# Abstract (250):

The prediction of *why* people travel when they move across cities remains an area within the broader mobility studies without extensive investigation. Arguably, this has been hindered by: (1) an absence of large datasets which detail the purposes of individual’s travel across a city; (2) the difficulty in accurately representing space and time within models used to predict

why people travel across cities.

Regarding (1), in recent years, Volunteered Geographic Information (VGI) provided by smartphones travel surveys have provided researchers an opportunity to study the attributes characterising urban mobility patterns within a city at increasingly fine temporal and spatial scales. This study makes uses of one such source of VGI: the *2017 MTL Trajet* travel survey app – a project with the aim to study *how* and *why* people move within the City of Montreal, Canada. Regarding (2), this project builds upon a small body of research to uncover and categorise spatial and temporal interdependencies of GPS data provided from the MTL Trajet project, before assessing the performance of three machine-learning classification models used to classify this GPS data: Random Forests, Support Vector Machines and Artificial Neural Networks. Specifically, these models are built to classify *why* people travel based on spatial and temporal characteristics of individual trip.

**Key Words:** Travel purpose classification, Mobility, Random Forest, Spatial and Temporal Error, Volunteered Geographic Information.

# Introduction (800-1000):

The purposes by which populations 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). In the past, this has primarily been due to an absence of large data sets which combine both the geographically coordinates of people’s movement (i.e. a GPS trace) and the activities for *why* people make these movements (i.e. for Work, Leisure, etc).

In recent years, improvements to GPS within smart-phones has provided researchers a new opportunity to study and record the large scale geospatial movement of people (Zhao *et al.*, 2019). Travel survey apps created for smart phones require much less effort from their participants than traditional travel surveys (i.e. where a separate GPS device is required to record movement) (Li et al., 2016). Therefore, it has become increasingly easy to collect qualitative information about movement within a city – including information about *how* and *why* people travel.

The ability of smartphone users’ to create a large amount of geographically-referenced data in these travel survey apps can help researchers generate unique insight into transport behaviour at much finer scales than ever before. This form of participatory data creation is known as Volunteered Geographic Information (hereafter, VGI) (after Goodchild, 2007).

Despite the potential to produce more VGI that can be used to generate insight into urban mobility patterns within a city, there are many cities globally that have no form of formal research initiated within them (Attard *et al.*, 2016). One exception to this, is Montreal, Canada, where a number of mobile travel survey applications have been created to study *how* and *why* people move along the city’s transport network.

This report makes use of the most recent available dataset from one of these studies: The *2017 MTL Trajet* travel survey project (Ville de Montréal) The *MTL Trajet* project was carried out between 18th September 2017 and 18th October 2017 and is used in this dissertation to following assess the following research questions:

*Main Research question:*

Can we effectively classify the purpose of trips without qualitative information from respondents?

*Background:*

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, is the by-product of the interaction between an origin and destination, and can thus is best considered a ‘derived demand’ for a given destination (Golledge & Gärling, 2001). Studying the patterns in the types of destinations that people demand to travel to, thus, underpins our comprehension of behavioural patterns within a city (Kwan & Neutens, 2012).

If we are able to discern the activities for which individual’s make movements (referred to as their ‘*trip purpose’*), we may be able to use this information to inform policy and account for demand in essential (e.g. health & educational services) and non-essential (e.g. leisure & commercial) services throughout a city (Attard *et al.*, 2016).

To better understanding and classify trip purpose, we first need to understand the temporal and spatial scales at which people travel for certain activities. The motivation of this study is thus to evaluate whether we can use spatial and temporal dependencies as key indicators for trip purpose classification models.

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, 2017a). 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 (available on both iOS and Android platforms) which automatically recorded a location trace using GPS provided from a user’s phone (**Figure 1.1A**; Patterson & Fitzsimmons, 2017a). When users were stopped in a given location for more than intervals of 120 seconds the app would prompt the user to ‘complete’ that trip and would be asked:

-  ‘Which travel modes did you use for this trip?’

-  ‘Why did you make this trip?’ (see similar example in **Figure 1.1B**).

Both the responses to these questions and GPS trace available in the data from the MTL Trajet are used to test three types of classification models that look to characterise the *purpose* of movement. Various temporal and spatial generalisation techniques will be used to represent the space and time trends seen within the data before being used in the models as predictors for trip purpose.

*Why this study is different:*

This study differentiates itself from other research by focusing on trip purpose identification without extensive qualitative information. For example there is no information about the respondent available in the dataset used in this study, the data also excludes information about the respondents age, employment status or which trips where carried out by the same app user. The findings of this research may then be useful for other anonymous trip-purpose related datasets.

*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 MTL Trajet survey.  
*Chapter 3* details the steps carried out in the data pre-processing and collection, the development of space and time metrics from the MTL Trajet data, and the set-up for each trip-purpose classification model.  
*Chapter 4*, presents the results from the analysis procedure and compares the performance of the classification models.  
*Chapter 5* discusses the extent to which the research objectives (set out in 1.1) have been achieved in the results and highlights uncertainty within them the analysis procedure. Finally, *Chapter 6,* draws conclusion from the research carried out in this project and suggests areas of further research.

# Existing Studies (1000-1500):

Used models fall within the following categories: rule-based, probabilistic,machine learning.

What have other studies done!

Literature review of all existing literature summarized in a table.

Current focus in the literature:

There exists a focus on trip purposes for small datasets tied to one city/urban region at a given time.

Also research that is

State gaps in the literature:

Examination of spatiality and temporality of the error terms.

Oversampling method.

Removing work and home.

(for each of above: state use in other forms of transport research)

Data and variables used in literature:

p\_nearby, POIs, various forms of spatial and temporal information.

In this study we remove work and returning home trips, (these classes by far out-balance the other ones). ‘Eating out’, ‘Personal’, ‘Cafe’, ‘Work’ previously used in the literature (TABLE of which variables have been used).

# Data:

*MTL Trajet*

Detail the study, its findings and the variables.

Contains: mode, purpose, spatial and temporal signature

*ERA-5 (weather):*

Why use? Link to other transport research

*Extraction of other information:*

Why not land-use, POI or clusters.

# Research Design and Methodology:

- remove duration and distance outliers

- remove modes not shared between years

# Model estimation:

# Accuracy assessment:

# Conclusions: