DSSS Coursework Write-up

Intro

London if famous for being an international metropolitan where people, capital, and resources from all over the world and cultural backgrounds gathered, which lead to a society with great diversity. The diversity in London is an important urban and social topic that has been measured and mapped in census report (<https://data.london.gov.uk/dataset/2011-census-diversity>). Based on census data, the report analyzed the diversity in various aspects including ethnicity, nationality, religion, socio-economical status, age structure, household composition, housing tenure, etc., providing a scope to understand the characteristics of local areas and societies. However, more perspectives and data resources can be used to study this phenomenon with other components of the urban society that could have also contributed to the diversity in London. With this being said, in this coursework, the diversity will be evaluated in terms of the available choices of cuisine types in an area.

In the article by Park (2017), the author comprehensively evaluated the ethnic food diversity of several major U.S. cities at census tract level with the Factual API and Google Places API, and analyzed the possible linkage between the food diversity and other neighborhood characteristics including population, ethnicity diversity, income, knowledge, and education. Referring to some ideas of constructing the analytical processes in Park’s paper, this study will map the cuisine diversity in London at ward level with restaurant information retrieved through Yelp Fusion API and discuss the relationship between this diversity and other possibly related attributes of a ward. To specify the definition of cuisine type discussed in this coursework, based on available information from Yelp, it refers to dishes related to a certain country, such as Italian cuisine or Chinese cuisine, and is indicated by the category label of restaurants on Yelp.

The objective of this study is to explore the pattern of food diversity in London, and provide a different scope to understand the diversity and evaluate the quality of communities. Besides, it is also expected to examine the effectiveness and validity of the open source data from Yelp in providing evidence for urban and social phenomena.

Research Q

The analysis in this study is designed to answer the following research questions:

1. How diverse are restaurants in London overall?

2. How different are wards in food diversity?

3. What is the relationship between the food diversity and other attributes of the neighborhood?

Dataset

The main data used is the businesses information retrieved from Yelp Fusion API Business Search Endpoint. With the specified search parameters, the data is returned in json format, including various attributes of businesses on their id, names, locations, categories, ratings, etc.

The ward level census data used for the relationship analysis was selected from the Census 2011 dataset available on London DataStore and reorganized to a single table. Referring to the related social aspects mentioned in the article of Park (2017), the following attributes were chosen (in the parenthesis are the column names shown on the table):

* Ethnicity diversity score (eth\_div\_score)
* Percent of population at working age 18-64 (pct\_wk\_age)
* Percent of population not born in UK (pct\_not\_born\_uk)
* Percent of household in which English is the first language of no one (pct\_esl\_household)
* Employment Rate (emp\_rate)
* Rate of new registered migrant workers (new\_migrant\_workers\_rate)
* Medium household income (med\_house\_income)
* Average GCSE score (avg\_gcse)
* Percent of population with qualification of Level 4 or higher (pct\_level4)

Besides, the boundary shapefile of London wards used to map the diversity score is retrieved from London DataStore.

All of the dataset loaded into this notebook and the json files extracted from API are also available in this repository.

Workflow

To answer the questions above, the analysis will be conducted with the general steps and Python packages specified below.

1. Retrieve information of London restaurants from Yelp Fusion API using requests 2.21.0, then convert to and save as json with json 2.6.0
2. Transform the json data to DataFrame and process it with pandas 0.23.4 and numpy 1.15.4
3. Summarize the cleaned data, and visualize it using matplotlib 3.0.2 and squarify 0.4.2
4. Join the restaurant data with boundary data, perform spatial aggregation using geopandas 0.4.1 and shapely 1.6.4
5. Calculate the diversity score with Simpson’s Diversity Index
6. Analyze the relationship between diversity score and other attributes with train\_test\_split module, LinearRegression module, and SVR module of sklearn 0.20.1

Data prep