

Comparative Analysis between Hamburg (Germany) and Top 10 USA cities

Applied Data Science Capstone Project

Agenda

- Introduction
- Data
- Methodology
- Analysis
- Conclusion

Introduction

- Immigration trend
- Barriers of immigration
- Business problem
- Data Science to reduce barriers
- Case study: immigration from Hamburg (Germany) to USA

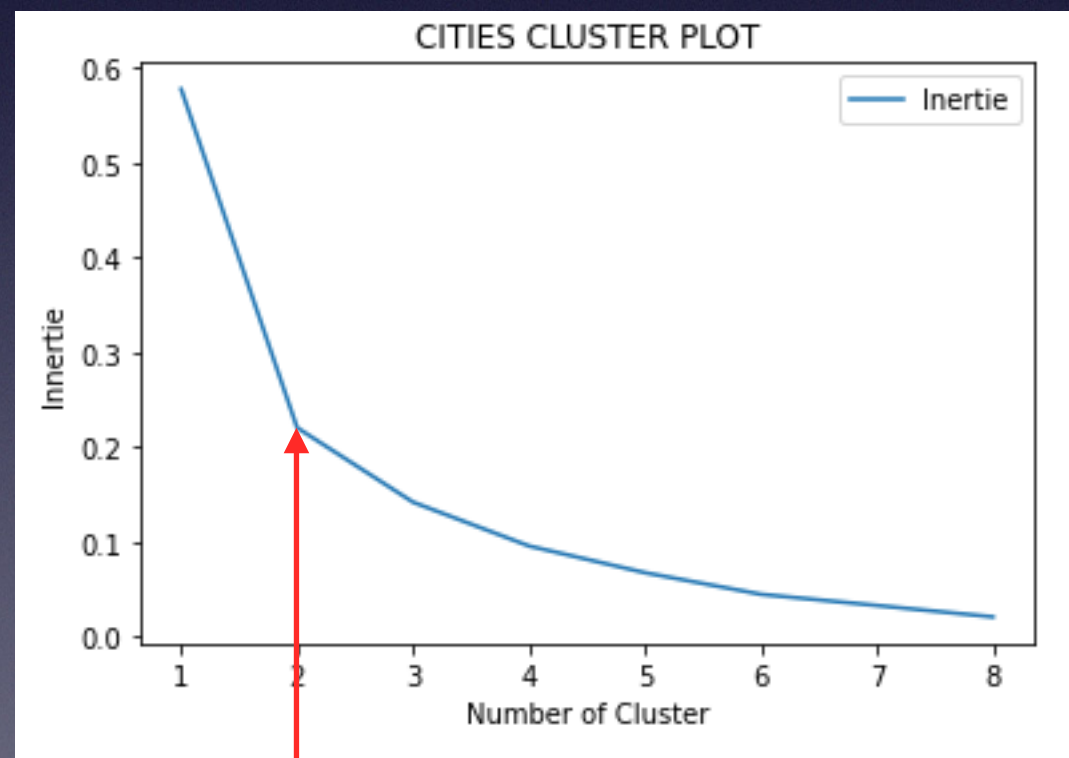
Data

- A list of the cities and they geo coordinates in Wikidata
- Foursquare API provide venues data
- focus on branch
'school','university','hospital','restaurant','coffee','entertainment','shops','playground','nightlife','lodging'

city	cityLabel	population	pa_s	pa_sLabel	coordenadas
Q wd:Q60	New York City	8398748	Q wd:Q30	United States of America	Point(-73.94 40.67)
Q wd:Q65	Los Angeles	3976322	Q wd:Q30	United States of America	Point(-118.24368 34.05223)
Q wd:Q1297	Chicago	2722389	Q wd:Q30	United States of America	Point(-87.627777777 41.881944444)
Q wd:Q16555	Houston	2195914	Q wd:Q30	United States of America	Point(-95.383055555 29.762777777)
Q wd:Q16556	Phoenix	1626078	Q wd:Q30	United States of America	Point(-112.076388888 33.528333333)
Q wd:Q1345	Philadelphia	1580863	Q wd:Q30	United States of America	Point(-75.163611111 39.952777777)
Q wd:Q975	San Antonio	1436697	Q wd:Q30	United States of America	Point(-98.493888888 29.425)
Q wd:Q16552	San Diego	1394928	Q wd:Q30	United States of America	Point(-117.1625 32.715)
Q wd:Q16557	Dallas	1197816	Q wd:Q30	United States of America	Point(-96.808888888 32.779166666)
Q wd:Q16553	San Jose	1025350	Q wd:Q30	United States of America	Point(-121.872777777 37.304166666)

Methodology I

- KMean cluster analysis
- intra-cluster inertia is used for choose optimal K value (elbow point)

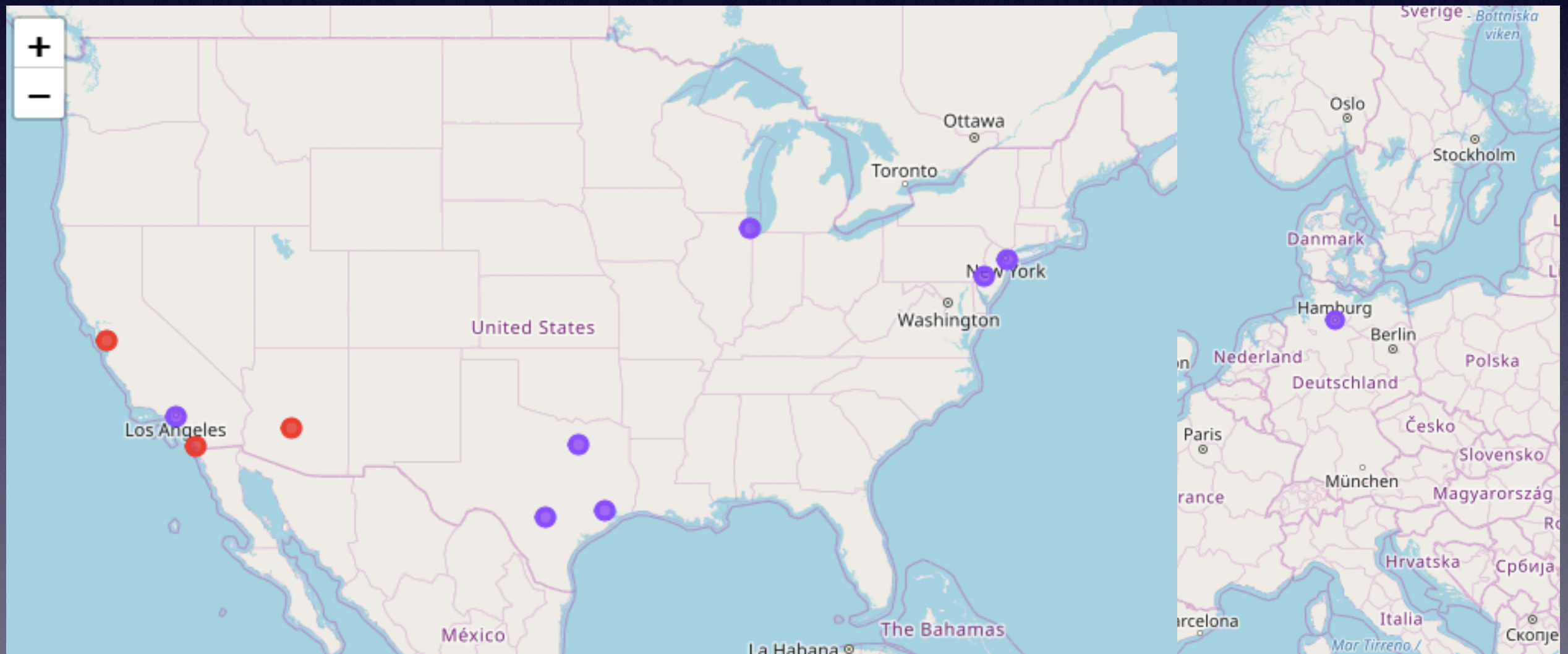


$K = 2$

Analysis

Cluster 1 shop type : *New York City, Los Angeles, Chicago, Houston, Philadelphia, San Antonio, Dallas, Hamburg*

Cluster 2 school type : *Phoenix, San Diego, San Jose*

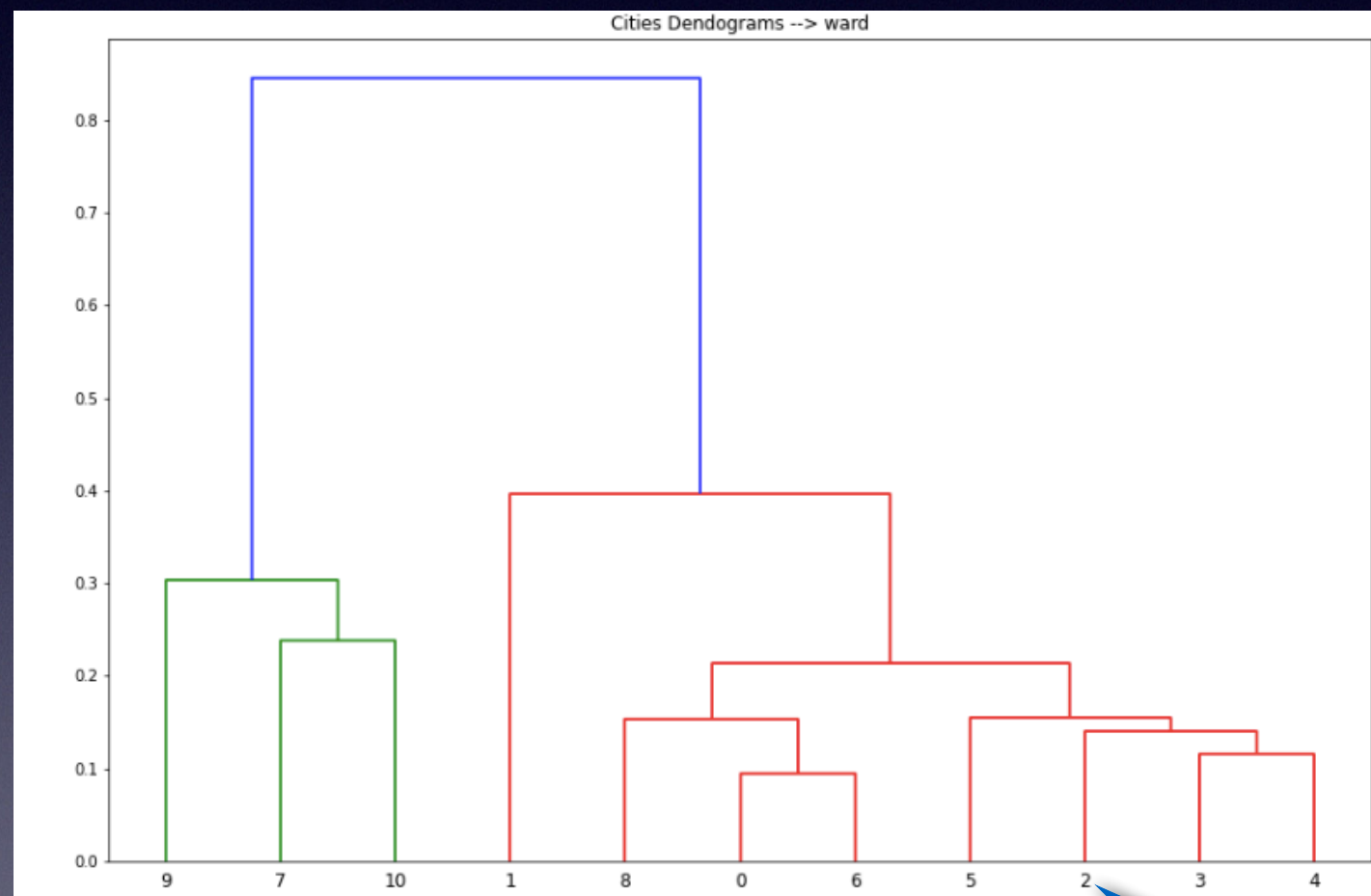


Top 5 Most common venues of each cities

	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Chicago	Shop/Store	Hospital	General Entertainment	Other Food Place	School
1	Dallas	Adult Education Center	Shop/Store	Preschool	Community College	Daycare
2	Hamburg	Shop/Store	Coffee Shop	European Restaurant	Bar	Other Food Place
3	Houston	Shop/Store	Bar	Other Food Place	European Restaurant	General Entertainment
4	Los Angeles	Shop/Store	Bar	General Entertainment	Asian Restaurant	Other Food Place
5	New York City	Shop/Store	School	Coffee Shop	Bar	Other Food Place
6	Philadelphia	Shop/Store	Hospital	School	Other Food Place	Bar
7	Phoenix	School	Elementary School	Shop/Store	Church	School/Education
8	San Antonio	Shop/Store	Hospital	Other Food Place	General Entertainment	European Restaurant
9	San Diego	School	Elementary School	Shop/Store	Language School	Adult Education Center
10	San Jose	Elementary School	School	Church	Preschool	Shop/Store

Methodology II

- Hierarchical cluster analysis to find similar city



2: Hamburg
3: Houston
4: Los Angeles

Conclusion

Combine both method (K-mean and Hierarchical cluster) , which give us precise and same result . That is the most similar as city 2 (Hamburg) is city 3 (Houston) and 4 (Los Angeles). In respect population, Houston has almost the same as Hamburg, LA has more than double than Hamburg. It could be also a reason for people to choose.

Furthermore, we can analyse the difference between Houston, Los Angeles, Hamburg more in detail, which can be studied in the future