

The Battle of Neighborhoods: Travel Edition

Richie Rivera

Can we
recommend a
neighborhood
based on previous
travel behaviours?

Background

Why?

- To suggest travel and relocation neighborhoods to potential customers
 - To use known and liked features about a given neighborhood to understand where similar populations are
-

Data



1. Scraped data from Airbnb's website to determine the neighborhoods in a given city.
2. The latitude and longitude data from GeoPy's Package
3. Venue information mined using the Foursquare API

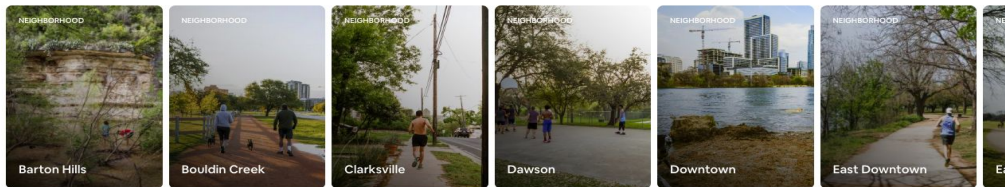
```
{
  "meta": {
    "code": 200,
    "requestId": "5ac51d7e6a607143d811cecb"
  },
  "response": {
    "venues": [
      {
        "id": "5642aef9498e51025cf4a7a5",
        "name": "Mr. Purple",
        "location": {
          "address": "180 Orchard St",
          "crossStreet": "btwn Houston & Stanton St",
          "lat": 40.7213744277209,
          "lng": -73.98800687282996,
          "labeledLatLngs": [
            {
              "label": "display",
              "lat": 40.7213744277209,
              "lng": -73.98800687282996
            }
          ]
        },
        "distance": 8,
        "postalCode": "10002",
        "cc": "US",
        "city": "New York",
        "state": "NY",
        "country": "United States",
        "formattedAddress": [

```

Neighborhoods

Austin

This tender-hearted city's deep grooves and Tex-Mex tacos promise boot-stomping, finger-licking good times.



Methodology



Target Variable

The Similarity Score

In order to create a recommendation, a similarity score was used. The inputs to this similarity score is an average of the nearest 100 venues from the geographical center of the neighborhood. The average was chosen to account for neighborhoods which don't have 100 venues within a 500 meter radius from the center.

In order to determine similar neighborhoods, we used the euclidean distance

In order to calculate the similarity score the euclidean distance was used. The euclidean distance was used because it prefers neighborhoods with similar relative frequencies and also was a good measure for neighborhoods with unique venues.

The Process

- Used Downtown Brooklyn, NYC to recommend a neighborhood in San Francisco
- Used the relative frequency
- Iterated through each San Francisco neighborhood
- Determine the most similar neighborhood based on the one with the shortest euclidean distance

Results

—

SoMa, SFO

- The analysis provided that SoMa, San Francisco is the most similar to Downtown Brooklyn
- The euclidean distance between Downtown Brooklyn and SoMa is 0.105
- There are similarities between the most common venues between both neighborhoods

	neighborhood	euclidean_distance
34	SoMa	0.105086
2	Bernal Heights	0.110682
9	Duboce Triangle	0.115594
0	Alamo Square	0.115635
15	Hayes Valley	0.117898

----SoMa----

	venue	freq
0	Pizza Place	0.05
1	Spa	0.04
2	Coffee Shop	0.04
3	Gym / Fitness Center	0.03
4	Nightclub	0.03

	Neighborhood	variable	value
12	Downtown Brooklyn	Burger Joint	0.05
49	Downtown Brooklyn	Pizza Place	0.05
17	Downtown Brooklyn	Coffee Shop	0.04
34	Downtown Brooklyn	Hotel	0.03
28	Downtown Brooklyn	Grocery Store	0.03

Discussion

Unique local venues

Regional and local differences cause for various differences in many venue types. For example, there are restaurant types in Tokyo which are unique to Japan. This caused for potential unnecessary omissions in venues.

Venue Categories are too granular

To increase the relevance of this analysis, it is recommended to use more general venue categories. For example, rather than using "Burger Joint", "Taco Stand", "Hotdog Cart", etc it may be better to label them all as "Quick Eats".

Conclusion



Conclusion

- By using the euclidean distance between the relative frequency of venues in neighborhoods we can find a relevant method to find similar neighborhoods.
- SoMa in San Francisco is the most similar neighborhood to Downtown Brooklyn in New York City.
- Regional differences may have caused for omitted neighborhoods in the baseline and target neighborhoods.