# Final Report: Identifying Urban Functional Regions

Author: Joshua Rimes c22002134

Supervisor: Padraig Corcoran

Key: more info on website | to do | ask Padraig

# Abstract

# Acknowledgements

# Table of Contents

# Introduction

* Define socio-economic
* Define Quality of life (Mercers chart)
* Define Functional region
* Define POI

# Background

## 2.1 Why is Urban Planning and Layout Important

<https://www.archistar.ai/blog/the-importance-of-urban-planning-the-seven-key-reasons/#:~:text=Cities%20rely%20on%20urban%20planning,city%20area%20layout%20and%20density>

* "Cities rely on urban planning to remain functional, grow in population, and attract businesses. Every crucial aspect of an urban environment is under the effect of how its layout is planned."
* Urban planning matters cos: planned city growth, improved health and quality of life, less environmental impact, better economy and resource utilisation, national development, disaster prevention and greater credibility, more efficient problem solving

<https://www.revistaespacios.com/a17v38n24/17382416.html>

* "the characteristics of an "ideal urban layout system" that would have existed in the absence of external disturbances: political, social and economic"

## 2.2 Existing Views on the Socio-Economic Impact of Urban Layout

<https://www.archistar.ai/blog/the-importance-of-urban-planning-the-seven-key-reasons/#:~:text=Cities%20rely%20on%20urban%20planning,city%20area%20layout%20and%20density>

* "Highly developed cities largely contribute to the economic health and productivity of society. On the other hand, poor urban planning can bring on opposite, harmful effects like constant heavy traffic, insufficient infrastructure, inadequate housing options."
* "With a thoroughly planned layout, a city can provide its residents with access to all essential services, points of interest, and amenities. At the same time, the unfavourable aspects of urban life are reduced, leading to an overall healthier lifestyle and improved quality of life."

<https://www.centreforcities.org/blog/layout-city-affects-economic-success/>

* Street 'accessibility' score – a more accessible town creates better economy
* Lots of focus on street layout rather than POI or regional layout

<https://www.sciencedirect.com/science/article/pii/S2210670712000455>

* "Although the concept of quality of life has been in the development discourse for some time now, measuring it in a city is quite difficult as the aspects to be measured are still questionable"
* Table of quality of life criteria (Mercer)
* "The project adopts a decentralized and integrated approach to address three main substantive areas; shelter, basic urban services and local economic development;"
* A priority must be found for there to be a direct conclusion

<https://www.tandfonline.com/doi/full/10.1080/13574800903435651>

<https://www.tandfonline.com/doi/full/10.1080/13574801003638111>

* Views and critiques on public spaces seem to be partisan and impartial to the distortion of evidence to suit a specific thesis.
* Whether negative or positive, there seems negligible evidence to base upon

## 2.3 Research on Region Identification

<https://www.jstor.org/stable/43617893?seq=5>

* "Traditionally, geographers and other scientists have used two classificatory viewpoints in defining regions."
* " A region may be composed of areas or locational entities which in some specified respect are homogeneous."
* "the variable(s) upon which the region is defined are attributes of the area being grouped. – this is known as a formal region.
* "a region may be composed of areas or locational entities which have more connection with each other than with other in outside areas … this suggests that the areas comprising the region differ so as to be functionally complementary to each other. Such a region is termed a functional region."
* If one was to look at the interaction between areas and entities, it might skew the relation and location of the entity within a region in a way that is less binary (this or that region).
* (Brown and Horton 1970) – the flows used for delimiting functional regions are surrogate for the functional distance separating entities. Functional distance derives from mapping the n properties of each entity in an n-dimensional space, the computing a measure of distance separating any two nodes. This reflects the net effect of entity properties upon the propensity of the entities to interact.
* In practice, functional distance is derived directly from interaction patterns.
* Markov chain analysis to flow matrices. MFPT as a measure of functional distance.
* The resulting functional regions from this method was used to find much larger regions than what I would like to focus on in this project.

<https://www.sciencedirect.com/science/article/pii/S0303243422000794#ab005>

* Classification for buildings included residential buildings – not present in the POI data ill be looking at
* Morphological feature extraction: graph constructed for buildings in a block within an urban district, GCNN (Graph Convolutional Neural Network) built to extract formal features of the block from the buildings in it.
* Socioeconomic feature extraction: The W2V model used to analyse the block-POI structure to find a representation for POI categories. Socioeconomic features for each block calculated by weighing the representations of all the inner POIs.
* Word2Vec (W2V) model creates a representation of each word present into a vector. Words in similar context or similar semantics are closer in vector value.
* Stacking ensemble learning: SE model with 2 base and 1 meta classifiers is constructed to process the morphological and socioeconomic features of each block. An output of the probabilities that the block belongs to a different functional region.

## 2.4 Research on POI & Socio-Economic Datasets

<https://location.foursquare.com>

* Foursquare place engine – machine learning to synthesise 16 billion data points to ensure accuracy
* Is not focused on geographical POI data
* Not free for students

<https://www.openstreetmap.org/#map=12/51.4828/-3.1610>

* Open street map provides detailed mapping for POI and city data
* Not downloadable as far as I'm aware

<https://www.precisely.com/product/precisely-points-of-interest/precisely-points-of-interest?utm_medium=cpc&utm_source=Online-Advertising&utm_campaign=Data-Integrity_Global_Digital-Ads-Google-Paid-Search-Brand-Campaign_2025&utm_content=>

* POI data
* Not free

<https://digimap.edina.ac.uk/roam/map/os>

* Map and data delivery service
* Available by subscription to Cardiff University
* Range of datasets for educational and research purposes (educational or research?)
* Loads of POIs
* Downloadable poi dataset
* Level 1: 9 groups, level 2: 52 categories, level 3: 600+ classes
* Data is positionally accurate to 4 levels of accuracy
* Record structure of dataset includes 25 columns (look on website)

<https://digimap.edina.ac.uk/roam/map/society>

* Societal data: population density, age ranges, professions, employment, health

<https://www.ons.gov.uk/search?topics=9731,6646,3845,9497,4262,4128,7755,4994,6885,9724,7367&filter=datasets>

* Data from 2021 census: residence type by age

<https://www.ons.gov.uk/economy/datalist?filter=datasets&page=4>

* <https://www.ons.gov.uk/economy/grossdomesticproductgdp/datasets/regionalgrossdomesticproductallnutslevelregions> - Regional gross domestic product: all ITL regions
* <https://www.ons.gov.uk/visualisations/labourmarketlocal/W06000015/#inactivity> - Employment, unemployment and economic inactivity in Cardiff

## 2.5 Research on POI Data

<https://datarade.ai/company/blog/what-is-poi-data>

* Representations:
* Coordinate-based - This representation allows for accurate geospatial positioning and mapping of various points of interest. Usually uses latitude & longitude
* Address-based - Each POI is associated with a specific postal address. This method allows users to search and identify POIs based on their textual addresses. More user friendly.
* Categorical - POIs can be categorized based on different criteria, such as their industry, purpose, or classification. This representation allows for grouping and organizing POIs into different categories or types, making it easier to search, analyse, and present the data.
* Hierarchical – POI data can be organised hierarchically based on levels of granularity. This structure allows for efficient organisation and navigation of the data.
* Attributes-based - POI data can include additional attributes or metadata associated with each point of interest. These attributes can provide further information about the POIs.
* Graph-based - In some cases, POIs and their relationships can be represented using graph structures. Each POI is considered a node, and the connections between POIs (e.g., proximity, adjacency) are represented as edges. This representation allows for analysing spatial relationships and performing network-based algorithms on the POI data.

<https://www.ordnancesurvey.co.uk/documents/product-support/support/points-of-interest-classification-scheme.pdf>

* Existing POI classification
* ‘Accommodation, eating and drinking’, ‘Commercial services’, ‘Attractions’, ‘Sport and entertainment’, ‘Education and health’, ‘Public infrastructure’, ‘Manufacturing and production’, ‘Retail’, ‘Transport’
* 52 categories within the 9 groups
* Could have a colour/symbol for each 9 classes, then a label for the specific categories

# Implementation

Functional requirements:

* POI data can be inputted and processed
* POI data can be clustered based on variable proximity and density, and classification level
* Clustered data is visualised on a map
* Overlays show socioeconomic data on map

PostgreSQL

Dash

Pandas

Numpy or Shapely(?)

Scikit-learn

Folium or dash-leaflets

Airflow (if data can be automatically accessed from digimaps)

## Week 4 – Preprocessing

* Installed dependences Dash, Pandas, Flask, Psycopg2, Numpy
* Set up basic dash web interface
* Upload feature takes in 3 types of data file

A screenshot of a computer

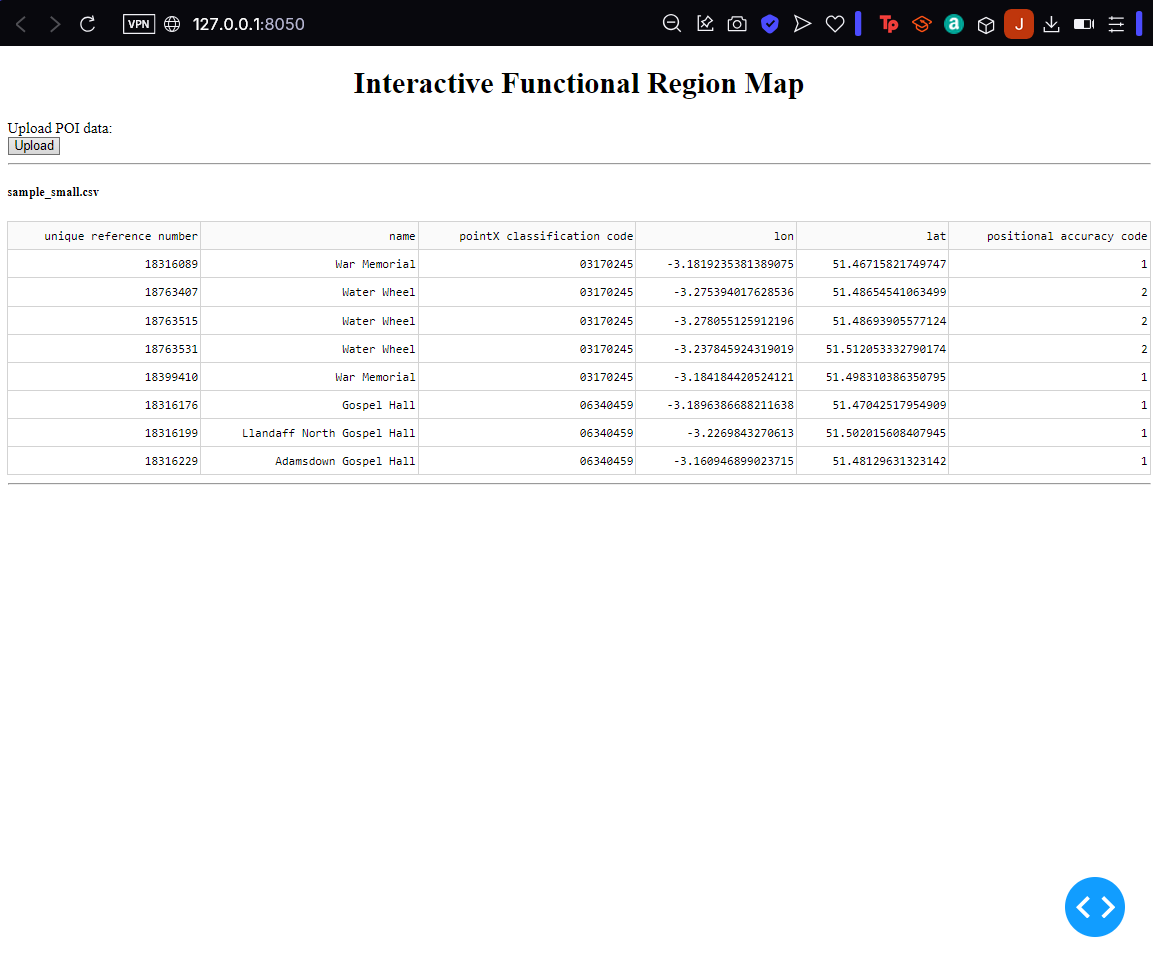
AI-generated content may be incorrect.

Cardiff northing and easting:

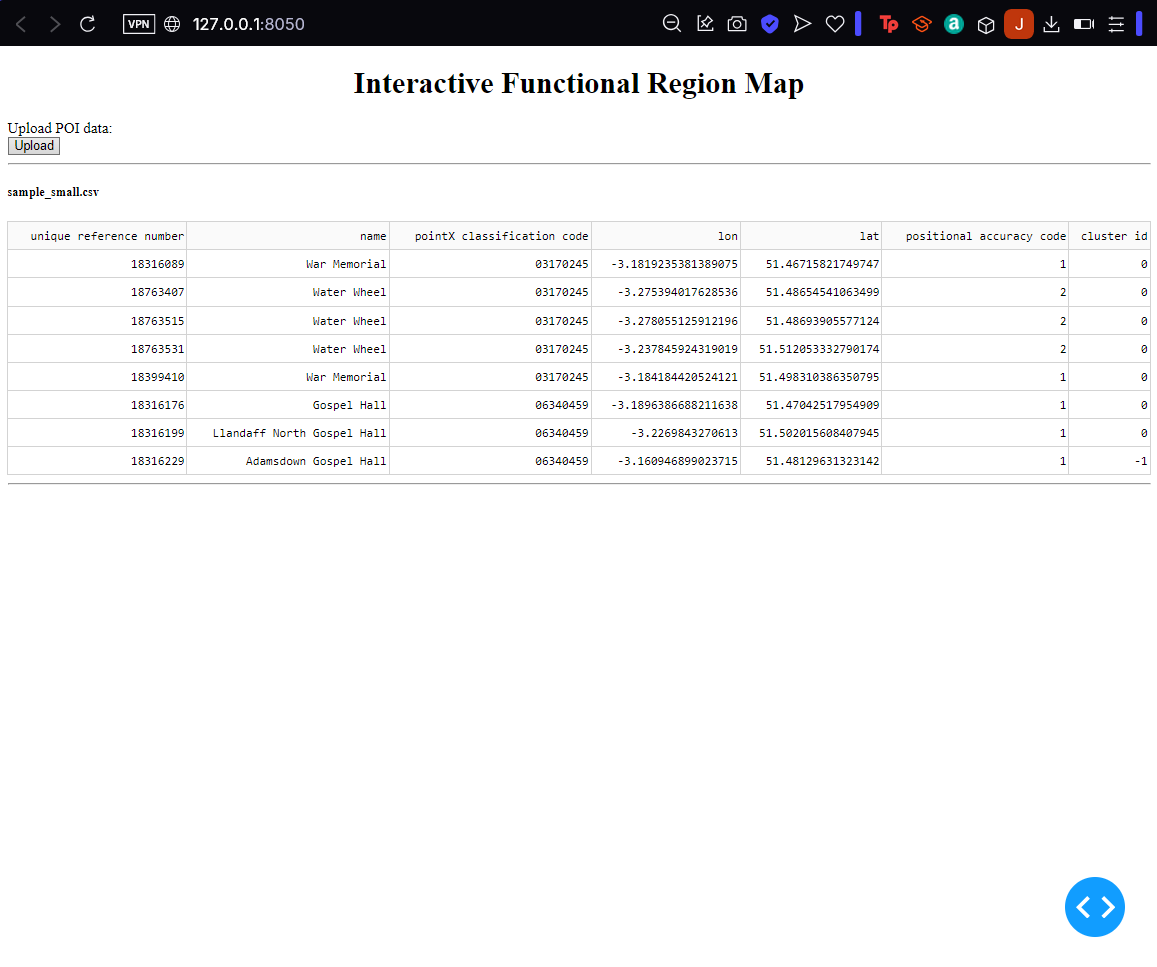
E= 325069.3 N = 183790.4

E= 311275.8 N = 173146.4

* Downloaded data for Cardiff using above northing and easting for future referencing
* Mention and justify website used for data
* Had to change column heading in POI data file manually because the file reader thought there was only one column.
* Data is stripped and stored in separate cells
* Easting and Northing had to be transformed into Lat and Lon
* Made it slower to upload each row
* Moved transformer instance outside of for loop to reduce time for upload
* Decided that a separate database is not necessary at this stage
* App can display data table of POI data
* Following image uses a sample of data



* Made dictionaries for POI classification
* Made function for classifying a POI
* Distance and clustering functions completed
* Cluster ids added to data table
* I need to set the parameters with careful consideration before I start clustering based on classification



* Pipeline for clustering:



* Made clustering for group level
* Had problem with indexing while cycling through the clustered POIs
* Explain why I used DBSCAN – justify

# Results & Analysis

# Conclusion

# Reflections on Learning

# Appendices

# References