EXPLORING SOCIAL DISPARITIES ACROSS NEIGHBORHOODS IN CHICAGO

Applied Data Science Capstone by IBM/Coursera
Jun 2021 | Junjun Zheng

GROWING CONCERNS OF SOCIAL INEQUALITY

- Imbalanced economic development and contributes to gaps in quality of life across urban neighborhoods

- Quantitative assessments for such social disparities are still lagging

- Barrier to sustainable development

RESEARCH STRATEGIES

- Data science toolkits

- A novel framework for integrating diverse analytic approaches.

- The relationship among various neighborhood characteristics, including household income, spatial clusters, and land uses.

- Aiming to provide nuances to equitable community investment and place-based development initiatives.

DATA AND METHODS

- City of Chicago as the case study area.

- 56 zip codes (neighborhoods)

Data Acquisition and Cleaning

- Polygons of all Chicago zip codes

- Median household income from U.S. Census: American Community Survey (2019, 5-year estimates)

- Foursquare location data: all venues within 800 meters of each neighborhood

TEN LAND USE CATEGORIES

- Offices

- Professional Services

- Retails

- Restaurants

- Schools

- Sport and fitness

- Leisure

- Civic Services

- Transportation

- Other Land Uses

METHODOLOGY

- K-means clustering analysis

- Silhouette scores

- 13 variables

- Comparing the characteristics of identified neighborhood typologies/clusters.

DESCRIPTIVE ANALYSIS

- Gamma or normal distributions

- Some variables (i.e., retails and restaurants) may show a second peak after the main peak
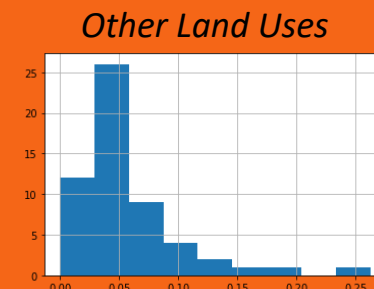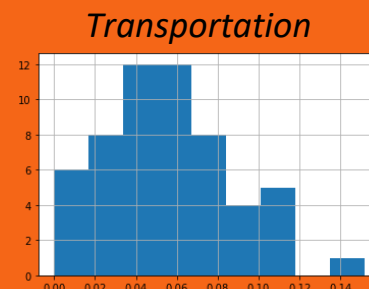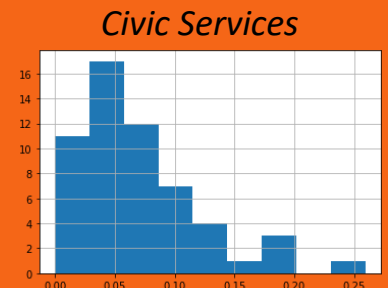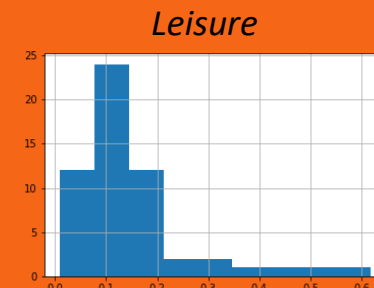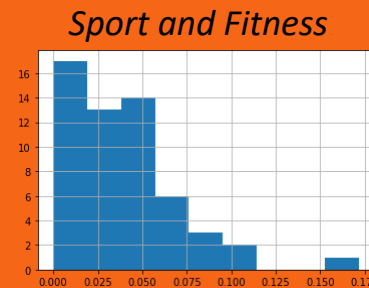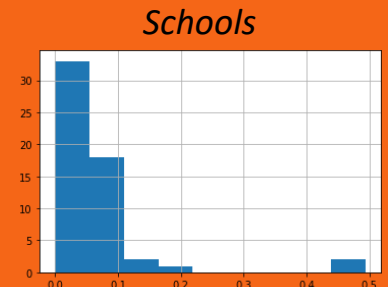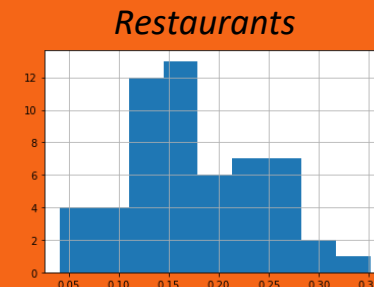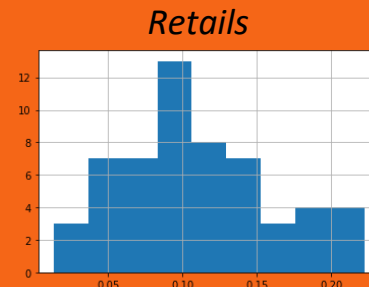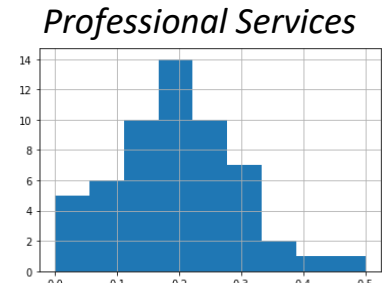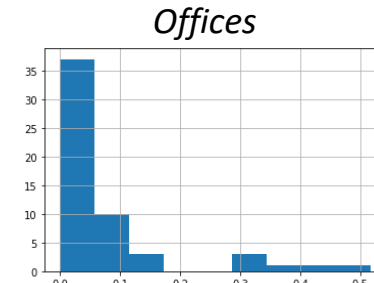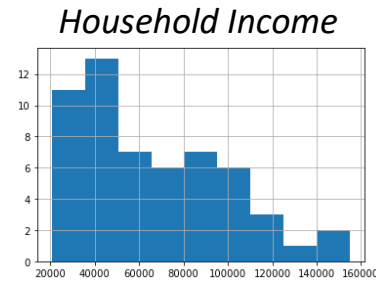
NEIGHBORHOOD CLUSTERS

- Optimal k = 5

- Average MHIs of these neighborhood clusters are statistically different

- CBD & historically affluent neighborhoods (Group 3 & 4)

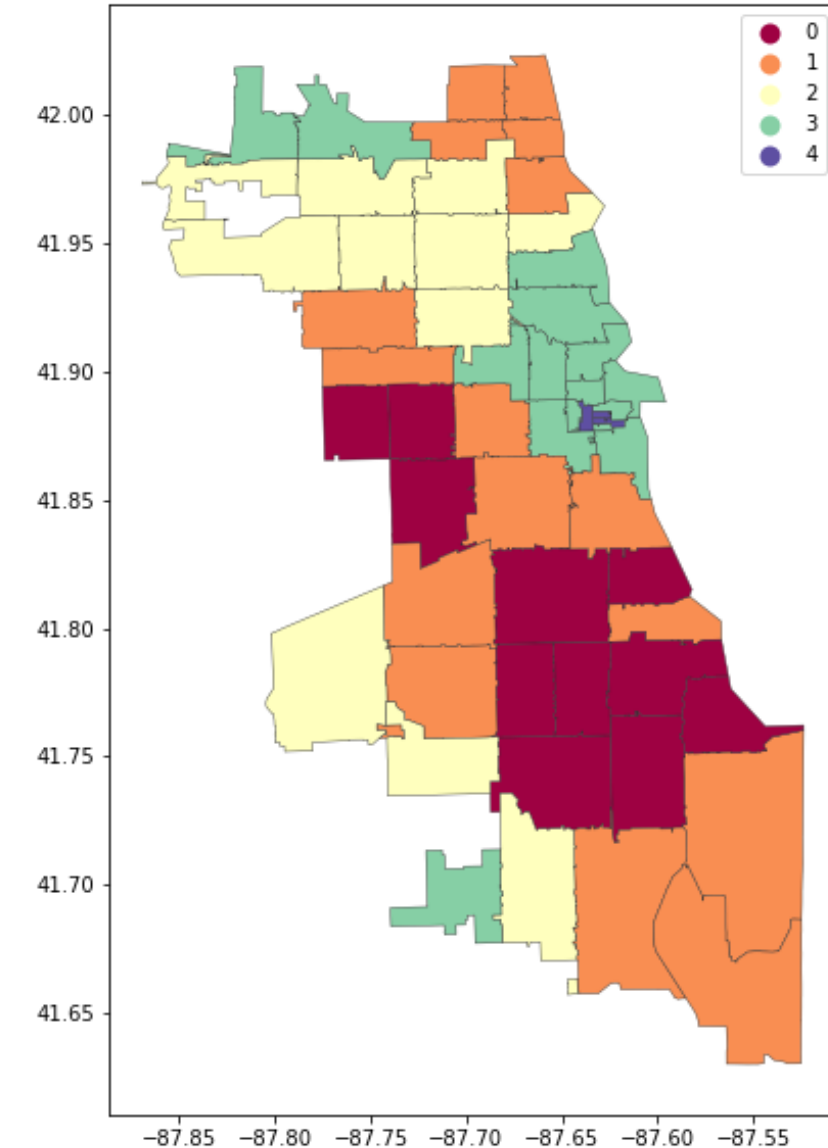- Western and southern city – the disadvantaged neighborhoods.
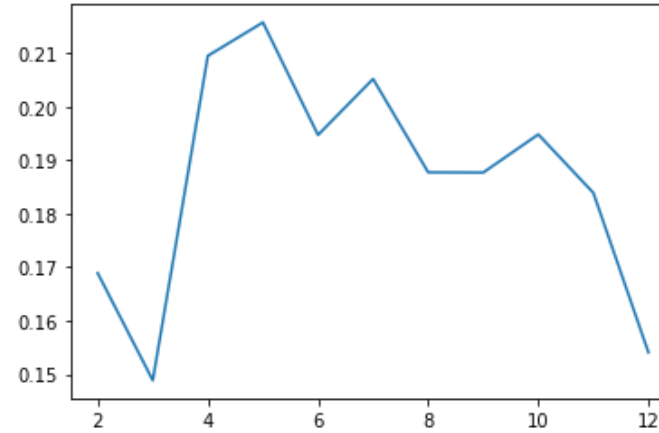
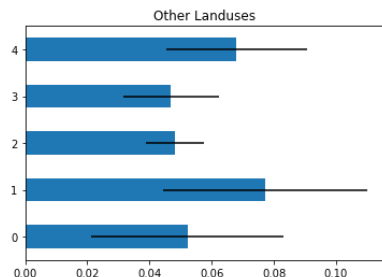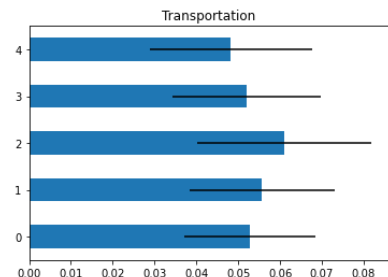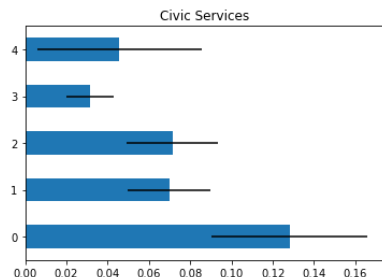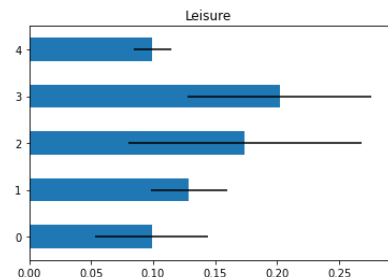(a) Silhouette Scores for the Clustering Results (*k = 2 to 12*)
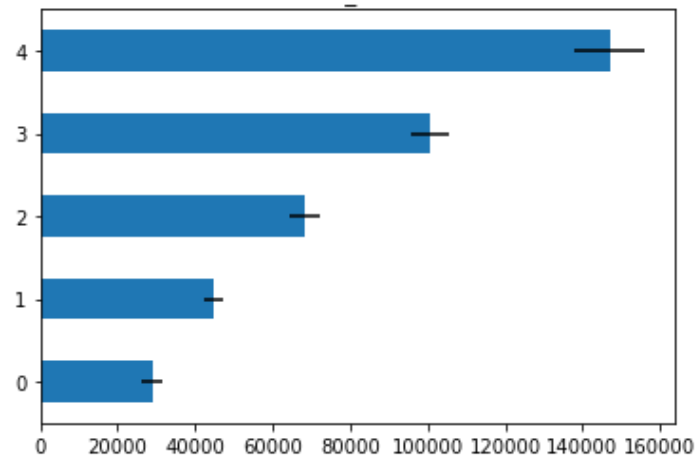
(b) Median Household Income of Neighborhood Clusters.
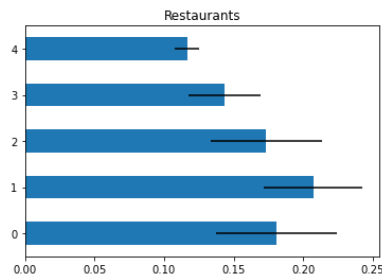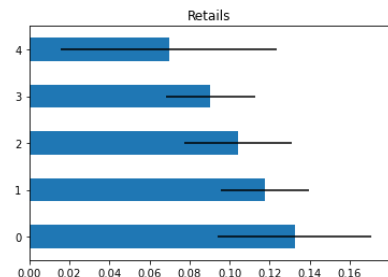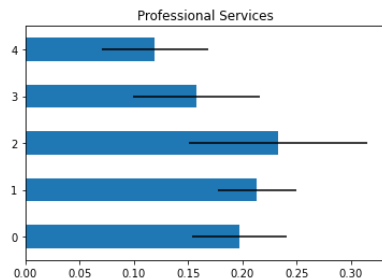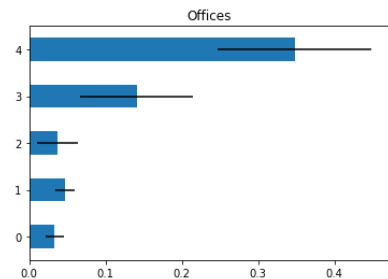*Note: error bars indicate 95% CIs (or 1.96* standard errors of the mean)*

(c) Neighborhood Clusters when *k = 5*
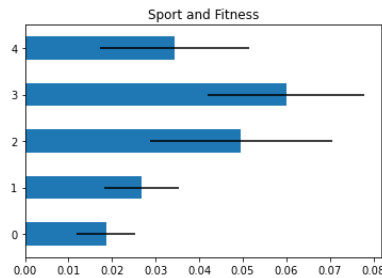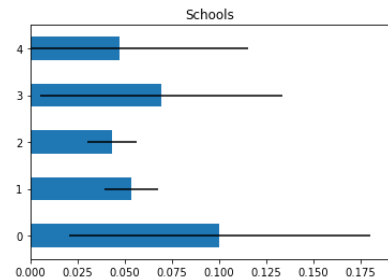
EXPLORATORY ANALYSIS

CONCLUSION AND DISCUSSION

- The income disparities are also associated with varying numbers of infrastructure and services in the neighborhood.

- The sport and fitness venues, such as basketball courts and activity centers, are notably lacking in low-income neighborhoods.

- Leisure places, such as parks, are also fewer in the historically disadvantaged area.