

Clustering of Houston City Neighborhood

For Coursera capstone project

Jason Chen, June 2020

Introduction

According to Wikipedia ^[1] Houston is the most populous city in TX with an estimated population in 2019 of 2.3 million ^[2]. **Figure 1** shows that from 1950 to 2010 the population of Houston city has a steady annual growth rate of 20,000 people/year. People move to Houston for its plentiful jobs, education, and favorable housing price ^[3]. There are approximately 90 so called super neighborhoods (**Figure 2**) in Houston city ^[4]. The neighborhoods have diverse attributes regarding population characteristics, age of population, ethnicity, income, education status, median household income, median housing value, and so on. For anyone who wants to move to Houston or anyone who wants to open a small business (for example, a day-care center, spider-smart training center, or a restaurant with a typical food type) it would be important to know the clustering and distribution of these diverse neighborhoods, so that an optimal location can be chosen for the small business. This study is about the clustering of Houston city neighborhoods using the various demographic and other data obtained online and the Foursquare location data.

Data

On its main website ^[4], the super neighborhood's demographics, land use map, and resource assessment are summarized in PDF files accessible through hyperlinks. **Figure 3** shows an example of the first page of the PDF file that contains the feature data regarding each neighborhood, including population characteristics (total population, and persons per sq. mile), age of population (percentage of under 5 years, 5-17 years, 18-64 years, 65 and over), ethnicity (non-Hispanic Whites, non-Hispanic Blacks, Hispanics, Non-Hispanic Asians, and non-Hispanic others), income categories (under 25,000, 25,001 – 50,000, 50,001 – 100,000, over 100,001), median household income, educational status (no diploma, high school diploma, some college, bachelor's or higher), and housing and households (total housing units, occupied, vacant, total households, family households, and median housing value). In total 76 PDF files are downloaded excluding those neighborhoods that do not have such online data.

The other major data type is the Foursquare search results of the various vendors close by each neighborhood. The longitude and latitude of the neighborhood for the Foursquare search is obtained through Google search and summarized in an Excel file. Three neighborhoods are excluded in this process as they are in the airport or lake area.

Data preparation

Data preparation is performed with the following steps:

- Latitude and longitude data for each neighborhood is obtained through Google search
- 73 PDF files are parsed using PDFMINER to convert PDF into text
- The feature data of each neighborhood is obtained through string search of the above text file
- The various data from above is combined into a data frame.

The resulting data frame is shown in **Figure 4**. This data frame along with the Foursquare search results will be used to cluster the neighborhood using different approaches.

Figures

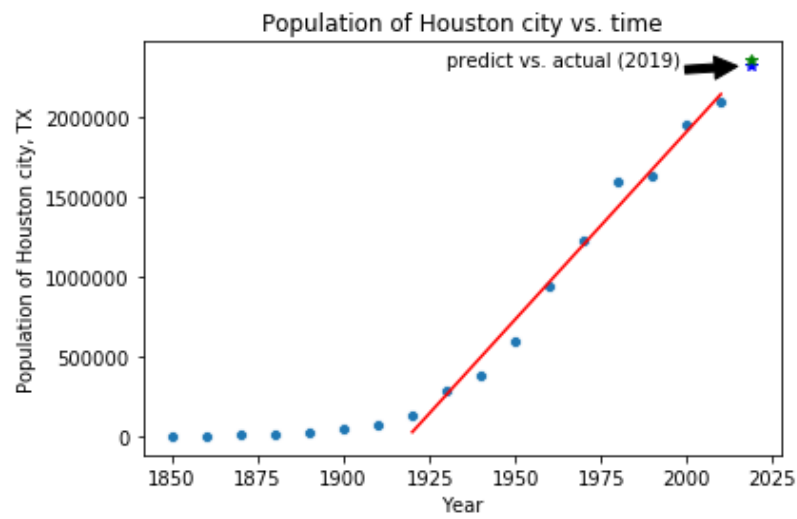


Figure 1. population of Houston from 1850 to 2010. A linear regression was performed using the historical data. The prediction of the population in 2019 matches well with the actual data.

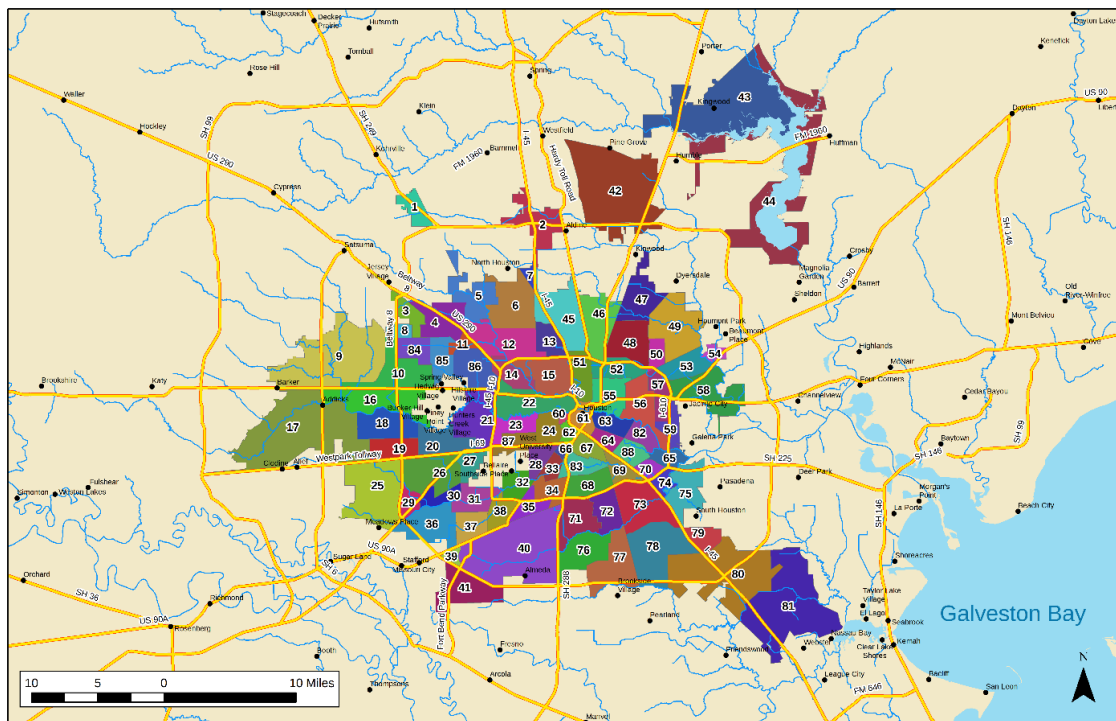


Figure 2. Neighborhoods of Houston city, TX ^[1].

SUPER

Neighborhood

RESOURCE ASSESSMENT



	Super Neighborhood		Houston	
Pop. characteristics	2000	2015	2000	2015
Total population	2,741	8,509	1,953,631	2,217,706
Persons per sq. mile	908	2,818	3,166	3,314
Age of Population				
Under 5 years	8%	9%	8%	8%
5- 17 years	13%	18%	19%	17%
18- 64 years	60%	66%	64%	65%
65 and over	19%	7%	9%	10%
Ethnicity				
Non Hispanic Whites	56%	27%	31%	26%
Non Hispanic Blacks	19%	26%	25%	22%
Hispanics	15%	27%	37%	44%
Non Hispanic Asians	8%	17%	6%	7%
Non Hispanic Others	2%	3%	1%	1%
Income				
Under \$25,000	37%	23%	33%	27%
\$25,001 to \$50,000	32%	20%	31%	25%
\$50,001 to \$100,000	26%	37%	24%	26%
Over \$100,001	5%	19%	12%	22%
Median Household Income	\$32,366	\$58,713	\$36,616	\$46,187
Educational Status				
No Diploma	13%	9%	30%	23%
High School Diploma	28%	19%	20%	23%
Some College	28%	36%	23%	24%
Bachelor's or Higher	31%	36%	27%	31%
Housing and Households				
Total housing units	1,660	3,660	782,378	909,336
Occupied	85%	91%	92%	88%
Vacant	15%	9%	8%	12%
Total households	1,416	3,320	717,945	799,714
Family households	546	1,107	457,548	491,778
Median Housing Value	\$64,976	\$153,557	\$79,300	\$131,700

Description

Willowbrook is in northwest Harris County. The area generally surrounds Willowbrook Mall and is primarily commercial, with about 63.9% undeveloped land. The area includes apartment complexes, office buildings, a major Houston Lighting and Power electric generating station and several retail shopping centers in addition to the regional mall. The area was annexed by the City in 1993.

Highlights

- ✓ Houston City Council District A
- ✓ Cypress-Fairbanks Independent School District
- ✓ 1 Police beat (includes bordering beats)
- ✓ 1,933 acres (3.02 sq. miles)



PLANNING &
DEVELOPMENT
DEPARTMENT

Figure 3. Example of PDF file of super neighborhood data sheet ^[4].

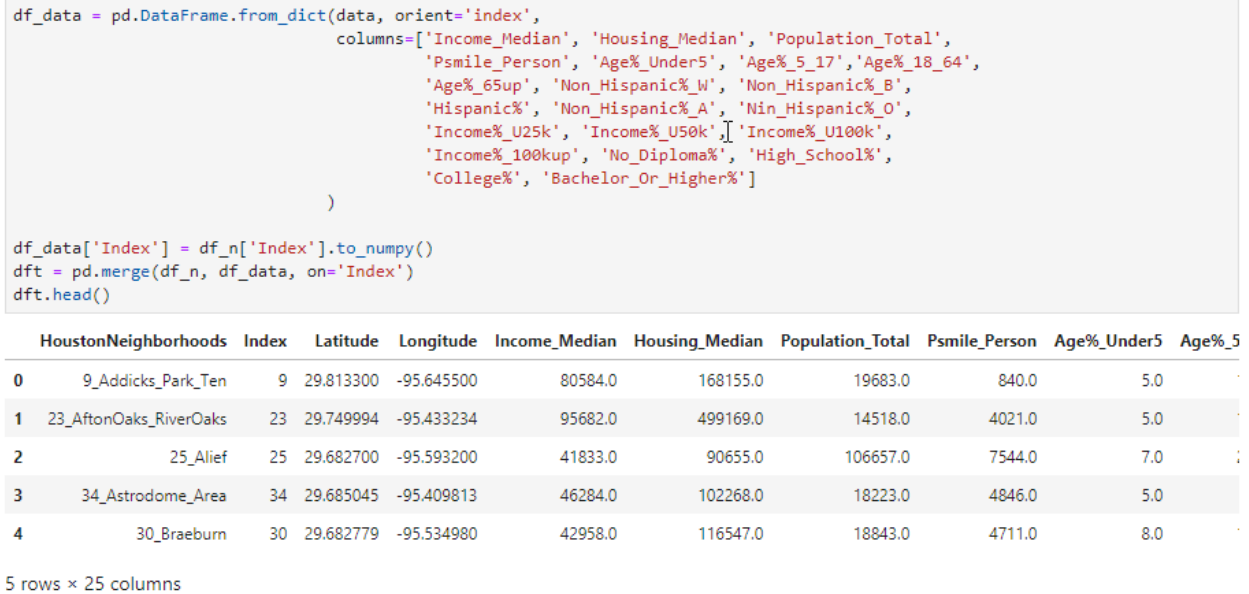


Figure 4. The head information for the combined data frame.

References

- [1]. <https://en.wikipedia.org/wiki/Houston>, accessed on June 2020.
- [2]. Annual Estimates of the Resident Population for Incorporated Places of 50,000 or More, Ranked by July 1, 2019 Population: April 1, 2010 to July 1, 2019". United States Census Bureau, Population Division. Retrieved May 21, 2020.
- [3]. <https://www.houstonchronicle.com/news/houston-texas/houston/article/Houston-growth-shows-no-sign-of-waning-4540887.php>, accessed on June 14, 2020.
- [4]. Super neighborhood, <https://www.houstontx.gov/superneighborhoods/recognized.html>, accessed on June 14, 2020.