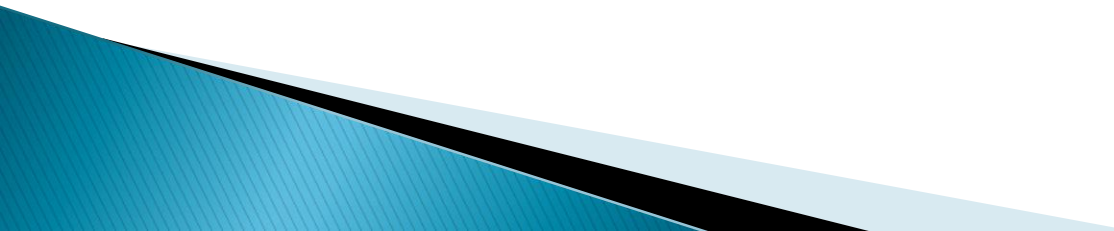


# Investigation of Houston and New York



A City Preview for  
City Migrant and  
Business Investor

# Outline

- ▶ Background
  - ▶ Data Acquisition and cleaning
  - ▶ Exploratory Data Analysis
  - ▶ Methodologies
  - ▶ Results
  - ▶ Conclusion
- 

# Background

## ▶ Motivation

- Many people in US each year move to another location.
- It is difficult to make moving decision before going there.
- Comparison of cities can help people “see” the city before they go.
- Business investor can also gain insight from this city investigation.

## ▶ Problem

- Investigate venues data to find out the two cities' difference

# Data Acquisition and cleaning

## ▶ New York Data

- Neighborhood name and latitude and longitude from this website:

[https://geo.nyu.edu/catalog/nyu\\_2451\\_34572](https://geo.nyu.edu/catalog/nyu_2451_34572)

- Venue data from Foursquare

## ▶ Houston Data

- Neighborhood name from Wikipedia page
- Neighborhood latitude and longitude from geopy package
- Venue data from Foursquare

# Data Feature Regroup

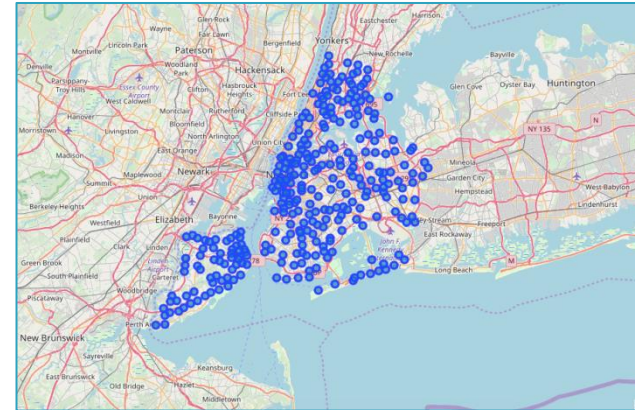
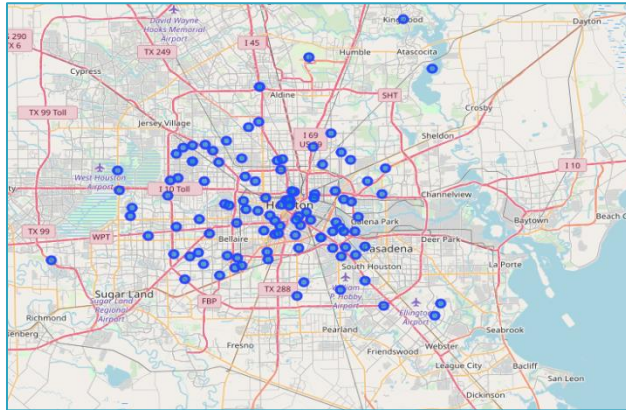
Venue Super Category	Venue Category (Foursquare API)
<b>Food and Drink</b>	afghan restaurant, african restaurant, airport food court, american restaurant, arepa restaurant, argentinian restaurant, asian restaurant, australian restaurant, austrian restaurant, bbq joint, bagel shop, bakery, bar, beach bar, beer bar
<b>Daily Essential</b>	accessories store, animal shelter, automotive shop, bank, big box store, board shop, butcher, cemetery, church, comic shop, convenience store, discount store, dog run, dry cleaner, duty-free shop, electronics store, eye doctor, fabric shop, fish market
<b>Fashion</b>	antique shop, boutique, clothing store, cosmetics shop, department store, design studio, event space, general entertainment, government building, health & beauty service, jewelry store, massage studio, men's store, nail salon, residential building
<b>Education</b>	college academic building, college rec center, high school, school, elementary school
<b>Entertainment</b>	aquarium, arcade, art gallery, art museum, arts & crafts store, bookstore, castle, circus, comedy club, concert hall, gun range, gun shop, historic site, history museum, indie movie theater, indie theater, jazz club, martial arts dojo
<b>Indoor Recreation</b>	athletics & sports, bowling alley, boxing gym, climbing gym, cycle studio, dance studio, gym, gym / fitness center, gymnastics gym, indoor play area, motorcycle shop, pilates studio, recreation center, rock club, skating rink, sporting goods shop
<b>Outdoor Related</b>	baseball field, baseball stadium, basketball court, basketball stadium, beach, bike shop, boat or ferry, botanical garden, bridge, campground, college baseball diamond, college basketball court, farm, farmers market, field
<b>Inner-City Transportation</b>	bus station, bus stop, gas station, harbor / marina, pier, stationery store, trail, train station
<b>Inter-City Transportation</b>	airport lounge, airport service



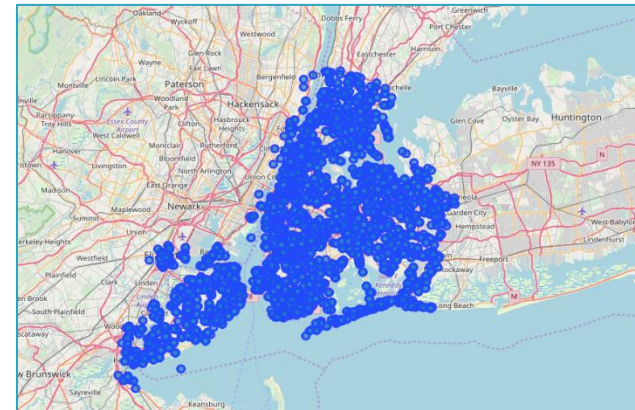
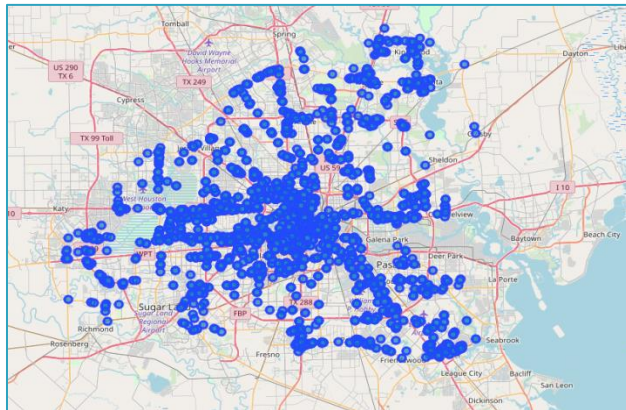
# Exploratory Data Analysis

## Houston and New York neighborhoods and Venues

Neighborhoods



Venues



Houston

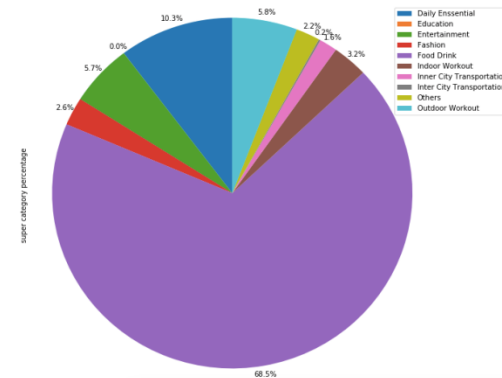
New York

# Exploratory Data Analysis

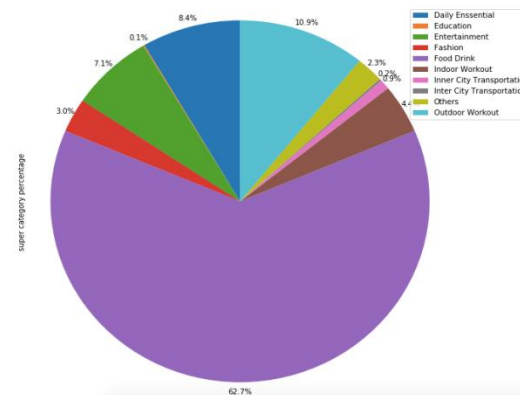
## Venue ratios in super category

Venue Super Category	Houston super category percentage(%)	New York super category percentage(%)
Daily Essential	10.28617	8.430337
Education	0.00886	0.091812
Entertainment	5.696819	7.089222
Fashion	2.587047	3.026527
Food Drink	68.47701	62.73732
Indoor Workout	3.18951	4.370922
Inner City Transportation	1.612475	0.888612
Inter City Transportation	0.159476	0.180346
Others	2.152919	2.331377

Houston Venues in each super Category

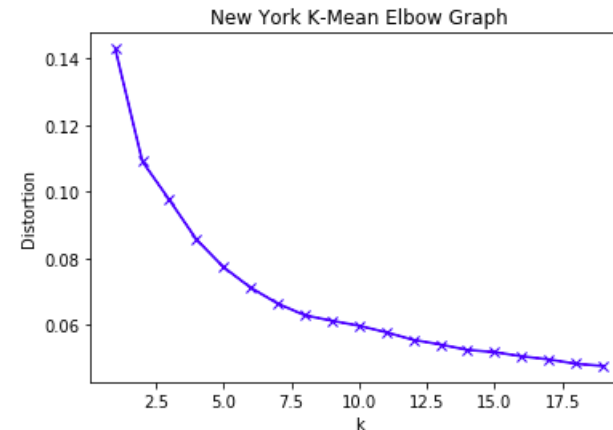
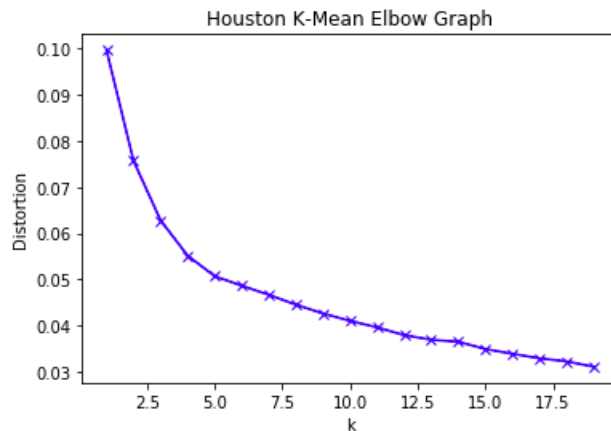


New York Venues in each super Category



# Methodologies

- ▶ I clustered each city with K-mean clustering algorithm
- ▶ Elbow plot is used to find the best k
- ▶ For easy to comparison, both city used  $k=4$

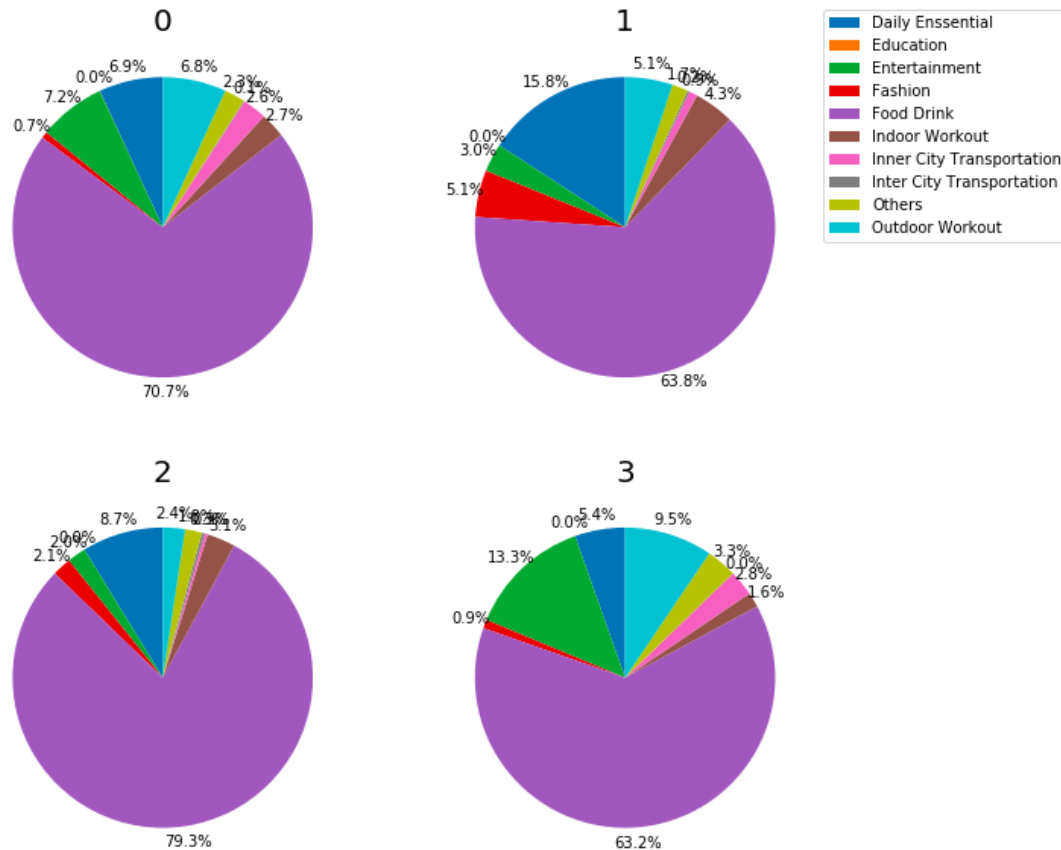




# Results

## Houston Venue Category Ration in Each Cluster

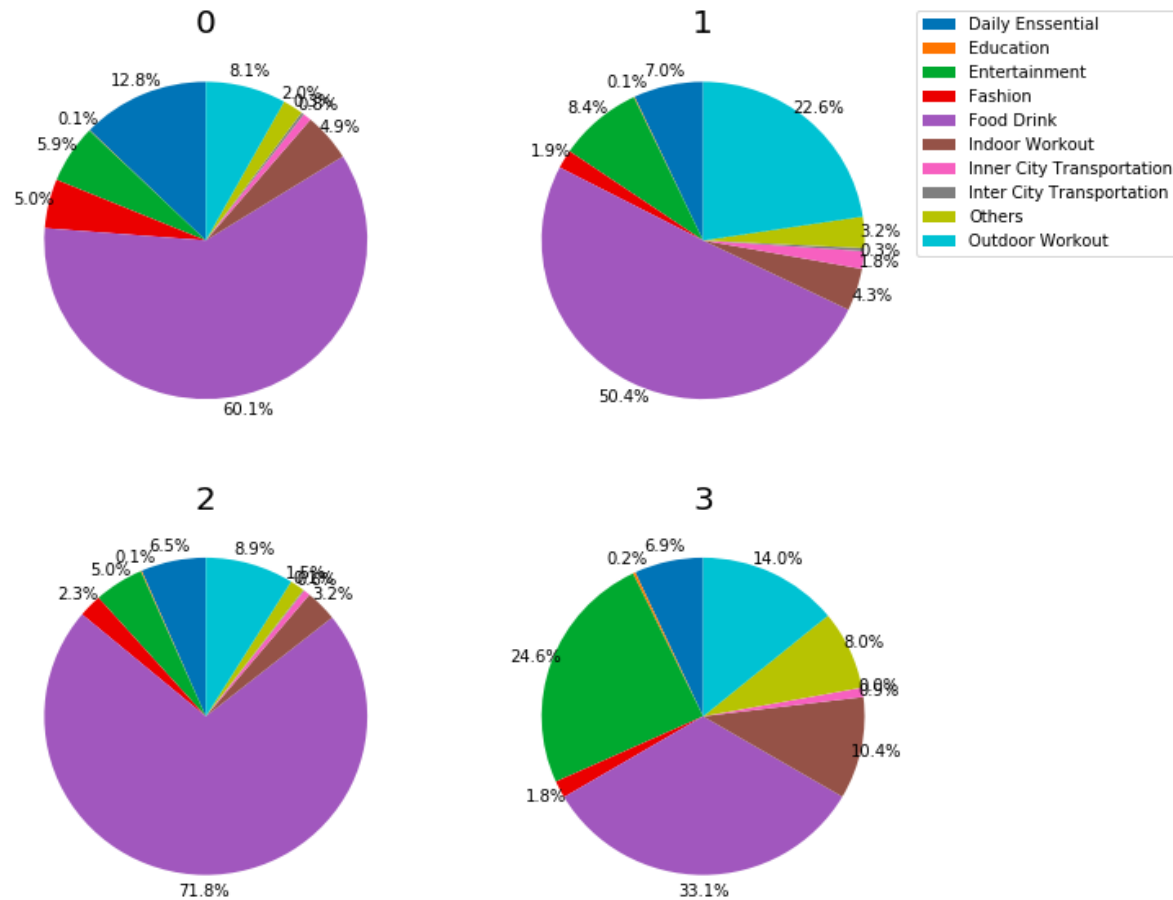
Houston Neighborhood Clusters



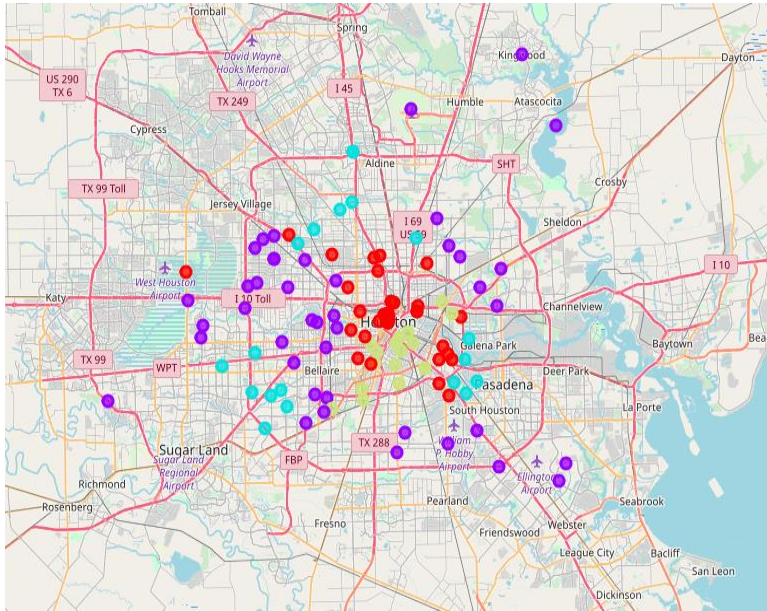
# Results

## New York Venue Category Ration in Each Cluster

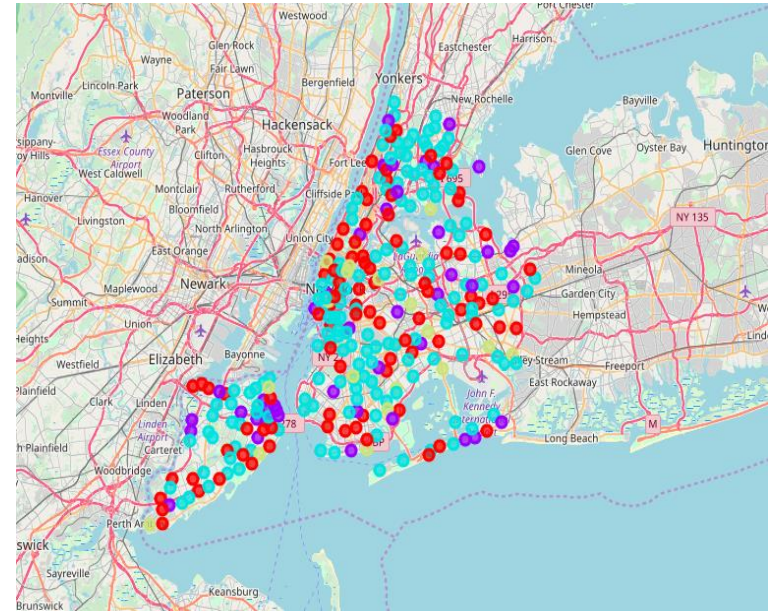
New York Neighborhood Clusters



## Map View of Neighborhood Clusters in Houston and New York

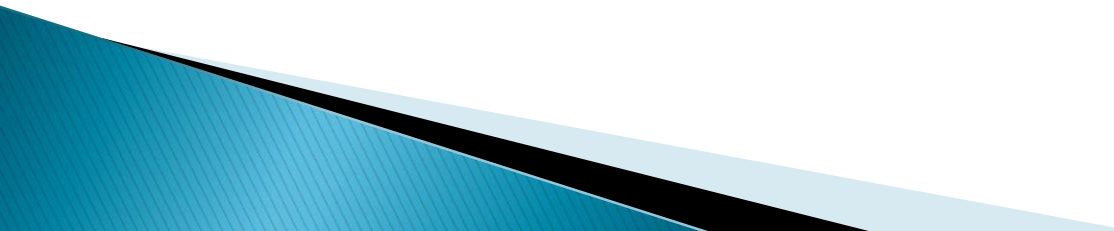


# Houston



New York

# Conclusion

- ▶ Two cities are big mature cities and similar in venue category ratio
  - ▶ The venues are more intertwined and mixed in New York than in Houston, which is correlated with higher density of population and larger population in New York
- 



# Recommendation for mover

- ▶ For people who want more convenience, New York is better choice because geographically, it has more diversity in terms of functionally of nearby neighborhood which can save some travel time, for example and live in a more dynamic environment.
  - ▶ For people who like to have more boundary of living environment and most of time stay clear from some functionality, such as entertainment, Houston is their best choice.
- 