Figure

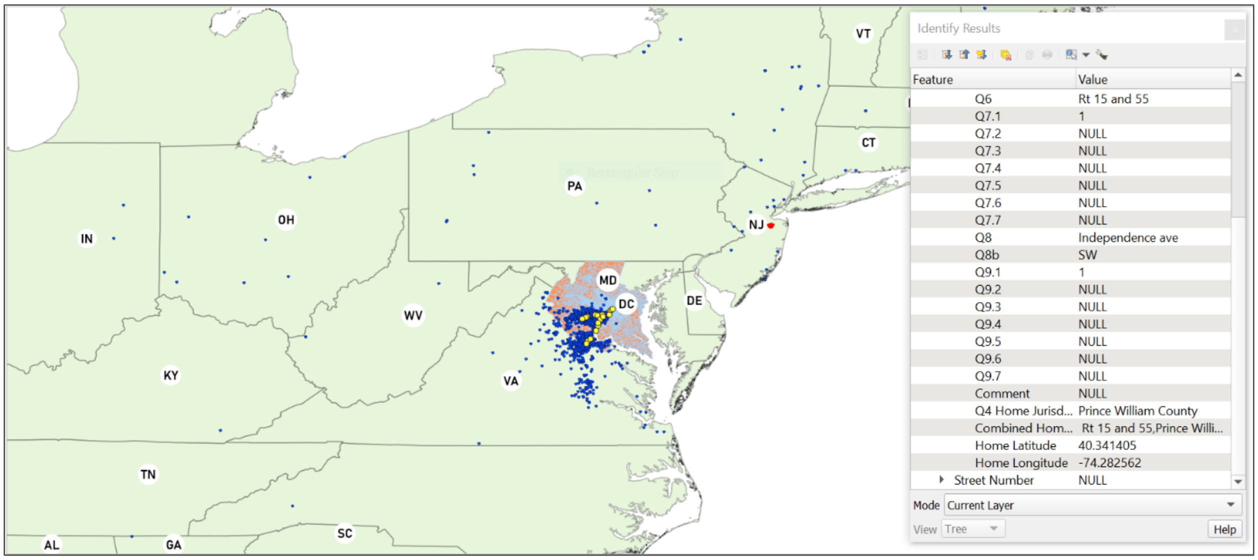
3-10 also shows an example home location (red dot) that should have been geocoded inside the modeled area (shown in orange) and not outside the region based on the home county information (Prince William County, VA for the example trip record). These incorrect geocodes were likely due to incomplete street address, county, or zip code information.

Due to the widespread geocoding errors in the raw data, we deployed a Google API-based geocoder (Geocode by Awesome Table[16]) to geocode the home addresses. We used all available location information in the survey data for geocoding such as home zip code and the associated city, county, state, and country information. Figure 3-11 shows the home locations with updated geocodes. The pink dots represent the location of VRE riders using correctly geocoded home information.

FIGURE 3-9: VRE SYSTEM MAP

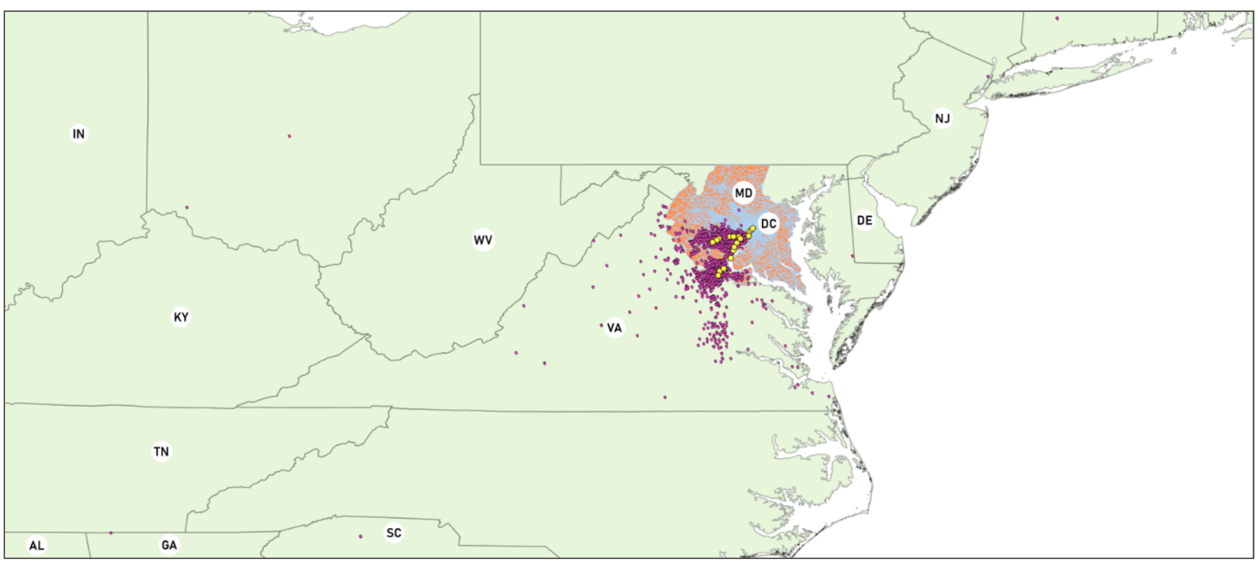
16 https://workspace.google.com/marketplace/app/geocode_by_awesome_table/904124517349

FIGURE 3-10: ORIGINS OF VRE TRIPS USING INCORRECT HOME LATITUDE-LONGITUDE IN THE ORIGINAL DATA



NOTE: The navy dots represent the location of the VRE rider's home location based on the original data provided. The yellow dots represent the location of VRE stops.

FIGURE 3-11: ORIGINS OF VRE TRIPS USING CORRECTLY GEOCODED HOME ADDRESS



NOTE: The pink dots represent the location of the VRE rider's home location based on the corrected geocodes. The yellow dots represent the location of VRE stops.

The next step in the process involved assigning a TAZ to the origin location. Because the survey was collected in the AM peak period on northbound trains, we assumed that most of these trips would be commute trips from home to work. Therefore, the home location was used as the proxy for the trip origin location. The tagging of TAZs was performed using the point-in-polygon feature in QGIS. This feature requires a polygon layer and a point layer. The following files were used for this exercise:

- Polygon layer: MWCOG TAZ shapefile (COGTPB3722.shp)
- Point layer: Geocoded home locations from the VRE OBS

Similarly, the trip destinations were also geocoded using the corrected address in the geocoder and a TAZ was assigned using QGIS (none of the trip destinations in the raw data were geocoded). To reduce the overall effort of performing QA/QC on destination addresses, geocoding was performed only for trips with a non-walk egress mode. Walk egress distances are typically short and the trip destination and alighting station in most cases will be within the same TAZ. Therefore, in the case of walk-egress, it is assumed that the destination TAZ is the same as the TAZ of the alighting stop. Table 3-8 shows the split of trips by egress mode and how a destination TAZ was assigned in each case.

TABLE 3-8: PROCESS OF DESTINATION TAZ ASSIGNMENT

DECISION CRITERIA	WALK EGRESS MODE	NON-WALK EGRESS MODE	TOTAL RECORDS
Destination TAZ = Alighting Station	4,094 (See Note 2)	304 (See Note 1)	4,398
Destination TAZ = Geocoded Actual Destination	None	1,901	1,901
Total Count	4,094	2,205	6,299

Note 1. Out of these 304 trips,158 trips had missing Destination zip-codes and 146 trips had only Destination zip-code information. These 146 trips were found to be distributed over 69 different zip codes. If these 146 trips were distributed over 69 zip-codes (and many TAZs) by employment weights, it would have many fractional trips. Therefore, this step was avoided. Consequently, the alighting stations' TAZ was assigned as Destination TAZ. Also, very few of these trips are EI/IE trips (12 out of 146).

Note 2. For walk egress mode(Q9.1=1), alighting stations' TAZ has been assigned as destination TAZ. Only 8 out of 4094 records have a missing or incorrect station. These 8 records were manually corrected to the nearest stations based on their Destination Address.

Visitor Trip Identification

For the VRE survey, visitor trips were identified based on home locations and how far they are from the MWCOG modeled area. We assumed that the respondents with a home location well outside the MWCOG region are probably visitors and not commuters. A buffer area of 75 miles around the study area was drawn to separate visitor trips from the external trips. Those trips falling outside the 75-mile buffer area were categorized as visitor trips. This visitor trip identification process yielded 160 visitor trips, which amounted to approximately 1% of total

VRE daily trips. However, most of these visitor trips are EI/IE trips. A separate visitor EI/IE trip table was not created. Instead, these trips were included as part of the regular non-visitor EI/IE trip table.

EI/IE/EE trip identification

Any trip with either origin or destination outside the modeling region is tagged as an EI/IE/EE trip. The MWCOG model has 3,669 internal TAZs. The locations which are outside the modeling region are not assigned a TAZ by the point-in-polygon process described in the previous section. The trip records with missing location information are also not assigned a TAZ. The TAZ IDs of valid external trip locations were set to zero while those with missing locations were set to NA. Next, the EI/IE/EE trips were identified. For example, a trip record with origin tagged as 0 and destination tagged as 3333 is tagged as an EI trip. Similarly, a trip record with origin 3318 and destination tagged as 0 is labeled as an IE trip. There were no EE trips in the VRE OBS data.

After external trips were identified, a TAZ was assigned to the external end using the following rules:

- The external end on the Manassas line was assigned to the TAZ of the Broad Run/Airport Station (TAZID 2609).
- The external end on the Fredericksburg line was assigned to the TAZ of the Spotsylvania Station (TAZID 3583).

Trip Weights Computation

After the AM peak period data was processed, the PM peak period data was synthetically generated by transposing the AM period trips. This is under the assumption that symmetry exists between AM and PM travel. Transposition of AM period trips involved the following changes to key attributes:

- Time Period: AM was changed to PM.
- Origin and Destination information were swapped.
- Access and Egress Mode were swapped.
- Boarding Stations and Alighting Stations were swapped.

Next, the AM and PM (synthetic) trips were combined into a daily set of 12,588 trips. Daily boardings were computed by line (Fredericksburg and Manassas) and boarding station. This was compared against the observed VRE boarding data and a weighting factor was computed for each line and station pair. The average monthly boarding data for October 2017 was used for the reweighting process. The daily ridership was calculated by dividing the average monthly ridership by 22 (number of weekdays in that month). For stations serving both lines, the monthly ridership data was combined before calculating the daily ridership. The weighting factors were assigned to trip records based on the boarding station and line combination.

While VRE does operate reduced service during the MD period, an MD trip table could not be estimated using the available information. This is a potential source of error since the MD trips are being assumed to be returning in the PM period. However, this should not have a significant impact since the share of MD trips is likely small.

OD Table Preparation

EI/EI matrices in origin-destination format were prepared for AM and PM peak periods. An access/egress mode segmentation was created, namely WALK (walking), PNR (park and ride), and KNR (kiss and ride). Table 3-9 shows how different modes were aggregated into these three access mode groups. Each trip mode is coded as ACCESS-TRANSIT-EGRESS to trace both access and egress modes. While a person may use auto mode on both access and egress end of the transit trip, typically, travel models allow drive access (either PNR or KNR) only at one end of the transit trip. Less than 1% of the trips reported walk access at the home end. Therefore, walk access trip records were recoded as KNR. Consequently, each trip is coded into the following mode combinations:

- PR-TR-WK
- KR-TR-WK
- WK-TR-PR
- WK-TR-KR

The drive access mode (PNR or KNR) was allowed only at the home/access end of the trip for AM trips. For PM trips, the drive access mode (PNR or KNR) is allowed at the home/egress end of the trip.

TABLE 3-9: VRE ACCESS/ EGRESS MODE SEGMENTATION

	ACCESS			EGRESS		
GROUP	**WALK**	**PNR**	**KNR**	**WALK**	**PNR**	**KNR**
Internal-Internal II, AM	Bus, Bicycle, Walk, Other [SEE NOTE 1]	Drove alone	Dropped Off, Carpool/Vanpool	Assigned to WALK regardless of egress mode for simplification		
Internal-Internal II, PM	Assigned to WALK regardless of access mode for simplification			Bus, Bicycle, Walk, Metrorail, Shuttle Bus, Other [SEE NOTE 1]	Drove alone	Dropped Off, Carpool/Vanpool, Taxi
External-Internal EI, AM	(SEE NOTE 2)	Drove alone	Dropped Off, Carpool/ Vanpool, Bus, Bicycle, Walk, Other [SEE NOTE 1]	Assigned to WALK regardless of egress mode for simplification		
Internal-External IE, AM	Bus, Bicycle, Walk, Other [SEE NOTE 1]	Drove alone	Dropped Off, Carpool/Vanpool	Assigned to WALK regardless of egress mode for simplification		
Internal-External IE, PM	Assigned to WALK regardless of access mode for simplification			(SEE NOTE 2)	Drove alone	Dropped Off, Carpool/Vanpool, Bus, Bicycle, Walk, Metrorail, Shuttle Bus, Taxi, Other [SEE NOTE 1]
External-Internal EI, PM	Assigned to WALK regardless of access mode for simplification			Bus , Bicycle, Walk, Metrorail, Shuttle Bus, Other [SEE NOTE 1]	Drove alone	Dropped Off, Carpool/Vanpool, Taxi

NOTE 1: Any response under 'Other' is placed under WALK unless specified. If a mode of transportation is specified in 'Other', it is placed under the WALK, PNR or KNR based on professional judgement.
NOTE 2: For EI/IE trips, WALK access/egress at the external ends are not allowed regardless of the reported mode because less than 1% of the trips reported using Walk access mode.

3.4 ON-BOARD SURVEY CODING RESULTS

This section provides a summary of three transit On-Board Survey (OBS) processing results to provide a comparison of the trip characteristics of the Metrorail, MARC, and VRE commuter rail riders. The transit trip characteristics have been categorized into five travel markets (as previously defined in Table 3-1), namely Internal-Internal (II), Internal-External (IE), External-Internal (EI), External-External), and Visitor.

Table 3-10 provides a breakdown of the five travel markets in terms of the number of unexpanded trip records from each of the transit OBS. For example, the 2016 Metrorail OBS resulted in 61,402 valid trip records, of which 58,246 II trips (95%) had both origin and destination trip-ends inside the COG model region, 1,867 IE/EI/EE trips (3%) had at least one trip-end outside the COG model region, and 1,289 trips (2%) were made by visitors within the COG model region. Similarly, for the three transit systems combined, the results show 71,742 trips (93%) as internal trips, 3,692 trips (5%) as external trips, and 1,316 trips (2%) as Visitor trips.

TABLE 3-10: COMPARISON OF TRANSIT TRAVEL MARKETS – UNEXPANDED TRIP RECORDS

Survey	Line	II	IE	EI	EE	Visitor	Total
Metrorail	Total	58,246	629	1,222	16	1,289	61,402
	Percent	94.9%	1.0%	2.0%	0.0%	2.1%	100.0%
MARC	Brunswick	1,043	59	71	2	2	1,177
	Camden	314	62	63	1	0	440
	Penn	287	641	186	22	7	1,143
	Total	1,644	762	320	25	9	2,760
	Percent	59.6%	27.6%	11.6%	0.9%	0.3%	100.0%
VRE	Fredericksburg	6,128	269	271	0	14	6,682
	Manassas	5,724	89	89	0	4	5,906
	Total	11,852	358	360	0	18	12,588
	Percent	94.2%	2.8%	2.9%	0.0%	0.1%	100.0%

Survey	Line	II	IE	EI	EE	Visitor	Total
All	Total	71,742	1,749	1,902	41	1,316	76,750
	Percent	93.5%	2.3%	2.5%	0.1%	1.7%	100.0%

Table 3-11 provides a breakdown of the same five travel markets in terms of the number of expanded 2018 Weekday Boardings for each of the transit OBS[17]. For example, the 2016 Metrorail OBS was expanded to 641,227 weekday boardings for year 2018, the 2016 MARC OBS was expanded to 39,498 weekday boardings for year 2018, and the VRE OBS was expanded to 18,491 weekday boardings for year 2018. With three transit systems combined, the results show 641,227 II trips (92%), 42,589 IE/EI/EE trips (6%), and 15,401 Visitor trips (2%).

TABLE 3-11: COMPARISON OF TRANSIT TRAVEL MARKETS – EXPANDED 2018 WEEKDAY BOARDINGS

Survey	Line	II	IE	EI	EE	Visitor	Total
Metrorail	Total	604,825	7,115	14,082	165	15,041	641,227
	Percent	94.3%	1.1%	2.2%	0.0%	2.3%	100.0%
MARC	Brunswick	6,690	485	189	11	12	7,387
	Camden	4,199	703	494	16	0	5,411
	Penn	8,083	10,351	7,133	812	322	26,700
	Total	18,972	11,539	7,816	838	334	39,498
	Percent	48.0%	29.2%	19.8%	2.1%	0.8%	100.0%
VRE	Fredericksburg	9,242	425	355	0	21	10,044
	Manassas	8,188	126	128	0	5	8,448

[17] The expanded weekday boradings were generated using the following weight fields in each OBS:

1. Metrorail: Avg_Daily_2018
2. MARC: Avg_Daily_2018
3. VRE: Trips_2018

	Total	17,431	551	483	0	26	18,491
	Percent	94.3%	3.0%	2.6%	0.0%	0.1%	100.0%
All	Total	641,227	19,206	22,381	1,003	15,401	699,217
	Percent	91.7%	2.7%	3.2%	0.1%	2.2%	100.0%

Figure 3-12 shows the Metrorail stations experiencing moderate to high number of external trips, along with the Union Station, Farragut North, Metro Center, Gallery Place, and L'Enfant Plaza stations showing as the Top 5 stations in terms of external travel markets (EI, IE, and EE). Similarly, Figure 3-13 shows the other Metrorail stations experiencing low number of external trips, or less than 200 trips.

FIGURE 3-12: METRORAIL STATIONS WITH MODERATE TO HIGH NUMBER OF EXTERNAL TRIPS

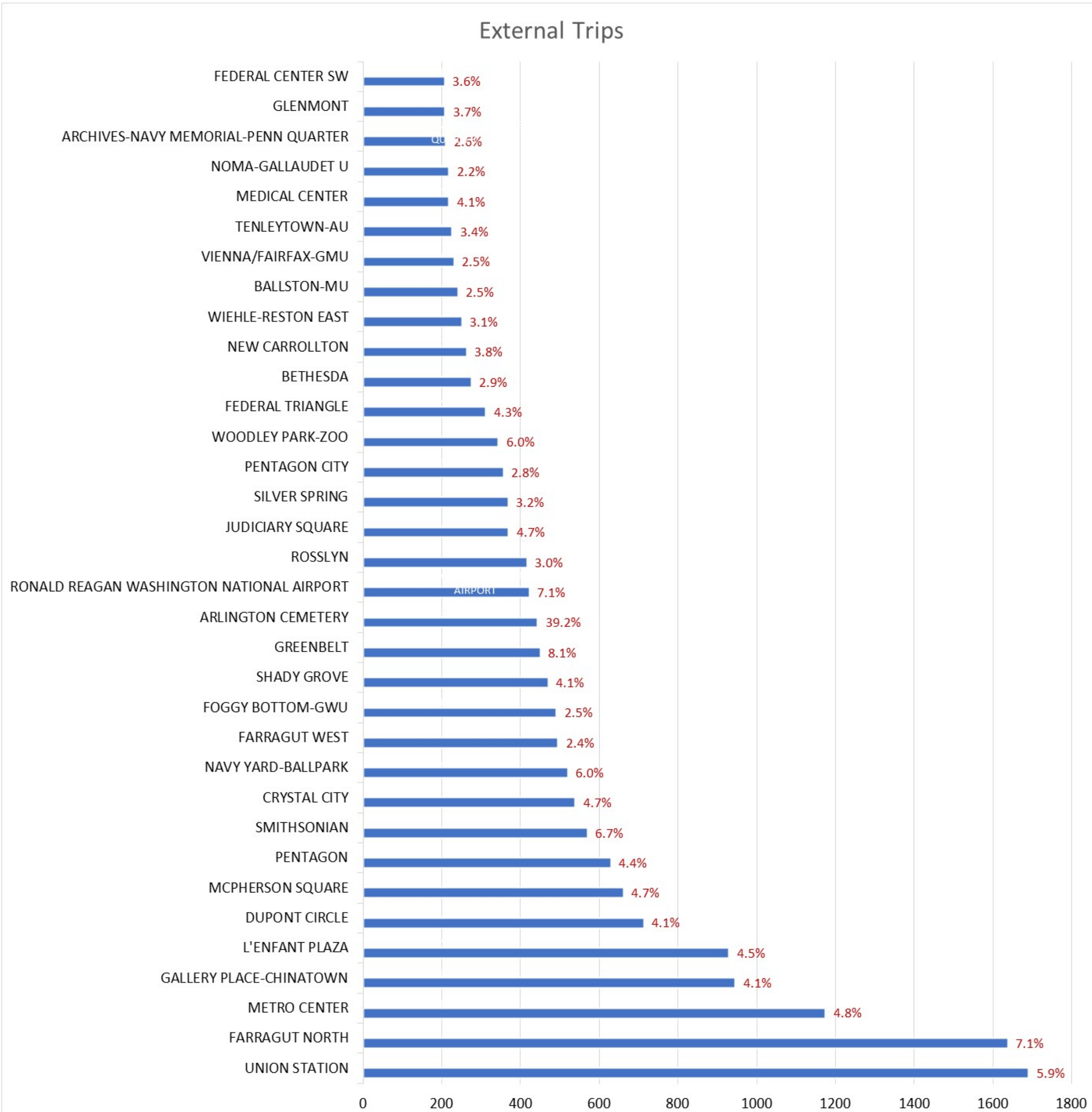


FIGURE 3-13: METRORAIL STATIONS WITH LOW NUMBER OF EXTERNAL TRIPS

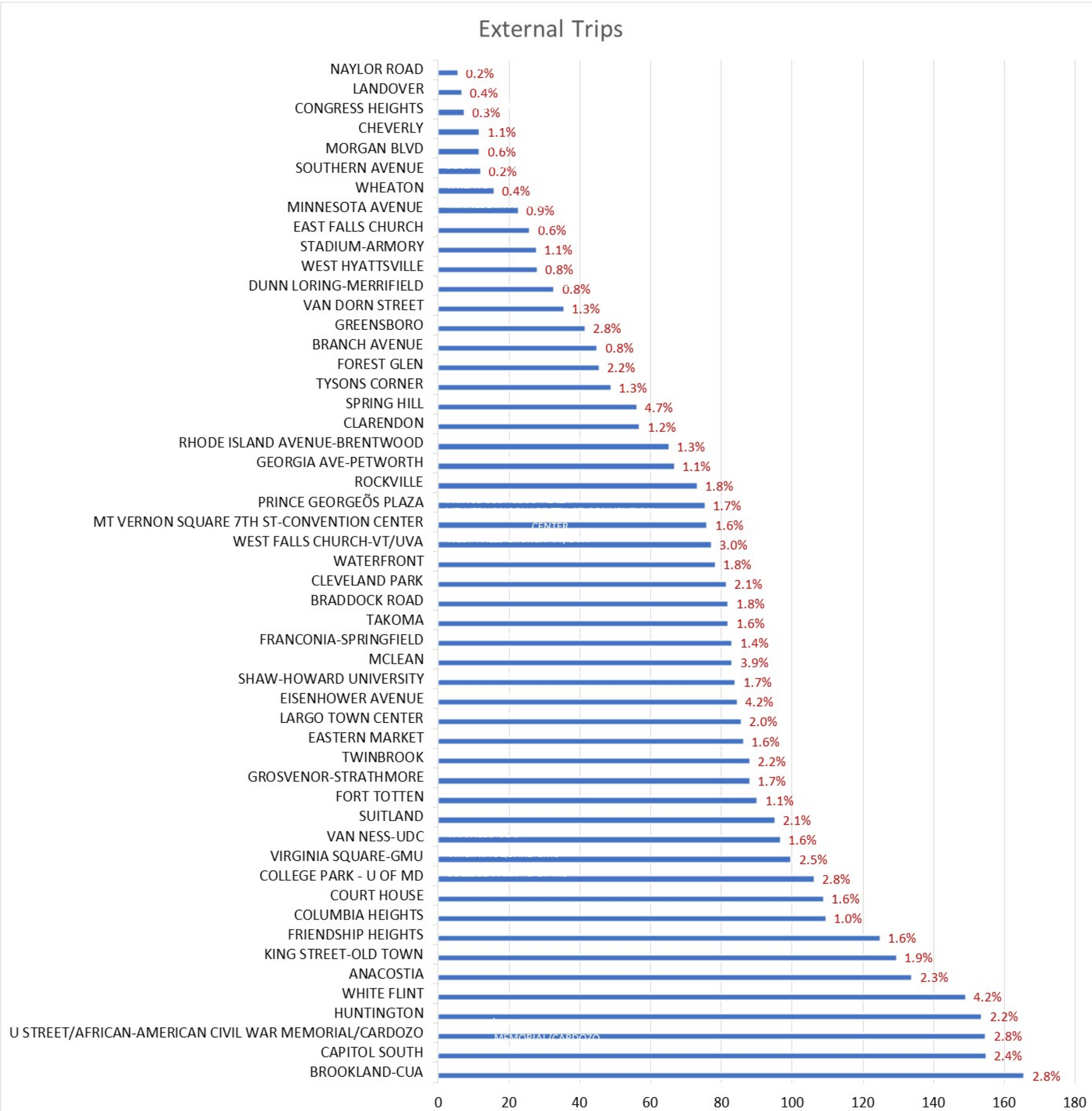


Figure 3-14 shows the MARC stations with very low to very high number of external trips, along with the Union Station for the Penn Line that has the highest external travel market with 8,639 EI/IE/EE trips. The Pennsylvania station along the Penn line also has high external travel market with 3,947 EI/IE/EE trips.

FIGURE 3-14: MARC STATIONS WITH VERY LOW TO VERY HIGH NUMBER OF EXTERNAL TRIPS

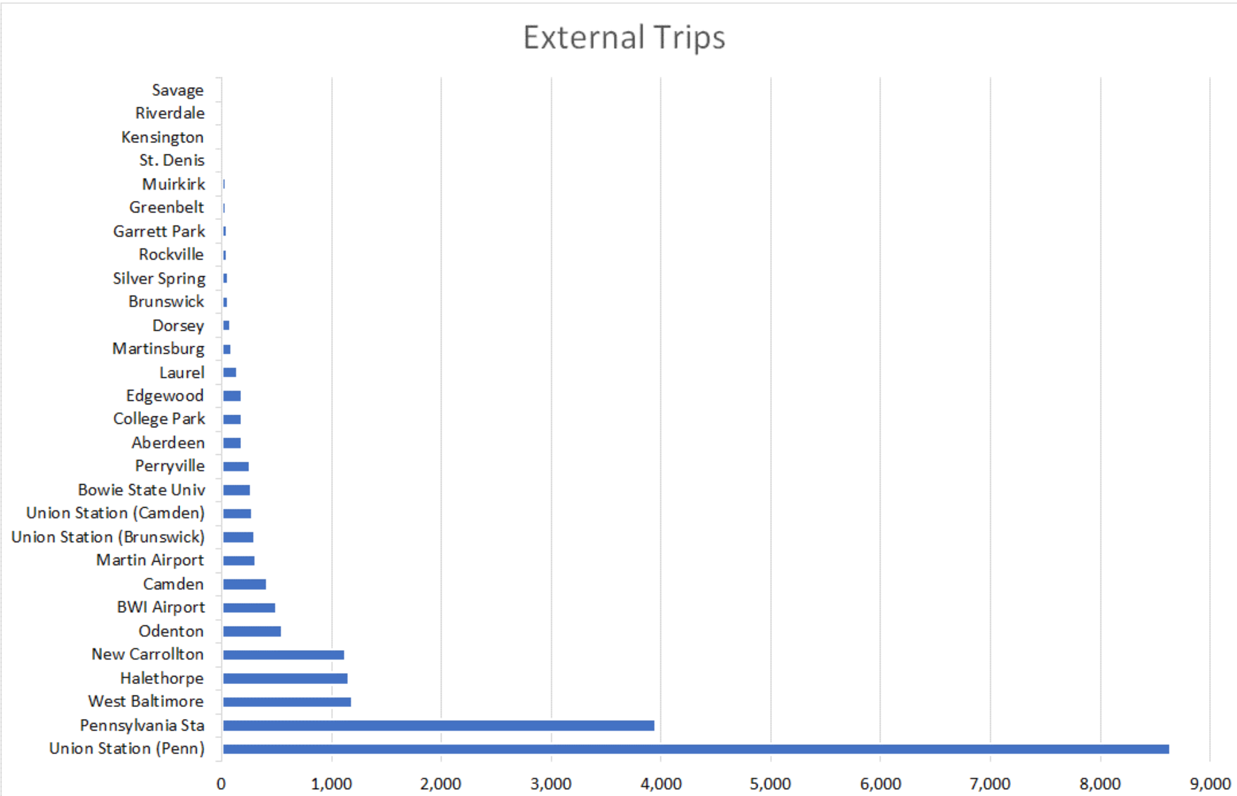


Figure 3-15 shows the VRE stations with very low to low number of external trips, along with the Spotsylvania station along the Fredericksburg Line that has the highest external travel market with 199 EI/IE/EE trips.

FIGURE 3-15: VRE STATIONS WITH VERY LOW TO LOW NUMBER OF EXTERNAL TRIPS

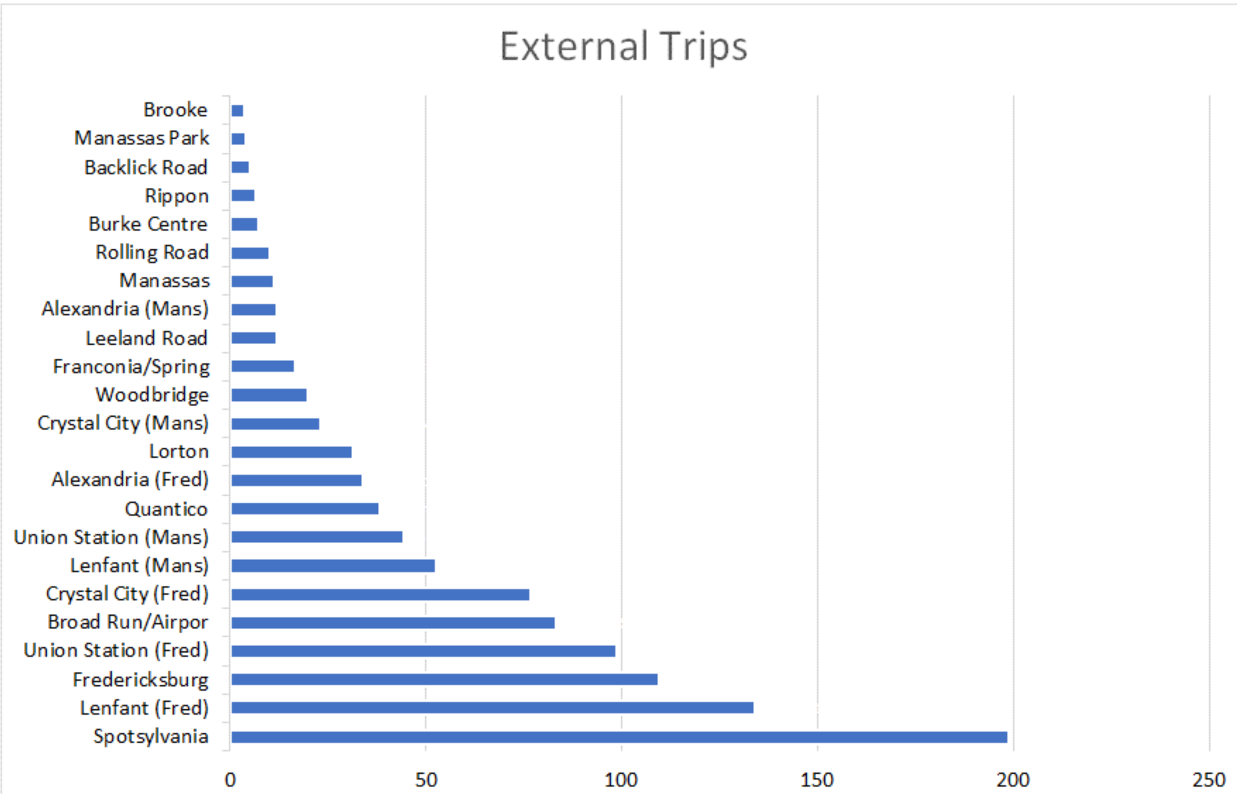


Figure 3-16 shows the Metrorail stations with moderate to high number of visitor trips, along with the Union station that has the highest visitor travel market of 2,616 visitor II trips. The Regan National airport station also experiences a high number of visitors. Similarly, Figure 3-17 shows the Metrorail stations with low number of visitor trips.

FIGURE 3-16: METRORAIL STATIONS WITH MODERATE TO HIGH NUMBER OF VISITOR TRIPS

FIGURE 3-17: METRORAIL STATIONS WITH LOW NUMBER OF VISITOR TRIPS

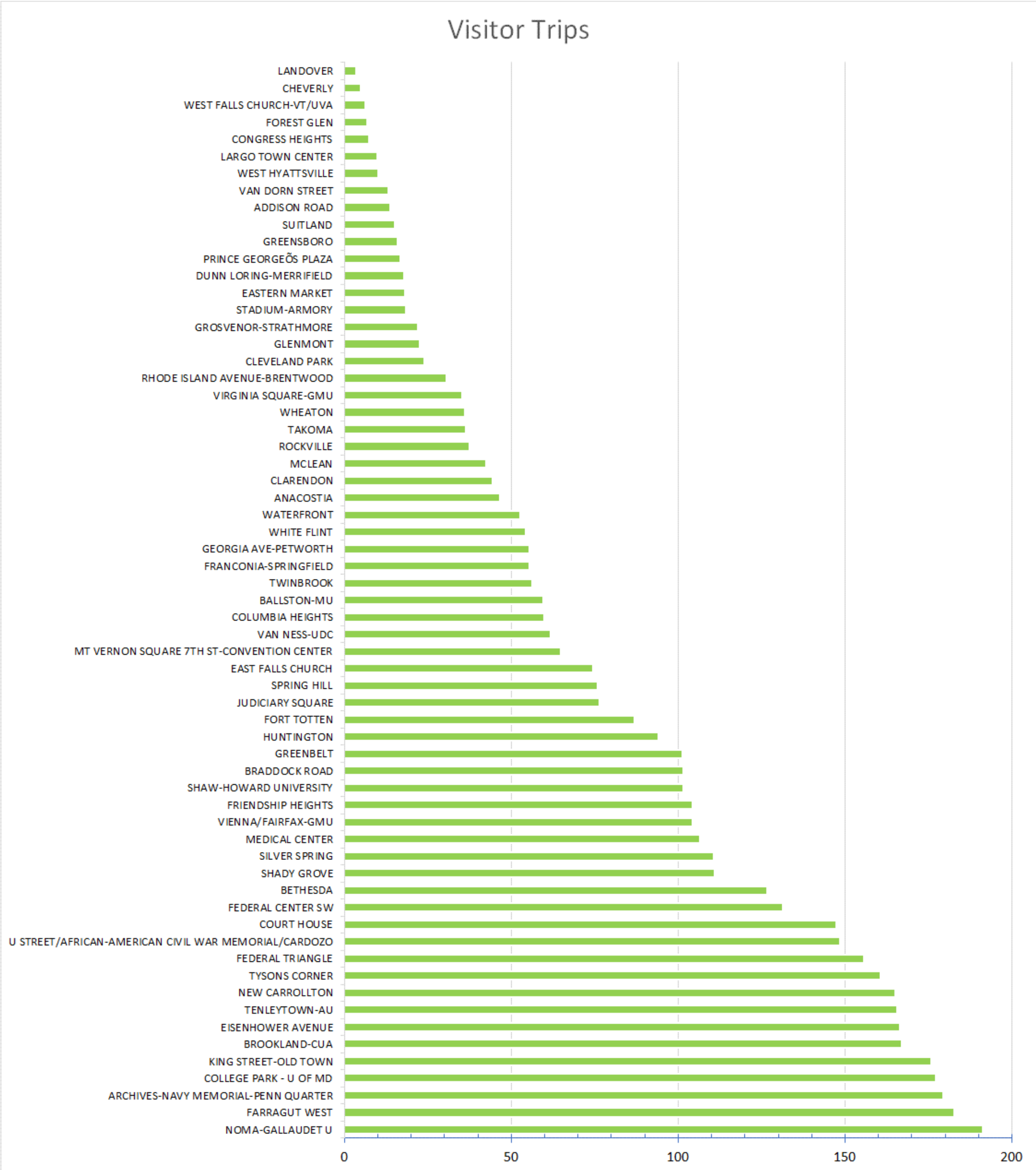


Figure 3-18 shows the MARC stations where some visitor trips were noted. The Odenton station and the BWI station, both along the Penn line, experience over 130 visitor trips in an average weekday.

FIGURE 3-18: MARC STATIONS WITH VISITOR TRIPS

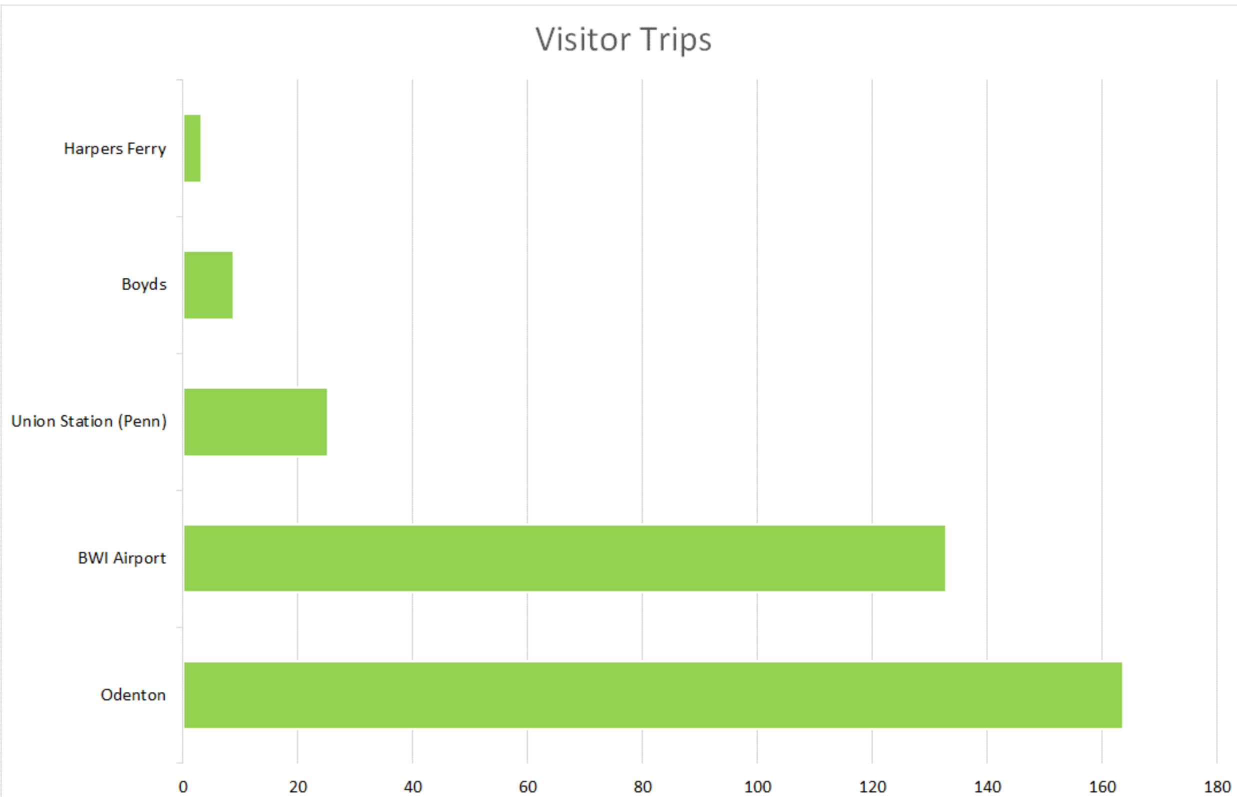


Figure 3-19 shows the VRE stations where very low number visitor trips were noted. The stations along the Fredericksburg and Manassas lines observed less than five visitor trips in an average weekday.

FIGURE 3-19: VRE STATIONS WITH VISITOR TRIPS

RSG

the science of insight

55 Railroad Row
White River Junction, VT 05001
802.295.4999
www.rsginc.com

White River Junction & Burlington, VT | Arlington, VA | Chicago, IL | Portland, OR | San Diego, CA

RSG promotes sustainable business practices that minimize negative impacts on the environment. We print all proposals and reports on recycled paper that utilizes a minimum of 30% postconsumer waste. RSG also encourages recycling of printed materials (including this document) whenever practicable. **For more information on RSG’s sustainability practices, please visit www.rsginc.com.**