**Houston - Targeted Food Advertising**

**Introduction:**

Due to the diverse nature of the Houston area the targeted advertising has a clear business value. In this project the focus is on food.

Houston is the fourth most populous city in the United States. It is diverse in many aspects. At least 145 languages are spoken by city residents. Houston's multiculturalism, fueled by large waves of immigrants. Houston's economy has a broad industrial base in energy, manufacturing, aeronautics, and transportation. Leading in healthcare sectors and building oilfield equipment, Houston has the second most Fortune 500 headquarters of any U.S. municipality within its city limits . Thus, the household income varies a lot too (https://en.wikipedia.org/wiki/Houston).

Houston has more than 80 types of food places and restaurants. and if one would like to adverse a new place in of the 80+ Houston neighborhoods it should be done wisely.

Choosing to eat at a restaurant is often a decision influenced by habit and proximity. With a wide distribution in area, ethnicity, income, etc identification of the most representative audience is not a straight forward task. However, once you have the dietary landscape in place it can be used to attract the audience and drive the business.

The purpose of this project is to explore the food preferences landscape of Houston area.

**Data:**

I am planning to use publicly available information regarding popular places in each neighborhood in combination with demographic data to break Houston into thematic groups.

The following sources of information were used:

* Houston geospatial data-Geogson file with geographical information about each SuperNeighoorhood in Houston ([https://cohgis-mycity.opendata.arcgis.com](https://cohgis-mycity.opendata.arcgis.com/))
* Census data on Houston demographics, in particular the average income per capita in each of Houston SuperNeighoorhoods ('Census\_2010\_By\_SuperNeighborhood.csv’, <https://toolbox.google.com/datasetsearch>)
* Foursquare Labs for information about popular food choices in each SuperNeighoorhoods .

**Methodology:**

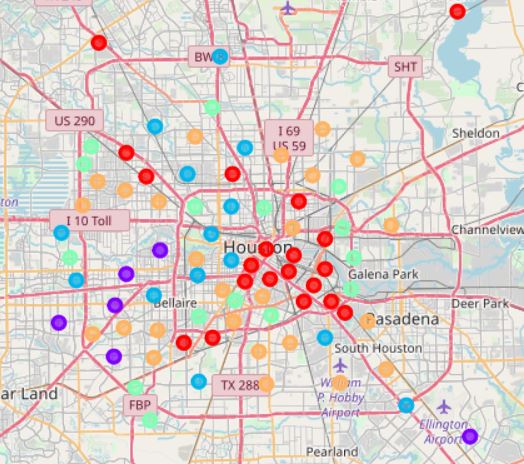
1. Identify location of each neighborhood in Houston . First I had to convert the zip file with geospatial data from official site of Houston into jeojson format. One of the online converters were used for this purpose. The file contains polygon coordinates so using geopandas the center of centroid was identified and used for further work. The information used to generate folium map of Houston colored by the area size of each neighborhood.
2. Load income data (census data set, read as the csv file). The census data set gives this information. The information used to generate folium map of Houston colored by per capita income
3. Collect information about food venues in the neighborhood (Foursquare). Foursquare allow to collect data in the certain proximity to the point of interest and also for the purpose of this project the venue ID was limited to food only (4d4b7105d754a06374d81259). The catalog is available at the <https://developer.foursquare.com/docs/resources/categories> and is very handy for the detailed exploration of the area.
4. The metadata about each neighborhood was collected using Foursquare Api and loaded into pandas data frame. Information about 1457 venues, 87 unique categories and 88 neighborhoods was collected. The top 3 venues in each neighborhoods were identified.
5. Use machine learning to cluster neighborhoods based on venues preferences index and income data. The set was divided into 5 clusters using k-means clustering approach. k-means clustering aims to [partition](https://en.wikipedia.org/wiki/Partition_of_a_set) n observations into k clusters in which each observation belongs to the [cluster](https://en.wikipedia.org/wiki/Cluster_(statistics)) with the nearest [mean](https://en.wikipedia.org/wiki/Mean), serving as a prototype of the cluster.
6. Based on the clustering results the folium map was generated and each cluster was analyzed.

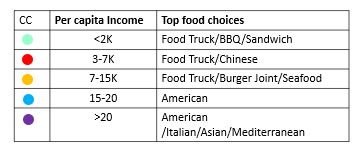
**The following visual results were obtained (summarized in the table below):**

It is interested that certain types of the restaurants were popular all over Houston, places like Pizza, Mexican Restaurant and Fast food. However, clustering gave pretty good results too. Apart from mentioned places the choices unique for the cluster were identified.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cluster# | Color Code | 1st Most common venue | 2nd Most common venue | Per capita income |
| 0 | Red | Fast Food  Mexican  Pizza | Mexican  Chinese  Food Truck | 3-7K |
| 1 | Purple | Mexican  Pizza  Italian  DimSum  American | FastFood  American  Asian  Deli  Japanese  Mediterranean | >20 |
| 2 | Blue | American  FastFood  Mexican  Pizza | Mexican  Pizza | 15-20 |
| 3 | Green | BBQ Join  Mexican | Fast-food  American  Food Truck  Sandwich Place | <2K |
| 4 | Yellow | Fast food  Mexican  Pizza  Food Truck | Pizza  Fast-food  BurgerJoint  Seafood restaurant | 7-15K |

One can expect in the areas with lower average income people tend to choose more affordable options, like Food Truck, BBQ and Sandwiches. While people with higher income tend to choose the variety of restaurants and it was hard to determine the top choice in the purple cluster. People tend to do more fine dining in variety of places.

****

****

**Conclusions:**

Overall the results of the project make a lot of sense. The segmentation worked quite well and can be used for the targeted food advertising in the Houston neighborhoods. Houston has sharp food landscape. With distinct preference in various neighborhoods. The segmented neighborhoods are highlighted on the map. Top food preferences in each neighborhood are identified and summarized in the legend. It became clear that Pizza, Mexican and fast food are liked buy people from all neighborhoods in Houston. The results of the analysis information can be used as a back bone for the targeted ads placement in the neighborhoods