

Predicting the Most Suitable Place for a Restaurant in Paris

Shubham

December 17, 2020

1. Data acquisition and cleaning

1.1 Data Sources

First, all the postal codes for Paris were scraped from [here](#) (Figure 1) and major tourist attractions of Paris were scraped from [here](#) (Figure 2). Further, latitude and longitude for each postal code and tourist attraction were obtained from python's geopy library using Nominatim class. Next, Foursquare API was used for exploring each postal code.

	Postal Code	Latitude	Longitude
0	75000	48.852921	2.345716
1	75001	48.865575	2.331444
2	75002	48.867520	2.343930
3	75003	48.864286	2.359562
4	75004	48.856132	2.357131

Figure 1. List of postal codes of Paris

	Tourist Location	Latitude	Longitude
0	Place des Vosges	48.855619	2.365542
1	Moulin Rouge	48.884079	2.332408
2	Conciergerie	48.855949	2.346026
3	Pantheon	48.846191	2.346079
4	Pere Lachaise Cemetery	48.861217	2.393929

Figure 2. List of top 25 tourist attractions in Paris

1.2 Data Cleaning

All irrelevant data were removed from websites during scarping. Entries whose coordinates could not be obtained by Nominatim, were manually corrected. Foursquare explore API's JSON data was filtered for only food category results. Further, it was filtered by only selecting food outlets, containing restaurant in their name.

Original data frame, obtained from foursquare API using food category, consisted of 2028 samples and many features. Upon examining the features, it could be concluded that many features were redundant for our analysis. Thus, they were removed and most prominent 4 features were selected. Further, it was filtered for restaurants and revamped to 1465 samples and 4 features (Figure 3). Features selected for a venue were - name, latitude, longitude and category. Next, 3 additional features were added namely postal codes and corresponding latitudes and longitudes. So, final data frame used for analysis contained 1465 samples and 7 features (Figure 4).

	name	categories	lat	lng
0	Sourire Tapas Françaises	Tapas Restaurant	48.851167	2.347728
1	Le Petit Châtelet	French Restaurant	48.852637	2.346919
2	Loubnane	Lebanese Restaurant	48.851237	2.347605
3	Sola	Japanese Restaurant	48.851569	2.348391
4	Chez Le Libanais	Lebanese Restaurant	48.853285	2.341673

Figure 3. List of restaurants received from Foursquare API

	Postal Code	Area Latitude	Area Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	75000	48.852921	2.345716	Sourire Tapas Françaises	48.851167	2.347728	Tapas Restaurant
1	75000	48.852921	2.345716	Le Petit Châtelet	48.852637	2.346919	French Restaurant
2	75000	48.852921	2.345716	Loubnane	48.851237	2.347605	Lebanese Restaurant
3	75000	48.852921	2.345716	Chez Le Libanais	48.853285	2.341673	Lebanese Restaurant
4	75000	48.852921	2.345716	Sola	48.851569	2.348391	Japanese Restaurant

Figure 4. Final list containing postal code and restaurant venues