# Notes:

* 1500-2000 words

## 3. Methodology:

This section first details the major data processing phases involved with handling the data including cleaning, translation.

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

*3.1 Study Area*

Translated data?

*3.2 Data Pre-processing:*

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).

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

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 Column Creation*

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

*3.3 Exploratory Spatial-Temporal Data Analysis*

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 Development of Metrics:*

Directionality

Merging of other pertinent data (i.e.

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)

*3.3.2 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)

*3.4 Modelling:*

*3.4.1 LDA*

*3.4.2 Other*

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 Methodological Limitations*

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)

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)