# Chapter 4. Results

This chapter is divided into three sections, the first (4.1) examines general trends in the model inputs (detailed in section 3) and identifies key areas of analysis, the second (4.2) reviews the space, time and structures and interdependencies within the model inputs before the third reviews the trip purpose classification models and their outputs (4.3).

## 4.1 Overview of model inputs

### 4.1.1 Trip distance & duration

After calculating the distances and duration of the individual trips of the MTL Trajet survey, our analysis finds a total of 7,594 trips which are removed from the analysis based on the outlier strategy adopted in 3.4.3. As shown in **Table 4.1**, the majority (6,709) of these were from trips that were less than 50 m in length. These trips are potentially from cases where the app had switched on for slight movements or the user had mistakenly ended a trip while in traffic for more than 2 minutes (Patterson & Fitzsimmons, 2017b).

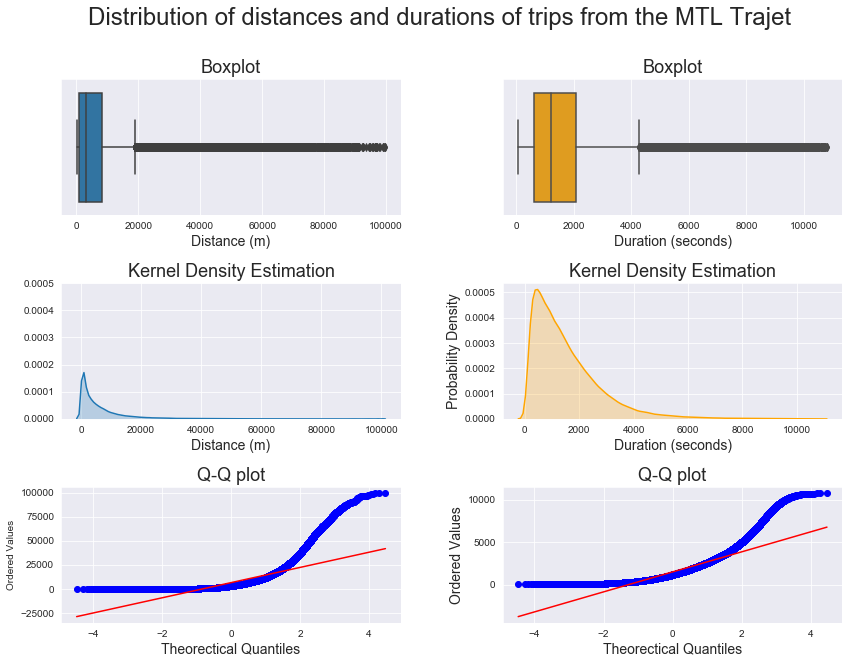
**Table 4.1** Outline of trips removed from the analysis

|  |  |  |
| --- | --- | --- |
| *Outlier Type* | | *Number removed* |
| *Distance below 50 m* | | 6,709 |
| *Distance above 100 km* | | 62 |
| *Duration below 60 seconds* | | 412 |
| *Duration above 3 hours* | | 411 |
| *Total* | 7,594 |

The resulting trips are shown to have a mean distance and duration of around 6.6 km and 26 mins, respectively (**Table 4.2**). Here, we see that both these variables are positively skewed, although distance is more so. The disparity between the mean and median in both trip distance and duration is indicative of both these variables exhibiting a long-tailed distributions, and this can be visually identified by examining the univariate kernel density estimations and Quantile-Quantile plots shown in **Figure 4.1**.

**Table 4.2** Summary statistics of distance and duration of trips from the 2017 MTL Trajet travel survey (converted to km and minutes; *N=177,938*)

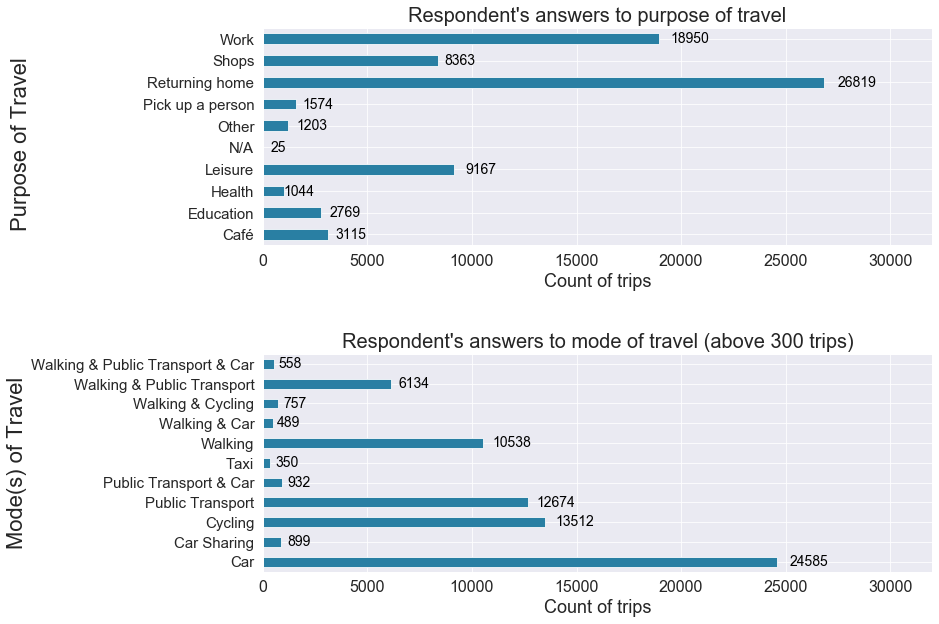
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | *mean* | *STD* | *min* | *25%* | *Median* | *75%* | *95%* | *max* | *kurtosis* | *Skewness* |
| *Distance (km)* | 6.63 | 9.92 | 0.05 | 0.84 | 3.14 | 8.09 | 25.25 | 99.81 | 15.216 | 3.355 |
| *Duration (min)* | 25.62 | 21.42 | 1.00 | 10.27 | 20.07 | 34.68 | 65.81 | 179.98 | 6.097 | 1.967 |



**Figure 4.1** Boxplots (top), Kernel Density Estimation (middle) and Quantile-Quantile (bottom) plots showing the distribution of distance and duration of trips from the 2017 MTL Trajet travel survey.

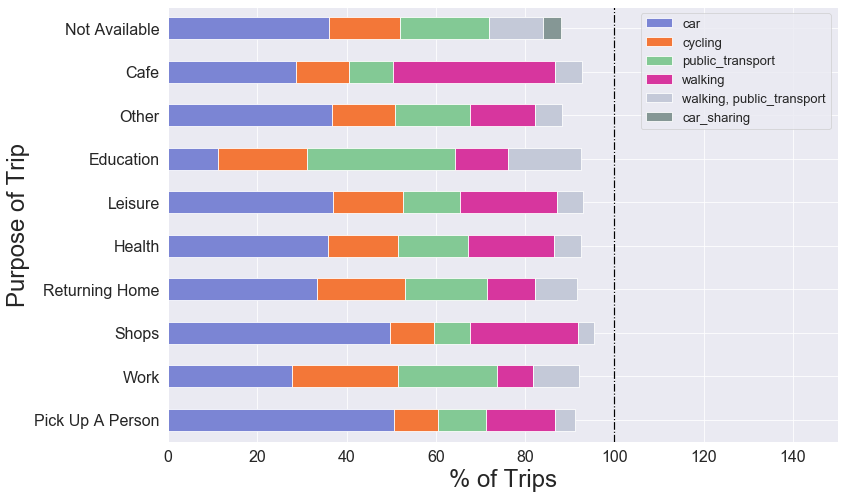
### 4.1.2 Travel purpose & mode

There are total of 73,029 trips from the MTL Trajet survey containing both a travel mode and purpose label. As shown in **Figure 4.2**, the categories of these variables have not been selected by the respondents in equal proportions. It is shown that severe class imbalance exists within both of these categories, with around 63.7% (45,769) of the trips that have been labelled as either trips to work and back to home, and 33.6% of the trips being taken by car. This finding is not unexpected for a survey that has taken place in North American city with a high level of employment, however, so we can argue that this study is relatively representative of trips occurring across Montréal (Meng *et al.*, 2019).



**Figure 4.2** Bar charts showing the type of trip purpose and travel mode selected by respondents to the 2017 MTL Trajet survey.

When the travel mode is broken down by purpose, in **Figure 4.3**, we see that there is higher usage of cars in trips for shopping and picking people up and lower usage of cars in trips for education.Notably, a higher proportion of respondent have walked or cycled when taking trips to work, cafés and places of education.



**Figure 4.3** Bar chart comparing the proportion of each unique trip purposes accounted for by each unique travel modes.

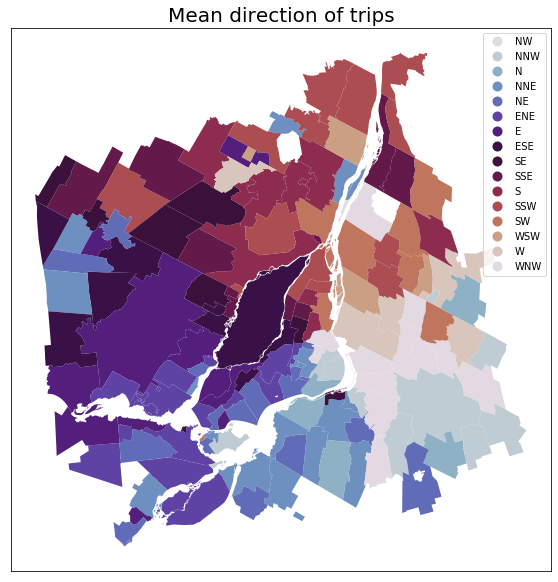
Comparing the distances and duration of the trips as grouped by trip purpose, in **Table 4.3**, we see that trips to cafés and shops are shorter in both mean distance (4.5 & 4.8 km) and duration (23 & 20 mins) compared to the other forms of trips such as work and returning home. When cross-referencing with travel mode (from **Figure 4.3**), it could be proposed that these values are a product of the fact that a higher proportion of these trips are walked.

**Table 4.3** Summary statistics of trip distance and duration per trip purpose (Note: trips that are classed as ‘Not Available’ have been omitted from this table)

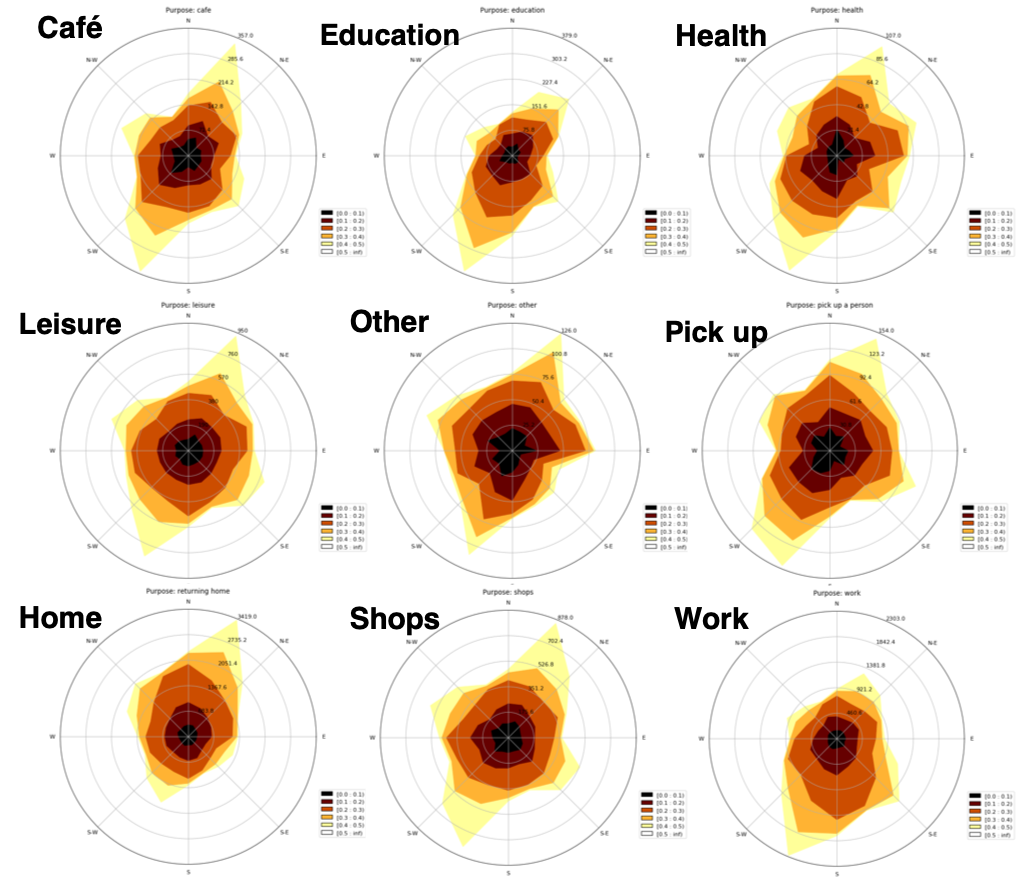
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *Trip Purpose* | *Trip distance (km)* | | | *Trip duration (mins)* | | |
|  | μ | σ | Skew | μ | σ | Skew |
| Café | 4.5 | 7.6 | +4.5 | 22.9 | 19.5 | +2.4 |
| Education | 5.8 | 6.8 | +3.5 | 28.9 | 19.4 | +1.3 |
| Leisure | 6.8 | 10.3 | +3.1 | 35.2 | 20.8 | +2.0 |
| Health | 6.2 | 8.0 | +2.8 | 25.2 | 19.7 | +2.0 |
| Other | 8.7 | 12.9 | +3.1 | 31.3 | 25.3 | +1.9 |
| Returning home | 7.5 | 9.8 | +3.2 | 29.0 | 22.2 | +1.7 |
| Pick up a person | 7.8 | 10.9 | +2.9 | 25.1 | 20.9 | +2.1 |
| Shops | 4.8 | 7.2 | +3.5 | 20.4 | 17.6 | +2.4 |
| Work | 7.6 | 8.3 | +2.6 | 28.8 | 19.8 | +1.5 |

### 4.1.3 Trip direction:

The mean direction of all trips taken across all 91 regions of Greater Montreal is shown in **Figure 4.4**.Here, we see that the direction is general towards the city of Montreal (see **Figure 3.2**) indicating we can be somewhat confident in assuming we have accounted for a some degree of the MAUP – with the directional dynamics of trips facing ‘inward’ towards downtown and the study area chosen versus out of the study area (ref). Across the individual purpose class, in **Figure 4.5,** the mean direction of the trips are generally shown to be in the NNE and SSW directions, something which is similar to the morphology of the island of Montreal. Notably, work and returning home trips shown to be more directionally dependent, in the SSW & NNE respectively, than the other purpose classes. Directional independence is shown in the trips categorized by other, café and purposes.



**Figure 4.4** Map showing the mean direction of trip within each region of Greater Montreal.



**Figure 4.5** Circular contour plot showing the mean direction of trips for each trip purpose.

### 4.1.4 Rush-hour & City Labels

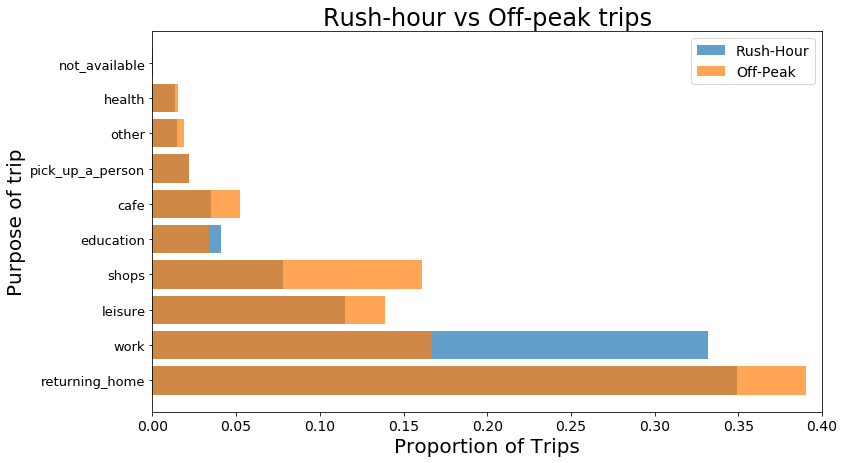
After applying city and rush hour labels to the origin and destination points (see 3.3.1), the majority of trips are found to have occurred within the City of Montreal (93.5%) and are evenly split between rush hour and off-peak (**Table 4.4**).

**Table 4.4** Results from the application of Rush-hour and City labels to the data.

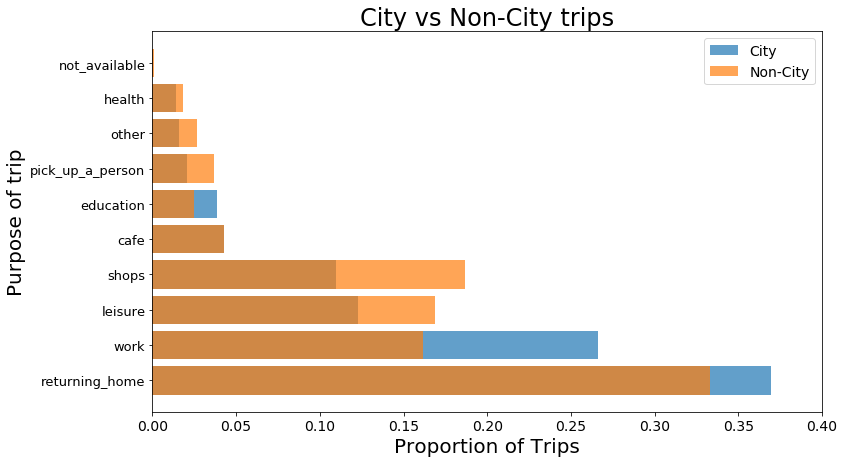
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Rush hour? | | City? | |
|  | Yes | No | Yes | No |
| Origin of trip | 36785 | 36244 | 63811 | 9218 |
| Destination of trip\* | 38650 | 34490 | 64136 | 8893 |

\* including trips that have passed through rush hour or city

When separated by purpose class, a higher proportion of trips are discovered to be carried out for work and education at rush-hour times versus trips to shops which are proportionally carried out at off-peak times (**Figure 4.6**). **Figure 4.7**, highlights that work and home-bound trips are disproportionately represented in trips occurring in the city as opposed to outside the city, where shopping and leisure trips are more proportionally represented.

**

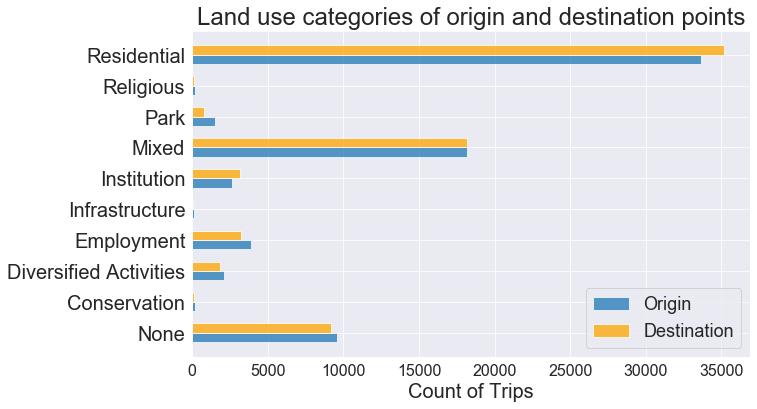
**Figure 4.6** Bar chart showing the proportion of trips carried out during rush-hour and off-peak as grouped by purpose.

**

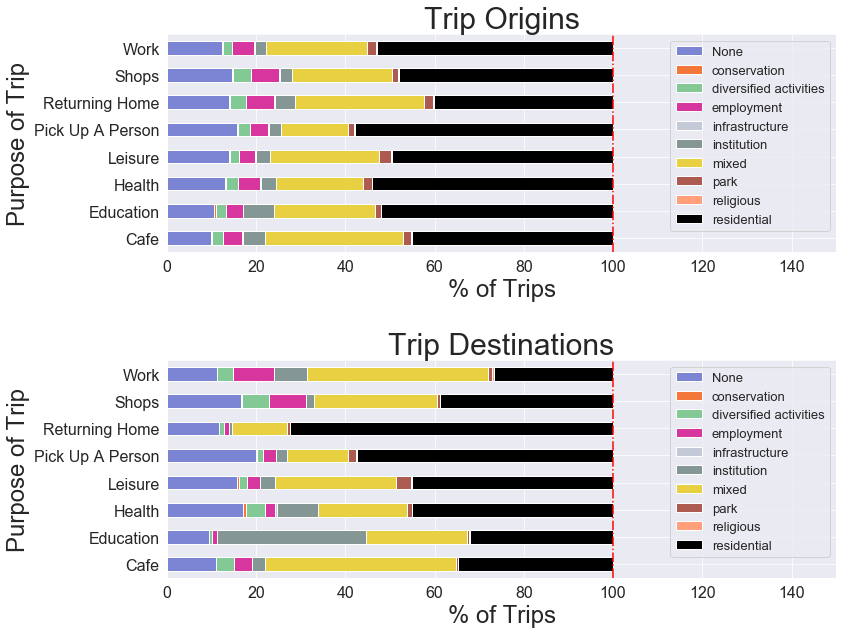
**Figure 4.7** Bar chart showing the proportion of trips carried out within and outside the City of Montreal as grouped by purpose.

### 4.1.5 Land Use

The majority of trips are found to have their origins and destinations in areas of the City of Montreal classified as residential and mixed use categories (Ville de Montreal, 2014; **Figure 4.8**). Most trips are found to begin in areas of mixed and residential land use (around 65-75% of trips), whereas most trip end in a variety of land use categories dependent on trip purpose (see **Figure 4.9**).



**Figure 4.8** Bar chart showing number of trips that have their origins or destinations in each land use category (as defined by Ville de Montreal, 2014).

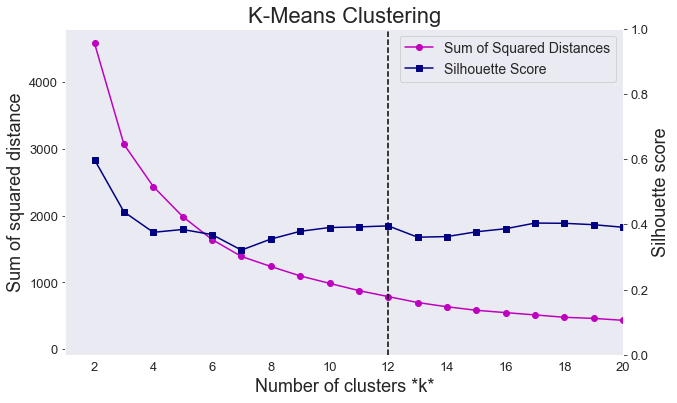


**Figure 4.9** Bar charts comparing the proportion of each unique trip purposes accounted for by each unique land use category (as defined by Ville de Montreal, 2014) in the trip origins and destinations.

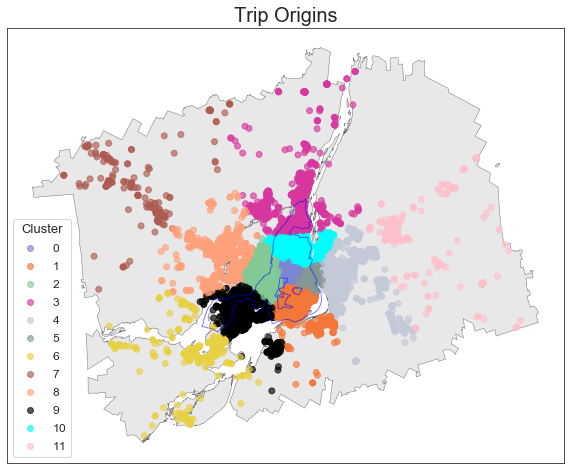
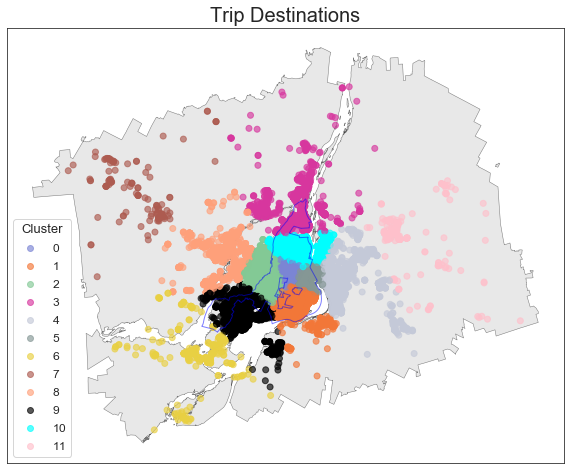
### 4.1.6 Clustering

### 4.1.6.1 Spatial

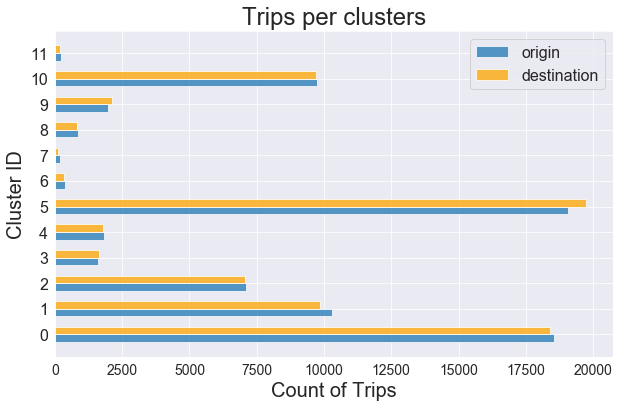
After fine-tuning values of *k* between 2-20 and evaluating the sum of squared distances and silhouette score within each k-number of clusters (**Figure 4.10**), we select a total of 12 for the k-means clustering algorithm to be built upon. These clusters mapped across the study region in **Figure 4.11** and a summary of how many trips have been assigned to each cluster is shown in **Figure 4.12**. Note that, the algorithm separates a region containing Downtown Montreal (*cluster-id=0*) and the clusters of 0,5 and 10 are most prominent in the data.



**Figure 4.10** Line graph comparing sum of squared distances and silhouette scores of k-means clustering algorithm for k between 2-20.



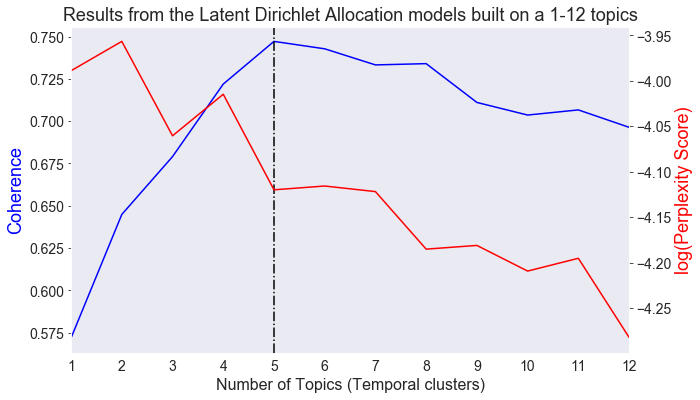
**Figure 4.11** Map of origin and destination points from the MTL Trajet trips coloured by cluster label across the study region.



**Figure 4.12** Bar chart showing number of trips per spatial cluster identified by the k-mean clustering algorithm.

### 4.1.6.1 Temporal

After examining coherence and log perplexity of LDA models modelled between 1-12 topics (here, temporal clusters), we select 5 topics to build from the data. As shown in **Figure 4.13**, at 5 topics we maximise the coherence of the LDA model, suggesting that the words within the topics are most similar at this point (Kumar, 2018). Notably, perplexity of model continues to drop passed this point, indicating that the model is better at predicting the topics, however in practice it has been found that coherence is a more stable metric for an LDA (Kumar, 2018).

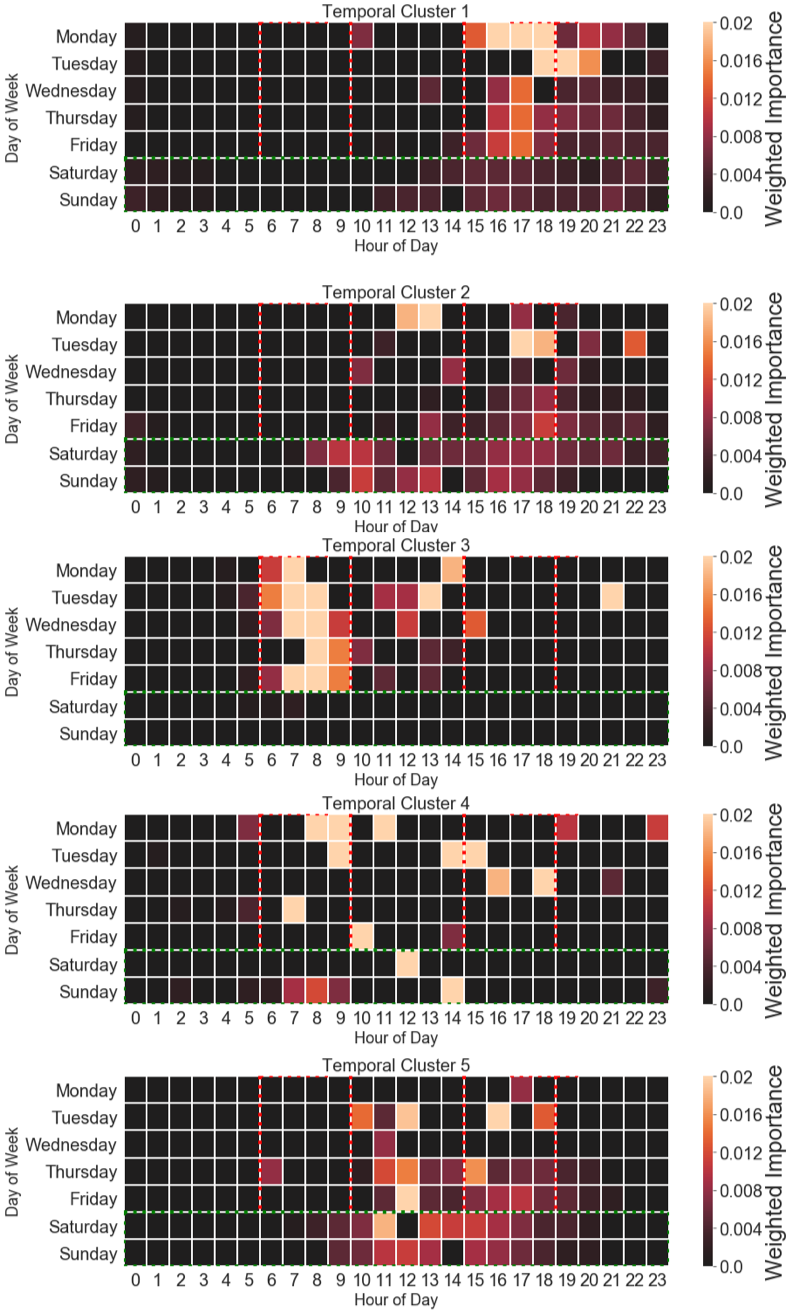


**Figure 4.13** Line graph comparing coherence score and log perplexity of LDA models using a topic count of between 1-12.

Results showing the temporal profile of the 5-topic LDA model are shown in the form of calendar plots in **Figure 4.14**. Note that, each ‘temporal profile’ is a 7-day by 24-hour matrix, which indicates the probability that a given ‘temporal word’ (i.e. “Sunday\_7”) is associated with that temporal cluster or TC (known as its *weighted importance*). The trip purposes classes associated with each TC along with their weighted importance are outlined in **Table 4.5**.

**Table 4.5** Outline of trip purposes associated with each temporal cluster found by a 5-topic LDA model.

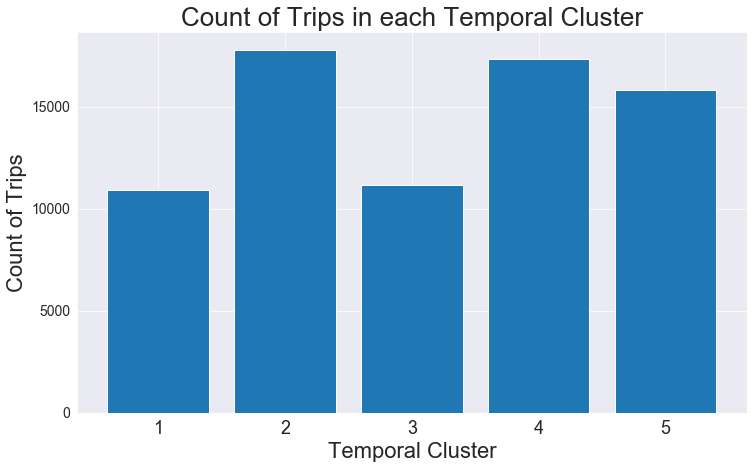
|  |  |  |
| --- | --- | --- |
| *Temporal cluster (TC)* | *Associated trip purpose* | *Weighted importance* |
| 1 | Returning Home | 0.549 |
| 2 | Leisure | 0.349 |
| Education | 0.111 |
| Other | 0.041 |
| 3 | Work | 0.488 |
| 4 | Work | 0.276 |
| Not Available | 0.002 |
| 5 | Shop | 0.306 |
| Café | 0.113 |
| Pick Up a Person | 0.060 |
| Health | 0.004 |



**Figure 4.14** Calendar plot showing the weighted importance of each ‘temporal word’ in each of the 5 temporal clusters (rush hour periods as defined by this study are outlined in **red** and weekends are outlined in **green**).

Work is found to be the only trip purpose existing in more than one TC as it is found in 3+4. Notably, TC3, is a topic containing temporal words mainly within the weekdays and morning rush hour –these are times we expect to be populated with work trips. The strongest association found between trip purpose classes and the TCs is found with the returning home class (with a probability of 0.549 to be found in cluster 1). On examining the temporal profile of TC1, 14 out of the 20 evening rush hour segments have been associated. **Table 4.5**, also indicates the relatively weak temporal clustering of Leisure & Education in TC2 and Shop, Café & Pick up a Person trips in TC5. This suggests that these purposes broadly share similar temporal characteristics within the data.

Finally, after joining the TCs back to the data (using step set out in 3.3.3), we find that TC2, TC4 and TC5 are most frequent in the data (**Figure 4.15**).



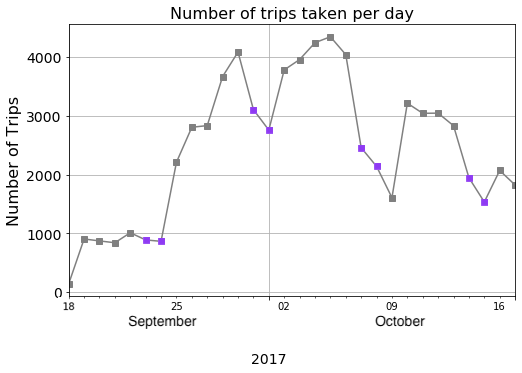
**Figure 4.15** Count of trips associated with each temporal cluster identified for this analysis.

## 4.2 Exploratory Space and Time Data Analysis of Model Inputs

This section highlights the methods carried out to investigate space, time and space-time signatures in the data. It is hoped that the identification of these forms of trends will assess the ability for the purposes to be modelled inform the modelling process (detailed in 4.3).

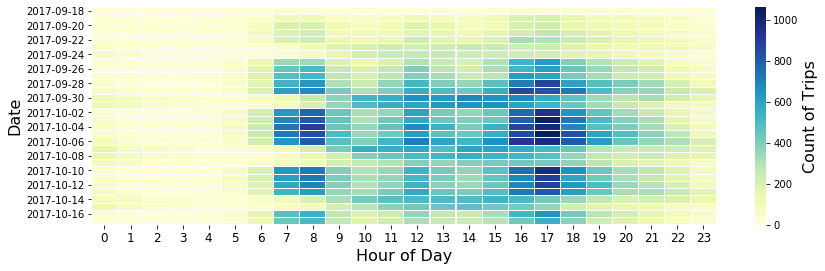
### 4.2.1 Temporal trends

A total of 73,029 trips were recorded across the study period (18th September 2017– 17th October 2017), but there is significant variation in the amount of recorded trips per day. As shown in **Figure 4.16**, during the first 7 days of the study less than around 1000 trips were recorded per day compared with more than 1500 trips in the remaining days (with the most amount of trips being recorded on Thursdays/Fridays). Here, less trips are recorded on weekends versus weekdays, other than Monday 9th October, which was the day Thanksgiving was celebrated that year in Canada.



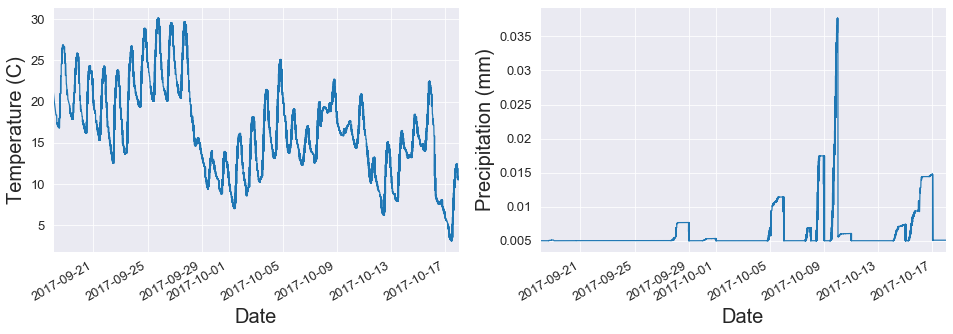
**Figure 4.16** Line plot showing the amount of recorded trips taken from the MTL Trajet app between 18th September 2019– 18th October 2019 (weekends indicated in **purple**).

The temporal profile of the count of trips per hour across the study period is shown in **Figure 4.17**. Here we see a clear temporal tendency for trips occurring within week days and specifically the rush hour periods.



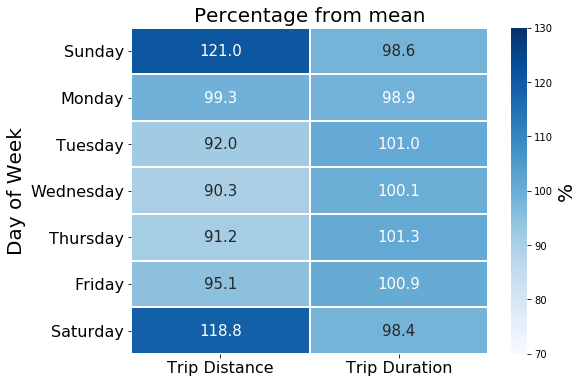
**Figure 4.17** Calendar plot showing the count of trips recorded per hour across the study period (18th September 2017–17th October 2017).

Further, there is also temporal deviation in weather patterns during the study period. As shown in **Figure** **4.18**, the temperature is shown to generally decrease with more rain falling towards the latter half the study. We expect that change in weather to have had an effect on the modes of transport chosen by the survey participants (Gong *et al.*, 2018), and thus may hinder the ability of our classification models to generalise (Xie *et al.*, 2016). Correspondingly, we find moderately strong and statistically significant (p-value>0.005) negative correlation between cycling (-0.36ρ) and walking (-0.44ρ) usage and temperature.

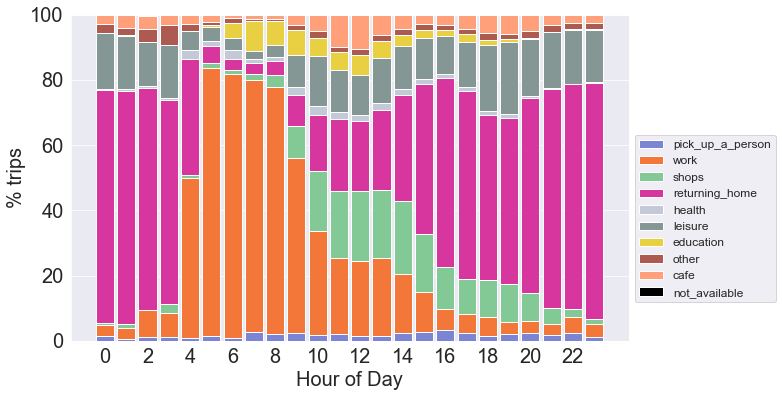


**Figure 4.18** Time series plot showing the average temperature (in Celsius) and precipitation (mm) recorded during the study period

As broken down by week, on average, trips of longer distances are taken on the weekends versus weekday (**Figure 4.19**). Arguably this could result from the influence of work, with people travelling further into rural areas during weekends. Notably, there is no deviation from the mean travel duration across the week.

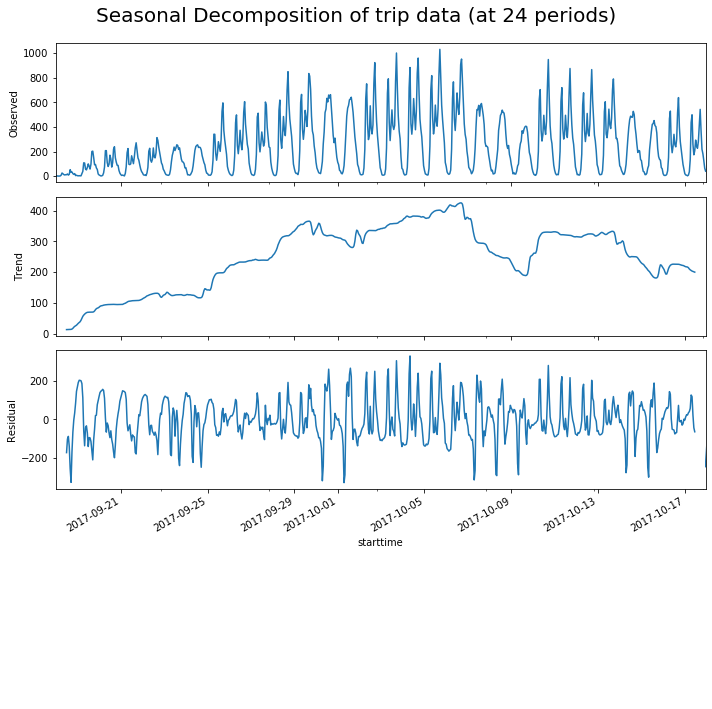


**Figure 4.19** Average trip distance and duration represented as a percentage of the mean.

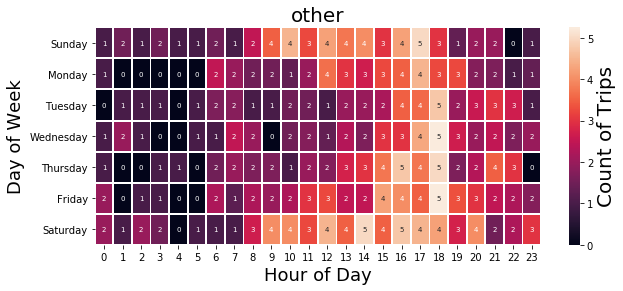


**Figure 4.X**

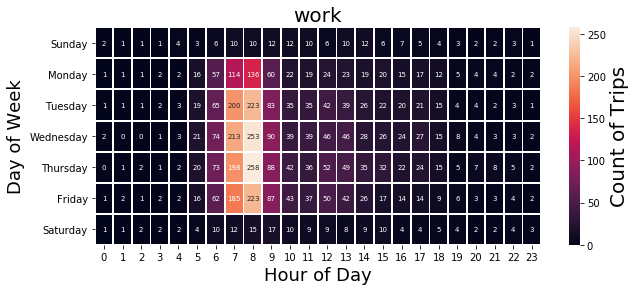
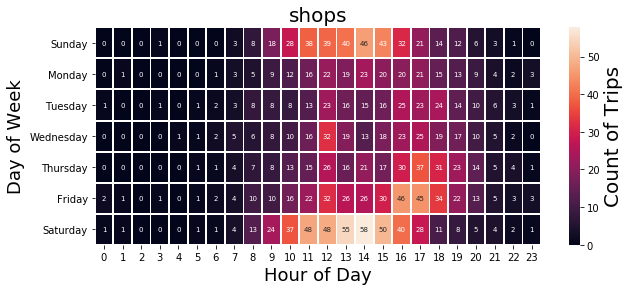
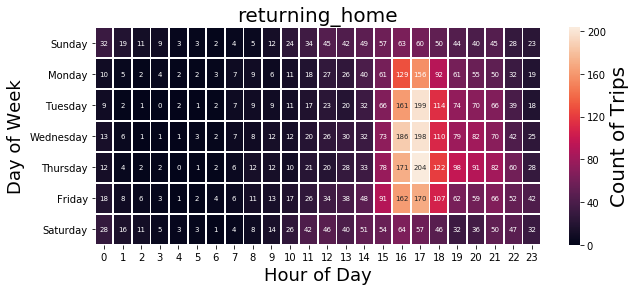
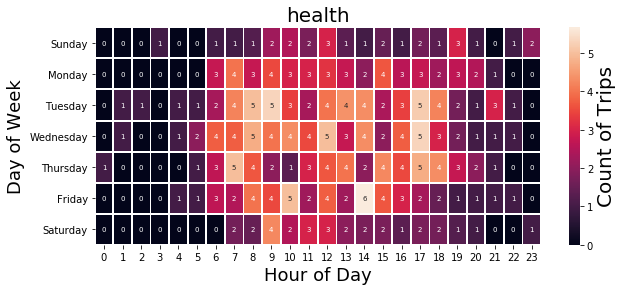
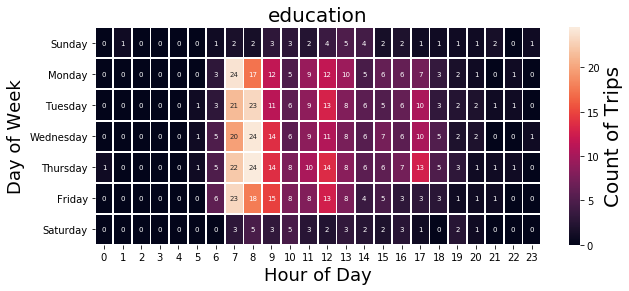
Clear diurnal a Temporal Decomposition



**Figure 4.X**







**Table 4.X** Augmented Dickey-Fuller Test (significant below 0.005 shown in **bold**)

|  |  |  |  |
| --- | --- | --- | --- |
| Purpose | ADF | p-value | n |
| All | -2.7261 | 0.0696 | 185285 |
| Cafe | -2.7386 | 0.0676 | 3189 |
| Education | -2.8689 | 0.0491 | 2830 |
| health | -4.1338 | **0.0009** | 1061 |
| Leisure | -1.8601 | 0.3511 | 9379 |
| Not available | -4.7958 | **0.0001** | 25 |
| Other | -2.4963 | 0.1164 | 1219 |
| Pick a person up | -2.8686 | 0.0491 | 1592 |
| Returning home | -2.8543 | 0.0509 | 27128 |
| Shops | -1.9669 | 0.3013 | 8554 |
| Work | -2.2594 | 0.1854 | 19241 |

Spatial Analysis:

Distribution:

|  |  |  |  |
| --- | --- | --- | --- |
| Purpose | ADF | p-value | n |
| All | -2.7261 | 0.0696 | 185285 |
| Cafe | -2.7386 | 0.0676 | 3189 |
| Education | -2.8689 | 0.0491 | 2830 |
| health | -4.1338 | **0.0009** | 1061 |
| Leisure | -1.8601 | 0.3511 | 9379 |
| Not available | -4.7958 | **0.0001** | 25 |
| Other | -2.4963 | 0.1164 | 1219 |
| Pick a person up | -2.8686 | 0.0491 | 1592 |
| Returning home | -2.8543 | 0.0509 | 27128 |
| Shops | -1.9669 | 0.3013 | 8554 |
| Work | -2.2594 | 0.1854 | 19241 |

Global Moran’s I

pick\_up\_a\_person 0.5629000152701666 0.0 0.001

work 0.5921787733479512 0.0 0.001

shops 0.5921624780237638 0.0 0.001

returning\_home 0.6169009566510006 0.0 0.001

health 0.547517193189426 0.0 0.001

leisure 0.5448749744913084 0.0 0.001

education 0.5872240653154563 0.0 0.001

other 0.5518371624439252 0.0 0.001

cafe 0.5733876521027782 0.0 0.001

not\_available 0.4356105321751141 0.0 0.001

Count of purpose per mtl trajet KS tests:

cafe KstestResult(statistic=0.5, pvalue=0.0)

education KstestResult(statistic=0.5, pvalue=0.0)

health KstestResult(statistic=0.5, pvalue=0.0)

leisure KstestResult(statistic=0.5, pvalue=0.0)

not\_available KstestResult(statistic=0.5, pvalue=0.0)

other KstestResult(statistic=0.5, pvalue=0.0)

pick\_up\_a\_person KstestResult(statistic=0.5, pvalue=0.0)

returning\_home KstestResult(statistic=0.5567048189947053, pvalue=0.0)

shops KstestResult(statistic=0.5, pvalue=0.0)

work KstestResult(statistic=0.5, pvalue=0.0)

### 4.1.4 Travel Purpose

Purpose Correlation:

KS test vs norm for grouped 1 hour

cafe KstestResult(statistic=0.5551535507713675, pvalue=3.287213099129875e-205)

education KstestResult(statistic=0.5, pvalue=9.82488208886262e-164)

health KstestResult(statistic=0.5, pvalue=9.82488208886262e-164)

leisure KstestResult(statistic=0.7251252221594695, pvalue=0.0)

not\_available KstestResult(statistic=0.5, pvalue=9.82488208886262e-164)

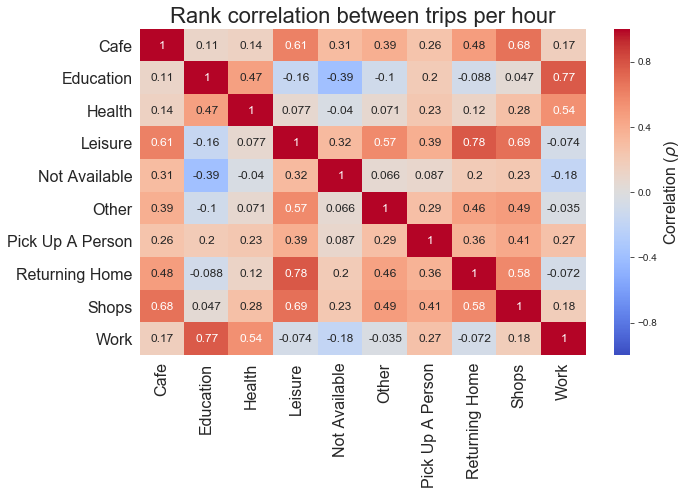
other KstestResult(statistic=0.5, pvalue=9.82488208886262e-164)

pick\_up\_a\_person KstestResult(statistic=0.5, pvalue=9.82488208886262e-164)

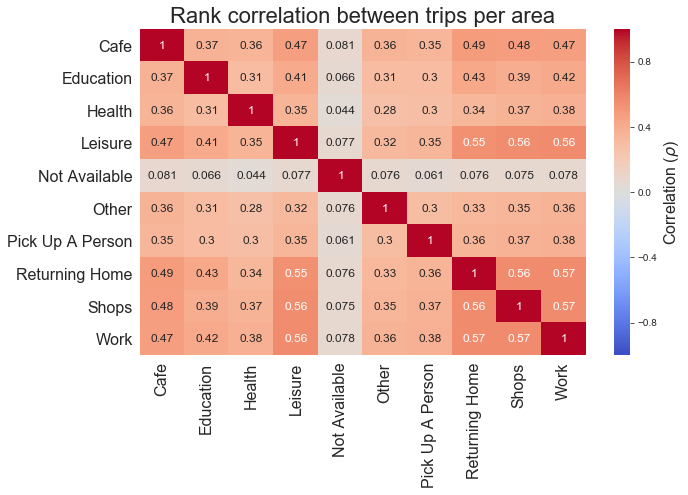
returning\_home KstestResult(statistic=0.8370232391566367, pvalue=0.0)

shops KstestResult(statistic=0.6061450521877981, pvalue=3.586226223248267e-249)

work KstestResult(statistic=0.749204542272784, pvalue=0.0)



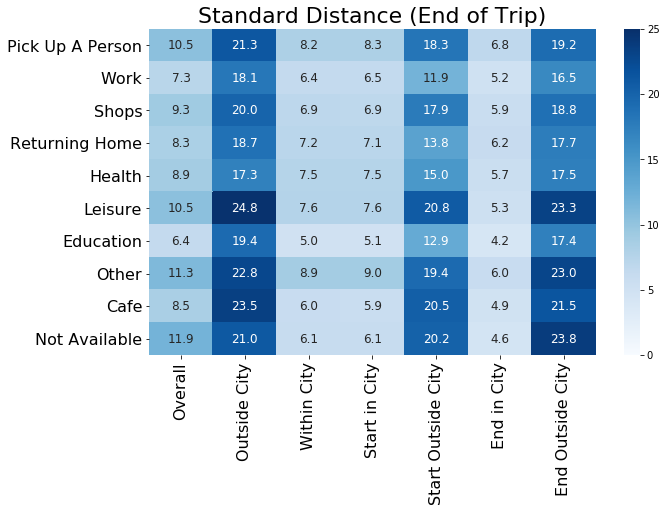
**Figure 4.X**



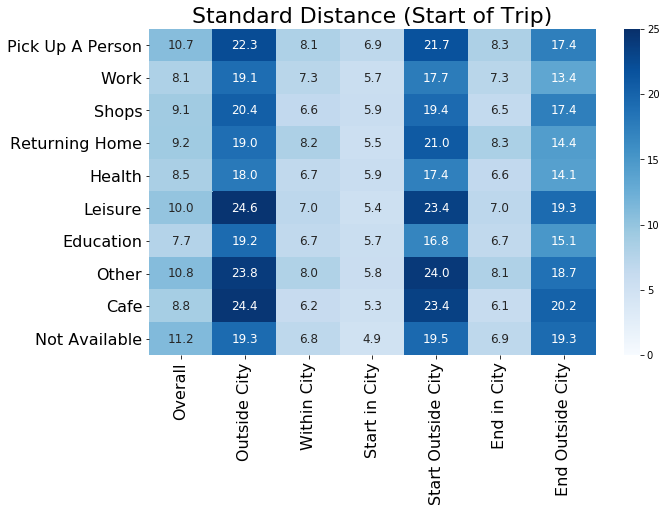
**Figure 4.X**

Start in the city cluster end

Start outside the city cluster end



**Figure 4.X**



**Figure 4.X**

## 4.3 Trip Purpose Classification Models

Based on discovery of trip purposes that share spatial and temporal dependencies we further subset the 10 unique trip purpose classes into . We also opt to remove trip classified as ‘Other’ and ‘Not Available’. It is hoped that these will (Jahromi *et al*., 2016).

Table of New Groups:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Label | Purpose Types |  |  |  |  |
| 0 |  |  |  |  |  |
| 1 |  |  |  |  |  |
| 2 |  |  |  |  |  |

Examining the forecast-ability of the network: To build upon the structure and findings of

the ESTDA.

- “omitted-variable bias (OVB) occurs when a statistical model leaves out one or more relevant variables” (i.e. purpose?)

**Table 4.X**

|  |  |  |
| --- | --- | --- |
| *Subset* | *N* | *% of total trips* |
| Rush Hour | 40,945 | 56.1 |
| Off-Peak | 32,084 | 43.9 |
| City | 68,275 | 93.5 |
| Non-city | 4,754 | 6.5 |

New Purpose labels:

Based on correlation

"1":["shops","leisure","cafe","returning\_home"],"2":["education"],\

"3":["health"],"4":["pick\_up\_a\_person"],"5":["work"]}

1 47464

5 18950

2 2769

4 1574

3 1044

*Classification of purpose of travel:*

Take 1st week -> predict then Take 2nd week -> predict then … compare accuracy scores

Sizes of data:

len(purp\_gdf), len(purp\_city), len(purp\_noncity), len(purp\_rush), len(purp\_nonrush), (71801, 66029, 5772, 39695, 32106)

CANT DO LSTM or RNN as time is not regular

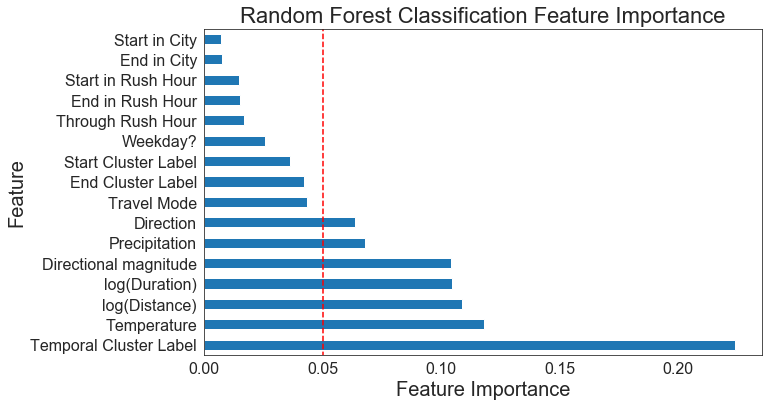
Can do CNN because of grid

*Random Forest:*

Random Forest:

ADD CONFUSION MATRIX OF PREDICTIONS FOR EACH CLASS WITH PROPER NAMES

* Feature importance
* Plot residuals



**Figure 4.X**

|  |  |  |  |
| --- | --- | --- | --- |
|  | RF |  | n |
| All | 0.8116480270099177 |  |  |
| City | 0.8131174162118273 |  |  |
| Non-City | 0.8086500655307994 |  |  |
| Rush | 0.8489446405768798 |  |  |
| Non-Rush | 0.7686380273550376 |  |  |

RF normal:

'purp\_gdf': array([0.80827185, 0.81544217, 0.8151112 , 0.81278586, 0.81742566]),

'purp\_city': array([0.81263883, 0.8150411 , 0.82103977, 0.80835463, 0.8143127 ]),

'purp\_noncity': array([0.80515298, 0.81320451, 0.80322581, 0.82200647, 0.82524272]),

'purp\_rush': array([0.84744822, 0.84464583, 0.84739179, 0.84680851, 0.84749213]),

'purp\_nonrush': array([0.76879298, 0.76944972, 0.76826376, 0.77100142, 0.77171334])

Cross-validated multi-class

{'purp\_gdf': array([0.73622947, 0.73838478, 0.73453903, 0.74275023, 0.74586841]),

'purp\_city': array([0.73694735, 0.74294601, 0.74016885, 0.72891901, 0.73780691]),

'purp\_noncity': array([0.70645161, 0.7516129 , 0.6983871 , 0.69789984, 0.73505654]),

'purp\_rush': array([0.79670736, 0.78838328, 0.7891232 , 0.80281169, 0.79278446]),

'purp\_nonrush': array([0.67220114, 0.66816888, 0.66500593, 0.67852906, 0.66476868])}

Predictions RF:

1.0 21108

0.0 2419

2.0 167

3.0 1

*Comparison of Models:*

Dimensions:

'purp\_gdf': (71801, 22),

'purp\_city': (67177, 22),

'purp\_noncity': (4624, 22),

'purp\_rush': (40342, 22),

'purp\_nonrush': (31459, 62)}

*SVC:*

cv\_scores\_rf

{'purp\_gdf': 0.7364844903988184,

'purp\_city': 0.7402228336866796,

'purp\_noncity': 0.7070773263433814,

'purp\_rush': 0.7959137684969578,

'purp\_nonrush': 0.6675014448083221}

cv\_scores\_svc

{'purp\_gdf': 0.6609411268200042,

'purp\_city': 0.6556903784564031,

'purp\_noncity': 0.7450851900393185,

'purp\_rush': 0.7543003079696537,

'purp\_nonrush': 0.7545752263533038}

cv\_scores\_nn

{'purp\_gdf': 0.7864950411479215,

'purp\_city': 0.7898867788353106,

'purp\_noncity': 0.7234600262123198,

'purp\_rush': 0.8369263126267558,

'purp\_nonrush': 0.7356963976112503}

num\_dims

{'purp\_gdf': (71801, 23),

'purp\_city': (67177, 23),

'purp\_noncity': (4624, 23),

'purp\_rush': (40342, 7),

'purp\_nonrush': (31459, 63)}

Hyper Parameter for SVC:

For the purposes of the classification models, the data is normalised for the purpose of the

Cs = [0.1, 1, 10]

gammas = [0.01, 0.1, 1]

Best == {'C': 0.1, 'gamma': 0.01}

SVC:

ADD CONFUSION MATRIX OF PREDICTIONS FOR EACH CLASS WITH PROPER NAMES

Results on the test set: RF:

precision recall f1-score support

0.0 0.00 0.00 0.00 0

1.0 0.85 0.89 0.87 15661

2.0 0.00 0.00 0.00 955

3.0 0.00 0.00 0.00 322

4.0 1.00 0.00 0.00 510

5.0 0.76 0.57 0.65 6247

accuracy 0.74 23695

macro avg 0.43 0.24 0.25 23695

weighted avg 0.78 0.74 0.74 23695

Results on the test set: SVC:

precision recall f1-score support

1.0 0.66 1.00 0.80 15661

2.0 0.00 0.00 0.00 955

3.0 0.00 0.00 0.00 322

4.0 0.00 0.00 0.00 510

5.0 0.00 0.00 0.00 6247

accuracy 0.66 23695

macro avg 0.13 0.20 0.16 23695

weighted avg 0.44 0.66 0.53 23695

Results on the test set: NN:

precision recall f1-score support

0.0 0.00 0.00 0.00 0

1.0 0.83 0.93 0.88 15661

2.0 0.00 0.00 0.00 955

3.0 0.00 0.00 0.00 322

4.0 0.00 0.00 0.00 510

5.0 0.73 0.64 0.68 6247

accuracy 0.79 23695

macro avg 0.26 0.26 0.26 23695

weighted avg 0.74 0.79 0.76 23695

NN:

10,000 iterations/epochs for NN

ADD CONFUSION MATRIX OF PREDICTIONS FOR EACH CLASS WITH PROPER NAMES

Best estimator:

ctivation='relu', alpha=0.1, batch\_size='auto', beta\_1=0.9,

beta\_2=0.999, early\_stopping=False, epsilon=1e-08,

hidden\_layer\_sizes=5, learning\_rate='constant',

learning\_rate\_init=0.001, max\_iter=500, momentum=0.9,

n\_iter\_no\_change=10, nesterovs\_momentum=True, power\_t=0.5,

random\_state=3, shuffle=True, solver='lbfgs', tol=0.0001,

validation\_fraction=0.1, verbose=False, warm\_start=False

'alpha': 0.01,

'hidden\_layer\_sizes': (50, 50, 50),

'max\_iter': 500,

'solver': 'lbfgs'

Predictions NN:

1.0 23136

0.0 559

CNN:

### 4.3.1 Spatial and Temporal Dependency in Error Terms:

After classification of trip purpose, we examine the spatial and temporal distribution of any mis-classified trips.

Error Terms & Scores

len(all\_true), len(no\_true), len(only\_svc), len(only\_rf), len(only\_nn), len(all\_but\_svc),len(all\_but\_rf),len(all\_but\_nn)

(13773, 3801, 857, 218, 588, 3427, 848, 183)

[Example results writing] “Most of the misclassified trip were …]

What I am trying to show:

* Idea about the MTL Trajet
* What specs for purposes (i.e. which modes, where, etc.)
* Space, time and space-time trends
* Classification results (how well we can classify purpose and most important things)