



# EFFECT OF NEIGHBORHOOD VENUE COMPOSITION TO HOUSING PRICE

IBM CAPSTONE PROJECT

11-22-2019

# EXECUTIVE SUMMARY

- Revealed housing price is mostly affected by the composition of Lounge, Aquarium, Beach, French Restaurants, Grocery Store, South American Restaurant and hardware Store within 5kM of a neighborhood
- Demonstrated the capability to conduct data analytical work using following tools and environment
  - Jupiter Lab / Python
  - IBM Cloud / Watson Studio
  - Github
- Original code location  
[https://github.com/bjb96/Coursera\\_Capstone/blob/master/capstone-project-final.ipynb](https://github.com/bjb96/Coursera_Capstone/blob/master/capstone-project-final.ipynb)

# AGENDA

- Introduction
- Data Source
- Data Exploration
- Result
- Discussion



The background is a dark blue gradient with faint, large concentric circles. In the corners, there are white line art elements resembling circuit traces or neural network connections, with small circles at the endpoints.

# INTRODUCTION

# INTRODUCTION

- Housing price varies dramatically depending on the house's neighborhood
- Knowing the influencers of housing price is critical to buyers and investors
- Many studies already discussed impact from school zones, crime rates et.al
- This study reveals the impact from neighborhood venue composition
  - The venue composition is the percentage of venues belongs to a category (e.g. 3% venues in Boca Raton, FL is "American Restaurant" vs 6% in Boynton Beach, FL)



DATA SOURCE

# DATA SOURCE

- House Price
  - Median price by neighborhoods in US. (<https://www.zillow.com/research/data/>)
- Venue Data
  - FourSqaure API (<https://developer.foursquare.com/>)
- Geolocation Data
  - Geopy (<https://geopy.readthedocs.io/en/stable/>)



# DATA EXPLORATION



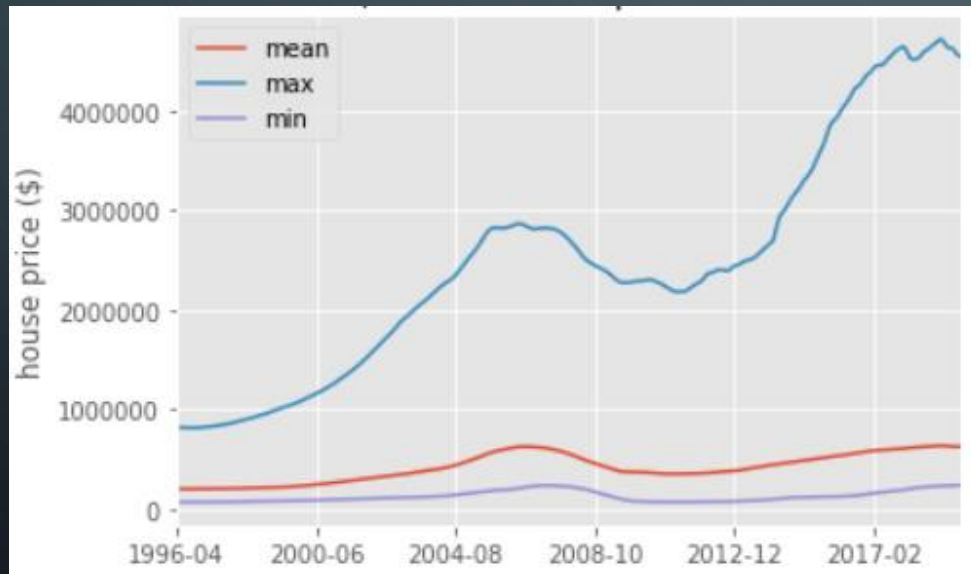
# STUDY ONLY USES A SUBSET OF THE ORIGINAL DATA FOR SOUTH FLORIDA

- House Price Dataset
  - Original dataset contains 9579 records with 287 attributes
  - This study focused on south Florida cities subset of data with 72 records

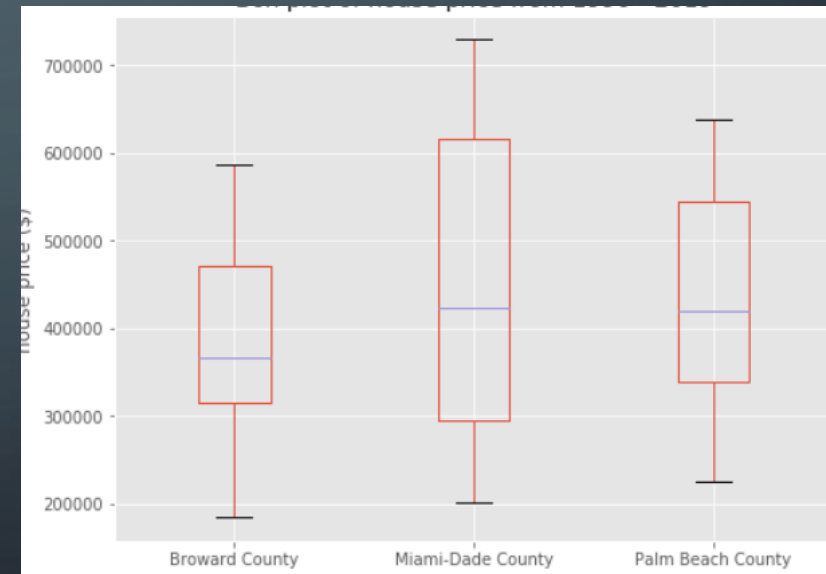
PostalCode	Borough	Neighborhood	City	State	1996-04	1996-05	1996-06	1996-07	1996-08	1996-09	1996-10	1996-11	1996-12
33010	Miami-Dade County	Miami-Fort Lauderdale-West Palm Beach	Hialeah	FL	115400.0	115800.0	116300.0	116600.0	117000.0	117400.0	117700.0	118100.0	118500.0
33015	Miami-Dade County	Miami-Fort Lauderdale-West Palm Beach	Country Club	FL	128000.0	128200.0	128300.0	128400.0	128500.0	128600.0	128800.0	128900.0	129000.0
33019	Broward County	Miami-Fort Lauderdale-West Palm Beach	Hollywood	FL	263600.0	263400.0	263000.0	262600.0	262200.0	262000.0	261900.0	262300.0	262400.0
33023	Broward County	Miami-Fort Lauderdale-West Palm Beach	Miramar	FL	92500.0	92600.0	92700.0	92800.0	92900.0	93000.0	93200.0	93400.0	93500.0

# HOUSE PRICE IS SIGNIFICANTLY AFFECTED BY TIME AND LOCATION

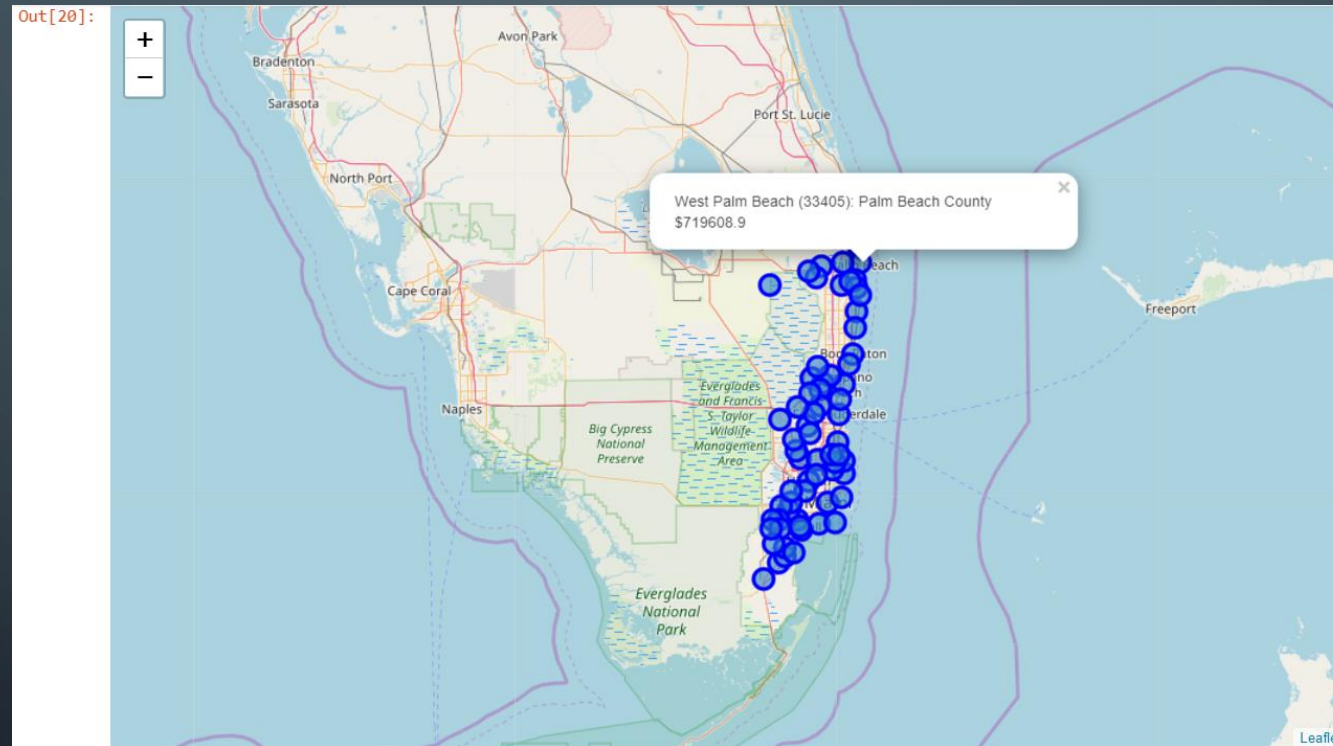
## AVERAGE HOUSE PRICE VS TIME



## HOUSE PRICE VS COUNTY



# HOUSE PRICE IS VISUALIZED BASED ON GEOLOCATION (LATITUDE, LONGITUDE)



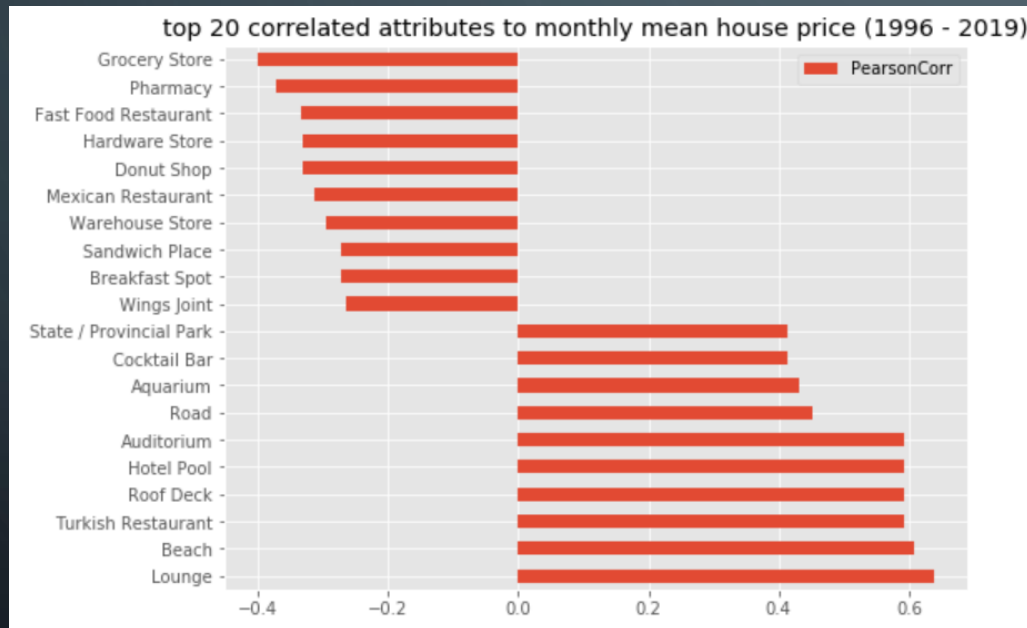
# 100 VENUES WITHIN 5 KM RANGE FROM THE GEOLOCATION IS RETRIEVED

Address	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Category
Hialeah,FL	25.857596	-80.278106	Publix	Grocery Store
Hialeah,FL	25.857596	-80.278106	Canton Village	Chinese Restaurant
Hialeah,FL	25.857596	-80.278106	Molina's Ranch Restaurant	Café
Hialeah,FL	25.857596	-80.278106	Chipotle Mexican Grill	Mexican Restaurant
Hialeah,FL	25.857596	-80.278106	Yoyito	Latin American Restaurant

# VENUE COMPOSITION IS VERY DIFFERENT AMONG CITIES

Neighbor	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Aventura,FL	Clothing Store	Italian Restaurant	Furniture / Home Store	Department Store	Cosmetics Shop	Shoe Store	American Restaurant	Juice Bar	Grocery Store	Hotel
Bal Harbour,FL	Beach	Italian Restaurant	Hotel	Coffee Shop	Grocery Store	Park	Resort	Peruvian Restaurant	Boutique	Dog Run
Boca Raton,FL	Bar	Grocery Store	Italian Restaurant	Pizza Place	Coffee Shop	American Restaurant	Beach	Sushi Restaurant	Resort	Steakhouse
Boynton Beach,FL	American Restaurant	Beach	Brewery	Bakery	Park	Seafood Restaurant	Italian Restaurant	Deli / Bodega	Sushi Restaurant	Mexican Restaurant
Century Village,FL	American Restaurant	Sandwich Place	Grocery Store	Mexican Restaurant	Clothing Store	Sports Bar	Indian Restaurant	Hardware Store	Asian Restaurant	Diner
Coconut Creek,FL	Park	Grocery Store	Coffee Shop	Latin American Restaurant	Pharmacy	Sandwich Place	Big Box Store	Bar	Donut Shop	Fast Food Restaurant
Cooper City,FL	Grocery Store	Italian Restaurant	Café	Pizza Place	BBQ Joint	Park	Donut Shop	Bar	Pharmacy	Coffee Shop

# TOP 20 CORRELATED VENUES TO HOUSE PRICE

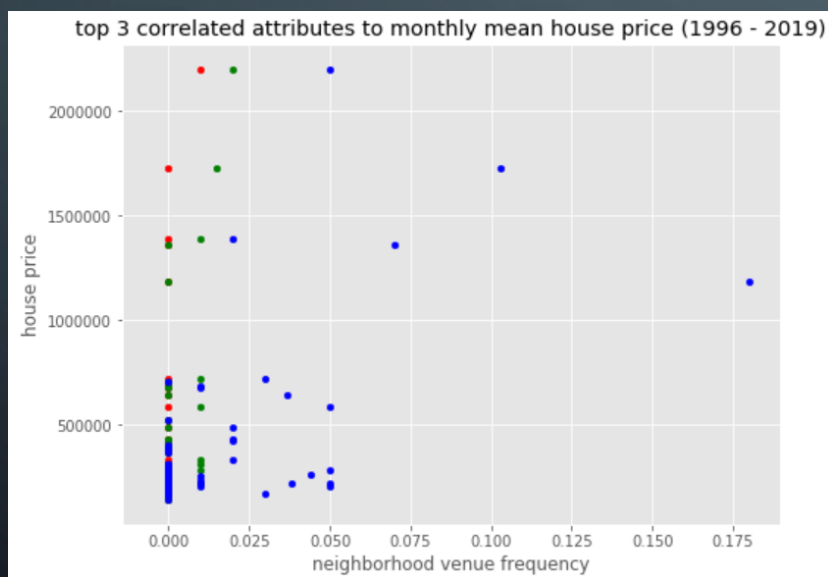


- Pearson correlation coefficient is calculated
- Positive correlation means house price is higher when the venue category composition is also high
- Negative correlation means house price is lower when the venue category composition is high

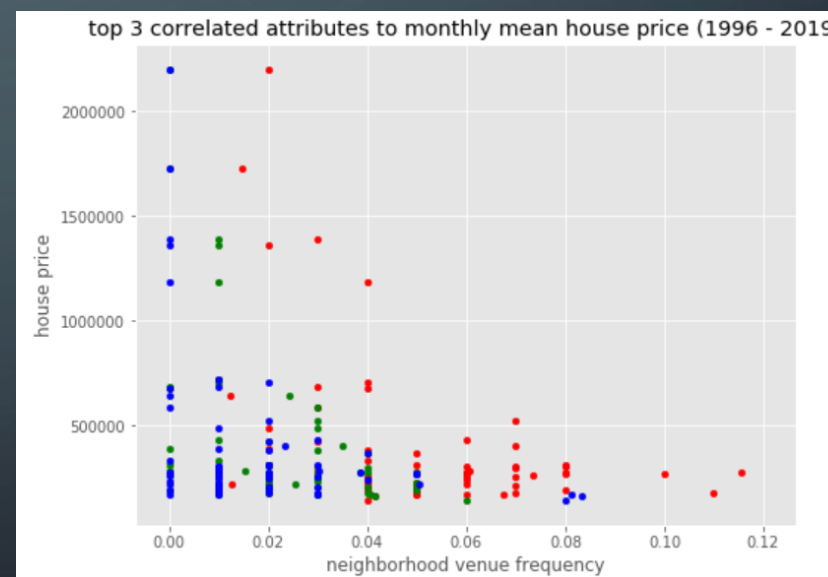


# HOUSE PRICE VS VENUE COMPOSITION

## POSITIVE CORRELATED



## NEGATIVE CORRELATED





RESULT



# FEATURE SELECTION

- Multiple feature selection methods are applied to extract most important features to predict housing price
  - Correlation
  - Mutual information
  - F test
- Ensemble approach is used to generate final selected feature list

# 11 FEATURES THAT HAVE BEEN SELECTED BY AT LEAST 2 OF 3 METHODS ARE USED IN REGRESSION

## CORRELATION

```
Out[63]: ['Aquarium',  
         'Art Gallery',  
         'Auditorium',  
         'Beach',  
         'Beer Store',  
         'Boutique',  
         'Breakfast Spot',  
         'Caribbean Restaurant',  
         'Chocolate Shop',  
         'College Academic Building',  
         'Convenience Store',  
         'Cupcake Shop',  
         'Discount Store',  
         'Food Court',  
         'French Restaurant',  
         'Gastropub',  
         'Grocery Store',  
         'Hardware Store',  
         'History Museum',
```

## MUTUAL INFORMATION

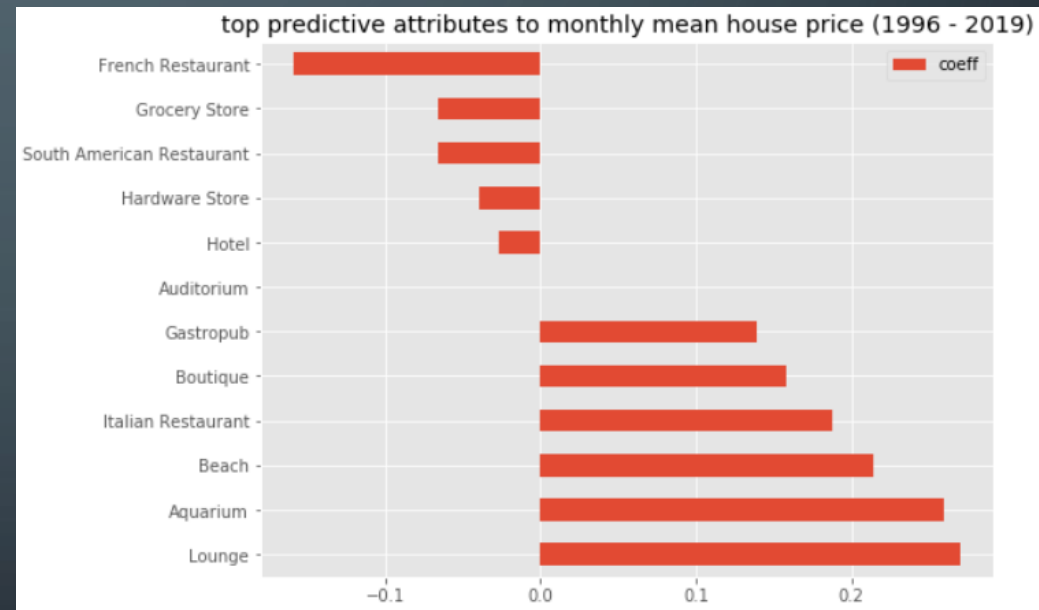
```
Out[60]: ['Donut Shop',  
         'Beach',  
         'Hardware Store',  
         'Greek Restaurant',  
         'Grocery Store',  
         'Pet Store',  
         'Performing Arts Venue',  
         'South American Restaurant',  
         'Southern / Soul Food Restaurant',  
         'Golf Course',  
         'Italian Restaurant',  
         'Tapas Restaurant',  
         'Lounge',  
         'Cosmetics Shop',  
         'Hotel',  
         'Fried Chicken Joint',  
         'Seafood Restaurant',  
         'Gas Station',  
         'Asian Restaurant',  
         'Movie Theater']
```

## F TEST

```
Out[56]: ['Lounge',  
         'Beach',  
         'Turkish Restaurant',  
         'Auditorium',  
         'Roof Deck',  
         'Hotel Pool',  
         'Road',  
         'Aquarium',  
         'Cocktail Bar',  
         'State / Provincial Park',  
         'Nudist Beach',  
         'French Restaurant',  
         'Hotel',  
         'Grocery Store',  
         'Italian Restaurant',  
         'Boutique',  
         'Boat or Ferry',  
         'Moving Target',  
         'Bed & Breakfast',  
         'Snack Place']
```

# LINEAR REGRESSION MODEL IS USED TO DETERMINE THE PREDICTIVE POWER OF FEATURES

- Linear regression model based on the 11 features selected shows
  - $R^2 = 48\%$
  - $MSE = 0.02$
- Coefficients of features are observed



The background is a dark blue gradient with faint, large concentric circles. In the corners, there are white line art elements resembling circuit boards or neural networks, with lines and small circles connecting them.

# DISCUSSION

# REGRESSION MODEL RESULT SHOWS THE MOST INFLUENTIAL VENUE COMPOSITION FOR HOUSING PRICE

- Lounge, Aquarium and Beach have the most positive predictive power to housing price. This means the higher is their compositions in a neighborhood, the higher is the average housing price, and housing price is most sensitive to their composition variations
- French Restaurants, Grocery Store, South American Restaurant and hardware store have the most negative predictive power. This means the higher is their compositions in a neighborhood, the lower is the average housing price, and housing price is most sensitive to their composition variations
- Lower  $R^2$  means the venue composition is not the only housing price influential factor. As mentioned earlier, other factors such as school zones and crime rates also affect the housing price. A dataset incl. all the factors is expected to yield more accurate housing price prediction model