# Applied Data Science Capstone project report



Louise Cooper

17/11/2018

## 

## Introduction

* The problem I have chosen to solve is to discover the best hotels for an IBM team to recommend, for visitors to IBM Hursley. The team regularly run workshops and briefings for business partners at IBM Hursley, and due to Hursley’s location and the duration of these events, many visitors need to stay somewhere local overnight.
* The stakeholders in this exercise are the IBM Hursley team, who have many events and briefings to run, and the business partners who are invited to these events, or to other meetings and courses at IBM. The IBM Hursley team want to provide the best recommendations to their visitors, and the business partners would benefit from having a hotel recommendation that is near venues that closely match their interests.

## Data

* The data to solve the problem includes Foursquare location data, to understand the location and popularity of venues near to IBM Hursley, such as hotels, cafes, shops and bars. This will be combined with data from a sample of 100 business partners indicating what food preferences and dietary restrictions they have, and their team size, age and visit duration.

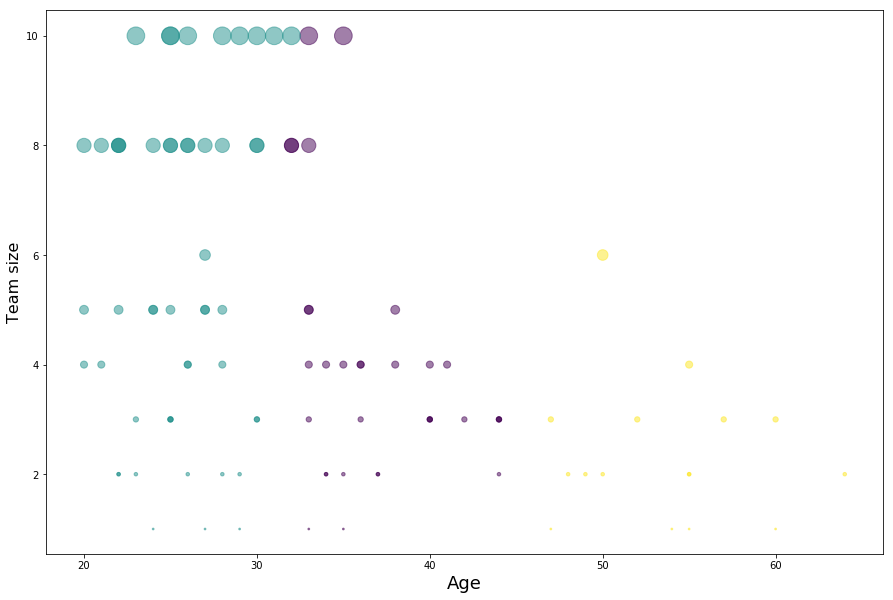


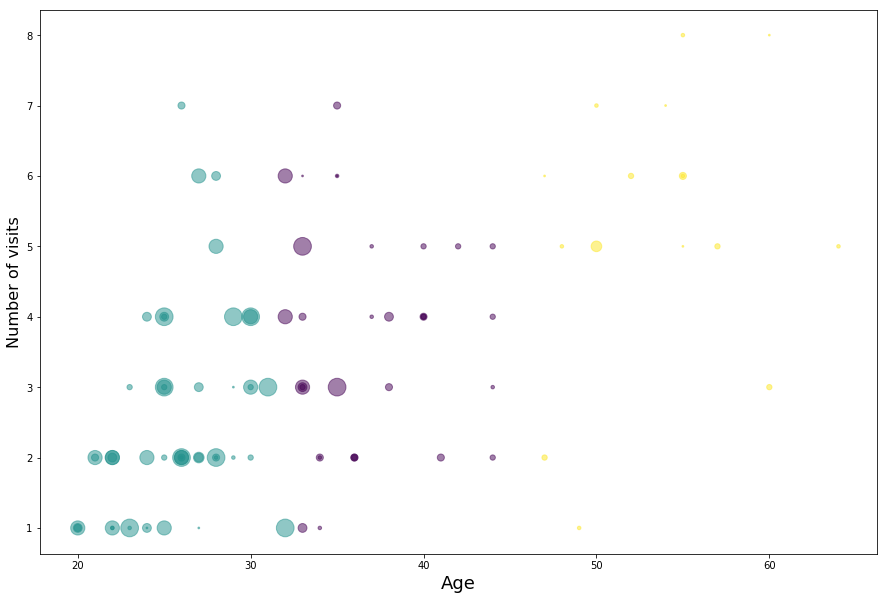
## Methodology

* k-means clustering is used to segment the business partner data into 3 segments, and identify each group’s food preferences
* These food preferences are used to provide the search criteria for queries on the Foursquare location data, to find the best area to recommend hotels
* Team size can also be used to infer that for larger teams, it would be better to find a hotel near to bars, so that they can socialise in the evenings
* The Foursquare review criteria is used to find the best hotel in the identified area, to provide the recommendation

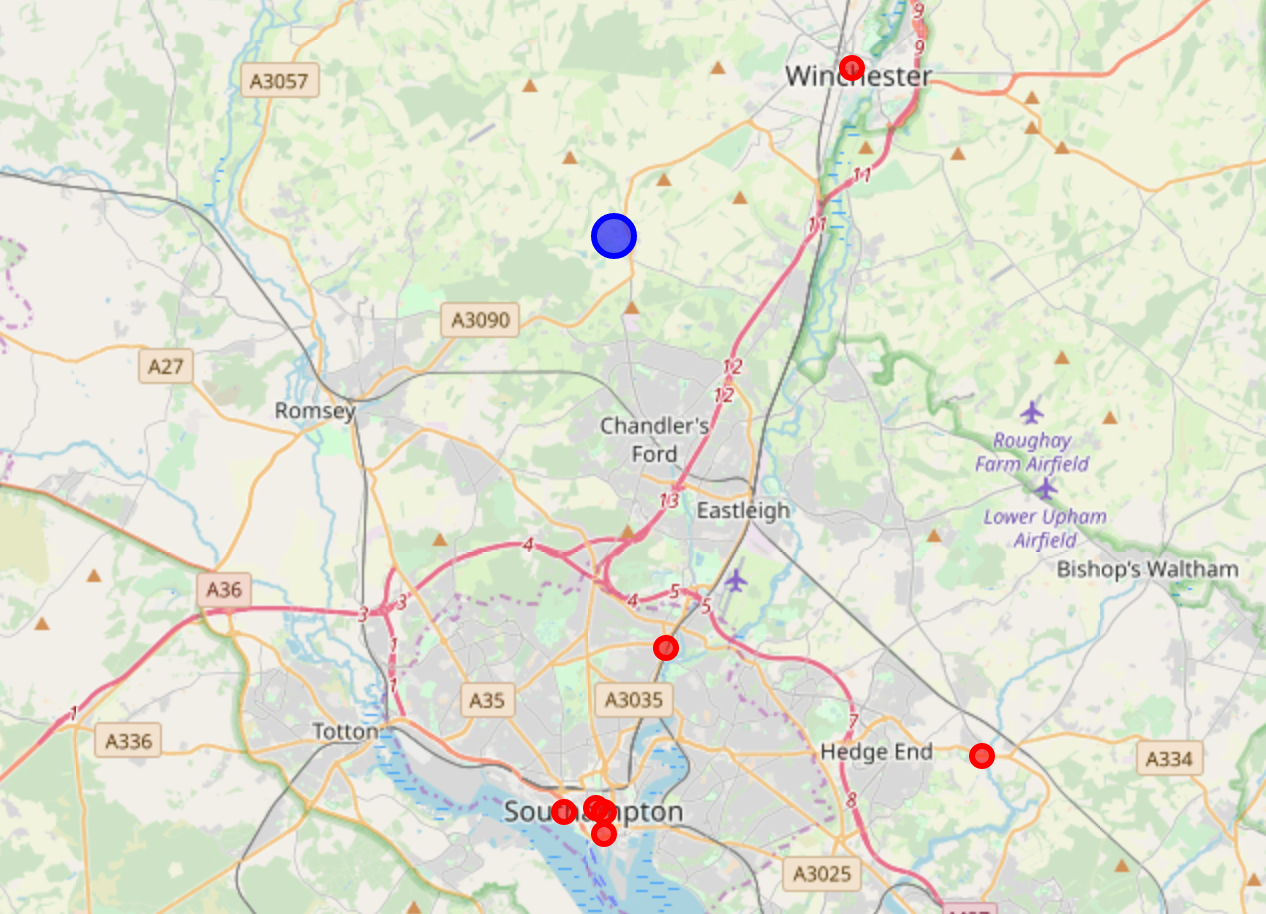
## Results

* The k-means clustering results indicate the age and size of team are significant to the segmentation. The three clusters can be characterised:
  1. older, small team, frequent visitor, prefers English, then French, then Italian food
  2. young, large team, fewer visits, prefers English, then Italian, then Indian food
  3. middle aged, medium team, medium visits, prefers English, then Italian, then Indian food

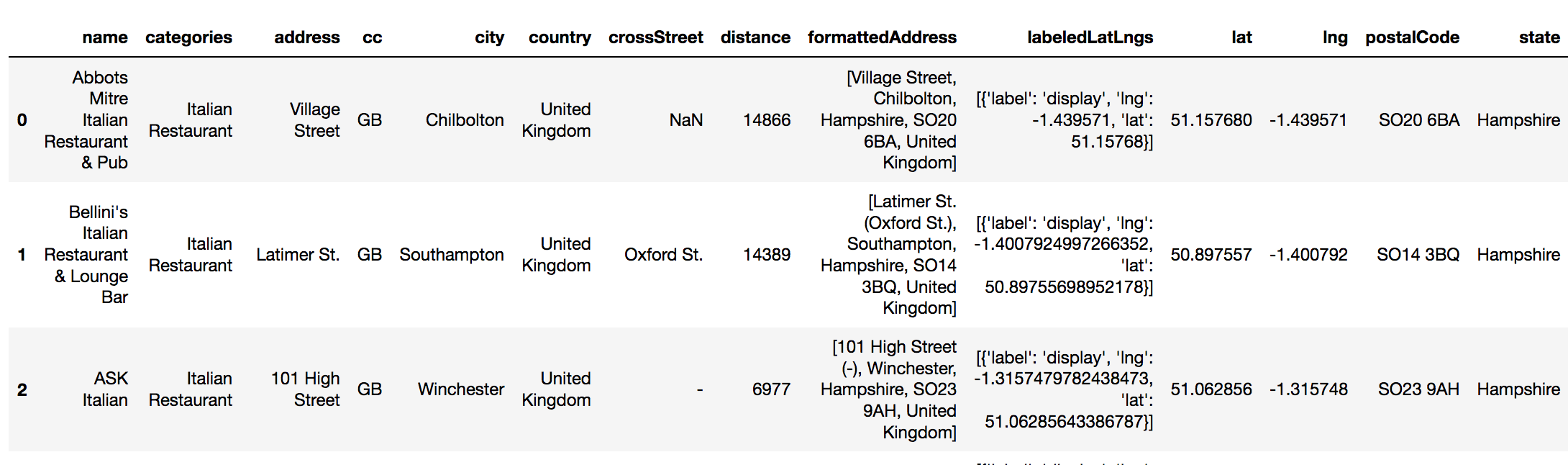




* Since two clusters prefer Italian food after English, we search the Foursquare location data for Italian restaurants near IBM Hursley
  1. older, small team, frequent visitor, prefers English, then French, then Italian food
  2. young, large team, fewer visits, prefers English, then Italian, then Indian food
  3. middle aged, medium team, medium visits, prefers English, then Italian, then Indian food



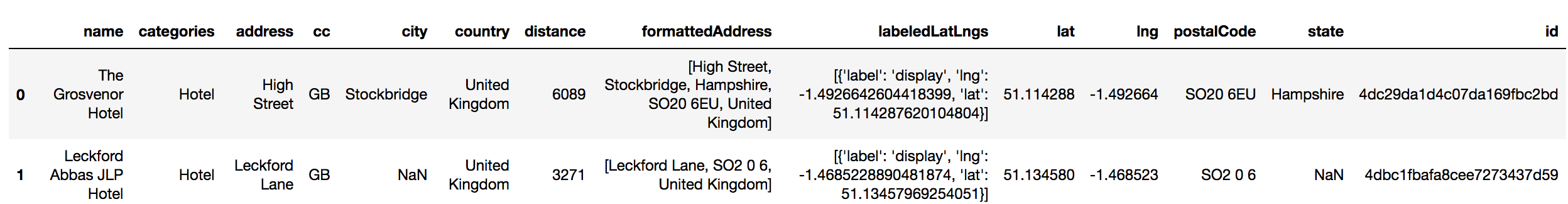
Abbots Mitre is the nearest Italian restaurant to Hursley.



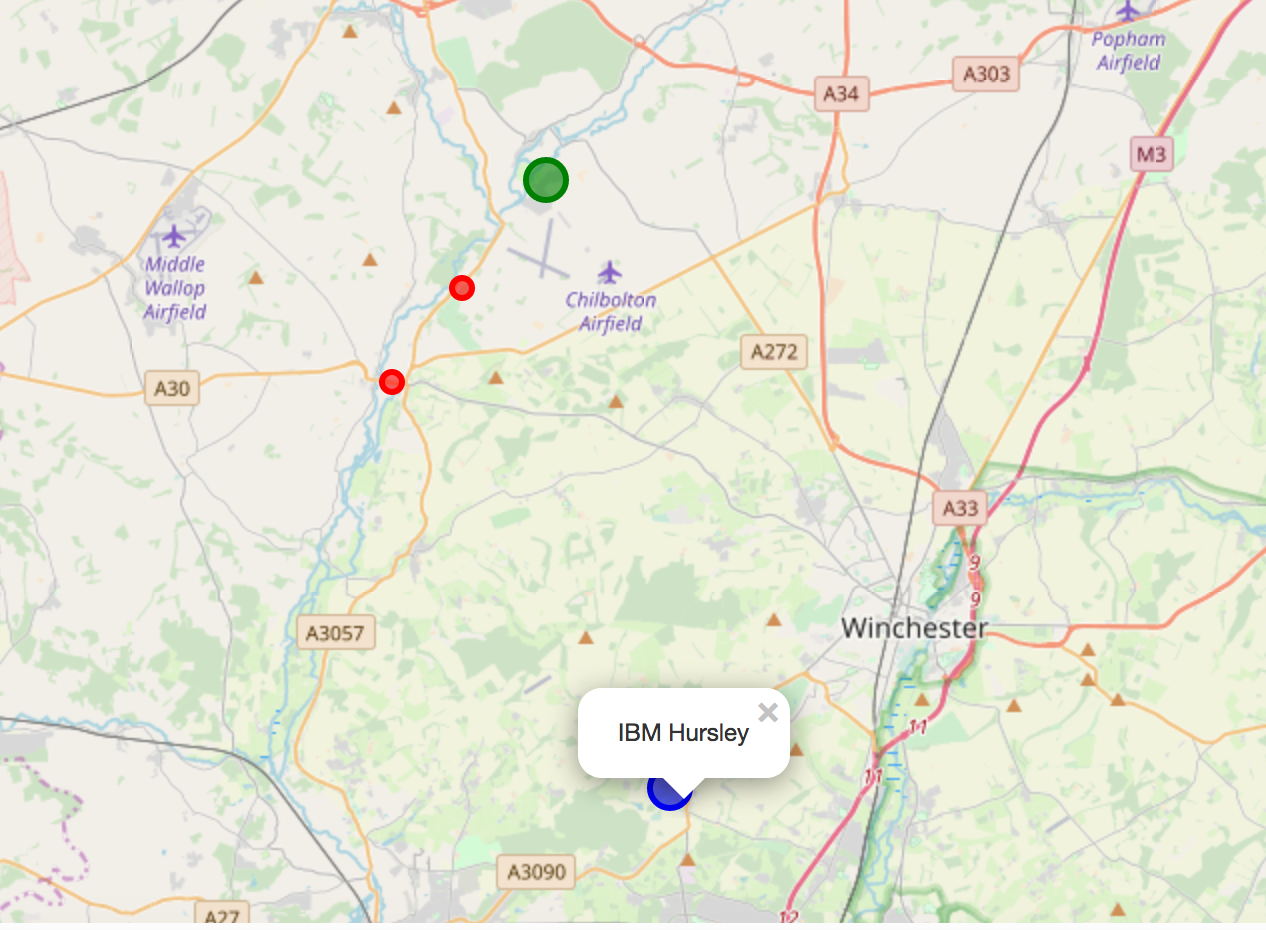
Using the Foursquare data, the rating for Abbots Mitre is 7.4.

Bellini’s is unrated, and ASK Italian is rated 6.0.

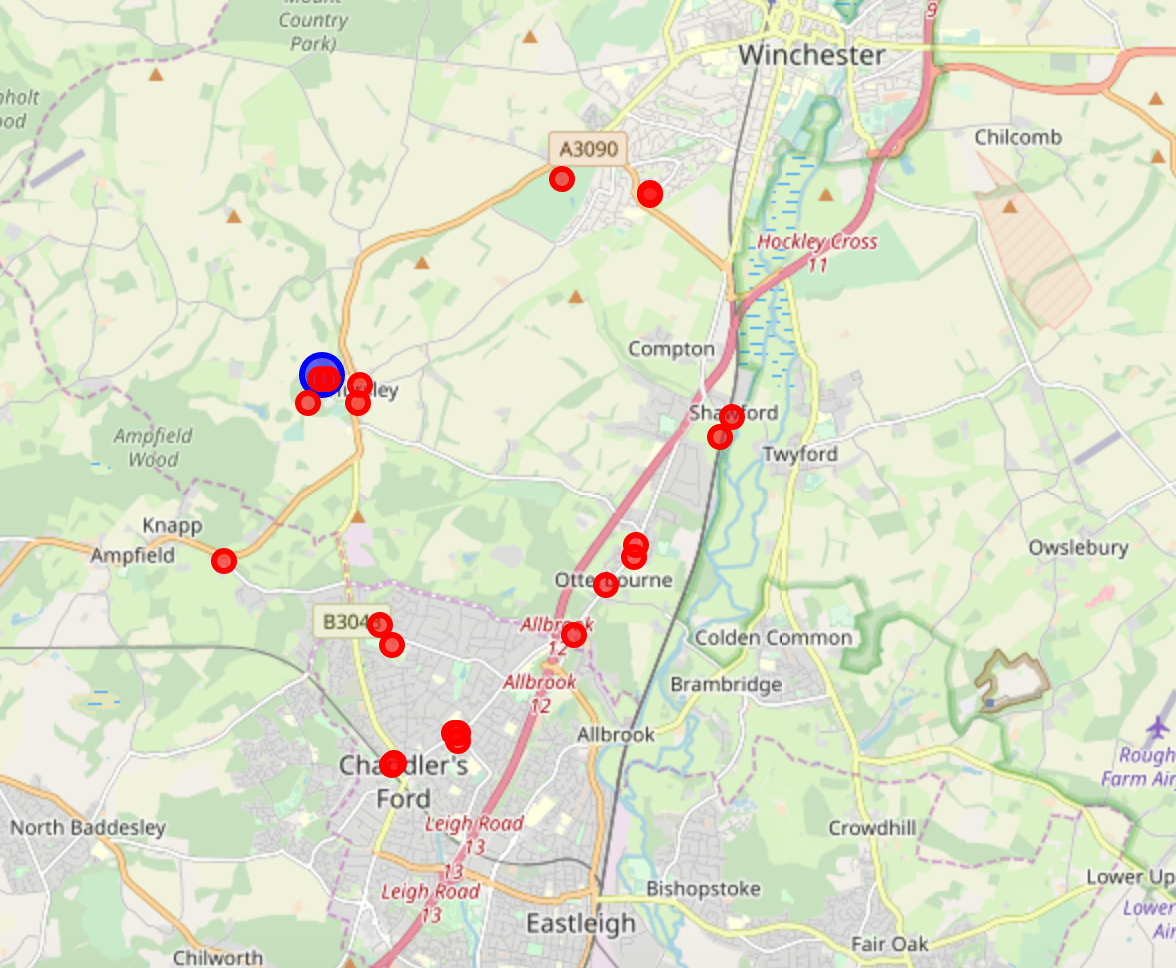
There are two hotels near Abbots Mitre restaurant.



This map shows the Abbots Mitre restaurant and the two nearby hotels, in relation to IBM Hursley:



This map shows popular venues by footfall, near to IBM Hursley:



## Discussion

* The average values for the 3 clusters show a clear difference between the clusters for age, team size and number of visits. Length of stay and vegetarian are not as differentiated between the clusters.



## Conclusion

* This investigation has shown that it is possible to segment the business partners visiting IBM Hursley, based on their age and team size, and that for the data sample used so far, this data can be used to predict the most likely food preferences.
* The Foursquare data can then be used to make a better hotel recommendation to the business partners, and also show them suggestions of places to visit in the evenings.