WCC

Table of Contents

[1. Introduction 1](#_Toc173446775)

[2. Literature Review 1](#_Toc173446776)

[3. Methodology 1](#_Toc173446777)

[3.1 Data and Methodology Workflow 1](#_Toc173446778)

[3.2 Data re-classification 2](#_Toc173446779)

[3.3 Clustering 10](#_Toc173446780)

[3.4 Data matching 10](#_Toc173446781)

[3.5 Access urban accessibility 11](#_Toc173446782)

[3.5.1 Accessibility measures 11](#_Toc173446783)

[3.5.2 r5r r package 12](#_Toc173446784)

[4. Results 12](#_Toc173446785)

[4.1 Clustering 12](#_Toc173446786)

[4.2 Data matching 14](#_Toc173446787)

[4.3 Urban accessibility 14](#_Toc173446788)

[5. Discussion 14](#_Toc173446789)

[5.1 Implications 14](#_Toc173446790)

[5.2 Next step 14](#_Toc173446791)

[5.3 Limitations 14](#_Toc173446792)

[6. Conclusion 14](#_Toc173446793)

# Introduction

# Literature Review

# Methodology

## Data and Methodology Workflow

The research methodology has been divided into specific sub-processes, each addressing various objectives crucial to the final research result (Figure !). Table ! lists all datasets and their sources, along with the sub-processes where they were applied. The subsequent sub-sections will detail the methodology of each process within the methodology schematic.

A diagram of a diagram

Description automatically generated with medium confidence

Figure !: Research Methodology Outline, divided into separate parts of 1.) data re-classification based in SIC, 2.) clustering for different cultural sections, then 3.) data matching of poi point data to spatially referenced building polygons, and finally 4.) generate the urban accessibility of each cluster



Table !: Sources, information extracted, and application of data used in methodology, with number referencing the methodological stage depicted in Figure 1

Add a subsection like Study area?

## Data re-classification

The original Point of Interest (POI) dataset encompasses all POI data within the Westminster area. However, since this paper specifically examines the cultural aspect, it was necessary to filter the POI data accordingly. The selection of culture-related POIs was based on documents provided by Westminster City Council (WCC), specifically the Creative Sector Standard Industrial Classification (SIC) classifications.

During the selection process, it was observed that the original POI data adhered to the Digimap grouping method, which is organized into three hierarchical layers: The first layer comprises nine groups; the second layer includes 52 categories, and the third layer contains more than 600 classes. However, it was noted that the scope of some Digimap classes is significantly broader than the corresponding SIC classifications. For instance, Digimap consolidates all publications under a single class 'Published Good,' whereas the SIC provides a more detailed breakdown, including books, newspapers, journals, and other types of publications.

To address these discrepancies, each POI within this kind of relatively broader Digimap classes was individually examined and reclassified according to SIC classifications, based on their commercial content and records available on gov.uk. The Digimap classification report is accessible in the zip file when downloading poi data, and the SIC classification report refers to the UK SIC 2007 version available in PDF format

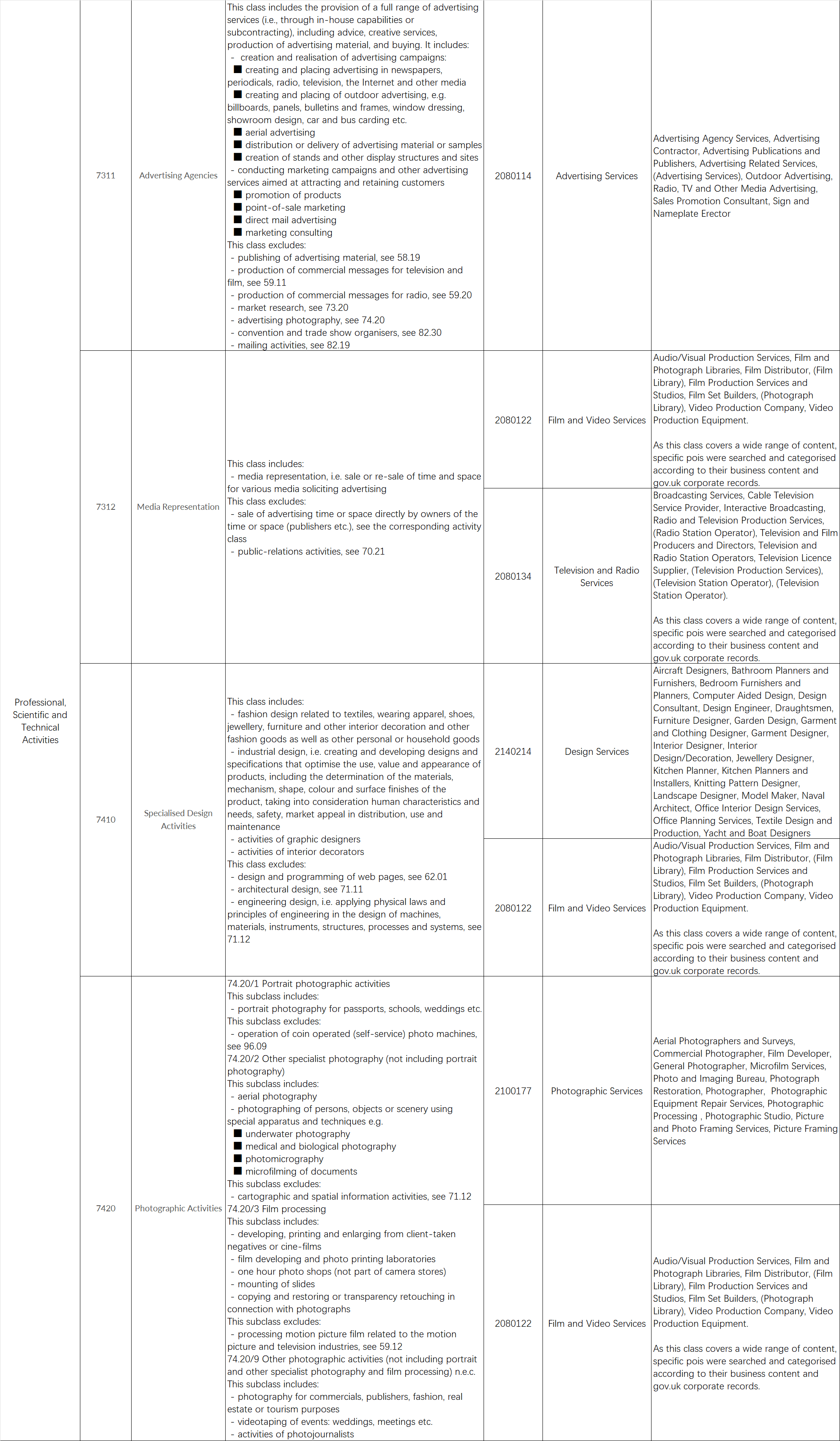
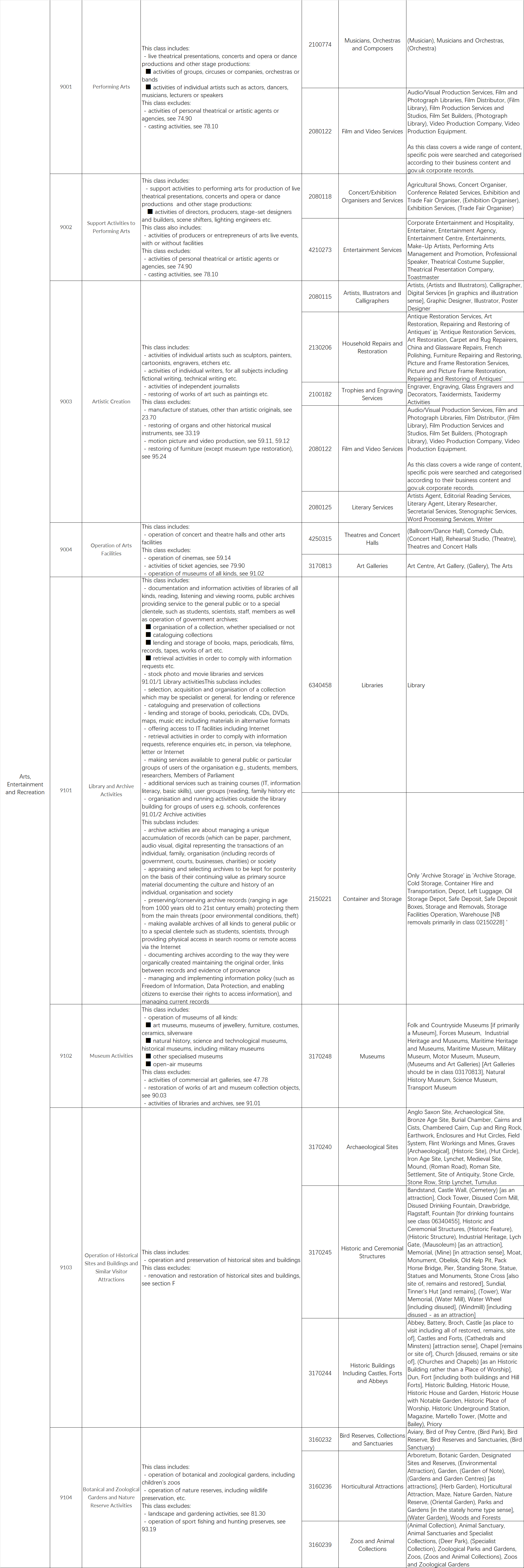
Table ! covers a detailed comparison of the two sets of classifications. (Consider putting this table in the appendix )

Table! : Detailed comparison between Digimap classification scheme and the SIC classification scheme

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA white paper with black text

Description automatically generated

## Clustering

All the culture-related poi data is re-classified into 4 sections. Subsequently, Ripley's K test is performed on each dataset segment to ascertain whether the data exhibit clustering.

Ripley's K test compare the observed distribution of points with the Poisson random model for a whole range of different distance radii. It can be expressed as this formula:

A black and white symbol

Description automatically generated

Need explanation of the formula?

Figure ! shows the graphs of Ripley’s K test for each section. The Kpois(r) line in Red is the theoretical value of K for each distance window (r) under a Poisson assumption of Complete Spatial Randomness. The Black line is the estimated values of K accounting for the effects of the edge of the study area. Where the value of K falls above the line, the data appear to be clustered at that distance. Where the value of K is below the line, the data are dispersed.

A graph of a number of people

Description automatically generated with medium confidence

Figure! : Ripley’s K test graphs for Manufacturing section, Information and Communication section, Professional, Scientific and Technical Activities section and Arts, Entertainment and Recreation section

If clustering is detected, a DBSCAN analysis is conducted on the respective dataset segment to identify and obtain detailed information about the specific clusters for each section data.

## Data matching

Once the clustering is obtained, each cluster is given a hexagon grid and get its centroids csv file through QGIS. Then accessibility accessed is between the centroids of the hexagon for the whole of London and the centroids for these clusters.

## 3.5 Access urban accessibility

### 3.5.1 Accessibility measures

A diverse array of measures is created by researchers due to the request for easily communicable, methodologically sound, and computationally feasible accessibility metrics (Páez, Scott, & Morency, 2012). These metrics are broadly categorized into two main types: place-based measures and person-based measures (Dijst, Jong, & Van Eck, 2002). Place-based measures assume that everyone within a specific location has equal access to activities distributed throughout the city, while person-based measures take into account the influence of individual characteristics. Due to the challenges associated with obtaining individual-level data, this paper will focus on place-based measures. For placed-based measures, the most commonly used accessibility metrics in the scientific paper and by transport agencies are as follows:

#### 3.5.1.1 Minimum travel cost

This metric refers to the lowest cost required to reach the nearest opportunity from a given origin, such as the travel time from one community to the closest school. This can be calculated with:



The measure requires minimal data and is straightforward to calculate and convey. However, it has two drawbacks: it neither considers the quantity of accessible opportunities at the destination nor accounts for competition for those opportunities.

#### 3.5.1.2 Cumulative opportunity measures

This metric measures the number of opportunities that can be reached within certain travel time limit, such as the number of health centers accessible by walking up to 30 minutes. This is the formula for this metric:

A group of math equations

Description automatically generated

Since it is easily communicable and requires little data, it is one of the most commonly used indicators by transport agencies in accessibility analysis (Papa et al., 2015; Boisjoly and El-Geneidy, 2017). However, like the minimum travel cost, it also ignores the competition for opportunities.

#### 3.5.1.3 Gravity measures

Similar with the cumulative opportunities measure, gravity measures count the number of opportunities accessible from a certain region. However, the number of opportunities is discounted with higher travel cost. In this case, this metric considers the ease of reaching different opportunities. The discount rate is determined by the decay function. This is the formula for this metric:



The decay functions have many options including linear and exponential decay function. A list of decay functions commonly used by transport agencies and researchers are provided by Levinson and King (2020).

This metric successfully takes into account the fact that people usually prefer the closer opportunity under the same conditions. In contrast, the discount rate makes it more difficult to interpret the accessibility level obtained and extra data is needed to make the discount rate more reliable to predict reality.

#### 3.5.1.4 Accessibility measures with competition: floating catchment area

The floating catchment area (FCA) family of indicators could be representative of this metric. These indicators take into account the competing demand.

The main advantage of this metric is that it incorporates competition into accessibility estimates, but similar with the gravity measures, the results can be hard to interpret.

### 3.5.2 r5r r package

In this paper, accessibility is estimated by the r5r r package. For this package, the data to be used include OSM's street network.pbf, public transport's GTFS data and topography. tif data. After collecting the data, travel time matrix will be calculated first. Then, the accessibility from each origin will be calculated and visualized, taking into consideration the transport costs between origin-destination pairs and the number of opportunities in each destination.

# Results

## Clustering

From the graphs of Ripley’s K (Figure !), the data for manufacturing section seems to be dispersed. Consequently, the DBSCAN analysis is conducted for Information and Communication section, Professional, Scientific and Technical Activities section and Arts, Entertainment and Recreation section.

Figure ! is the cluster information for Information and Communication section.

eps = 500

Minpts = 4

A diagram of a map

Description automatically generated with medium confidence

Figure !: The cluster information for Information and Communication section.

Figure ! is the cluster information for Professional, Scientific and Technical Activities section.

A map of a country with different colored triangles

Description automatically generated

Figure !: The cluster information for Professional, Scientific and Technical Activities section.

Figure ! is the cluster information for Arts, Entertainment and Recreation section.

A map of a country with red triangles

Description automatically generated

Figure !: The cluster information for Arts, Entertainment and Recreation section.

## Data matching

## 4.3 Urban accessibility

It's stuck. Looks like a GTFS data issue, error message says rJava is having trouble linking to R and reading in the data, still working on it. If you have any advice, really appreciated!

‘Error in .jcall("RJavaTools", "Z", "hasField", .jcast(x, "java/lang/Object"), :

java.lang.NullPointerException: Cannot invoke "Object.getClass()" because "<parameter 1>

# Discussion

## 5.1 Implications

## 5.2 Next step

## 5.3 Limitations

# 6. Conclusion