GTAModel Version 4.2 “1000 Run Experiment”

Lisa Guseva
Gozde Ozonder, Ph.D.
Eric J. Miller, Ph.D.

August 30, 2021

UNIVERSITY OF TORONTO
FACULTY OF APPLIED SCIENCE & ENGINEERING
Transportation Research Institute

UTTRI

Table of Contents

1. INTRODUCTION

GTAModel Version 4.2 is an agent-based microsimulation model of out-of-home activities and travel in the Greater Toronto-Hamilton Area. The newest version was recently developed at the University of Toronto and it is mainly based on a set of complex Monte Carlo simulations that generate a 100% synthesized population of the GTHA. It is in operational use by the cities of Toronto, Mississauga, Brampton and Vaughan, as well as the Regions of Durham and Peel.

The model system's statistical characteristics, on the other hand, have never been thoroughly investigated. To investigate the run-to-run variability in output, the model system was run for 1000 replications with a given base scenario. The model's findings were saved in files, which were then utilized to conduct computations on the data and produce graphs.

This work entails calculating and comparing a variety of statistics using the run results. To decide how to more efficiently and effectively evaluate the data, a vast database comprising model run results is used. This work requires using Python to calculate the mean, standard deviation, and coefficient of variation for a collective of variables. These variables are:

- Number of out-of-home work activities generated by occupation sector, by employment status, by spatial categories.
- Number of out-of-home school activities generated by student status, by spatial categories.
- Number of out-of-home shopping activities generated by spatial categories.
- Number of out-of-home other activities generated by spatial categories.
- Trip length frequency distributions by travel time by mode by spatial categories.
- Average auto ownership by spatial categories.
- Driver's license ownership by spatial categories.
- Total number of trips by time-of-day by spatial categories.

Using the outputs of the 1000 runs, bar graphs using Python or Excel were generated to detect and interpret any variability or anomalies in the data and determine the cause of those fluctuations. For the purpose of this report, only the results associated with the first 50 runs are examined and documented.

The aim of this document is to outline the progress of the study and to present the results. This report is organized into four major categories: Data, Methods, Results, and Conclusions. The Data section presents a description of the simulation outputs from 1000 runs of the GTAModel as well as the files what were used in the code. The Methods section includes the descriptions of the coding script used to perform calculations on the data. It explains the rationale that was employed, as well as the reasons for why it was a successful approach. The Results section documents the results that were acquired from calculating the mean, standard deviation, and coefficient of variation, as well as the resulting graphs.

2. Data

The data collected from the 1000 runs of the GTAModel was saved in folders containing zip files. The first folder contains data collected from 50 runs, while the subsequent folders contain data from 10 runs up to 1000 runs. Each folder contains zip files which have compressed copies of the data for a single run. Each zip file has directories that further split down the contents. These folders are: AssignedNetworks, Demand, LOS Matrices, Microsim Results, Screenlines, StationAccess, and Validation.

For the purpose of this research, the files used are solely the ones contained in the Microsim Results folder. For each run, this folder has multiple csv files: facilitate_passenger, households, persons, trip_modes, trip_stations, and trips. This is where the model's data are saved. The files utilized in this research for the variables being examined were households, people, trip modes, and trips. Each file involved in the study is listed and discussed in Table 2.1.

Table 2.1: Microsim Results File Descriptions [4]

File	Information Stored & Descriptions

households.csv	• household_id
	  ○ The unique identifier for the household
	• home_zone
	• weight
	  ○ This is the same as the household expansion factor in TTS
	• persons
	  ○ The number of people living within the household
	• dwelling_type
	• vehicles
	  ○ The number of vehicles assigned to the household from the auto ownership model
	• income_class
	  ○ The income class according to TTS2016 that the household belongs to
	  ○

persons.csv	• household_id
	  ○ The unique identifier for the household
	• person_id
	  ○ The unique identifier for the person within the household
	• age
	• gender
	  ○ Male of Female
	• driver's license
	  ○ TRUE, or FALSE

	• transit_pass ○ TRUE, or FALSE • employment_status [2] ○ F = Full-time ○ H = Work at home full-time ○ J = Work at home part-time ○ O = Not employed ○ P = Part-time ○ 9 = Unknown • occupation [2] ○ G = General office / clerical ○ M = Manufacturing / construction / trades ○ P = Professionsal / management / technical ○ O = Not employed ○ 9 = Unknown • free_parking ○ TRUE, or FALSE • student_status ○ F = Full-time, P = Part-Time, O = Not a student • work_zone ○ If the zone is 0, then they have not been assigned a work zone • school_zone ○ If the zone is 0, then they have not been assigned a school zone • weight ○ The person's expansion factor to scale the agent to the global population
trip_modes.csv	• household_id ○ The unique identifier for the household • person_id ○ The unique identifier for the person within the household • trip_id ○ The unique identifier for the trip for the given person • mode ○ The name of the mode that was used • o_depart ○ The time that the trip started / departed the origin • d_arrive

	○ The time that the trip ended / arrived at the destination • weight ○ The number of times that this mode was selected for the given trip
trips.csv	• household_id ○ The unique identifier for the household • person_id ○ The unique identifier for the person within the household • trip_id ○ The unique identifier for the trip for the given person • o_act ○ The activity at the origin of the trip • o_zone ○ The zone number of the trip's origin • d_act ○ The activity at the destination of the trip • d_zone ○ The zone number of the trip's destination • weight ○ The factor to apply to scale this trip to scale the trip for the global population ○ Same as the person's weight

Because each variable is categorized by spatial category, understanding which section of the GTHA each spatial category relates to is useful. Each spatial category represents a planning district or a region. This relationship is listed below, and the layout of Toronto's planning districts is represented in Figure 1.

- Spatial category 1 → PD1
- Spatial category 2 → PD2, PD3, PD4, PD6
- Spatial category 3 → PD7, PD8, PD9
- Spatial category 4 → PD13, PD14, PD15, PD16
- Spatial category 5 → PD5, PD10, PD11, PD12
- Spatial category 6 → Durham
- Spatial category 7 → York

- Spatial category 8 → Peel
- Spatial category 9 → Halton
- Spatial category 10 → Hamilton

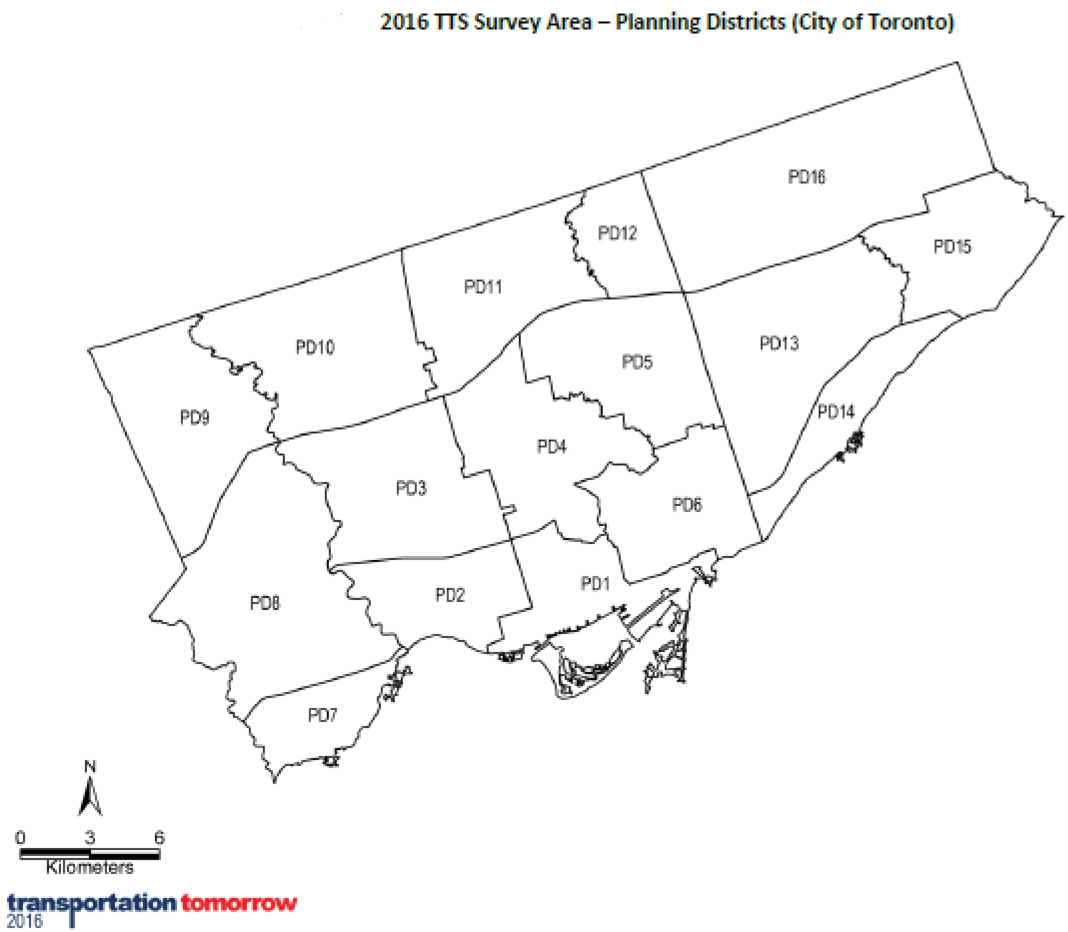


Figure 1: City of Toronto Broken Up Into Planning Districts

3. Methods

This section describes the logic that was utilized to program each variable as well as the strategy.

3.1. General Strategy

In general, two functions were used to process each variable. The first function is responsible for data extraction and organization. Data are obtained from the relevant files and stored in a way that allows computations and graph generation to be performed conveniently. The second function is the one that performs those calculations.

There are a few functions that are employed consistently in every variable analyzed. The Conversions function, for example, is used to convert zones into planning districts and then into spatial categories. It returns a tuple with two dictionaries as elements. The first dictionary is the zone to planning district conversion where the zones are the keys, and the planning districts are the values. The second dictionary is the conversion of planning districts to spatial groupings. The keys are the planning districts, and the values are the spatial categories that correspond to them.

Get Archive is the next function that is used to obtain the files containing model run outputs. This function accepts two parameters: the location to the needed zip file and the name of the desired csv file. This function locates and returns the required csv file from the zip file. If the file cannot be located, an exception is thrown by the function.

These two functions are used within other scripts in order to gather data and sort it by spatial categories.

3.2. Out-of-Home Work Activities

This variable computes the amount of work trips that occurred in each spatial category, sorted by employment status as well as occupation sector.

In the first function, data is extracted from the trips file and a list of household IDs, person IDs, activities, and spatial categories are made. The persons file is read and two dictionaries were created. The first dictionary is a nested dictionary where the keys are the household IDs, and the values are a nested dictionary. The keys are the person IDs and the values are the employment statuses. The keys of the second dictionary are the household IDs and the values are nested dictionaries where the keys are the person IDs and the values are the occupation sectors.

A for loop is utilized to go through the list of household IDs. If the corresponding activity is “PrimaryWork”, the household ID, person ID, and spatial category get added to a dictionary. This dictionary contains information only for work trips. Using the three dictionaries, a final dictionary was created to store the output.

The second function utilizes information collected from 50 runs to calculate the means, standard deviations, and coefficients of variation, storing it in a format like that from the first function.

3.3. Out-of-Home School Activities

The purpose of this variable is to calculate the amount of school trips that occur, organized by spatial category and student status.

Data is first extracted from the trips file. 4 lists are made using data from each row in the file: activities, household IDs, person IDs, as well as spatial categories. Using the persons file, a dictionary is created where the keys are the household IDs, and the value dictionary has people IDs as keys and their student status as values.

Another dictionary is created for all trips which were made to school. The keys are the household IDs, and the values are a dictionary where the keys are the person IDs and the values are the spatial categories. Using these two dictionaries, a final dictionary was created which counts the total amount of school trips that occurred in each spatial category for each student status.

The second function utilizes information collected from 50 runs to calculate the means, standard deviations, and coefficients of variation, storing it in a format similar to that from the first function.

3.4. Out-of-Home Shopping Activities & Out-of-Home Other Activities

The purpose of these variables is to identify and compare the amounts of shopping activities and other activities that occur across spatial categories. This has been accomplished through the usage of a function which takes in two parameters: the folder number as well as the file number, which are used to extract data from a single file. The trips.csv file is read, and all the destination spatial categories are stored in a list. Another list contains the activities that relate to those spatial categories. A dictionary is then created where the keys are the spatial categories and the values are a list of activities done in that spatial category. Finally, a dictionary is created with the keys being the spatial categories and the values being the number of “Market” activities that occurred in that spatial category.

For each spatial category, another function is utilized to compute the means, standard deviations, and coefficients of variation. The first function is executed for each of the 50 runs, and the results are saved in a list. To compute the mean for each spatial category, add the data and divide the sum by the total number of runs. Similarly, the standard deviation was determined using the standard deviation equation. Finally, by dividing the standard deviations by the means, the coefficient of variation was obtained. All of the outcomes were saved in dictionaries.

The statistical measures for the out-of-home other activities variable were computed in a similar manner, hence not discussed any further.

3.5. Trip Length Frequency Distributions

This variable calculates the trip lengths using each mode of transportation in each spatial category.

While looping through the file, the trip mode from each row was added to a list. The durations were computed by subtracting the departure time from the arrival time for each row in the dataset and adding them to another list. Finally, the weights for each row were added to a third list. This was utilized to readily access the corresponding length, trip mode, and weight as the indexes in the lists line up perfectly.

Two dictionaries were prepared. The first dictionary has spatial categories as keys while the value dictionary has trip modes as keys and total durations as values. The second dictionary is the same as the first, except that the nested value is the number of trips made with that specific mode in a spatial category. Dividing the first dictionary's values by the second dictionary's values gives the mean durations for one run.

To obtain the results for 50 runs, the second function called the first function. It calculated the new mean over the 50 runs using the mean durations for each of the 50 runs, which may be referred to as the "mean of the means." The standard deviation and coefficient of variation were computed using the same procedure.

3.6. Auto Ownership

The goal of this variable is to determine how many households in each spatial group own a given number of cars (0, 1, 2, 3, or 4).

A list of automobile quantities and a list of spatial categories was created in the first function. A for loop was used to iterate over each spatial category and a dictionary was created to hold the results. The keys are the spatial categories, and the value is a nested dictionary with keys representing the number of cars in a household and values representing the total number of households with that number of cars.

The second function collects data from 50 iterations of function 1 and stores it in a list. This list is then used to compute the mean, standard deviation, and coefficient of variation over all 50 runs.

3.7. Driver's License Ownership

This variable includes looking at adults over the age of 16 and computing the fraction of people in each spatial group who have a driver's license.

If a person is over the age of 16, they are added to a dictionary where the keys are the spatial categories and the values are the total number of people over 16 in that spatial category. A second dictionary was made with the keys being the spatial categories and the values are the number of people over 16 who have a license in that spatial category. The fraction of people who have a license in each spatial category was calculated by dividing the values of the second dictionary by the values of the first.

The second function collects and saves data from the first 50 repetitions of function 1 in a list. Using this list, the mean, standard deviation, and coefficient of variation for all 50 runs are computed.

3.8. Total Number of Trips

This variable counts the number of trips that occurred in a given day for each hour in each spatial category.

A list of trip start times, weights, and spatial categories was created. A for loop was implemented to go through the list of spatial categories. Two dictionaries were made: the first one which has spatial categories as keys and a list of start times as values. The keys in the second dictionary are spatial categories, and the values are a list of weights. An array of 24 elements, symbolizing 24 hours, was created. Each array member represents one hour from 12 a.m. to 12 a.m. the next day. Using the trip weights, the number of trips were added to the appropriate position in the array based on the trip start timings. This process was repeated for each spatial category and the results were saved as a dictionary.

3.9. Graphing

Depending on the number of factors involved in the graphs, some variables were graphed using different software than others. Graphs with two parameters were created in Excel, while graphs with three or more parameters were created in Python, specifically plotly.

Means, standard deviations, and coefficients of variation were kept in distinct data frames to construct 3+ parameter graphs. They were then transformed to csv files in the format shown in Figure 2.a. The file was then updated to make it easy for plotly to operate. Figure 2.b depicts this.

	VFH	WAT	Carpool	Auto	Passenger	DAT	Bicycle	Walk	Schoolbus	PAT	PET	RideShare
1	7.275733	35.17015	18.34861	9.39226	6.375642	59.83638	10.56199	22.62396	6.527427	53.12821	50.58183	6.233111
2	9.021092	42.78699	15.6176	11.31862	8.913124	52.57375	20.99068	25.42394	7.296914	47.14747	43.99099	7.750377
5	8.406243	49.54171	14.30871	10.95005	8.429236	54.79293	19.65885	23.98274	7.481718	48.74908	40.69389	7.587428
3	9.186337	54.40677	14.67347	11.63053	8.698457	59.57815	20.35813	24.66755	7.761018	50.78856	44.17916	7.869512
4	7.369708	52.26139	16.06586	11.77745	8.245761	56.84239	18.83872	23.59377	6.964736	51.1159	40.83684	7.231884
6	9.068201	59.19267	19.78503	15.65627	7.479675	84.09538	17.38844	26.16746	5.171806	66.55594	58.04574	6.769318
7	9.390679	73.04147	18.59746	14.42826	7.967975	88.57036	18.72402	25.41411	5.075404	67.74532	58.71976	7.701463
8	9.139433	58.04429	16.60896	12.8511	7.388879	77.10153	19.30552	26.6831	5.418677	65.70501	54.73671	7.501506
9	9.217712	60.39455	17.51402	14.20366	6.98803	83.66708	17.64192	27.61898	4.513867	63.68734	53.35514	7.274632
10	9.048731	40.1586	16.16085	12.72231	6.77399	93.23556	17.316	27.9025	6.206711	66.81808	72.07921	6.435446

Figure 2.a: csv file example made by data frame

	A	B	C
1	Mode	Spatial_Cate	Mean
2	VFH	1	7.27573332
3	WAT	1	35.1701541
4	Carpool	1	18.348614
5	Auto	1	9.39226027
6	Passenger	1	6.37564211
7	DAT	1	59.8363783
8	Bicycle	1	10.5619934
9	Walk	1	22.6239621
10	Schoolbus	1	6.52742658
11	PAT	1	53.1282143
12	PET	1	50.5818298
13	RideShare	1	6.23311062
14	VFH	2	9.02109221
15	WAT	2	42.7869891
16	Carpool	2	15.6176036
17	Auto	2	11.3186246
18	Passenger	2	8.91312409
19	DAT	2	52.5737547
20	Bicycle	2	20.9906829
21	Walk	2	25.4239396
22	Schoolbus	2	7.29691435
23	PAT	2	47.1474701
24	PET	2	43.9909931
25	RideShare	2	7.75037749
26	VFH	3	9.18633692
27	WAT	3	54.4067718
28	Carpool	3	14.6734675
29	Auto	3	11.6305286
30	Passenger	3	8.6984567
31	DAT	3	59.5781507
32	Bicycle	3	20.3581297
33	Walk	3	24.6675476
34	Schoolbus	3	7.76101826
35	PAT	3	50.7885591

durations_means_50_runs

Figure 2.b: csv file used for plotly graphs

The file is read and stored in another data frame. From there, a graph is generated using plotly.

4. Results

The graphs in this section show the means, standard deviations, and coefficients of variation for each variable. Variables with one parameter have only one version of each graph, but variables with two parameters have two versions of each graph. The means represent the averages of the 50 runs that were processed. The standard deviation is the average amount of variability in the data collection and indicates how much each data point deviates from the mean on average. A high standard deviation in a normal distribution suggests that values are generally distant from the mean, whereas a low standard deviation shows that values are grouped close to the mean. [5]. The standard deviation was calculated using the following equation:

$$\mathrm{SD} = \sqrt{\frac{\sum |\mathrm{x} - \mathrm{m}|\mathbf{2}}{N}} \quad [6]$$

where x is a certain value from the data set, μ is the mean, and N is the amount of data points.

Finally, the coefficient of variation is a statistical measure of the dispersion of data points in a data series around the mean. The coefficient of variation is the ratio of the standard deviation to the mean, and it can be computed using the following formula:

$$\mathrm{CV} = \frac{s}{m} \quad [1]$$

where s is the standard deviation and m is the mean.

4.1. Out-of-Home Work Activities

Because this was the only variable that was structured using three parameters, the graphs have been generated for each employment status and occupation sector combination. This section will only cover a handful of the graphs. The remainder may be accessed in the Appendix and evaluated using the same techniques explained below.

Figures 3.a and 3.b show the work activity means for a full time and part time employment status, and a general office / clerical occupation sector. Even though there are a variety of employment statuses (full-time, work at home full-time, work at home part-time, not employed, unknown), the results only include full time workers as well as part time. This is because trips to the workplace only occur for those two employment statuses. As expected, people who work full time make more trips to work than those working part time. The majority of full-time work trips occur in planning district 1 (spatial category 1), while the lowest are in Durham (spatial category 6). For part time workers, the majority of trips happen in Peel (spatial category 8) while the least occur in planning districts 7, 8, and 9 (spatial category 3).

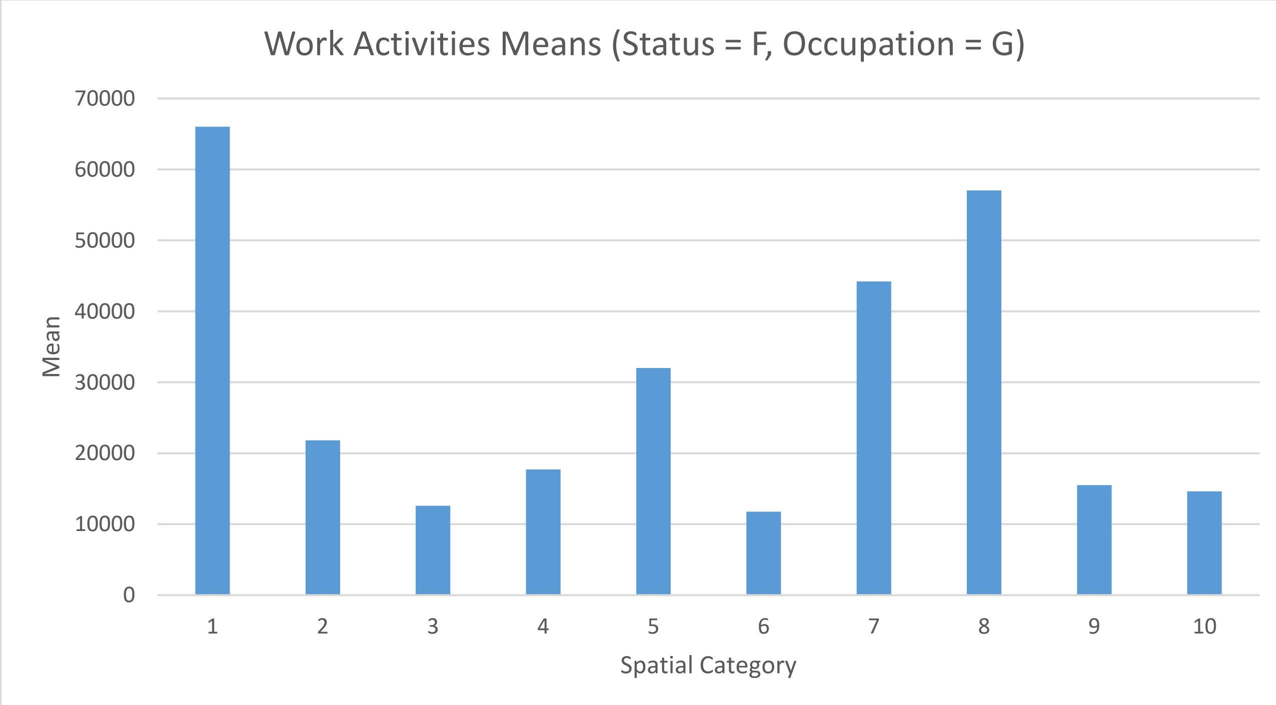


Figure 3.a: Graph for Mean Work Activities Per Spatial Category (Full-Time Status and General Office Occupation Sector)

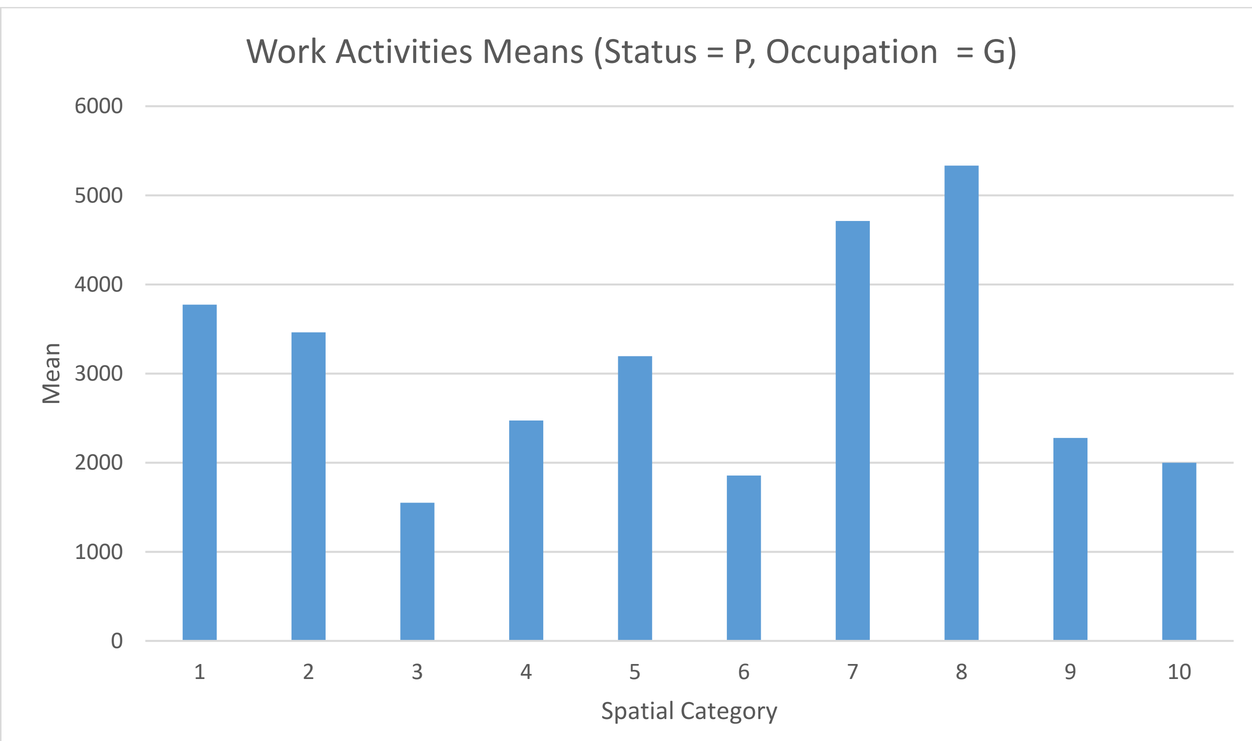


Figure 3.b: Graph for Mean Work Activities Per Spatial Category (Part-Time Status and General Office Occupation Sector)

Next, Figure 4.a and Figure 4.b show the work activity standard deviations for a full time and part time employment status, and a general office / clerical occupation sector. The standard deviations for a full-time employment status are in the hundreds while the values for part time workers are all below 70. This is plausible because the mean values for a full-time status were larger than for part time, indicating that the standard deviations would follow the same pattern.

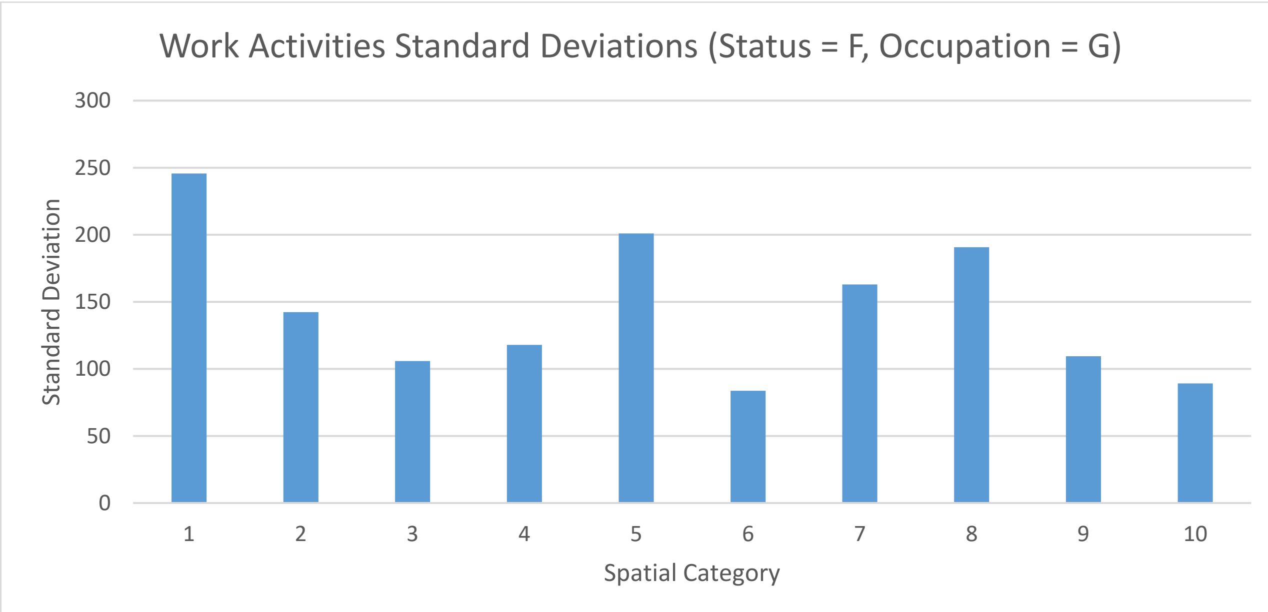


Figure 4.a: Graph for Work Activities Standard Deviations Per Spatial Category (Full-Time Status and General Office Occupation Sector)

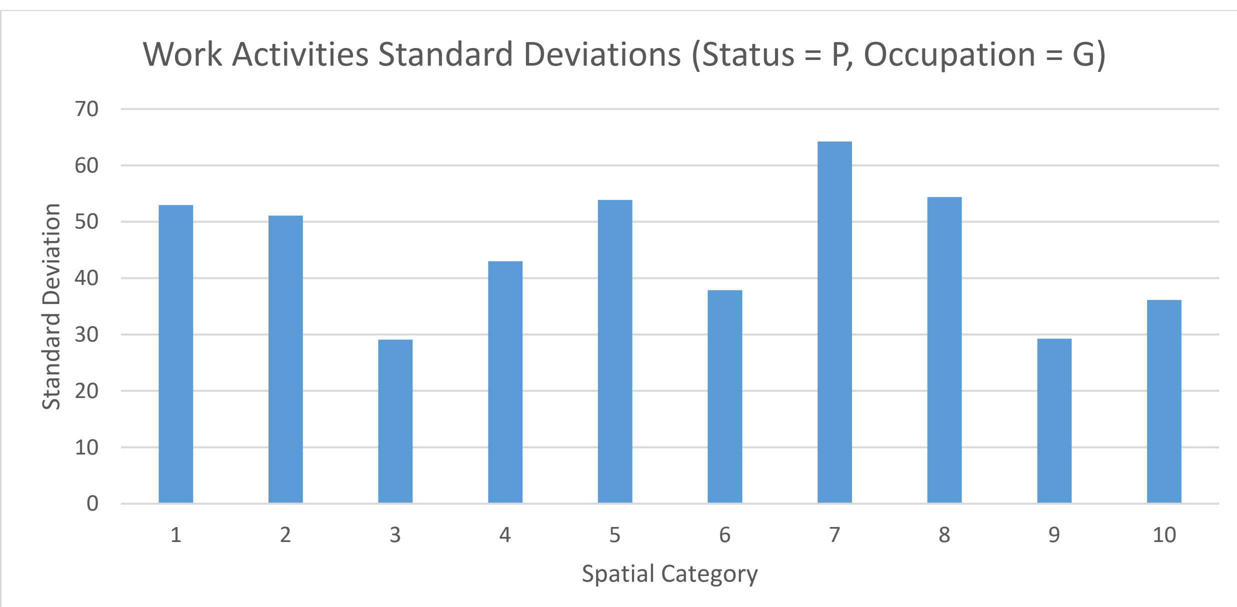


Figure 4.b: Graph for Work Activities Standard Deviations Per Spatial Category (Part-Time Status and General Office Occupation Sector)

Figures 5.a and 5.b depict the coefficients of variation for a full time and part time employment status, and a general office / clerical occupation sector. The values are extremely small, which is expected when comparing the standard deviations graphs to the means graphs. This outcome indicates that the model produces consistent results across the 50 runs. The coefficients of variation are smaller for full time trips than for part time trips, which could be because the mean values were smaller for part time trips, allowing room for more variability.

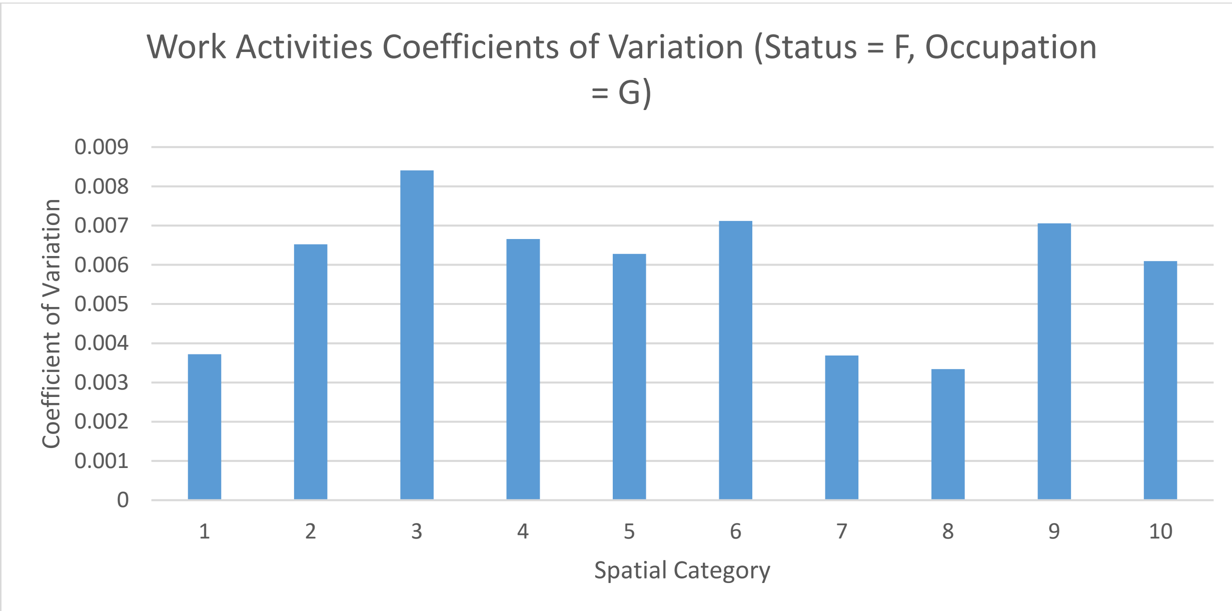


Figure 5.a: Graph for Work Activities Coefficients of Variation Per Spatial Category (Full-Time Status and General Office Occupation Sector)

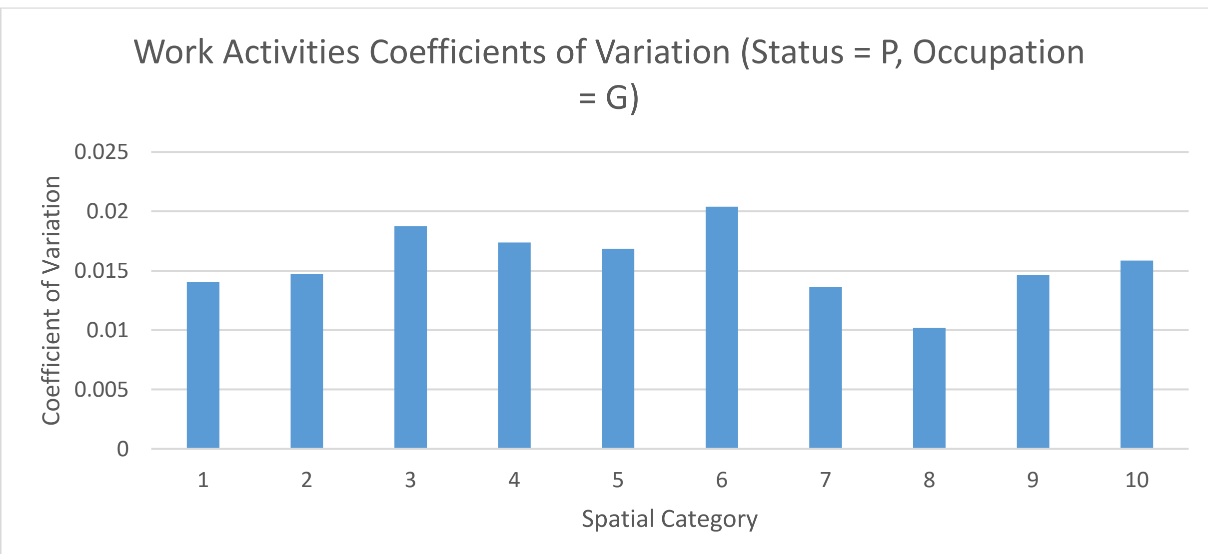


Figure 5.b: Graph for Work Activities Coefficients of Variation Per Spatial Category (Part-Time Status and General Office Occupation Sector)

4.2. Out-of-Home School Activities

Figures 6.a and 6.b illustrate two variations of the graphs for school activities. According to the data, those with a full-time student status make more journeys to school than students with a part-time student status. Peel (spatial category 8) has the highest number of school trips for kids with a full-time student status, whereas planning districts 7, 8, and 9 have the lowest number of school trips (spatial category 3). The statistics displayed in this graph are coherent given that the Peel District School Board is one of Canada's largest school boards, with over 150,000 students [3].

The findings vary slightly for students with a part-time student status, since the biggest quantity of school excursions are in planning district 1 (spatial category 1) and the fewest in Halton (spatial category 9).

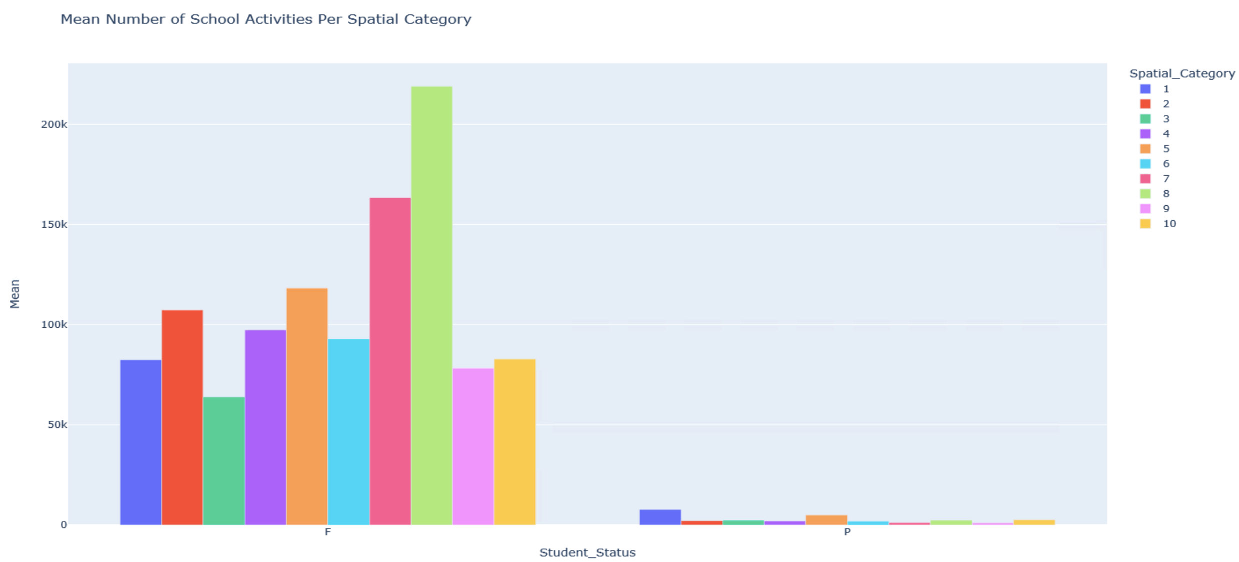


Figure 6.a: Version 1 of Graph for Mean School Trips Per Spatial Category

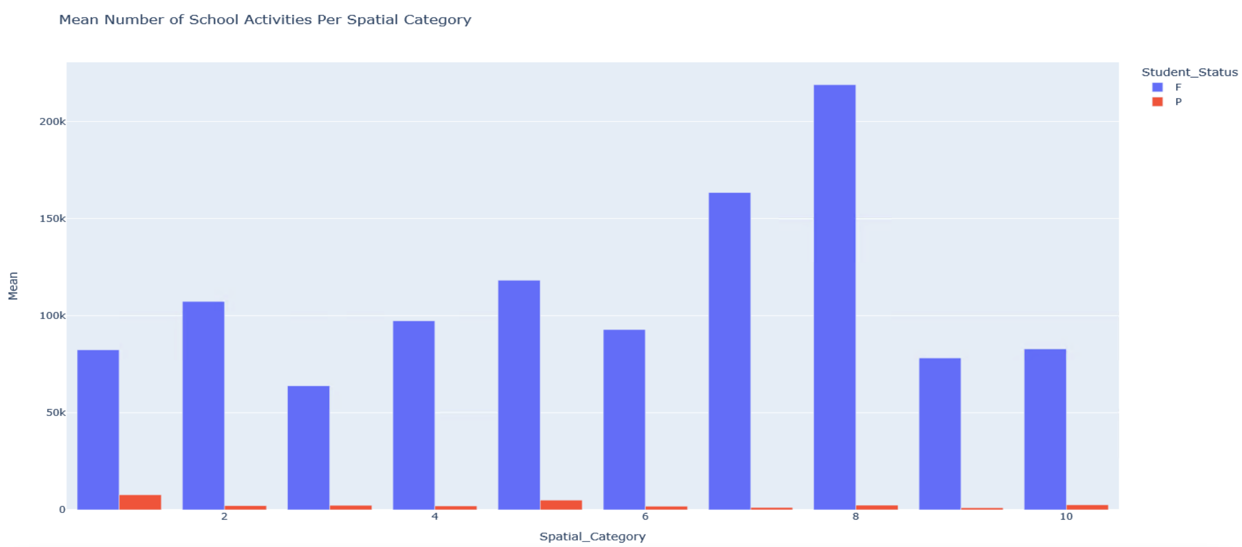


Figure 6.b: Version 2 of Graph for Mean School Trips Per Spatial Category

The standard deviations for school trips in each spatial group are illustrated in Figures 7.a and 7.b. It is apparent that the values are around 1000 times lower than the means, indicating that there is relatively minimal variability between the processed runs. The standard deviations for full-time students are higher than for part-time students which is owing to the difference in means between the two.

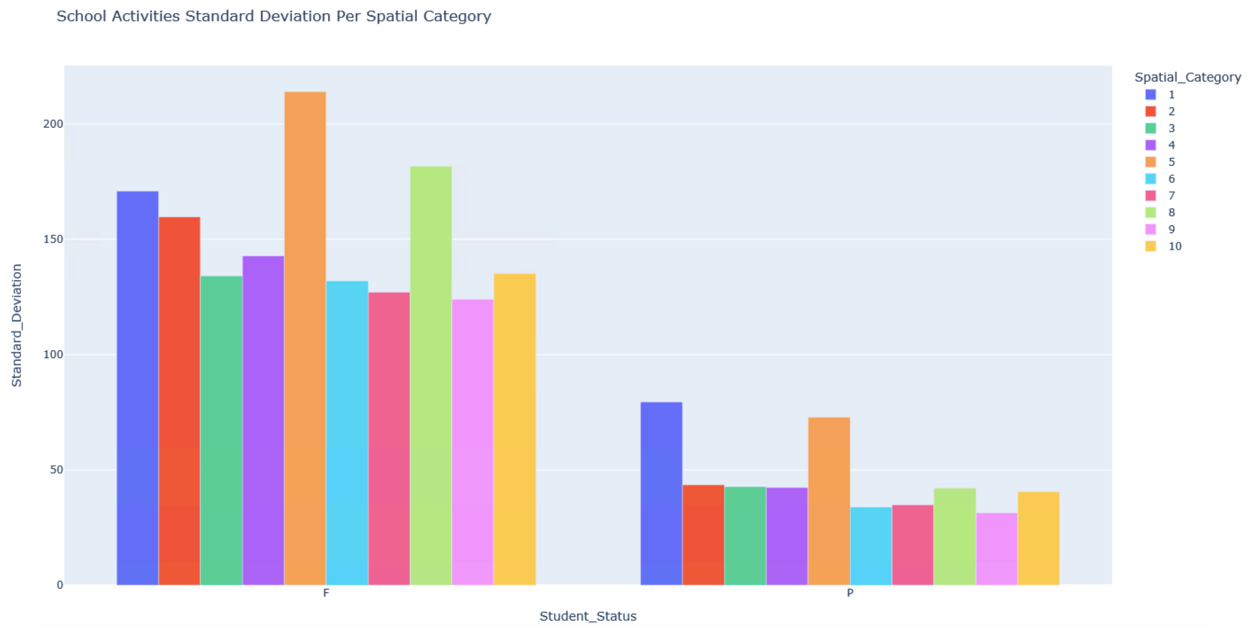


Figure7.a: Version 1 of Graph for School Trips Standard Deviations Per Spatial Category

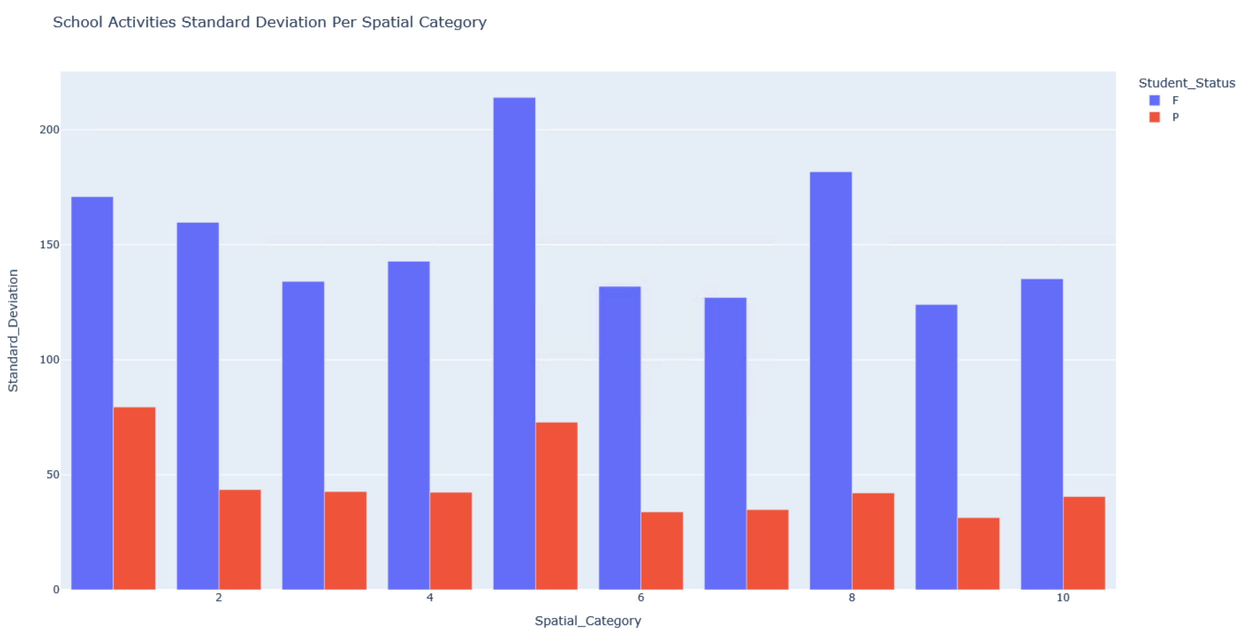


Figure 7.b: Version 2 of Graph for School Trips Standard Deviations Per Spatial Category

The coefficients of variation for school trips are shown in Figures 8.a and 8.b for each spatial group. In general, the values are relatively small; nevertheless, full-time students' values are substantially smaller than part-time students' values. The highest standard deviation is about 0.03 and the smallest is around 0.001.

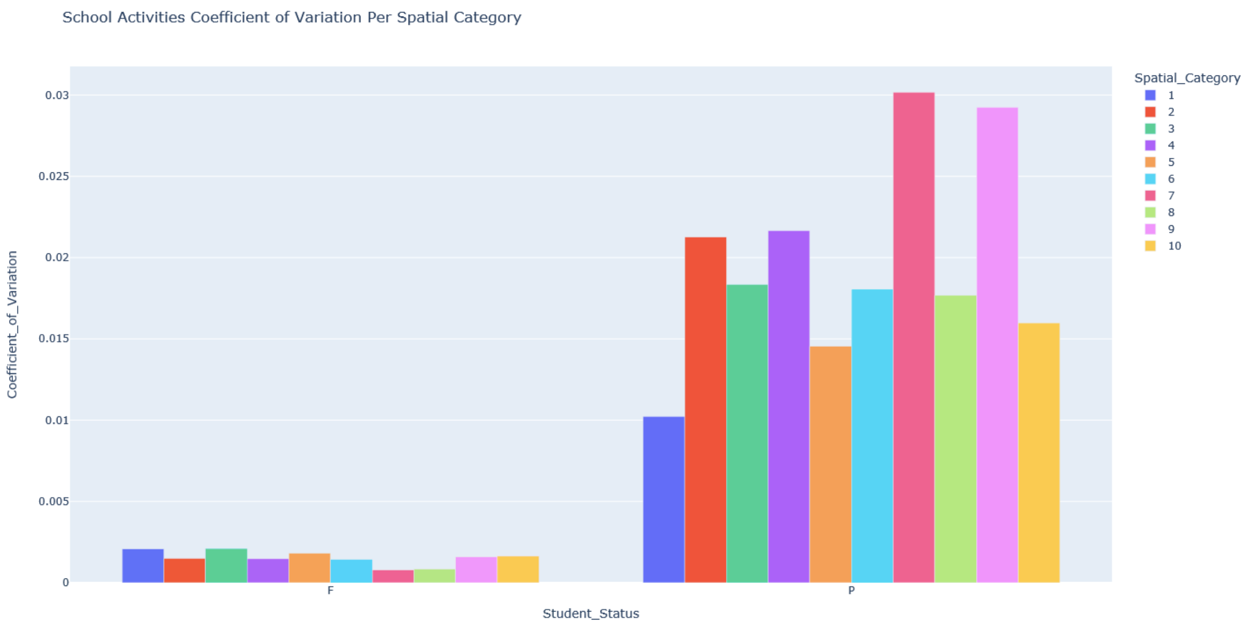


Figure 8.a: Version 1 of Graph for School Trips Coefficients of Variation Per Spatial Category

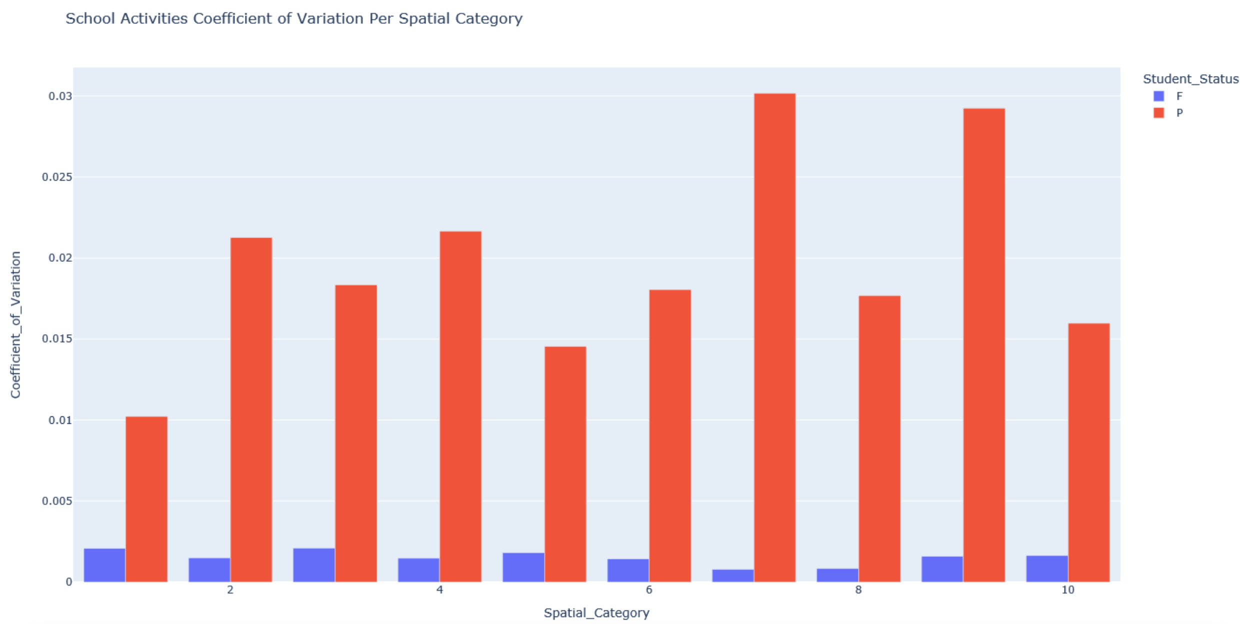


Figure 8.b: Version 2 of Graph for School Trips Coefficients of Variation Per Spatial Category

4.3. Out-of-Home Shopping Activities

According to the graph in Figure 9, the average number of shopping trips ranges from roughly 70000 to around 210000. The York region (spatial category 7) and the Durham region appear to have the largest number of shopping excursions (spatial category 6). The lowest number of retail trips occurs in planning district 1 (spatial category 1).

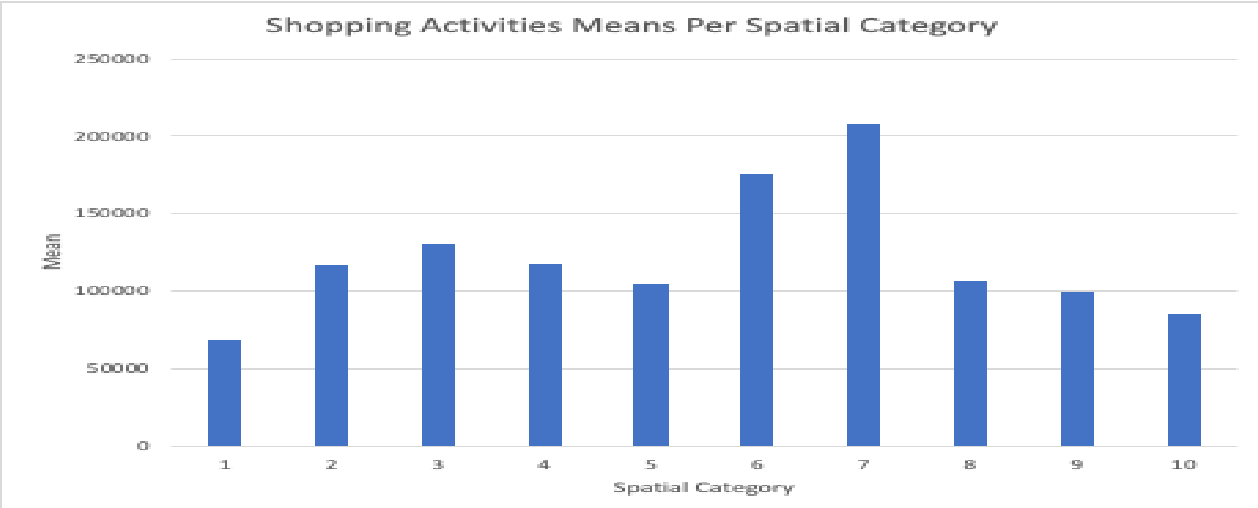


Figure 9: Graph for Mean Shopping Activities Per Spatial Category

Looking at Figure 10, it is evident that the standard deviation ranges approximately from 300 to 500, numbers that are significantly smaller than the mean values, indicating that the model produces quite consistent data throughout all 50 runs. The figure below indicates that the Durham region has the largest standard deviation (spatial category 6) while the Halton region has the lowest standard deviation (spatial category 9).

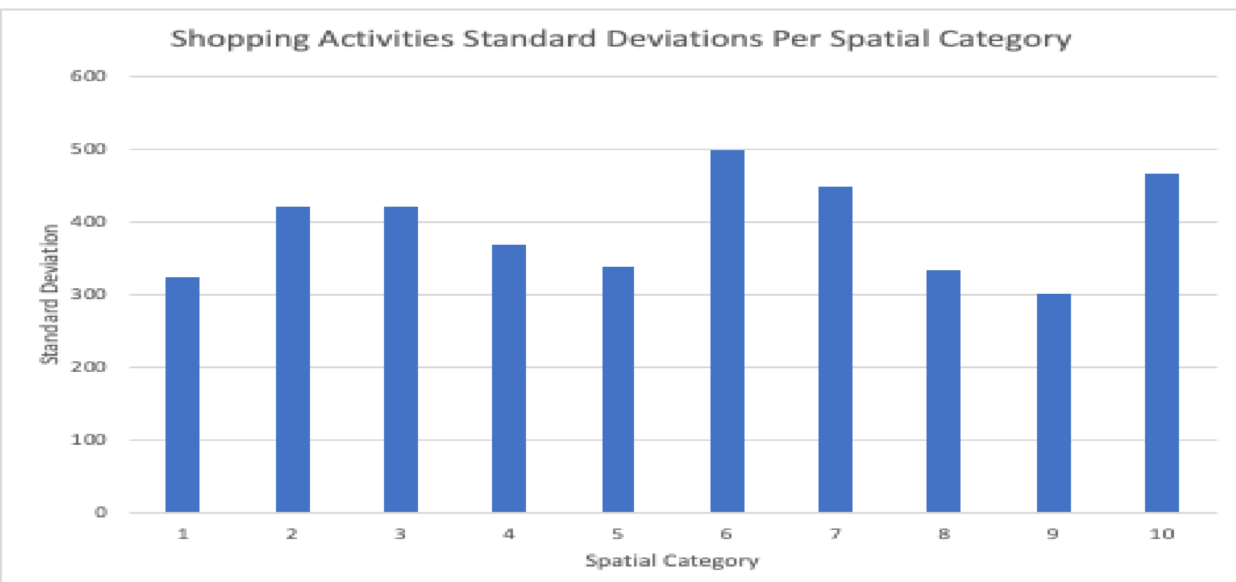


Figure 10: Graph for Shopping Activities Standard Deviations Per Spatial Category

Figure 11 depicts the coefficients of variation for each spatial category, and as can be seen, the values are very insignificant. They range between 0.002 to 0.0055, showing that data point dispersion is incredibly small. The York region (spatial category 7) has the smallest coefficient of variance, whereas the Hamilton region has the biggest (spatial category 10).

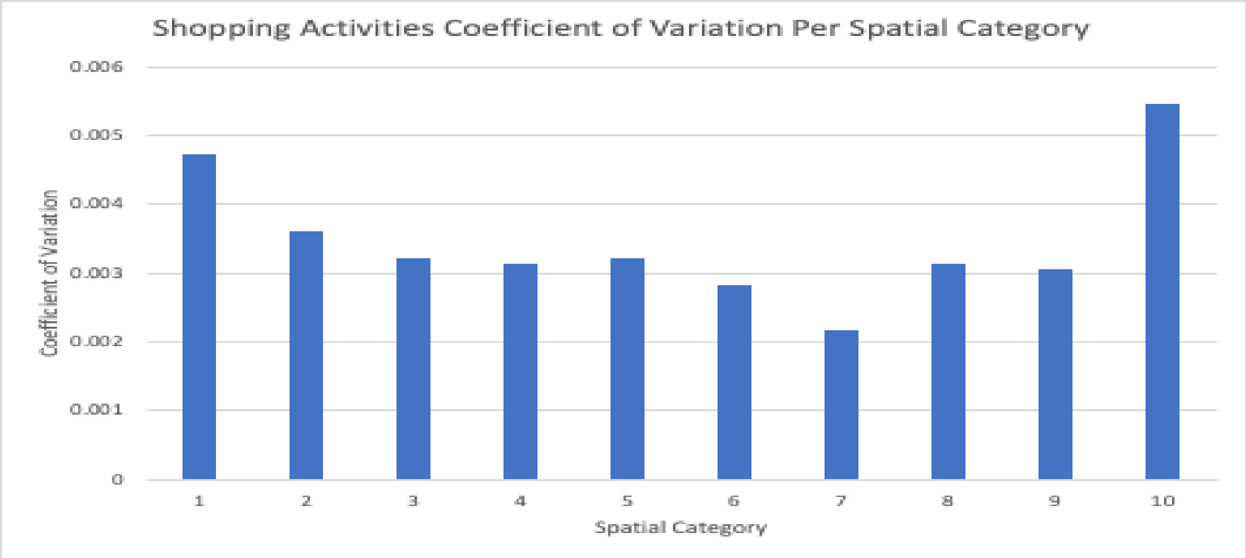


Figure 11: Graph for Shopping Activities Coefficients of Variation Per Spatial Category

4.4. Out-of-Home Other Activities

The graph in Figure 12 shows that the average number of other visits ranges from approximately 100000 to approximately 340000. These values are larger when compared to the means from the shopping activities graph, which makes sense given that "other" activities is a much broader group. Other activities tend to be the most prevalent in the York (spatial category 7) and Durham (spatial category 6) regions.

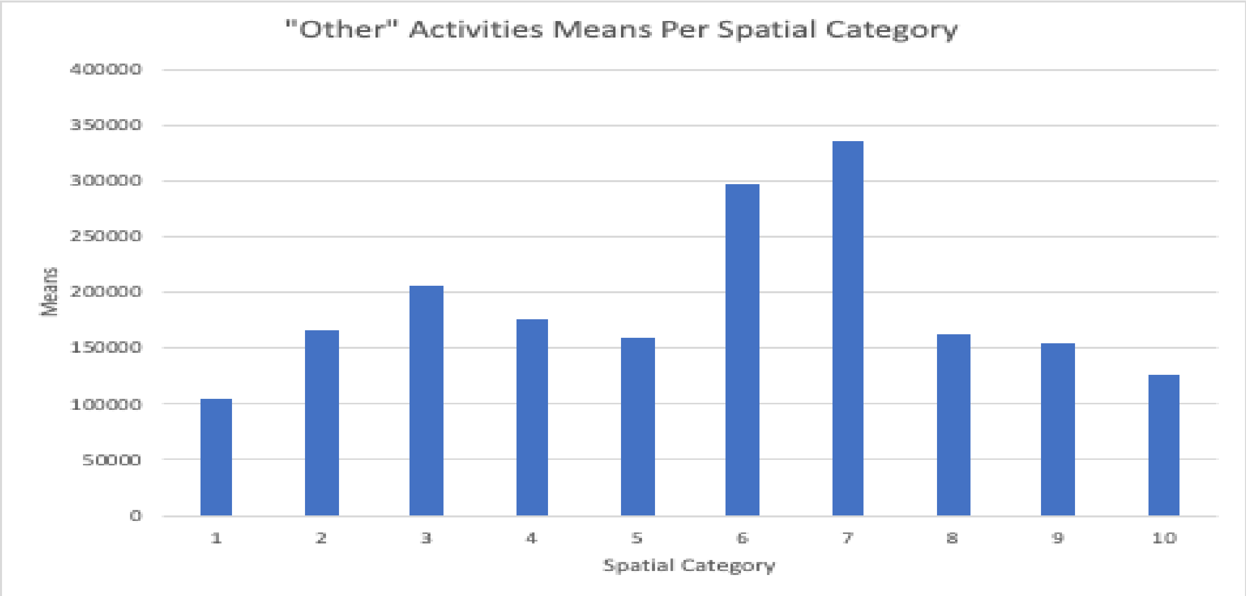


Figure 12: Graph for Mean Other Activities Per Spatial Category

In Figure 13, it is clear that the standard deviation ranges between 360 and 710, which are significantly lower than the mean values, indicating that the model produces reasonably continuous data over all 50 trials. These figures are marginally higher than those for shopping. Furthermore, in contrast to shopping activities, the Durham region (spatial category 6) has the highest standard deviation, while planning districts 13, 14, 15, and 16 have the lowest (spatial category 4).

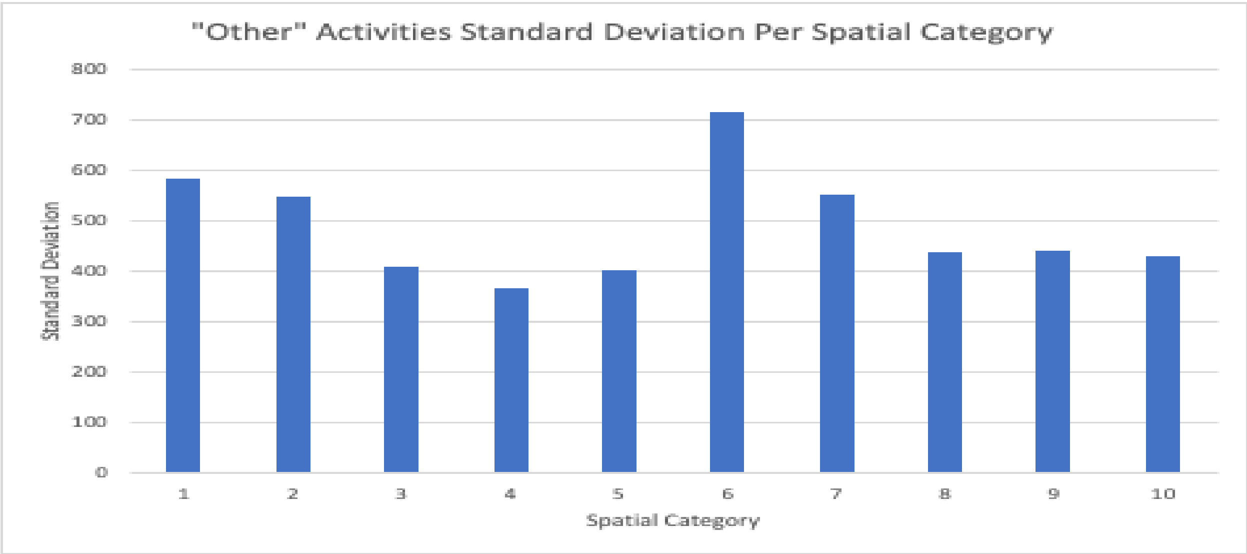


Figure 13: Graph for Other Activities Standard Deviations Per Spatial Category

The coefficients of variation for each spatial category are shown in Figure 14, and the values are exceedingly low. They are in the range of 0.0016 to 0.0055, suggesting that the data point dispersion is minor. The York region has the lowest coefficient of variance (spatial category 7), whereas planning district 1 has the greatest (spatial category 1).

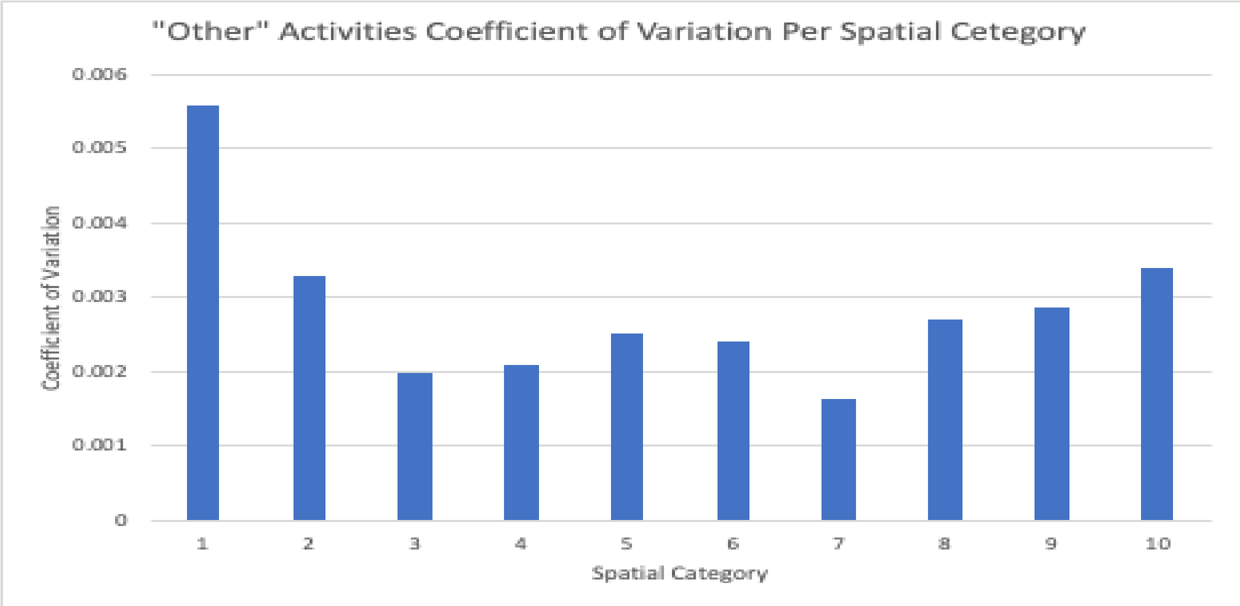


Figure 14: Graph for Other Activities Coefficients of Variation Per Spatial Category

4.5. Trip Length Frequency Distributions

The mean trip durations for each spatial group, segregated by trip mode, are shown in Figures 15.a and 15.b. It is evident that depending on the form of transportation, trip times may range from 5 minutes to 95 minutes. Trips on the school bus take less than 10 minutes, as expected, because schools are frequently near to people's homes. Walking, on the other hand, averages over 30 minutes across all spatial groups, which is another expected outcome. DAT and WAT are the modes of transportation that take the most time.

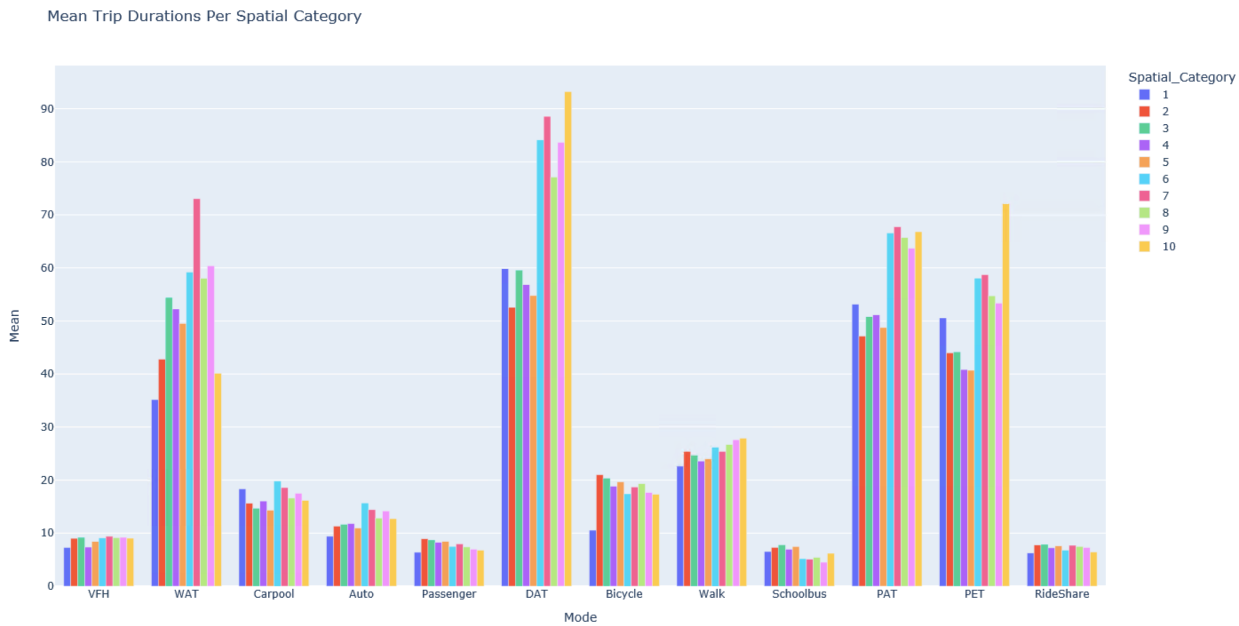


Figure 15.a: Version 1 of Graph for Mean Trip Durations Per Spatial Category

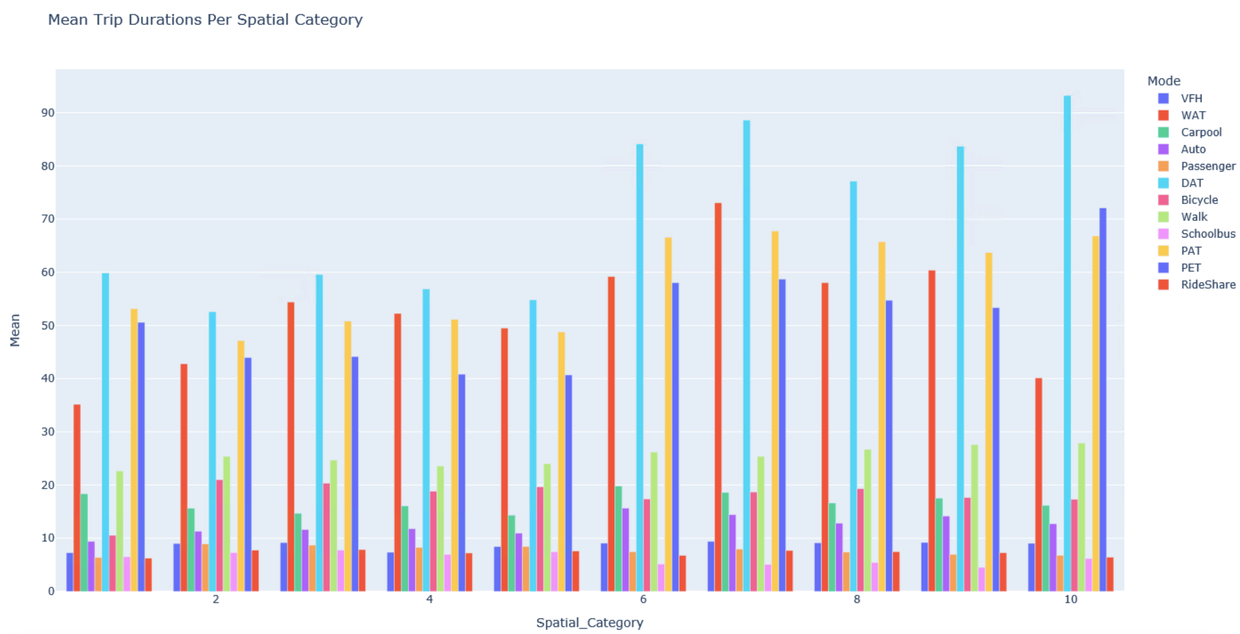


Figure 15.b: Version 2 of Graph for Mean Trip Durations Per Spatial Category

Figures 16.a and 16.b show the standard deviations for trip durations organized by spatial category and trip modes. The values range from approximately 0.05 to 0.75, which are significantly lower than all the trip durations. This illustrates that the model produces results that are extremely comparable throughout the 50 runs. One noticeable aspect about this graph is that the standard deviation for PET in spatial category 10 is significantly higher than the rest. This could signal a run in which that data point deviates significantly from the mean and should be studied more in the future.

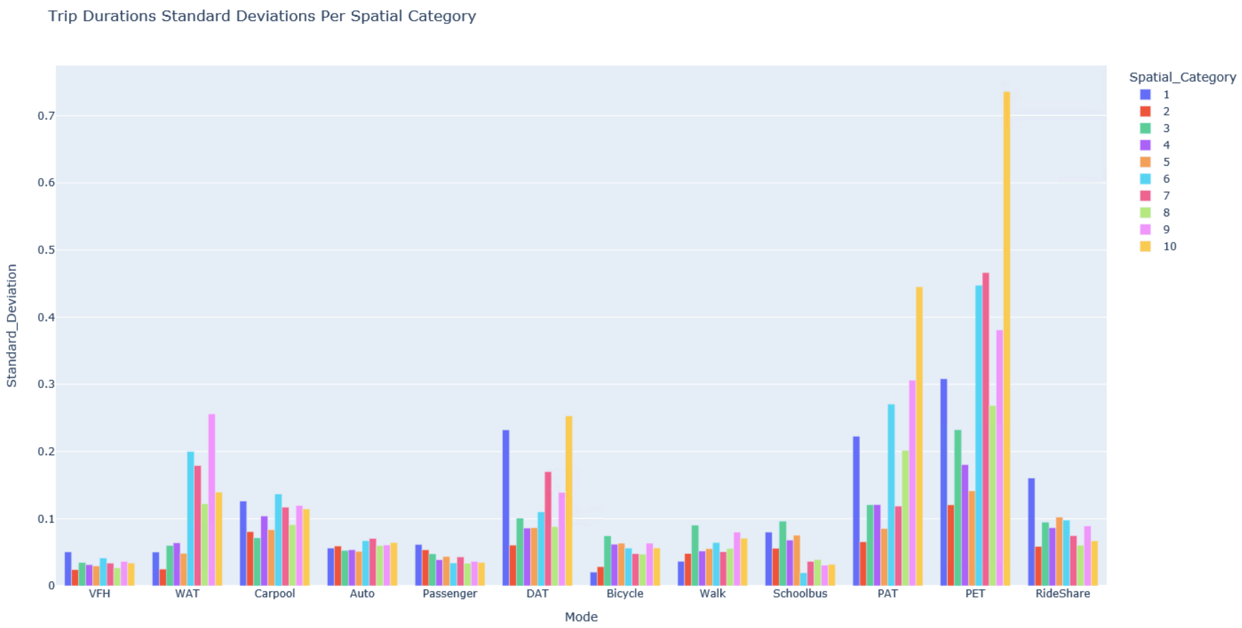


Figure 16.a: Version 1 of Graph for Standard Deviations of Trip Durations Per Spatial Category

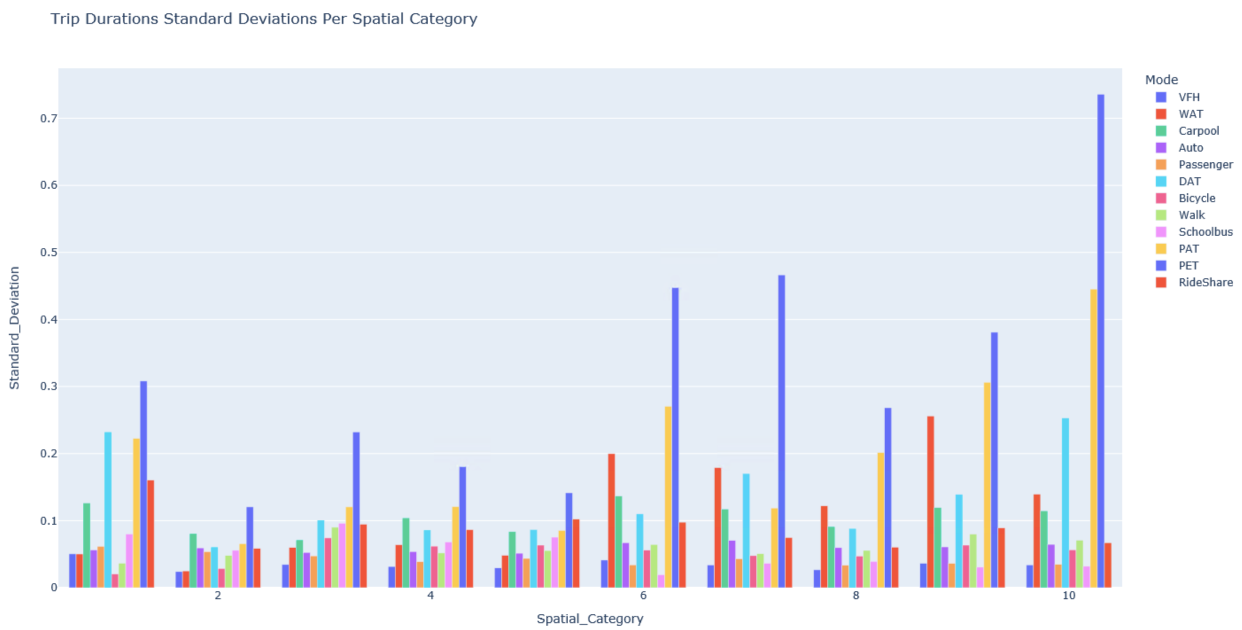


Figure 16.b: Version 2 of Graph for Standard Deviations of Trip Durations Per Spatial Category

Figures 17.a and 17.b depict two alternative types of trip duration coefficients of variation graphs. The values vary from about 0.001 to 0.03. This demonstrates, in addition to the minimal standard deviations, that the model delivers consistent results throughout 50 runs. In this case, unlike in the graph for standard deviations, RideShare from spatial category one gives a substantially greater coefficient than the other data, which could indicate a potential flaw in the model.

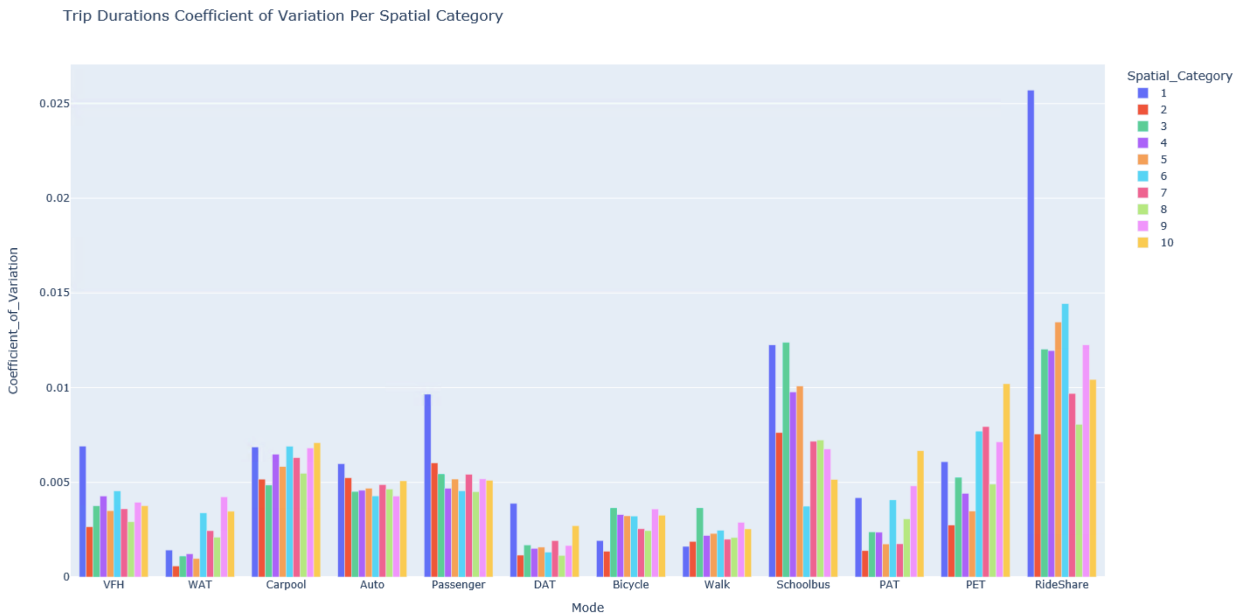


Figure 17.a: Version 1 of Graph for Coefficients of Variation of Trip Durations Per Spatial Category

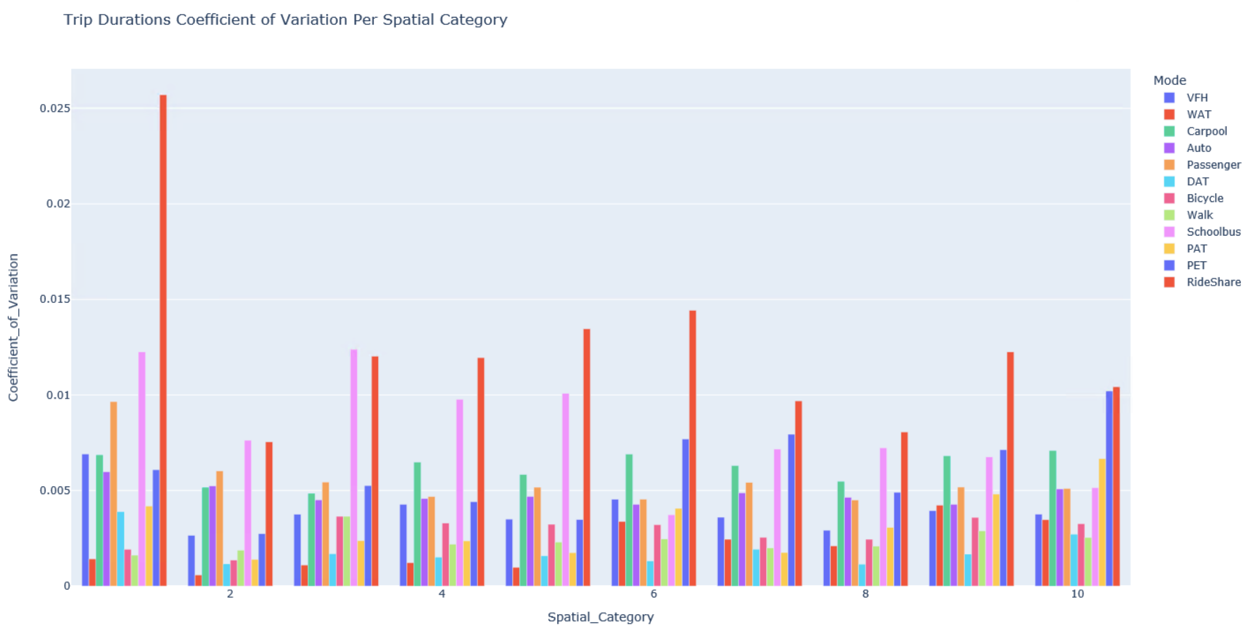


Figure 17.b: Version 2 of Graph for Coefficients of Variation of Trip Durations Per Spatial Category

4.6. Auto Ownership

Figures 18.a and 18.b show that the majority of persons in planning districts (spatial categories 1–5) have no cars or only one car for a single household. This might have been influenced by a number of factors, the first of which was the area's revenue. With the exception of spatial group 10, it is clear that the majority of households in higher spatial categories own two cars. Another point of concern is the proximity of the buildings. Having needs such as food stores, hospitals, pharmacies, and so on close to one's home would lessen the need to go by car. Traveling by automobile is undoubtedly the most convenient and practical way to get around in locations such as Durham, York, Peel, and Halton, as evidenced by the reported data.

The number of persons over the age of 16 living in each household is the final element that would influence the number of cars in a household in different regions. In the future, it would be particularly fascinating to examine the average number of adult household residents in each spatial category to determine if there is any association with the number of vehicles.

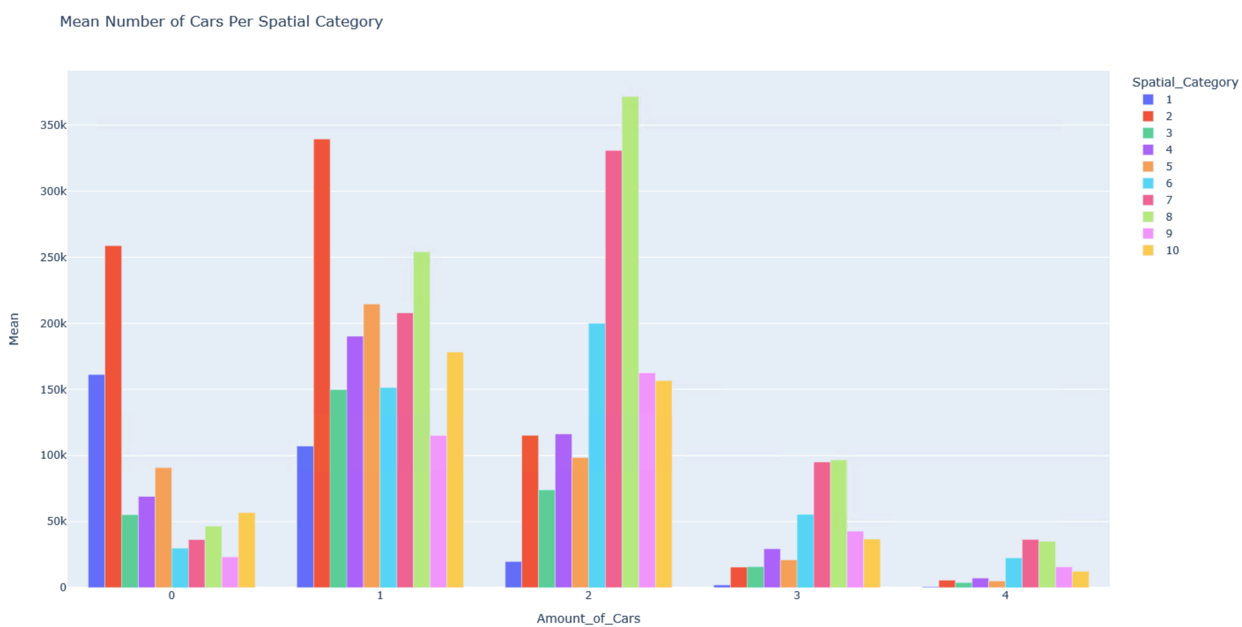


Figure 18.a: Version 1 of Graph for Mean Auto Ownerships Per Spatial Category

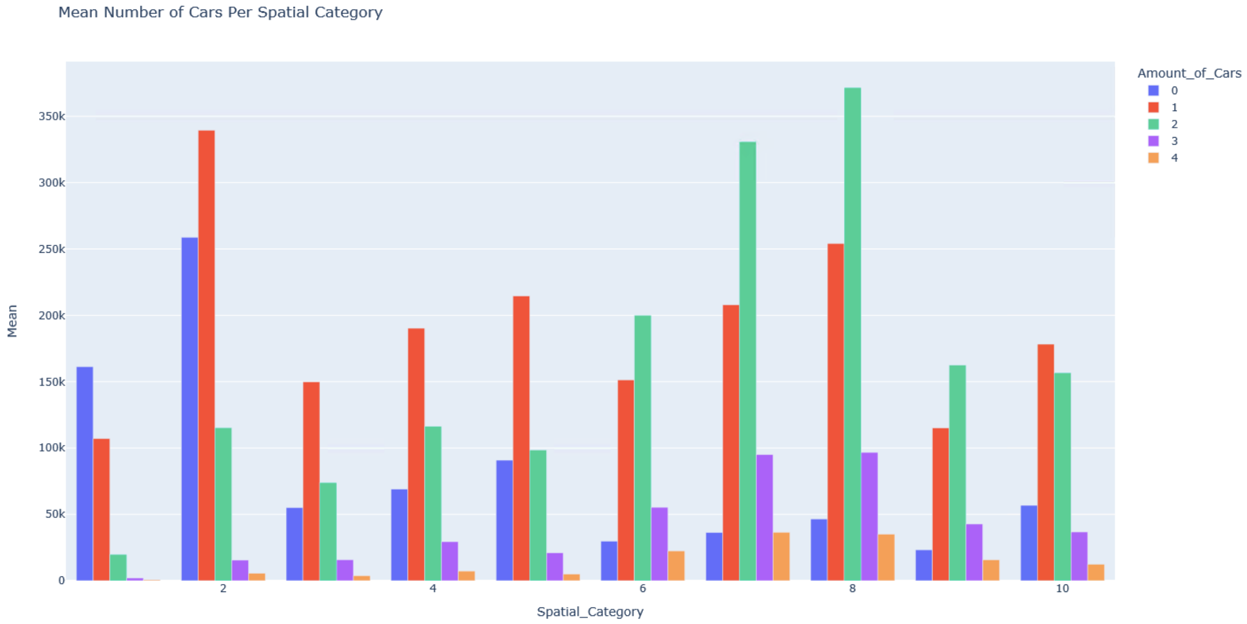


Figure 18.b: Version 2 of Graph for Mean Auto Ownerships Per Spatial Category

The standard deviations of auto ownership for each spatial category are displayed in Figures 19.a and 19.b respectively. The graphs for means and standard deviations are nearly identical in shape, but the standard deviation values are about twice as small as the means.

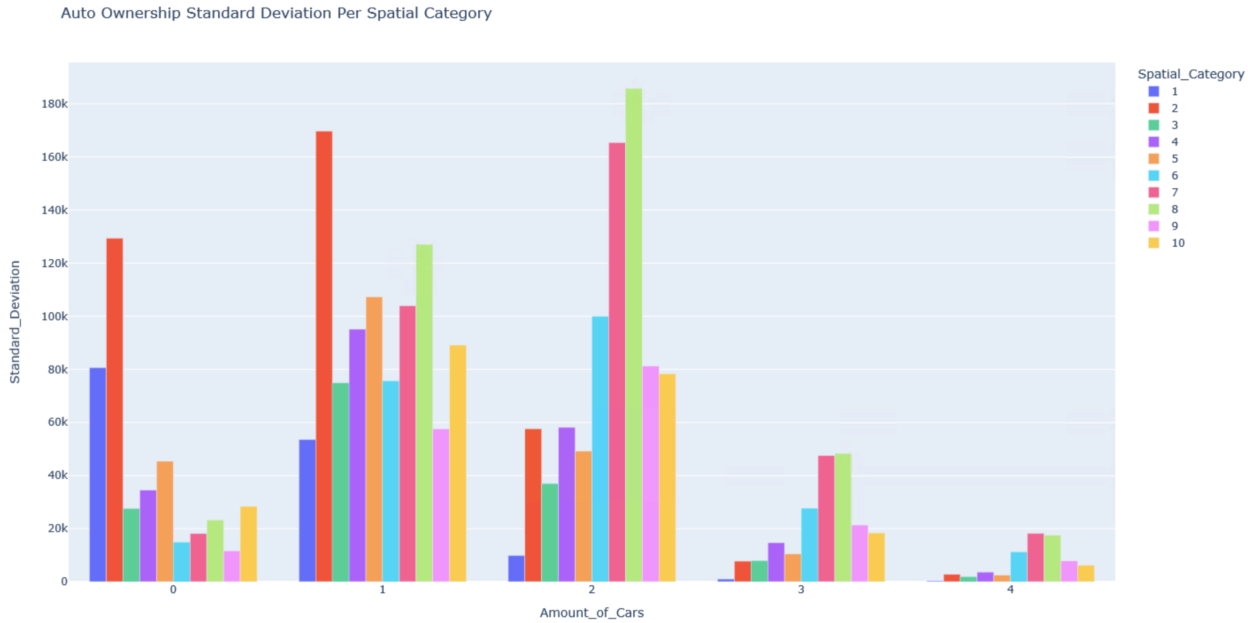


Figure 19.a: Version 1 of Graph for Standard Deviations of Auto Ownerships Per Spatial Category

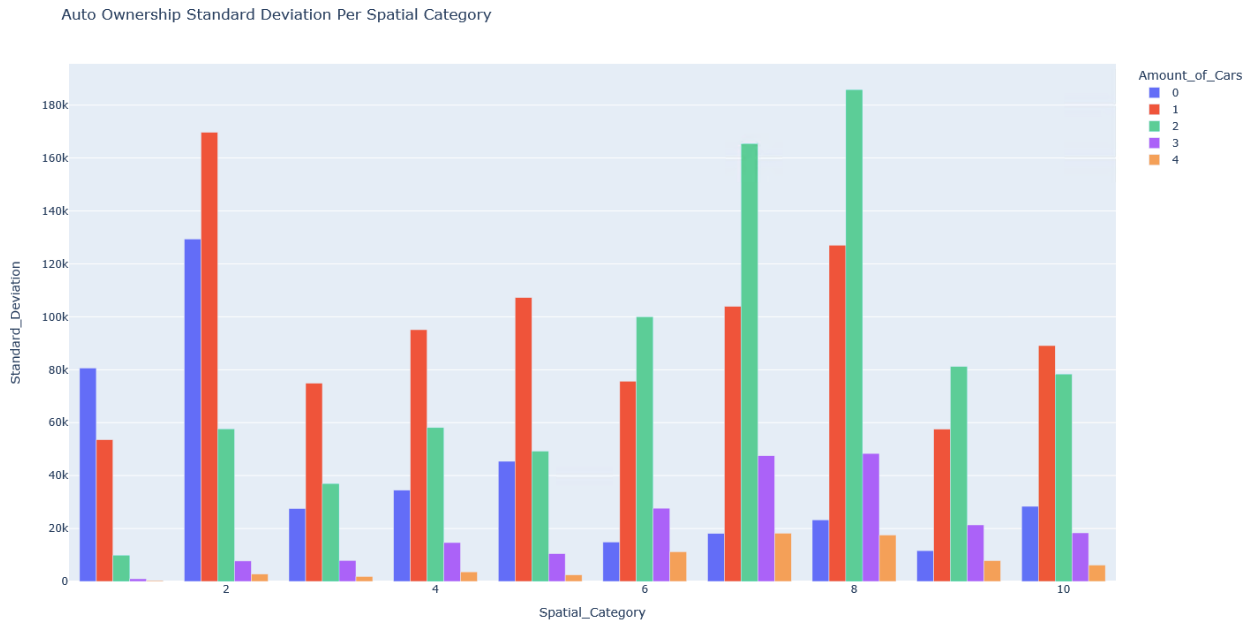


Figure 19.b: Version 2 of Graph for Standard Deviations of Auto Ownerships Per Spatial Category

Figures 20.a and 20.b support the standard deviation plots since each coefficient of variation is approximately 0.5. This is noteworthy because the values are higher than for other variables, and the coefficients of deviation are relatively steady when compared to other variables.

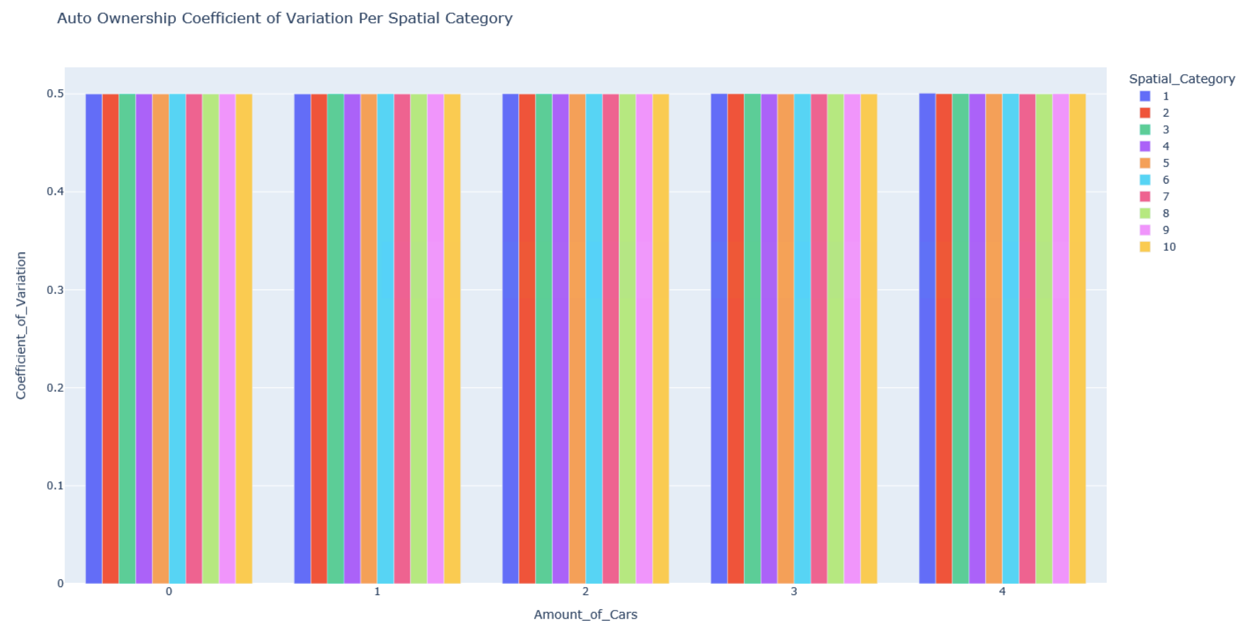


Figure 20.a: Version 1 of Graph for Coefficients of Variation of Auto Ownerships Per Spatial Category

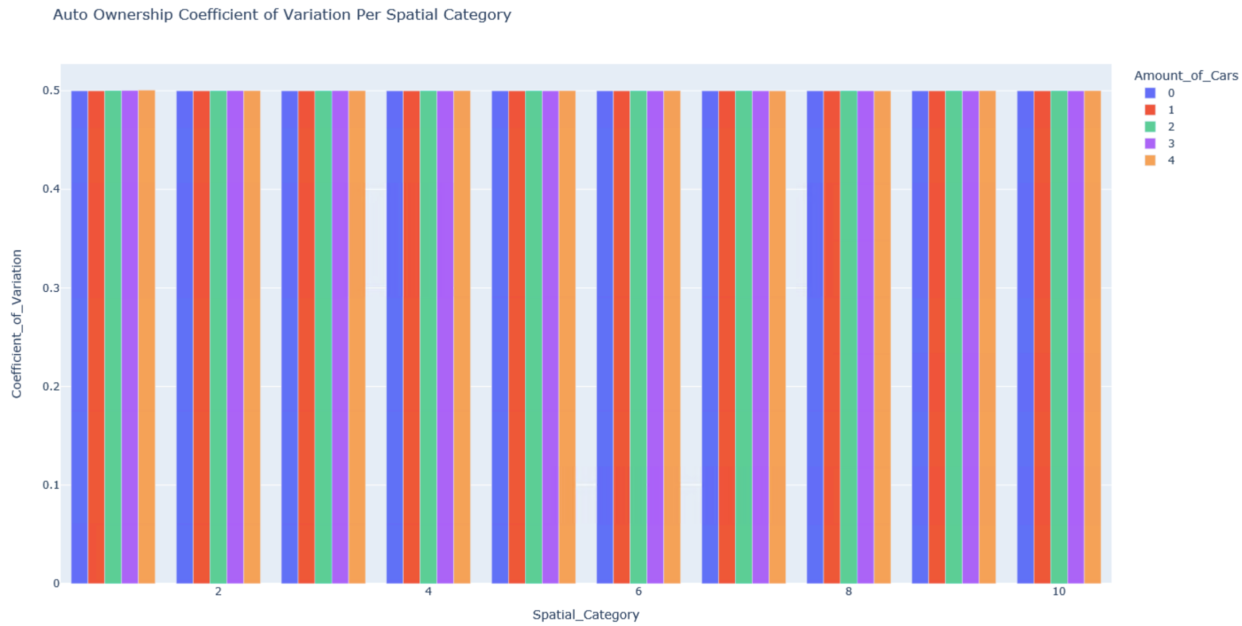


Figure 20.b: Version 2 of Graph for Coefficients of Variation of Auto Ownerships Per Spatial Category

4.7. Driver's License Ownership

According to the graph in Figure 21, the mean fraction of people over the age of 16 who have a driver's license fluctuates between 0.74 and 0.85. The majority of persons in the York (spatial category 7) and Halton (spatial category 9) regions have a driver's license. Planning districts 7, 8, and 9 have the fewest license holders (spatial category 3).

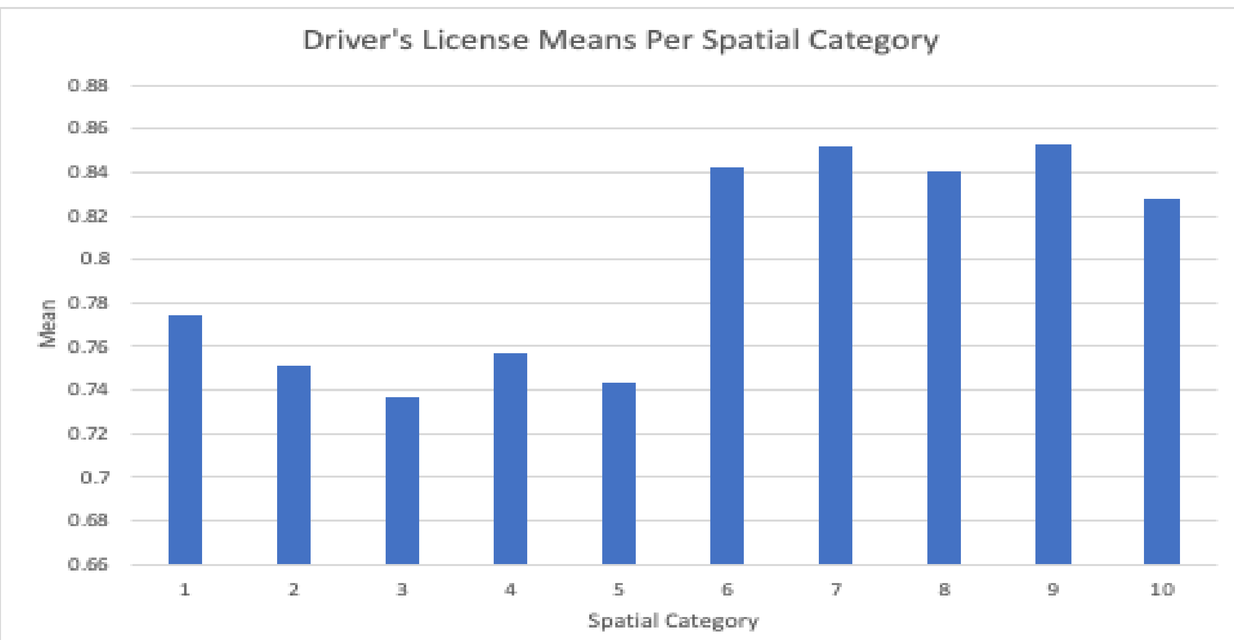


Figure 21: Graph for Mean Fraction of People With Driver's License Per Spatial Category

Looking at Figure 22, it is clear that the standard deviation ranges approximately from 300 to 500, numbers which are significantly smaller than the mean values, indicating that the model produces pretty accurate data across all 50 runs. The Durham region has the largest standard deviation (spatial category 6) while the Halton region has the lowest standard deviation (spatial category 9).

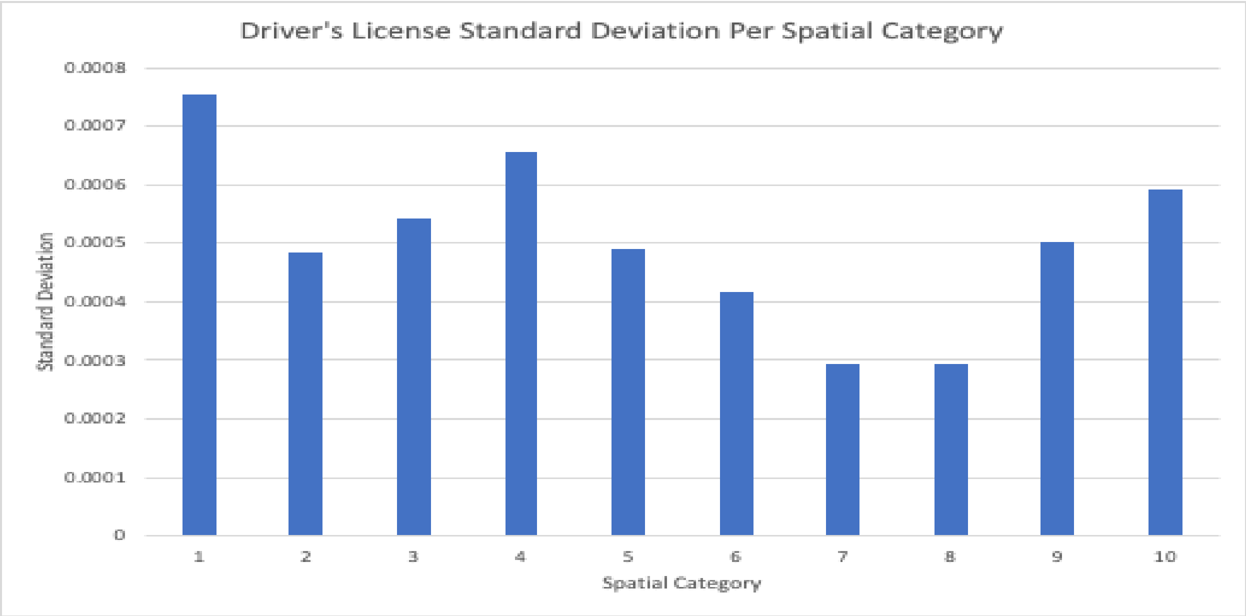


Figure 22: Graph for Driver's License Ownership Standard Deviations Per Spatial Category

Figure 23 exhibits the coefficients of variation for each spatial category, and as shown, the values are quite small. They vary from about 0.0004 to 0.001, implying that the data point dispersion is extremely low. The York region (spatial category 7) has the lowest coefficient of variation, whereas planning district 1 has the greatest (spatial category 1).

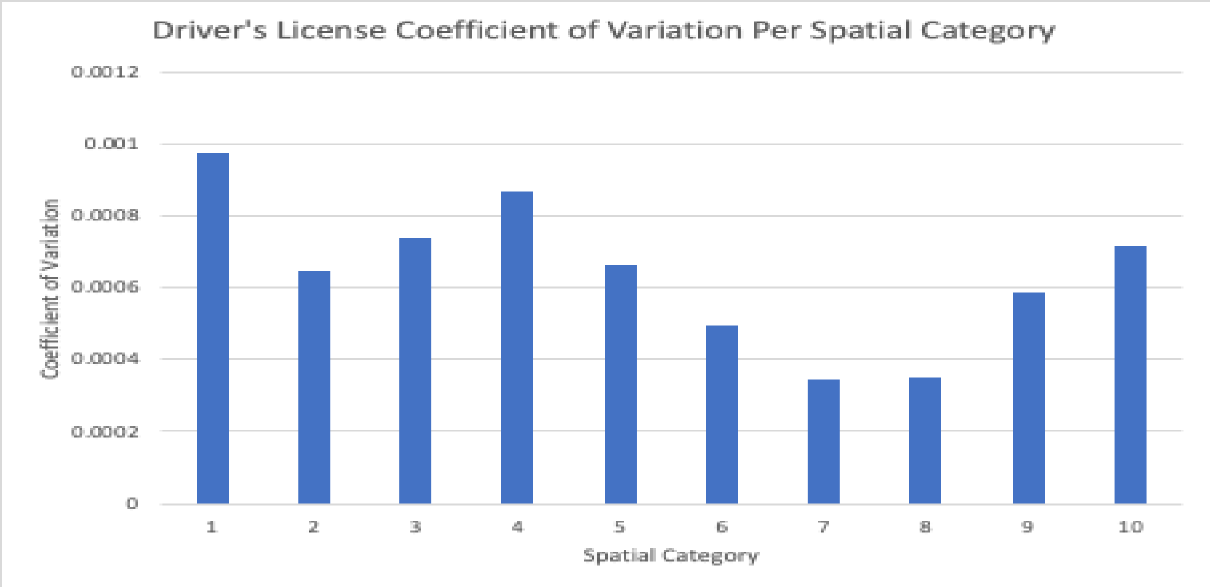


Figure 23: Graph for Driver's License Ownership Coefficients of Variation Per Spatial Category

4.8. Total Number of Trips

Figures 24.a and 24.b show the mean trip quantities for each spatial category for each hour from 12 a.m. to 12 a.m. the following day. Hour 1 is 12 a.m., and the time progresses by hour until hour 24 is 12 a.m. the next day. As expected, the least number of trips occur between 12 a.m. and 5 a.m., while the majority of travels occur between 7 a.m. and 9 a.m. and 4 p.m. to 6 p.m. This occurs as a result of individuals leaving for school/work in the morning and returning home in the evening, as illustrated by the model.

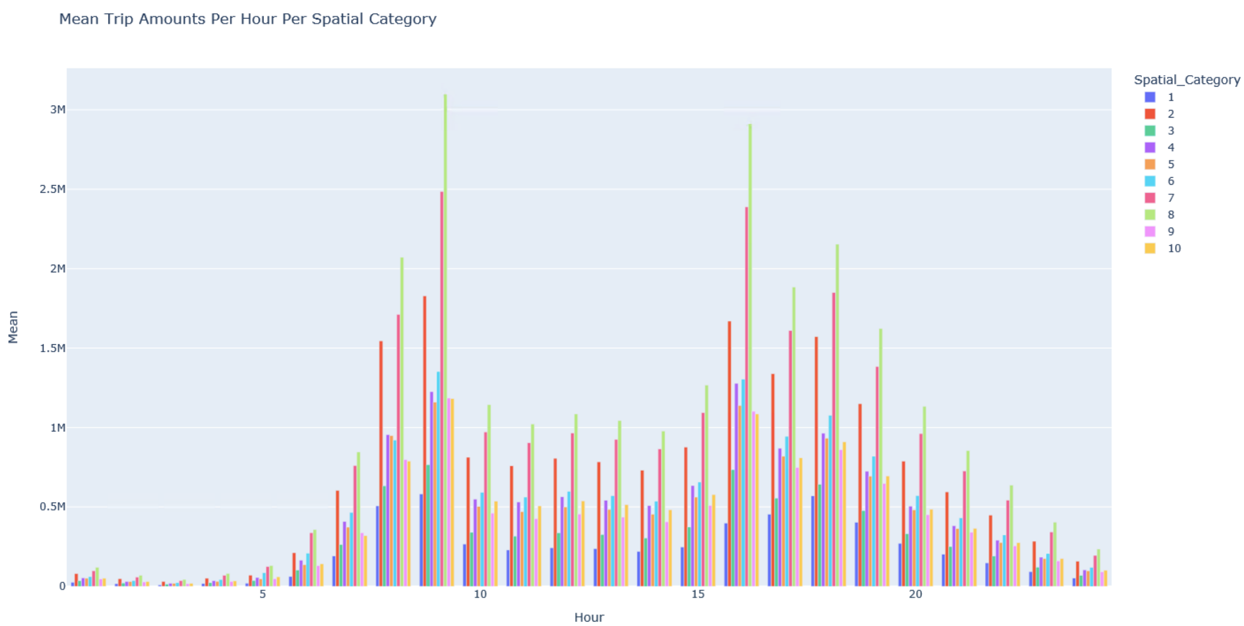


Figure 24.a: Version 1 of Graph for Mean Trip Amounts Per Spatial Category

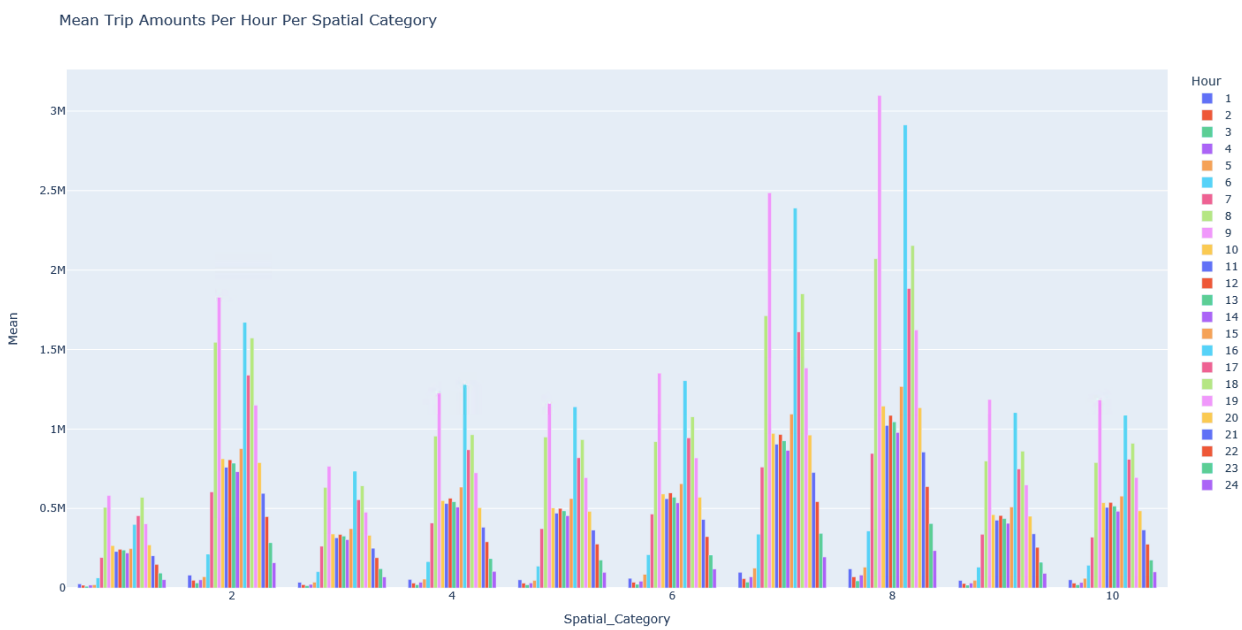


Figure 24.b: Version 2 of Graph for Mean Trip Amounts Per Spatial Category

The standard deviations for trip quantities are shown in Figures 25.a and 25.b. All of the values are below 7000, which is significantly less than the means, which are in the millions. This, once again, indicates that the model is consistent with the data for the first 50 runs.

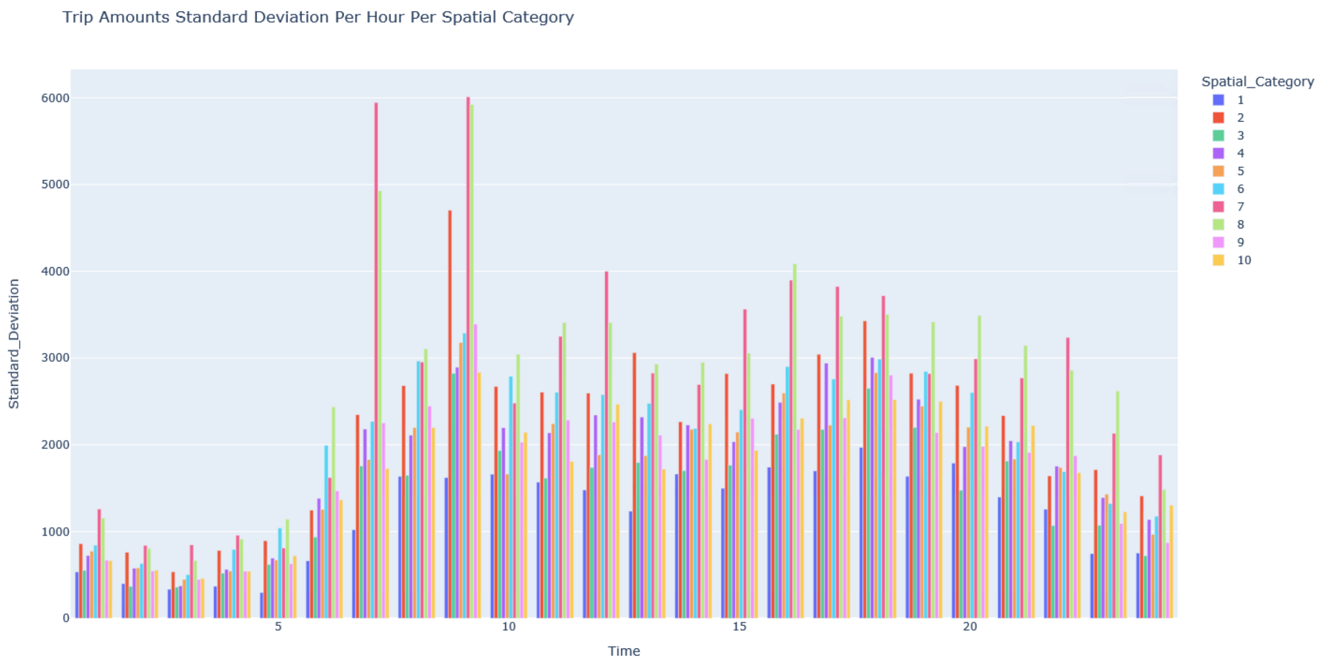


Figure 25.a: Version 1 of Graph for Trip Amounts Standard Deviations Per Spatial Category

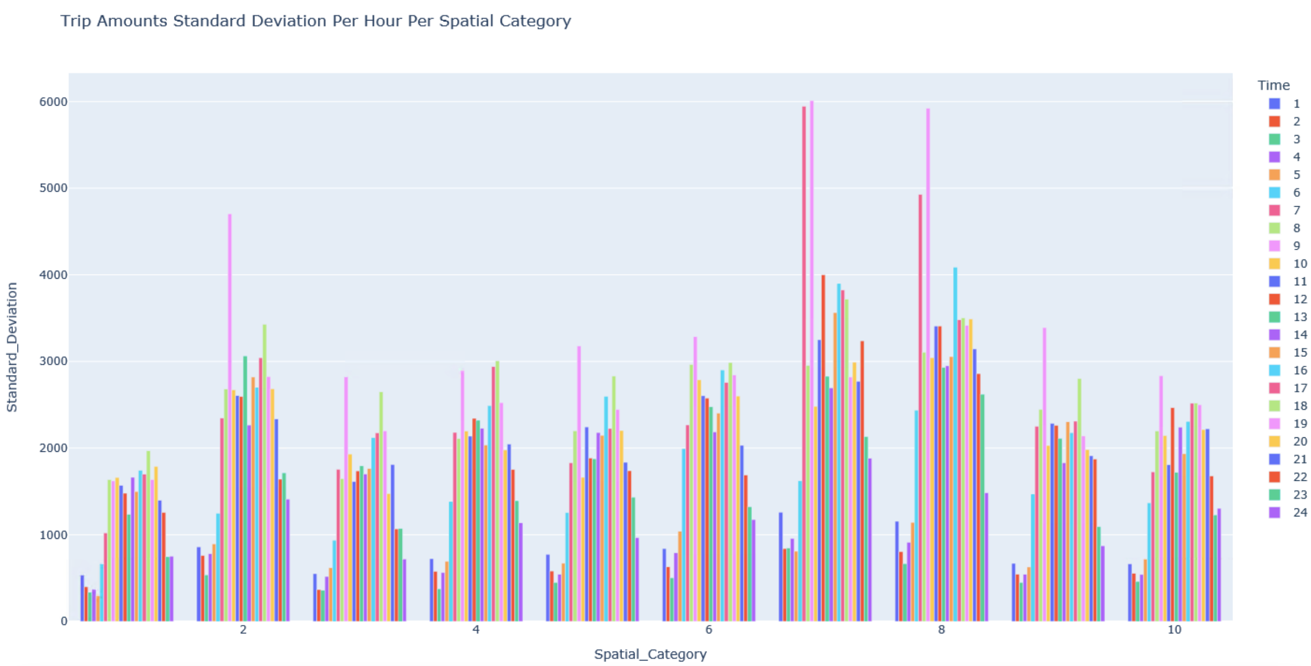


Figure 25.b: Version 2 of Graph for Trip Amounts Standard Deviations Per Spatial Category

The coefficients of variation for trip amounts in each spatial group are shown in Figures 26.a and 26.b. The numbers are relatively modest, but an interesting observation is that all of the coefficients are significantly higher in the morning than anywhere else. This is something that should be explored further into in the future.

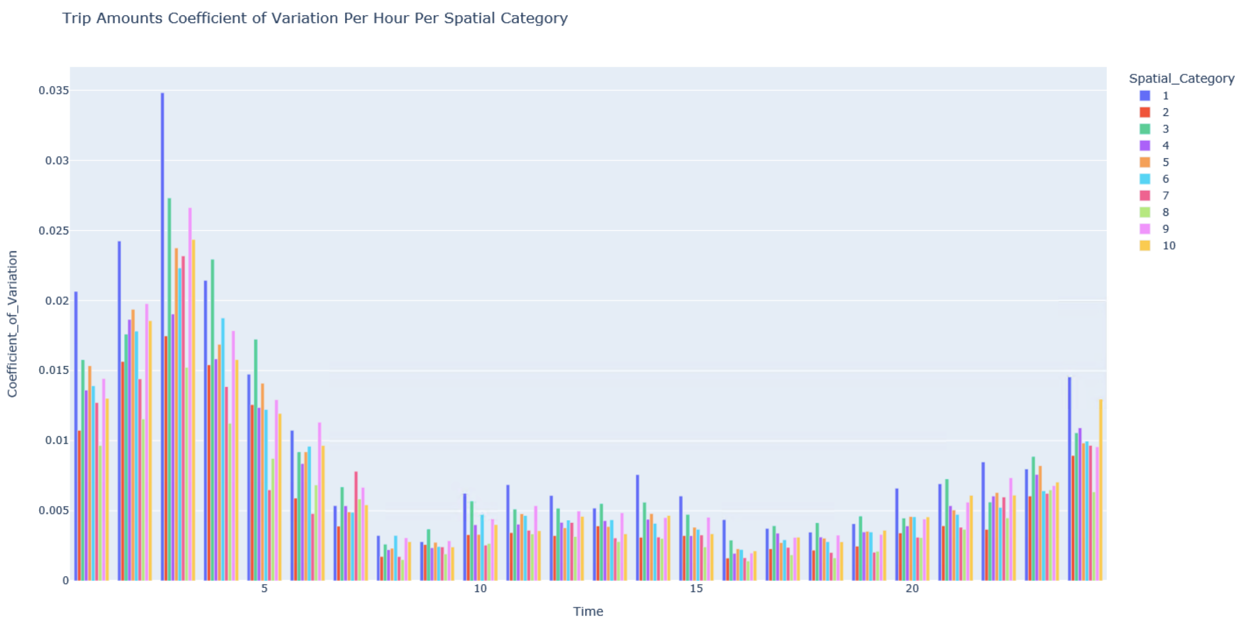


Figure 26.a: Version 1 of Graph for Trip Amounts Coefficients of Variation Per Spatial Category

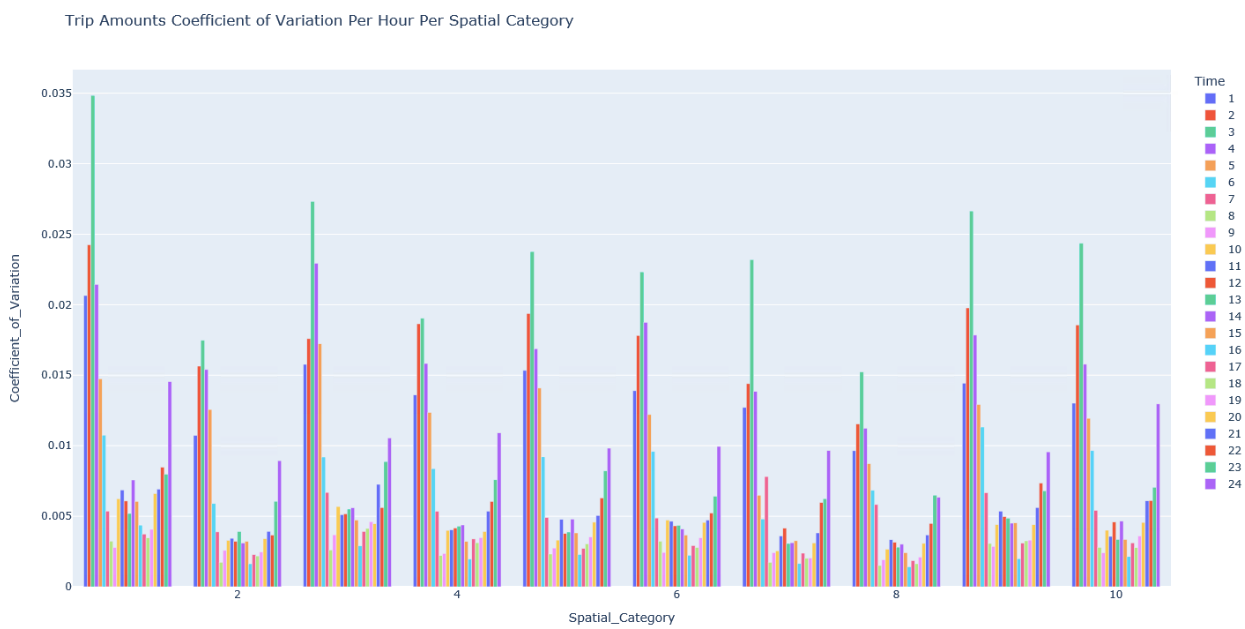


Figure 26.b: Version 2 of Graph for Trip Amounts Coefficients of Variation Per Spatial Category

5. Conclusions

To summarize, the purpose of this study was to examine the data generated by GTAModel across a vast quantity of runs and detect any aberrant outcomes, as well as to discover the likely causes of them. The initial experiment was designed to collect data from 1000 runs, however owing to a data collecting error, only the first 50 runs produced correct results. As a result, it was easier to execute the analysis on 50 runs rather than 1000.

Each variable was extracted using two functions, the first extracting the essential information and the second using the first to save the results of 50 runs. The mean, standard deviation, and coefficient of variation were calculated based on these findings.

The findings shown in the graphs above show that GTAModel delivers data with a high degree of accuracy across several runs, as evidenced by the coefficients of variation values.

Once the issue with data collection has been resolved, the most critical next step in this research would be to test the model for 1000 runs. 1000 runs is a considerably larger database to work with and would provide more accurate insight into the variability of the model's results. Some more parameters to consider in the future are as follows:

- Activity start time distributions by spatial category & purpose.
- Activity duration distributions by spatial category.
- PoRPoW & PoRPoS → linkage matrices (distance frequency distribution).
- Mode split by time-of-day by purpose by spatial categories.
- Average speed, volume on several main corridors by time-of-day.
- Transit ridership (boarding by major lines).
- Average tour length frequency by spatial categories.

6. References

[1] A. Hayes, “Coefficient of variation (cv) definition,” *Investopedia*, 22-Aug-2021. [Online]. Available: https://www.investopedia.com/terms/c/coefficientofvariation.asp. [Accessed: 22-Aug-2021].

[2] B. Ashby, “TTS 2016 Data Guide,” *transportation tomorrow survey 2016*, 2016. [Online]. Available: http://dmg.utoronto.ca/pdf/tts/2016/2016TTS_DataGuide.pdf.

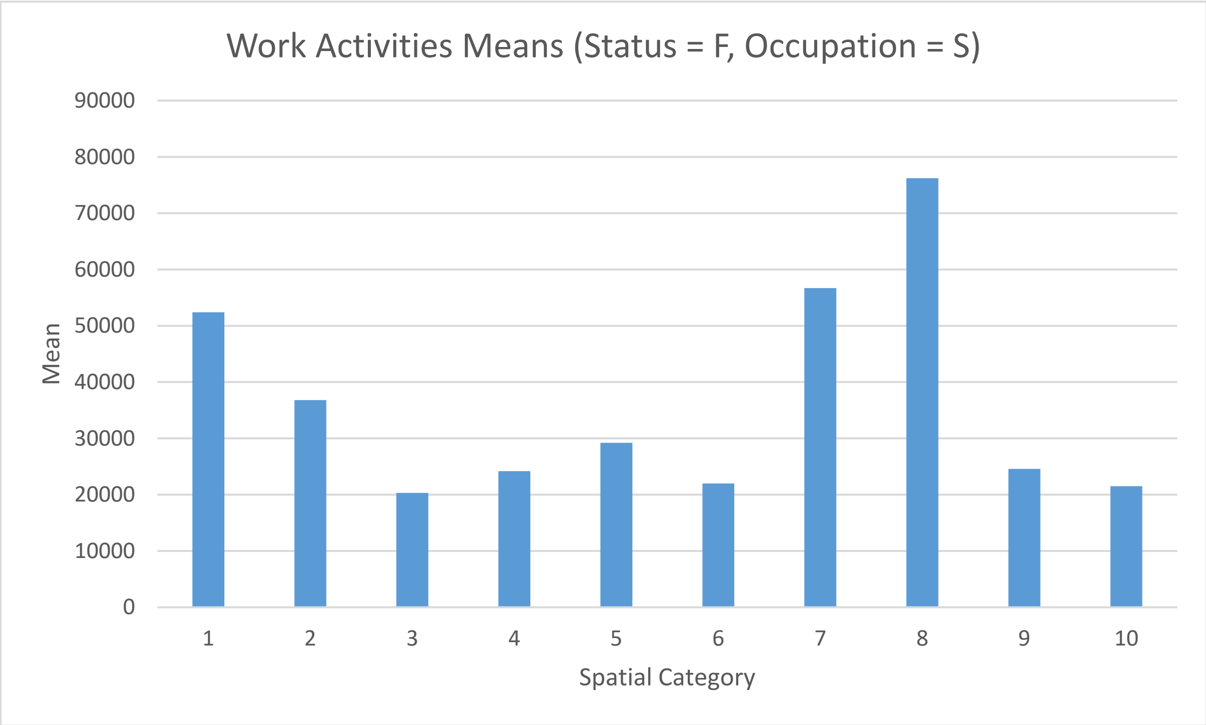
[3] “Doing business with the Peel District School Board,” *Bid opportunities*. [Online]. Available: https://www.biddingo.com/*.main?toPage=StBO.jsp&searchType=O&bidOrgId=11000116&selectedOrgId=11000116&ownerName=Peel+District+School+Board. [Accessed: 23-Aug-2021].

[4] “Microsim,” *Travel Modelling Group Documentation*. [Online]. Available: http://tmg.utoronto.ca/doc/1.6/gtamodel/user_guide/file_formats/microsim.html.

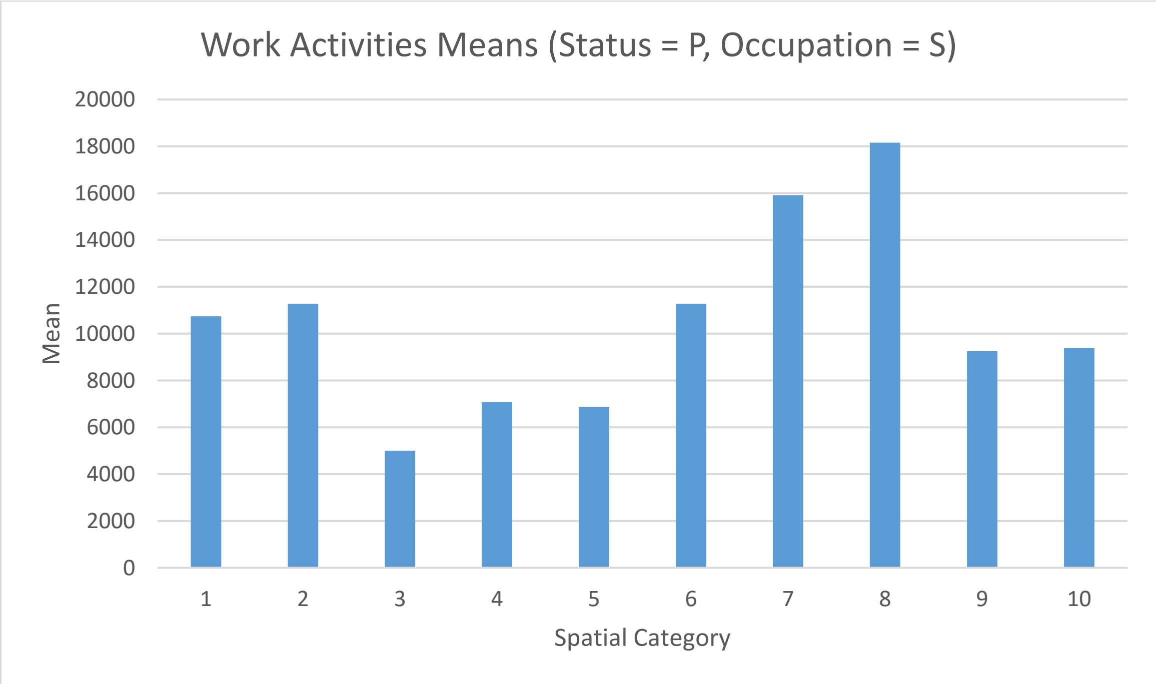
[5] Probability Book: J. L. Devore, *Probability and Statistics for Engineering and the Sciences*, Eighth. Brooks/Cole, Cengage Learning, 2012.

[6] “Standard deviation: Calculating step by step (article),” *Khan Academy*. [Online]. Available: https://www.khanacademy.org/math/statistics-probability/summarizing-quantitative-data/variance-standard-deviation-population/a/calculating-standard-deviation-step-by-step. [Accessed: 22-Aug-2021].

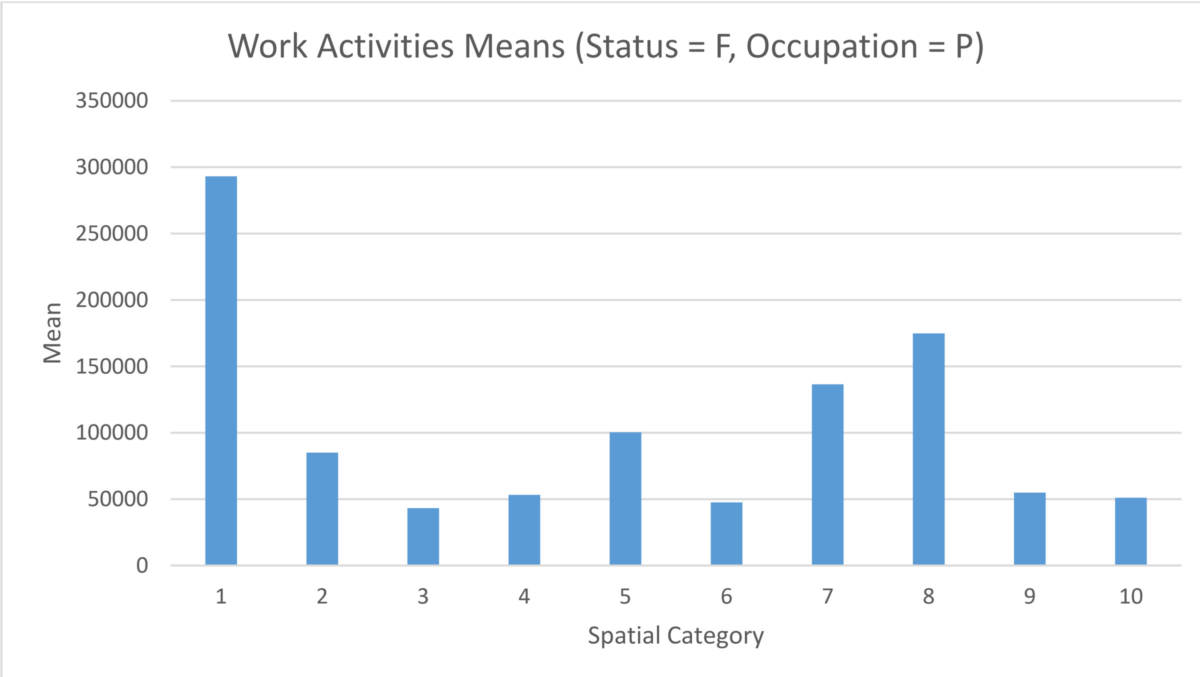
6. Appendix



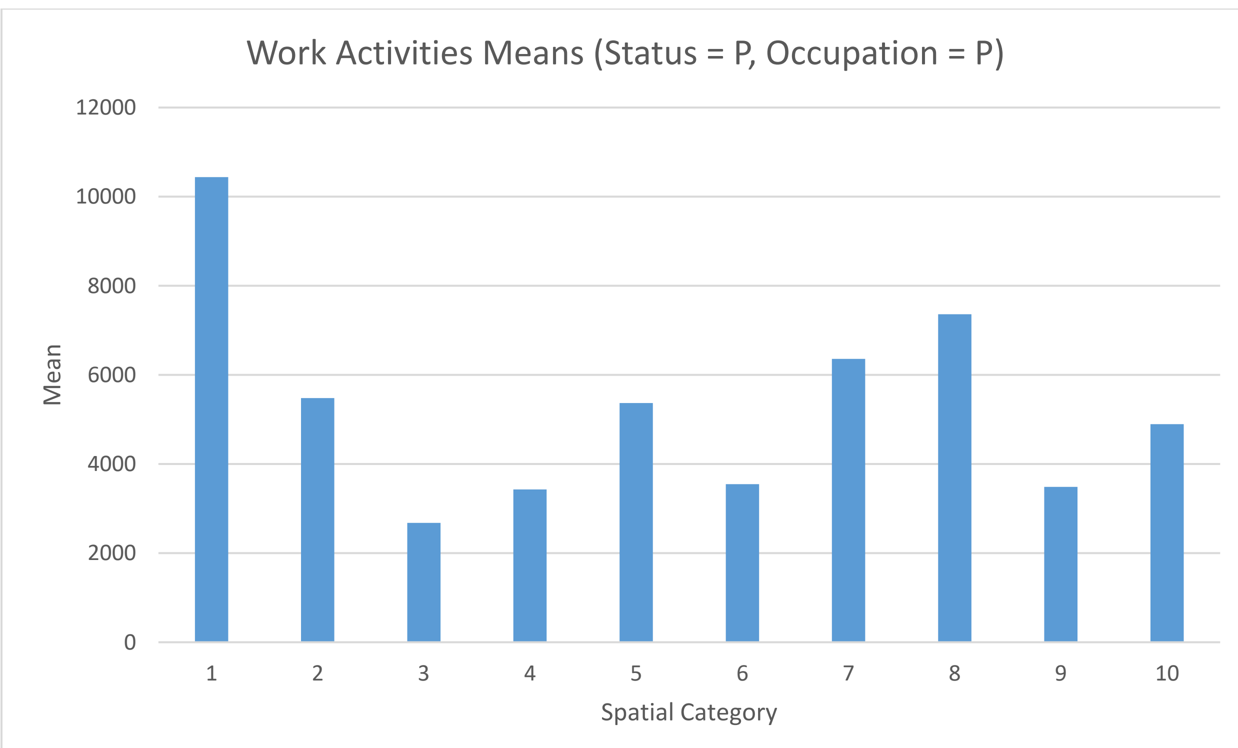
A.1: Graph for Mean Work Activities Per Spatial Category (Status = F, Occupation = S)



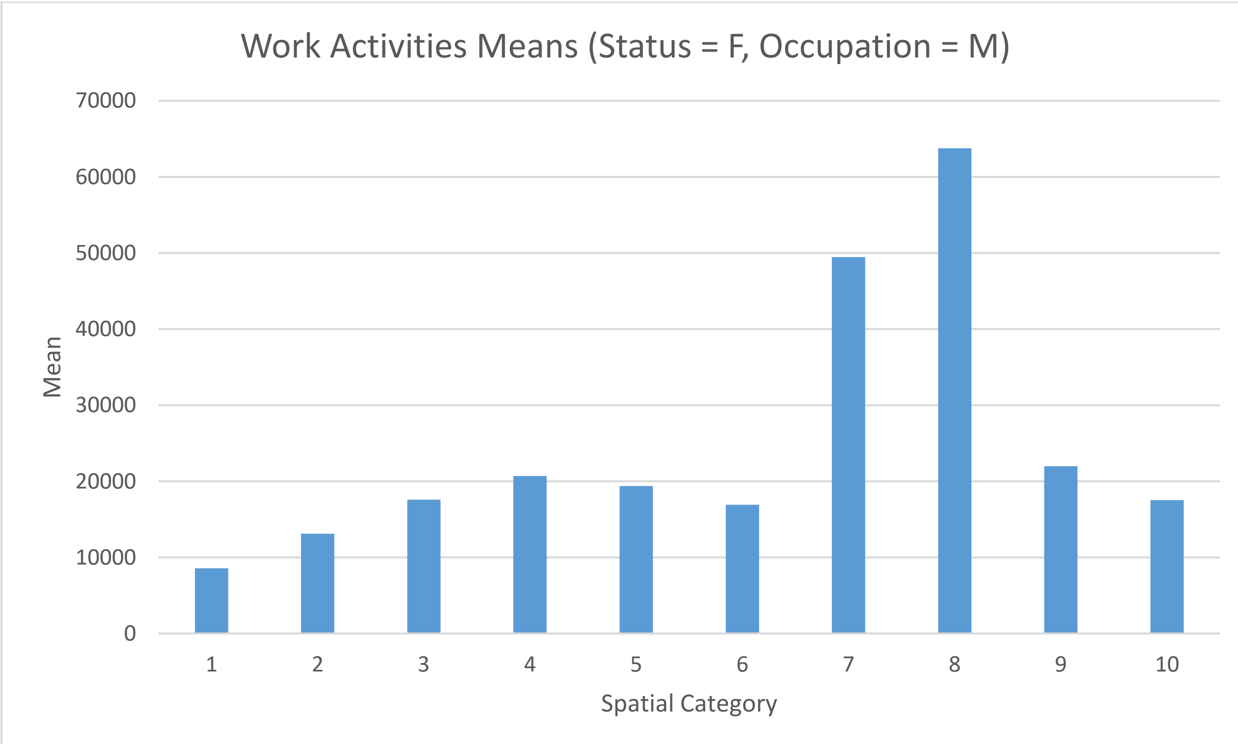
A.2: Graph for Mean Work Activities Per Spatial Category (Status = P, Occupation = S)



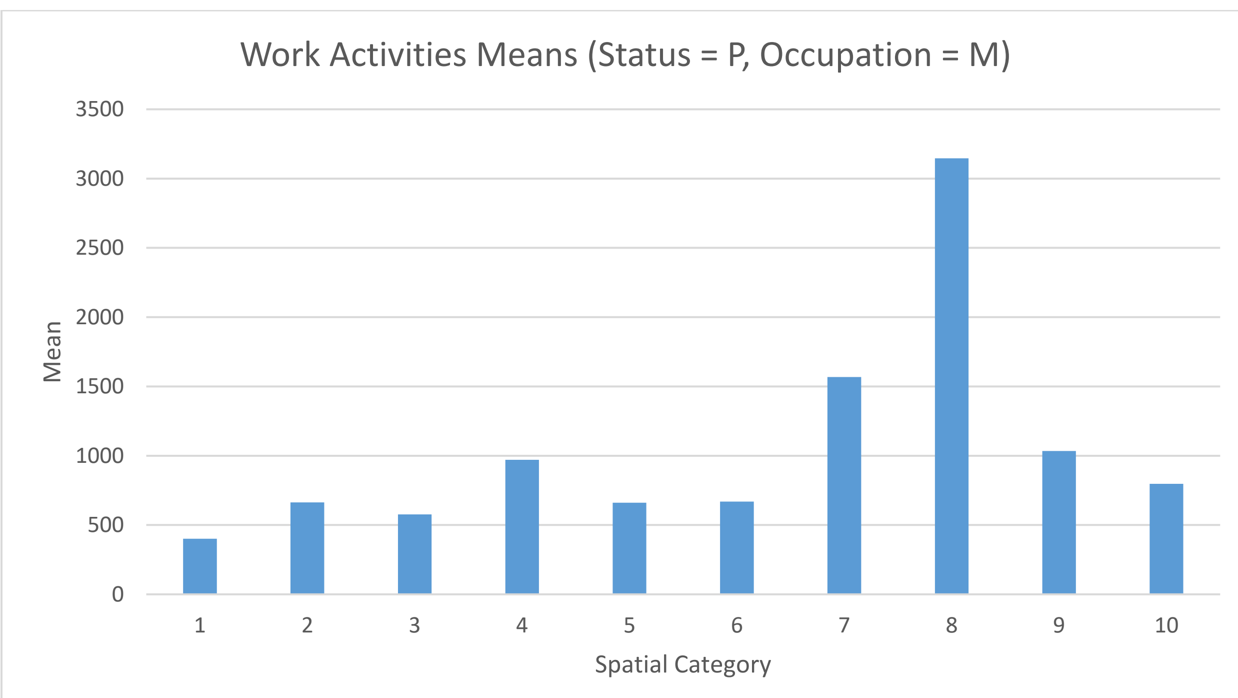
A.3: Graph for Mean Work Activities Per Spatial Category (Status = F, Occupation = P)



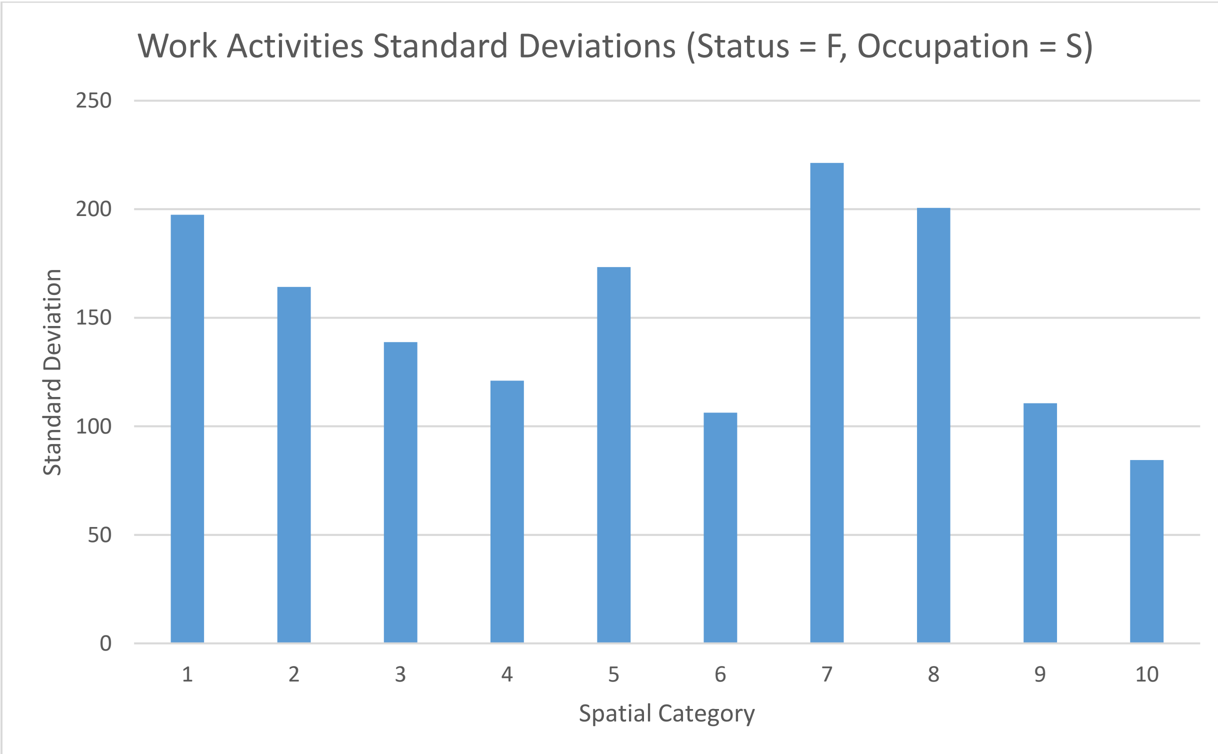
A.4: Graph for Mean Work Activities Per Spatial Category (Status = P, Occupation = P)



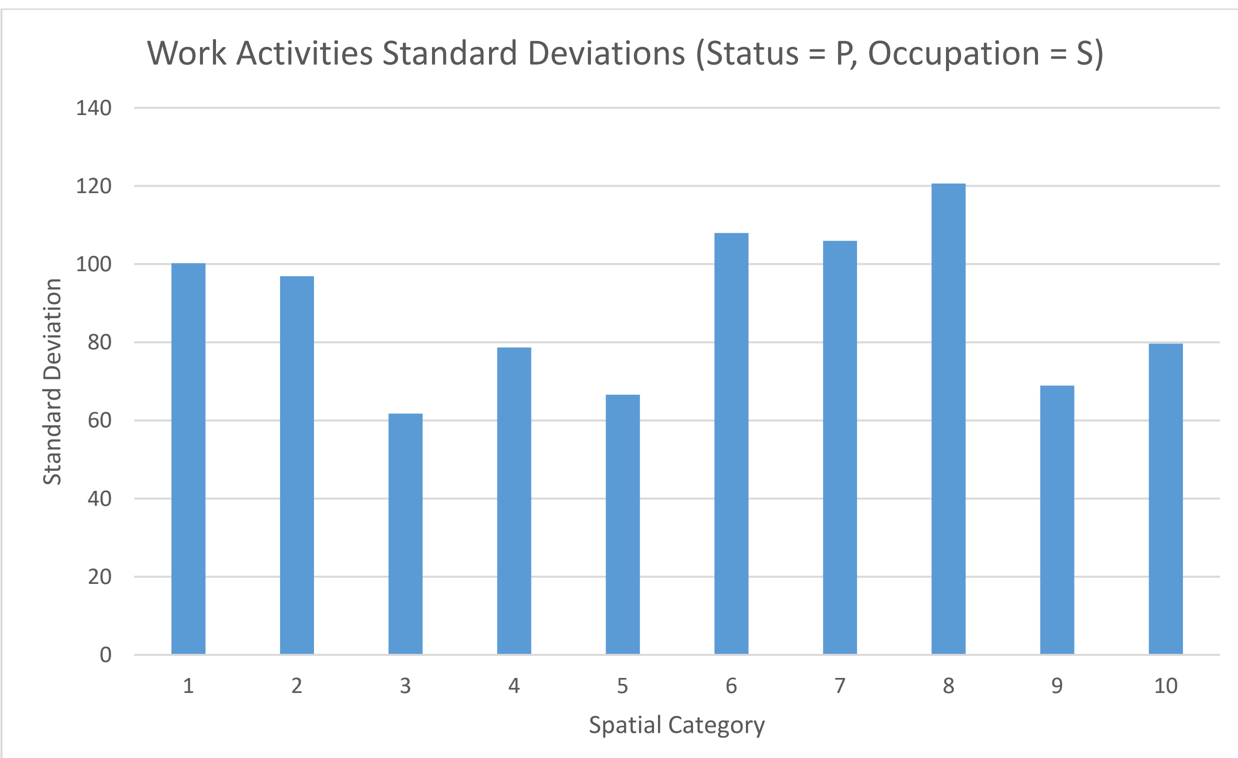
A.5: Graph for Mean Work Activities Per Spatial Category (Status = F, Occupation = M)



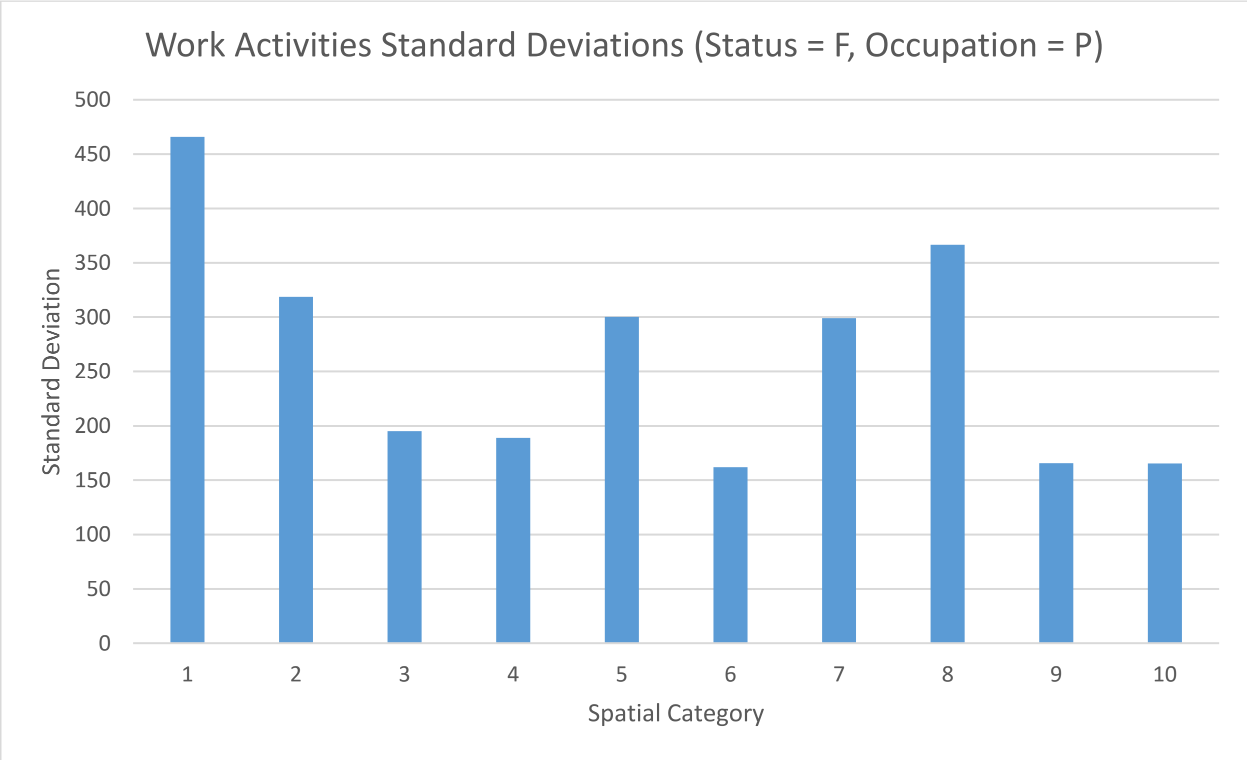
A.6: Graph for Mean Work Activities Per Spatial Category (Status = P, Occupation = M)



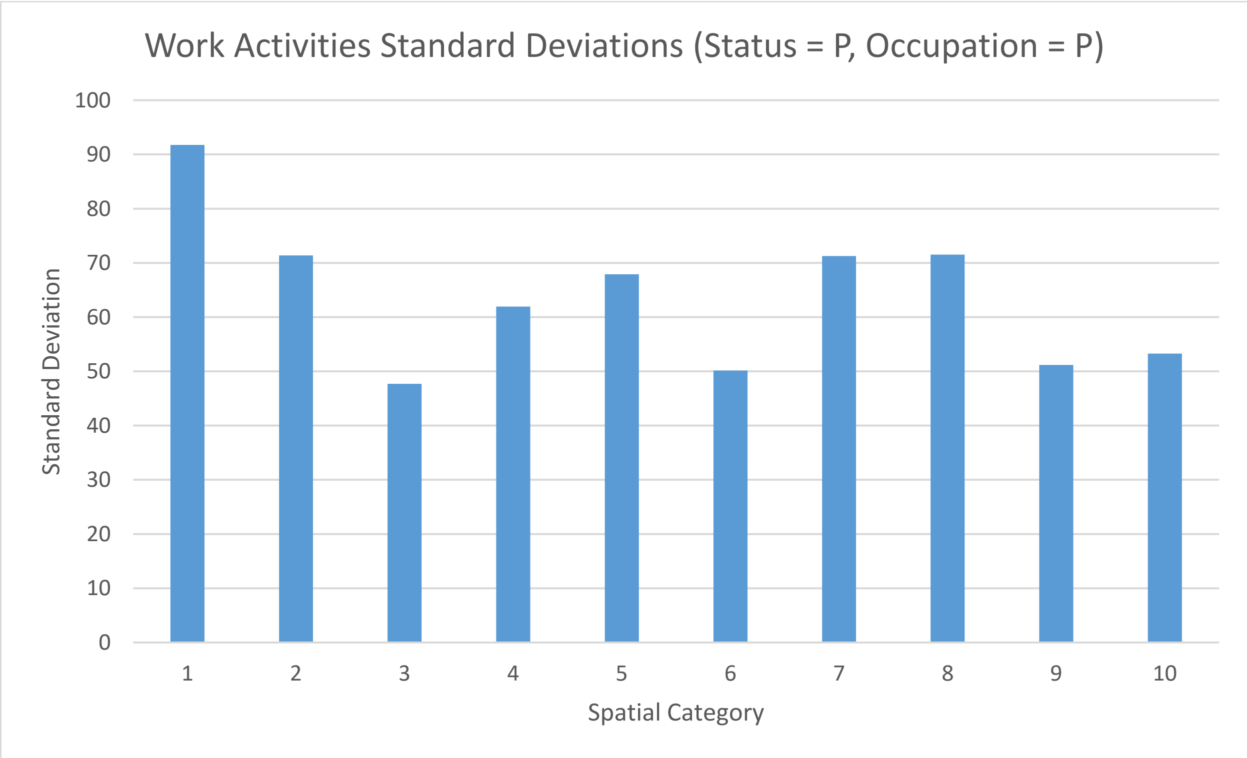
A.7: Graph for Work Activities Standard Deviations Per Spatial Category (Status = F, Occupation = S)



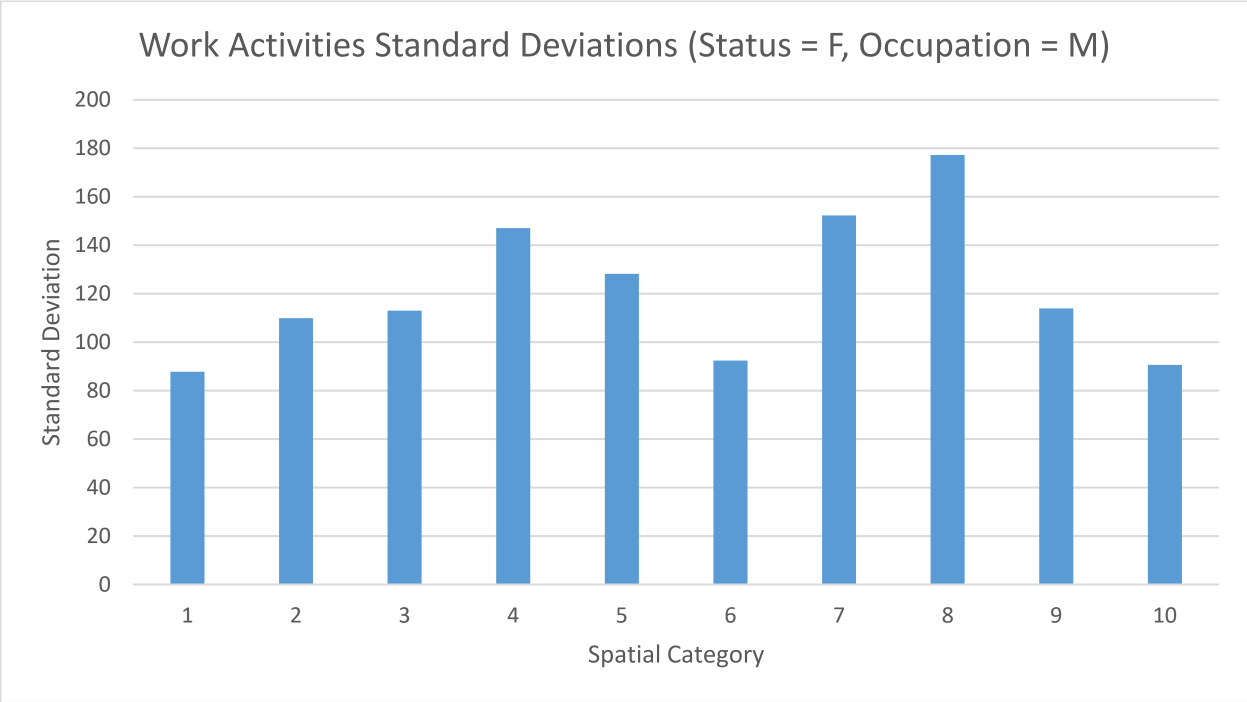
A.8: Graph for Work Activities Standard Deviations Per Spatial Category (Status = P, Occupation = S)



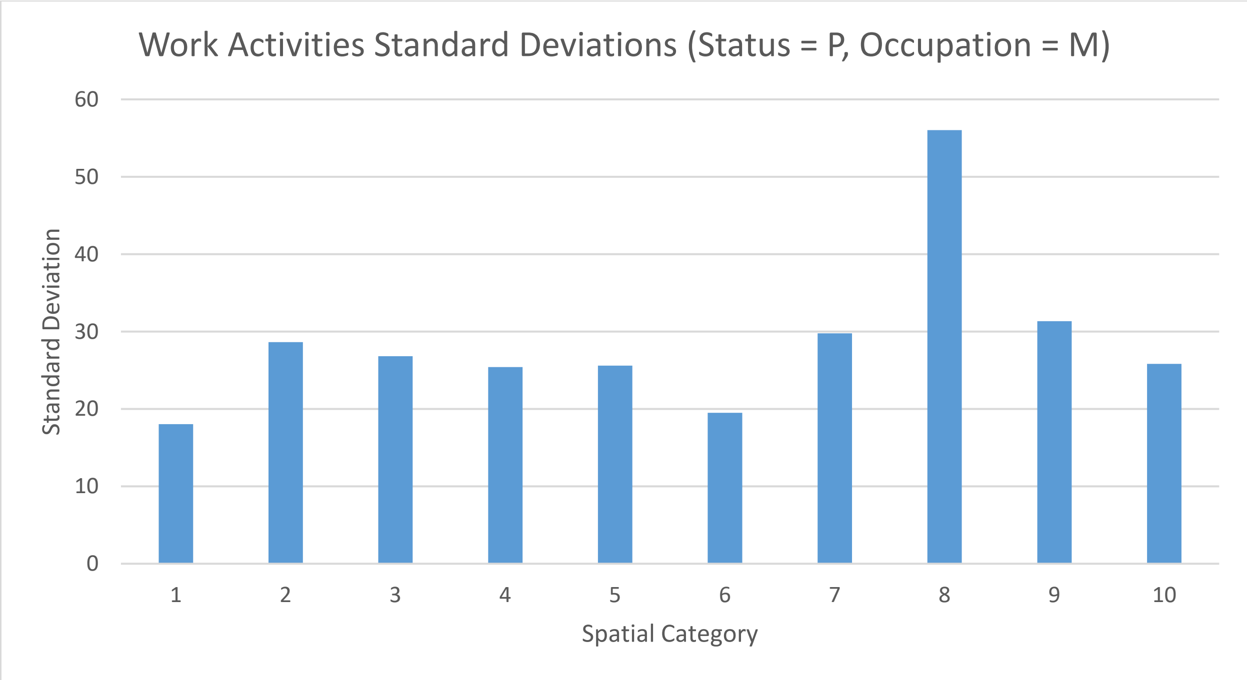
A.9: Graph for Work Activities Standard Deviations Per Spatial Category (Status = F, Occupation = P)



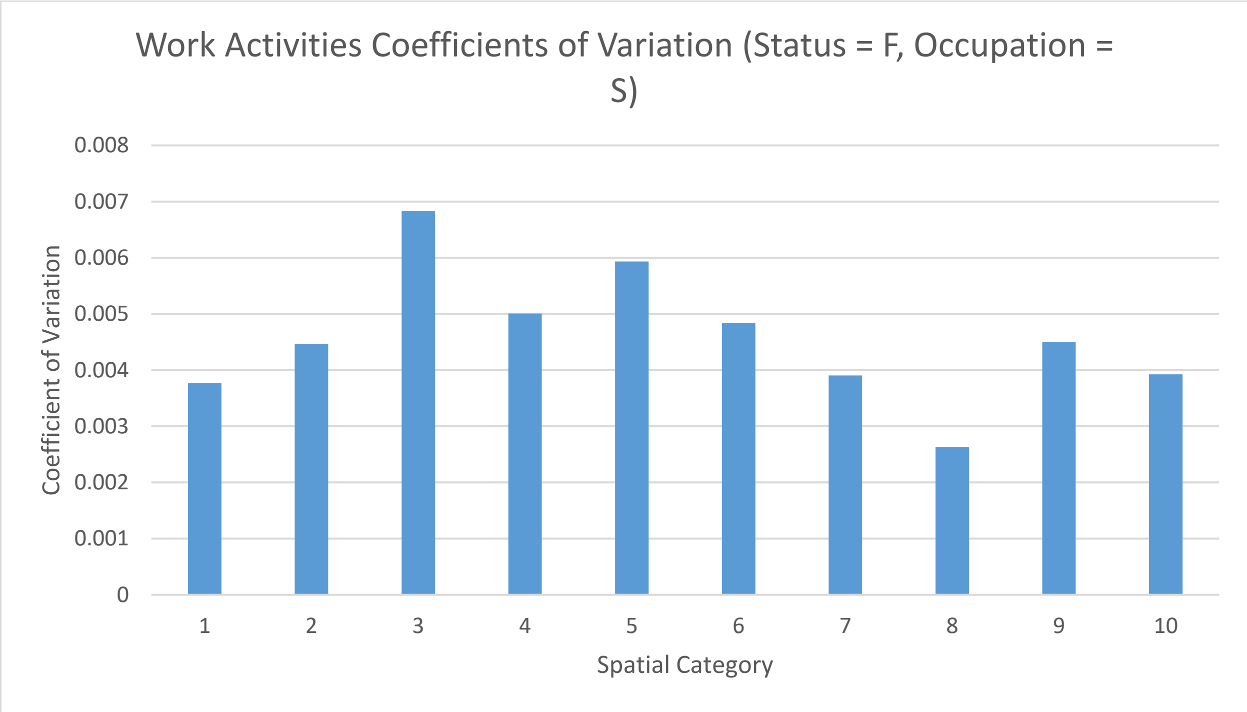
A.10: Graph for Work Activities Standard Deviations Per Spatial Category (Status = P, Occupation = P)



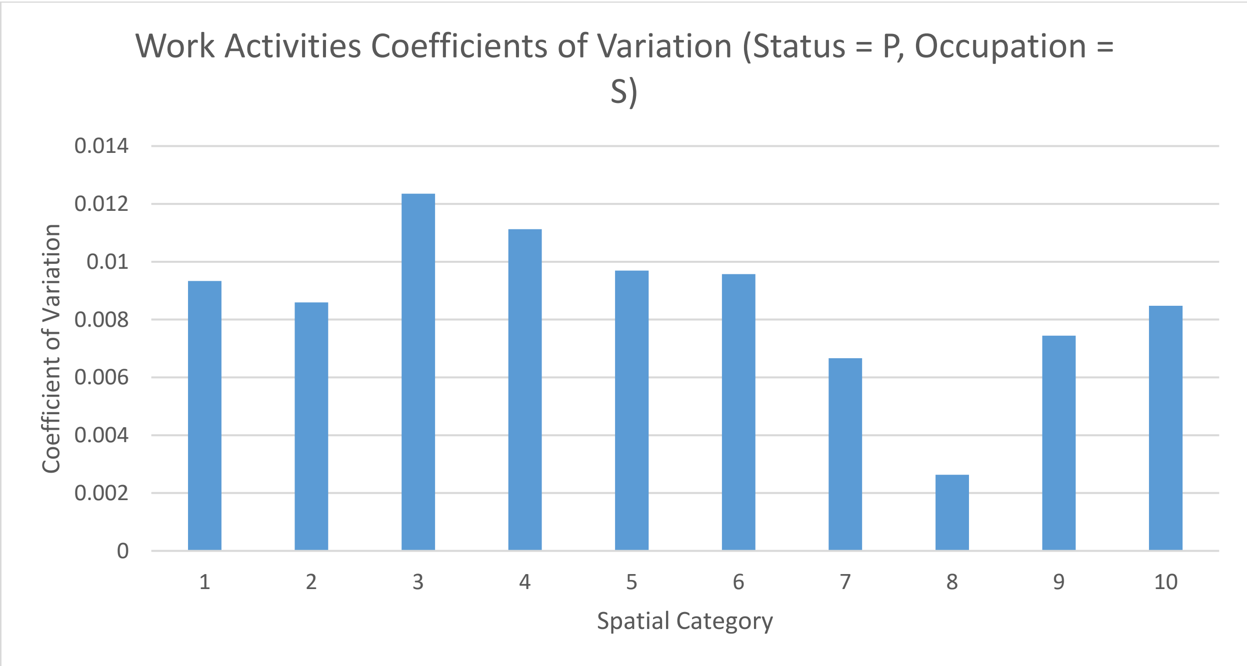
A.11: Graph for Work Activities Standard Deviations Per Spatial Category (Status = F, Occupation = M)



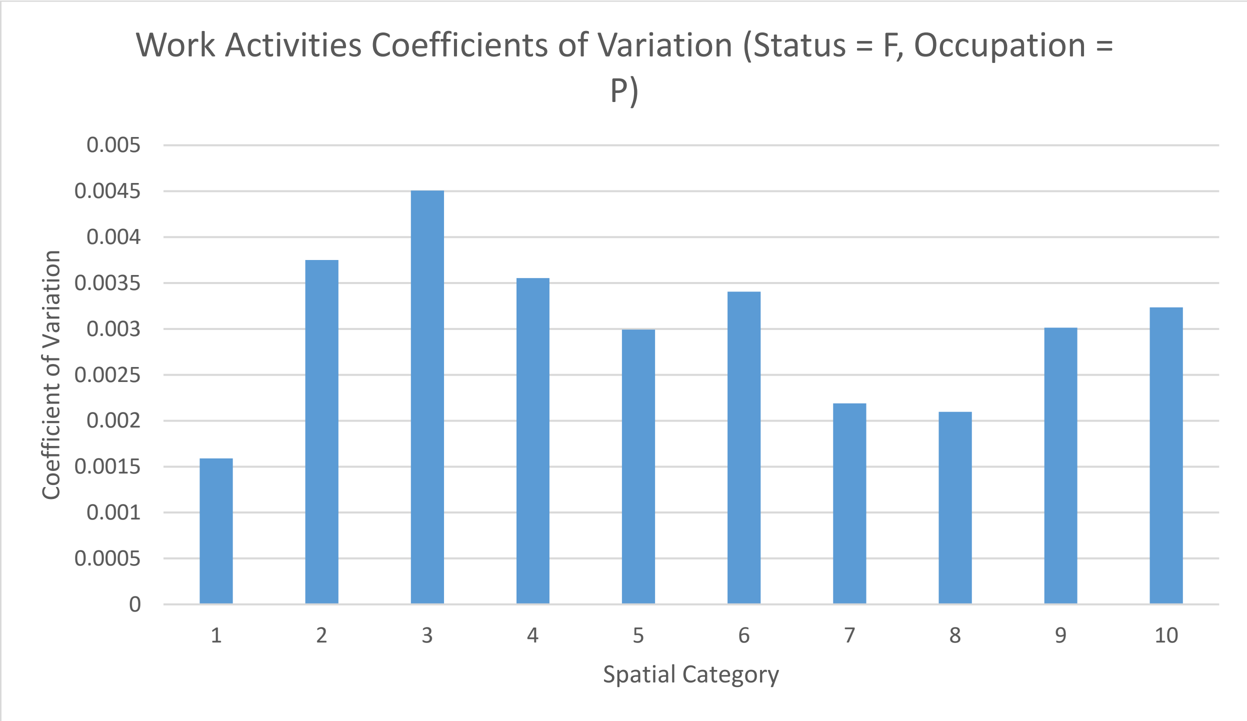
A.12: Graph for Work Activities Standard Deviations Per Spatial Category (Status = P, Occupation = M)



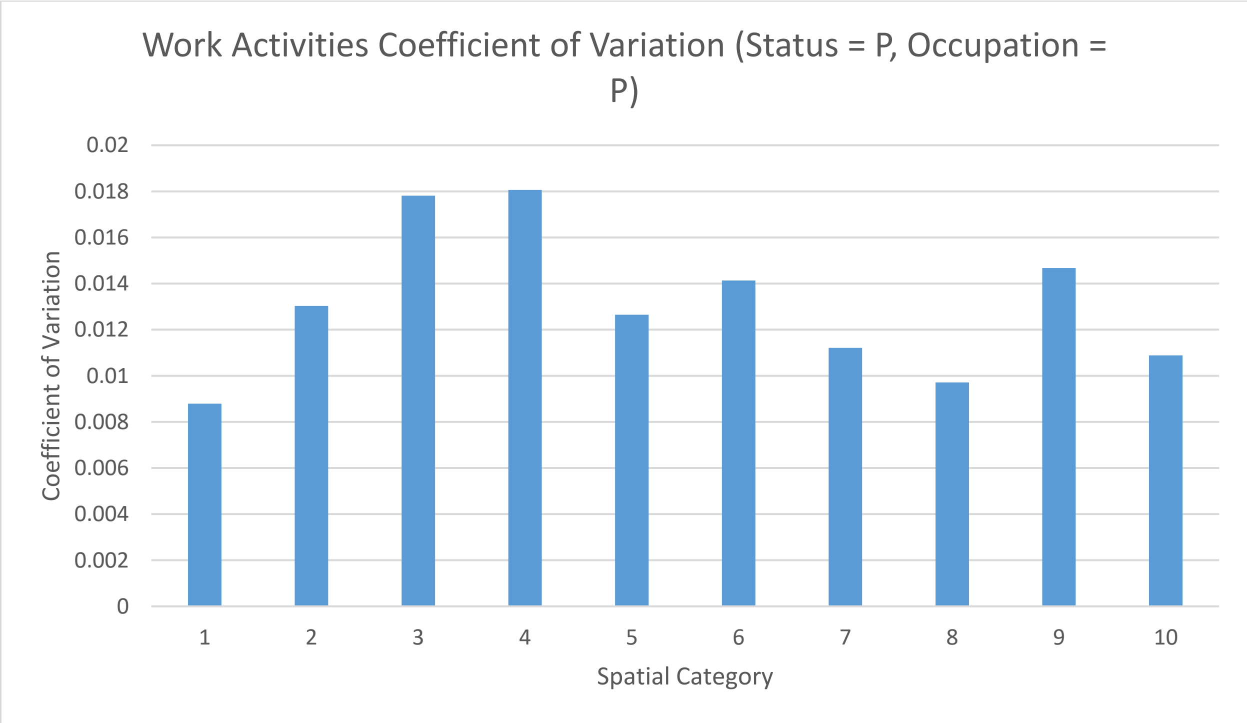
A.13: Graph for Work Activities Coefficients of Variation Per Spatial Category (Status = F, Occupation = S)



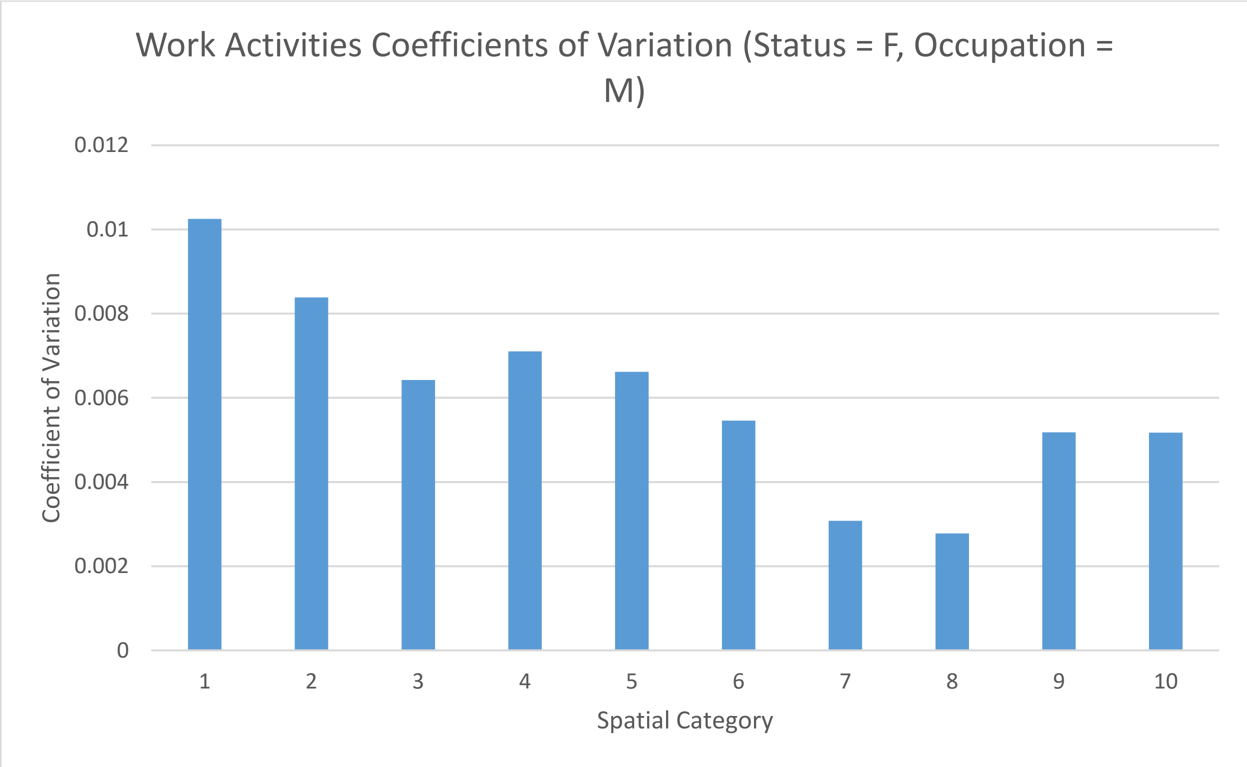
A.14: Graph for Work Activities Coefficients of Variation Per Spatial Category (Status = P, Occupation = S)



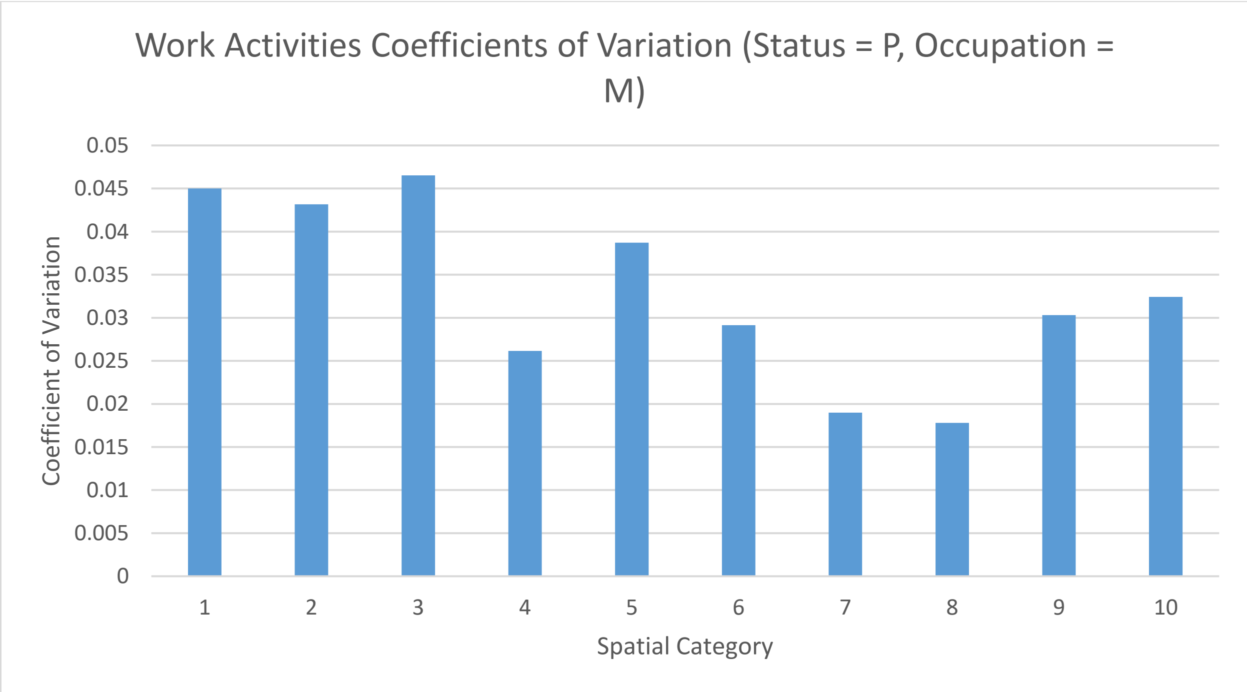
A.15: Graph for Work Activities Coefficients of Variation Per Spatial Category (Status = F, Occupation = P)



A.16: Graph for Work Activities Coefficients of Variation Per Spatial Category (Status = P, Occupation = P)



A.17: Graph for Work Activities Coefficients of Variation Per Spatial Category (Status = F, Occupation = M)



A.18: Graph for Work Activities Coefficients of Variation Per Spatial Category (Status = P, Occupation = M)