NEW YORK CITY DOT

CITYWIDE MOBILITY SURVEY

Survey User Guide

CONTENTS

1.0 Overview

New York City Department of Transportation (NYC DOT) conducts the Citywide Mobility Survey (CMS), a household travel survey designed to measure New York City residents' travel behavior, preferences, and attitudes. The survey was first launched in 2017, with subsequent waves conducted in 2018, 2019, 2020, 2022, and 2024. This continuity enables tracking of travel patterns over time and analysis of how changing conditions, policies, and urban dynamics influence mobility across the city. The objectives of the survey are to better understand:

- New Yorkers' basic transportation conditions
- Changes in travel behavior over time
- The equity of access to transportation across the city
- New Yorkers' experience of and opinions of transportation

The survey also provides valuable data for building models of transportation behavior.

The CMS collects two primary types of data: 1) demographic information and 2) travel diaries that document all trips made by a respondent during a set period of time.

This guide is based on the 2024 CMS methodology; however, the same approach applies to both the 2022 and 2019 CMS surveys, which use updated methods and weights consistent with this framework. Note that because of the timing of the American Community Survey (ACS) data release, the 2024 CMS targeted the 2023 ACS, while the 2022 and 2019 surveys targeted the ACS data corresponding to their respective reference years, as they were reweighted later.

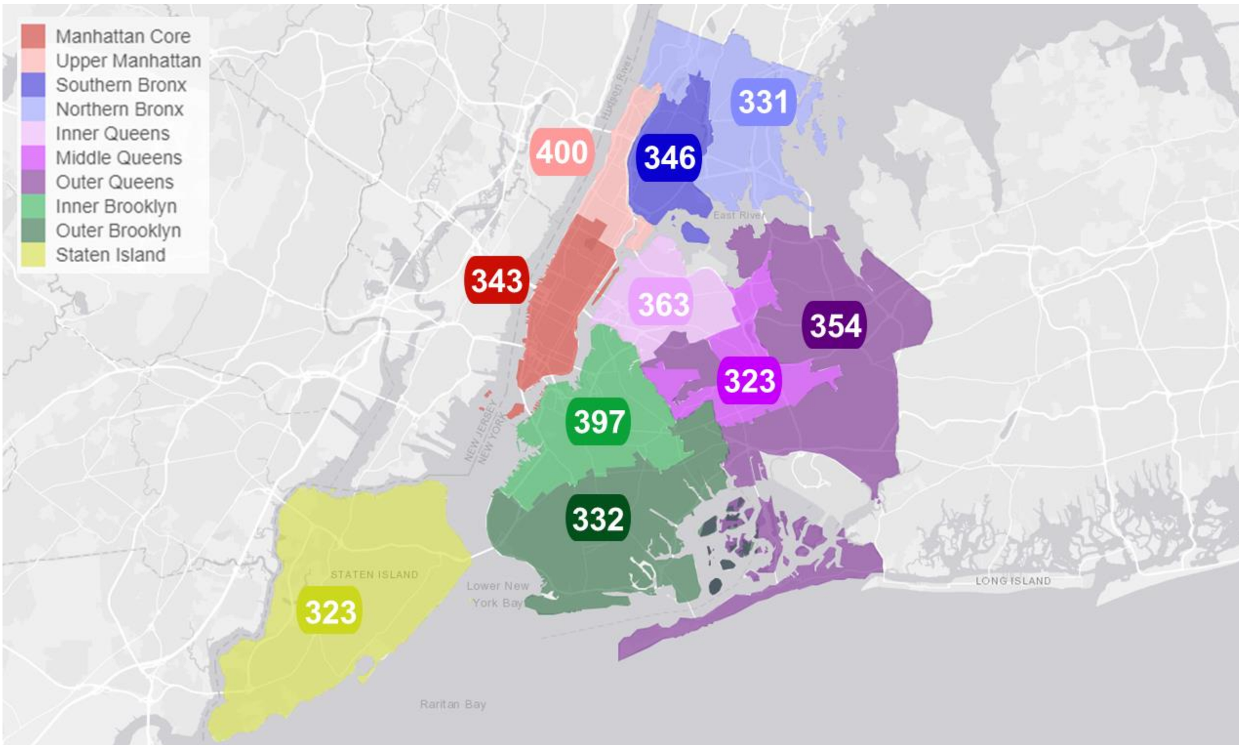
2.0 Survey Design & Methodology

As in the 2019 and 2022 survey waves, the 2024 CMS primarily used address-based sampling (ABS) to recruit participants, which accounted for approximately 69% of the total sample. To address a lower-than-expected response rate, NYC DOT supplemented recruitment by re-inviting some participants from the 2019 and 2022 CMS cohorts. To support statistically robust analysis at the sub-borough level, the sampling design targeted 300 completed responses within each of the ten geographic CMS zones defined by NYC DOT. These survey zones are aggregations of neighborhood tabulation areas (NTAs) based on 2010 Census boundaries. A shapefile of the CMS zones is publicly available through NYC Open Data.

Among respondents, about 81% completed the survey using the smartphone application, 18% participated online, and 2% responded via telephone interviews conducted through a call center. Smartphone participants completed a real-time seven-day travel diary, while online and call center participants completed a one-day travel diary. Despite these differences in diary length, all participants responded to the same questionnaire, enabling the integration of responses into a single weighted dataset.

The 2022 CMS effort collected complete travel behavior data from 3,512 resident adults from May 6, 2024, through June 22, 2024. The survey was conducted by RSG, which had previously conducted the 2019, 2020, and 2022 surveys. Figure 1 shows the actual number of respondents in each CMS Zone.

Figure 1: Citywide Mobility Survey region and complete records by zone



3.0 Weighting Methodology

This section summarizes how survey weights were developed to make the dataset representative of New York City residents across key geographic and sociodemographic dimensions that influence travel behavior. The weighting process is designed to correct for potential biases related to data collection methods, including differences in response patterns between participants completing the survey by phone and those using the smartphone app.

Two separate sets of weights have been developed for this survey. The first set is calculated at the city level and is designed to ensure that the results align with the overall citywide targets. The second set is developed at the zonal level, where the targets correspond specifically to the CMS zones. It is important to use the appropriate set of weights depending on the level of analysis being conducted, since applying city-level weights to zonal results (or vice versa) could lead to incorrect conclusions.

The weighting procedure produces four distinct weight variables, each intended for a specific unit of analysis::

- **A household-level weight**, whose sum represents the total number of households in the survey region.

- **A person-level weight**, whose sum represents the total number of adult persons in the survey region.
- **A day-level weight**, which distributes each person's weight across the number of complete diary days and is used for household-day and person-day analyses to approximate an average travel day per person. The sum of these weights also reflects the total persons in the survey region (and matches the sum of the person-level weights).
- **A trip-level weight**, whose sum represents the total number of trips taken on a typical day in the survey region (Monday–Sunday).

The weighting has two primary objectives: first, to expand the survey so that it represents the full population rather than only the sampled respondents, and second, to reduce response bias by adjusting record weights so they align as closely as possible with multiple demographic control totals derived from ACS 1-year PUMS data. The survey weighting process includes three primary steps that are covered in more detail throughout this section:

1. Initial expansion
2. Weighting for non-response bias
3. Calculate day and trip level weights

For analyses intended to infer travel behavior across the city rather than just describe survey respondents, weighted data should be used. Applying the appropriate weight makes each record represent a specific number of similar households, persons, days, or trips in the region; for example, a household with a weight of 250 represents 250 comparable households. Analysts typically work with weights by summing them for the group of interest instead of counting raw records. Note that the resulting weighted dataset represents resident travel within the sampling frame over the survey period and does not include commercial vehicle travel, travel by people living in group quarters outside the address-based frame (such as dormitories or institutional housing), trips made by non-residents, or seasonal/holiday travel occurring outside the survey period.

The remainder of this section expands on the concepts above. The outputs from the 2024 NYC CMS Travel Survey weighting process are in Appendix A.

3.1 Initial Expansion

The first step of the weighting process is to calculate an initial weight for each respondent based on their probability of being sampled. This step reverses the initial sampling plan for the survey, which segmented the survey region into several distinct groups or sampling segments.

The 2024 NYC CMS study primarily sampled persons for data collection using traditional address-based sampling (ABS) whereby a single person from each household is invited by mailed invitation. However, in this current survey wave a subset of respondents were reinvited from previous waves and must be treated as a supplemental non-probability sampled (NPS) segment.

Unlike most RSG weighting procedures, the process developed for this project differs in two ways. The study unit is person-based and includes travel for all days in a week. RSG considered a response as complete and eligible for weighting if one, and only one person, from each household provided complete data on at least one day.

The initial expansion weight is the inverse of the probability of inclusion in the study where the probability of inclusion is the product of the probability of selection and the probability of a response.

$$P_{selection} = \frac{m}{N}$$

$$P_{response} = \frac{r}{m}$$

$$P_{inclusion} = P_{selection} \times P_{response} = \frac{m}{N} \times \frac{r}{m} = \frac{r}{N}$$

Where *m* is the number of mailed invites, *r* is the number of responses, and *N* is the total population. Thus, the initial weight calculation for a given segment (IW_s) can be calculated as the number of actual persons in a segment (N_s) divided by the number of responses in the same segment (r_s), or

$$IW_s = \frac{1}{P_{inclusion}} = \frac{N_s}{r_s}$$

For example, if segment A contains 100 persons and the survey collected responses from 5 persons (r_s), each person in that segment would have an initial weight of 20 (100 / 5 = 20). The expected number of actual people is 5-year ACS data at the block group level, excluding children under 18 and group quarter populations.

When dealing with a non-probability sample, the above calculation is done separately for the ABS and non-probability samples (convenience-based sample or CBS), resulting in two sets of initial weights. The initial sample weights must then be adjusted using a blending factor, *α,* for sampling strata *k,* and combined with the ABS records.

$$\mathrm{IW_s} = \alpha \times \mathrm{IW}_{\mathrm{cbs},s} + (1 - \alpha) \times \mathrm{IW}_{\mathrm{abs},s}$$

Where α_s is the blending factor between 0 and 1. Determining an appropriate blending factor can be iterative depending on the variance and imbalance inherent in the initial weights. A simple blending factor may be calculated from the sample size proportionality:

$$\alpha = \frac{r_{cbs}}{r_{abs} + r_{cbs}}$$

An optimal blending factor is not strictly necessary since initial weights are meant to be an approximate starting point to be further adjusted in the weight optimizing algorithm. The main objective is to try and bring the two sets of initial weights in-balance with each other to facilitate the weighting algorithm.

For population targets, RSG uses the 2023 ACS 1-year Public Use Microdata Sample (ACS PUMS) to determine the target proportion of each weighting category. The 1-year ACS provides the most recent and up to date demographic data available from ACS. However, block group level data are required to calculate the initial expansion weights and block group level data are only available from the 5-year ACS, which can vary slightly from the 1-year ACS. The 5-year ACS-based initial weights are then adjusted to match the 1-year ACS totals. This adjustment is done by creating a crosswalk from block groups to PUMAs, aggregating both the 1-year and 5-year ACS data into PUMAs, and proportionally adjusting the initial weights to match the 1-year ACS for each PUMA. Table A1 in Appendix A shows the initial weights by sample segment.

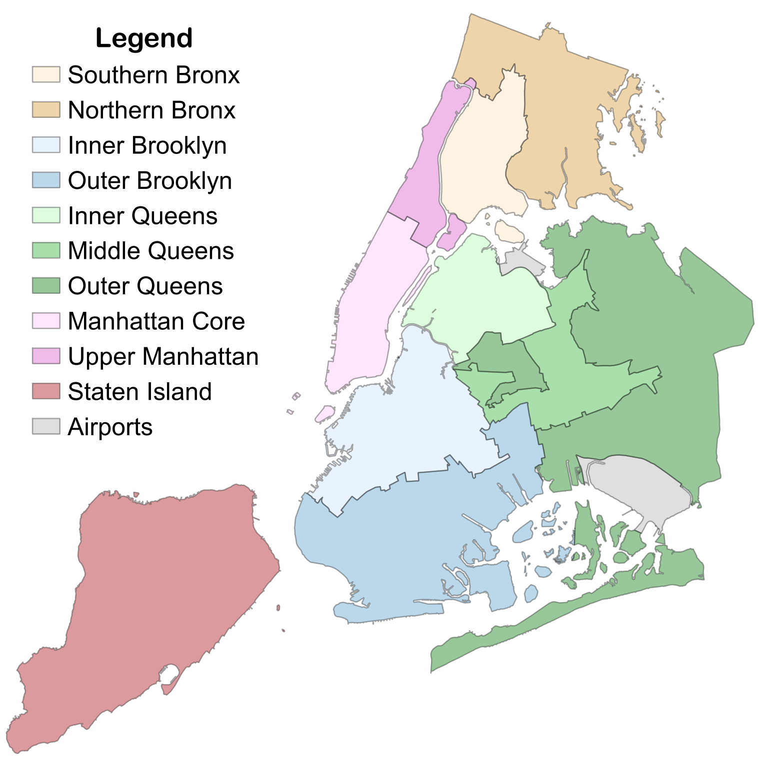
3.2 Weighing for non-response bias

After creating the initial weights to account for each respondent's likelihood of being sampled, RSG conducted a reweighting process to adjust for non-response bias across geographies and demographics. Non-response bias refers to biases in the unweighted data that occur because different types of people respond to surveys at different rates.

Geographic Target Areas

Weighting targets are based on Public Use Microdata (PUMS), which are ACS data in disaggregated form. This enables creation of more flexible aggregated targets as opposed to using the traditional ACS tables. This flexibility is critical to supporting cross-tabulated targets, such as gender by ethnicity, rather than simply gender and ethnicity independently. The PUMS data are spatially allocated to areas called Public Use Microdata Areas (PUMAS), which have populations in the range of 100,000 to 200,000 and tend to be smaller than counties but larger than block groups. Although PUMAs are larger than block groups, survey data can still be sparse when split by PUMA. This can result in insufficient representation (i.e., missing person/household types) in the survey sample data to be weighted to desired target totals, yielding poorly weighted data. For this reason, RSG typically aggregated PUMAs until a sufficient sample and broader target totals yielded a satisfactory result. For NYC CMS, pre-determined target areas were provided based on New York CMS zones, as shown in Figure 2.

Figure 2: Geographic Target Areas (CMS zones)



Using non-PUMA geographic target areas requires the aggregated PUMS data in PUMAs to be "converted" to the CMS zones using a crosswalk. To achieve this, the PUMAs are proportionally allocated by population, based on the aggregated block group population in the overlapping areas. For the example crosswalk shown in Table 1, 92% of PUMA 3604107 will be allocated to Manhattan Core and 8% is allocated to Upper Manhattan.

Table 1: Example PUMA to CMS zone Crosswalk

PUMA	PUMA population	CMS zone	Overlapping 5-year ACS population	Proportion	CMS zone population
3604103	118,545	Manhattan Core	146,049	100.00%	118,545
3604104	99,863	Manhattan Core	113,403	100.00%	99,863
3604107	186,104	Manhattan Core	196,720	91.78%	170,808
3604107	186,104	Upper Manhattan	17,617	8.22%	15,296
3604412	192,734	Middle Queens	58,433	23.10%	44,525
3604412	192,734	Outer Queens	194,502	76.90%	148,209
3604413	155,416	Outer Queens	202,547	100.00%	155,416
3604414	91,013	Outer Queens	128,263	100.00%	91,013

Demographic Targets

Different household and personal attributes affect survey response, which can introduce bias in unweighted survey data. For example, lower income or older residents may have lower response rates to the surveys. To correct these types of biases, RSG selected a variety of household- and person-level target categories. Table 2 shows the weighting categories for each target. As part of this process, RSG also imputed missing data where applicable (described later in the memo). Some categories were aggregated for specific CMS zones where sample size was insufficient. Conversely, some target categories were combined to provide cross-tabulated targets to account for interaction effects.

Table 2: Demographic Targets

Variable	Categories
Household Size	1 person 2 person 3 person 4 person 5 person or more
Income ***(Imputed if non-response)***	Under $25,000 $25,000–$49,999 $$74,999–$99,999 $100,000–$199,999 $200,000 or more
Adults	1 adult 2 adults 3 adults or more
Presence of Children	0 children 1 or more children
Gender ***(Imputed if non-response)***	Male Female
Age	Under 5 5–17 years 18–34 years 35–64 years 65 years or older
Employment	Employed full-time Employed part-time Not employed
University Student Status	University student Not a university student
Educational Attainment	Some college education No college education
Race	Asian or Pacific Islander Black White Other
Ethnicity ***(Aggregated to Hispanic/Not Hispanic for Middle Queens CMS Zone)***	Mexican Puerto Rican Dominican Other Hispanic Not Hispanic
Gender by Ethnicity	Female and Hispanic Female and Not Hispanic Male and Hispanic Male and Not Hispanic
Gender by Race	Female and Asian or Pacific Islander Female and White Female and African American or other Male and Asian or Pacific Islander Male and White Male and African American or other
Total Persons	*Not applicable*
Total Households	*Not applicable*

Data Imputation

The income, gender, race, and ethnicity questions in the survey allowed participants to respond with "prefer not to answer." To facilitate data weighting, RSG imputed missing values for these variables when a participant selected "prefer not to answer." Imputation was performed exclusively to support the weighting procedure and was not incorporated into the analytical results.

Income Imputation for Declined to Respond

RSG imputed income using a model-based approach where missing income was predicted based on a set of independent variables including:

- Income distribution of the block group based on ACS 2022 5-year data.
- Number of non-working adults in the household.
- Number of children in the household.
- Employment status of household members.
- Educational attainment of household members.
- Whether the household is a single-family home.

RSG has tested this model across many travel survey projects. The model adequately matches the income values that were reported, which indicates it is reliable to predict the missing income values. Table A3 in Appendix A includes the model specification and coefficients.

Gender Imputation

RSG probabilistically assigned missing gender responses using a Monte Carlo procedure based on the sample data's gender distribution within the respondent's age category.

Race and Ethnicity Imputation

RSG probabilistically assigned missing race and ethnicity responses using a Monte Carlo procedure based on the ACS data's race and ethnicity distribution within the respondent's reported home block group.

Weighting Process

Using the geographies and targets outlined, RSG adjusted the initial weights using an entropy-maximization (EM)[1] algorithm using PopulationSim. This approach is beneficial because it reduces the variance in the final weights, which in turn reduces the margins of error when using the weighted data.

During the process, RSG constrained the ratio of the initial weight to final weight to be in a range of 0.01–6.0 for each person with a maximum weight of 15,000. RSG determined this range after testing alternative limits and judging the best tradeoff between accuracy and variability. Although removing these constraints would enable the process to match the targets more closely, keeping the constraints reduces variance in the final weights. Furthermore, the PUMS targets are based on census survey data, so it is best practice to not try to match the targets too precisely by allowing the weights to vary widely. Table 3 provides the distribution of the final weights RSG calculated for each PUMA group. Table 4 summarizes the ratio of citywide final weight to the initial weight.

---

[1] For more information, see Multi-level Population Synthesis Using Entropy Maximization-Based Simultaneous List Balancing by Paul et al. (2018).

Table 3: Summary of Final Weights

PUMA Group	Sample Size	Min	Mean	Median	Max
Inner Brooklyn	397	26.95	2,354.53	864.81	14,995.70
Inner Queens	363	13.49	1,463.29	18.37	10,521.89
Manhattan Core	343	148.96	2,525.87	2,078.18	12,338.72
Middle Queens	323	10.48	1,162.34	18.59	9,293.60
Northern Bronx	331	14.42	1,539.01	768.13	11,011.64
Outer Brooklyn	332	20.78	3,088.57	628.17	14,989.93
Outer Queens	354	29.29	2,491.52	968.98	14,982.60
Southern Bronx	346	12.32	1,370.67	260.01	10,547.94
Staten Island	323	86.34	1,171.55	758.97	6,358.53
Upper Manhattan	400	9.6	1,108.67	293.08	9,460.09
CMS Overall	**3,512**	**9.6**	**1,828.21**	**603.36**	**14,995.70**
Citywide	**3,512**	**43.54**	**1,829.02**	**939.18**	**9,999.35**

Table 4: Ratio of Final to Initial Weights

Geography	Sample Size	Min	Mean	Median[2]	Max
Inner Brooklyn	397	0.01	1.00	0.40	6.00
Inner Queens	363	0.01	0.96	0.01	5.75
Manhattan Core	343	0.06	1.07	0.87	4.78
Middle Queens	323	0.01	0.87	0.01	5.08
Northern Bronx	331	0.01	0.94	0.46	5.68
Outer Brooklyn	332	0.01	1.19	0.22	6.00
Outer Queens	354	0.01	1.06	0.42	6.00
Southern Bronx	346	0.01	1.00	0.20	5.77
Staten Island	323	0.08	0.93	0.59	5.30
Upper Manhattan	400	0.01	0.90	0.22	5.17
CMS Overall	3,512	0.01	0.99	0.35	6
Citywide	3,512	0.03	1.03	0.51	5.00

To demonstrate how this step impacts the weights, consider the example in Table: 5. If there were only a single weighting geography with only a single target (e.g., household size), this step would simply adjust the initial weights such that the distribution of the adjusted weights matches the distribution of the targets.

In this example, assume each of 5 persons has an initial weight of 20.0 and a distribution of household sizes as follows:

[2] In many cases, the median ratio is extremely close to the minimum ratio because there is a significant skew toward a low ratio. This is a necessary consequence of minimizing the variance across the weights.

- 40% of households have 1 member
- 40% of households have 2 members
- 20% of households have 3 members

The distribution of household sizes in the target population is as follows:

- 20% of households have 1 member
- 40% of households have 2 members
- 40% of households have 3 members

The initial weight for households with a size of 1 would be scaled down to match the targets, and initial weights for households with a size of 3 would be scaled up to match the targets.

The maximum entropy method achieves a similar outcome but addresses multiple targets at once.

Table: 5 Example weights after reweighting

	Household Size	Initial Weight	Weight After Reweighting
Household 1	1	20.0	10.0
Household 2	2	20.0	20.0
Household 3	1	20.0	10.0
Household 4	2	20.0	20.0
Household 5	3	20.0	40.0

3.3 Adjusting for trip non-response

Trip-Rate Adjustments

In reviewing all rMove™ (the smartphone app) survey data, RSG has found that the trip rates from the smartphone-based survey data are frequently 15–20% higher than those from online survey data. There are three main reasons for this:

- Smartphone-owning households have different socio-demographic characteristics than non-smartphone households and tend to make more trips.
- There are about twice as many "stay at home" days with no reported trips in the online and call center-based data in comparison to the smartphone-based data.
- Even on days with one or more reported trips, there are more trips per day reported on average in the smartphone-based data than in the online and call center-based data.

In this study, the trip rates for the online and call-center surveys were adjusted to better match those of the smartphone-based surveys. While the adjustment factor would ideally treat call-center surveys independently of online browser surveys, there was insufficient sample size in this case to support an adjustment mode.

Table 6: Respondent Frequency by diary platform

Diary Platform	Browser	Call Center	rMove
Home-based other	137	14	18,829
Home-based mandatory	9	0	2,182
Non-home-based other	2,472	310	77,127
Non-home-based mandatory	243	10	8,128

The starting point for the trip-rate bias correction was the person-day weights. The following steps were then taken to adjust trip rates:

1. Trips were segmented into the following four trip types that have different levels of underreporting. Then for each person-day in the sample, the number of trips were counted by type.
    a. Home-based work/school trips
    b. Home-based other trips
    c. Non-home-based work/school trips
    d. Non-home-based other trips
2. For each trip type, a Poisson regression model was estimated where the dependent variable was the number of trips of that type for the person-day. The independent variables were the set of household and person variables, including age, income, employment, student status, education, telework frequency, and dummy variables for online and call center-based person-days.

For each person-day and for each trip type, the estimated regression model was applied with and without the bias coefficients. The ratio of the two estimates resulted in a factor to apply to the trip weight for that person-day. For example, if the model predicted 1.10 trips with the estimated model and 1.32 trips with the bias parameters set to 0 for an online or call center-based person-day, then a factor of 1.32/1.10 = 1.2 was used to multiply the person-day weight to get an adjusted trip weight. For smartphone respondents, the bias coefficients do not apply, so the factor was always 1.0 and the trip weight equaled the person-day weight. A lower bound of 1.0 and an upper bound of 2.0 was placed on ratios to avoid extreme adjustment to the weights. The specifications for each of the four regression models are shown in Table B4 in Appendix B. The resulting trip adjustment factors by diary method and trip type are shown in Table 7.

Table 7: Trip Adjustment Factors

Trip type	rMove	Call Center / Online
Home-based mandatory	1	2.00
Home-based other	1	2.00
Non-home-based mandatory	1	1.00
Non-home-based other	1	1.00

4.0 Data User Guide

Transportation planners rely on detailed travel data to inform their planning. The 2024 CMS dataset includes highly detailed information about how, where, when, and why New Yorkers make trips. The CMS allows NYC DOT to obtain descriptive statistics about citywide travel and to analyze trends over time.

Data users can derive many key figures from the CMS dataset, including person-trip rates, travel mode shares, vehicle occupancy factors, geographic travel patterns, and more – with rich demographic detail. Beyond traditional travel survey uses, the CMS program is also somewhat unique in its frequency, with five complete travel surveys since 2017. This allows for trend analysis over time, which is often not feasible when collection occurs less often.

4.1 Dataset Overview

The data can be seen as composed of two parts:

- Part one, also called the “recruit survey,” collected information about household composition, demographics, and typical travel behavior.
- Part two, also called the “travel diary,” required participants to record their travel during an assigned travel period.

“Complete” households met the following conditions:

- The participant completed the recruit survey in full.
- The participant completed a travel diary for at least one weekday. Only one person in the household (identified as person 1) completed a travel diary.

The household member who completed the recruit survey is referred to as “person 1.”

The 2024 CMS dataset includes six data tables, often referred to as their own “level” of data. These tables include all user-input survey variables, passively collected GPS and location data, survey metadata, and derived variables to support data analysis. The tables included in the dataset include:

- **Household**: contains data about the characteristics of the participant’s household, including household size, income, type of residence, number of vehicles, bicycles, and micromobility devices, package delivery location, and other demographic and transportation-related information.
- **Person**: contains characteristics of individual members of the participant’s household, including age, race, gender, disability status, education, employment, remote work, typical commute mode, biking frequency, shared services, and other demographic and transportation-related information.
- **Vehicle**: contains characteristics of each vehicle in the participant’s household.
- **Day**: contains information on each day of the participant’s assigned travel period, including deliveries, remote work duration, online shopping, reasons for no travel if applicable, and start/end location of the day, with between one and seven travel days recorded for each participant.

- **Trip**: contains characteristics of each trip made by the participants, including origins and destinations, mode, purpose, transportation related fees such as transit, parking, or taxi pay, parking location for the bikes or micromobility devices, number of travelers, trip duration, and trip distance. All location information in the Open Data, such as home and work location, is only provided at the CMS zone level. The Open Data also does not include any questions to which users are able to submit responses
- **Location**: Two or more records for each trip collected via smartphone (if any). To protect the privacy of survey participants, the Location table is not posted on Open Data.

4.2 Data Preparation

This section summarizes the methods used to prepare the data. Given that all data were collected in a "controlled" environment (e.g., survey answers are validated in real-time), data preparation was primarily focused on coding variables and deriving new fields to facilitate analysis.

Initial Data Review

Before reviewing the data for completion, RSG removed households from the dataset that met the following exclusion criteria:

- Household reported a home location outside the survey region. Most households dropped during initial review were excluded for this reason.
- Household reported contact information that matches other households (indicating duplicates). In these cases, RSG kept the first "household" to report their travel diary and removed the subsequent records.

Additional Trip Review

In addition to the standard QA/QC checks, the trip trace data were reviewed manually to check for the following potential issues:

- Walk trips occurring over water.
- Ferry trips occurring over land.
- Transit trips that could not feasibly be using the system reported (bus, subway, LIRR, etc.).

Trips that were flagged during this review had their mode changed to the appropriate response where possible, either by comparing with shapefiles of the transportation system or the context of the respondent's other trips in their travel period. Where no appropriate mode could be identified, the mode was assigned "Other". The mode_1, mode_2, etc. set of variables capture the mode information with these changes made, and the original answers are preserved in the "_app" set of variables. Trips that had their mode(s) changed during this review can be seen by filtering the trip table to flag_mode=1.

Geographic Variables

Census 2010 and 2020 block group shapefiles, as well as 2012 and 2022 PUMA (Public Use Microdata Area) shapefiles, were downloaded via the R Tigris package and spatially joined to the reported home, work, school, origin, and destination coordinates. The final reported block group may not always match the block group ascribed to the household's sample address (which is used to determine the sample segment) for a few reasons: Sample addresses are geocoded differently than survey addresses, sample addresses sometimes are coded to a mailbox location rather than a home location, and home addresses in the survey are not always geocoded to a person's exact home (e.g., a cross street nearby). Because a person's reported home address is considered to be more recent and typically more accurate than the sample address, the geographic variables are derived using this address. All location information in the Open Data, such as home and work location, is only provided at the CMS zone level.

Households retain their initial sample segment assignment, as this is what determines their probability of being invited according to the information in the sample address file.

4.3 Data Coding and Labeling

Time and location standards

The timestamp information is broken out into date, hour, minute, and second in the trip table. This is done to preserve the desired timezone (Eastern Time) across all trips, regardless of where the trip took place geographically (e.g., if a trip took place in another time zone, the time information for that trip are still in Eastern Time).

To convert the data from its broken out format into a timestamp, the following code can be run using the dplyr, stringr, and lubridate packages in R:

```
trip = trip %>%
  mutate(depart_hour    = str_pad(depart_hour,    width=2, side='left', pad='0'),  # stringr function
         depart_minute  = str_pad(depart_minute,  width=2, side='left', pad='0'),
         depart_seconds = str_pad(depart_seconds, width=2, side='left', pad='0'))

trip = trip %>%
  mutate(depart_time = paste0(depart_date, ' ', depart_hour, ':', depart_minute, ':', depart_seconds),
         depart_timestamp = as_datetime(depart_time, tz = 'US/Eastern'))          # lubridate function
```

Missing values and gaps in the data

A survey data table cell may be missing data for one of four reasons:

1. Value or response is missing due to survey logic, participant non-response, or error.

Example: Participants who traveled by bus were not asked if they were the driver or passenger on the trip.

Coded as: 995 for categorical variables, blank/NA for continuous variables

2. A respondent indicated that the question was not applicable and skipped that question.

Example: Some participants did not share how they pay to park at work because they do not park at work (e.g., carpool).

Coded as: 996 (often labeled as "Not applicable")

3. A respondent indicated that they didn't know the answer and skipped that question.

Example: Some participants who made a vehicle trip and paid to park the vehicle may not remember the amount they paid.

Coded as: 998 (Don't know)

4. A respondent indicated that they preferred not to answer a question and skipped that question.

Example: Some participants chose not to provide their household income.

Coded as: 999 (Prefer not to answer)

Other notes about missing survey data

- For a survey to be complete, all survey questions asked of the participant must have been answered in the app or online instrument.
- Continuous variables (e.g., trip distance, trip duration) are not coded with missing value codes and are instead left empty when missing to avoid interfering with statistical calculations.

Outliers

Continuous variables (e.g., trip distance, trip duration, parking cost) in the dataset may contain outliers. Data users should be aware of these outliers when calculating summary statistics (e.g., mean) for these variables.

RSG has flagged trips with unreasonably high speeds, distances, or durations given their reported mode (flag_speed, flag_distance, flag_duration). These flags were created for single-mode trips, and a modified z-score of 4 was used as the cut-off for each of these metrics (i.e., trips more than 4 standard deviations away from the median for a given metric are considered outliers).

4.4 Derived and Recoded Variables

This dataset includes a combination of variables that were actively collected via survey questions, passively collected by the rMove or other metadata, implicitly assigned (e.g., administrative variables such as ID numbers), and derived or recoded (calculated from some combination of other variables). Key derived or recoded variables in this dataset are summarized below. Note that all location information in the Open Data, such as home and work location, is only provided at the CMS zone level.

Household-level Derived Variables

- Completion status
- Home geographies (block group, zone)
- Aggregate income (based on the initial and follow-up income questions)

Person-level Derived Variables

- Completion status
- Number of complete days
- Work/school geographies (state, county, block group, zone)

Day-level Variables

- Completion status

- Number of trips per day
- Day completion status

Trip-level Variables

- Trip speed
- Trip path distance (based on the GPS location data)
- Trip origin and destination geographies (state, county, block group, zone)
- Departure time (imputed in some cases)
- Trip purpose (imputed in some cases)
- Mode type and purpose categories

4.5 Imputation

Departure Time

In some cases, the rMove app may have detected the start of a trip after its true start time, which can yield invalid or extreme values for trip duration and speed. In these cases, the fields depart_date, depart_hour, and depart_minute were adjusted for “late pickup” conditions using the following approach:

- Departure time was imputed using the median speed between all locations along the trip, excluding the origin point, and the distance between the origin and the next point on the trip. For trips with fewer than three recorded locations, imputed departure time is set three minutes earlier than the original departure time to compensate for rMove’s 3-5-minute ping interval. Note that some trips that are the result of split loop trips may only have three or fewer points but will use the imputed depart time from before the loop trip was split and thus may not be included in this rule.
- If the imputed departure time overlaps with the previous trip’s arrival time, the previous trip’s arrival time was instead used as the departure time. Regardless of the number of locations along a trip, if the imputed departure time was later than the initially reported departure time, the imputed departure time is set to the original departure time. User-added trips as well as long distance passenger mode trips are also set to the original departure time, as user-added trips are not subject to “late pickup” conditions, and long-distance passenger modes are often plane trips where all collected traces contain speed information from other modes and thus are less reliable (as rMove cannot collect locations when a phone is in “airplane mode”).

Duration and speed are calculated based on the imputed departure time.

Purpose

Respondents report the purpose of the trip destination in each trip survey. The origin purpose is derived from the destination purpose of the previous trip, except for the first trip in the travel period or where an rMove trip occurs after a trip with item non-response. For the first trip in the travel period, the origin purpose can be inferred from “begin_day” in the day table.

When purpose was not asked because an analyst split a user-reported trip during data cleaning (creating a new destination along a trip), purpose values are derived where possible based on proximity (within 150 meters) to estimated home, work, or school locations. If the location is not proximate to home, work, or school locations, the purpose is set to "other."

The purpose category variables (o_purpose_category, d_purpose_category) contain aggregated purpose values based on the type of purpose at the origin/destination of each trip. Dataset users are welcome to perform their own recoding of the purpose categories as well.

Trip purposes have been imputed in cases where a purpose reported by the user is assumed to be inaccurate based on information about that person's reported habitual locations and other trips (primarily to home, work, and school locations). The trip purpose imputation approach was applied to all rMove trips in person-days with at least 1 complete trip and no more than 10 incomplete trips. ("Incomplete" trips are trips for which the respondent did not answer the trip-specific survey questions about purpose, mode, etc. for the given trip.)

The approach was to apply various "tests" in logical sequence to trips for which the stated purpose is not consistent with the location type based on the reported habitual locations. In general terms, the tests were designed to:

- Check the respondent's reported destination purpose when it conflicts with the destination location type. (The details of the tests depend on the trip purpose, with different criteria used for change-mode trips, pick-up/drop-off trips, linked transit trips, trips with home destinations but other reported purposes, etc.)
- Identify cases where respondents swapped the order of two or more trips when reporting their details.
- Identify cases where respondents may have omitted a trip and shifted remaining reported trip details by one trip when reporting the rest of their trips.
- Fill in missing data by sampling destination purposes from other trips made to the same locations, either by the same respondent or by other respondents.

4.6 Reminders for Data Users

Although travel survey data provides many opportunities for interesting analysis, data users should consider the context and best applications of the data. Data users should **keep the "universe" of data collection in mind** to ensure the analysis is logical for the data source. For example, users may try to treat the survey as a full population dataset, compare results directly to sources that are based on a different "universe" (such as traffic counts or census-based commuting data), or draw conclusions about groups or types of travel that were not fully captured in the survey design. We would want to remind users that the dataset represents a specific survey universe and should be interpreted within that scope. In particular, the survey does not include commercial vehicle travel (including travel by delivery bicyclists), travel by persons residing in group quarters outside of the address-based sampling frame (such as college dormitories or institutional housing), travel by nonresidents (for example, commuters traveling into the city from outside the study area or visitors to the city), trips made by children, or seasonal and holiday travel that occurred outside of the survey fielding period.

Second, data users should **use the weighted survey data** in any analysis wishing to draw conclusions about the city as a whole or within a zone (as opposed to describing only the survey respondents themselves). Applying weights ensures that the final analysis is regionally representative. Unweighted data may still be appropriate when a user is examining survey responses for quality control, methodological review, or exploratory analysis that focuses specifically on the survey participants rather than the population. In general, citywide weights should be used when producing results intended to represent New York City, while zonal weights should be used when the goal is to produce results that are representative within a CMZ zone or to compare results across zones in a way that reflects the population distribution within each zone.

Finally, data users should **ensure a sufficient sample size (and acknowledge margins of error)** in any analysis. The smaller the sample size, the larger the margin of error. For example, travel survey data users can generally draw reasonable conclusions about trip rates by mode on an average day but should consider the sample size for modes with small shares of overall citywide travel.

4.7 Practical Tips for Data Analysis

Many ways exist to view, join, summarize, and map CMS data. To achieve the full data benefits, data users may need multiple tools. Data users can apply three key data analysis mechanics to make the most of their analyses.

Joining Tables on Unique IDs

All data tables can be joined into a single database as needed. Some unique IDs are a combination of two variables. In these cases, joining on only one of the variables will create duplicate records.

Table name	Variable(s) to join to other survey data tables
Household	hh_id
Person	hh_id, person_id
Vehicle	hh_id
Day	hh_id, person_id, day_num
Trip	hh_id, person_id, day_num, trip_id
Location	trip_id

Data table Relationships

Records in each data table can be linked to records in other tables, for example:

- Each person belongs to one household; each trip belongs to one person
- One household may have many vehicles, persons, and trips;
  a household may have no vehicles and trips, but must have at least one person

Records are linked through ID numbers in each table:

HH ID	Person #	Person ID	Trip #	Trip ID
24000430	2	2400043002	1	2400043002001
24000460	1	2400046001	1	2400046001001
24000460	2	2400046002	2	2400046002002

Applying and Interpreting Weights

Analyses designed to draw conclusions about travel behavior in the city (as opposed to just the survey respondents) should use weighted data. When applied, the weights make the dataset representative of personal citywide or zone-level travel for the time period studied. This means it does not include commercial vehicle travel (including by delivery bicyclists), travel for persons residing in group quarters outside of the address-based sampling frame (e.g., college dorms, institutional housing), travel from nonresidents (e.g., commuters from outside the city, and visitors to the city), or seasonal/holiday travel outside of the survey fielding period.

Using weighted data generally involves summing the weights for the groups of interest. The sum of weights in each table represents the following groups:

Household: Represents the total number of households within the survey region.

Person: Represents the total number of adult residents (residents 18 or older) within the survey region.

Vehicle: Represents the total number of personal vehicles of households in the survey region.

Day: Represents one average day for all adults residing in the survey region. This is equal to the number of adult residents in the region.

Trip: Represents the total number of trips all adults residing in the survey region make on a typical day. This differs from the number of trips made by adult residents *in the survey region* on a typical day, given that some residents make trips outside the region and that some trips within the survey region are made by non-residents or by resident children.

To calculate weighted crosstabs or descriptive statistics, sum the weights for that table. Grouping variables or filters will provide weighted totals for specific subgroups, like two-person households, or by age group. Keep in mind the following when creating weighted statistics and summaries from travel survey data:

1. **Filter to the data relevant to your analysis (e.g., complete travel days).**
   - For example, focusing on travel days with complete survey information (i.e., no unanswered survey questions) is best. Also note that not all people are asked every question, so understanding the 'missing value' codes can be important.

2. **Remember the survey design when using and interpreting weighted values.**
   - When working with trip and day data, it is important to consider that participants using smartphones app recorded more days and trips compared to one-day online and call center participants. To ensure equal representation, the trip rates (not attributes) for the online and call-center surveys participants were adjusted to better match those of the smartphone-based surveys. For more details, refer to the detailed methodology in section 3.
3. **Make sure to use the correct weight for the analysis you intend to perform.**
   - CMS is a weighted sample, and it is necessary to use the weighting variables to generate estimates and standard errors that accurately represent the population of the city. In general, use household weight for household-level analyses; use person weight for person-level analyses. Apply day weight for household-day and person-day analyses, and trip weight for trip-level analyses. When working with a merged file that includes household and person records, use person weights to estimate person characteristics. Remember to exercise caution when analyzing person and household characteristics from merged files.

Generating Trip Rates

Trip rates – or the number of trips per day among groups – are useful for comparing several travel behaviors (e.g., travel by mode, travel by age group). Trip rates can be weighted or unweighted, but this section focuses on the former.

To calculate a weighted trip rate, data users must divide the number of weighted trips by the number of weighted travel days. For example, if there are 300,000 weighted person-trips across 75,000 person-days, then the average person-trip rate is 4.0 per day. If there are 225,000 person-trips by car across 75,000 person-days, then the person-trip rate for car trips is 3.0. This is different than calculating vehicle trip rates, which would require calculating the weighted vehicle trips taking place (in this example, if the average vehicle occupancy is two people, then the vehicle trip rate would be 1.5).

Data users should always calculate the number of weighted travel days using the day table rather than the trip table given that persons with zero-trip travel days do not have any records in the trip tables for those days.

Trip Unlinking

The trip unlinker is run on trips that have no auto mode and there is a bus or rail trip. The qualifying trips are queried through the Google API and the route with the best spatial overlap with the collected trip's trace data is selected. Trips are "split" into their access, transit, and egress components as a result of this unlinking process- seen under the "unlinked_trip" field in trip data. If a trip was split in manual trip editing, the directions will be deleted and the trip will be re-routed based on the new start/stop points.

Synthetic Trips

Synthetic trips are created for transit trips that are missing an access/egress leg. Synthetic trips are identified with the transit_quality_flag variable in the trip table. These trips have distance and duration set to NULL (to avoid skewing summary statistics). These trips identify where we know that there was a trip, but we are unsure of the details. Synthetic access trips will have the same depart and arrive time as the depart time for the subsequent trip and the same origin/destination as the origin of the subsequent trip.

Synthetic egress trips will have the same depart and arrive time as the arrive time of the prior trip and the same origin/destination as the prior trip's destination.

Mode type (mode_type in trip table)

Mode_type synthesizes mode_1 to mode_9 down to a single, easier-to-use variable for analytical purposes (so that data users can avoid always referencing all modes on a multimodal trips). Survey respondent may have specified multiple modes for a trip, but trip record doesn't contain attributes such as duration or cost for each of those modes, it's common to estimate which of these modes was the primary or main mode of the trip. Mode_type is our best estimate of the main mode and is determined by considering how a typical user might describe a trip that combines mode. In general, the mode normally used for longer-distance trips takes precedence – for example, a trip that includes both walking and subway modes is normally characterized as a subway trip. Table D1 in Appendix D shows the full crosswalk of which detailed modes correspond to which mode_types. Higher values of mode_type are prioritized over lower mode_type values in the derivation. For example, transit trips, with mode_type 13, are prioritized over walk trips, with mode_type 1. When transit trips were unlinked using the Google API during cleaning, the non-transit legs of the trip were recoded using Google's suggested mode (most frequently "walk" or "bike).

Purpose Categories (d_purpose_category and o_purpose_category in trip table)

Purposes are up-coded (e.g., d_purpose_category) to purpose categories guided by the questionnaire for convenience. A full crosswalk of purposes and purpose categories are shown in Table D2 in Appendix D.

5.0 Further Information

For further information about the Citywide Mobility Survey program, please visit the CMS information page on the NYC DOT website.

CMS information page: https://www.nyc.gov/html/dot/html/about/citywide-mobility-survey.shtml

Appendix A: 2024 NYC CMS Travel Survey Weighting Output

Table A1: Initial Weights

Invitation	Alpha	Segment	Samples	1-Year ACS	Initial Weight
ABS	0.69	General - Inner Brooklyn	215	770,708	2,455.80
	0.69	General - Inner Queens	198	388,739	1,345.04
	0.69	General - Manhattan Core	211	795,664	2,583.38
	0.69	General - Middle Queens	179	313,601	1,200.23
	0.69	General - Northern Bronx	192	402,117	1,434.80
	0.69	General - Outer Brooklyn	154	779,637	3,468.27
	0.69	General - Outer Queens	195	788,983	2,771.88
	0.69	General - Southern Bronx	115	225,498	1,343.34
	0.69	General - Staten Island	227	351,656	1,061.29
	0.69	General - Upper Manhattan	236	330,212	958.57
	0.69	HTR - Inner Brooklyn	33	164,328	3,411.45
	0.69	HTR - Inner Queens	61	143,027	1,606.31
	0.69	HTR - Manhattan Core	19	70,750	2,551.01
	0.69	HTR - Middle Queens	42	63,568	1,036.88
	0.69	HTR - Northern Bronx	36	107,348	2,042.84
	0.69	HTR - Outer Brooklyn	67	246,350	2,518.95
	0.69	HTR - Outer Queens	34	93,217	1,878.27
	0.69	HTR - Southern Bronx	139	248,915	1,226.81
	0.69	HTR - Staten Island	4	26,782	4,586.91
	0.69	HTR - Upper Manhattan	49	113,432	1,585.91
		Subtotal	2,406	6,461,952	
NPS	0.31	Reinvited from previous surveys	1,106	6,461,952	1,840

Table A2: PUMA to CMS Zone allocation Table

PUMA	CMS Zone	Target Population	Allocation
3604107	Manhattan Core	160,976	92%
	Upper Manhattan	13,472	8%
3604207	Southern Bronx	87,656	92%
	Northern Bronx	7,561	8%
3604305	Outer Brooklyn	121,915	85%
	Inner Brooklyn	21,600	15%
3604312	Outer Brooklyn	105,417	86%
	Inner Brooklyn	17,222	14%
3604317	Outer Brooklyn	105,414	90%
	Inner Brooklyn	12,185	10%
3604405	Outer Queens	85,891	61%
	Middle Queens	56,019	39%
3604407	Outer Queens	139,910	72%
	Middle Queens	54,144	28%
3604408	Outer Queens	81,536	75%
	Middle Queens	27,258	25%
3604412	Outer Queens	143,133	77%
	Manhattan Core	160,976	92%
3604103	Manhattan Core	146,049	100%
3604104	Manhattan Core	113,403	100%
3604108	Manhattan Core	208,015	100%
3604109	Upper Manhattan	102,040	100%
3604110	Upper Manhattan	131,544	100%
3604111	Upper Manhattan	117,326	100%
3604112	Upper Manhattan	210,715	100%
3604121	Manhattan Core	148,677	100%
3604165	Manhattan Core	183,399	100%
3604204	Southern Bronx	149,130	100%

PUMA	CMS Zone	Target Population	Allocation
3604205	Southern Bronx	131,760	100%
3604208	Northern Bronx	92,722	100%
3604209	Northern Bronx	170,272	100%
3604210	Northern Bronx	124,893	100%
3604211	Northern Bronx	112,984	100%
3604212	Northern Bronx	157,588	100%
3604221	Southern Bronx	150,425	100%
3604263	Southern Bronx	165,844	100%
3604301	Inner Brooklyn	199,114	100%
3604302	Inner Brooklyn	119,565	100%
3604303	Inner Brooklyn	178,989	100%
3604304	Inner Brooklyn	112,934	100%
3604306	Inner Brooklyn	119,251	100%
3604307	Inner Brooklyn	122,203	100%
3604308	Inner Brooklyn	107,462	100%
3604309	Inner Brooklyn	97,637	100%
3604310	Outer Brooklyn	125,296	100%
3604311	Outer Brooklyn	187,329	100%
3604313	Outer Brooklyn	108,089	100%
3604314	Outer Brooklyn	156,955	100%
3604315	Outer Brooklyn	155,697	100%
3604316	Inner Brooklyn	90,397	100%
3604318	Outer Brooklyn	198,391	100%
3604401	Inner Queens	169,933	100%
3604402	Inner Queens	123,263	100%
3604403	Inner Queens	159,759	100%
3604404	Inner Queens	173,921	100%
3604406	Middle Queens	121,859	100%
3604409	Middle Queens	149,688	100%

PUMA	CMS Zone	Target Population	Allocation
3604410	Outer Queens	137,707	100%
3604411	Outer Queens	122,436	100%
3604413	Outer Queens	202,547	100%
3604414	Outer Queens	128,263	100%
3604501	Staten Island	178,934	100%
3604502	Staten Island	139,228	100%
3604503	Staten Island	167,324	100%

Table A3: Income Imputation Model Summary

Parameter	Description	Estimate	std error	t-stat	P-value
f_0_24999	Fraction of people in block group with incomes under $25,000	-1.332	0.407	-3.27	0.001
f_25000_49999	Fraction of people in block group with incomes $25,000–$50,000	-0.766	0.468	-1.64	0.102
f_50000_74999	Fraction of people in block group with incomes $50,000–$75,000	-1.003	0.496	-2.02	0.043
f_75000_99999 (reference level)	Fraction of people in block group with incomes $75,000–$99,999	NA	NA	NA	NA
f_100000_199999	Fraction of people in block group with incomes $100,000–$199,000	0.402	0.451	0.89	0.372
f_200000_plus	Fraction of people in block group with incomes more than $200,000	2.567	0.406	6.32	<0.001
nonworking_adult_n	Number of non-working adults in household	-0.853	0.129	-6.63	<0.001
full_time_graduate_degree_n	Number of full-time workers with graduate degrees in household	1.917	0.130	14.70	<0.001
part_time_graduate_degree_n	Number of part-time workers with graduate degrees in household	0.182	0.246	0.74	0.458
full_time_bachelor_degree_n	Number of full-time workers with bachelor's degrees in household	1.348	0.127	10.64	<0.001
part_time_bachelor_degree_n	Number of part-time workers with bachelor's degrees in household	-0.523	0.234	-2.23	0.026
full_time_no_college_n	Number of full-time workers with no advanced degrees in household	0.484	0.141	3.45	<0.001
part_time_no_college_n	Number of part-time workers with no advanced degrees in household	-1.074	0.198	-5.42	<0.001
single_family_home	Lives in single family housing	-0.124	0.076	-1.64	0.102
head_under_35	Age of respondent is under 35	0.216	0.112	1.94	0.052
head_65_plus	Age of respondent is 65 or over	1.181	0.080	14.79	<0.001
own_home	Respondent owns or rents their home	-0.187	0.086	-2.17	0.03

Appendix B: Trip Rate Adjustment Models (CMS Zone Weights)

Table B1: Home-based mandatory trip model (CMS zone weights)

parameter	Description	estimate	std error	t-statistic
(Intercept)		-3.4209	0.1103	-31.0155
I(diary_online | diary_call)TRUE	Online & Call-center diary	-2.3037	0.1623	-14.1922
lower_income	Income less than 50k	-0.4984	0.1216	-4.0974
higher_income	Income over 200k	-0.0383	0.0557	-0.6866
sf_home	Lives in single family home	0.2266	0.0539	4.2070
age_under_25	Age is under 25	-0.4090	0.1210	-3.3791
age_over_65	Age is over 65	-0.4537	0.1008	-4.4992
bachelors	Has bachelor's degree	-0.2207	0.0612	-3.6073
grad_degree	Has masters/PhD	-0.3734	0.0664	-5.6235
is_student	Is student	-0.4880	0.1159	-4.2102
employed_ft	Employed full-time	1.8293	0.1047	17.4779
employed_pt	Employed part-time	1.4163	0.1171	12.0985
work_loc_varies	Has masters/PhD	0.6659	0.0663	10.0444
two_plus_jobs	Works 2+ jobs	0.1701	0.0743	2.2901
telework_everyday	Teleworks >4 days a week	-0.1945	0.1282	-1.5173
telework_rarely	Telework <1 day per month	0.1055	0.0883	1.1956

R-squared: 0.23

Table B2: Home-based other trip model (CMS zone weights)

parameter	Description	estimate	std error	t-statistic
(Intercept)		-0.0631	0.0295	-2.1422
I(diary_online | diary_call)TRUE	Online & Call-center diary	-1.4319	0.0341	-41.9364
lower_income	Income less than 50k	-0.0587	0.0277	-2.1231
higher_income	Income over 200k	0.1193	0.0195	6.1083
sf_home	Lives in single family home	0.1654	0.0180	9.2070
age_under_25	Age is under 25	-0.1832	0.0332	-5.5258
age_over_65	Age is over 65	-0.1857	0.0241	-7.7051
bachelors	Has bachelor's degree	0.2340	0.0213	11.0006
grad_degree	Has masters/PhD	0.2188	0.0221	9.9202
is_student	Is student	-0.2191	0.0359	-6.1031
employed_ft	Employed full-time	-0.2389	0.0272	-8.7823
employed_pt	Employed part-time	-0.1117	0.0328	-3.4042
work_loc_varies	Has masters/PhD	0.0827	0.0303	2.7296
two_plus_jobs	Works 2+ jobs	-0.0310	0.0253	-1.2256
telework_everyday	Teleworks >4 days a week	0.0706	0.0419	1.6855
telework_rarely	Telework <1 day per month	0.0094	0.0350	0.2686

R-squared: 0.24

Table B3: Non-home-based mandatory trip model (CMS zone weights)

parameter	Description	estimate	std error	t-statistic
(Intercept)		-1.7353	0.0543	-31.9322
I(diary_online | diary_call)TRUE	Online & Call-center diary	0.0719	0.0295	2.4394
lower_income	Income less than 50k	-0.5782	0.0588	-9.8295
higher_income	Income over 200k	0.0019	0.0264	0.0717
sf_home	Lives in single family home	-0.0364	0.0273	-1.3331
age_under_25	Age is under 25	-0.3348	0.0529	-6.3266
age_over_65	Age is over 65	-0.7150	0.0524	-13.6337
bachelors	Has bachelor's degree	-0.1281	0.0284	-4.5131
grad_degree	Has masters/PhD	-0.4483	0.0329	-13.6401
is_student	Is student	-0.0099	0.0475	-0.2081
employed_ft	Employed full-time	1.4917	0.0516	28.8915
employed_pt	Employed part-time	0.9966	0.0575	17.3279
work_loc_varies	Has masters/PhD	0.7574	0.0308	24.5868
two_plus_jobs	Works 2+ jobs	-0.2065	0.0400	-5.1601
telework_everyday	Teleworks >4 days a week	-0.8852	0.0895	-9.8869
telework_rarely	Telework <1 day per month	0.0939	0.0432	2.1718

R-squared:0.33

Table B4: Non-home-based other trip model (CMS zone weights)

parameter	Description	estimate	std error	t-statistic
(Intercept)		1.2626	0.0139	90.7039
I(diary_online | diary_call)TRUE	Online & Call-center diary	0.1008	0.0090	11.2074
lower_income	Income less than 50k	-0.0052	0.0120	-0.4307
higher_income	Income over 200k	-0.0134	0.0088	-1.5358
sf_home	Lives in single family home	-0.1317	0.0087	-15.1268
age_under_25	Age is under 25	-0.0089	0.0143	-0.6251
age_over_65	Age is over 65	-0.1506	0.0108	-13.9781
bachelors	Has bachelor's degree	0.1568	0.0096	16.3381
grad_degree	Has masters/PhD	0.1254	0.0101	12.4753
is_student	Is student	-0.0660	0.0152	-4.3385
employed_ft	Employed full-time	0.0907	0.0129	7.0557
employed_pt	Employed part-time	0.0338	0.0148	2.2821
work_loc_varies	Has masters/PhD	0.0961	0.0136	7.0842
two_plus_jobs	Works 2+ jobs	0.0713	0.0116	6.1249
telework_everyday	Teleworks >4 days a week	-0.0909	0.0218	-4.1770
telework_rarely	Telework <1 day per month	0.0553	0.0155	3.5680

R-squared: 0.06

Appendix C:
Trip Rate Adjustment Models (Citywide weights)

Table C1: Home-based mandatory trip model (Citywide weights)

parameter	Description	estimate	std error	t-statistic
(Intercept)		-3.5859	0.1161	-30.8753
I(diary_online | diary_call)TRUE	Online & Call-center diary	-1.8223	0.1390	-13.1061
lower_income	Income less than 50k	-0.5817	0.1237	-4.7018
higher_income	Income over 200k	-0.0965	0.0554	-1.7412
sf_home	Lives in single family home	0.3380	0.0520	6.4951
age_under_25	Age is under 25	-0.3997	0.1198	-3.3353
age_over_65	Age is over 65	-0.4574	0.0997	-4.5860
bachelors	Has bachelor's degree	-0.1907	0.0609	-3.1330
grad_degree	Has masters/PhD	-0.2889	0.0654	-4.4194
is_student	Is student	-0.3045	0.1111	-2.7414
employed_ft	Employed full-time	1.9249	0.1115	17.2577
employed_pt	Employed part-time	1.5425	0.1216	12.6867
work_loc_varies	Has masters/PhD	0.6856	0.0651	10.5311
two_plus_jobs	Works 2+ jobs	0.0855	0.0749	1.1416
telework_everyday	Teleworks >4 days a week	-0.1436	0.1307	-1.0986
telework_rarely	Telework <1 day per month	0.3139	0.0811	3.8722

R-squared: 0.23

Table C2: Home-based other trip model (Citywide weights)

parameter	Description	estimate	std error	t-statistic
(Intercept)		-0.0604	0.0299	-2.0194
I(diary_online | diary_call)TRUE	Online & Call-center diary	-1.5697	0.0376	-41.7018
lower_income	Income less than 50k	-0.0983	0.0276	-3.5625
higher_income	Income over 200k	0.1539	0.0194	7.9166
sf_home	Lives in single family home	0.1713	0.0176	9.7495
age_under_25	Age is under 25	-0.1436	0.0325	-4.4207
age_over_65	Age is over 65	-0.1414	0.0234	-6.0483
bachelors	Has bachelor's degree	0.1974	0.0211	9.3653
grad_degree	Has masters/PhD	0.1717	0.0221	7.7837
is_student	Is student	-0.1895	0.0353	-5.3651
employed_ft	Employed full-time	-0.2364	0.0278	-8.4915
employed_pt	Employed part-time	-0.0668	0.0326	-2.0489
work_loc_varies	Has masters/PhD	0.0082	0.0310	0.2639
two_plus_jobs	Works 2+ jobs	-0.0221	0.0255	-0.8684
telework_everyday	Teleworks >4 days a week	0.0624	0.0459	1.3577
telework_rarely	Telework <1 day per month	-0.0145	0.0349	-0.4158

R-squared: 0.24

Table C3: Non-home-based mandatory trip model (Citywide weights)

parameter	Description	estimate	std error	t-statistic
(Intercept)		-2.0316	0.0578	-35.1322
I(diary_online | diary_call)TRUE	Online & Call-center diary	0.0990	0.0309	3.2070
lower_income	Income less than 50k	-0.5019	0.0586	-8.5721
higher_income	Income over 200k	-0.0447	0.0269	-1.6593
sf_home	Lives in single family home	-0.1117	0.0276	-4.0536
age_under_25	Age is under 25	-0.1317	0.0506	-2.6035
age_over_65	Age is over 65	-0.4619	0.0488	-9.4562
bachelors	Has bachelor's degree	-0.1656	0.0292	-5.6703
grad_degree	Has masters/PhD	-0.3716	0.0326	-11.3837
is_student	Is student	0.0329	0.0471	0.6975
employed_ft	Employed full-time	1.8271	0.0554	32.9746
employed_pt	Employed part-time	1.2827	0.0602	21.3181
work_loc_varies	Has masters/PhD	0.6363	0.0324	19.6280
two_plus_jobs	Works 2+ jobs	-0.1763	0.0396	-4.4487
telework_everyday	Teleworks >4 days a week	-0.6467	0.0847	-7.6386
telework_rarely	Telework <1 day per month	0.0703	0.0441	1.5947

R-squared:0.33

Table C4: Non-home-based other trip model (City-wide weights)

parameter	Description	estimate	std error	t-statistic
(Intercept)		1.2673	0.0140	90.2299
I(diary_online | diary_call)TRUE	Online & Call-center diary	0.0524	0.0095	5.5295
lower_income	Income less than 50k	-0.0193	0.0122	-1.5795
higher_income	Income over 200k	0.0060	0.0088	0.6840
sf_home	Lives in single family home	-0.1695	0.0086	-19.6539
age_under_25	Age is under 25	0.0603	0.0142	4.2574
age_over_65	Age is over 65	-0.0970	0.0107	-9.0599
bachelors	Has bachelor's degree	0.1601	0.0096	16.6280
grad_degree	Has masters/PhD	0.1228	0.0101	12.1155
is_student	Is student	-0.0864	0.0153	-5.6595
employed_ft	Employed full-time	0.1121	0.0130	8.6005
employed_pt	Employed part-time	0.0621	0.0148	4.1962
work_loc_varies	Has masters/PhD	0.0581	0.0138	4.2234
two_plus_jobs	Works 2+ jobs	0.0398	0.0116	3.4138
telework_everyday	Teleworks >4 days a week	-0.0117	0.0226	-0.5173
telework_rarely	Telework <1 day per month	0.0322	0.0158	2.0381

R-squared: 0.07

Appendix D: Mode and Purpose Categories

Table D1: Crosswalk of detailed mode and prioritization

Detailed mode	Detailed mode value	Mode type	Mode type value
Walk (or jog/wheelchair)	1	Walk	1
Standard bicycle	2	Bike	2
Other rented bicycle	4	Bike	2
Electric bicycle	82	Bike	2
Bicycle or e-bicycle	103	Bike	2
E-scooter, skateboard, unicycle	107	Bike	2
Citi bike share (standard bicycle)	201	Bike	2
Citi bike share (electric bicycle)	202	Bike	2
Standard cargo bicycle	203	Bike	2
Electric cargo bicycle	204	Bike	2
E-scooter share (e.g., Bird, Lime, Veo)	83	Scooter Share	4
Other moped (not my household's)	113	Scooter Share	4
E-scooter (e.g., Ninbot, GOTRAX)	205	Scooter Share	4
Other electric (e.g., e-skateboard, hoverboard, e-unicycle, or other small electric device)	206	Scooter Share	4
Other non-electric (e.g., personal kick scooter, skateboard, rollerblade, or other small device)	207	Scooter Share	4
Regular taxi (e.g., Yellow Cab)	36	Taxi	5
Other hired car service (e.g., livery, black car, limo)	60	Taxi	5
Uber, Lyft, or other smartphone-app ride service	49	Smartphone-app ridehailing service	6
Smartphone-app ride service (Uber/Lyft), taxi, or car service	106	Smartphone-app ridehailing service	6
Other	5	Other	7
Golf cart	44	Other	7
ATV	45	Other	7
Medical transportation service	63	Other	7
Other boat (e.g., kayak)	80	Other	7
Snowmobile	81	Other	7
Other	104	Other	7
Household vehicle 1	6	Vehicle	8
Household vehicle 2	7	Vehicle	8
Household vehicle 3	8	Vehicle	8
Household vehicle 4	9	Vehicle	8
Household vehicle 5	10	Vehicle	8
Household vehicle 6	11	Vehicle	8
Household vehicle 7	12	Vehicle	8
Household vehicle 8	13	Vehicle	8

Detailed mode	Detailed mode value	Mode type	Mode type value
Household vehicle 9	14	Vehicle	8
Household vehicle 10	15	Vehicle	8
Other car or truck in household	16	Vehicle	8
Rental car	17	Vehicle	8
Other car or truck (not my household's)	22	Vehicle	8
Work vehicle	33	Vehicle	8
Friend/relative/colleague's car or truck	34	Vehicle	8
Other motorcycle in household	47	Vehicle	8
Household car, truck, moped or motorcycle	100	Vehicle	8
Other vehicle (e.g., friend's, rental, work)	101	Vehicle	8
Other moped in household	111	Vehicle	8
Other motorcycle (not my household's)	114	Vehicle	8
Carshare service (e.g., Zipcar)	18	Car Share	9
Peer-to-peer car rental (e.g., Turo)	112	Car Share	9
School bus	24	School bus	10
Shuttle (e.g., university, employer, hotel)	26	Shuttle/vanpoo l	11
Paratransit/Dial-A-Ride	27	Shuttle/vanpoo l	11
Vehicle ferry (took vehicle on board)	79	Ferry	12
Commuter Van (dollar van)	21	Transit	13
Local bus	23	Transit	13
Intercity bus (e.g., Greyhound, Megabus)	25	Transit	13
Other bus	28	Transit	13
NYC subway	30	Transit	13
Staten Island Railroad	39	Transit	13
Other rail	42	Transit	13
Express bus	55	Transit	13
Commuter rail (e.g., LIRR, Metro-North, NJ Transit)	58	Transit	13
Select Bus Service	61	Transit	13
Bus, or shuttle	102	Transit	13
Subway, train or rail	105	Transit	13
Ferry	108	Transit	13
PATH train	174	Transit	13
Airplane/helicopter	31	Long Distance Passenger	14
Intercity rail (e.g., Amtrak)	41	Long Distance Passenger	14

Table D2: Crosswalk of purposes and purpose categories

Purpose label	Purpose	Purpose category label	Purpose category
Went home	1	Home	1
Went to primary workplace	10	Work	2
Went to work-related activity (e.g., meeting, delivery, worksite)	11	Work Related	3
Volunteering	13	Work Related	3
Other work-related	14	Work Related	3
Attend K-12 school	21	School	4
Attend college/university	22	School	4
Attend other type of class (e.g., cooking class)	23	School	4
Attend vocational education class	25	School	4
Attend daycare or preschool	26	School	4
Attend other education-related activity (e.g., field trip)	24	School Related	5
Other activity only (e.g., attend meeting, pick-up or drop-off item)	44	Escort	6
Pick someone up	45	Escort	6
Drop someone off	46	Escort	6
Accompany someone only (e.g., go along for the ride)	47	Escort	6
BOTH pick up AND drop off	48	Escort	6
Grocery shopping	30	Shop	7
Got gas	31	Shop	7
Other routine shopping (e.g., pharmacy)	32	Shop	7
Shopping for major item (e.g., furniture, car)	36	Shop	7
Dined out, got coffee, or take-out	50	Meal	8
Exercise or recreation (e.g., gym, jog, bike, walk dog)	51	Social/Recreation	9
Social activity (e.g., visit friends/relatives)	52	Social/Recreation	9
Leisure/entertainment/cultural (e.g., cinema, museum, park)	53	Social/Recreation	9
Religious/civic/volunteer activity	54	Social/Recreation	9
Family activity (e.g., watch child's game)	56	Social/Recreation	9
Errand without appointment (e.g., post office)	33	Errand	10
Medical visit (e.g., doctor, dentist)	34	Errand	10
Errand with appointment (e.g., haircut)	37	Errand	10
Changed or transferred mode (e.g., waited for bus or exited bus)	60	Change Mode	11
Went to another residence (e.g., someone else's home, second home)	150	Overnight	12
Went to temporary lodging (e.g., hotel, vacation rental)	152	Overnight	12
Other	61	Other	13
Other	62	Other	13
Other reason	99	Other	13