

Peer-graded Assignment: Capstone Project - The Battle of Neighborhoods (Week 1)[1](#)

NEW HOME

1. Introduction

1. 1 Background Information

Almost everybody is moving to another place - at least once in a lifetime. Starting to study in a new city, during a semester abroad, after changing the job, or moving together with a new partner are only some of the reasons why people would find a new place to live. Besides the price of the housing, another important factor is its surrounding. When moving to another place, people most likely would like to find a place that is comparable to the current home to feel comfortable. Even though different people put emphasis on different venues: For example, for some it might be interesting to have a Gym nearby, while other prefer good restaurants or coffee stores nearby - they all have the same problem in common: How to find a place that meets the personal requirements? In this report we will answer this question using Data Science Methods.

1. 2 Problem statement: Where should John move within Miami

In this report, we are trying to find a borough in Miami for John, where he most likely would feel comfortable. John is 27 years old and is currently living in The Asia Brickell Key. He really enjoys living there. Unfortunately, he lost his wealth - that is why John needs to move to another place. He would like to use this as a chance to explore new neighborhoods in Miami. Brickell Key is an artificial triangular island at the mouth of the Miami River in Miami, Florida, United States. John like to see modern Art and grab coffee afterwards. That is why it would be nice to have a Art Gallery and coffee store close to the new home ,john likes to keep his body fit, When it is getting too late at work, he likes to grab dinner in a restaurant. That is why it would be a plus to have bars or restaurants close to his home. Summarized, John would most likely enjoy to have the following venues nearby:

- Art Gallery
- restaurants & bars
- Gym
- coffee store

All in all, John really loves his current borough, that is why we need to find out in addition, which of the other boroughs is most similar to the one he is currently living in. We will use data science methods to identify the most promising neighborhoods based on these criteria.

1. 3 Target audience for this report

This report is an analysis of the boroughs in Munich customized to the needs of John (park, coffee store... nearby). However, the approach is applicable to all people with any specific needs. The information gathered from Foursquare in combination with data science methods are a good basis to derive data driven decisions regarding boroughs that best fit the specific needs at hand. It would even be possible for real estate agents to use some similar approaches to find the perfect home for their customers.

2. Data

2.1 Description of the Data

In order to find the most promising borough for John the following data is needed:

1. Average price per m² of the apartments in Miami :This information is gathered through web scraping from this webpage [Miami Average price](#).
2. Information about the venues in all boroughs of Miami (including those around John's home): This information is gathered through web scraping from Wikipedia [List of neighborhoods in Miami](#).the Geocoder Python package (<https://geocoder.readthedocs.io/index.html>) is used to receive the latitude and longitude coordinate for all of the boroughs. The boroughs and their corresponding latitude and longitude are used as input for Foursquare to source information about the boroughs.

2.2 How will the data be used to solve the problem

We will start with an exploratory data analysis, where we intend to understand the underlying data. The describe method provides valuable insights for the "average price per m² of the apartments in miami" investigations.

To get a first impression about the distribution of venues in miami, they are visualized using Folium map. The chosen color code will give immediate yet superficial insight, how John's favourite venues are distributed across miami and how the surrounding of his actual location looks like.

For the further analysis, the venues will be divided into two types: Firstly, we have a data frame containing all venues of John's personal interest and one containing all other venues. This subdivision will let us analyse which boroughs are most similar to John's current neighborhood in terms of his personal preferences but also tells us the most common venues within the boroughs. This ensures to find the top borough in terms of John's interests - which is of course the most important criterion - but also provides a list for John with the most common venues in the neighborhoods that he could use for his final decision.

One hot encoding and k-means will narrow the list of the most promising boroughs to three. Combining these three with the pricing analysis lets us recommend the best borough match for John.

2.1 Data preparation

Several steps needed to be performed to use the data and derive meaningful recommendations from it. First of all we gathered the boroughs of Miami through two web scraping . The result is shown in figure(a). you'll notice, that there column **Borough** too but we want the Borough that John live in , so get it from another table . As we would like to get the Average Rent for the Borough seemlier to John Borough we Merge the two tables ,The result is shown in (c) , but before that we need to remove any symbols from the average column and renamed column names. The columns that were not needed for the investigation were dropped. After that, The price was displayed in e/m2 instead of Cents/m2. . The next step was to merge the data frames what turned out to be challenging, as we did not have a price per m2 for every borough. That is why we used a common

:

	Borough
0	Allapattah
1	Arts & Entertainment District
2	Brickell
3	Buena Vista
4	Coconut Grove

figure (b)

:

	Borough	Average Rent
0	Brickell Key	25.75
1	Brickell	22.62
2	Buena Vista	22.06
3	Miami Design District	22.06
4	Midtown - Edgewater	22.01

figure(a)

:

	Borough	Average Rent
0	Allapattah	15.73
1	Arts & Entertainment District	18.01
2	Brickell	22.62
3	Buena Vista	22.06
4	Coconut Grove	18.01
5	Coral Way	18.01
6	Design District	18.01
7	Downtown	18.01

Figure(c)

data science method: The NaNs were replaced by the median of the corresponding borough. The Geocoder Python package [4] was used to get the coordinates (latitude and longitude) for all of the neighborhoods of Miami. The latitude and longitude of Johns current home were received the same way. Details about the Python code can be found in the Jupyter notebook of this report [6]. The next step was to gather the venues within the boroughs in a separate data frame using the Foursquare API [1]. We chose a radius of 750m and 50 as limit of number of venues returned by Foursquare API1. The data frame received from this approach is shown in figure . This data frame was used as input to gather the venues in the boroughs in a separate data frame. Summarized, the following data(frames) were created which were used

for further analysis:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Allapattah	25.808706	-80.215468	Club Tipico Dominicano	25.809557	-80.218593	Nightclub
1	Allapattah	25.808706	-80.215468	Family Dollar	25.810178	-80.213342	Discount Store
2	Allapattah	25.808706	-80.215468	Rincon Progreseno	25.810013	-80.215118	Latin American Restaurant
3	Allapattah	25.808706	-80.215468	Mad Man Jacks	25.810157	-80.212053	BBQ Joint
4	Arts & Entertainment District	43.650597	-79.385537	Shangri-La Toronto	43.649129	-79.386557	Hotel

1- df: Contains the average price per m of the apartments in Miami. This data frame is used to compare the average price per square meter of the boroughs in Miami with the actual price per square meter of Johns apartment.

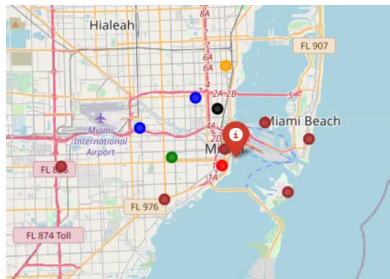
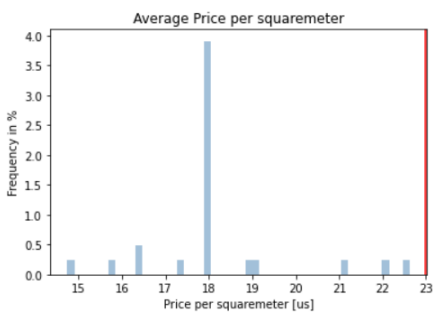
2- mis_venues: Contains all the venues in the boroughs of Munich. This data frame is used to find boroughs that fits most Johns requirements.

3 Methodology

This section represents the main component of the report. It starts with an exploratory data analysis before we dig deeper into solving the problem and applying machine learning algorithms. For the analysis, the venues are divided into two different data frames: Those who are Johns interests and all the other ones. We will perform one hot encoding to narrow the list of the most promising boroughs for both of the venue data frames. Combining the results with a k-mean cluster analysis of all venues in Munich will provide us the most promising neighborhoods for John.

3.1 Exploratory Data Analysis

We started with an exploratory data analysis of the average price per square meter of apartments in all boroughs of Miami which is shown in Figure ?? shows the frequency of prices in Miami. Figure shows the distribution of the prices across Miami and the corresponding color codes are specified in table 3.1.



3.2 One hot encoding

To analyse each neighborhood, one hot encoding is used for both data frames:

- 1- Johns favourite venues:
- 2- All venues in miami except of Johns favourite venues

3.2.1 Johns favourite venues

As we already know, that John prefers to have Gym, coffee stores, bars, restaurants and Art nearby. In this analysis, we used one hot encoding to identify how many of the favorite venues are located in the deferent neighborhoods.

	Neighborhood	Bars	Art Gallery	Cafés	Gym	Restaurants	FavoriteScore
0	Wynwood	5	7	1	1	12	0.117117
1	Brickell	3	0	0	2	18	0.103604
2	Downtown	9	0	1	2	10	0.099099
3	Miami	4	0	2	1	15	0.099099
4	Arts & Entertainment District	2	0	3	1	15	0.094595
5	Lummus Park	5	3	0	1	10	0.085586
6	Edgewater	2	0	0	1	13	0.072072
7	Park West	2	1	3	1	5	0.054054
8	Coconut Grove	4	0	0	4	4	0.054054
9	Midtown	0	1	0	2	8	0.049550

	Neighborhood	Bars	Art Gallery	Cafés	Gym	Restaurants	FavoriteScore
19	Wynwood	5	7	1	1	12	0.117117
17	Park West	2	1	3	1	5	0.054054

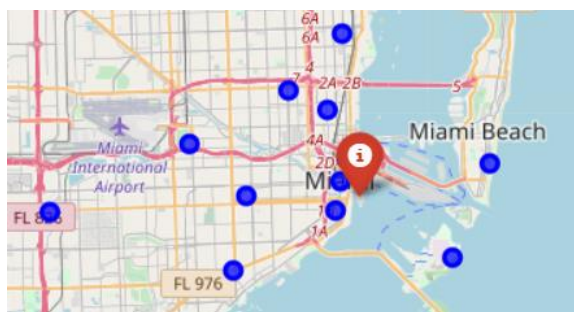
All venues in Miami except Johns favorite venues

In this analysis we identified the ten most common venues in all boroughs of Miami without taking into account the venues of Johns interest. John could use this list to explore the boroughs further before he makes a final decision.

	Neighborhood	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
0	Allapattah	BBQ Joint	Discount Store	Nightclub	Women's Store	Exhibit	Construction & Landscaping	Convenience Store	Cosmetics Shop	Creperie	Department Store
1	Arts & Entertainment District	Coffee Shop	Steakhouse	Concert Hall	Pizza Place	Theater	Bakery	Lounge	Salad Place	Burrito Place	Monument / Landmark
2	Brickell	Hotel	Burger Joint	Coffee Shop	Grocery Store	Lounge	Pharmacy	Spa	Ice Cream Shop	Smoke Shop	Bakery
3	Coconut Grove	Pizza Place	Sandwich Place	Pet Store	Bagel Shop	Boutique	Donut Shop	Mattress Store	Speakeasy	Sporting Goods Shop	Intersection
4	Coral Way	Ice Cream Shop	Cosmetics Shop	Snack Place	Coffee Shop	Arcade	Event Space	Clothing Store	Hotel	Creperie	Dessert Shop

3.2.3 Clustering of the neighborhoods

The k-means clustering was used to cluster the neighborhood into 4 clusters. In this analysis all venues in the boroughs of Miami were taken into account. To do that, we first used one hot encoding, then calculated the mean of each venue in each neighborhood and finally grouped the data frame based on the neighborhood. The distribution of the clusters is shown in figure (a)..



	Neighborhood
0	Allapattah
1	Arts & Entertainment District
2	Upper Eastside
3	Park West
4	Overtown
5	Midtown
6	Miami

4 Results and Discussion

Different analysis have been performed to find the most suitable new neighborhood for John. Considering the price per m John can move anywhere, as all of the boroughs have a lower price per m.⁹ The venues within the different neighborhoods have been divided in two different data frames: a) Johns favorite venues and b) All venues in miami except of Johns favorite venues. For the analysis where just Johns favorite venues have been considered (a), we defined a favorite score, taking into account, how often Johns favorite venues occurred in the specific neighborhoods. This analysis narrowed down the results to the following three neighborhoods:

1- Wynwood

2- Park Weast

For the analysis, where all venues in Miami except of Johns favorite venues have been considered (b), we identified the 10 most common venues within all of the boroughs which could be helpful for John, to further explore other neighborhoods. Based on the k-means cluster analysis, the following neighborhoods are most similar to the one John currently lives in:

1- Upper Eastside

2- Park Weast

3- Over town

4- Midtown

If you combine these findings, the most promising borough for John is identified: Wynwood as this is the one with the highest favorite score, but also the one which is most similar to his current place according to the k-means algorithm.

5 Conclusion

The purpose of this project was to identify a borough that is similar to Johns current one and has venues, that are important for John in order to narrowing down the search for optimal new borough as a place to live. For this report different analysis have been performed. Considering the average prices per m² for apartments in Miami showed, that no matter where John is moving he will most likely find a cheaper apartment.

By just taking into account the venues that are important for John, we could identify two boroughs, which seemed to be most promising as new neighborhood for Johns home. The k-means provided an insight into similar neighborhoods, compared to Johns actual one. After combining these results, we identified one single borough, that is most likely the best choice for Johns new area to live: Wynwood. Further dividing this neighborhood, showed, that the one Wynwood would be the best choice.