Comparative analysis of the wards in the city of Derby (United Kingdom) based on venue information, price house and criminality rate

Course: Applied Data Science Capstone Project

Author: MdN Calvo Mateo

06/07/2020

1. Introduction / Business Problem

The city of Derby, situated in the East Midlands region of United Kingdom, had a population of 256,906 habitants in 2019. The city of Derby is divided into 17 wards, or neighborhoods, as it is displayed in Figure 1.

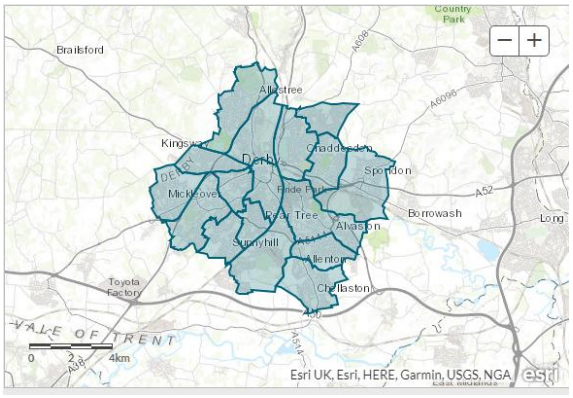

Figure 1. Map of Derby and its wards. Source: https://info4derby.derby.gov.uk/

Derby is one of the cities in United Kingdom with the highest proportion of advance manufacturing jobs across its population. Rolls-Royce plc, Toyota and Bombardier, which are 3 important multinationals dedicated to the manufacturing of aerospace engines, cars and trains, respectively, have manufacturing plants located in or around Derby.

These 3 companies employ a significant amount of employees which relocate to Derby from another countries or another places in United Kingdom for job reasons. One of the key decisions faced by people when relocating it is to decide where to rent or purchase a property to live in. Choosing the right neighborhood or ward according to personal preference is not an easy and straightforward task to accomplish in an unknown city.

This is the reason why this report will focus of comparing the 17 different wards in which the city of Derby is divided, and compare them based on the following criteria:

- Top 10 most common venues and services in each ward.
- Criminality rate.
- Median house price.

This report will be useful for those people moving to Derby or prospective property buyers in Derby since it will enable to have a better understanding of the various wards, the similarity amongst them and their main characteristics. This information will enable the potential property renters/buyers in Derby to make a more informed decision regarding where to choose to live.

2. Data Sources

With the prospects of understanding the selected characteristics of the 17 wards in which the city of Derby is divided, the following data and information has been gathered.

- The venues and attractions in each of the 17 Derby wards, based on the available information in the Foursquare API Database.
    - The information regarding the existing venues in each ward has been accessed via the Foursquare API application.
      Source: https://developer.foursquare.com/
    - The coordinates for the estimated central location for each ward have been obtained from Google Maps and manually inputted into an Excel Document.
      Source: https://www.google.com/maps

- The criminality rate in each of the yards for in the period May 2019 – May 2020, measured as rate of crimes per 1000 population.
    - The information regarding criminality rates per ward in Derby, during the period May 2019 – May 2020 has been downloaded from the "Info4Derby Portal", and this information has been added to the existing Excel document already containing the coordinates for each Derby ward.
      Source: https://info4derby.derby.gov.uk/crime-and-community-safety/reports/

- The mean price paid for all property types in each of the Derby wards during Q1 2019, expressed in GBP. All property types include Bungalows, Flats/Maisonettes, terraced houses, semi-detached houses and detached houses.
    - The information regarding mean property prices per ward in Derby for the period Q1 2019 has been downloaded from the "Info4Derby Portal". This information has been added to the existing Excel document already containing the coordinates for each Derby ward and the associated criminality rates.
      Source:
      https://info4derby.derby.gov.uk/housing/report/view/d867ba1e909244c8ac75f97294cb7b93/E05001767

Once the required data for the analysis has been gathered, this project will analyse this information using a Jupyter Notebook with the programming language Python. The previously described data will be imported as a Pandas Dataframe and once prepared, this project will apply unsupervised Machine Learning methodologies in order to cluster the 17 wards in Derby in various clusters, with the objective of identifying similar clusters based on the criteria describe in the lines above. In this case, the methodology that will be used to analyze the dataset will be K-means Clustering. A detailed explanation of the methodology used in this project will be presented on the next sections of this document.