Where to open a new Restaurant (and what type)

1. Introduction to problem:

   1.1. Background

   Food and restaurant are big business world over, more so in urban areas. However, there are so many choices these days. With an ever more connected world, cuisines from all corners of the world are more easily available to consumers. Also, the awareness amongst customer is increasing about sourcing, nutrition and they expect more from a restaurant visit. Ease of access, rarity of a cuisine, type of food influence customer decision to pick a restaurant for a meal alone or with friends and family.

   Nottingham is one of the largest urban centres in East Midlands region of England with a mid-sized but very bustling and active city centre. The choice of restaurants is many, however there is an ever-growing demand. With more and more global students joining in the two large universities, establishment and growth of a some of the large technology companies the city is becoming more cosmopolitan. If someone wants to open a new restaurant:

   - ***The location*** needs to be smart enough to be easily accessible, however at the same time it can't be lost in the sea of other restaurants already crowding the city centre
   - Also, ***the type*** (cuisine) of the restaurant can make or break the business. Just like the location, A popular cuisine will face tough competition while a less available one may be a flop for want of experimenting customers

   If there is some data available to indicate where to open a restaurant and what will be a preferred cuisine, it'll help a new restaurateur. This body of work will try to help an enthusiastic entrepreneur with the help of empirical data and data science.

2. Data

   2.1. Data source

   Foursquare API is the primary source of the data. However, since there is a limit of ***50 maximum values return for the call***, we must be innovative get a better result. This is especially true if we increase the search radius.

   One of the methods that deployed is to merge results from separate calls to the API. However, this needs to be used judiciously due to a limit on maximum number of API calls per day by foursquare

   2.2. Challenges

   The following are the possible challenges in the data received from a search results from API calls

   - The search result basis will be based on the collection and smartness of the Foursquare API. The address and searchability is heavily skewed for US location address type. ***For locations outside US*** It may return a lot of unusable data or data which'll need significant massaging and extrapolating
   - Initial investigation with this has shown the following issues which need to be solved in due course of this project

   2.2.1. It ***doesn't return the 'cuisine type'*** as a field in the JSON. Some of the fields such as pluralName or shortName fields have values which may be useful to extract the cuisine

2.3. Some of the ***crucial fields such as Address, and Postcode return NaN data values.*** If the count is statistically insignificant, we can drop these rows. However, if they are too many we'll need to deploy some other method to obtain the postcode and address. One of the promising methods is to use *geocoder to obtain the postcode and address via the Latitude and Longitude*

2.3.1. NaN is defined as a float, we need to identify those data values and decide whether to drop those rows, extrapolate the data or do something else. Based on the importance and feasibility of extrapolating the data the following rule was adopted if a field contained unusable value, i.e. NaN

- Cuisine: Drop the row as without the cuisine there can be no recommendations made
- Address: Keep it and check if it can be extrapolated from the Latitude, Longitude
- Latitude/Longitude: Keep it and check if these can be extrapolated from the Address and Postcode
- Postcode: Keep it and check if it can be extrapolated from Address, Latitude and Longitude

2.4. After cleaning up the data to desired degree there'll need to be two primary recommendations

- Location: If we take the city centre as ground zero, this project will try and find out the location clusters of most restaurants nearby. We can use k-means or any other suitable method to define these clusters
- Cuisine: We'll also try to forecast the most and least popular cuisines in a given radius

Hopefully armed with this scientific information, the to be restaurant owner can take an informed decision on the location and type of the new restaurant.