Opening A New Italian Restaurant in Leeds (UK) – A Location Analysis

Lewis Lawton

May 2020

Introduction

The city of Leeds is located in the north of England, 272 km north west of the capital city London. In fact, it is the largest city in the county of West Yorkshire and has become the largest legal and financial centre outside of the capital. It first began as a small borough in the 13th Century, but by the 18th Century it had become a major hub for the production and trading of many goods, such as wool. By the mid-20th Century; Leeds had expanded and absorbed the surrounding villages, leading the city to sit amongst the fourth-most populous urban area in England – with a population of roughly 2.6 million.

Known nowadays for its wide array of food and drink and in such a highly populous area; there are, of course, many restaurants to be found in the city of Leeds. The city is also served by 5 universities, ultimately lowering the average age of its residents compared to the rest of Yorkshire or England as a whole. Roughly 10% of the residents of Leeds are between the age of 20 - 24, many of which are likely to go out to eat regularly. Leeds is constantly growing. With a bustling city centre, a young demographic and vibrant nightlife, it is a great place in which to set up a new business.

Business Problem

A local business owner is looking to open a new restaurant in the city of Leeds, the issue is that they are unsure on a specific location in which to open. They wish to open as soon as possible and have spoken to several stakeholders concerning this task. Through conversation and their own market research, they have identified that they wish to open an Italian restaurant specifically. A stakeholder has suggested using location data to decide upon a suitable location to open the restaurant. They have reached out to a Data Scientist in order to complete this task.

Leveraging the Foursquare location data API, this report will endeavour to highlight potential areas where it may be profitable to open an Italian restaurant. A clustering approach will be taken to segment the different areas around Leeds in order to locate an optimal location to open this business.

It will also aim to inform any future strategy for similar business owners considering Leeds as a potential area to set up a new and developing restaurant business. Potential implications of the

appropriateness and use of location data for similar projects in the future will be discussed and highlighted.

Data

A web scrape of the Wikipedia article containing the postcodes (LS) for the Leeds area forms the basis of the investigation. The link to the Wikipedia article can be viewed here (https://en.wikipedia.org/wiki/LS_postcode_area). The table contained within this article also lists the specific neighbourhoods assigned to each postcode within the Leeds area.

The latitude and longitude for each postcode is to be obtained in CSV format from the freemaptools.com website, linked here (https://www.freemaptools.com/download-uk-postcode-lat-lng.htm). This data will be cleaned, and the relevant features and data will be obtained. This data can then be merged with the table containing the postcodes and neighborhoods, allowing further analysis.

The Foursquare Places API (https://developer.foursquare.com/) will be utilised to determine nearby venues located within each postcode and neighbourhood, in order to determine the frequency and distance of similar businesses. This data will be collected for each neighbourhood within the postcodes and form the basis of map plots to help visualise the data. The data from Foursquare will also allow detailed data tables to be populated for use in the segmentation of different areas around Leeds.

A clustering approach using the K-means clustering algorithm will be taken to discover how the different areas around Leeds cluster. The sum of squared distances of samples to the nearest cluster centre will be calculated for multiple values of K in order to ascertain the appropriate number for K. This will be plotted and the “elbow method” will be implemented to decide upon the K value.

The effectiveness of the algorithm will be evaluated following analysis, supplemented by a discussion of the results achieved. Future implications and refinements to the model will be discussed and suggested later in the report.

All statistical and exploratory analysis is to be completed within a Jupyter Notebook running Python (version 3.6). The Pandas library will be utilised to display and store tabular data, Numpy will allow specific stats to be run on this data. Geopy, Matplotlib and Folium will also be implemented in the report in order to visualise results and communicate significant findings.