Final Year Project Report

Spatial Profiling for Residential Property Search

Aaron Cullen

A thesis submitted in part fulfilment of the degree of

BSc. (Hons.) in Computer Science

Supervisor: Gavin McArdle

UCD School of Computer Science

University College Dublin

April 6, 2018

Project Specification

The housing market in Ireland is currently in a state of crisis. House prices are set to rise by 20% [1] over the next three years and the average cost of rent has risen by 11% countrywide in 2017 alone, making it 16% higher than in 2008 before the economic crisis began [2]. The result of this situation is that more and more people are being priced out of areas to live and are searching for new areas to reside. However while cost may be a large reason for many, it is not the singular factor that determines a person's choice of area, there are many other circumstances that influence a person's choice in location. Different people are attracted to different areas for different reasons, for some it may be the abundance of work in their chosen field, the locality of nearby amenities such as schools and hospitals, commute times, broadband speed, the types of accommodation available and even the people who live there are all important factors.

There is a considerable amount of data available to match up to the considerations people take. For example, the census alone covers 15 topics that provide detailed data on wide spectrum of information from defined areas. Data from mapping services such as OpenStreetMap can provide the locations of useful amenities. Garda crime data from stations can give us a relative idea about safety and level of crime in an area. We cannot forget house prices, for which information based on the Residential Property Price Register gives us up to date information on the prices of homes sold.

While the data are available, they are not easily findable and they are not in formats and layouts that are user friendly. The main goal of this project is to change this by creating an understandable and straightforward solution to presenting this data in the Republic of Ireland. This will be accomplished with a web-based interface that will allow individuals to tailor queries to their specific needs, in order to find their ideal living areas or to find and recommend similar areas to another desired location; the results are then displayed using an interactive map. To make this possible, spatial profiles of ‘small areas’, which are locations defined by the census, made up of between 80 to 120 dwellings, around the country will be created by using the previously mentioned datasets. To extend this tool further, it will be integrated with popular property websites such as Daft.ie to show listings of available properties to buy or to rent in the given areas.

Abstract

The basis of this project is to create a spatial profiling system for a residential property search within the Republic of Ireland. The data used to create the spatial profiles consist mostly of information from the Central Statistics Office Census 2016, along with a small number of other datasets. The locations that are spatially profiled belong to Census small areas; these areas are generally small and populated by between 80 and 120 dwellings. Users are able to specify their ideal living area either by a set of location-based questions or by comparing locations to a selected area. The result of this process is a map that displays a choropleth map layer detailing the suitability of areas on the map. This should allow individuals to relocate to the best location that meets their requirements.

Acknowledgments

I would like to thank my supervisor, Dr. Gavin McArdle for all the valuable guidance and feedback provided throughout the research and development of the project.

Table of Contents

1 Introduction

1.1 Motivations

Residential housing is currently a topic in Ireland that consistently fills the pages of newspapers and sparks debate among comment sections online. With housing prices and rents soaring not just in Dublin and the other major cities but surrounding areas, the difficulty of finding a place to live is becoming ever increasing. The revival of the economy and the decrease in unemployment since the severe recession, in combination with the lack of new property development has led to a shortage of housing. Currently an increasing amount of people are limited to finding employment within and around the major urban areas as the countryside takes longer to recover from the recession. As many people move into the cities, the prices have risen with demand, pricing many out of the cities and to the commuter belts. This has resulted in more cars on the road, increasing the commuting times of those travelling in and out of the urban areas.

This has affected my life as well, as the difficulty of finding affordable accommodation near UCD is much harder than in 2014, as a result I now face a long daily commute from Wicklow. This problem required the need for a solution that could find a location that compromises between needs and wants, an area with reasonable rent and a shorter commute time. However, there was not any available service or tool that compiled all the relevant information to present recommended locations. It is possible to visit property sites to find places for rent and then use services such as Google Maps to see the potential travel times and nearby transport, but this is time consuming and in today's chaotic rental market you are likely to lose out on a property if you don't engage in the renting process immediately.

The more I thought of the idea of an “ideal area locater” the more scenarios emerged, initially for my own case it’s a fixed location I need to be around, but what if someone wanted to live in any area of the country with a specific number of requirements? By creating increasingly specific and detailed scenarios that, an individual may have the clearer the need for a tool that could handle a range of queries and return locations that would ideally suit the given set of input parameters. After some research into the available datasets, it was also clear that this was a task that was both possible and worth pursuing.

1.2 Description and Implementation

The concept of spatial profiling is to use data from specified areas and to extract the important information that is desired into a format that is easily understandable. For the residential property search that will be developed, the areas will be defined by the “Small Areas” that are used by the census. Created by The National Institute of Regional and Spatial Analysis, these areas range in physical size as they consist of areas with 80 to 120 dwellings, meaning towns and cities have a much higher density of areas while remote parts of the countryside may have large sparse boundaries. These areas are useful as they are small enough to give specific information on locations but they also link directly to the census data that are the main source of data that will be used.

In order to create a spatial profile of these small areas, we need to gather datasets that provide the information regarding the aspects of the area that should be required by a user's search to be queried.

The census is an extremely large and detailed dataset that records accurate data from every part of the country. The data covers a wide range of topics that deal with important everyday aspects of people's lives. Details such as Sex, age, marital status, nationalities, religion, work and level of education of the people living in an area that are recorded. Information around housing is also recorded through the census; this includes the house type, age and the utilities available to the home. Other useful information includes data on commuting times and the number of cars owned in an area. The census will be the main dataset used in the spatial profiling due to its direct link to the small areas boundaries being used and its large amount of data make it the most useful dataset for information within each area.

While the census contains much of the information needed for the spatial profile, it is lacking some important aspects that make up what people seek in an area. The amenities in an area, the public transport available, the prices of homes and the safety and crime levels in the area are all examples of further data required to build a worthwhile profile. Fortunately, this data is available too. By using three other datasets, these important aspects are covered. These datasets are the OpenStreetMap data that includes, points of interests and worship along with public transport locations, secondly the Residential Property Price Register, giving us the house prices of homes sold in an area and finally the Recorded Crime Offences by Garda Station, showing the frequency of the types of crimes reported to stations.

Equipped with the data we are able to start evaluating areas in a meaningful way. There are many common methodologies applied to ranking areas when constructing "best places to live" rankings or guides. Using weighted ranking systems to value aspects of the area such as job market, the cost of living and the general quality of life. However, the flaw in these lists is that the methodology is static, so what the list really reflects is what the creator determines to be the most important attributes of an area rather than an individual assessment that shifts to a user's suggested inputs.

This project differs from the traditional methods by giving users a series of questions about a multitude of topics related to both the requirements people look for in a desired area and the given implemented datasets. Then using the inputted information it is possible to create unique rankings that are suited to the individual. This is accomplished by dividing the topics up into different sections and giving users the ability to expand the sections for more detailed questions if that section is of particular importance to them, giving a greater weighting importance to that section. Using employment as an example topic, it should not have significant importance for an elderly retired couple, while for a young professional it may be very important to be located in close proximity to employment opportunities and so should be considered by the tool to be of increased importance when weighting the different sections that calculate the areas.

When the user completes the questions, they are presented with a page that displays the map of Ireland with an overlay using a choropleth map to represent the rankings of areas. The interactive map would also allow users to click on an area to see details of that area. Another option for users rather than enter the questions is to find areas that are similar to another given area, this would take all the details of the selected area and try to find another area that matches the profile, similarly displaying the results in the same method as the question option.

The technical implementation of the specified setup involves a spatial database at the centre of the setup. The spatial database is used to both store the datasets and to query the data. Once the user has inputted data, it then becomes transformed into queries the database can be called in order to query the data and then the areas can be ranked in terms of suitability to the user. Once the result are generated, they are displayed using a Tile Map Service to display the map of Ireland and then using the geometry layer files to display the small areas as an overlay with a choropleth map.

1.3 Report Summary

The remainder of this report is structured as follows:

- Background Research: Consisting of material from background readings that have contributed to shaping the design of the project.

- Project Design: Detailing the approach taken and a discussion of the design aspects of the project.

- Project Implementation: Describing in detail the implementation of technology used in the project.

- Testing and Evaluation: Demonstrating the testing and evaluation methods of the project.

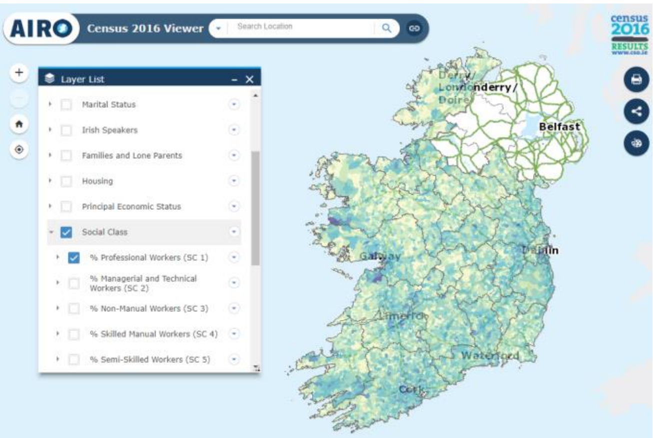
- Conclusion and Future Work: Reflection on completeness of the project criteria with criticisms and discussion of possible future work.

2 Background Research

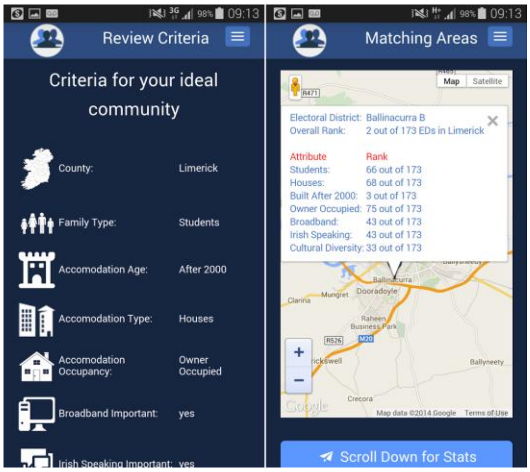
The background research looks at a number of different topics that assisted in the development of the design and planning for the project. The research will firstly look at existing systems that are similar to the project to get an idea of what is already available. Then the methodology behind evaluating areas will be examined to understand how areas are evaluated and ranked. The aspects of an area which individuals desire and the reasons why they move will be investigated to better understand what the requirements of someone who is relocating. In terms of data the background research, will discuss the datasets that have been selected based off the research done in previous sections and why other datasets have not been included. Finally, the technical implementation will detail the different software tools used to create and support similar systems to find the best-suited solution.

2.1 Existing Similar Systems

The initial research performed was an investigation of pre-existing systems that completed similar spatial related tasks and to compare them to the planned solution for the project.



1.a Screenshot of AIRO displaying census data



1.b Screenshots of Census Explorer application

Figure 1. Collection of screenshots from similar systems.

Teleport.org is one such system; it allows users to choose from a number of tags to find the location from across the globe that would be most suitable for them. The tags are under a handful of headers that relate to quality of life, living costs and career. It follows these selections up with a handful of personal financial questions to get the financial status of the user. A list of cities that best match a user's choices and the difference in the cost of living in comparison to their current location is produced. Teleport provides an excellent web based interface but the focus of the site is clearly skewed towards young professionals who are looking for cities across the globe to find skilled employment. With many of the questions very focused on income and the field of work the user is currently involved in this system is not user friendly for a wide spectrum of different users. The global nature of the search also means that results are limited to large urban areas and the not a specific area with the given location.

Dwellr (play.google.com/store/apps/details?id=air.gov.census.phone.dwellr) was another system that recommended locations based on a user’s given living preferences, it was created by the United States Census Bureau and used the data from the US census to formulate its results, meaning of course that the results were confined to the USA. Dwellr was delivered through a mobile application rather than a web based interface, similarly to teleport it gave users a list of large urban areas closest matching to their preferences. Once again, the areas returned were not highly specific; instead, a higher-level approach of giving information about the entire urban area rather than small areas within the city is used. In late September 2017, the US Census Bureau stopped supporting census applications they had developed, including Dwellr.

The Census Explorer (play.google.com/store/apps/details?id=com.censusExplorer) is another mobile application much alike Dwellr, however it is using data from the Irish census. The questions The Census Explorer provides the user with are based around the type of accommodation and some details regarding the people in the area. Unlike the other two systems that have been examined, The Census Explorer returns information from much more specific locations rather than just metropolitan areas, and provides the user with the statistics in relation to the questions asked in the input (see Figure 1.b). The one drawback of the application is that you must select a county to choose the location from rather than being able to find the best locations countrywide in a single search.

Another Irish based spatial resource is the All-Island Research Observatory (AIRO) (http://airomaps.nuim.ie/id/Census2016/). While this website does not deal with recommending users where to live, it does a particularly good job of displaying the data, as the visualizations are high quality and make it easy to understand the displayed dataset at a glance. Much of that data AIRO uses will be the same data that is used for the residential property search of this project; however, the AIRO site does not combine this data but rather displays it individually.

Reviewing existing systems has been a useful contribution to the project as it has highlighted the elements of other systems that are successful and worthwhile keeping. It also demonstrated that while there are a number of similar spatial systems already created, currently none of them provides the required specifications of the residential property search that will be created.

2.2 Methodology for Evaluating Areas

In order to be able to evaluate and rank areas we must have a methodology in place. By researching and analysing other methodologies deployed by ranking lists and discovering the factors that people desire in an area, the end goal is to determine what data is both necessary and important.

Looking at a high-level ranking system was the first step of this process; a well-known location-ranking index is the Human Development Index (HDI) that is used in the United Nations Development Programme's annual Human Development Report to measure the development of countries. The index is formulated using three different key factors, life expectancy, education and income. In terms of what would be useful to apply to an evaluation for areas within Ireland, the life expectancy is not of much use as the difference between life expectancy rates across the country is negligible as seen by the health and care table in the regional quality of life in Ireland report [3]. Education level and income may be more useful tools for an Irish index as these factors have trends that are more distinguishable amongst areas within Ireland.

While HDI ranking gives a good overview of a process behind evaluating areas, it does so with very large areas, dealing at the nation state level rather than city or even neighbourhood. In order to understand smaller sized areas it is best to examine more localised ranking systems. As there is a lack of detailed ranking systems for areas available in Ireland, an American system was examined instead. The US News real estate, 2017 Best Places to Live is a leader board of the "100 largest metropolitan areas based on affordability, job prospects and quality of life" [4]. Using five different indices with different weightings, they take data from the indexes and standardise it using a standard deviation to differentiate each location from the mean of each dataset.

The five indices that the methodology makes use of are, the job market index, the value or affordability index, the quality of life index, the desirability index and the net migration to the area. Looking at these individually, there is a lot of useful information to how they formulate their evaluation.

The job market index accounts for 20% of the overall score and consists of the employment rate and the average salary, both given an equal weighting. The value index takes into account the cost of living in an area and is weighted at 25% of the score. Their quality of life index consists of a number of different factors, the most significant being the crime rate, with the quality of education and health along with commuting times making up the rest of the score. The final two indexes are less relevant, and are created from data that is not as freely available or reliable as the other data groups. While the weightings are static in this methodology unlike the proposed plan for the project, this is not a problem as we can simply adjust the weighting to match the user's preferences; so that the data they find to be significant can have the biggest impact on the ranking. The methodology is also functional as similar Irish datasets can be found to match majority of those used.

Reading technical research papers on evaluating the data was worthwhile too, in particular papers detailing the measuring of the quality of life in an area. Two such papers were Li, G., and Q. Weng (2007) [5] and Brereton, Clinch and Ferreira, (2006) [6]. The latter of these papers has a well-constructed formula used to calculate the level of well-being of an individual *i* in a location *k* (see Formula 1).

$$u_{i,k} = \alpha + \beta' \mathbf{x}_{i,k} + \gamma' \mathbf{a}_{i,k} + \varepsilon_{i,k} \qquad i = 1 \ldots I,\ k = 1, \ldots, K$$

Formula 1. Quality of life calculation from Brereton, Clinch and Ferreira, (2006).
'a' is a vector made up of spatial factors such as commuting time or proximity to a hospital while 'x' is a vector of demographic characteristics such as age and gender.

Similar to residential property search that is being developed, the paper measuring the quality of life in Indianapolis uses a selection of data such as population density and the employment rate from a census. Factor analysis is then applied to the data of each area to create maps showing the mapping of the quality of life in the city showing another way to turn data into evaluations.

Reviewing methodologies was an important aspect of the background research of the project as it provided examples of systems that used formulaic approaches to creating an evaluation of an area given a certain set of data; this will be useful when dealing with the results that the data returns from queries.

2.3 Why People Move

In the section above, we looked methodologies in order to use data to rank locations. However, humans are not robots that simply choose to live in an area, which a formula decided was most suitable, there are many factors that are not stored plainly in census details or found by looking at income statistics. A report by Thomas, Serwicka and Swinney (2015) [7] looked into the reasons people from different age groups live where they do. In the report, it is clear to see that there is a difference in what people in different age groups desire. For young adults it is important to live closer to their workplace and closer to social amenities. Among age groups that are more likely to be raising families the availability of good schools and safe neighbourhoods is most valuable trait of an area, while for older retired age groups, moving to the countryside is a key desire. Equally as important as what makes an area desirable for someone is what makes it avoidable. Neighbours, the distance from a workplace and the cost of housing are common issues that people dislike about where they live. Similarly, we see differences between the age groups, for example, many younger people find the lack of transport to be a frustration, and while so do the older population who may no longer be able drive or have moved out to the countryside where there is a lack of public transport.

In Dennett, A. R. (2010) [8] the role of age and life course stage are examined, younger people want to be near leisure and social centres as well as be in nearby proximity to the workplace, likely living in a city apartment. As people get older and start to have families they move out of the city to suburban life, where issues such as good schools and the safety of an area become a greater concern. Finally, as people move towards retirement the need for schools diminishes and a life in the countryside becomes a greater desire. From looking at the data from the different groups, it is important to see that one size does not fit all, and that is why it is important to allow people to tailor their options when finding an area to reside in.

A report published by the United States Census Bureau, Reason for Moving: 2012 to 2013 [9], is a comprehensive look into the motivations behind people who moved home in the United States between 2012 and 2013. 11.7% of the US population moved home in that year period, of these 48% stated that they moved for housing related reasons, 30% for family and almost 20% for employment. Of the housing related reasons, we can see some similarities to reasons previously mentioned. For some it is to reduce the price of their housing, while others want to upgrade their living conditions and desire safer areas. In terms of job related moves, a change in job circumstance was the greatest factor, while others decided to relocate to cut down on commuting times. Family related moves were down to a change in marital status and people moving out of their parent's homes to establish their own household. Other reasons such as moving for college made up a small percentage of the total relocations.

The CSO in Ireland released a similar report on the "Population Distribution and Movements" [10] Based on Census 2016 data. Between April 2015 and April 2016, 5.7% of the Irish population moved homes. The report shows the trends that the population within cities and large towns are moving at a proportionally higher rate than the population in the countryside, which only made up for 25% of moves despite nearly 40% of the population residing there. While 80% of the Dublin population that moved stayed within Dublin, the majority that left the county moved to commuter belt counties such as Meath, Kildare and Wicklow. Issues mentioned previously such as commuting times and the desire to live in the countryside are causations for such moves. Employment is another important factor in migration. Counties based in the North West have the some of the highest unemployment rates [11] and they have some of the highest rates of movement out of the county, with many likely moving to find work elsewhere.

The information collected from the research of migration played an important part in the dataset selection for the project. What individuals desire and do not desire in terms of their home and its surrounding area along with the statistics for moving simplified the selection of data as it made clear what range the data would have to cover.

2.4 Data

There is a massive amount of data on a wide range of different topics, which is publicly available for use nowadays. However, it is important not to select every dataset available when trying to create a database solution this will result in a huge collection of redundant and useless data that will unlikely be used. Through the background research completed, it was possible to identify the collections of data required for the spatial profiling of areas. The main dataset consists of the Census 2016 data released by the Central Statistics Office. While the secondary data exists of datasets from OpenStreetMap, Residential Property Price Register and the Recorded Crime Offences by Garda Station.

2.4.1 Census Data

The census is an official survey of the entire population of a country. The detailed nature of the census makes it a very useful dataset to have as it enables easy access to lots of specific information in very small areas of the population, which is required for an accurate residential property search. The census also provides the spatial data to define these small areas onto a map, which makes the technical implementation section of the project robust. The spatial data of the census are provided through a shapefile. The shapefile contains information about both the census statistics within the small area and the geometry of the boundaries for these areas. The data in the census is composed of forty-five different tables that are organised by fifteen different themes. (See Table 1)

Sex, age and marital status	Migration, ethnicity, religion and foreign languages	Irish language	Social class and socio-economic group	Motor car availability, PC ownership and internet Access
Housing	Communal establishments	Disability, carers and general health	Families	Education
Commuting	Principal status	Occupations	Industries	Private households

Table 1. Table displaying the themes contained in the 2016 Census.

Data in the census is stored with the table contents stored horizontally, the small areas and their ID is stored horizontally. Each 'gid' is linked to an individual small area. The 'gid' is also used in the spatial data that contains information on the small area. Using the themes of the census a wide range of topics are covered. Information around demographics, socio-economic issues and utilities are provided for the queries that will be created. However, the census does not cover everything that is important for the property search as it lacks information on items such as amenities and crime in the area; to compensate for this we will introduce extra datasets.

2.4.2 Secondary Data

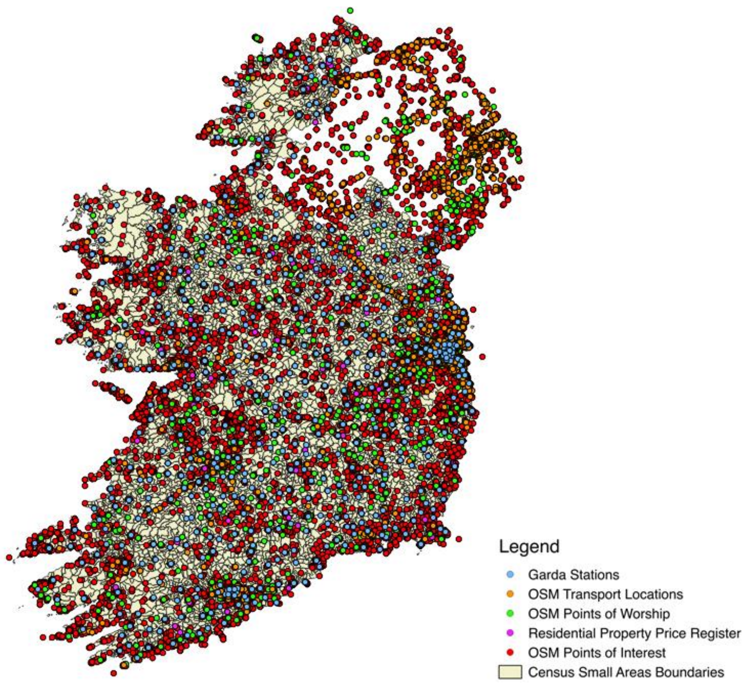


Figure 2. Census Shapefile Map with the secondary data plotted on the census small area map. Different datasets are represented by different colour dots with each dot being a location from the data. Census statistics are contained within each small area.

OpenStreetMap, Residential Property Price Register and the Recorded Crime Offences by Garda Station make up the secondary data to cover the areas the census cannot.

The **OpenStreetMap** data being used consists of three datasets; these are points of Interest that include amenities such as hospitals, schools, parks and restaurants. 'Points of worship' that show the location of religious establishments and their denominations. The third dataset is the locations of transport infrastructure. The datasets are available as shapefiles so there is no complicated conversion process required to make the data compatible with the census data.

The **Recorded Crime Offences by Garda Station** lists crimes such as burglary, theft and assaults. Since the location of the crime is not recorded at the point of the incident, instead it is recorded at the station that responded to the incident. This makes the data less accurate than similar data used in ranking the American cities. However, it is still useful information to use so long as we do not put too much emphasis on it. The spatial positions of the Garda stations are listed using the Irish Grid Coordinates rather than Latitude and Longitude. Although it is possible to convert them to create a shapefile.

The final data set is the **Residential Property Price Register**; this data has the list of all sold properties in Ireland since 2010. It tracks the address, the price and the type of house that is sold. The fact that the location data are stored using addresses rather than a coordinate system makes it difficult to convert them on mass. In order to obtain the coordinates from the address we need to convert it using Google Maps API, this is a slow process however, but once the backlog of necessary data is converted, the number of updates to the data each week is minimal, making it manageable to handle.

Other datasets such as pollution levels and climate were also considered as possible datasets. However, the difference level in values across different areas in Ireland was insignificant. Including these datasets would not have contributed significantly while increasing the amount of data by a considerable amount. With all this information now in the system, we can view the data graphically (See Figure 2).

2.5 Technical Implementations of the Spatial System

With the data selected, a system is required to store and manage the data in order for it to be analysed. The best method to achieve this is to use a geographic information system (GIS). The United States Geological Survey states, "A GIS is a computer system capable of capturing, storing, analysing, and displaying geographically referenced information; that is, data identified according to location" [12]. By using the basic geometries of lines, polygons and points [13], a spatial system can represent spatial data such as coastlines, regions and buildings. The setup of the GIS enables the creation of spatial queries that allow the spatial attributes and the non-spatial attributes to combine to evaluate results. For example, a spatial query using the datasets from the previous section would be to ask for all the areas with over 20 children that also contain a school.

The GIS also plays and important role in displaying the data on the internet. The GIS links to a mapping server, these servers overcome the technical limitations that regular web based technologies face when dealing with spatial data [14]. The result of this is that it is possible to rapidly render the data to represent a visual map with our information. There are many GIS systems available for use, along with web services that work in combination with them. The project will require using a map and layers to display the data. The layers will display the choropleth map so the evaluation of the areas is available to view. When considering the colours of the choropleth map, it is important a colour scheme that is friendly to colour vison impairment, as all users should be able to view the data clearly [15].

Implementing a GIS system that is widely used with lots of documentation and is compatible with many tools is a good step in order to ensure that the system will be able to accomplish the intended tasks without unnecessary complication. An example of a robust system is PostGIS, as an open source piece of software there is an abundance of information covering all areas of the software. Similarly, it works with a wide variety of web-based tools that can meet the requirements for the project specification. This makes it an ideal GIS system to use for this implementation.

Finding a usable technical implementation was a very important aspect for the project as if no technical solution existed, the information and data from the other sections of the background research would not be of any use to the project.

2.6 Technical Implementations of the Mapping System

There is a wide range of options available when creating an online interactive map. After looking at a number of different options for mapping tools to create the online interactive map. A solution using, PostGIS, OpenLayers and GeoServer was identified as a potential implementation [16]. The PostGIS implementation that is mentioned above is compatible with the system as the spatial data is already stored in the spatial database. GeoServer then takes the data from the database and then OpenLayers is used to display the map. Another option that maybe explored if the planned PostGIS, OpenLayers and GeoServer method runs into issues is to use QGIS2Web as the QGIS software is already connected to the PostGIS database that has the spatial data. While not as powerful as the other solution from examining its capabilities, it is capable of the task required [17].

3 Project Design

The project design firstly details the goals of the projects and then builds on the background research by discussing the approach taken in the design and considerations behind the ideas and problems that transpired in the process of solving and developing the goals for project.

3.1 Project Goals

As previously mentioned the overall goal for this project was to combine available datasets to produce an understandable and straightforward solution to presenting this data in the Republic of Ireland. The specific goals of the project where broken down as follows:

Core:

- To develop a spatial profile for meaningful statistical units within Ireland. A set of parameters to describe each area will be designed.
- To develop a web-based interface, which allows users to describe their ideal living area, based on questions or by naming an area, they like and develop suitable techniques to recommend similar areas based on the results.

Advanced:

- To develop an interactive map to allow individuals to query an area and view its complete profile.
- To extend the recommender tool to include a property recommender by integrating it with property sites such as Daft.ie and MyHome.ie to recommend individual properties.

3.2 Time Management

In order to create a map of the milestones required to finish the project a Gantt chart as seen in Figure 3, was created to keep track of the goals that needed to be completed and the planned dates within which they were started and finished. The Gantt chart was a useful measurement for keeping the development of the project on schedule. Combining the Gantt chart and the weekly meetings with the supervisor helped keep the project progressing and prevented issues from taking a considerable amount of time.

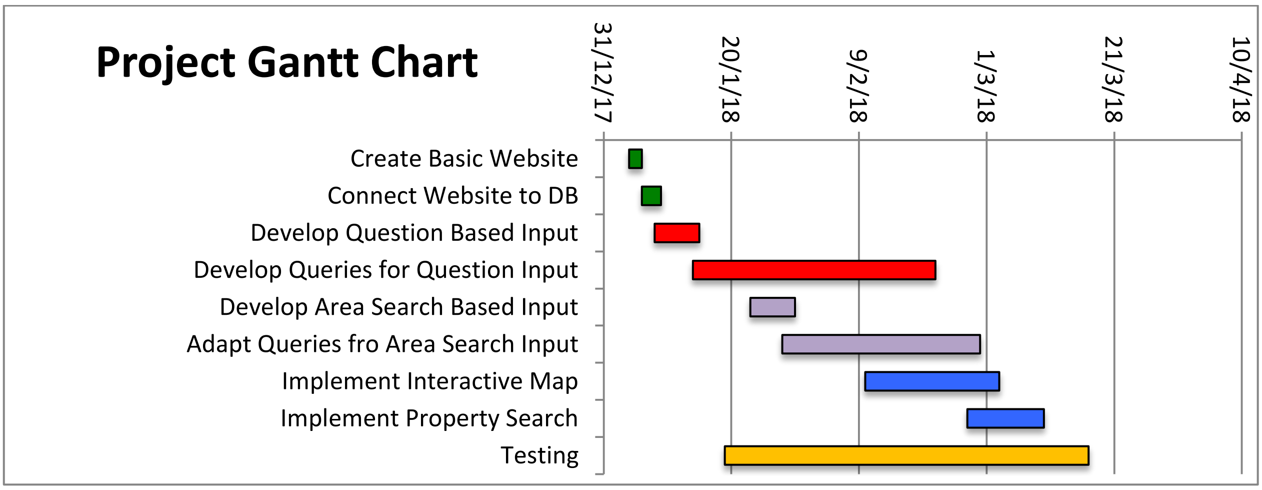


Figure 3. Gantt chart displaying the tasks and timeframes from the tasks to be completed.

3.3 Project Approach

The project approach details the decisions taken and the reasons for these decisions during the development of the project specification.

3.3.1 Webpages

All the interaction with the project takes place through webpages. As many devices with a wide range of screen shapes and sizes are used, it is important to have a highly responsive design that is compatible with not only desktop computers but also devices such as smartphones and tablets. In order to achieve this webpages were setup using Bootstrap 3 [17], which implements a mobile first approach to ensure compatibility among all devices. It was also important consideration to ensure all the technologies used for the user interaction would operate on as many devices as possible.

3.3.2 User Input

The project specification required that the user be able to search for areas through a web-based interface based either on questions or by naming an area. This ultimately led to two different forms of user input.

For the question-based search, a series of questions split into a range of different topics such as housing, location, education and healthcare. These questions are used to find the suitability of the Census small areas within Ireland given the users preferences. The questionnaire uses a Likert scale [19] approach for answering questions, rather than binary yes or no questions; the user responds to a given statement on a sliding scale that ranges from agree to disagree with a neutral response in the middle. This style provides a more comprehensive response to how the user feels about the statement. The questionnaire consists of 28 questions, the user can also expand different sections to reveal more questions in that topic if it is an area of significance to them. In total there are 40 possible questions. The questions are displayed with the statement above a slider as seen in figure 4. In regards to how the questions where created and selected, a number of user stories were created in order to choose the datasets required, these user stories also helped drive the creation of the user questions. An example of a user story is as follows - *"Mary, 70, Retired and widowed, wants a smaller home closer to health care facilities and wants to live near to shops. She currently lives rurally in Offaly and does not want to be more than an hour from her current location. She would like to live in an area with other elderly person that is safe"*. It's not feasible to have a question input for every single requirement a user may have. However by developing questions that make use of the datasets available and then combining these with the functionality that the visual of a map displays, users can obtain more information. For example a user who wants to live near the coast or a certain location can find the highest scoring area in the location they desire by examining the areas on the map.

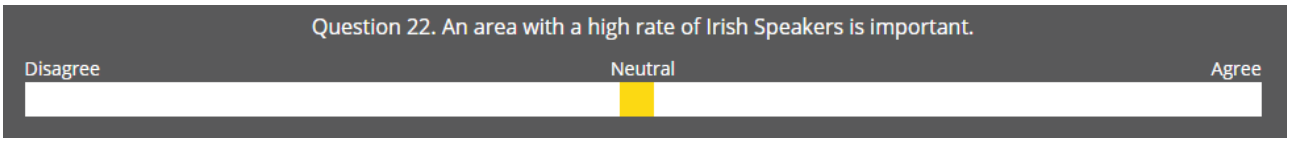


Figure 4. Slider showing how the questions in the user input are displayed.

The second option for searching areas allows users to search for an address within Ireland. The small area that contains the given address is compared to all other small areas to find the similarity between them. The search bar provided has an autocomplete dropdown list to both simplify address entry for users and to make sure that the address entered is valid (See Figure 5). Another design choice considered for this search was to implement a map and allow users to drag and drop a marker to the desired location, however the address search was ultimately chosen as it was a faster, less cumbersome method of input.

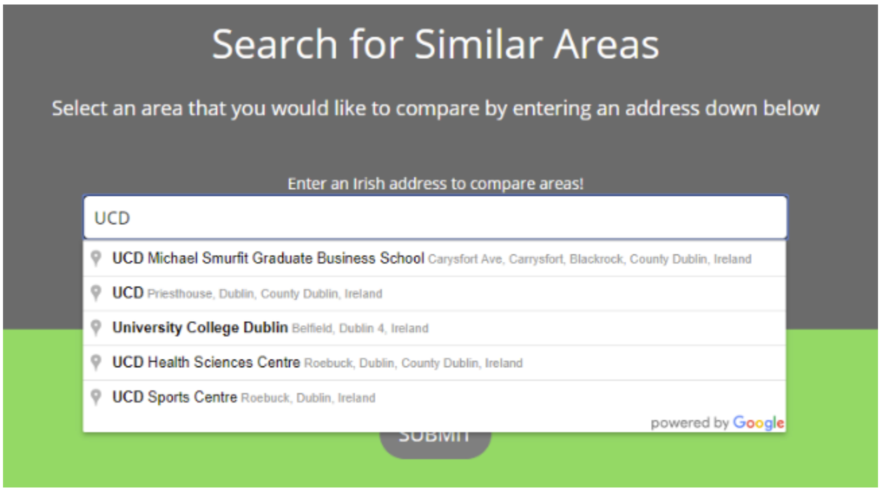


Figure 5. Search bar displaying the dropdown menu for the address search user input option.

3.3.3 Data

The management of the data storage using PostGIS with a PostgreSQL database is the same as described in the background research. The datasets used were largely the same as those discussed in the background research. However there were two changes to the dataset selection, firstly the 'Points of worship' dataset was removed due to the data being inexplicit. The majority of buildings listed in the dataset where listed simply as Christian, with only a minority detailing their denomination. There were only a handful of non-Christian places of worship listed too. As a result, the census data relating to the religious population within each small area was used instead.

Secondly, 'The Residential Property Price Index' data was not used; instead, it was decided to use The Central Statistics Office 'House Prices by Eircode' dataset that provides the median price in each Eircode routing area. There are two reasons for the change; firstly the data from the 'Residential Property Price Index' did not include spatial properties, only an address for the location. As a result, this meant that all the addresses would have to be geocoded to produce the latitude and longitude values for each location. In order to batch geocode such a large dataset a paid service costing hundreds or possibly thousands of euros would have been required to process the data in a reasonable time and manner. Another problem that arose from testing a sample of 1000 property locations from 'Residential Property Price Index' dataset was that it was possible for many small areas to have no properties with a recent sale history located within its boundary. Other small areas contained only a small number of recorded house sales; this could have misrepresented realistic house prices in a given area by having an outlying sale value that would skew the average price for that small area. While the Eircode areas are larger than the majority of Census small areas, it does still provide differentiation of housing prices among over 150 regions while still providing coverage for all the small areas.

3.3.4 Processing the Data

In order to manage the interaction between the database and the website, PHP's PostgreSQL extension was implemented. This extension provides an extensive API that allows for web interaction with the database, and most importantly for this project, enables execution of quires to retrieve information from the database. While other languages such as Python also provided similar API's for database interaction, PHP was chosen as its documentation was straightforward and understandable. There is also a large amount of resources online available such as tutorials of implementations which aided the development.

In terms of processing the data, the initial idea was to directly run spatial queries each time the user submitted an input. However, it quickly became apparent that the time taken to run a spatial query would be too long to have a reasonable search time as for each question from the input there would be 18641 spatial queries ran, resulting in a query that could possibly take hours to complete. The solution to the problem was to create tables in the database that had the spatial data pre-computed. While this reduced the flexibility of the queries, it greatly reduced the time taken to process the data. The pre-processed data for point-based locations included two types of table; firstly, for the types of locations that are commonly found within a small distance such as bars or transport stops, a count of the number of those locations within a given small radius was recorded for each small area. The radius was based off the distance a person would generally travel to such a location, the radius for transport locations was set at 1.5km while the radius for nearby schools was 10km as people are more likely to drive there compared to a transport stop. Secondly, for less commonly found buildings such as hospitals or universities the minimum distance to the closest location from each small area was recorded in the table. For the area datasets such as the Garda Crime Data and House Prices by Eircode, the values they contained where joined to the Census small areas contained within them.

The PHP not only queries the data but is also used to evaluate the data returned by the query. A score is calculated for each question for each small area. The question-based input includes a weighting system used to adjust the overall area score in favour of the areas the user most strongly responded. The reason behind the weighting system was to ensure the features most important to the user had the largest effect on the overall score, so for an area to be recommended, it would have to score highly in those sections. The area-selected input simply compares the difference between the values of the selected area. For example taking hospital distance as the value, a lower score is not necessarily better but rather one that is closer to the distance of the selected area. While it may not seem reasonable to score an area lower for being closer to a hospital, ultimately it is scored this way as when comparing areas, the similarity between the areas is the key evaluator.

The final step in the data processing is to create the output. The output generated is a GeoJSON file that contains information about the small area such as its name along with the score values and most importantly the geometry data that is used to display the information on the map.

3.3.5 Mapping Technology

The background research concluded that a combination of OpenLayers and GeoServer would be the best option for an interactive map implementation for this project. The plans were altered however after discovering that spatial queries would take far too long to run as discussed in the previous section. The initial attempts to run the OpenLayers and GeoServer setup also led to some frustrations. This resulted in a reconsideration of the technologies used, after some further research there were two important changes made. First of all, rather than using GeoServer to serve the small areas layer, the PHP could generate a GeoJSON file that could be used by the web-mapping client. Secondly, the choice of web mapping client changed from OpenLayers to Leaflet. Leaflet was not an option that had been considered in the background research as most of the options explored were based on an implementation with GeoServer and there were not many Leaflet examples for that setup. After some comparison and examination between both Leaflet and OpenLayers, the final choice was made to go with Leaflet. The main reason behind this choice was that Leaflet offered all the functionality that was required to implement the necessary parts of this project while being a lighter weight package than OpenLayers which has an extensive library of functionality built in. In terms of size, the Leaflet source file is less than 250KB while the OpenLayers is 3.5MB. Similar to the other technologies chosen, the Leaflet documentation was excellent; this made the learning curve for using Leaflet far less steep. The API reference list was precise and very easy to navigate, while they provided tutorials that worked without the need for modifications, in comparison the OpenLayers API was verbose and difficult to work with, coupled with the fact that tutorials and examples often did not work as shown.

3.3.6 Interactive Map

Figure 6. Interactive map showing the results of comparison of areas similar to Bray, Co. Wicklow.

There were many design choices made in regards to the features and styling of the interactive map as seen in Figure 6.

3.3.6.1 Information Displayed on the Interactive Map

One of the decisions made in the background research was to use a choropleth map. This style of map colours the areas or regions on the map to represent a visual for the data they are displaying. For this project the score is based off the evaluation of each Census small area. A single sequential hue using a palette of green colours was selected as it is colouring blind friendly [20]. The Tile Map Service Mapbox provides the background map using their 'Light' styled map. This style provides a full cartographic map without any colouring that might affect or distract the visual of the layers rendered above.

The other display features that are displayed in Figure 6 are the '**Score Legend'** in the bottom right hand corner and the '**Area Info**' found in the top right. The 'score legend' was implemented in order to illustrate to users the grading of the colours in relation to the score. The 'area info' displays the information relating to the area that is currently being hovered over by the user; this information includes the name of the area as well as a score breakdown so the user can see how the area scores in the different sections that are used in the ranking. The area that the user's cursor is hovering over is also highlighted on the map with a grey border around the area.

When an area is clicked on, the map will also zoom in to fit this area on screen to give the user a closer look at the selected region, a popup is also opened that displays three options in regards to searching for properties in the area as seen in Figure 7.

3.3.6.2 Property Locations

Figure 7. Set of images showing the process of searching for properties on the interactive map.

The aim for the house property search was to use the API of a housing property website such as Daft.ie to return the location of properties and display them on the map. It was not possible to use the Daft API unfortunately, as the API is only available to letting agents and not to the public. Other property listing sites such as MyHome.ie and Property.ie also did not have accessible APIs. As there were no APIs available a different approach to the problem was taken, the idea of popup iframe was the main idea. This would use the name of the area from the Census, to try generate a URL that would load a webpage with properties in that area and display them in an iframe popup on the map. However, this approach was not very successful either as many of the names given to the areas from the Census did not match up to the locations that the property sites used for locations. Ultimately, this approach was not satisfactory as it did not fit the original aim and did not work for a large amount of areas.

After exhausting these options, it appeared, as there was not a satisfactory solution to solve this problem, however further research provided information on gathering data from Daft. Using an open endpoint that Daft uses for ajax requests on its map overlay a limited amount of data can be gathered from generating a URL that finds properties between two coordinate locations that are either for sale, rent or sharing. The information returns the coordinate locations of the properties and some details about the listing along with a link to the listing on Daft.ie. This provided enough information in order to meet the original goal of displaying the properties on the map, although the lack of full API access restricted the ability to filter the properties recommended to the user.

The process of displaying properties is shown in Figure 7, the user first selects the type of listing for which they are searching. All the locations listed in the area are displayed as markers, which can be clicked on to display more information about that individual property. Rather than just load properties listed within the selected Census small area, the area returned is approximately 3 kilometres in diagonal diameter. The reason behind this is that many of the small areas can be very small in area and are unlikely to have a listing within the bounds of the small area but may have many in close proximity.

3.3.6.3 Top 15 Properties

A listing of the top 15 best areas are displayed below the map, this is to help the user locate the highest scoring locations on the map in regards to their search as it can be difficult to pinpoint the exact best location from just viewing the map. A button is provided which zooms to the location of the area on the map when so the area can be easily located (See Figure 8). The list is limited to just 15 as the map is the intended use method to search for locations rather than having an 18641 long list with all the areas ranked.

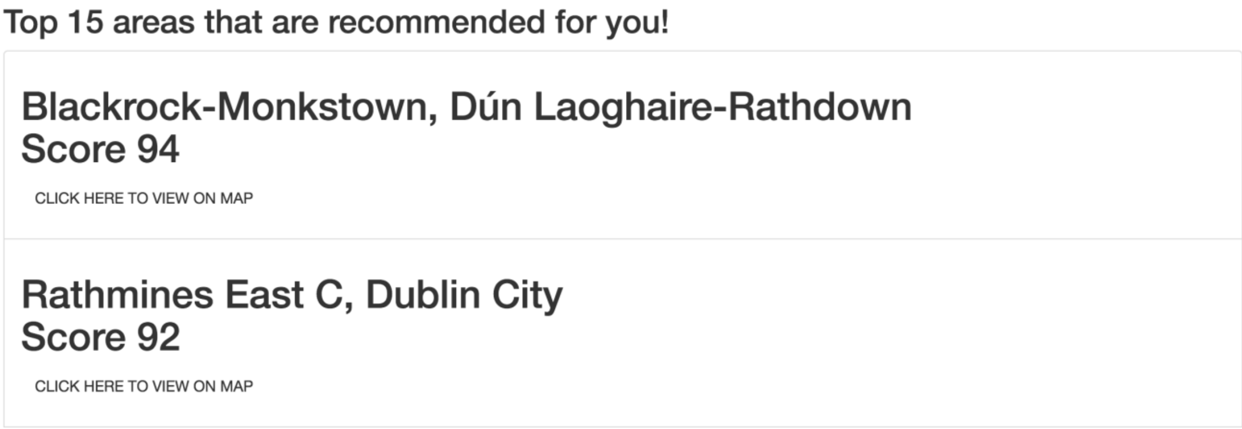


Figure 8. Top two areas displayed from the top 15 areas list that provides the highest scoring areas to the user in a list along with a button to locate them on the map.

4 Project Implementation

Building on the discussion in the previous section, the project implementation aims to provide a detailed discussion surrounding the implementation of the project code for each of the areas discussed in the project design.

4.1 System Overview

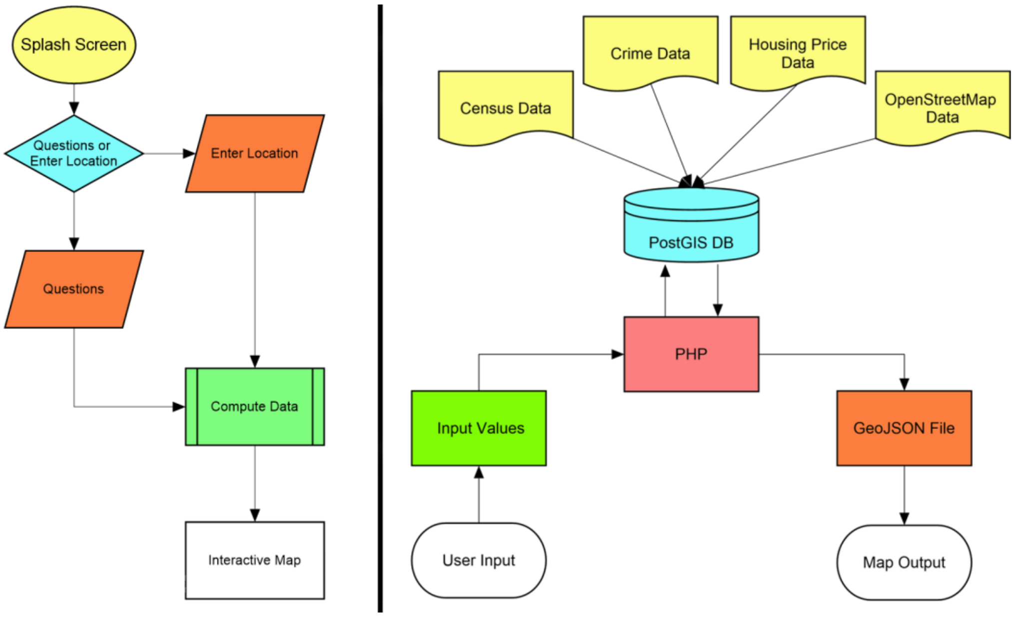


Figure 9. On the left hand side the basic flowchart of the webpage is shown, while on the right a detailed view of the sub process 'Compute Data' is shown.

Figure 9 shows a flow chart for the structure of the components of the implementation. As mentioned in the previous chapter, users have the option between taking a set of questions to determine their ideal location and using a location to find locations similar to the given area. The sub process in Figure 9 shows the flow of data through the system in the construction of the final output to map.

4.2 Data Implementation

As discussed earlier, the tables that are queried from the database are made up of pre-processed spatial information in order to reduce the time taken to process all the queries that are ran. These tables were produced by connecting all the datasets to the Census small areas using the 'gid' which is primary key of the Census dataset as the unique identifier for each area. There are two different types of spatial dataset found in the database, the OpenStreetMap datasets where the geometry is represented using coordinate locations as points and the Garda Crime Data and House Prices by Eircode that represent bounded areas with the use of polygons, similar to the Census small areas.

The point data was connected to the small areas by firstly finding the distance between every point within the data and every Census small area. This was done using the **ST_Distance_Spheroid()** function, which returns the minimum distance between two coordinate geometries. This query was obviously very inefficient and took several hours to complete, however it only had to be ran a single time so it has no effect on any part of the project. The table produced has columns to store the identifier for the Census small area, the name of the structure, the type of structure it is and then finally the distance it is from the small area as seen in Figure 10. Following this more specific tables were created based on the type of building. As mentioned above, the table consisted of either the count of all the structures nearby or the minimum distance to a given structure. These tables consist of exactly 18641 rows and are the pre-processed tables accessed by the queries from the user input. These tables only store the 'gid' of the small area and count or minimum distance of the related structures, this allows for the rapid access of the table.

	gid integer	name character varying(100)	fclass character varying(20)	distance double precision
1	1	Tesco Metro	supermarket	104133.827807854
2	1	Applegreen	kiosk	208753.416405191
3	1	Kiosk	kiosk	205463.362456823
4	1	SuperValu	supermarket	193498.040329605
5	1	Tesco Clare Hall	supermarket	199987.812465983

Figure 10. Table showing the distance between buildings from the OpenStreetMap Points of Interest and the Census small area identified by the 'gid'.

For the datasets with bounded areas a different approach had to be taken. As the boundaries of the different datasets do not match each other, the data between the layers does not directly match the small areas. In order to find the areas from the other datasets that contained the Census small areas the **ST_PointOnSurface()** function was used, this returns a point guaranteed to lie on the surface of the given area. By creating a table containing a point for each small area it was then possible to link both the Garda Crime Data and House Prices by Eircode to each individual small area by using the **ST_Contains()** function. This query found all the small area points within each area of the other datasets and joined each small area 'gid' to that data from the other dataset to create the house pricing and crime statistics tables.

4.3 Question Input and Scoring

As previously mentioned the question input method uses sliders to input the user's feedback to the questions. The sliders represent an integer value between 1-100. Each sliders value is submitted through the use of the PHP POST method. For each question a query is run to return the relevant data. The slider input value between 1-100 does not directly relate to the values found within all the datasets, so in order to scale the range of the values to reflect the queried data, a calculation is performed for each question. A number of different types of calculations are used for the different sets of data retuned by the queries.

Exponential functions are one method used to modify the range. The questions that use this method query the percentage rate of an area such as unemployment levels or population density, as well as the datasets storing location count information such as the number of schools near the small area. By finding the maximum, minimum and average values in the dataset we can choose suitable values to use for the range. Using these points we can get the line of the equation and then find the A and B values for the general form on the exponential equation $\mathbf{y=AB^x}$, where 'X' is the value the user inputted from the slider. The value obtained from the formula gives a value relevant to the dataset being queried. The score is then calculated by dividing the value returned from the equation by the value from the query, if the area has a higher value than the result of the formula the score is capped at 100. For the distance questions the value is converted into metres by multiplying the slider input by 1000 to convert the values into metres. The value is converted to metres as this is the unit of length used by the query. Before converting the slider value to metres however the value is inversed so that a higher score equals to a lower distance. In order to calculate the score, the value inputted by the user is compared to each areas distance. Similarly to exponential function method, the score is capped at 100. The final method for evaluating the scores is mostly used by the data found in the Census. This simply involves selecting the data values we want from the dataset and comparing those values to the total amount of values within that dataset. The percentage found is then directly compared to the slider value to generate the score.

The scoring system has two steps, first a non-weighted score for each question is generated using the slider input and the returned values from the query as mentioned. Then the weighting calculation is performed. This is accomplished by firstly adding all the slider values together so there's a total input value of all the slider values. This value along with the slider input and unweighted score are sent to the **calc_weighted_score** function. The weight for each question is calculated by dividing the input value by the total input value, then multiplying the unweighted score by the input weight. The value returned from this function is the weighted score for the question. All the weighted scores are stored in an array and the sum of this array is found after all the scores have been calculated. This sum ranges between 0-100 and is the score used to rank the areas shown on the map.

4.4 Area Compare Input and Scoring

As discussed previously the input for the area compare search method uses a search bar with an autocomplete dropdown list, which then finds the coordinates for the address entered. This was created by modifying Googles Place Autocomplete API [21] to include the Google Geocoding Service API [22]. This was done by taking the **geocodeAddress()** function from the Geocoding API and implementing to the Autocomplete API so that the selected address from the dropdown list was automatically geocoded. When the user submits the address they have chosen, the coordinates that have been generated are sent using the PHP POST method to the PHP file. The program then attempts to locate the small area containing the coordinates. If the coordinates are not found to be within a small area, the user is returned to the input page and alerted to the unsuccessful search.

Unlike the question based input there are no questions directly asked to the user, however the same library of questions are used to evaluate the areas compared to the selected area. As the comparison is between the profiles of two areas, two queries are performed for each question. The first query is used to acquire the information for the selected area and then the second for all of the areas similarly to the question input. There is no user weighting in this method unlike the question method, as a result in order to evaluate the scores of the area, the similarity of the two areas is compared by finding the percentage difference between the values returned by the queries. The percentage difference between the two values is measured using the calculation as seen in Formula 2. The score is then found by subtracting the percentage difference from 100. The selected area will always have a perfect score for all the evaluations performed. To prevent negative scores that occur during significant differences between values, negative values are assigned the lowest score which is zero. All the scores are added to an array and the overall score for each area is calculated using the average of all the values contained in the array.

$$\frac{|V_1 - V_2|}{\frac{(V_1+V_2)}{2}} \times 100$$

Formula 2. Percentage difference calculation between two values V_1 and V_2

4.5 PHP Generated Output

The output of the PHP script for both methods of input is a generated GeoJSON structure. The GeoJSON structure is made up of objects called 'features' that that combine geometric data and attributes. This makes it possible to store the scoring and name information alongside the spatial data for each small area. The structure is created by constructing an array for which the content can be read as valid GeoJSON. A loop that contains all the scoring evaluations previously mentioned increments through each Census small area. Each area is represented as a feature object. A string is generated once the scoring evaluation for the area has been completed. This string creates the feature object that contains the unique identifier for the area, the scores, the name of the area and the geometry data. Once the string is created it is added to the array storing the GeoJSON structure.

As the features are stored using a 'FeatureCollection' they can be stored as single string. Using the PHP **implode()** function a single string is created from all the strings in array. This string is then passed into a JavaScript variable that is used to display the data on the interactive map.

4.6 Interactive Map

The Leaflet JavaScript library provides the necessary functions required to produce the map and also handle the interactive properties for the map. The GeoJSON layer generated from the input is parsed using the **L.geoJson()** function. Each feature is loaded from the GeoJSON file, before adding to the map a style is applied, this creates the properties for the visual appearance of the feature, most importantly the **getColour()** function is used to assign the colour of the feature according to the range in which the score for the feature is found. The colour scale increments in values of 10, with scores above 90 receiving the darkest shade of colouring and scores below 30 receiving the lightest.

Interactive features such as highlighting featured areas and zooming to them are also implemented using functions from the Leaflet library. The other elements displayed on the map are also managed by the Leaflet libraries, the 'Score Legend' and 'Area Info' displays are displayed in a fixed positon within the map container by using the **L.Control()** function with the **onAdd()** extension to display a HTML element that displays information relevant to the data displayed on the map.

4.7 Property Location Finder

The property location was built using JavaScript along with some features of the Leaflet library to display the information on the map. As stated earlier, access to the Daft API was not available so instead an open endpoint is being accessed via a URL. In order to generate the URL to return the properties listed in a given area, three pieces of information are required, the type of property listing, the south west coordinates and the north east coordinates. As discussed in the project design, when the user clicks on a location a popup is opened that contains three buttons for the different types of property listing. When one of these buttons is selected the **createLink()** function is called. The type of property listing is found from the value sent from button that is selected. The coordinate location of the popup is recorded when it is first clicked, however as we need to find a south west and north east coordinate values form this point the **getSWNE()** function is called. This function offsets the coordinate values to find a location approximately 1.5km diagonally away from the centre point to find the other two coordinates. The south west coordinates have the offset values subtracted from the initial coordinates while the north east coordinates are found by adding the offset values. With the three necessary pieces of data to generate the link obtained, the values can be concatenated into a string to form the URL.

The generated URL is sent to the **findHomes()** function, here the JSON encoded-data that is found from the link is obtained using a GET HTTP request using jQuerys **getJSON()** function. An issue that occurred during this process was that cross-domain requests using jQuery are blocked by Daft. In order to get around this problem the Any Origin service provided by anyorigin.com was used to bypasses the same origin policy to allow cross-domain requests to be returned. This solution is not ideal; however there was not enough time that could be devoted to developing a more elegant solution. The JSON data returned lists all the properties found within the given zone of the north east and south west coordinates. The details for each property returned includes the address and coordinates of the property location, a link to the listing on the Daft.ie website, a small photo of the property, a description of the property and pricing information.

In order to display the data on the map, we create a marker point using leaflets **L.marker** for each property listing at its given coordinates and then include the remaining information within the marker popup that is shown when the marker is selected. All the markers produced are added to a FeatureGroup created to store the markers together. This allows for the markers to be easily removed from the map when the user runs another property search. This prevents the map from being cluttered with markers.

5 Testing and Evaluation

In order to ensure that the results being produced by the output are valid the various components of the system are tested to ensure that no errors are produced during the input and calculation of values. Then evaluating the output by confirming the scores produced are rational results by comparing with the results to known sets of data.

5.1 Testing Input

In terms of testing the system, the majority of the focus was on the user inputs for the search features. For the question based search all the individual questions where tested. The method for testing involved firstly testing the query being ran to ensure it returned the correct data from the database. Then using the maximum, minimum and median value from the input for from the slider, the calculation of the score for the question was produced to ensure that the value generated was within the expected range. Mathematical errors such as division by zero where also tested at this stage. Similar tests were done for the area search input method scoring calculations.

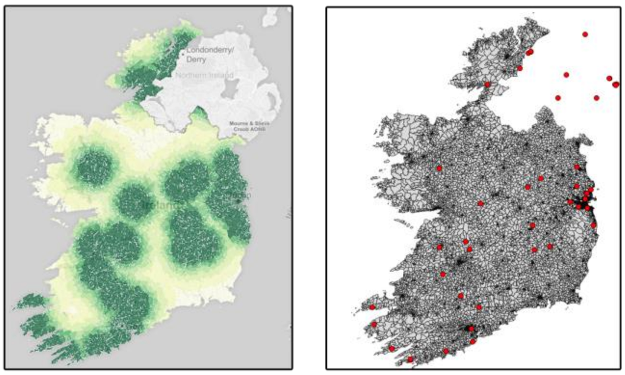
The features of the interactive map were tested in a straightforward manner, for each action available to the user the functionality was tested in a test case manner. For each test, the assessment passed if the expected outcome occurred when the activity was used and failed if that outcome did not occur. An example can be seen in Table 2.

Action	Expected Outcome	Pass/Fail
Area on map clicked	1. Map zooms to fit to that area. 2. Popup with option to search for different types of properties appears.	Pass

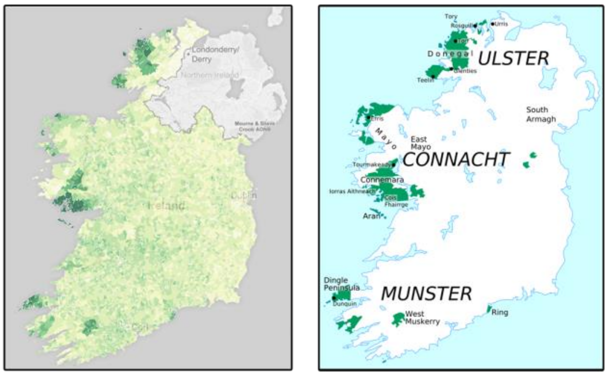
Table 1. Table displaying style of test cases used for the interactive map testing.

5.2 Evaluating Output

As there is 18641 area scores generated it is hard to evaluate the output for all the areas as plain text values. To evaluate the output produced the scores for each individual question were generated and displayed on the interactive map. This allows for the visual comparison of the output to a known dataset that is related to the output. Looking at Figure 11.a we see the comparison of the output of the hospital distance question from the input to the hospital locations that are contained in the dataset used in the query. It’s visible to see that the locations within the user’s desired distance of 25km score high, while the score for areas located further away from hospitals reduce gradually as expected. Similarly in Figure 11.b we see the areas that score the highest for the percentage of Irish speakers are mostly found within the bounds of the known Gaeltacht regions where Irish is the primary language spoken.



(11.a) Maps comparing the output scores of the hospital evaluation to the locations of the hospitals in the dataset.



(11.b) Maps comparing the output scores of the Irish speaking areas to the known locations of the Gaeltacht region.

Figure 11. Set of maps illustrating the evaluation of the outputs compared to a known dataset.

5.2.1 Evaluating Score Weighting System

The evaluation of the scoring weighting system was measured using the same method. Figure 12 shows the output of a user's search whose main interest was to find an area nearby a university or college. The highest scoring areas as expected can be found surrounding the locations of the universities and colleges. However unlike in the previous example seen in Figure 11, there is not a clear gradual reduction of the scoring from the centre points of the buildings as all the other questions are also evaluated too even if they not significantly weighted.

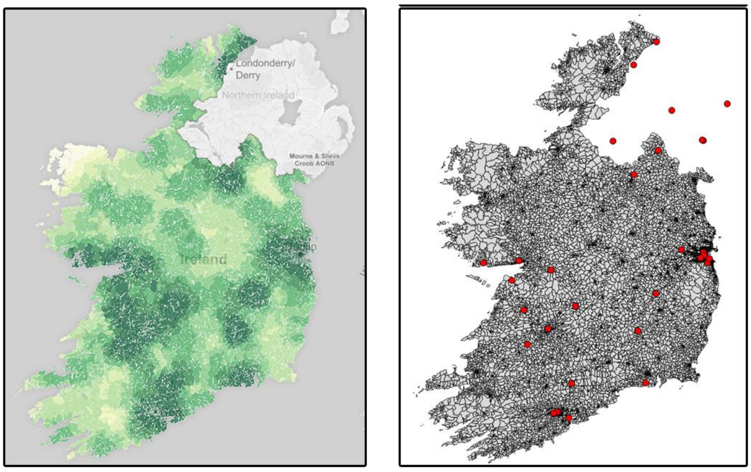


Figure 12. The map on the left displays the total weighted score for a user that valued living near a university or college highly. The map on the right displays the locations of the universities and colleges found within the dataset.

5.2.2 Evaluating Area Compare Search

For the area comparison search, the first step of the evaluation is to ensure that the area containing the inputted location has a score of 100. The selected area should always score 100 as it is directly compared to itself. In order to make sure that the areas that score highly in the comparison are in fact similar, the raw results of the query data can be viewed for each question to ensure that the values are in fact similar. Finally the comparison score for two areas should be bidirectional. This means that if we search for an area A and area B is found to be 79% similar, when we search for an area B, we should expect area A to also be 79% similar.

6 Conclusions and Future Work

The conclusion for this report will assess what has been accomplished from the project specification, the criticisms of the implementation and finally will discuss the possibility for future additions to the project.

6.1 Work Accomplished

The core specifications of the project firstly called for the creation of a set of parameters to create a spatial profile for areas within Ireland. This requirement was achieved by using the Census small areas as the meaningful unit areas to profile. The Census data was then combined with a number of other datasets to further describe the spatial profile of each area.

As the Census small area data is made up of 18641 different areas, many of these areas are very small. The large number of different areas resulted in two issues later in development. The time taken to calculate the meaningful value for each area and then load each area onto the interactive map took a longer period of time than desired. It takes about 5-7 seconds to load the entire process; this time could be reduced by merging many of the very small areas found in the dataset, especially within cities and towns where they are smallest. The small size of many of the areas also makes it difficult to locate many of them on the map. In a re-approach of the project all small areas with an area less than $0.25km^2$ would be merged with similar nearby areas. As over 60% of the areas in the dataset are below $0.25km^2$, this would be a significant optimization that would both increase the process time of the output and also make it easier to view all the individual areas, while not sacrificing much detail.

The second core implementation was to create the web-based interface to allow users to find recommend areas. This requirement was accomplished by creating a questionnaire based input and an area compare search. While the address search bar approach taken for the area compare based search is ideal, the questions found in questionnaire based search could be improved. Further research on how to properly phrase a question would have produced a clearer set of questions that could increase the users understanding of each question. In terms of the calculation of the areas that are recommended, the structure of the code is not ideal. A more uniform approach to the construction of the code used to calculate the evaluated score would have resulted in cleaner code that would have allowed for faster implementation of additional questions.

The advanced features of the project specification required an interactive map be developed to allow users to view the scores produced by the inputs and view the complete profile of the area. The Leaflet JavaScript library provided the ideal tools to allow for the successful development of the interactive map and the displaying of information in regards to the details of the area profiles.

The final specification for the project regarded the development of a property recommender integrated with property listing websites such as Daft.ie, to extend the area recommender. While the inputted data from the user provides information about the type of home they are looking for, due to the lack of a public API tool for the any of the Irish based property sites it was not possible to create a fully-fledged recommender tool. The solution created however, does return all properties within a given range of a selected point, which given the available tools was the best option available.

6.2 Future Work

An extension of the project that could improve the accuracy of the area evaluations would be the use of proprietary spatial data. While the open source data produced by the Government such as the census and crime data are well maintained and accurate, open source data such as the OpenStreetMap location datasets used in this project are created and maintained by volunteers. These datasets accurately contain the majority of the information however it is not uncommon to find mistakes and missing data. There are many proprietary spatial datasets similar to these open source alternatives that are much more reliable and are professionally maintained, however as they are very expensive to access this is a route that would require some substantial funding behind the project.

With full access to a property listing API or if the project itself was to be embedded within a property listing site it would be possible to create a fully developed property recommender. This would make it possible for users to give a greater level of detail in regards to the properties of property that they are seeking. Such a recommender could significantly improve the accuracy of the areas recommended as ultimately people looking to move to a new area are likely to be searching for a combination of both properties and an area, rather than finding the ideal area and then looking for properties within it.

Another possible future use of the tool would be to recommend areas based on other factors rather than solely the most suitable places to live. For example using the system in place and with additional datasets, it would be possible to adapt the search tools created in this project to instead recommend suitable areas to go on holidays.

7 References

[1] Daft.ie. (2017). Rental Price Report, An analysis of recent trends in the Irish rental market 2017 Q3. Retrieved 1st December 2017 https://www.daft.ie/report/2017-Q3-rentalprice-daft-report.pdf

[2] Pope, C. (2017, Novemeber). House prices to rise 20% over the next three years, ESRI says. The Irish Times, Retrieved 1st December 2017 https://www.irishtimes.com/business/economy/house-prices-to-rise-20-over-the-next-three-years-esri-says-1.3290711

[3] “Table 5.1: Regional Quality of Life in Ireland 2013. Central Statistics Office. Central Statistics Office of the Republic of Ireland. Retrieved 1st December 2017. http://www.cso.ie/en/media/csoie/releasespublications/documents/otherreleases/2013/regqualityfull.pdf

[4] USNews. 2017. U.S. News & World Report Unveils the 2017 Best Places to Live. Retrieved 1st December 2017. https://www.usnews.com/info/blogs/press-room/articles/2017-02-07/us-news-unveils-the-2017-best-places-to-live.

[5] Li, G., and Q. Weng. "Measuring the quality of life in city of Indianapolis by integration of remote sensing and census data." International Journal of Remote Sensing 28.2 (2007): 249-267.

[6] Brereton, Finbarr, J. Peter Clinch, and Susana Ferreira. "A spatial econometric analysis of well-being using a geographical information systems approach." (2006).

[7] Elli Thomas, Ilona Serwicka & Paul Swinney (2015) Urban demographics Why people live where they do: centreforcitites.org. Retrieved 2nd December 2017. http://www.centreforcities.org/wp-content/uploads/2015/11/15-11-02-Urban-Demographics.pdf

[8] Dennett, Adam Raymond. Understanding internal migration in Britain at the start of the 21st century. University of Leeds, 2010. Retrieved 7th December 2017. http://www.homepages.ucl.ac.uk/~ucfnard/AdamDennettPhDThesis.pdf

[9] U.S. Census Bureau (2013). Reason for Moving: 2012 to 2013 Population Characteristics. Retrieved 3rd December 2017, https://www.census.gov/prod/2014pubs/p20-574.pdf

[10] Census of Population (2016). Profile 2 Population Distribution and Movements. Retrieved 3rd December 2017. http://www.cso.ie/en/releasesandpublications/ep/p-cp2tc/cp2pdm/im/

[11] Census of Population (2016). Chapter 2 Unemployment. Retrieved 3rd December 2017. http://www.cso.ie/en/media/csoie/newsevents/documents/census2016summaryresultspart2/Chapter_2_Unemployment.pdf

[12] United States Geological Survey. What is a geographic information system (GIS)? . Retrieved 6th December 2017. https://www.usgs.gov/faqs/what-a-geographic-information-system-gis

[13] University of Washington (2013). The GIS Spatial Data Model Retrieved 4th December 2017. https://courses.washington.edu/gis250/lessons/introduction_gis/spatial_data_model.html

[14]Obe, Regina O., and Leo S. Hsu. PostGIS in action (2011). Manning Publications Co. (pp.313-316).

[15] Gardner, S.D. Evaluation Of The Colorbrewer Color Schemes For Accommodation Of Map Readers With Impaired Color Vision (2005). Retrieved 7th December 2017. http://www.personal.psu.edu/cab38/ColorBrewer/Steve_Gardner_thesis_PSU.pdf

[16] Open Source Geospatial tools workshop (2008). GeoServer Tutorial. Retrieved 4th December 2017. http://www.gise.cse.iitb.ac.in/wiki/images/0/02/GeoServer_tutorial.pdf

[17] Qgistutorials.com Web Mapping with QGIS2Web — QGIS Tutorials and Tips. Retrieved 4th December 2017. http://www.qgistutorials.com/en/docs/web_mapping_with_qgis2web.html

[18] Bootstrap.com Designed for everyone, everywhere. Retrieved 4th April 2018. http://getbootstrap.com/docs/3.3/

[19] SurveyMonkey.com What is a Likert scale? Retrieved 4th April 2018. https://www.surveymonkey.com/mp/likert-scale/

[20] Harrower, M., & Brewer, C. A. (2003). ColorBrewer. org: an online tool for selecting colour schemes for maps. The Cartographic Journal, 40(1), 27-37.

[21] developers.google.com - Maps JavaScript API, Place Autocomplete. Retrieved 4th April 2018. https://developers.google.com/maps/documentation/javascript/examples/places-autocomplete

[22] developers.google.com - Maps JavaScript API, Geocoding service. Retrieved 4th April 2018. https://developers.google.com/maps/documentation/javascript/examples/geocoding-simple