# Notes:

* 1500-2000 words

## 3. Methodology:

*3.1 Study Area*

The study area chosen for this project spans across the Greater Montreal region. Shapefiles outlining the geographical boundaries of the Greater Montreal region, (Statistics Canada, 2016; Statistics Canada, 2019).

QGIS software was used to select

To-Add:

* About Montreal and its POI
* Shapefiles and selection in QGIS
* Shapefiles: greater, dissemination areas,
* Number of arrondissements
* Translated map of data?

*3.2 Data:*

Data were retrieved from the MTL Trajet data available on the Montreal Open Database (Portail données ouvertes Montréal) for the years 2016, 2017. This data is derived from an questionnaire app (*MTL Trajet*, ref). The MTL Trajet survey was conducted from September 18 to October 15, 2017

and included X amount of participants

|  |  |  |  |
| --- | --- | --- | --- |
| Column | Data Type | Spatial Units | Temporal Units |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Spatial join was used in QGIS to select the dissemination across the Montreal region where the study passed into.

*3.2.1 Translation:*

For purposes of

Time from UTC to local

*3.2.2 Direction, Distance and Speed*

set one-hour intervals, peak/off-peak and workday/weekends labels (after Liu & Cheng, 2018)

The data were then subset

*3.2.3 Development of Metrics:*

Directionality

Temperature:

Temperature and precipitation data for the dates (18th)were retrieved ERA-5 1\*1 degree average over the area of Montreal per hour (precipitation and 2m temperature). Reanalysis data . 1464\*1\*1

*3.2.4 Outlier Detection:*

Is a form of knowledge discovery (KD)﻿. Anomaly detection is inherently challenging as it requires a clear definition of what is considered to be normal and abnormal (Li *et al.*, 2016)

*3.3 Exploratory Spatial-Temporal Data Analysis*

[On visualisation of big geodata] We should carefully generalize, e.g., emphasize the important while removing the unimportant, (Li *et al.*, 2016)

Reason for: Insight into those spatial and temporal trends can improve the performance of Intelligent Transportation Systems (ITS).(Tayyab *et al.*, 2014)

*3.3.1 Spatial*

[intersection between network and dissemination areas for grouping]

*3.3.2 Temporal*

*3.3.3 Spatio-Temporal*

Merging of other pertinent data (i.e. land-use)

LISA:

Merge the data into the dissemination areas

[Clusters] not only suggests characteristics of the pattern itself but also of its background processes. (Yamada & Thill, 2010)

ST-KDE:

See Wei *et al.* (2018) (Also uses grid-ified data)

*3.4 Modelling:*

*3.4.1 Latent Dirichlet Allocation*

*3.4.2 Classification*

Significant class-imbalance exists in the MTL Trajet data (show histogram of classes)

Spatial regression between key areas

difficult to quantify space-time clusters → At what point does a cluster of crimes become a hotspot? (Li *et al.,* 2016)

Including external sources: POI and importance

After Ren *et al.* (2019) does both spatial and temporal distribution of errors

Directionality model

*3.5 Limitations:*

*3.5.1 Data*

Representativeness present in all forms of VGI, especially this one

Difficult to falsify VGI (Elwood *et al.*, 2012)

Compared with data from professional vendors, spatial UGC [VGI] faces greater authenticity issues, such as mistaken data due to carelessness or inadequate (Shi *et al.*, 2018)

*3.5.2 Methodological*

Vital to consider all forms of uncertainty in a spatial big data investigation (Shi *et al.*, 2018)

Higher accuracy of deep learning classification relies on much larger training data sets with labelled classes than conventional classifiers (Shi *et al.*, 2018)

Weather has impact on mode (Xie *et al.*, 2016)