A picture containing colorful, city

Description automatically generated

A Summer in Houston, TX:

Must-See Venues

**Introduction / Business Problem:**

I am going to analyze neighborhoods in Houston, Texas and try to find the hottest areas for socializing and finding great food. Houston is the largest city in Texas, so I will constrain the analysis to neighborhoods within 15km of Downtown Houston, TX

Interested parties for this analysis would be people looking to move into the area, or just explore Houston's finest! As one of the most diverse cities in Texas, Houston is a mecca for exploring and experiencing different cultures by their eateries.

## The Data Used, and How We Will Use it to Solve Our Problem:

Simply using google and our GitHub community, we can locate a real estate website that lists Houston neighborhoods, zips, and housing information.

Once we webscrape, we can clean the data and merge with a master USA zip code/lat/long csv located on GitHub.

After a quick mapping utilizing Folium, we can use FourSquare to gather venue data on our locations of choice; that again being neighborhoods within 15km of Downtown. This will allow for a quick Uber ride to the venue of our choice from one central location!

Here is a snip of the data we used for our initial analysis.

<https://www.houstoniamag.com/home-and-real-estate/2017/03/neighborhoods-by-the-numbers-real-estate-data-2017>

Table

Description automatically generated

## Methodology:

Utilizing a csv online with all US zip codes and coordinate data, I merged the two dataframes based on their common value of ZIP and excluded data that was not pertinent to our workflow.Once this is complete and we clean the data, I used my developer’s license in FourSquare to gather venue data for my area of interest, which is, again venues within 15 km of Downtown Houston. Here is a snip of the tail of my newly created dataframe showing the successful merge of the geospatial and neighborhood data.

Graphical user interface, text

Description automatically generated

My dataframe was initially too large for the scope that I wanted to analyze, and here is a snip of the neighborhood locations before constraining to 15 km.

Map

Description automatically generated with medium confidence

ThisText

Description automatically generated snip shows the classification based on distance from downtown. The last columns have calculated distance and a Boolean column that describes the same.

Below, we can see the new map with a much smaller dataset displayed. This will be the data we conduct our analysis on.

Chart, map

Description automatically generated

The snip below shows the neighborhoods, coordinates, and venue names and categories. I will group this dataframe by neighborhood name next.

A screen shot of a computer

Description automatically generated with medium confidence

This one below shows the grouped neighborhoods and their top 10 venues.

A screenshot of a computer

Description automatically generated

The map below represents our neighborhood clustered using the Kmeans method of machine learning. The map is interactive and highly intuitive.

Map

Description automatically generated

Here we have a final product of one of my favorite clusters and categorized by most common type.

A screenshot of a computer

Description automatically generated

**Results, Discussion and Conclusion**

Based on the results above, we can see JUST how much there is to do near Downtown Houston! Cocktail bars and health venues seem to be the majority of the popular venues.This data makes sense due to the population near downtown being young professionals, while the more tenured professionals often move the the suburbs. Outside activities and socialization are driving the scene in Houston, as people want to look good and have fun!

Depending on your preferences and lifestyle, you can make your own decision based on where you would want to live or explore. One of the reasons I retiained the median home price in the `houston\_df` dataframe is so that could help influence your decision. Mind you - it is circa 2016, which is likely quite outdated, but could still be helpful as a relational comparison tool.

I hope that you can glean some valuable insights from this exercise and feel free to drop a note with any questions!

Thanks for reading along!