Data

Data Source

For the purpose of this study, location data is required that describes hotels and restaurants around Birmingham city center. Specifically, location data, number of restaurants around hotels, foot traffic in restaurants and their category. To retrieve this data, the Foursquare API was utilized in two stages, one for hotel information, and one for restaurant information.

Data Cleaning

For the hotels dataset, a “Search” call was made to the Foursquare API with the search query “Hotel”, with the aim was to identify hotels around Birmingham city center. After the call is made, results need to be organized into a Pandas data frame.

Data wrangling is performed to keep only the required data attributes related to the name and location of each venue as well as extracting the venues’ categories to verify the results. Results are then organized and viewed in a Pandas data frame, displaying some categories other than “Hotel”. These results were dropped from the data frame in order to focus on hotels only.

For restaurant data, the “Explore” call was made to the API to retrieve data on which venues were trending. Likewise, results were organized in a data frame and cleaned from any unnecessary features other than venue name, cuisine type and location data.

Feature Selection

To find which areas have groups of hotels close to one another, and the number of trending restaurants in each area, two new data frames were created one for hotels and for restaurants, each holding only the latitude and longitude coordinates of the original datasets to aid the Machine Learning algorithms explained in the next section.