Applied Data Science Capstone: Capstone Project

- The Battle of Neighborhoods:

Opening a pub in Aberdeen, Scotland, United Kingdom

Bodø, 23.12.2019

Aleksander Jenssen

Number of pages: 17

Table of Contents

1.0 Introduction......1
1.1 Background......1
1.2 Problem/research question, and interested stakeholders......2
2.0 Data......2
2.1 Part 1: Initial analysis......2
2.1.1 Necessary data......2
2.1.2 Data sources......3
2.1.3 Data preprocessing......3
2.1.4 Data usage/analysis......5
2.2 Part 2: Final analysis......5
2.2.1 Necessary data......5
2.2.2 Data sources and preprocessing......5
2.2.3 Data usage/analysis......6
3.0 Methodology......6
3.1 Part 1: Initial data analysis: Exploratory data analysis......6
3.1.1 Exploring *ab_popdata* distributions......6
3.1.2 Visualizing the locations of Aberdeen City´s wards......7
3.2 Part 1: Data analysis......8
3.3 Part 2: Exploratory data analysis......9
3.3.1 Relationship between median income and self employment rate......9
3.3.2 Relationship between crime rate and self employment rate......10
3.3.3 Relationship between crime rate and median income......11
3.3.4 Evaluation and final dataset......12
4.0 Results......12
5.0 Discussion......14
6.0 Conclusion......15

List of tables

Table 1: First three rows of the processed dataframe *ab_popdata....4*
Table 2: More information about wards in Aberdeen City.....6
Table 3: *ab_popdata* distributions....6
Table 4: Initial comparison of wards....12
Table 5: Final comparison of wards......13
Table 6: Population of neighborhoods in *Kincorth/Nigg/Cove.....13*

List of figures

Figure 1: Population of Aberdeen´s wards....7
Figure 2: Visualization of Aberdeen City´s wards....7
Figure 3: Clusters of Aberdeen City´s wards....8
Figure 4: Median income vs Self employment rate....10
Figure 5: Crime rate vs Self employment rate......10
Figure 6: Crime rate vs Median income......11

1.0 Introduction

This report is a part of the final deliverables of the capstone project of the Applied Data Science Capstone course from IBM, delivered through Coursera. Through the following chapters I will present a business problem, describe the data which will be used to find a prospective solution to the problem, describe the research methodology used during the capstone project, present my findings, before finally discussing the findings in light of the business problem and concluding on a possible solution.

1.1 Background

The city of Aberdeen is the third most populated city in Scotland, with a population of 227,560 in 2018, according to the Aberdeen City Council (2019). Being regarded as one of the best cities to start a new business in the United Kingdom, due to it´s high survival rate among newly established enterprises, Aberdeen has become a cradle for economic growth in Scotland.

One category of businesses which are abundant in Aberdeen are pubs. The vast majority of these pubs, however, are located close to the city centre, with few pubs situated at the outskirts of the city. This opens up a business opportunity for prospective entrepreneur who are interested in running their own pub in Aberdeen, if they are able to find a suitable area to start their business. A premise for succeeding in a new venture is the ability to attract a sufficient amount of customers to your business, which means that there are certain criteria that needs to be met when determining whether a given area is suitable for starting a business.

In the case of opening a pub in Aberdeen there are two main criteria that should be met, namely that:

1. *There should be few other pubs in the area, and;*
2. *There should be a sufficient amount of potential customers in the area.*

With these criteria in mind, a business problem has been outlined, as presented in the following sub-chapter.

1.2 Problem/research question, and interested stakeholders

With the aforementioned business opportunity in mind, the research question that this capstone project aim to answer is: *In which area of Aberdeen should an entrepreneur open a pub if they seek to minimize the number of other pubs in the area, while maximizing their reach to customers?*

Finding a solution to the business problem/opportunity, by answering the research question, could be of great interest to several stakeholder. The two most prominent stakeholders in this regard are the potential entrepreneurs who can achieve their goal of successfully starting and running their own pub, in addition to the inhabitants of the selected area(s), who then will have access to pubs in their local area and therefore will not have to travel to the city centre for that particular purpose.

2.0 Data

The purpose of this chapter is to introduce and explain the data that is being used during the capstone project. In the following sub-chapters I will introduce, explain and discuss which data was necessary in order to find a prospective solution to the research question, how and where the data was gathered, how the data was pre-/processed and finally used and analyzed. The data analysis performed during the project took place in two stages. Therefore, the chapter is divided into two parts, one for each stage of data analysis.

2.1 Part 1: Initial analysis

2.1.1 Necessary data

In order to perform the necessary data analysis to determine where one should open a new pub in Aberdeen, in order to minimize the number of nearby pubs, while maximizing the number of potential customers in the area, there are certain types of data which are necessary. First, data containing information about the different areas in Aberdeen is needed. This dataset would ideally contain information about Aberdeen´s neighborhood, their respective populations, and location data in the shape of coordinates. In addition to this data, information about different venues in each area of Aberdeen is needed to properly answer the research question and give recommendations on where to open a pub.

2.1.2 Data sources

A rather large amount of information about the areas of Aberdeen can be found in csv-format here, at doogal.co.uk, a website which provides public domain licensed data on postcodes in the United Kingdom. This dataset includes area/ward names, coordinates, respective districts and population, among other things. However, the quality of the population data in this dataset was not deemed of sufficient quality to be used in the data analysis. Therefore, this dataset from Aberdeen City Council was used to generate the correct population for each area. Location data from Foursquare API was used to explore venues in the different areas of Aberdeen. Finally, having selected the best area (ward) to start a new pub in Aberdeen, this dataset on Aberdeen´s neighborhoods was used to determine the best neighborhood in the chosen area/ward, according to previously mentioned criteria.

2.1.3 Data preprocessing

The first major step in preprocessing the different datasets was to load them all into my Jupyter Notebook. Since the population data from Aberdeen City Council was stored in pdf-format, I first had to write it into an excel-file, Population_data.xlsx, containing one row for each area/ward (13 in total) which also displayed the local population, before loading it into the notebook. A similar process was done to load the neighborhood data of Aberdeen towards the end of the capstone project, when all that remained was to determine the best neighborhood to set up a pub, having already determined the best area/ward (see Note/clarification below).

After loading the datasets containing information about Aberdeen´s areas and population into the Jupyter Notebook, they were saved to the variables *ab_data* and *ab_pop*, respectively. Next followed the different stages in preprocessing the *ab_data* dataset: having created the ab_pop dataset myself, there was no need for preprocessing.

First, I examined the shape of the dataset, and discovered that it consisted of 17,064 rows/entries and 46 columns/features. The majority of these features are redundant with respect to the data analysis of this project, so the dataset was reduced to the following features:

1. District
2. Ward
3. Latitude
4. Longitude

To make a single entry for each area/ward I decided to group the dataset by 'Ward' and 'District', and to use the mean value of latitude and longitude as a proxy for locating the geographic center of each area/ward.

Next, I decided to narrow the dataset down to all areas located in the district called Aberdeen City, since the original dataset contain information about postcodes in all of Aberdeen county. The resulting dataset contained 13 rows and 4 columns, a significant reduction. Furthermore, this now meant that my two datasets could be merged, since they both consisted of a single entry for each ward in Aberdeen. The resulting dataframe, *ab_popdata*, now contained one entry for each of Aberdeen city´s 13 wards/area as well as their corresponding districts, latitudes, longitudes and populations. Table 1 shows the structure of the first three rows of *ab_popdata*, excluding its indices.

Table 1: First three rows of the processed dataframe *ab_popdata*

Ward	District	Latitude	Longitude	Population
Airyhall/Broomhill/ Garthdee	Aberdeen City	57.128143	-2.135057	16,664
Bridge of Don	Aberdeen City	57.189205	-2.108137	19,212
Dyce/Bucksburn/ Danestone	Aberdeen City	57.190166	-2.181515	18,827

Note/clarification

Up until this point I have been talking about areas of Aberdeen, occasionally speaking of Aberdeen ´s wards/areas. In Aberdeen a ward represents a voting district for use in elections. I have decided to use Aberdeen´s wards as a proxy for the city´s areas, mostly because of the fact that information regarding Aberdeen´s wards was readily available. I believe that this is justified, in the context of this capstone project, because of the relatively small geographic sizes of these wards: one ward typically is made up of two to five neighborhoods. At the end of the capstone project, when determining the ideal area to open a pub in Aberdeen, I will first determine which ward is the most suitable, before determining the best neighborhood in that ward, based on the criteria mentioned in the introduction.

2.1.4 Data usage/analysis

The dataset(s) presented above were used to make a suggestion for the most suitable area(s) in Aberdeen to start a new pub. In order to determine an appropriate location, I implemented machine learning, specifically k-means clustering, to separate areas with an abundance of local pubs from areas with few or no pubs nearby. If one or more areas appear to be equally well suited, other measures would be taken into consideration. In this regard, the criteria highlighted in the previous chapter will act as a guideline. Should there still be uncertainties related to determining which area is preferable, yet new information must be explored.

2.2 Part 2: Final analysis

This was indeed the case. The following sections will present the data used in the second and final stage of the data analysis related to the capstone project.

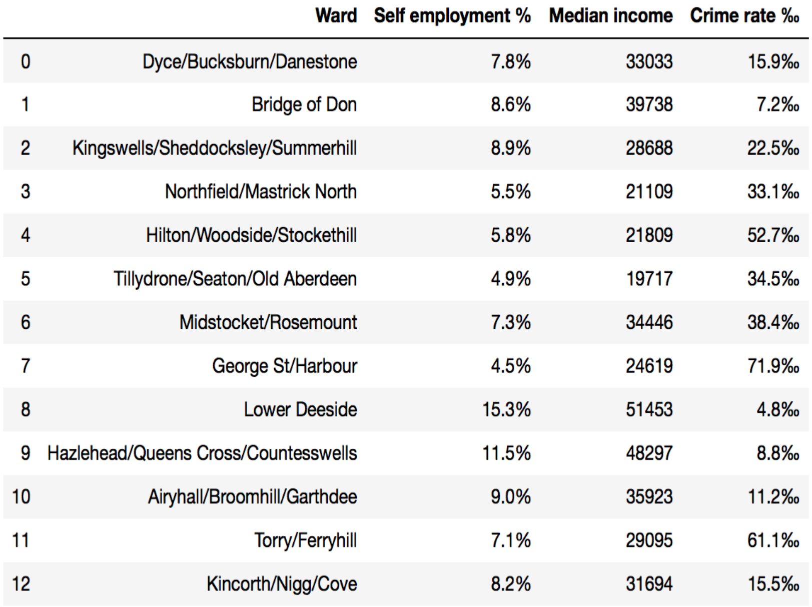
2.2.1 Necessary data

In order to determine which ward of Aberdeen is the most suitable for opening a pub, given that there are similar amounts of pubs/bars in the areas, as well as similar abilities to reach customers, more information about the given areas must be acquired. Specifically, information about the conditions of entrepreneurship is needed. Generally, the rate of self employment is a good indicator of the general conditions of entrepreneurship, as one would generally not be self employed should this not be a viable way of making a living. The median income in an area is a good indicator of the purchasing power of the area´s inhabitants. A high median income could therefore lead to better condition for entrepreneurship, as more potential customers can afford your products/services, and each customer can purchase more products. Finally, the crime rate of a given area indicates the security of the area, and is considered a factor that affects the conditions for entrepreneurship. Data on self employment rates, median income and crime rates of each ward/neighborhood of Aberdeen is therefore necessary.

2.2.2 Data sources and preprocessing

All the necessary information mentioned in the previous section is readily available from Aberdeen City Council. The data on self employment rate, median income, and crime rate in the different wards of Aberdeen City can be found in this pdf-file. To get a better understanding of the data, I wrote it into an excel-file, before loading it into R for processing.

Table 2: More information about wards in Aberdeen City



	Ward	Self employment %	Median income	Crime rate ‰
0	Dyce/Bucksburn/Danestone	7.8%	33033	15.9‰
1	Bridge of Don	8.6%	39738	7.2‰
2	Kingswells/Sheddocksley/Summerhill	8.9%	28688	22.5‰
3	Northfield/Mastrick North	5.5%	21109	33.1‰
4	Hilton/Woodside/Stockethill	5.8%	21809	52.7‰
5	Tillydrone/Seaton/Old Aberdeen	4.9%	19717	34.5‰
6	Midstocket/Rosemount	7.3%	34446	38.4‰
7	George St/Harbour	4.5%	24619	71.9‰
8	Lower Deeside	15.3%	51453	4.8‰
9	Hazlehead/Queens Cross/Countesswells	11.5%	48297	8.8‰
10	Airyhall/Broomhill/Garthdee	9.0%	35923	11.2‰
11	Torry/Ferryhill	7.1%	29095	61.1‰
12	Kincorth/Nigg/Cove	8.2%	31694	15.5‰

The final dataset includes one entry for each of Aberdeen´s 13 wards, displaying their rates of self employment, median income, and crime rates. The dataset is shown in full in table 2, displayed through pandas´ interface in Python.

2.2.3 Data usage/analysis

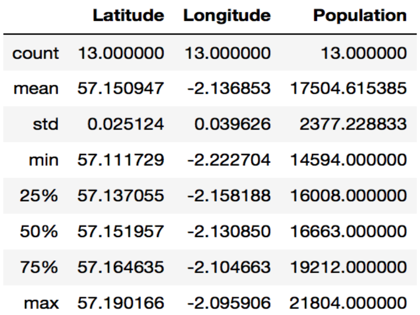
The dataset presented above was used to better be able to determine which area in Aberdeen would be most suitable for opening a pub, given that using the two main criteria gave inconclusive results. Self employment rate was given the highest weight, in front of median income and crime rate. In the end, the numbers suggested that one area was unequivocally better than the rest. This will be discussed later.

3.0 Methodology

Throughout this chapter, I will present the different steps that I have taken to analyze the data during the capstone project. This chapter will present the different types of methodological approaches taken to analyze the data in the two stages of data analysis conducted during the capstone project, and is split into two parts accordingly.

3.1 Part 1: Initial data analysis: Exploratory data analysis

Table 3: *ab_popdata* distributions

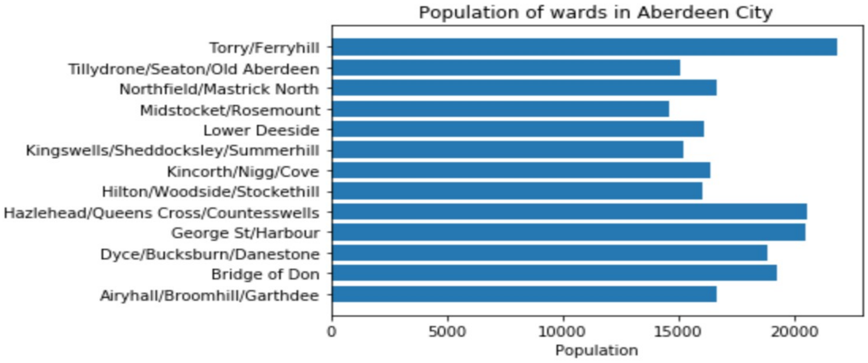


	Latitude	Longitude	Population
count	13.000000	13.000000	13.000000
mean	57.150947	-2.136853	17504.615385
std	0.025124	0.039626	2377.228833
min	57.111729	-2.222704	14594.000000
25%	57.137055	-2.158188	16008.000000
50%	57.151957	-2.130850	16663.000000
75%	57.164635	-2.104663	19212.000000
max	57.190166	-2.095906	21804.000000

3.1.1 Exploring *ab_popdata* distributions

The first step that I took as part of my exploratory data analysis was to uncover the distributions of the features of *ab_popdata*. Although the latitude and longitude distributions give some insight into the geographical locations of the different wards of Aberdeen City, the most interesting observations are related to the population distribution. Here we can see that the average population in Aberdeen´s wards are 17504, with a minimum population of 14594 and a maximum of 21804. Interestingly, while there is a significant difference between the median population and the third quartile (2549), the median population and the first quartile are very close to each other (655).

Figure 1: Population of Aberdeen´s wards

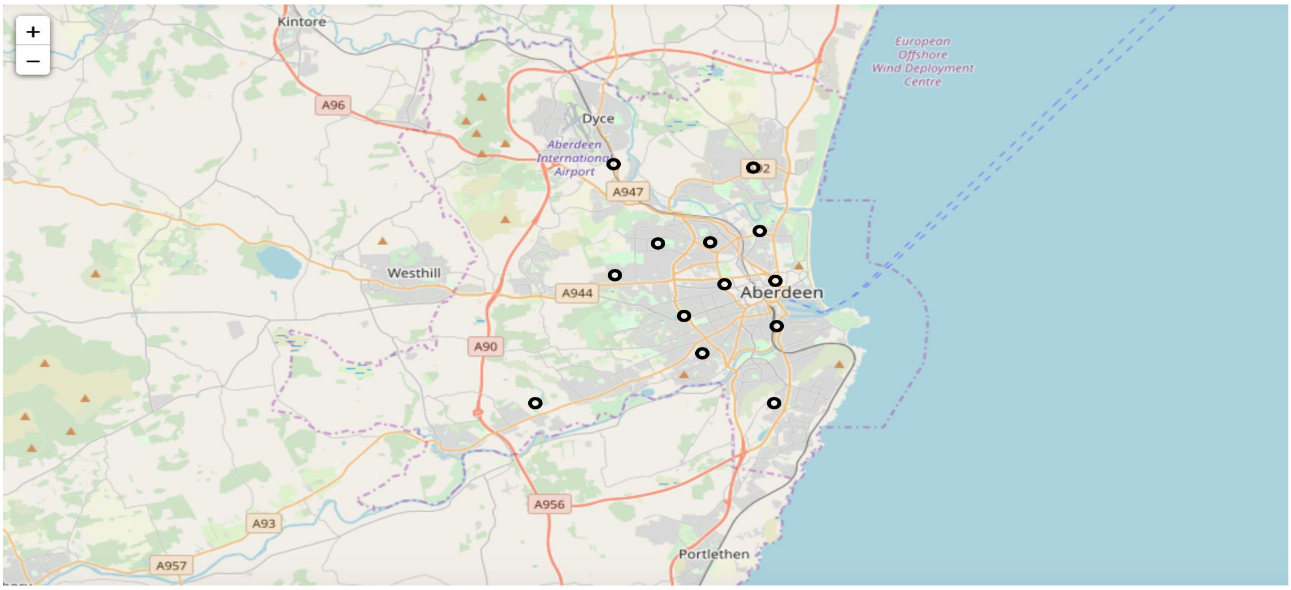


This suggests that there are multiple wards with a population in that range. Figure 1, to the right, shows precise that.

3.1.2 Visualizing the locations of Aberdeen City´s wards

The final step in the exploratory data analysis that took place in the first stage of data analysis was to visualize the location of each of Aberdeen City´s wards on a map of the city. In order to visualize the wards, the locations displayed in ab_popdata was used to create a map of Aberdeen with Python ´s Folium package, superimposing the different wards´ locations on top of the map. The map of Aberdeen City, and it´s wards is displayed in figure 2 below.

Figure 2: Visualization of Aberdeen City´s wards

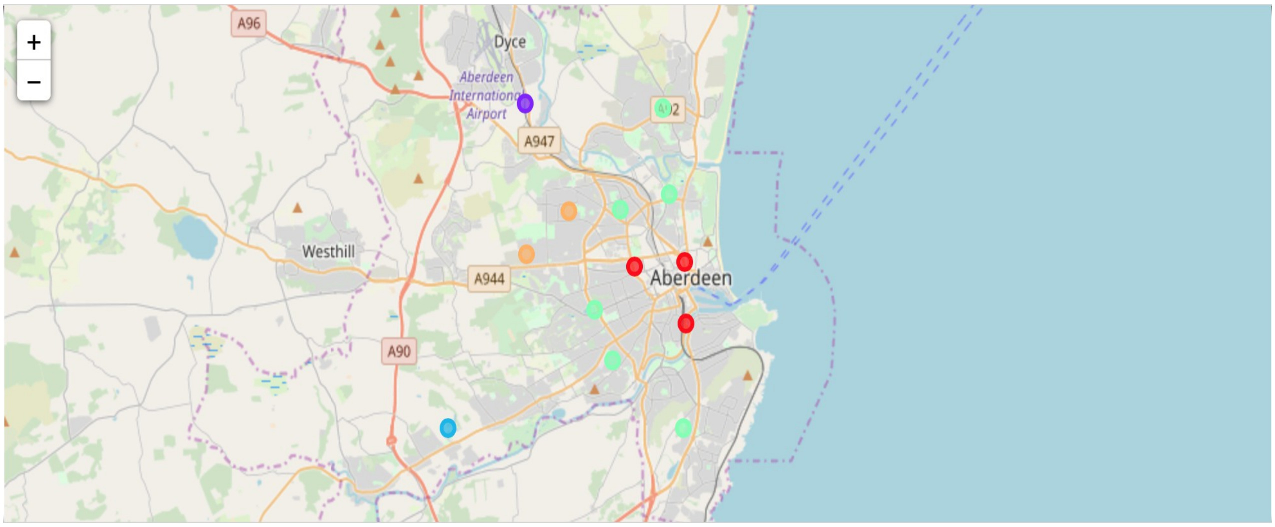

+
−
Kintore
European Offshore Wind Deployment Centre
A96
Dyce
Aberdeen International Airport
A947
2
Westhill
A944
Aberdeen
A90
A956
A93
A957
Portlethen

The most important insight to derive from the visualization is how the wards are spread throughout Aberdeen City. It is worth noticing that three wards make up the city centre, six wards make up the suburbs of Aberdeen City, while the final four wards make up the outskirts of the city. The latter four are the ones we are presumable interested in, though further analysis will reveal if this is indeed the case.

3.2 Part 1: Data analysis

In order to determine which of Aberdeen City´s wards would be the most suitable for opening a pub, I decided to apply K-means clustering to the location data from Aberdeen that I acquired from Foursquare. My decision of using machine learning, specifically K-means clustering, to attempt to answer the research question, came down to the functions of clustering algorithms; namely that they are well suited for finding and determining categories that unlabelled data can be placed into. This meant that such algorithms could find patterns in the types of venues in the different wards of Aberdeen. In an ideal situation, the algorithm could have exposed areas of Aberdeen where pubs and bars are abundant, and also area with few or no pubs or bars. Interestingly, when using the k-means clustering algorithms on the Foursquare data, clear patterns in the venues of Aberdeen´s different wards were visualized. Figure 3 shows how the k-means clustering algorithm placed Aberdeen´s wards into five different clusters, visualized on a map of Aberdeen City.

Figure 3: Clusters of Aberdeen City´s wards



Here, too, there are some very interesting observations to be made. The most apparent observation is that the clusters seem to be fitted according to their respective distances from the city centre (red

< green < orange < purple < blue), despite the fact that this data has not been used in the clustering. This suggests that similar venues are located in similar types of areas in the city, e.g. city centre, suburbs and outskirts.

Having determined the clusters using the K-means algorithm, and visualized their locations on a map of Aberdeen, it was time to inspect the different clusters. After a visual inspection, it was apparent that none of the clusters had manage to exclusively capture wards without pubs or bars. Through further inspection it was revealed that only two wards in Aberdeen City contained no pubs nor bars among their ten most common venues, *Hilton/Woodside/Stockethill* and *Kincorth/Nigg/Cove* (the latter contained at least one hotel bar, but this was disregarded because of the general dissimilarity between hotel bars and pubs). Examining these two wards in light of the projects two main criteria, shown in chapter 1.1, it was suggested that further analysis was needed in order to properly answer the project´s research question. Therefore, more criteria had to be considered before recommending the most suitable area to open a pub in Aberdeen.

3.3 Part 2: Exploratory data analysis

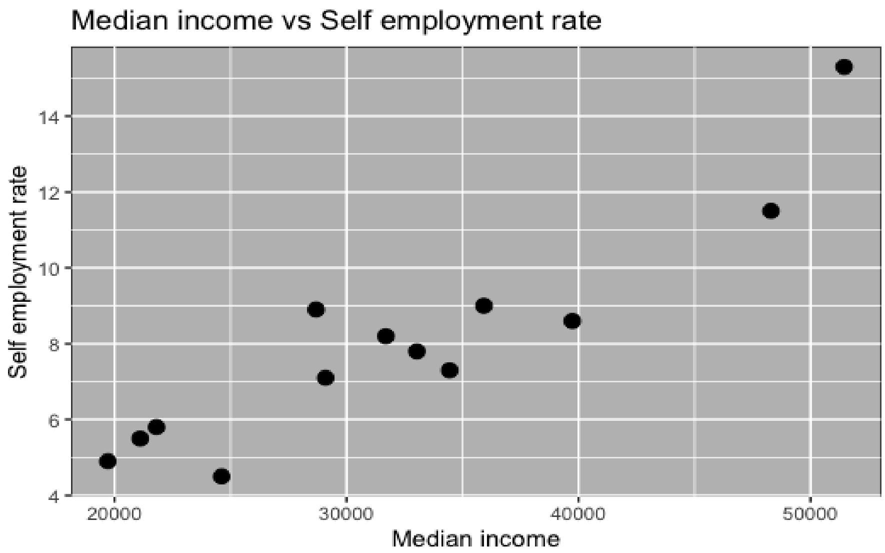
The second stage of data analysis was almost entirely of an exploratory nature. For that reason, the entire last stage of the data analysis will be covered during this subchapter. As previously explained, more criteria had to be taken into consideration before being able to determine which area of Aberdeen is the most suitable to open a pub in. Specifically, data concerning the conditions for entrepreneurship in the different wards would be examined.

In this regard, the self employment rate of the different wards is considered the best indicator for the conditions of entrepreneurship, as a high rate of self employment could suggest that entrepreneurship is a viable way of life in a given ward, whilst a low rate of self employment would suggest the opposite. Additionally, factors affecting the self employment rates of the different wards was explored, namely median income, and total crime rate (see section 2.2.1 for justification). To determine if these factors indeed could affect the rate of self employment in Aberdeen City´s wards, different regression analyses were performed, as presented in the following sections.

3.3.1 Relationship between median income and self employment rate

Figure 4 shows the relationship between median income and self employment rates in the thirteen wards of Aberdeen City.

Figure 4: Median income vs Self employment rate

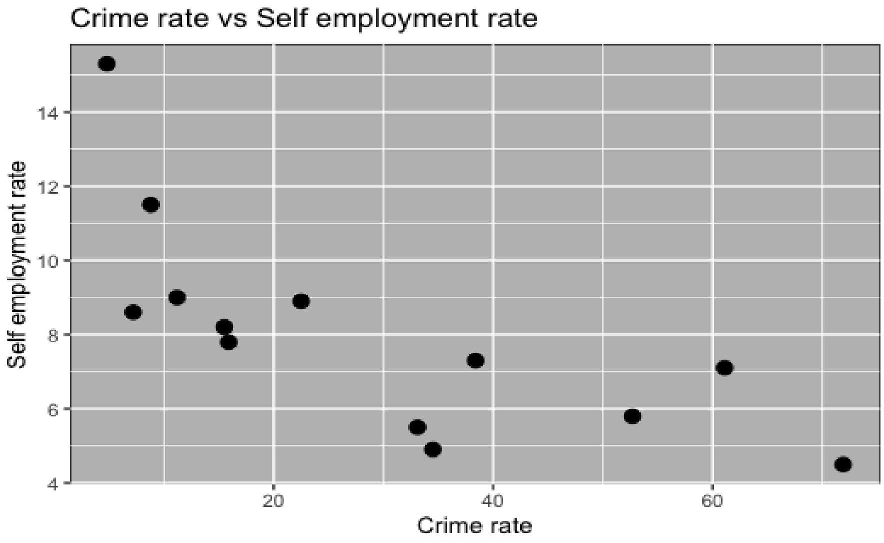


Immediately, there seem to be a somewhat linear relationship between the median income and self employment rates in the wards of Aberdeen. When building a linear regression model in R, I got the following results: R = 0.921, $R^2 = 0.849$, adjusted $R^2 = 0.835$, with a very high statistical significance (F-test, $p = 7.76 * 10^{-6}$). This signifies a relationship between median income and self employment rate that is approximately linear.

3.3.2 Relationship between crime rate and self employment rate

Figure 5 shows the relationship between crime rate and self employment rates in the thirteen wards of Aberdeen City.

Figure 5: Crime rate vs Self employment rate

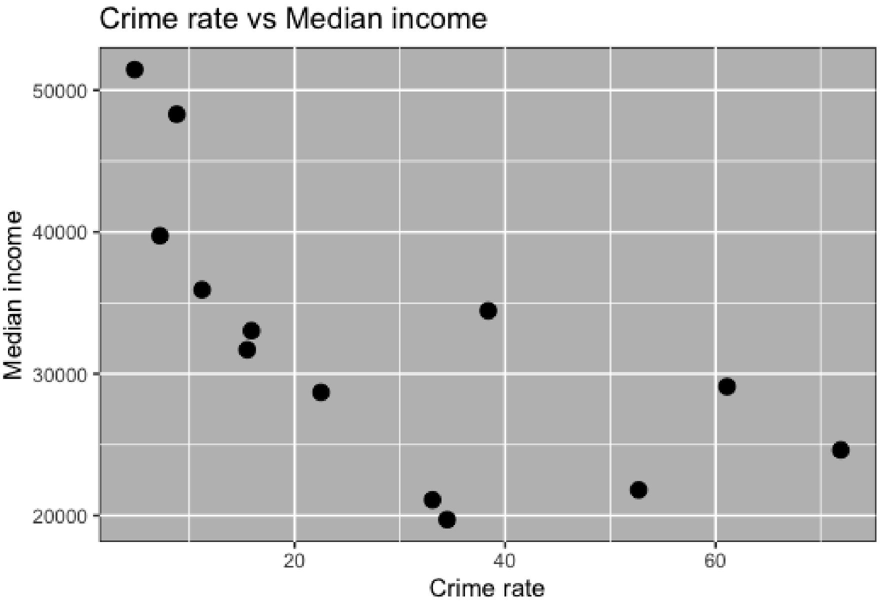


In this case, the linearity of the relationship between crime rate and self employment rate is not as obvious as in the last case, although there seem to be a distinctly negative (logarithmic) relationship between the two variables. After building a regression model, this is confirmed. There is a statistically significant negative relationship between the two variables (F-test, p = 0.0058), but it is only as decent linear fit: R = 0.717, R^2 = 0.514, adjusted R^2 = 0.470.

3.3.3 Relationship between crime rate and median income

I decided to test the relationship between crime rate and median income to check whether there exists collinearity between the two independent variables. Despite the fact that I am not building a model to be used to make predictions, it is still useful to know whether the two variables influence each other. Figure 6, on the following page, shows the relationship between crime rate and smedian income in the thirteen wards of Aberdeen City.

Figure 6: Crime rate vs Median income



Again, we see that the linearity of the relationship between the crime rates and median income is not strong, but that there is a general negative trend, with some outliers. There is still a statistically significant relationship between the two variables (F-test, p = 0.010), but the linear fit is not great: R = 0.680, R^2 = 0.463, adjusted R^2 = 0.414.

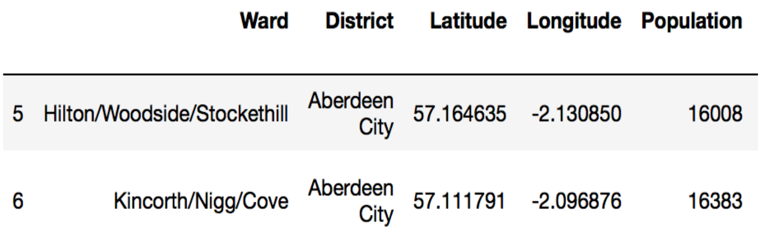
3.3.4 Evaluation and final dataset

Examining the results of the plots and regressions above, it is determined that median income and crime rate do indeed affect the rates of self employment in Aberdeen City´s wards. Therefore, these three variables will be used to determine which of the two aforementioned wards of Aberdeen City, *Hilton/Woodside/Stockethill* and *Kincorth/Nigg/Cove*, is best suited for opening a pub in. The final dataset, showing one entry for each of the two wards, with the three variables, population, drinking % and estimated drinking population as features, is shown in table 5, on the next page.

4.0 Results

Analyzing the data using the K-means clustering algorithm, with the purpose of determining which wards of Aberdeen are the most suitable to open a new pub, yielded the following results. First, while the algorithm did a good job placing wards with an abundance of pubs and bars in the same cluster, it did not find any clusters solely made up of wards without any pubs or bars as some of their most common venues. Second, after further examination, it was found that only two wards had neither pubs nor bars as one of their ten most common venues, *Hilton/Woodside/Stockethill* and *Kincorth/Nigg/Cove*. As these two ward both fulfilled the first criterium highlighted in the introduction, the next thing to examine was their respective populations. Table 4 summarizes the location and population data for these two wards.

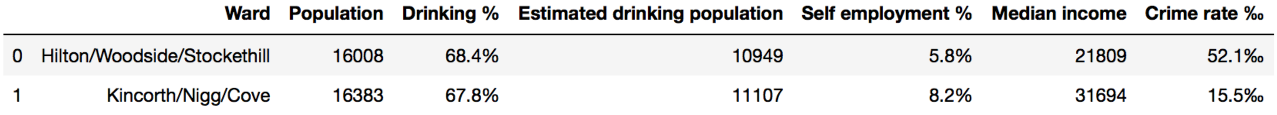
Table 4: Initial comparison of wards



	Ward	District	Latitude	Longitude	Population
5	Hilton/Woodside/Stockethill	Aberdeen City	57.164635	-2.130850	16008
6	Kincorth/Nigg/Cove	Aberdeen City	57.111791	-2.096876	16383

Since there were no significant difference in population between the two wards, further analysis was conducted. This analysis included establishing the estimated drinking population of the two wards, defined as their inhabitants between the ages of 16 and 64. Furthermore, the self employment rate, median income and total crime rate of the two wards were included in the analysis. The results, containing the analyzed features of the two wards, is shown in table 5.

Table 5: Final comparison of wards

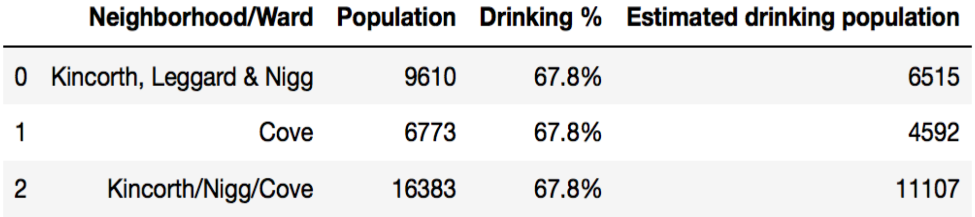


	Ward	Population	Drinking %	Estimated drinking population	Self employment %	Median income	Crime rate ‰
0	Hilton/Woodside/Stockethill	16008	68.4%	10949	5.8%	21809	52.1‰
1	Kincorth/Nigg/Cove	16383	67.8%	11107	8.2%	31694	15.5‰

From this data it is revealed that *Kincorth/Nigg/Cove* has a slightly higher estimated drinking population than *Hilton/Woodside/Stockethill*, as well as a higher self employment rate, significantly higher median income, and much lower total crime rate. According to the suggestion that a high rate of self employment, a high median income, and a low crime rate indicates good conditions for entrepreneurship, the only step remaining was to examine the different neighborhoods of *Kincorth/Nigg/Cove*.

The ward is commonly split into two areas/neighborhoods, namely *Kincorth, Leggart & Nigg* and *Cove.* Again, the two main criteria from chapter 1.1 will be the first ones used to determine if one of the neighborhoods is more suitable to open a pub in than the other. Since neither pubs nor bars are among the most common venues in all of *Kincorth/Nigg/Cove*, this is assumed to be true for the two neighborhoods in the ward as well. Accordingly, the populations of the two neighborhoods will be the primary factor for determining if one of the neighborhoods is more suitable than the other. Table 6 summarizes the population of the two neighborhoods.

Table 6: Population of neighborhoods in *Kincorth/Nigg/Cove*



	Neighborhood/Ward	Population	Drinking %	Estimated drinking population
0	Kincorth, Leggard & Nigg	9610	67.8%	6515
1	Cove	6773	67.8%	4592
2	Kincorth/Nigg/Cove	16383	67.8%	11107

Finally, we can see that the neighborhood of *Kincorth/Nigg/Cove* that has the highest population, and therefore also the highest estimated drinking population, is *Kincorth, Leggart & Nigg*, with approximately 2000 drinking inhabitants more than *Cove*. In the following chapter, I will discuss these results.

5.0 Discussion

The research question that was posed in this capstone project was: *In which area of Aberdeen should an entrepreneur open a pub if they seek to minimize the number of other pubs in the area, while maximizing their reach to customers?*

To answer the research question, two criteria that ideal areas/wards of Aberdeen would fulfill were outlined:

1. *There should be few other pubs in the area, and;*
2. *There should be a sufficient amount of potential customers in the area.*

After analyzing the different wards of Aberdeen City, two wards stood out as the most suitable to open a new pub in. However, when using the second criterium to determine which of the two wards were the most suitable, the results were inconclusive. As a result, three more criteria were added, to ensure that a specific recommendation could be made:

1. *The rate of self employment should be as high as possible*
2. *The median income should be as high as possible*
3. *The total crime rate should be as low as possible*

When these criteria were taken into consideration, *Kincorth/Nigg/Cove* was determined to be unequivocally more suitable area for opening a new pub than *Hilton/Woodside/Stockethill* was. Lastly, when determining which of the two neighborhood of *Kincorth/Nigg/Cove* was the best fit, it was only necessary to consider the respective populations of *Kincorth, Leggart & Nigg* and *Cove.* The former (6515) has an approximately 42% higher estimated drinking population than the latter (4592).

For that reason, the answer to the research question is that *Kincorth, Leggart & Nigg* is the most suitable area for opening a pub in Aberdeen, if your goal is to minimize the number of nearby pubs and bars, and maximize your reach to potential customers. It should though be noted that different approaches to examining the date might yield different results, that analyzing other types of data from the wards of Aberdeen could have the same effect, and that further analysis and inspections should be mad before deciding where to open a pub in Aberdeen.

6.0 Conclusion

The purpose of the capstone project has been to help prospective entrepreneurs who wants to open a pub in Aberdeen determine which area of the city is the most suitable to do so in. Through exploratory data analysis and some machine learning algorithms, it was determined that the most suitable neighborhood in Aberdeen city to open a new pub, if you want to minimize the number of other pubs in the area, while maximizing your reach to customers, is *Kincorth, Leggart & Nigg*, located to the south of the city centre, in a ward called *Kincorth/Nigg/Cove.* While this recommendation holds according to the data analysis performed during the project, it is recognized that further research and analysis should be made before deciding where to open a pub in Aberdeen, should one wish to do so.