# IBM DATA SCIENCE CAPSTONE PROJECT

LOCATIONS FOR A BOOKSTORE

## ABSTRACT

This project is to find best locations for starting a small enterprise of a bookstore by using the Foursquare API

Submitted by Athira Govindankutty
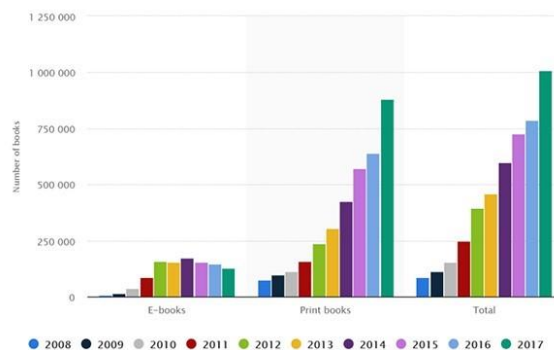
# Table of Contents

# INTRODUCTION

This project is aimed to solve an imaginary problem or a hypothetical situation with the application of Foursquare API and other machine learning techniques. Foursquare API gives the location data and could be used to get details of different venues like ratings, location coordinates etc. Hence it can be used to cluster locations according to their venue categories, finding best suited locations for various business enterprises, housing and many more. In this project I take the opportunity to find a good location to start a small enterprise, a Bookstore.

Books have always had a very crucial part of development of mankind. They are even considered the best companions of one's life as they provide unconditional knowledge. Hence a bookstore or a library that provides such resources are also equally important for a town to bring up civilized and educated youth. As Neil Gaiman said " A town is not a town without a bookstore". During this era of modernisation and digitisation new substitutes (audiobooks and eBooks) for printed books have come to forefront. This seemed to have a dissident impact on traditional bookstores that sell printed copies. But recent surveys show that the impact of eBooks have plateaued and audiobooks are not seen as an alternative but a complement to printed books that come handy while driving, jogging etc. According to the survey conducted by Storaenso a company of renewable materials, it is evident that printed books have a positive future and also as things go more digitised the demands for printed books will also increase.
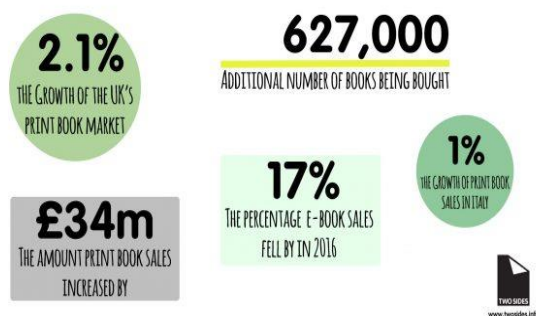


Fig 1: Growth of printed books                    Fig 2: Statistics of printed books and eBooks over the past years

Hence opening a bookstore will always serve the community nearby by providing books as a better option to counterweight the time that a person spends on digital media.UK is a country where a person spends an average of 5 hours and 18 minutes of a week by reading. Thus, opening a Highstreet bookstore chain will also be a good option. In this project I have tried to find locations to open a bookstore in Warwickshire, a county in United Kingdom. Entrepreneurs or High street bookstore chains who wish to open a bookstore in Warwickshire are the targeted audience of this project.

# DATA

Data required to complete the project:

1) Towns in Warwickshire county

2) Location coordinates of these towns

3) Accessibility of bookstores or libraries within a commutable distance

4) For towns without any bookstores or libraries, does the town have a big enough population to start a venture. Also, the presence of schools or colleges in the vicinity so that the store could be useful for students and teachers.

## Data Extraction

The dataset needed for the project is collected from various sources as listed below. Towns in Warwickshire and their population data is obtained from the following site and they were extracted using 'Beautiful Soup' library.

List of towns: https://worldpostalcode.com/united-kingdom/england/warwickshire

Population data: https://www.ebayinuk.co.uk/warwickshire-towns-villages-population/

Location coordinates of the towns were obtained using OpenCage API.

Foursquare API was used to collect details of bookstores, libraries and schools in these towns.

# METHODOLOGY

Using location coordinates of towns in Warwickshire to extract the details of existing bookstores, discount stores and libraries using Foursquare API.

Creating a segregated data-frame of towns without any bookstores or libraries in a radius of 5km from the location co-ordinates.

Extraction of population data for the towns in segregated data frame

Search for location of schools or colleges using the Foursquare API. The towns with more population and schools nearby can be opted as the best option for opening a store.

Towns with bookstores and libraries but are thickly populated are clustered using the K-means clustering to find towns that only have libraries which would be another smart option to open a discount store (Stores that sell second hand books) or a bookstore.

## Flowchart
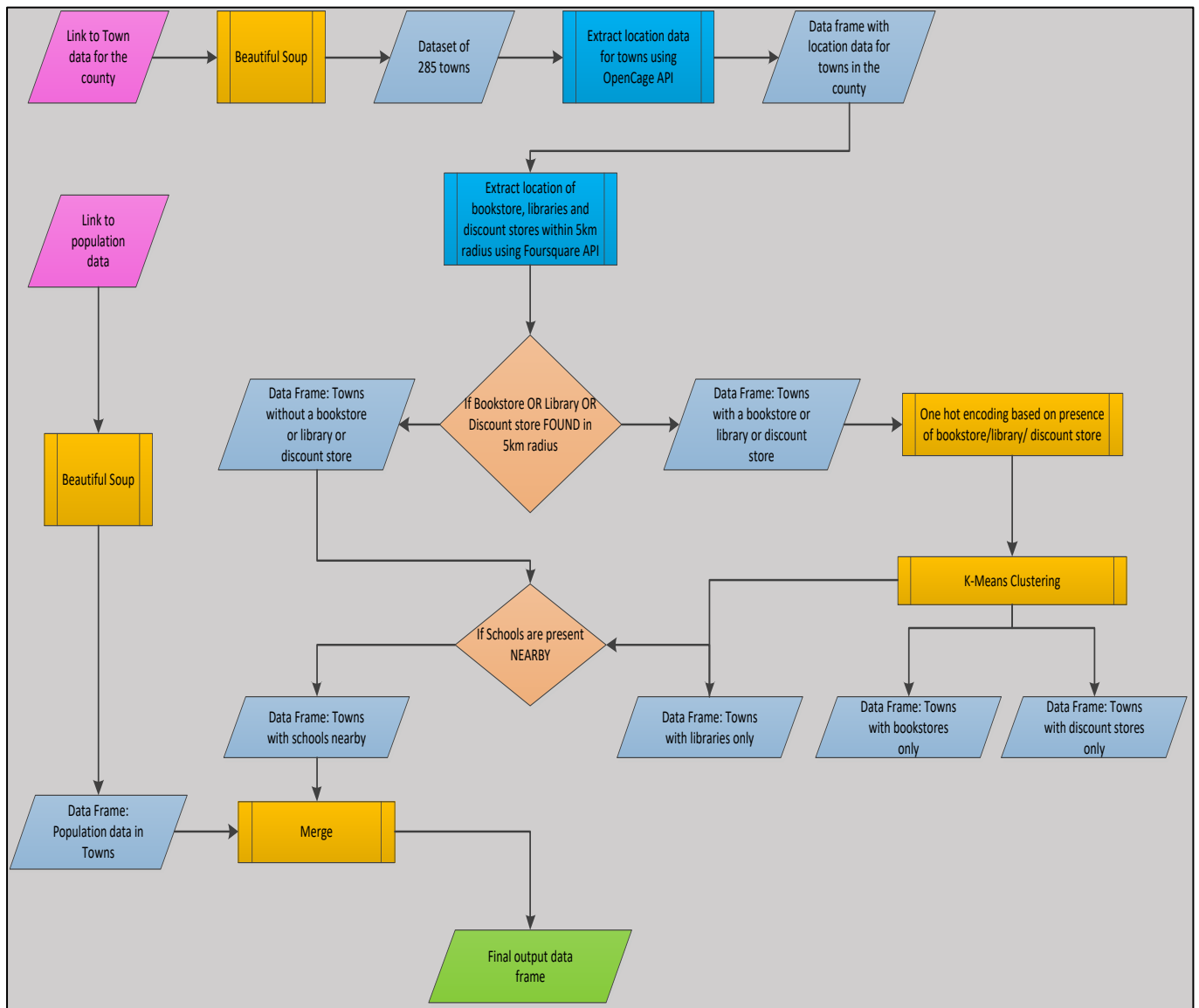


*Fig 3: Flowchart of the project*

## Description

List of towns in Warwickshire and their postcodes are scraped from the site mentioned above and created a data frame with this list. Location details like latitude and longitude of the towns which are necessary to run the Foursquare API were obtained by giving postcodes as the input and using the library pgeocode. It was noted that there were plenty of towns with the same postcode and hence they all gave the same venues/bookstores for all those towns with same postcode. Hence Open cage API was used to get location details of the towns. Town names and country code were given as the

input to get unique latitude and longitude values for each town. Hence a data frame of 285 towns with its location details were obtained.

| | Towns | lat | lng |
|---|---|---|---|
| 0 | Admington | 52.113355 | -1.707347 |
| 1 | Alcester | 52.217618 | -1.868442 |
| 2 | Alderminster | 52.135912 | -1.666328 |
| 3 | Alveston | 52.206518 | -1.660409 |
| 4 | Ansley | 52.521885 | -1.563154 |

*Fig 4: Data frame of towns with location details*

Bookstores in 5km radius of these towns were extracted using the Foursquare API. A data frame of all the bookstores in the town was obtained. After grouping the data with respect to count of bookstores in towns it was found that only 120 towns had bookstores within the specified radius. Similarly, libraries were also taken into account as all the sources that are available to the public had to be analysed. From this it was found that 140 towns had libraries nearby. Therefore, in total 169 towns had bookstores/libraries while 116 towns do not have any bookstore or library within the given distance. By considering the population and the vicinity to schools of these towns we could narrow down the towns which are appropriate to start the venture.

Population data of 140 towns were obtained from the other site specified earlier which is based on 2011 census data. Schools nearby the towns (within 5km radius) were also found using the foursquare API. Out of 116 towns that do not have bookstore/libraries only 49 towns population data was available and only 36 of them had schools nearby. By using the seaborn library, population and No: schools of these towns were graphically represented.

Then towns with bookstores and libraries were taken for further analysis.Bookstore category of all the bookstores and libraries were used to perform onehot encoding,and this data was used to get mean of all the bookstores and libraries in a town.

|   | Bookstore | Discount Store | Library |
|---|---|---|---|
| 0 | 0.000 | 0.000 | 1.00 |
| 1 | 0.625 | 0.125 | 0.25 |
| 2 | 0.500 | 0.000 | 0.50 |
| 3 | 0.500 | 0.000 | 0.50 |
| 4 | 0.000 | 0.000 | 1.00 |

*Fig 5: Data frame after one-hot encoding*

These towns were then clustered using unsupervised technique of K-means clustering .K value was determined by the elbow method as 5.
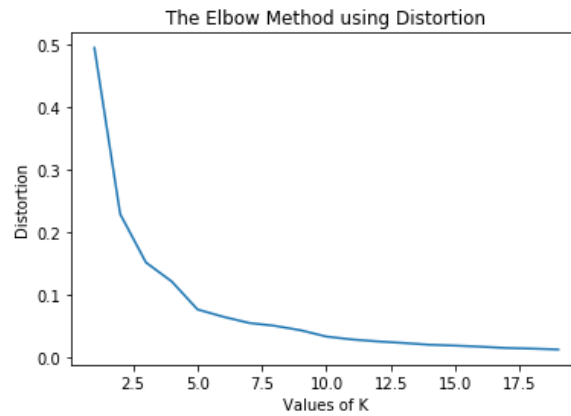


*Fig 6: Elbow method to get K*

| Cluster | No: of towns | Towns |
|---|---|---|
| Cluster 0 | 4 | Towns with discount stores only |
| Cluster 1 | 49 | Towns with libraries only |
| Cluster 2 | 25 | Towns with bookstores only |
| Cluster 3 | 39 | Towns with more libraries and discount stores than bookstores |
| Cluster 4 | 52 | Towns with more bookstores than libraries and discount stores |

The cluster with libraries alone was taken for further analysis as they are a great option to start new bookstores or discount stores.Population details and towns vicitnity to schools were considered as before to narrow down the favourable options and analysis were repeated.

# RESULT

The entire dataset of 285 towns were divided into two, the ones with bookstores/libraries and those without either of them. The latter one was then carefully analysed by combining population data and no: of schools near them.
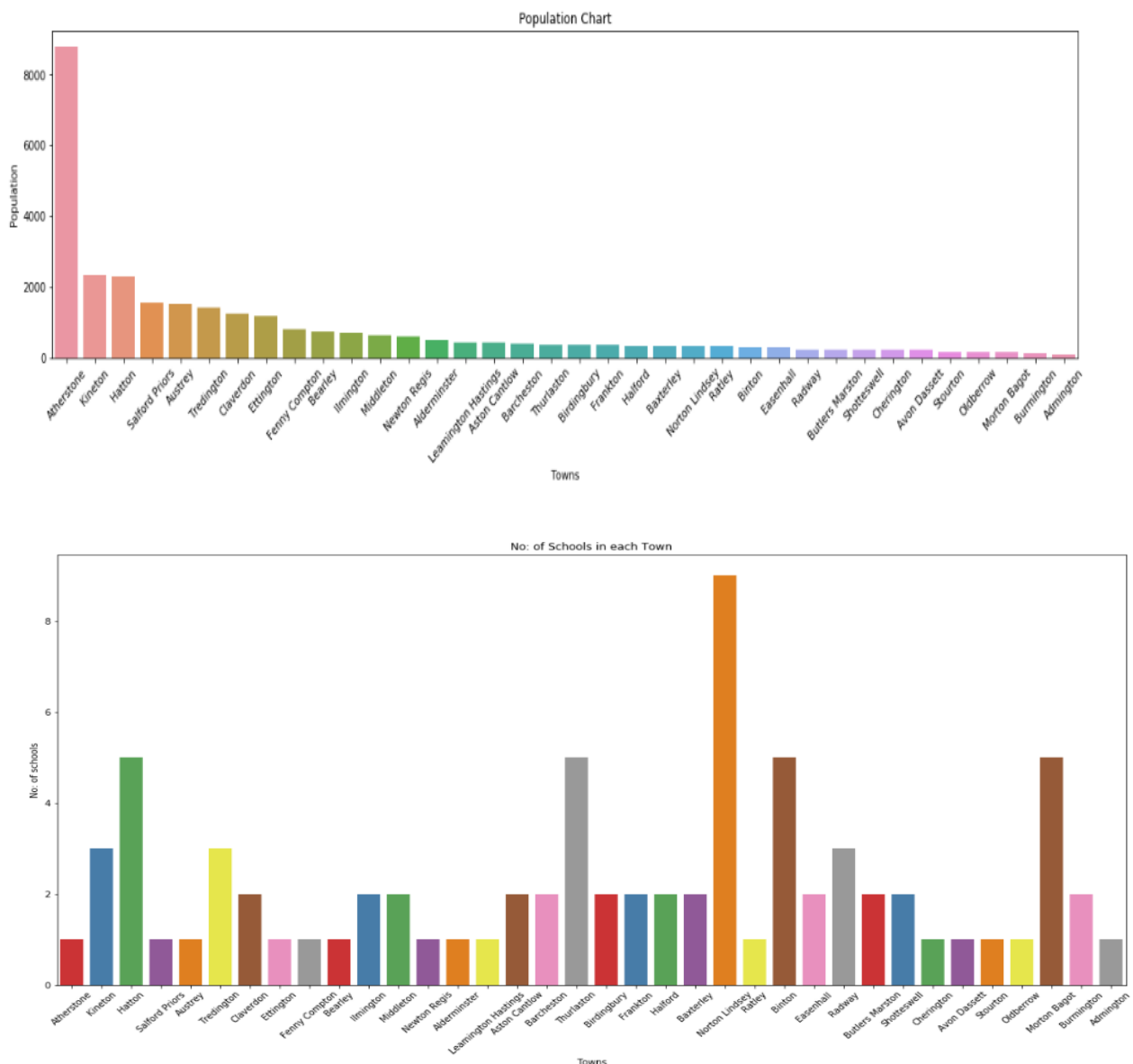




*Fig7:  Graph of population and no: of schools in towns*

From the visualisations it is clear that the best option for opening a bookstore is Atherstone as the population here is very high compared to other towns. However, the school near this town is a nursery. Kineton and Hatton with appreciable population have more than 2 schools nearby hence these towns can also be a suitable option. Hatton has 5 schools nearby and one of them is a high

school similarly Kineton also has a high school nearby. Hence these three towns are best for the purpose.

| Town | Town latitude | Town longitude | School |
|---|---|---|---|
| Atherstone | 52.577244 | -1.543395 | Ashleigh Nursery |
| Kineton | 52.157470 | -1.508069 | Kineton Primary School |
| Kineton | 52.157470 | -1.508069 | Temple Herdewyke Primary School |
| Hatton | 52.295213 | -1.672907 | Hopes And Dreams |
| Hatton | 52.295213 | -1.672907 | Claverdon Primary School |
| Hatton | 52.295213 | -1.672907 | Budbrooke Primary School |
| Hatton | 52.295213 | -1.672907 | Ferncumbe Primary School |
| Kineton | 52.157470 | -1.508069 | Kineton High School |
| Hatton | 52.295213 | -1.672907 | Henley-in-Arden High School |

*Fig 8: Schools list of Atherstone, Hatton and Kineton*

Now looking at the clustered dataset the cluster which consists of towns that only has libraries can be considered as some of their population is greater than the towns that we have already analysed. There are 49 towns that only have libraries and no bookstores but population data is available for 16 towns only.
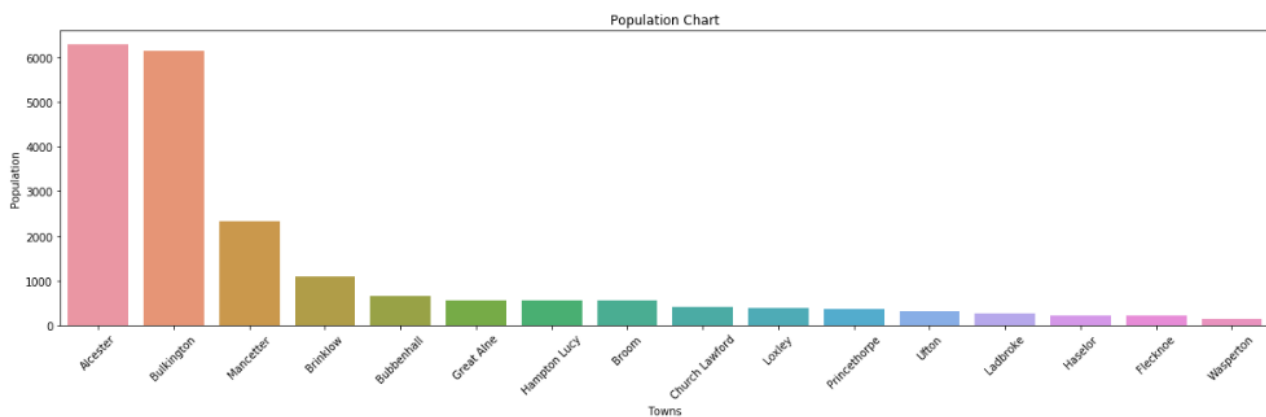


*Fig 9: Population of the towns*
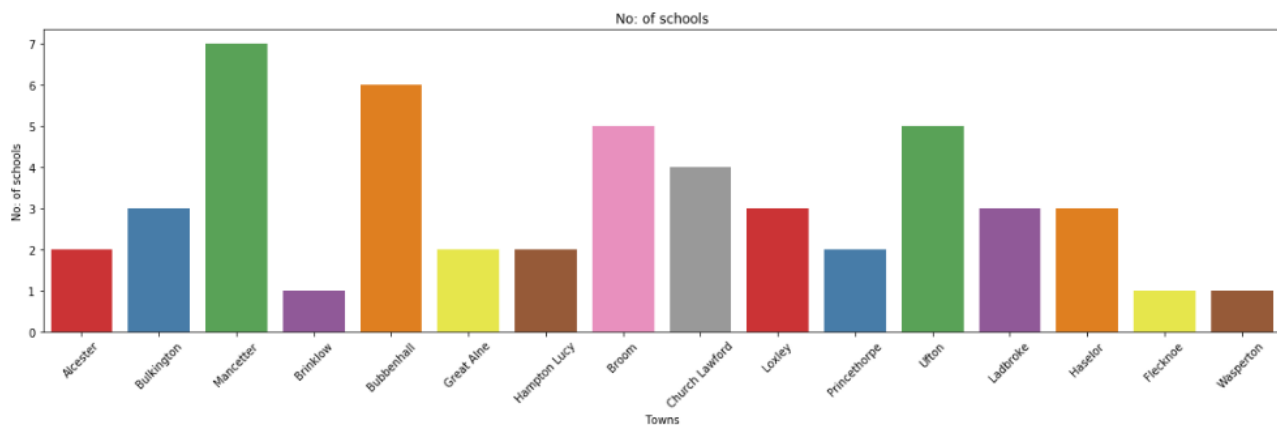
*Fig 10: No: of schools the towns*

| Town | Town Latitude | Town Longitude | School name |
|---|---|---|---|
| Alcester | 52.217618 | -1.868442 | St. Benedict's RC High School |
| Alcester | 52.217618 | -1.868442 | Alcester High School Technology College |
| Bulkington | 52.475791 | -1.423134 | Grace Academy |
| Bulkington | 52.475791 | -1.423134 | Nicholas Chamberlaine Technology College |
| Bulkington | 52.475791 | -1.423134 | Foxford school & community arts college |
| Mancetter | 52.569655 | -1.530255 | Queen Elizabeth School |
| Mancetter | 52.569655 | -1.530255 | Racemeadow |
| Mancetter | 52.569655 | -1.530255 | Michael Drayton School |
| Mancetter | 52.569655 | -1.530255 | Nathaniel Infant School |
| Mancetter | 52.569655 | -1.530255 | Brooklyn Day Nursery |
| Mancetter | 52.569655 | -1.530255 | Ashleigh Nursery |
| Mancetter | 52.569655 | -1.530255 | St Anne's RC Primary School |

*Fig 11: Dataset of schools in Alcester, Bulkington and Mancetter*

As we can see Alcester would be a good location to start a discount store or a bookstore as there are no bookstores in this area and the population is quite higher than the other two towns of Kineton and Hatton. It also has 2 high schools in 5km vicinity. Bulkington would also be a good option but properties for business are very less compared to Alcester in this region. Considering Mancetter which have 7 schools nearby, 5 of them are primary schools and nurseries and population is comparitively low hence the other two options are better.

# CONCLUSION and DISCUSSION

For this project towns of Warwickshire and their population data were scrapped. From these 285 towns, 6 towns that are best suited for starting a bookstore or a discount store was discovered with the help of different libraries and the K-means clustering technique. Using folium library these towns are located in a map.

| Towns | No.of schools | Population | No.of libraries |
|---|---|---|---|
| Atherstone | 1 | 8798 | 0 |
| Alcester | 2 | 6279 | 1 |
| Bulkington | 3 | 6147 | 1 |
| Mancetter | 7 | 2338 | 1 |
| Hatton | 5 | 2317 | 0 |
| Kineton | 3 | 2338 | 0 |

*Fig 12: Shortlisted towns*
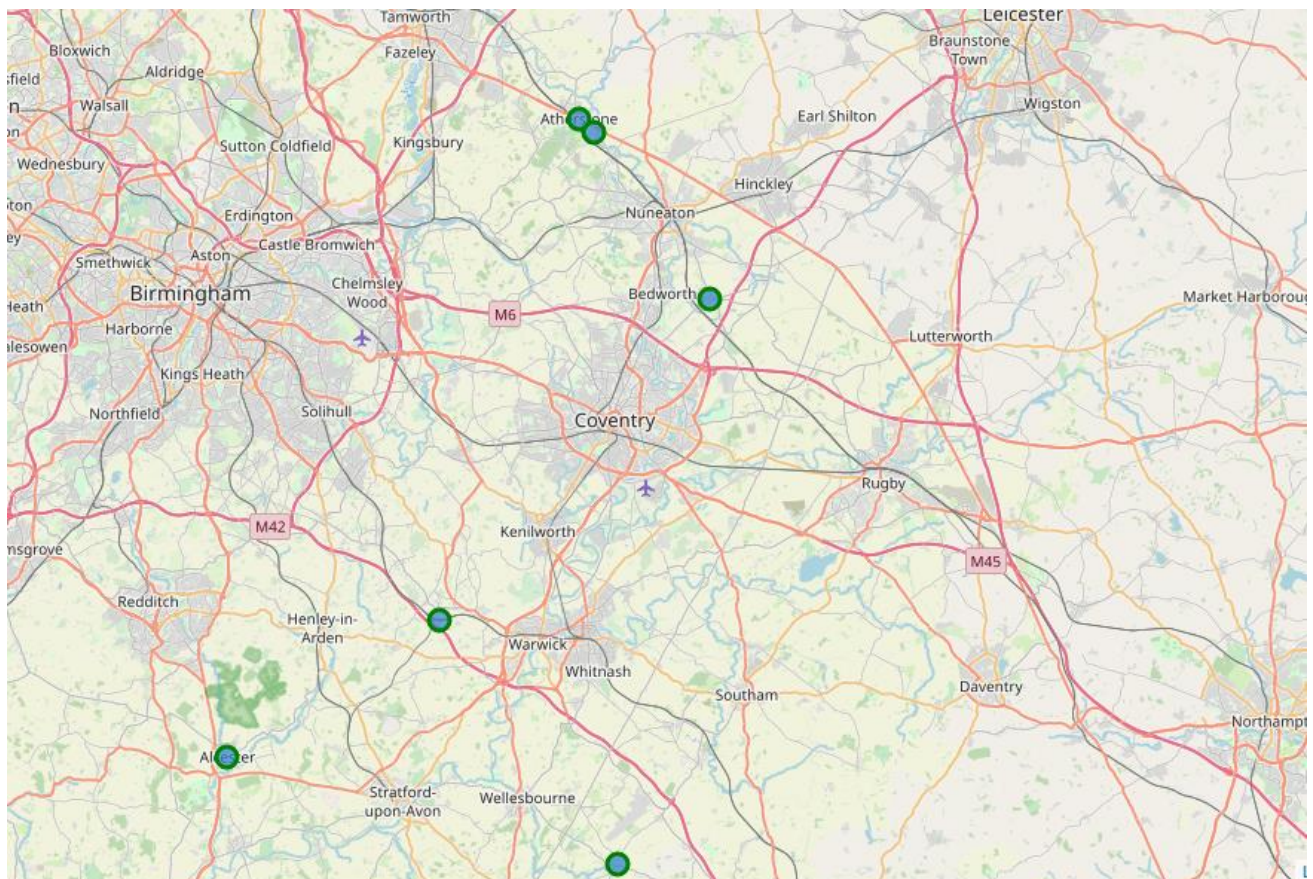


*Fig 13: Map of the 6 towns*

Population and vicinity to schools are the only factors considered for this project other than these various other data such as vicinity to malls, railway stations, supermarkets, rent for commercial buildings and many more can also be considered. Also, population data of all towns was not available. Thus, further developments can be made by considering all these factors. George R R Martin said "A mind needs books as a sword need whetstone, if it is to keep its edge." Bookstores helps the purpose of sharpening a man's mind with different books and they will always have a brighter future. Even now there are many towns that do not have any source of books like a library or a bookstore. These towns can be found using this method with a few improvements and new ventures can be started here.

# REFERENCE

- The Futures is bright for Print books:

  https://www.twosides.info/UK/print-book-sales-rise

- Printed books have a positive future:

  https://www.storaenso.com/en/newsroom/news/2018/5/printed-books-have-a-positive-future-a-new-survey-indicates