Zongrong Li
DSCI 510

Geospatial Analysis of Socioeconomic Status and Restaurant Distribution in Los Angeles

Short Description:

This project seeks to conduct a geospatial analysis of how socioeconomic status influences the distribution of restaurants in Los Angeles County. Utilizing data from multiple web sources, I will examine the interplay between the socioeconomic, educational, and income levels of different communities within the county and the variety and distribution of restaurants in those areas. This project utilizes data science techniques such as Data Extraction and Parsing and Data Visualization, to process this abundant information, aiming to highlight how different areas within LA County experience the cultural and economic phenomenon of food service distribution, providing insights into urban planning, economic development, and social equity.

Purpose:
The purpose of this project is to explore the correlation between the socioeconomic, educational, and income conditions of different communities and the impact these factors have on the types of restaurants prevalent in these areas, highlighting how different factors contribute to the diversity or homogeneity in food service options.

Data Source and Collection Method:

To study the interplay between community socioeconomic factors and restaurant distribution in Los Angeles County, I will use the Yelp API for restaurant data, the ArcGIS API for geographic and demographic insights, and the ADI API for socioeconomic status indicators. Automated data collection and processing will be executed using Python's BeautifulSoup for web scraping, requests for API interactions, and pandas for data structuring and cleaning, streamlining the extraction of complex datasets.

Yelp API: https://docs.developer.yelp.com/docs/fusion-intro
ArcGIS API: https://developers.arcgis.com/python/
Area Deprivation Index: https://www.neighborhoodatlas.medicine.wisc.edu/

Ideas about Analysis and Visualization:

To analyze and visualize the project's data, I will use BeautifulSoup for web scraping, access web APIs for real-time data retrieval, and leverage pandas.read_html() to extract information from structured web pages and undertake geospatial analysis using GIS tools and GeoPandas, a Python library. This approach will allow us to map restaurant distributions across Los Angeles County and correlate them with socioeconomic factors such as income, education, and employment levels. The intent is to uncover spatial patterns and understand how community characteristics influence restaurant variety.

Statistical modeling, specifically regression analysis, will be employed to quantify the relationship between these socioeconomic indicators and restaurant diversity. For visualization, we will use Python's Matplotlib and Seaborn libraries to produce interactive maps and charts. These visual tools will highlight the socioeconomic diversity in Los Angeles County and its impact on the restaurant industry, offering a clear and engaging way to present our findings.

Citation:

"Python Foundation for Spatial Analysis (Full Course Material)." n.d. Courses.spatialthoughts.com.
https://courses.spatialthoughts.com/python-foundation.html