Finding Similar Neighborhoods Based on Most Common Type of Venue

ALLISON WU

IBM COURSERA DATA SCIENCE PROFESSIONAL CERTIFICATE

CAPSTONE PRESENTATION

Introduction

BACKGROUND AND BUSINESS PROBLEM

Can location data be useful for relocation purposes?

- Can find neighborhoods in a new city that are similar to neighborhoods in old/familiar city
- Can know ahead of time that new neighborhood has resources and venues of interest
- Makes moving easier
- Takes less time to adjust to new surroundings

Business Problem

For someone who currently lives in one city, can we find the most similar neighborhood in another city based on the kinds of venues that are most common in that neighborhood in order to ease the relocation transition?

Specifically, we will explore Charleston, South Carolina and Pittsburgh, Pennsylvania. We will also consider a home-buying budget of $300,000.

Two Different Cities

Charleston, SC

- Total population (July 2019): **137,566**
- # households: **55,889**
- **74.1%** White
- **21.7%** Black/African American
- **1.9%** Asian
- **3.2%** Hispanic/Latino

Pittsburgh, PA

- Total population (July 2019): **300,286**
- # households: **138,058**
- **66.8%** White
- **23.0%** Black/African American
- **5.8%** Asian
- **3.2%** Hispanic/Latino

Charleston, SC

- Average peak summer temperature: **91° F**
- Average low winter temperature: **38° F**
- **2.3** craft breweries per capita
- Median home value: **$348,052**

Pittsburgh, PA

- Average peak summer temperature: **83° F**
- Average low winter temperature: **21° F**
- **4.1** craft breweries per capita
- Median home value: **$195,195**

Data

DATA SOURCES & ACQUISITION

DATA CLEANING & PROCESSING

Defined Factors

- Neighborhoods in Charleston
- Neighborhoods in Pittsburgh
- The number of breweries per neighborhood in Charleston
- The number of breweries per neighborhood in Pittsburgh
- Median home values per neighborhood in Charleston
- Median home values per neighborhood in Pittsburgh

Data Sources & Acquisition

- Zillow Home Value Index (2020)
  - Exported .csv file & uploaded to IBM Cloud Storage
- Foursquare API
  - JSON list of 100 venues within a 1 mile radius of each neighborhood's center based on obtained latitude & longitude coordinates

In [4]:
```
chs_zhvi=zillow_zhvi[zillow_zhvi['City']=='Charleston'].reset_index(drop=True)
chs_zhvi=chs_zhvi[chs_zhvi['State']=='SC'].reset_index(drop=True)
chs_zhvi=chs_zhvi.loc[:,['RegionName','State','City','Metro','2020-11-30']]
chs_zhvi.rename(columns={'2020-11-30':'Median Home Value 2020-11'},inplace=True)
chs_zhvi['neighborhood']=chs_zhvi[['RegionName', 'State']].agg(', '.join, axis=1)
print('there are',chs_zhvi.shape[0],"neighborhoods in Zillow's Home Index database for Charleston, SC")
chs_zhvi.head()
```

```
there are 26 neighborhoods in Zillow's Home Index database for Charleston, SC
```

Out[4]:

	RegionName	State	City	Metro	Median Home Value 2020-11	neighborhood
0	Harleston Village	SC	Charleston	Charleston-North Charleston	892637.0	Harleston Village, SC
1	Daniel Island	SC	Charleston	Charleston-North Charleston	876967.0	Daniel Island, SC
2	Cannonborough-Elliottbororugh	SC	Charleston	Charleston-North Charleston	568652.0	Cannonborough-Elliottbororugh, SC
3	Shadowmoss	SC	Charleston	Charleston-North Charleston	314499.0	Shadowmoss, SC
4	Wagener Terrace	SC	Charleston	Charleston-North Charleston	564916.0	Wagener Terrace, SC

We extracted information for 26 neighborhoods in Charleston.

In [5]:
```
pit_zhvi=zillow_zhvi[zillow_zhvi['City']=='Pittsburgh'].reset_index(drop=True)
pit_zhvi=pit_zhvi[pit_zhvi['State']=='PA'].reset_index(drop=True)
pit_zhvi=pit_zhvi.loc[:,['RegionName','State','City','Metro','2020-11-30']]
pit_zhvi.rename(columns={'2020-11-30':'Median Home Value 2020-11'},inplace=True)
pit_zhvi['neighborhood']=pit_zhvi[['RegionName', 'State']].agg(', '.join, axis=1)
print('there are',pit_zhvi.shape[0],"neighborhoods in Zillow's Home Index database for Pittsburgh, PA")
pit_zhvi.head()
```

```
there are 77 neighborhoods in Zillow's Home Index database for Pittsburgh, PA
```

Out[5]:

	RegionName	State	City	Metro	Median Home Value 2020-11	neighborhood
0	Mount Lebanon	PA	Pittsburgh	Pittsburgh	354231.0	Mount Lebanon, PA
1	Squirrel Hill South	PA	Pittsburgh	Pittsburgh	442777.0	Squirrel Hill South, PA
2	Shadyside	PA	Pittsburgh	Pittsburgh	612067.0	Shadyside, PA
3	Brookline	PA	Pittsburgh	Pittsburgh	163387.0	Brookline, PA
4	Squirrel Hill North	PA	Pittsburgh	Pittsburgh	669792.0	Squirrel Hill North, PA

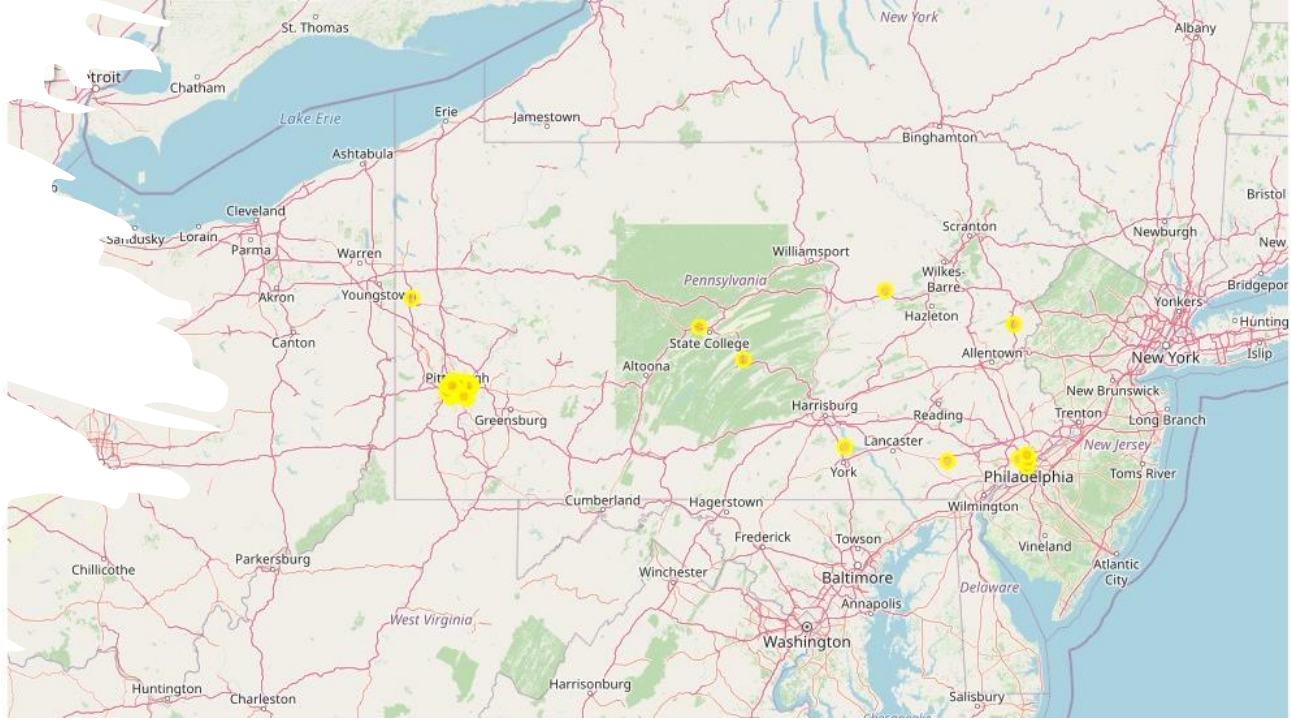
We extracted information for 77 neighborhoods in Pittsburgh.

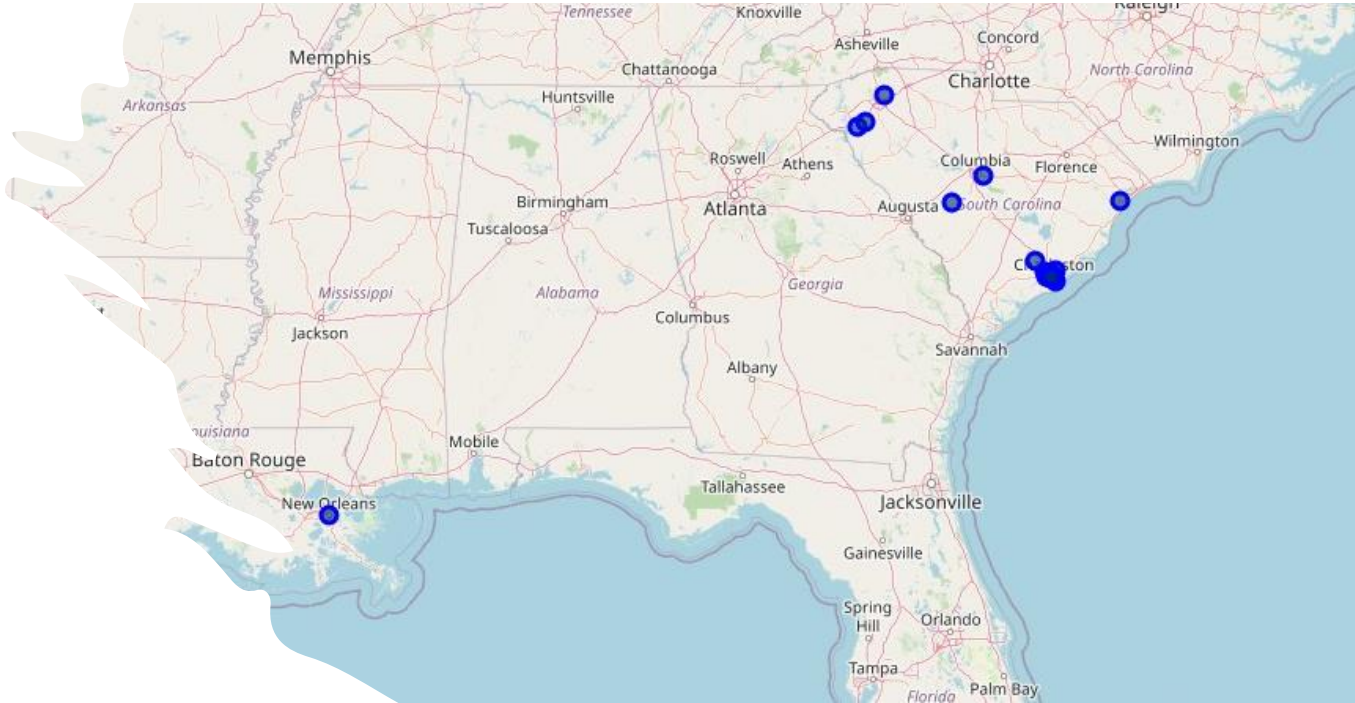
Data Cleaning & Processing

- Zillow Home Value Index data uploaded to IBM Cloud Storage
  - Converted from .csv format to Pandas dataframe
- Data cleaned to only include neighborhoods in Charleston, SC and Pittsburgh, PA
  - Data then split into 2 separate dataframes, one for each city

Data Cleaning & Processing (con't.)

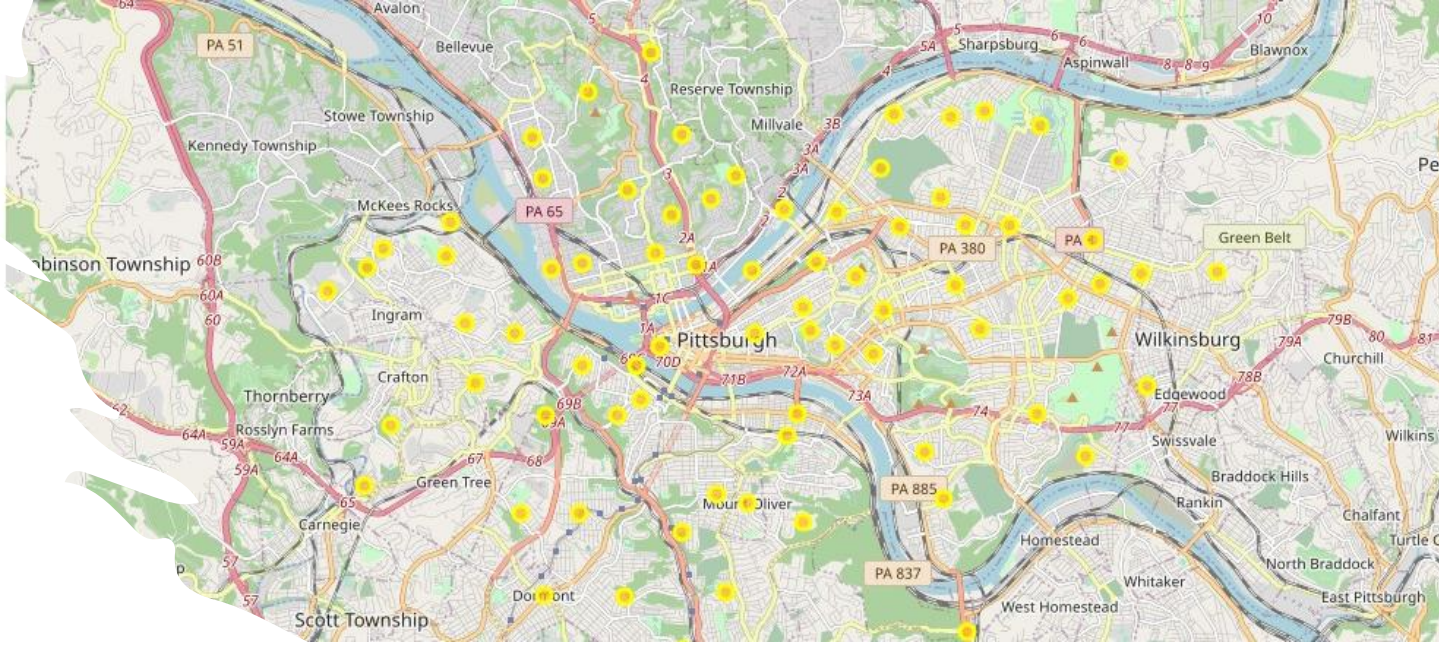
- Used Geopy Nominatim to find latitude & longitude of each neighborhood
- Used Folium to map neighborhoods to check accuracy of coordinates
- Found inaccuracies

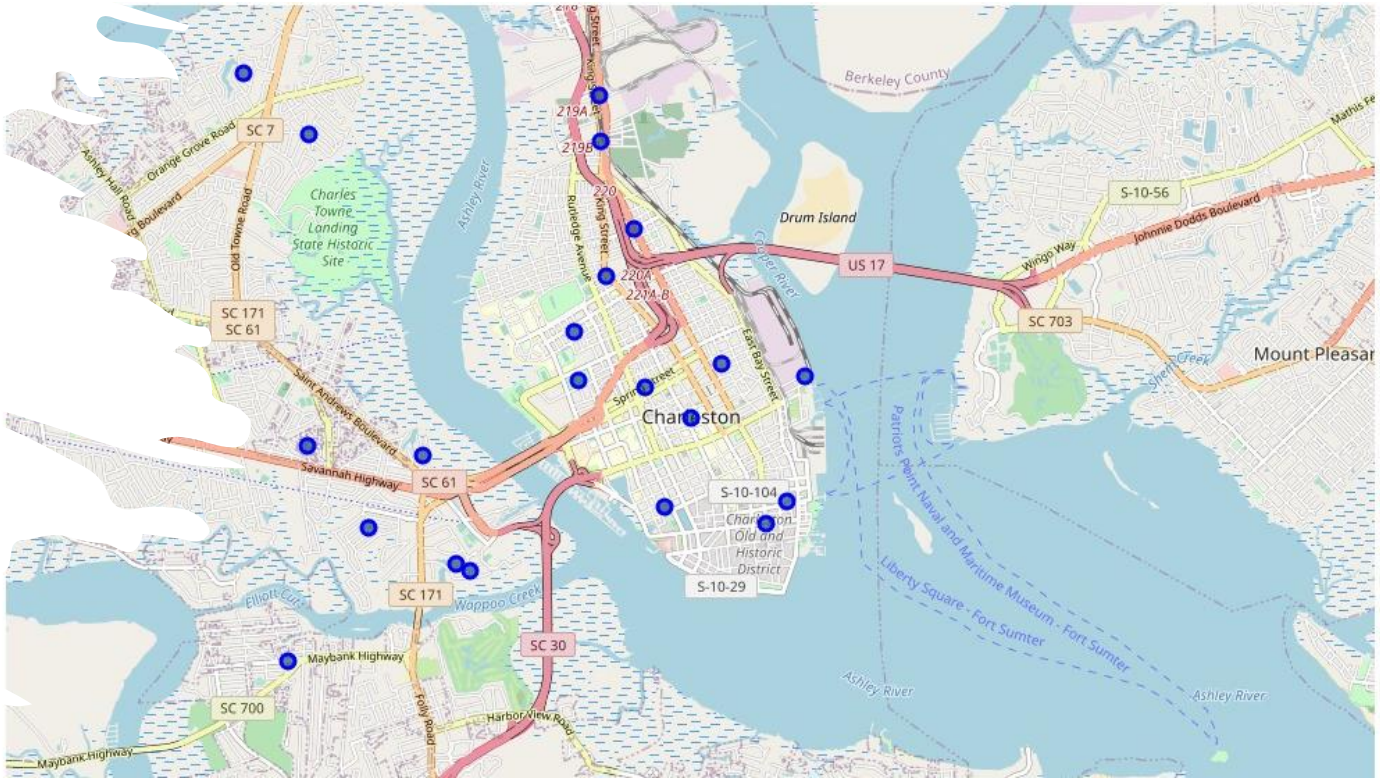




Data Cleaning & Processing (con't.)

- Manually input correct coordinates for neighborhoods outside of known city bounds
- Re-visualized maps to check coordinates again





Methods & Results

Methods

- Connected to Foursquare API & extracted JSON data for venues
  - Limit = 100 venues
  - Radius = 1 mile (1609.34 meters)
- From JSON data, made a dataframe for each city that included venue name, category, latitude, & longitude
  - Found 1,711 venues in Charleston neighborhoods
  - Found 4,451 venues in Pittsburgh neighborhoods

	RegionName	State	City	Metro	Median Home Value 2020-11	neighborhood	lat	long	Brewery
13	Silver Hill-Magnolia	SC	Charleston	Charleston-North Charleston	205635.0	Silver Hill-Magnolia, SC	32.820900	-79.951600	0.240000
4	Wagener Terrace	SC	Charleston	Charleston-North Charleston	564916.0	Wagener Terrace, SC	32.816200	-79.951400	0.162162
14	East Central	SC	Charleston	Charleston-North Charleston	343867.0	East Central, SC	32.807000	-79.947300	0.112903
6	North Central	SC	Charleston	Charleston-North Charleston	455721.0	North Central, SC	32.802100	-79.950700	0.063291
20	Sandhurst	SC	Charleston	Charleston-North Charleston	439524.0	Sandhurst, SC	32.823232	-79.995645	0.034483
19	Hampton Park Terrace	SC	Charleston	Charleston-North Charleston	696325.0	Hampton Park Terrace, SC	32.796300	-79.954700	0.034091
17	Northbridge Terrace	SC	Charleston	Charleston-North Charleston	405576.0	Northbridge Terrace, SC	32.816843	-79.987590	0.027778
2	Cannonborough-Elliotborough	SC	Charleston	Charleston-North Charleston	568652.0	Cannonborough-Elliotborough, SC	32.790600	-79.945900	0.020000
15	South Windermere	SC	Charleston	Charleston-North Charleston	627225.0	South Windermere, SC	32.776010	-79.980090	0.012195
24	Moreland	SC	Charleston	Charleston-North Charleston	511431.0	Moreland, SC	32.783510	-79.973423	0.011628

	RegionName	State	City	Metro	Median Home Value 2020-11	neighborhood	lat	long	Brewery
63	Spring Garden	PA	Pittsburgh	Pittsburgh	129360.0	Spring Garden, PA	40.471000	-79.988000	0.125000
41	Upper Lawrenceville	PA	Pittsburgh	Pittsburgh	251403.0	Upper Lawrenceville, PA	40.482037	-79.951025	0.125000
39	Spring Hill - City View	PA	Pittsburgh	Pittsburgh	132676.0	Spring Hill - City View, PA	40.466730	-79.993929	0.086957
21	Lincoln-Lemington-Belmar	PA	Pittsburgh	Pittsburgh	57826.0	Lincoln-Lemington-Belmar, PA	40.473523	-79.898046	0.076923
22	Central Lawrenceville	PA	Pittsburgh	Pittsburgh	291035.0	Central Lawrenceville, PA	40.472349	-79.953866	0.070000
20	Stanton Heights	PA	Pittsburgh	Pittsburgh	214786.0	Stanton Heights, PA	40.481407	-79.937572	0.066667
45	Lower Lawrenceville	PA	Pittsburgh	Pittsburgh	267422.0	Lower Lawrenceville, PA	40.464514	-79.964399	0.060000
71	Herrs Island	PA	Pittsburgh	Pittsburgh	650633.0	Herrs Island, PA	40.465062	-79.976818	0.060000
35	Morningside	PA	Pittsburgh	Pittsburgh	232856.0	Morningside, PA	40.482434	-79.929518	0.053571
69	Chartiers City	PA	Pittsburgh	Pittsburgh	118351.0	Chartiers City, PA	40.458012	-80.070694	0.043478

Methods (con't.)

- Generated dummy-codes for venue categories
  - 188 categories for Charleston venues
  - 285 categories for Pittsburgh venues
- Calculated frequency of venue category & grouped venues by neighborhood
- Merged dataframe with Zillow Home Value Index dataframe for each city

Results

- In Charleston, **Silver Hill-Magnolia** had the most brewery venues (frequency=**0.24**)
- In Pittsburgh, both **Spring Graden** & **Upper Lawrenceville** had the most brewery venues (frequency = **0.125**)
- Most similar Charleston neighborhood to Pittsburgh neighborhoods is actually East Central (frequency = **0.113**)
- Median Home Values
  - Spring Garden: **$129,360**
  - Upper Lawrenceville: **$251,403**

Conclusion & Future Directions

- Spring Garden & Upper Lawrenceville in Pittsburgh are most similar to East Central in Charleston
- Relocation to either neighborhood will fall below $300,000 budget
- ZHVI data is dated November 2020
  - Would be most helpful to have current home value prices
- Can take into account additional factors to compare neighborhood similaritiy
  - E.g., Proximity to known secondary location (ex., workplace) or school districts
- Can perform statistical analysis on multiple factors to generate objective similarity measure
  - E.g., Analysis of Variance, or *t*-test