Pittsburgh Pennsylvania Regional Dataset

Pittsburgh is a city in western Pennsylvania at the junction of 3 rivers. Its Gilded Age sites, including the Carnegie Museum of Natural History, the Carnegie Museum of Art and the Phipps Conservatory and Botanical Gardens, speak to its history as an early-20th-century industrial capital. In the North Shore neighborhood are the modern Andy Warhol Museum, Heinz Field football stadium and PNC Park baseball stadium.

Business Problem

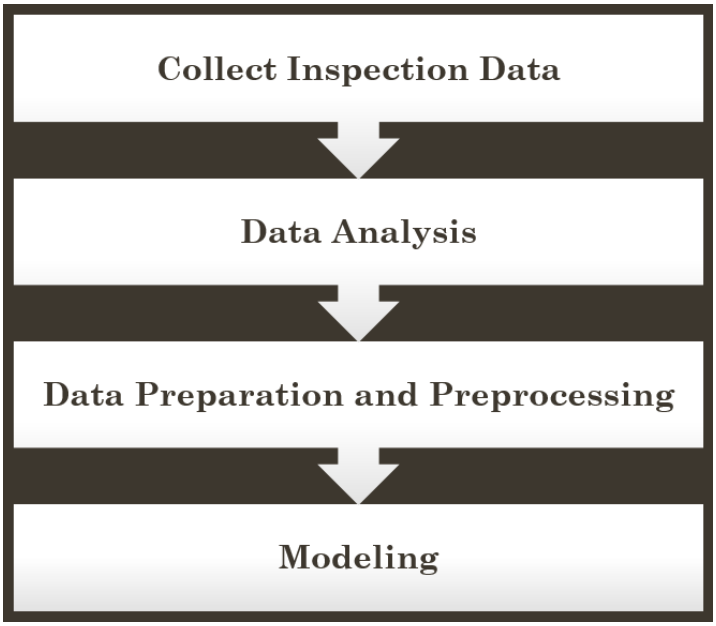
In this project the business problem is how we can move to a new neighborhood in Pittsburgh PA with similar venues providing information about the area along with similar house prices that clients can afford.

This project utilizes machine learning techniques to provide insights on how homeowners can choose a desired neighborhood based on their needs. This project retrieves all schooling, restaurant, bar, gym, supermarket and many more entertainment venues to divide these neighborhoods to colored clusters in maps to provide visual similarities in the city of Pittsburgh, PA.

Dataset: Allegheny County Property Sale Transactions

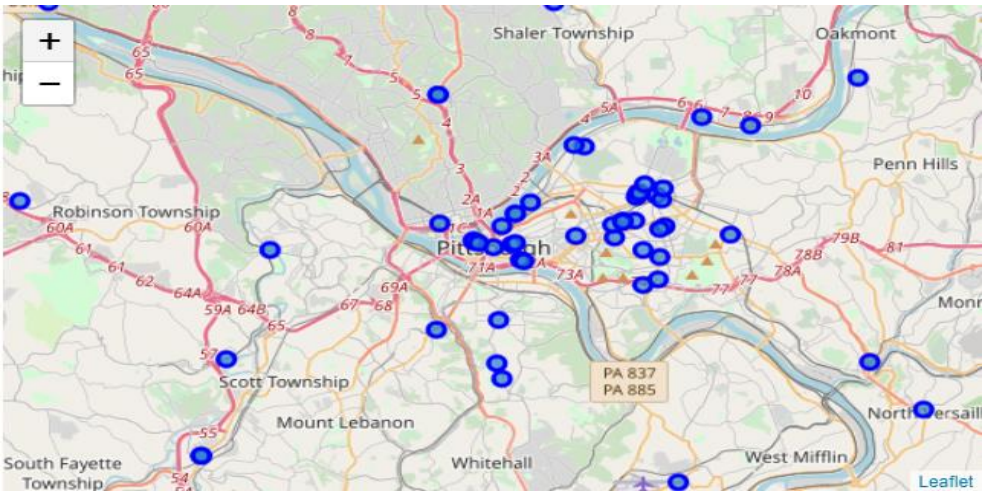
This dataset obtained from Allegheny County (Pennsylvania) and the City of Pittsburgh both publish their data through the Western Pennsylvania Regional Data Center. The Western Pennsylvania Regional Data Center supports key community initiatives by making public information easier to find and use. The Data Center is managed by the University of Pittsburgh's Center for Social and Urban Research, and is a partnership of the University, Allegheny County and the City of Pittsburgh. Contains data on all Real Property parcels that have sold since 2013 in Allegheny County, PA.

Methodology



A. Collect Inspection Data

This data was collected from Allegheny County (Pennsylvania) and the City of Pittsburgh both publish their data through the Western Pennsylvania Regional Data Center. The Western Pennsylvania Regional Data Center supports key community initiatives by making public information easier to find and use. The Data Center is managed by the University of Pittsburgh's Center for Social and Urban Research, and is a partnership of the University, Allegheny County and the City of Pittsburgh. Contains data on all Real Property parcels that have sold since 2013 in Allegheny County, PA.



B. **Data Analysis**

The Foursquare API was used to search for nearby venues of each neighborhood in radius of 1000 meters. Only the venues name and venues category (i.e. café, restaurant, school, etc.) are extracted. After obtaining all the venues, the total number of venues in each category is counted for each neighborhood. To clarify, I also build a table with top 10 most frequent venue categories for each neighborhood.

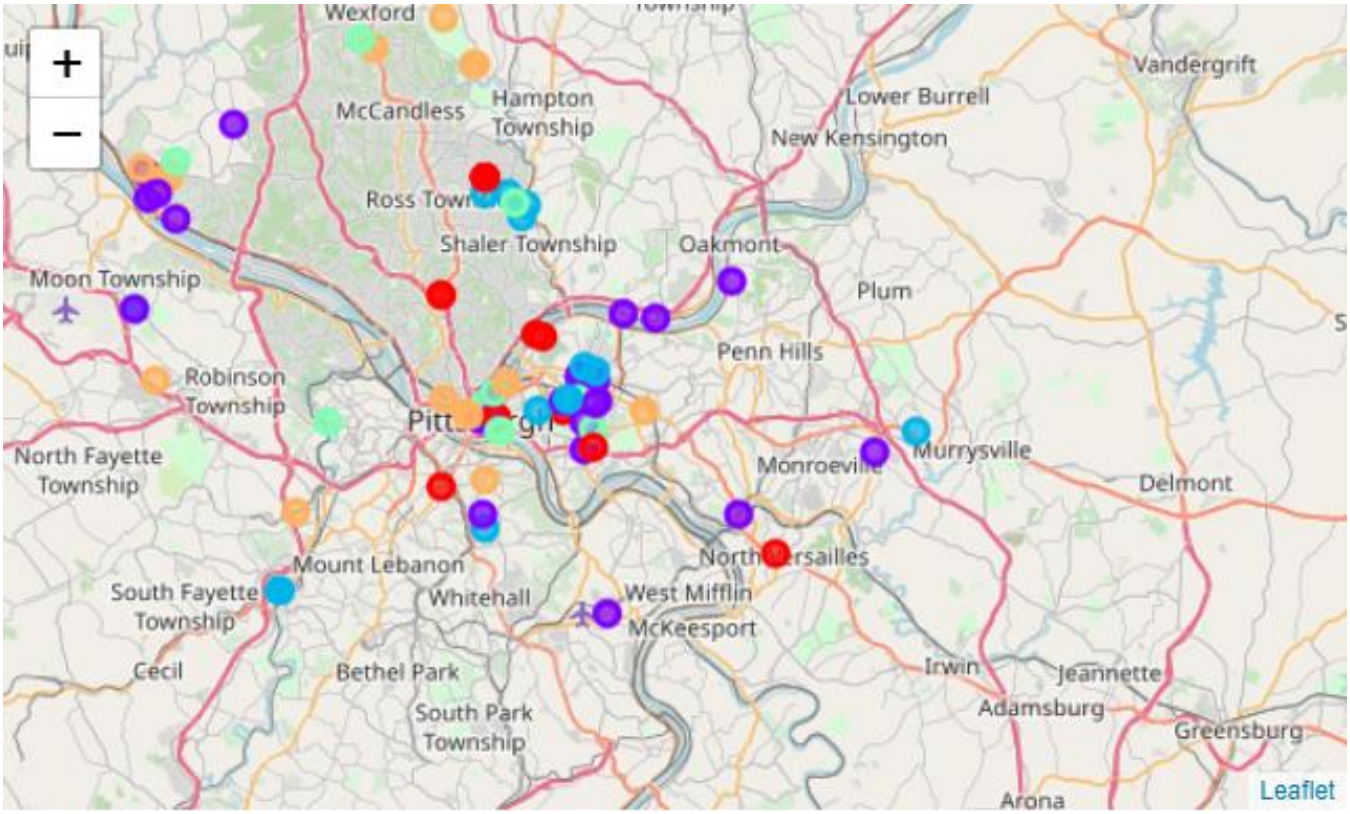
Street	Street Latitude	Street Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0 ALPHA DR ALLISON PARK PA 15101	2	2	2	2	2	2
0 AMBER ST PITTSBURGH PA 15206	49	49	49	49	49	49
0 BRANDYWINE CT MC DONALD PA 15057	2	2	2	2	2	2
0 BUTLER PLANK RD GLENSHAW PA 15116	6	6	6	6	6	6
0 LINCOLN HWY NORTH VERSAILLES PA 15137	6	6	6	6	6	6
0 MARYLAND DR GLENSHAW PA 15116	4	4	4	4	4	4
0 PENN AVE PITTSBURGH PA 15206	52	52	52	52	52	52
0 SHADY AVE PITTSBURGH PA 15232	57	57	57	57	57	57
0 STEUBENVILLE PIKE PITTSBURGH PA 15205	2	2	2	2	2	2
0 SUTTER RD GLENSHAW PA 15116	2	2	2	2	2	2
0 WOODLAND RD SEWICKLEY PA 15143	2	2	2	2	2	2

	Street	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	0 ALPHA DR ALLISON PARK PA 15101	Park	Pizza Place	Donut Shop	Fish & Chips Shop	Financial or Legal Service	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Exhibit	Event Space
1	0 AMBER ST PITTSBURGH PA 15206	Coffee Shop	Grocery Store	Gym / Fitness Center	Hotel	American Restaurant	Salon / Barbershop	Liquor Store	Plaza	Pharmacy	Pet Store
2	0 BRANDYWINE CT MC DONALD PA 15057	Brewery	Steakhouse	Zoo	Dumpling Restaurant	Fish & Chips Shop	Financial or Legal Service	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Exhibit
3	0 BUTLER PLANK RD GLENSHAW PA 15116	Supermarket	Pharmacy	Salon / Barbershop	Gym / Fitness Center	Bank	Gas Station	Dry Cleaner	Financial or Legal Service	Fast Food Restaurant	Farmers Market
4	0 LINCOLN HWY NORTH VERSAILLES PA 15137	Pawn Shop	American Restaurant	Gym	Video Store	Pharmacy	Flower Shop	Dance Studio	Dry Cleaner	Financial or Legal Service	Fast Food Restaurant

	Street	Price	Latitude	Longitude
74	0 ALPHA DR ALLISON PARK PA 15101	1950000.0	40.551668	-79.987459
79	0 AMBER ST PITTSBURGH PA 15206	1500000.0	40.460333	-79.931427
215	0 BRANDYWINE CT MC DONALD PA 15057	2200000.0	40.708323	-75.226708
262	0 BUTLER PLANK RD GLENSHAW PA 15116	2400000.0	40.538425	-79.962543
811	0 LAKE MARSHALL DR GIBSONIA PA 15044	1850000.0	40.601984	-79.994453

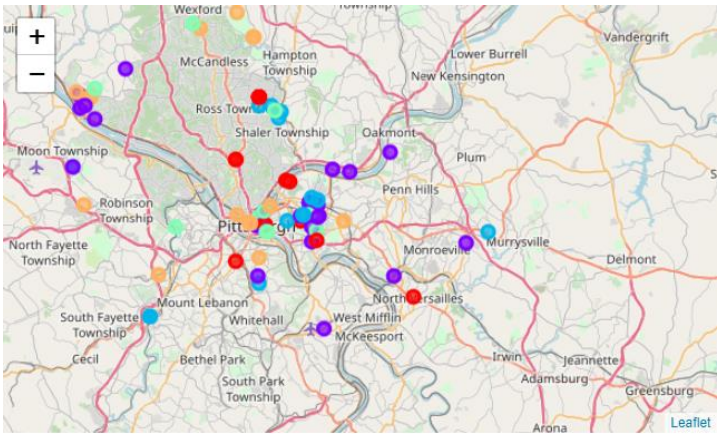
C. Data Preparation and Preprocessing

According to the total number of venues in each category, I applied hierarchical agglomerative clustering method to compare neighborhoods among cities. Unlike kmeans or other machine learning clustering methods, hierarchical agglomerative clustering doesn't require the number of clusters at the beginning. Moreover, it could also tell us whether the dataset is good for clustering at first glance. To gain deeper insight from the dataset I decide to separate the neighborhoods into 5 clusters by cutting at the distance of 31. Secondly, I applied hierarchical agglomerative clustering of the sklearn.cluster library to cluster these neighborhoods.



D. Modeling

At the end of this project, producing a descriptive map and models and their relationships between various types of information that are to be stored in a database are made.



Cluster 1 Venues Based on Price

	Price	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
74	1950000.0	Park	Pizza Place	Donut Shop	Fish & Chips Shop	Financial or Legal Service	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Exhibit	Event Space
857	2077032.0	Pawn Shop	American Restaurant	Gym	Video Store	Pharmacy	Flower Shop	Dance Studio	Dry Cleaner	Financial or Legal Service	Fast Food Restaurant
3421	2000000.0	Coffee Shop	Food Truck	Plaza	Performing Arts Venue	Hotel	Indian Restaurant	College Cafeteria	College Academic Building	Sculpture Garden	Vegetarian / Vegan Restaurant
5148	2030000.0	Bar	Thrift / Vintage Store	Brewery	Garden Center	Bakery	Juice Bar	Tattoo Parlor	Tapas Restaurant	Flower Shop	Bookstore
6241	1900000.0	Bar	Gay Bar	Vegetarian / Vegan Restaurant	Tattoo Parlor	Bakery	Building	Thrift / Vintage Store	River	Discount Store	Pizza Place

Cluster 2 Venues Based on Price

	Price	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
79	1500000.0	Coffee Shop	Grocery Store	Gym / Fitness Center	Hotel	American Restaurant	Salon / Barbershop	Liquor Store	Plaza	Pharmacy	Pet Store
1364	1597471.0	Coffee Shop	Mexican Restaurant	Pizza Place	Bar	Chinese Restaurant	Ice Cream Shop	Noodle House	American Restaurant	French Restaurant	Café
1612	1500000.0	Brewery	Art Gallery	Zoo	Dumpling Restaurant	Fish Market	Fish & Chips Shop	Financial or Legal Service	Fast Food Restaurant	Farmers Market	Falafel Restaurant
2824	1550000.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3866	1597471.0	Intersection	Park	Bus Station	Coffee Shop	Art Gallery	Dumpling Restaurant	Fish & Chips Shop	Financial or Legal Service	Fast Food Restaurant	Farmers Market

Cluster 3 Venues Based on Price

	Price	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
262	2400000.0	Supermarket	Pharmacy	Salon / Barbershop	Gym / Fitness Center	Bank	Gas Station	Dry Cleaner	Financial or Legal Service	Fast Food Restaurant	Farmers Market
949	2400000.0	Liquor Store	Beer Store	Garden Center	Bank	Electronics Store	Fish Market	Fish & Chips Shop	Financial or Legal Service	Fast Food Restaurant	Farmers Market
1193	2400000.0	Coffee Shop	Mexican Restaurant	Pizza Place	Noodle House	Grocery Store	Restaurant	American Restaurant	Ice Cream Shop	Gym / Fitness Center	French Restaurant
1455	2400000.0	Café	Thai Restaurant	Dry Cleaner	Fish & Chips Shop	Financial or Legal Service	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Exhibit	Event Space
1961	2462500.0	Gym / Fitness Center	Gym	Zoo	Dry Cleaner	Fish & Chips Shop	Financial or Legal Service	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Exhibit

Cluster 4 Venues Based on Price

	Price	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
215	2200000.0	Brewery	Steakhouse	Zoo	Dumpling Restaurant	Fish & Chips Shop	Financial or Legal Service	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Exhibit
1436	2150000.0	Women's Store	Rental Car Location	Zoo	Donut Shop	Fish & Chips Shop	Financial or Legal Service	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Exhibit
10255	2125000.0	Bar	Bakery	Coffee Shop	Gay Bar	Grocery Store	Sandwich Place	Diner	Art Gallery	Pizza Place	Gourmet Shop
11888	2200000.0	Gas Station	Grocery Store	Zoo	Dry Cleaner	Fish & Chips Shop	Financial or Legal Service	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Exhibit
15664	2200000.0	Coffee Shop	Bar	Bakery	Grocery Store	Seafood Restaurant	Boat or Ferry	Italian Restaurant	Pizza Place	Gourmet Shop	Diner

Cluster 5 Venues Based on Price

	Price	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
811	1850000.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5620	1800000.0	Hotel	American Restaurant	Theater	Coffee Shop	Bar	Restaurant	Deli / Bodega	Gym	Plaza	Mexican Restaurant
6688	1800000.0	Fast Food Restaurant	Liquor Store	Pizza Place	Auto Workshop	Mobile Phone Shop	Shoe Store	Sushi Restaurant	Food & Drink Shop	Dry Cleaner	Optical Shop
12081	1850000.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
13210	1750000.0	Hotel	Theater	Coffee Shop	Italian Restaurant	Plaza	Bar	American Restaurant	Restaurant	Mexican Restaurant	Deli / Bodega

Analytical Results

Business Solution:

Pittsburgh has become a world-class city; home prices are still lower than the US national average. The median home value in Pittsburgh is $164,317. The median list price per square foot in Pittsburgh is $144, which is higher than the Pittsburgh Metro average of $125. The median price of homes currently listed in Pittsburgh is $224,900 while the median price of homes that sold is $187,400. The median rent price in Pittsburgh is $1,450, which is higher than the Pittsburgh Metro median of $1,225. Machine learning tools provide essential insights to homebuyers and businesses to assist them with decision making based on their needs. This project solves the business problem with clustering 5 neighborhoods, recommending all sorts of venues surrounding these spots in Pittsburgh, PA. We recommended profitable venues according to amenities and essential facilities surrounding such venues i.e. elementary schools, high schools, hospitals & grocery stores.

In this process first we gathered data from Allegheny County (Pennsylvania) and the City of Pittsburgh which gathered their data through the Western Pennsylvania Regional Data Center. after data collection, we analyzed the data to gather statistical information on Pittsburgh housing data. Then we used geolocation library to import longitude and latitude of the properties and used Foursquare API interface to collect information on venues based on the location of the property. By shaping this comprehensive dataset, we are able to provide profitable insight on the location desired houses in Pittsburgh's area. Unsupervised machine learning, k-mean clustering method utilized to cluster 5 neighborhoods that have similar characteristics in the city of Pittsburgh.

In this project we conclude that clusters that have natural amenities are in high demand than the ones that are surrounded with Plazas including restaurants and

shopping centers. Cluster 2 and 4 are evident for this fact. In this study we also see that cluster 0 in East Pittsburgh is the best area for profitable purchases as it shows similar characteristics in cluster 2 and 4. One might target underpriced investments in these areas of Pittsburgh in order to make a business affair.