**Initial Plan: Identifying Urban Functional Regions**

Author: Joshua Rimes c22002134

Supervisor: Padraig Corcoran

Key: to be deleted | to do | ask Padraig

# Project Description

Urban functional regions are urban regions that perform different functions in cities. Examples include tourism regions, bank/financial regions, shopping regions, nighttime economy regions, food regions etc. The goal of this project is to identify the functional regions in a given city in an automated manner using Point of Interest (POI) data. A POI is a specific point location that someone may find useful or interesting. Examples include pubs, shops, gyms, restaurants etc. The POI data will be obtained from OpenStreetMap or the Ordnance Survey.

Successfully achieving the goals of the project will require skills in data mining and analysis. Real-world data from UK cities including Cardiff City will be used in the project.

* Cities thrive in different ways, many different industries competing against each other
* Regions located relative to each other could cause a negative or positive effect on growth or economy
* If we could figure out what layouts work best, could be used to improve each industry

Cities are a hive of activity that integrate functions and industries together for the growth of its population and economy. That being said, lots of industries compete with each other constantly for their position and importance within the urban hierarchy.

The urban layout likely has a substantial effect on which industries thrive and which are battling to survive; an area that focuses on a specific function or industry might have a crucial impact on the other areas located near or adjacent. These **functional regions** are the foundation of urban success and should be considered carefully in the design of the city. They can range from tourism and shopping regions to financial and corporate regions.

If we were able to find out the impacts of the layout of functional regions, we could change the approach of designing future cities to benefit urban growth. This project aims to give an insight into this by identifying the functional regions of the city of Cardiff in an automated manner using point of interest (**POI**) data. Using the resulting data, I will compare the layout of functional regions to data from research on the economy, urban growth, and quality of life, so that a conclusion can be made on the impacts of urban layout.

The challenge of automating data analysis and using that data along with research displays in-depth knowledge and application of computer science systems including spatial data, data mining, and data structures. In utilising my solution, it will prove that the use of computer science is imperative to the future of urban design and specifically to the problem I am posing.

* Problem: identifying urban regions based on functionality in an automated manner to compare cities that are successful/unsuccessful in different industries
* Solution: create a python data mining & visualising system using poi data
* By using premade libraries
* Use system and online research to compare cities and write up an analysis of findings
* references
* <https://www.pbctoday.co.uk/news/planning-construction-news/urban-design-principles-in-cities/109848/>
* <https://www.tandfonline.com/doi/full/10.1080/13574809.2013.854695#d1e681>
* <https://www.sciencedirect.com/science/article/pii/S0303243422000794#ab005>
* Can show my development skills through programming, data management & mining, reporting
* Go into depth on the way cities are designed to favour specific industries
* Automation through coding -> computer science
* Spatial data, data structures, python

# Aims and Objectives

* Overall: aim is to identify city layouts that help functional regions thrive
* Make system that can take in poi data about a city and output a visual representation of the functional regions within the city
  + Risk: system does not function in time to do the comparisons
  + Could be due to lack of expertise, time, software specifications etc…
* Compare cities using the visualisations and online research about the successes/failures of industries in those cities
* Reach a conclusion about how the layout of the city affects its growth and economy.
  + A conclusion may not be found if the city layout has zero impact on growth or economy.

The overarching aim of this project is to discover the impact of urban layout on the economy and quality of life. To do this I will create a system that automates the identification of functional regions in a city using POI data.

**Objective 1**

I will implement a python system that preprocesses geospatial POI data, automates the identification of urban functional regions using a clustering algorithm, and visualises the resulting data in an interactive manner.

A risk of making the system myself is that the code I will need to write may require me to acquire new knowledge of libraries and algorithms that I might not understand in time to implement into the system. If this proves too challenging then I would likely have to mitigate the automated part of the system to make the programming easier, however, I believe there is a low risk of this happening, and I feel confident that I'll be able to learn and integrate these new algorithms into my system.

**Objective 2**

I aim to collect data about the functional regions and layout of Cardiff and an undecided city by inputting their POI data into my system and recording the results. Alongside those results, I will acquire data from online research and censuses about quality of life, average salaries, and economic growth in the functional regions of both cities.

The risk in this objective revolves around the access I have to data about both cities. POI data should be available for me to download without cost or ethical issues because it is public secondary data. Data obtained through research and previous censuses might be harder to find given that articles may be subjective and not truly represent the facts I want to use in my comparison. Because the main aim to seek the impacts of urban layout requires this data, I may need to focus on census and numerical data which could be harder to find.

* Bristol? – ask Padraig if necessary to have chosen second city already
* There may be confounding factors that affect the result that are not entirely due to urban planning, like Cardiff and Bristol being in the same region of the United Kingdom, and likely have similar social-economic.

**Objective 3**

The final and most important objective of my project is to reach a conclusion on the impact of urban layout on the success of industries by comparing the success of different functional regions in Cardiff as well as another undecided city, to show how the location of certain types of functional regions can play a positive or negative effect on other regions.

Finding that there is no impact on the success of industries would be just as conclusive as an extremely positive or negative impact, therefore there is no risk in reaching a conclusion.

# Feasibility

* Any data collected is secondary so no need for ethical approval?

↓

* Need for licencing if using the geodata?
* No non-standard hardware or software needed
* Census data?
* Possibility that not able to access poi data, unlikely

The data I will be collecting in this project is all non-personal secondary data; therefore, I need no ethical approval to use it.

Data downloaded from Digimap OS services require agreement to an Ordinance Survey Educational User Licence. I can agree to this licence via my status as an undergraduate student at Cardiff University. I will also be using data from the 2021 UK census, this data requires no licencing to download and use.

To run the system once it has been built, there will be a few feasible requirements, these include a desktop or laptop with a processor and storage capable of dealing with small data sets, as well as the latest version of python installed and a provided list of required libraries.

My solution is feasible because the data I need is accessible, and the system needs a low spec computer to code and run.

# Work Plan

* Include tasks, milestones, deliverables when writing this section
* Look for different datasets that I can use. I can get POI data from the ordnance survey via [digimap](https://digimap.edina.ac.uk/) or from [openstreetmap](https://www.openstreetmap.org/#map=12/51.4877/-3.1804). Maybe can also integrate other datasets such as city boundaries or census data.
* Do a review of related works on the topic. This can be used to identify the data analysis methods that I plan to use.
* Implement methods; I should be able to use existing software libraries.
* Perform the analysis using above methods. I can compare different cities, different POI types, create nice visualisations etc.
* Write the report.

**Week 1**

Milestone: Complete Initial Plan

* Tasks: Write the four sections of the initial plan covering context, aims, feasibility, and work plan.
* Deliverable: An Initial Plan document that describes the outline and plan for my solution and how I will achieve it.

**Week 2**

Milestone: Background reading on current urban planning and layout standards

* Tasks: Research whether there are already views on the impact of urban layout. Write up what I've found in my research.

Milestone: Choose and download datasets for POI data and census data

* Tasks: Research the different datasets I can use to get POI data and which part of the 2021 census I want to analyse. Write up my research and decisions.

**Week 3**

Milestone: Review of related works on identifying functional regions

* Tasks: I will research previous projects and articles on region identification and POI data analysis to identify potential data analysis methods I could use in my system.

**Week 4**

Milestone: The system will be able to take in and clean data

* Tasks: I will use Python libraries to handle inputting geospatial data. I will normalise data entries so that processing the data is consistent.
* Deliverable: A Python system for collecting data (used via a terminal at this stage).

Milestone: Data in the system will be categorised and stored in a spatial database

* Tasks: I will make a meaningful categorisation system for the POI data. I will use a Python connection to PostgreSQL to create a storage for the data.

**Week 5**

Milestone: The system will be able to calculate relevant variables to be used in clustering

* Tasks: I will program methods for calculating features such as POI density and proximity using Python libraries.

**Week 6**

Milestone: Clustering algorithms will be implemented into the system

* Tasks: Using premade methods and libraries, I will create a clustering method that can define functional regions. I will make a classifier for assigning functional labels to the clusters.
* Deliverable: A system that clusters collected POI data into functional regions, executed from a terminal.
* (Very) rough 12 week plan:

1. Write up the initial plan
2. Look at datasets to use (poi, boundaries, census)
3. Background reading
4. Implement methods to address problem – preprocessing
5. Implement methods to address problem – clustering
6. Implement methods to address problem – clustering
7. Implement methods to address problem – visualisation
8. Implement methods to address problem – visualisation
9. Software testing & analysis of cities
10. Analysis of cities
11. Write report
12. Write report - draft before submission
13. Preprocess data – collection, cleaning, categorisation, storage
14. Automated identification – calculate features with parameters, clustering, classification
15. Visualisation – interactive map, changing variables, heatmaps?

* Gantt chart

# References

* Use Cardiff Harvard citing