

# An Analysis of Boston Neighborhoods

*William Dean*

## Overview

### Motivation

I really enjoy to travel especially visiting new cities around the world. I currently live in Boston, MA and have become very familiar with the area.

In a few of my most recent travels, I have found difficulty exploring bigger cities. When spending a few days in a certain city, I often end up spending a lot of time traveling within the city, not having enough time to connect with the people and the cultures.

### Goal

The *goal of this project* is to use data from areas and neighborhoods of cities to see how communities compare to others within the same city and also to others in different.

When looking each neighborhood, there a few things that I am interested in measuring:

- What are the people like? How are the people that live in the area and what do people who visit think of the area and their time there.
- What are the living situations like? What kind of houses are in the area.
- What is the culture like? What are these communities known for and what are people talking about?

### Data

I wanted data that would reflect all these measure. Here are a few sources I found.

#### Airbnb

The Airbnb Datasets provides insight into many cities in the US and around the world. Not only are there reviews written by people who use the service, but also the listings have information about the neighborhood, the house, and the host.

Using this information will provide information for the people who live in the neighborhood, details of the neighborhoods from the hosts, and the experiences of the people who visit.

#### Twitter

Twitter data gives insight to what is going on within a neighborhood, and what people are talking about. This accomplished trying to capture the neighborhood culture and potentially what people are like.

## Selected Data

I ended looking at the Boston data mainly and also using data from Chicago. There were a few limitations with the Airbnb data from Chicago where only the house listing was provided. With it, there was no information about the host, or the reviews themselves from the people who stayed there.

## Methods

Doing the analysis, I mainly used a bag of words approach to the text analysis. For a neighborhood's set of reviews, I collected all the defining words for a given neighborhood as well as important sets of two words for a neighborhood.

When comparing two or more neighborhoods to each other, I used latent Dirichlet allocation (LDA) to group into a set of  $k$  topics.

## Boston

### People from each Neighborhood

#### Hosts

Each Airbnb Host provides a description of themselves and their household. Using the methods stated above, I grouped terms from each neighborhood. This process appeared to catch terms associated with people and aspects of their lives. Consider the words chosen for Brighton, MA below:

---

Brighton Neighborhood Bigrams

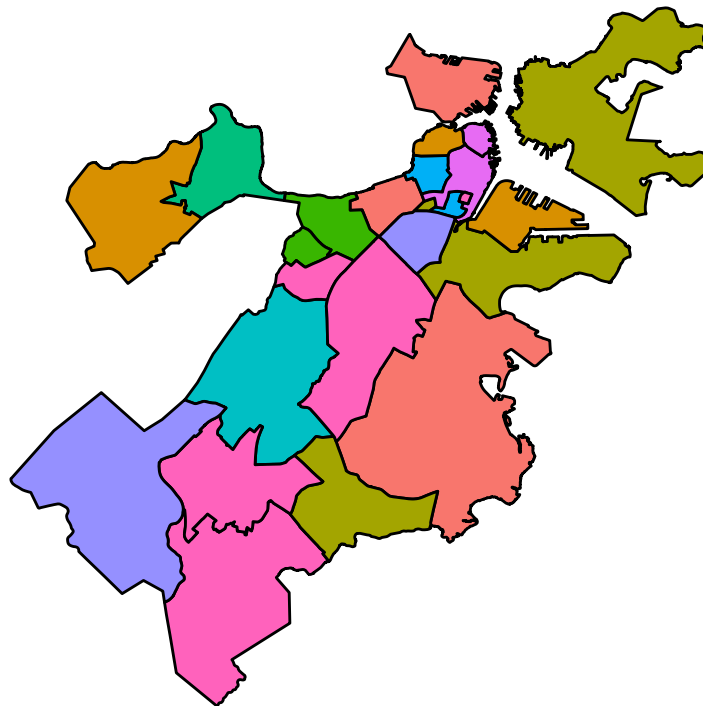
---

england france  
tech company  
accountant downtown  
action movies  
active person

---

The third row in context of its host description is: "Lorin works as an *accountant downtown* and takes the expressway bus which is a quick 20 minute ride."

Doing this for all of the neighborhoods, and grouping all 25 Boston neighborhoods into 15, we see the following result:



Regions of the same color (same grouping) appear either next to each other or are in similar proximity to downtown. This seems likely with people living outside of downtown and commuting into the city.

### People's Reviews

Considering all the reviews of the Airbnbs for the neighborhoods of Boston and which area's reviews are most similar, we see the results below:

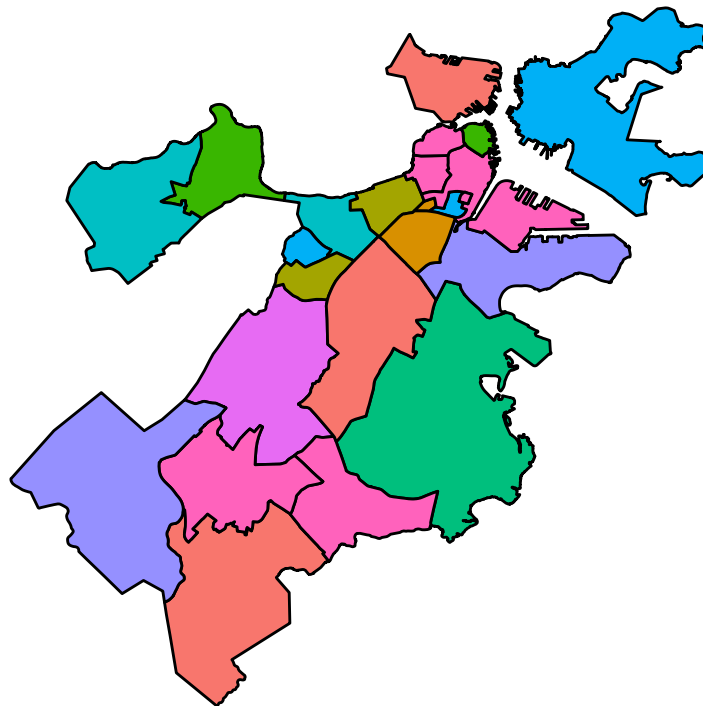


The contents of what people are saying into their reviews seems to depend on what neighborhood they are staying. Again, many neighborhoods far from the heart of Boston look to be grouped together, and as the neighborhoods get closer, there appear to stay grouped together. Southie and East Boston group together, the Fenway area, and JP Neighborhoods.

## **Boston Neighborhoods**

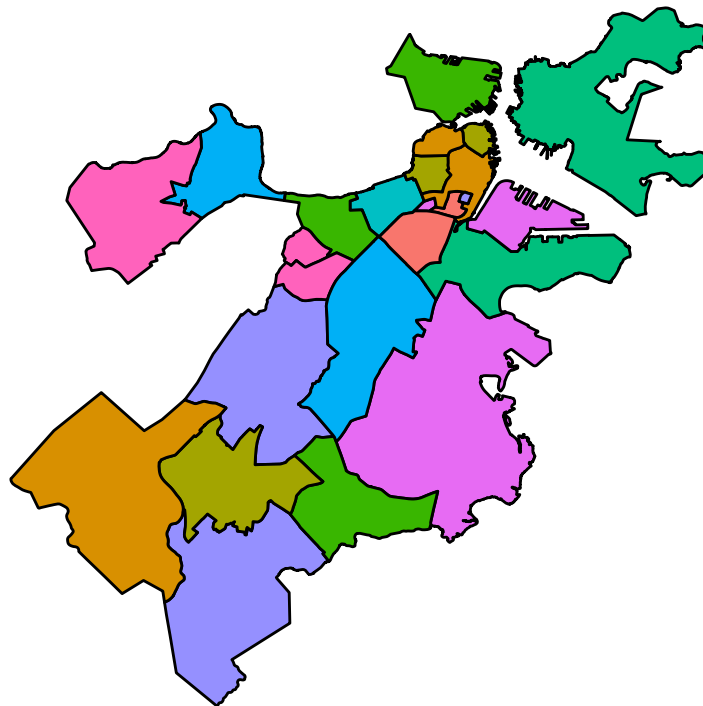
### **Neighborhood Descriptions**

Every Airbnb host provides a description of the neighborhood where their property. From the results, we see that many of the adjacent neighborhoods in Boston are grouped together. For example, Downtown and Seaport are grouped together, but Southie is grouped by itself.



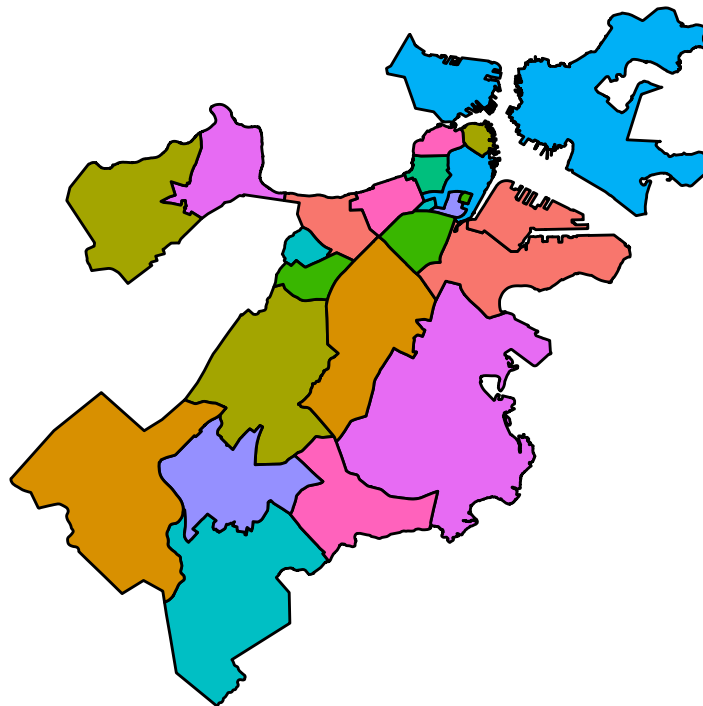
## Housing

Using each Airbnb listing's detailed description of the property, we gain information about the type of homes in each location.



## Culture

Looking at tweets associated with each neighborhood either through direct mention or in a hashtag, we can pick up on what people in each area are talking about. The tweets were sample during the week of November, 30th. Here are the clusters:



## Comparison of Boston and Chicago

After looking at Boston, let's consider how another city's neighborhoods compare to the ones we just explored.

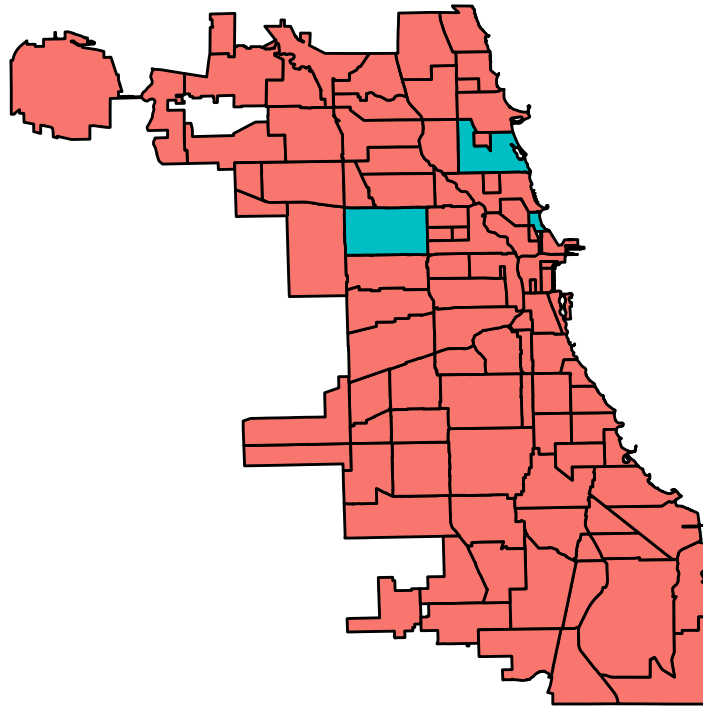
There are some limitations as stated before with the data from Chicago. We will only compare the tweets from both cities and the names of the Airbnb listings.

### Tweets

This data will help understand which neighborhoods in Boston are talking about similar topics as the neighborhoods in Chicago. Using LDA with 25 topics for all 25 Boston neighborhoods, we can check if any of the Boston neighborhoods match with neighborhoods in Chicago, or if the neighborhoods in Boston are completely different than Chicago.

Below are the areas in Chicago that are similar to Downtown Boston based of the twitter data. The projected similar locations in Chicago are in fact close to the downtown area of Chicago as well.

City	Neighborhood
Boston	West End
Boston	Downtown
Chicago	Gold Coast
Chicago	Lake View East
Chicago	Humboldt Park
Chicago	Lakewood - Balmoral



## Description of Houses

The Airbnb information for Chicago is quite limited however, comparing the listing names for the properties in Boston and Chicago give the following results.

Grouping	% Boston Neighborhood	Grouping1	% Boston Neighborhood1
1	0.4285714	14	0.6666667
2	0.1428571	15	0.0000000
3	1.0000000	16	0.5000000
4	0.6666667	17	1.0000000
5	0.3333333	18	0.5000000
6	0.1250000	19	0.0000000
8	0.0000000	20	0.5000000
9	0.0000000	21	0.5000000
10	0.0000000	22	0.1250000
11	0.5000000	23	0.0000000
12	0.0666667	24	0.3333333
13	0.0000000	25	0.5000000

Using property listing names to represent the type of homes in an area may be a stretch, but it does group neighborhoods in Boston with neighborhoods in Chicago.



## Conclusion

The neighborhoods in Boston have a interesting relationship with each other and are similiar in as many ways as they are different.

## Next Steps

Even though I was able to shed light behind the questions I had starting this project, I think this is just one approach to the question. Continuing on, I would like to see what other data I could use to expand on this and maybe separate each neighborhood into further classifications.

## What I Learned

With this project, this was the first time where I used only text data to answer questions. I got very familiar with the packages in R for text mining, mainly `tidytext`, `tm`, and `topicmodels`.