City in North America

A Clustering Analysis

1 Introduction

1.1 Background

As the evolvement of globalization, it becomes more and more common for people to relocate among cities and countries.

One major part of people's life is eating and entertaining. It is crucially important for people to have access to places they enjoy. Also, the weather and culture differs in different cities are important factors as well. It is helpful for people interested in moving to have a tool to compare the similarities among cities.

Therefore, I am going to use data like venues types, restaurant quantity, climate data, demographics, etc. And create a tool to directly compare cities and cluster similar into groups. Hope it would be helpful for those people.

Particularly, I am going to cluster cities in North America into several groups.

1.2 Stakeholder/Target Audience

The target audience is those who are currently searching for new cities to moving. Or someone wants to compare the living environment between different cities/neighborhoods. Or someone who is curious about similarities among different cities.

2 Data

2.1 Foursquare Data
For foursquare data, we just use 'Search API' to get the recommend venues for each specific city by using the city's geolocation.

	City	City Latitude	City Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	New York City	40.712728	-74.006015	The Bar Room at Temple Court	40.711448	-74.006802	Hotel Bar
1	New York City	40.712728	-74.006015	Four Seasons Hotel New York Downtown	40.712612	-74.009380	Hotel
2	New York City	40.712728	-74.006015	Korin	40.714824	-74.009404	Furniture / Home Store
3	New York City	40.712728	-74.006015	Aire Ancient Baths	40.718141	-74.004941	Spa
4	New York City	40.712728	-74.006015	9/11 Memorial North Pool	40.712077	-74.013187	Memorial Site
5	New York City	40.712728	-74.006015	One World Trade Center	40.713069	-74.013133	Building
6	New York City	40.712728	-74.006015	Washington Market Park	40.717046	-74.011095	Playground
7	New York City	40.712728	-74.006015	Crown Shy	40.706187	-74.007490	Restaurant
8	New York City	40.712728	-74.006015	Liberty Park	40.710384	-74.013868	Park
9	New York City	40.712728	-74.006015	sweetgreen	40.705586	-74.008382	Salad Place
10	New York City	40.712728	-74.006015	The Rooftop @ Pier 17	40.705463	-74.001598	Music Venue
11	New York City	40.712728	-74.006015	Battery Park City Esplanade	40.711622	-74.017907	Park
12	New York City	40.712728	-74.006015	Pier 25 - Hudson River Park	40.720193	-74.012950	Park
13	New York City	40.712728	-74.006015	Nelson A. Rockefeller Park	40.717095	-74.016716	Park
14	New York City	40.712728	-74.006015	Brooklyn Bridge	40.705967	-73.996707	Bridge
15	New York City	40.712728	-74.006015	La Compagnie des Vins Surnaturels	40.720448	-73.997969	Wine Bar
16	New York City	40.712728	-74.006015	Pier 25 Beach Volleyball	40.720380	-74.014860	Volleyball Court
17	New York City	40.712728	-74.006015	Metrograph	40.714999	-73.991035	Indie Movie Theater
18	New York City	40.712728	-74.006015	Stick With Me	40.721304	-73.995474	Chocolate Shop
19	New York City	40.712728	-74.006015	CAVA	40.721928	-73.996512	Mediterranean Restaurant
20	New York City	40.712728	-74.006015	Brooklyn Bridge Park	40.702282	-73.996456	Park
21	New York City	40.712728	-74.006015	Blacktail	40.704537	-74.017281	Cocktail Bar
22	New York City	40.712728	-74.006015	Brooklyn Bridge Park - Pier 1	40.702900	-73.995987	Park
23	New York City	40.712728	-74.006015	Elizabeth Street Garden	40.722138	-73.994744	Garden
24	New York City	40.712728	-74.006015	Rubirosa Ristorante	40.722706	-73.995957	Italian Restaurant
25	New York City	40.712728	-74.006015	Trader Joe's	40.725611	-74.004985	Grocery Store
26	New York City	40.712728	-74.006015	McNally Jackson Books	40.723493	-73.996048	Bookstore
27	New York City	40.712728	-74.006015	Pebble Beach	40.704329	-73.990265	Beach
28	New York City	40.712728	-74.006015	Brooklyn Bridge Park - Pier 2	40.699570	-73.997933	Park
29	New York City	40.712728	-74.006015	New York Pilates - Bowery	40.723310	-73.993146	Pilates Studio
...	...	...	...	...	...	...	...
1170	Houston	29.758938	-95.367697	Buffalo Bayou Loop	29.761345	-95.401556	Trail
1171	Houston	29.758938	-95.367697	Salata	29.773528	-95.397887	Salad Place
1172	Houston	29.758938	-95.367697	The Hay Merchant	29.744825	-95.392782	Beer Garden

2.2 Climate Data

For Climate Data, we scrape from Wikipedia and extract the table which contains the data we need. Then we Use the data that are available for all cities. Which are Average high, Average Low, Average precipitation days, Mean monthly sunshine hours.

Climate data for New York (Belvedere Castle, Central Park), 1981–2010 normals,[a] extremes 1869–present[b]													[hide]
Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Year
Record high °F (°C)	72 (22)	78 (26)	86 (30)	96 (36)	99 (37)	101 (38)	106 (41)	104 (40)	102 (39)	94 (34)	84 (29)	75 (24)	106 (41)
Mean maximum °F (°C)	59.6 (15.3)	60.7 (15.9)	71.5 (21.9)	83.0 (28.3)	88.0 (31.1)	92.3 (33.5)	95.4 (35.2)	93.7 (34.3)	88.5 (31.4)	78.8 (26.0)	71.3 (21.8)	62.2 (16.8)	97.0 (36.1)
Average high °F (°C)	38.3 (3.5)	41.6 (5.3)	49.7 (9.8)	61.2 (16.2)	70.8 (21.6)	79.3 (26.3)	84.1 (28.9)	82.6 (28.1)	75.2 (24.0)	63.8 (17.7)	53.8 (12.1)	43.0 (6.1)	62.0 (16.7)
Average low °F (°C)	26.9 (−2.8)	28.9 (−1.7)	35.2 (1.8)	44.8 (7.1)	54.0 (12.2)	63.6 (17.6)	68.8 (20.4)	67.8 (19.9)	60.8 (16.0)	50.0 (10.0)	41.6 (5.3)	32.0 (0.0)	48.0 (8.9)
Mean minimum °F (°C)	9.2 (−12.7)	12.8 (−10.7)	18.5 (−7.5)	32.3 (0.2)	43.5 (6.4)	52.9 (11.6)	60.3 (15.7)	58.8 (14.9)	48.6 (9.2)	38.0 (3.3)	27.7 (−2.4)	15.6 (−9.1)	7.0 (−13.9)
Record low °F (°C)	−6 (−21)	−15 (−26)	3 (−16)	12 (−11)	32 (0)	44 (7)	52 (11)	50 (10)	39 (4)	28 (−2)	5 (−15)	−13 (−25)	−15 (−26)
Average precipitation inches (mm)	3.65 (93)	3.09 (78)	4.36 (111)	4.50 (114)	4.19 (106)	4.41 (112)	4.60 (117)	4.44 (113)	4.28 (109)	4.40 (112)	4.02 (102)	4.00 (102)	49.94 (1,268)
Average snowfall inches (cm)	7.0 (18)	9.2 (23)	3.9 (9.9)	0.6 (1.5)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0.3 (0.76)	4.8 (12)	25.8 (66)
Average precipitation days (≥ 0.01 in)	10.4	9.2	10.9	11.5	11.1	11.2	10.4	9.5	8.7	8.9	9.6	10.6	122.0
Average snowy days (≥ 0.1 in)	4.0	2.8	1.8	0.3	0	0	0	0	0	0	0.2	2.3	11.4
Average relative humidity (%)	61.5	60.2	58.5	55.3	62.7	65.2	64.2	66.0	67.8	65.6	64.6	64.1	63.0
Mean monthly sunshine hours	162.7	163.1	212.5	225.6	256.6	257.3	268.2	268.2	219.3	211.2	151.0	139.0	2,534.7
Percent possible sunshine	54	55	57	57	57	57	59	63	59	61	51	48	57
Average ultraviolet index	2	3	4	6	7	8	8	8	6	4	2	1	5
Source #1: NOAA (relative humidity and sun 1961–1990)[237][249][233][250]													
Source #2: Weather Atlas[251]													
See Geography of New York City for additional climate information from the outer boroughs.													

2.3 Demographic Data

For demographic data, since it's unstructured data, we have to manually fill a table to represent each cities' race demography.

Demographics

Main articles: Demographics of New York City, New York City ethnic enclaves, and Demographic profile of New York City

Racial composition [show]   2010[270]   1990[272]   1970[272]   1940[272] | City compared to State & U.S. [show]

New York City is the most populous city in the United States,[283] with an estimated 8,398,748 residents as of July 2018,[6] incorporating more immigration into the city than outmigration since the 2010 United States Census.[284][285] More than twice as many people live in New York City as compared to Los Angeles, the second-most populous U.S. city,[283] and within a smaller area. New York City gained more residents between April 2010 and July 2014 (316,000) than any other U.S. city.[283] New York City's population is about 43% of New York State's population[286] and about 36% of the population of the New York metropolitan area.[287]

Population density

In 2017, the city had an estimated population density of 28,491 inhabitants per square mile (11,000/km²), rendering it the most densely populated of all municipalities housing over 100,000 residents in the United States, with several small cities (of fewer than 100,000) in adjacent Hudson County, New Jersey having greater density, as per the 2010 Census.[288] Geographically co-extensive with New York County, the borough of Manhattan's 2017 population density of 72,918 inhabitants per square mile (28,154/km²)[289] makes it the highest of any county in the United States[290][291] and higher than the density of any individual American city.[292]

Race and ethnicity

Further information: Category:Ethnic groups in New York City, Bangladeshis in New York City, Caribbeans in New York City, Chinese in New York City, Filipinos in New York City, Fuzhounese in New York City, Indians in New York City, Irish in New York City, Italians in New York City, Japanese in New York City, Koreans in New York City, Puerto Ricans in New York City, Russians in New York City, and Ukrainians in New York City

The city's population in 2010 was 44% white (33.3% non-Hispanic white), 25.5% black (23% non-Hispanic black), 0.7% Native American, and 12.7% Asian.[293] Hispanics of any race represented 28.6% of the population,[295] while Asians constituted the fastest-growing segment of the city's population between 2000 and 2010; the non-Hispanic white population declined 3 percent, the smallest recorded decline in decades; and for the first time since the Civil War, the number of blacks declined over a decade.[294] Throughout its history, New York has been a major port of entry for immigrants into the United States. More than 12 million European immigrants were received at Ellis Island between 1892 and 1924.[295] The term "melting pot" was first coined to describe densely populated immigrant neighborhoods on the Lower East Side. By 1900, Germans constituted the largest immigrant group, followed by the Irish, Jews, and Italians.[296] In 1940, whites represented 92% of the city's population.[272]

Approximately 37% of the city's population is foreign born, and more than half of all children are born to mothers who are immigrants.[297][298] In New York, no single country or region of origin dominates.[297] The ten largest sources of foreign-born individuals in the city as of 2011 were the Dominican Republic, China, Mexico, Guyana, Jamaica, Ecuador, Haiti, India, Russia, and Trinidad and Tobago,[299] while the Bangladeshi-born immigrant population has become one of the fastest growing in the city, counting over 74,000 by 2011.[42][300]

Asian Americans in New York City, according to the 2010 Census, number more than one million, greater than the combined totals of San Francisco and Los Angeles.[301] New York contains the highest total Asian population of any U.S. city proper.[302] The New York City borough of Queens is home to the state's largest Asian American population and the largest Andean (Colombian, Ecuadorian, Peruvian, and Bolivian) populations in the United States, and is also the most ethnically diverse urban area in the world.[204][205]

The Chinese population constitutes the fastest-growing nationality in New York State; multiple satellites of the original Manhattan Chinatown, in Brooklyn, and around Flushing, Queens, are thriving as traditionally urban enclaves – while also expanding rapidly eastward into suburban Nassau County[303] on Long Island,[304] as the New York metropolitan region and New York State have become the top destinations for new Chinese immigrants, respectively, and large-scale Chinese immigration continues into New York City and surrounding areas.[35][305][306][307][308][309] with the largest metropolitan Chinese diaspora outside Asia,[42][310] including an estimated 812,410 individuals in 2015.[311]
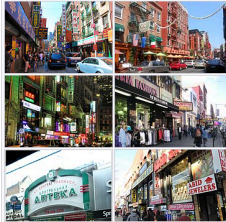
In 2012, 6.3% of New York City was of Chinese ethnicity, with nearly three-fourths living in either Queens or Brooklyn, geographically on Long Island.[312] A community numbering 20,000 Korean-Chinese (Chaoxianzu or Joseonjok) is centered in Flushing, Queens, while New York City is also home to the largest Tibetan population outside China, India, and Nepal, also centered in Queens.[313] Koreans made up 1.2% of the city's population, and Japanese 0.3%. Filipinos were the largest Southeast Asian ethnic group at 0.8%, followed by Vietnamese, who made up 0.2% of New York City's population in 2010. Indians are the largest South Asian group, comprising 2.4% of the city's population, with Bangladeshis and Pakistanis at 0.7% and 0.5%, respectively.[314] Queens is the preferred borough of settlement for Asian Indians, Koreans, Filipinos,[315] and Malaysians[35] and other Southeast Asians;[316] while Brooklyn is receiving large numbers of both West Indian and Asian Indian immigrants.

New York City has the largest European and non-Hispanic white population of any American city. At 2.7 million in 2012, New York's non-Hispanic white population is larger than the non-Hispanic white populations of Los Angeles (1.1 million), Chicago (865,000), and Houston (550,000) combined.[317] The non-Hispanic white population was 6.6 million in 1940.[318] The non-Hispanic white population has begun to increase since 2010.[319]

The European diaspora residing in the city is very diverse. According to 2012 Census estimates, there were roughly 560,000 Italian Americans, 385,000 Irish Americans, 253,000 German Americans, 223,000 Russian Americans, 201,000 Polish Americans, and 137,000 English Americans. Additionally, Greek and French Americans numbered 65,000 each, with those of Hungarian descent estimated at 60,000 people. Ukrainian and Scottish Americans numbered 55,000 and 35,000, respectively. People identifying ancestry from Spain numbered 30,838 total in 2010.[300]

People of Norwegian and Swedish descent both stood at about 20,000 each, while people of Czech, Lithuanian, Portuguese, Scotch-Irish, and Welsh descent all numbered between 12,000–14,000 people.[321] Arab Americans number over 160,000 in New York City,[322] with the highest concentration in Brooklyn. Central Asians, primarily Uzbek Americans, are a rapidly growing segment of the city's non-Hispanic white population, enumerating over 30,000, and including over half of all Central Asian immigrants to the United States,[323] most settling in Queens or Brooklyn. Albanian Americans are most highly concentrated in the Bronx.[324]

The wider New York City metropolitan statistical area, with over 20 million people, about 50% greater than the second-place Los Angeles metropolitan area in the United States,[3] is also ethnically diverse,[325] with the largest foreign-born population of any metropolitan region in the world. The New York region continues to be by far the leading metropolitan gateway for legal immigrants admitted into the United States, substantially exceeding the combined totals of Los Angeles and Miami.[35] It is home to the largest Jewish and Israeli communities outside Israel, with the Jewish population in the region numbering over 1.5 million in 2012 and including many diverse Jewish sects, predominantly from around the Middle East and Eastern Europe, and including a rapidly growing Orthodox Jewish population, the largest outside Israel.[313]

The metropolitan area is also home to 20% of the USA's Indian Americans and at least 20 Little India enclaves, and 15% of all Korean Americans and four Koreatowns;[326][327] the largest Asian Indian population in the Western Hemisphere; the largest Russian American,[305] Italian American, and African American populations; the largest Dominican American, Puerto Rican American, and South American[305] and second-largest overall Hispanic population in the United States, numbering 4.8 million;[320] and includes multiple established Chinatowns within New York City alone.[328]

Ecuador, Colombia, Guyana, Peru, and Brazil were the top source countries from South America for legal immigrants to the New York City region in 2013; the Dominican Republic, Jamaica, Haiti, and Trinidad and Tobago in the Caribbean; Egypt, Ghana, and Nigeria from Africa; and El Salvador, Honduras, and Guatemala in Central America.[329] Amidst a resurgence of Puerto Rican migration to New York City, this population had increased to approximately 1.3 million in the metropolitan area as of 2013.

Since 2010, a Little Australia has emerged and is growing rapidly representing the Australasian presence in Nolita, Manhattan.[330][331][332][333] In 2011, there were an estimated 20,000 Australian residents of New York City, nearly quadruple the 5,537 in 2005.[334][335] Qantas Airways of Australia and Air New Zealand have been exploring the possibilities of long-haul flights from New York to Sydney and Auckland, respectively, which would both rank among the longest non-stop flights in the world.[336][337]

Clockwise, from upper left: the Manhattan Chinatown; Lower Manhattan's Little Italy; Upper Manhattan's Spanish Harlem; Little India, Queens; Brooklyn's Little Russia; Midtown Manhattan's Koreatown

2.4 Geolocation Data

And we use Geolocation API to get each cities' geolocation data.

	latitude	longitude
New York City	40.712728	-74.006015
Toronto	43.653963	-79.387207
Vancouver	49.260872	-123.113953
Boston	42.360253	-71.058291
Montreal	45.497216	-73.610364
San Francisco	37.779281	-122.419236
Seattle	47.603832	-122.330062
Edmonton	53.535411	-113.507996
Calgary	51.025327	-114.049868
Los Angeles	34.053691	-118.242767
Chicago	41.875562	-87.624421
Houston	29.758938	-95.367697

3 Analyze Data Methodology

3.1 Analyze methodology

In general, we first draw the dendrogram of the data to hierarchically represent the relationship between different cities. Then we choose the distance that makes the most sense, then cluster them.
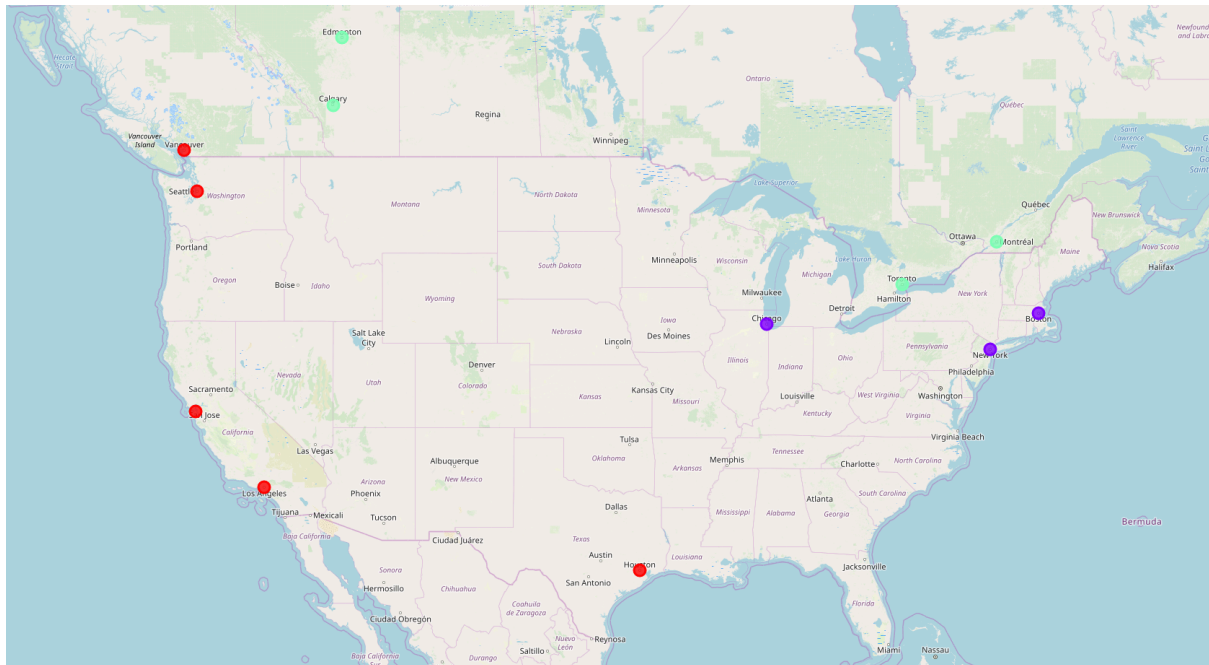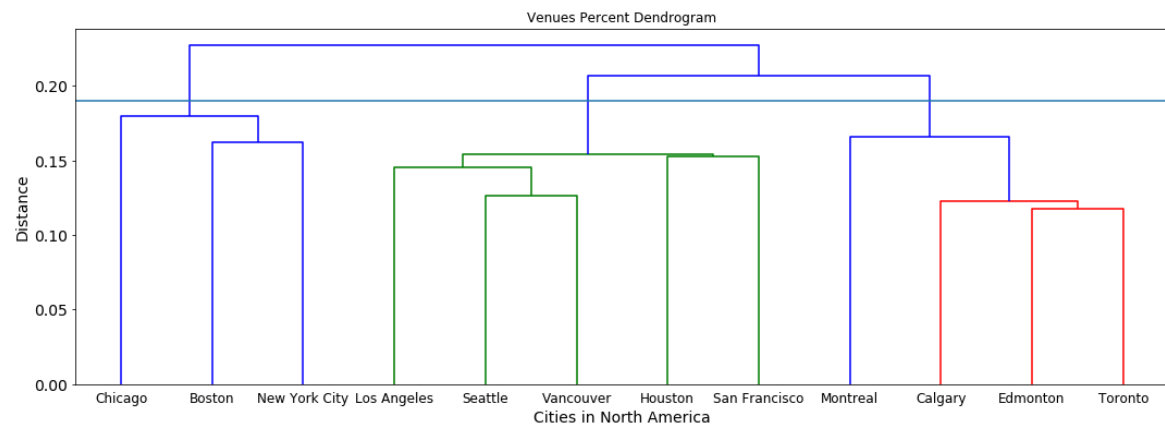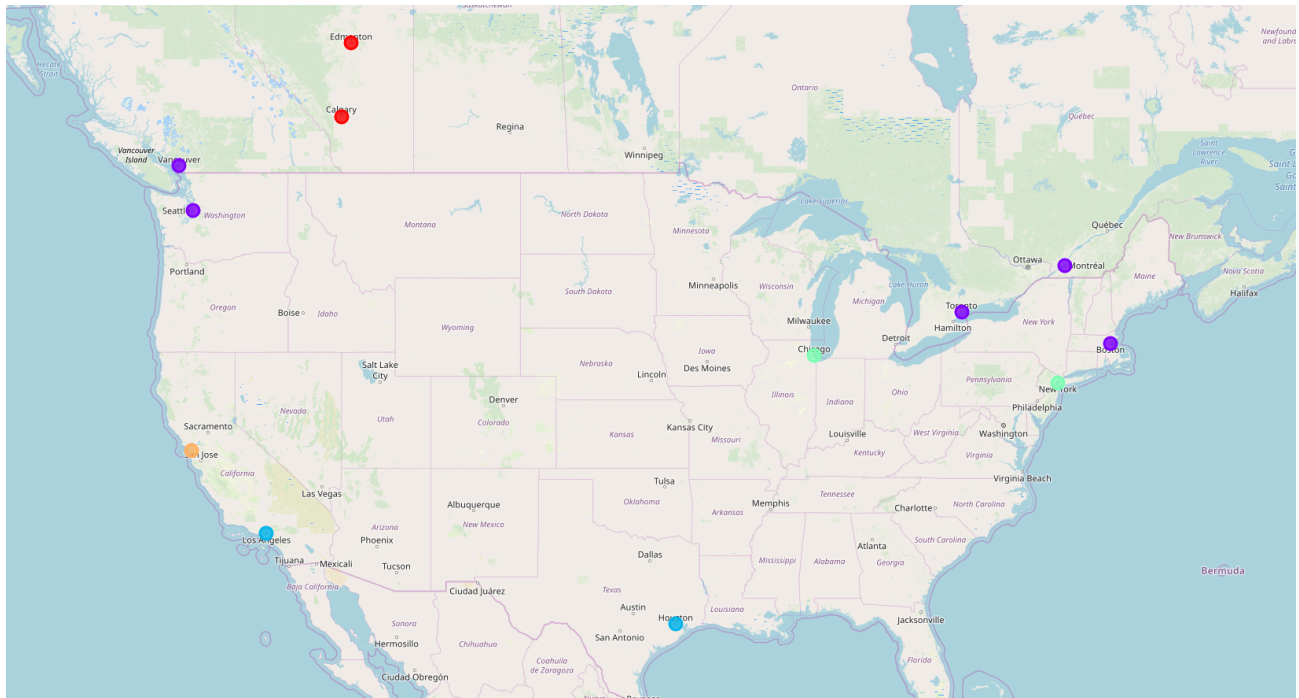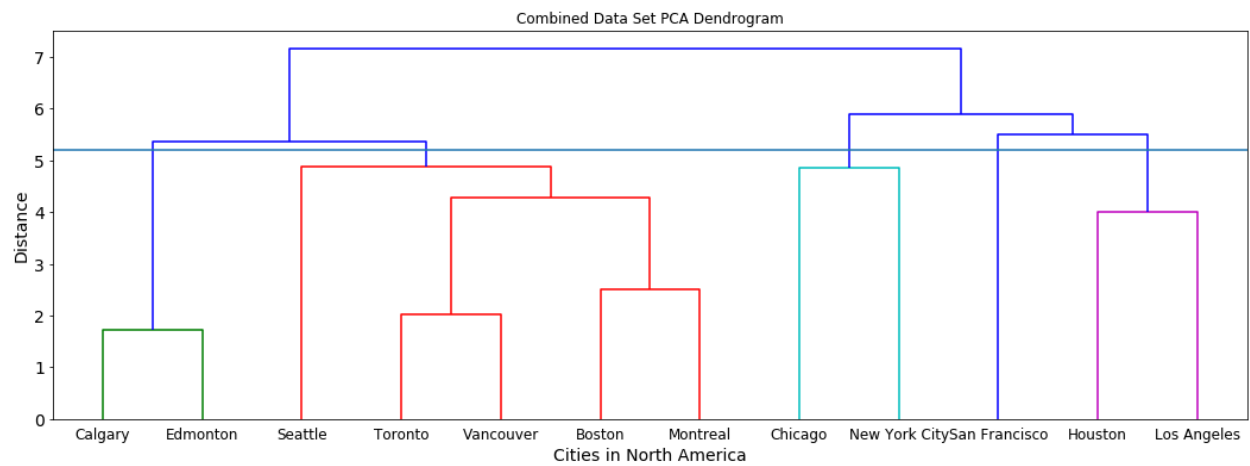
3.2 Climate Data

Climate Dendrogram

3.3 Demographic Data

Demographic Dendrogram

3.4 Venues Data

Venues Percent Dendrogram
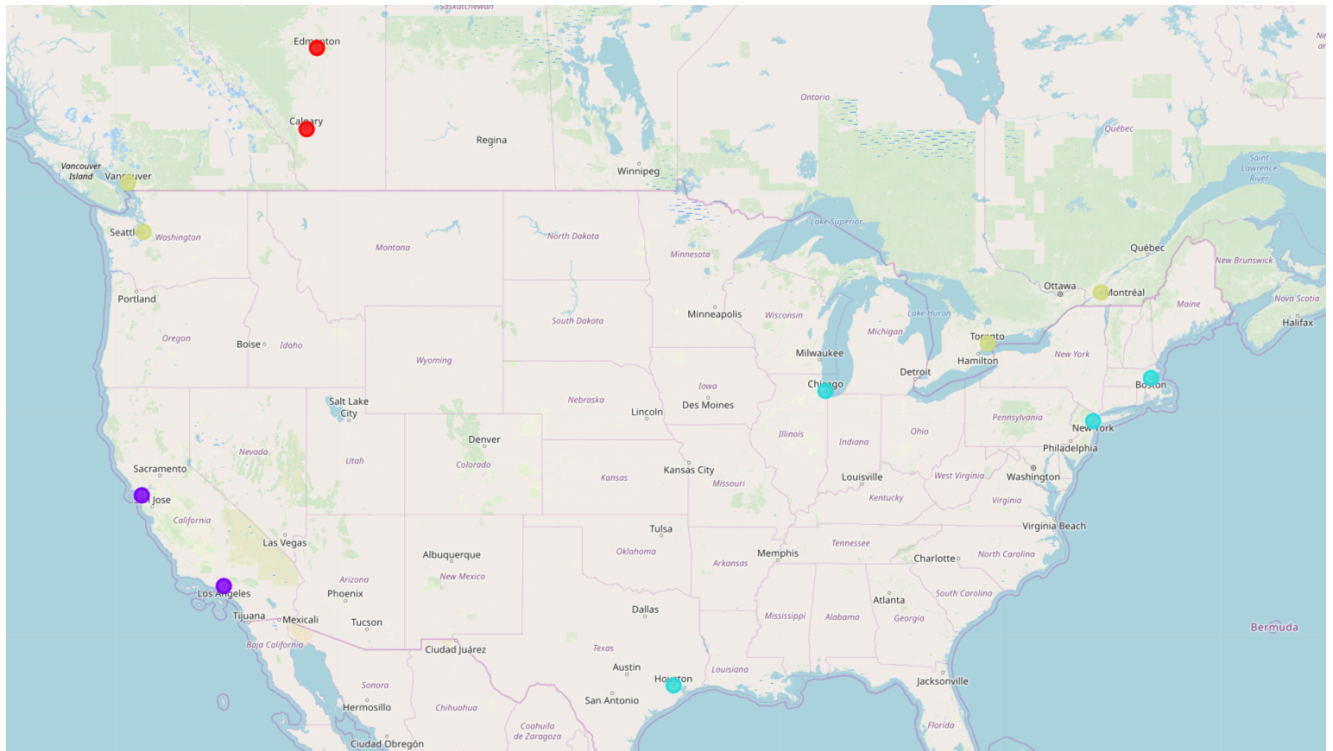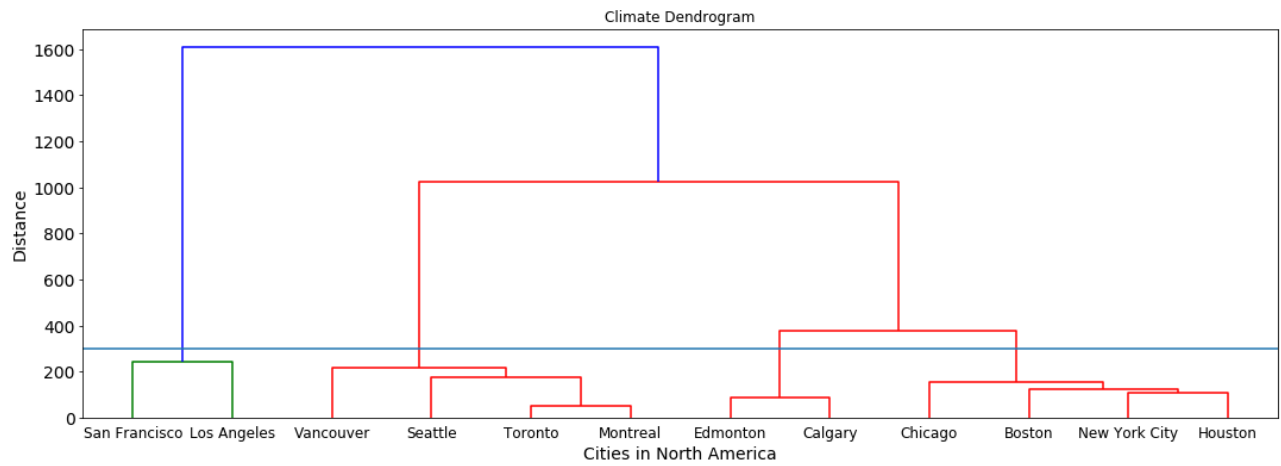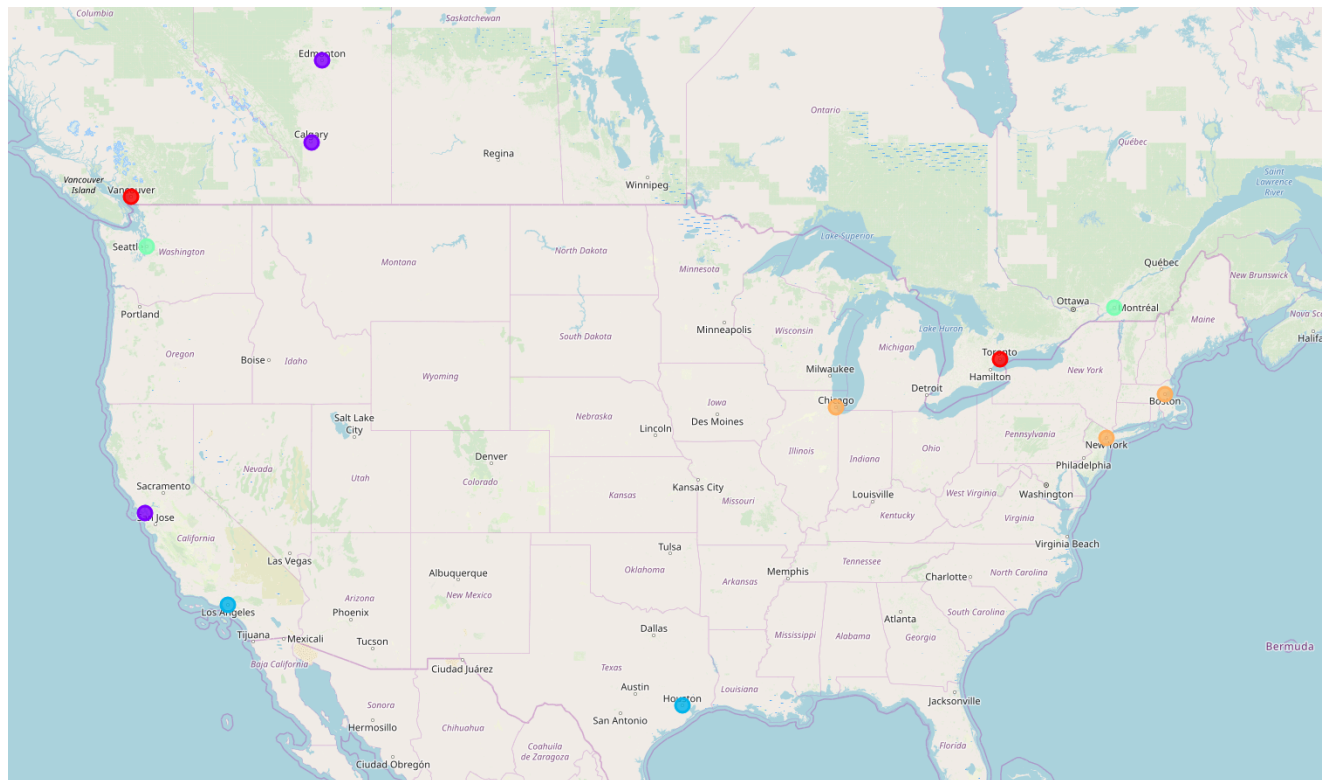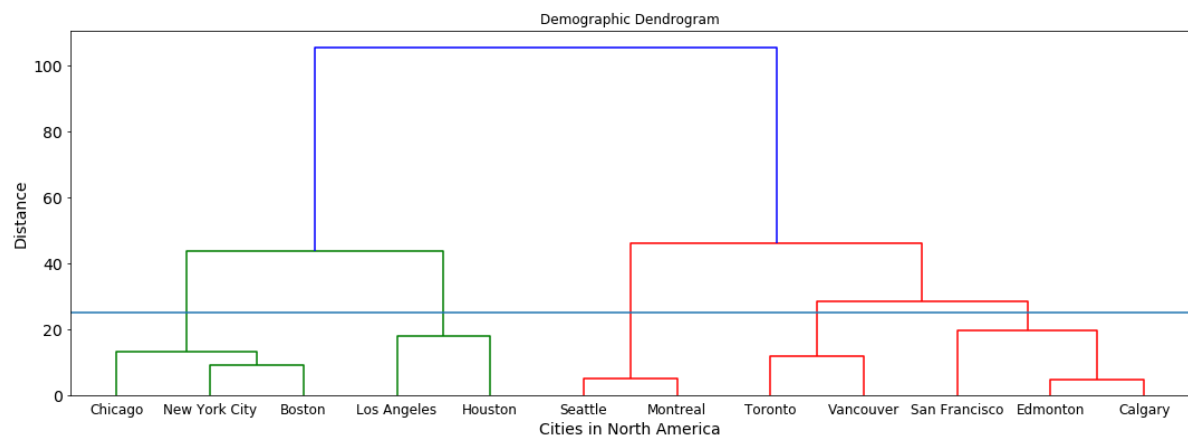
3.5 Combined Data

4 Results & Discussion

The graph clearly shows that the clustering has significant influence from geolocations. And the dendrogram of each data set is very helpful to understand the distance between different cities use different data measurement.

5 Conclusion

Climate, Cultural, Demographic all contribute to the cluster of cities in North America. And cities in North America has pretty diverse Characteristics in consider of those information.