Can we predict why people travel within a city?: A study analysing the spatial and temporal characteristics of travel intention within Montréal, Canada between September and October 2017.

*This dissertation is submitted in part requirement for the MSc in the Centre for Advanced Spatial Analysis, Bartlett Faculty of the Build Environment, UCL.*

*A colour version of this dissertation is available online.*

Word Count:11,986

Thomas J. Keel – 18110348

CASA0004: MSc Spatial Data Science & Visualisation University College LondonSupervisor: Huanfa Chen

*30th August 2019*

# Abstract

The quantification of trip purpose within movement in cities, remains an area within broader mobility studies without an extensive investigation. In the past, this investigation has been hindered by two factors:

(1) the absence of relevant data which details the purpose of people’s travel in a city;

(2) the difficulty in representing the space-time dynamics of mobility within models/metrics. Regarding (1), in recent years, smartphones travel surveys have provided researchers a platform to study travel within a city on a large scale. In turn, this has fuelled an eruption of volunteered geographic information (VGI), which provides qualitative information about travel at increasingly finer temporal and spatial scales in cities. This study makes uses of one such study: the *2017 MTL Trajet* project – a survey examining travel behaviour and intent patterns across Montreal between 18th September 2017 and 18th October 2017. Results from this survey provide insight into how and why people travel and could be used to inform the better planning of essential and non-essential services at the city-level. This project builds on a small body of research into methods that uncover spatial and temporal interdependencies of data before assessing the ability of three distinct classification models: Support Vector Machines, Random Forest and Artificial Neural Networks used to predict the trip purpose.

**Key Words:** Trip Purpose, Mobility, Volunteered Geographic Information, Spatio-Temporal Investigation, Classification Modelling.

# Declaration

I, Thomas Keel, hereby declare that this dissertation is all my own original work and that all sources have been acknowledged. It is 12,000 words in length

Signed:

Date: 28th August 2019

Chapter 2. Literature Review

Chapter 3. Methodology

Chapter 4. Results

Chapter 5. Discussion

Chapter 6. Conclusion

Bibliography

Appendix 1. Notification of intent not to apply for ethical approval

Appendix 2. Link to scripts used for the analysis

Table of Contents

[Abstract 1](#_Toc17552552)

[Declaration 2](#_Toc17552553)

[List of Figures 3](#_Toc17552554)

[List of Tables 4](#_Toc17552555)

[List of Acronyms and Abbreviations 4](#_Toc17552556)

[Acknowledgments 5](#_Toc17552557)

[Chapter 1. Introduction 6](#_Toc17552558)

[1.1 Research Overview and Questions 6](#_Toc17552559)

[1.2 Motivation 7](#_Toc17552560)

[1.3 Approach 8](#_Toc17552561)

[1.4 Outline 10](#_Toc17552562)

# List of Figures

**Figure 1.1** Screenshot from the MTL Trajet App (left; source: Patterson, 2017) and example of prompt within Itinerum apps (right; source: Patterson *et al.*, 2019)………………………..……….1

**Figure 2.1** Comparison of Trip Purpose classification model accuracy within the literature (ANN=Artificial Neural Network; SVM=Support Vector Machine)…………………………………………2

**Figure 2.2** Framework of the Itinerum Mobile and Desktop app (Source: Patterson & Fitzsimmons, 2017)…………………………………………………………………………………………………...…………2

**Figure 2.3** Example of GPS trace with location collection priorities within an Itinerum platform app (A); Example of the on screen prompt after an Itinerum platform app stops recording movement (B) (Source: Patterson *et al.*, 2019)……………………………………………………..4

**Figure 2.4** Example of *Map Matching* done by the Open Source Routing Machine when processing the raw GPS trace from user devices (Source: Hamouni, 2018)…………………………..4

**Figure 3.1** Montreal within Quebec, Canada………………………...………………………..……………………5

**Figure** **3.2** GPS routes from the MTL Trajet plotted within the study area…………...………...……5

**Figure 3.3** Location of Montreal within the study area. ……………………………………………………….5

[**Figure 3.4** Bar chart showing the count of each category of land use]

**Figure 3.5** Map showing land use changes within the City of Montreal………..…………...…………5

**Figure 3.6.** Circular contour plot (windrose; left) and circular histogram (right) showing the direction of trips (circle bands indicate count of trips)

**Figure 3.7** Example of the spatial join between a route and the underlying dissemination areas (route in **blue**; overlapping dissemination areas in **red**)

**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.

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

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

**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.

**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.

**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.

**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.

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

**Figure 4.14** Calendar plot showing the weighted importance of each ‘temporal word’ in each of the temporal clusters (1-5).

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

**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**).

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

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

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

# List of Tables

**Table 3.1** Description of the key variables from data from the MTL Trajet survey before pre-processing.

**Table 3.2** Categories of mode and purpose allowed for trips in the MTL Trajet survey.

**Table 3.3** Description and cover of Land Use categories within the City of Montreal.

**Table 3.4** Definition of Rush hourand Off-peak hours used in this study.

**Table 3.5** Description of the key variables from the MTL Trajet survey after pre-processing.

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

**Table 4.2** Summary statistics of overall distance and duration of trips from the 2017 MTL Trajet travel survey

**Table 4.3** Summary statistics of trip distance and duration per trip purpose

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

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

# List of Acronyms and Abbreviations

**ANN** – Artificial Neural Networks

**CNN** – Convolutional Neural Networks

**DA** – Dissemination Areas

**GPS** – Global Positioning System

**LDA** – Latent Dirichlet Allocation

**MAUP** – Modifiable Areal Unit Problem

**MLP** – Multi-Layer Perceptron

**MTUP** – Modifiable Temporal Unit Problem

**RF** – Random Forest

**SVM** – Support Vector Machine

**VGI** – Volunteered Geographic Information

# Acknowledgments

I would like to thank my family for their thorough support

Secondly, I would like to thank my supervisor Huanfa for their comments and advice and the staff at CASA for the support this year.

Finally, I am extremely grateful for those people I have met throughout this year at CASA, who have

Especially my flat mates James and George.

All dissertations are submitted electronically, but we also require ***two printed and bound copies to be submitted as well***. Double spacing must be used, except for indented quotations, tables, bibliographies and footnotes, which should be single-spaced. The left- side margin should be not less than 40mm (1.5 inches) and other margins not less than 20mm (0.75 inch).

# Chapter 1. Introduction

## 1.1 Research Overview and Questions

The purposes by which people use transport networks on a large scale remains an area with a distinct lack of investigation within the broader mobility studies (Yazdizadeh *et al.,* 2019). This is primarily due to an absence of comprehensive data available to study trip purpose at such a scale.

In recent years, improvements to smartphone technology has provided an opportunity to study and record geospatial movement of people on an increasingly large scale, with smartphones being able to better record similar mobility behaviour as their carriers (Jahromi *et al.*, 2016). Mobile survey apps created to make use of this have provided researchers a platform to collect qualitative information about movement for example how and why people are moving (Li et al., 2016). This ability to create a large amount of geo-referenced data from these surveys – which can be referred to as Volunteered Geographic Information (hereafter, VGI) – can help us generate unique insight into a population’s transport behaviour patterns throughout a city and throughout time.

Arguably, spatial and temporal information provided from this form of VGI can be integrated into city-level decision-making to help inform planning a variety of essential and non-essential services (Attard *et al.*, 2016). For example, if we knew that people tended to cycle to cafés during lunch breaks, policy could be implemented to introduce bike racks near the cafés.

Despite this, there is still a gap in knowledge in understanding of travel intent in most cities. This is owing to the fact that no in-city research has been initiated there (ref). One exception is in Montreal, where a number of mobile applications have been created in recent years (since 2016) to study how people move across the city based on their smartphones. This report makes use of the most recent available data from one of these studies: The *2017 MTL Trajet* study carried out between 18th September 2017 and 18th October 2017.The data is used in this study to assess an over-arching research question along with two sub-questions:

**Main Research Question:**

Can we effectively classify the purpose of trips using spatial and temporal indicators from the same trips?

**Sub-Questions:**

1. Which spatial and temporal indicators are most important for the classification of trip purpose?
2. Which type of classifier model is most effective in the classification of trip purpose?

## 1.2 Motivation

Movement can be thought of as an interaction between an origin and destination (Murray *et al.*, 2012). People move across space and through time to go from where they are to where they want to be. Transport, on the hand, is a by-product of the demand for movement i.e. it is best considered as a derived demand for a particular destination (Golledge & Gärling, 2001). Studying the purpose of movement individual’s travel between an origin and destination within a transport network underpins our comprehension of the human behaviour within a city (Kwan & Neutens, 2012). If we are able to discern the activities for which people travel to, and at which temporal and spatial scales they are travelling at, we may be able to better plan and manage essential and non-essential services within a city (i.e. transport infrastructure, leisure activities, etc.).

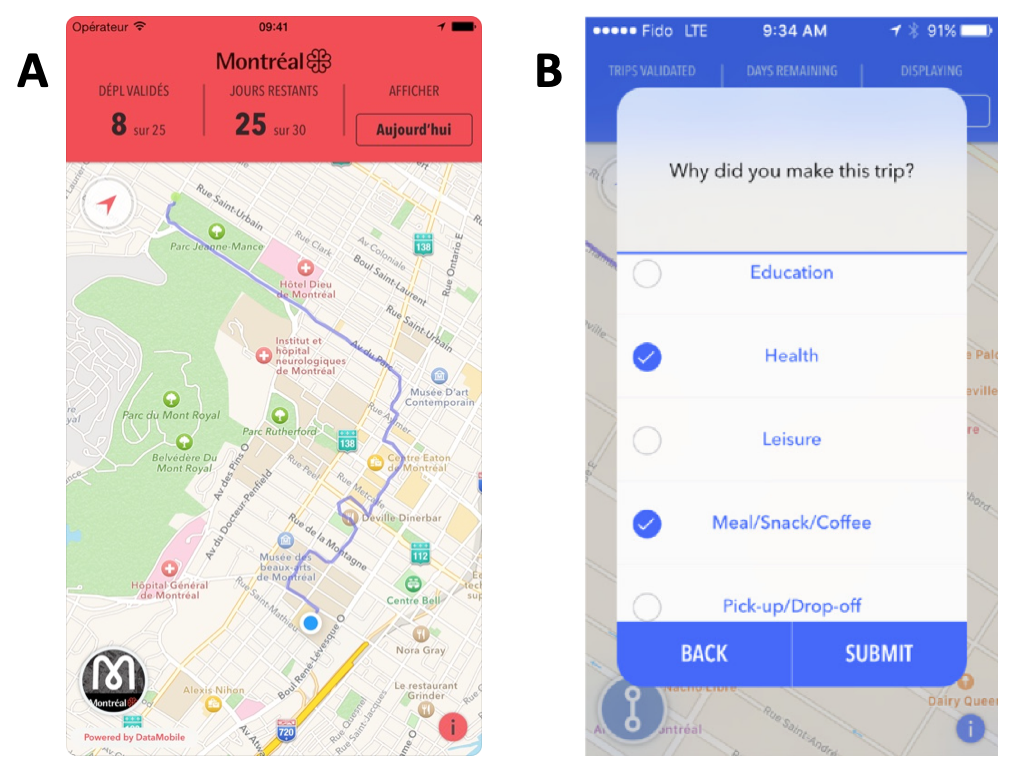
* “Getting a deeper understanding of human mobility is a prerequisite for a broad range of possible studies on smart cities and related research areas”. (Xie *et al.*, 2016)
* Understanding the interactions within complex system such as a city is a prerequisite for predicting changes within it… big data offers us an opportunity to study this (Cheng *et al.*, 2017)
* :”In practice, people’s trip purposes are very important in understanding travel behaviors and estimating travel demands” (Meng *et al.*, 2019)

The motivation of this study is thus to deconstruct the spatial, temporal and spatial-temporal profiles of trip purposes and try to model them.

Big geographic data allows us to not only study the spatial and temporal interactions but also interactions of socio-economic factors [this is what this research aims to do] (Cheng *et al.*, 2017).

## 1.3 Approach

This study makes use of data from the *2017 MTL Trajet* survey originally collected by researchers at the Transportation Research for Integrated Planning (TRIP) lab, Concordia University (Patterson & Fitzsimmons, 2017). This survey was part of the 2015-2017 Montréal Smart and Digital City Action Plan and was created to study travel behaviour across the city (MTL Trajet, 2017). Data collection for this survey was carried out through a mobile app which automatically recorded a location trace using Global Positioning System (hereafter, GPS) provided from a user’s phone (**Figure 1.1A**). When users were stopped in a given location for more than intervals of 120 seconds the app would prompt the user to end the trip and insert a reason behind why they had made this trip. An example of a similar app using the same framework is shown in **Figure 1.1B**.

****

***Figure 1.1*** *(A) Screenshot from the MTL Trajet app showing recorded GPS trace (source: Patterson, 2017). (B) Example of prompt similar to one used in the MTL Trajet app (source: Patterson et al., 2019).*

Data from the *MTL Trajet* forms the backbone of the information used to test classification models that look to characterise the purpose of movement. Temporal and spatial clustering techniques will be used to simplify the space and time profiles of the trips before analysing the …

It is hoped, in combination with a spatio-temporal investigation, the analysis presented can infer something about movement at a higher scale within a city.

## 1.4 Outline

The following chapters of the report are organised as follows:

*Chapter 2* reviews literature relating to trip purpose classification, the use of VGI in mobility studies and the use of the MTL Trajet Project.

*Chapter 3* details the steps carried out in the data pre-processing and analysis procedures for this report.

*Chapter 4*, presents the results from the analysis procedure and classification models. Finally, *Chapter 5* and *6* discuss and draw conclusions from the results, highlight uncertainty within the analysis procedure and suggest further research around the development of trip purpose classification models.

Extra:

The improvement of mobile networks and the geolocation abilities of smartphones in recent years, means we currently have an opportunity to study cities through data resulting from mobile GPS traces (Li *et al.*, 2016; Patterson *et al.*, 2019).

Mobile phones as sensors (ref)