

Introduction:

Where are the optimal regions to invest within the historical Houston wards and Houston Heights areas?

Data section:

The data is separated into regions based on the six historical wards and Houston ("The") Heights geographical area.

Datasource:

The Wards and Houston Heights neighborhoods were obtained from the City of Houston Super Neighborhoods website and googling the latitude and longitude.

<https://www.houstontx.gov/superneighborhoods>

FourSquare locational data was used to obtain all commercial customer facing venues within each region through use of the API connection.

Methodology section:

Overview:

- The six historical wards and Houston Heights were chosen due to the 7 primarily centers around the Downtown central business district and these regions have experienced growth in urban residences, specifically medium to large scale apartment units over the last 5-7 years, including in Downtown. The Downtown region was excluded given that the vast majority of workers do not live Downtown therefore, the variety of venues and businesses would look different than the 7 regions.

Explanatory Data Steps:

1. Collected latitudinal and longitudinal coordinates for the 7 regions and convert into a dataframe.
2. Individually, using the FourSquare locational query the venues within each region was pulled and added to a dataframe followed by merging the regions into a large dataframe consisting of 599 venues.
3. The data was then analyzed to find the frequency of each venue/business based on category as defined in FourSquare.
4. This allowed the data to be classified by the Top 10 most common venues per the 7 regions.
5. Finally, the data is clustered into four (4) clusters and plotted onto a map of Houston using the Folium visualization tool.

Results section:

- The Second, Fifth and Sixth wards lack standalone Coffee Shop within the Top 10 most common venues. A more expansive dataset that includes median socioeconomic status might reveal there is a lower amount of discrepancy value within these two regions. Moreover, a cost competitive Coffee Shop, store front or a mobile Coffee Van/Truck, might have more success within these regions.
- There are no Gyms or Fitness Center in the Top 10 most common venues for the First, Second, Third and Fifth wards. Gyms, specifically Cross-Fit/pod-based fitness classes, typically consist of small spaced sites, modified commercial garages. However, the monthly membership costs tend to be a multiply higher than a YMCA or medium to large scale fitness center. If a small scale Gym were able to manage a competitive business model for the respective region there is the potential for a good investment.

Discussion section:

- The Second, Third and Fifth wards are larger and less dense. This required the radius variable in the FourSquare query to be increased to 3000 meters to return at least 90 venues per ward. The radius for the remaining wards and Houston Heights was set to 1000 meters which was sufficient to return a minimum of 90 venues each.
- A future analysis could consider all neighborhoods within the Interstate 610 loop to obtain more data within Houston proper which is denser per capita relative to the Greater Houston region.
- Furthermore, more rigorous venue data covering building size, outdoor or rooftop patio sitting and area walkability would provide deeper insights would be quite beneficial for an investment analysis.

Conclusion:

In light of Covid-19, a comparison between pre-pandemic on 12 March 2020 and 12 months later on 12 March 2021 would, I assume, have different clustering configuration. The least dense regions, Second, Third and Fifth wards, may have likely experienced more store front closures relative to the denser First, Fourth and Sixth wards along with Houston Heights. A future analysis with recent venue status would shed light on this theory.