

# CITY CLUSTERS

NON-SPATIAL PROXIMITY

Klebert Toscano de S. Cintra  
<https://github.com/KTSC>



# RELEVANCE



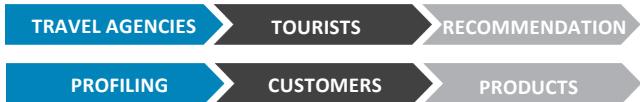
## DECISIONS INCREASE COMPLEXITY WITH THE NUMBER OF VARIABLES

Reducing the number of options can lower the decision making demands, improve satisfaction with the choice made and make all the process faster, which means revenue is more certain and in a shorter time. To illustrate one such application of a classification method this project seeks patterns that can be used to classify entire cities based on its venues.



## RECOMMENDATIONS FOR CUSTOMERS

The project is based on the problem of travelers who seek similar or contrasting destinies when compared to their city of origin. With a few changes it is also applicable for grouping real estate by its multidimensional similarities that are not obvious nor easy to visualize or for recommendation systems in general.



## USE BUSINESS CATEGORIES TO DESCRIBE GEOGRAPHICAL SIMILARITY

We are getting data from the Foursquare Places API.



# DATA SOURCES



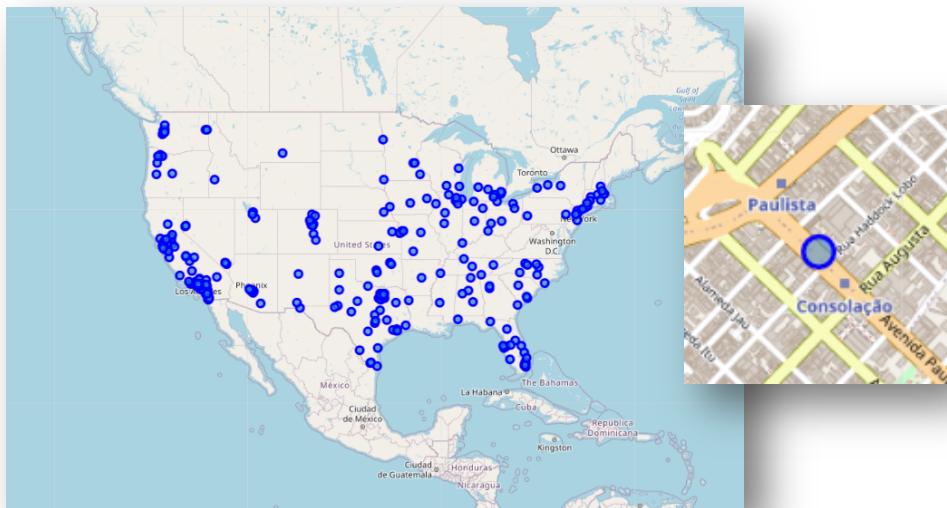
## GEOGRAPHICAL COORDINATES



WIKIPEDIA

NOMINATIM

[https://en.wikipedia.org/wiki/List\\_of\\_United\\_States\\_cities\\_by\\_population](https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population)



## VENUES



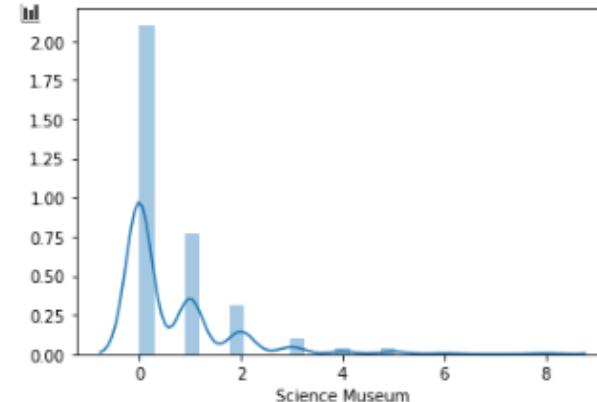
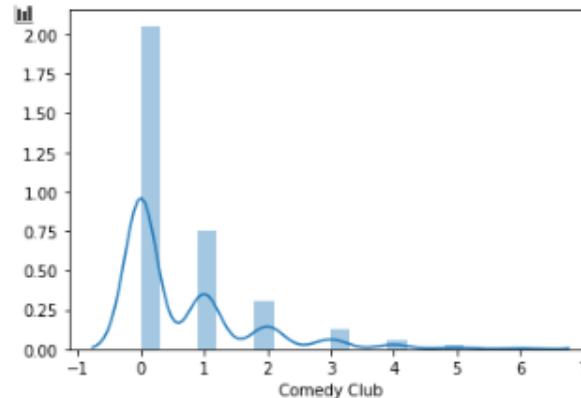
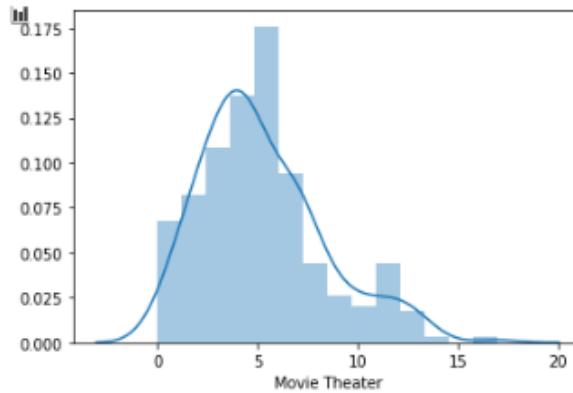
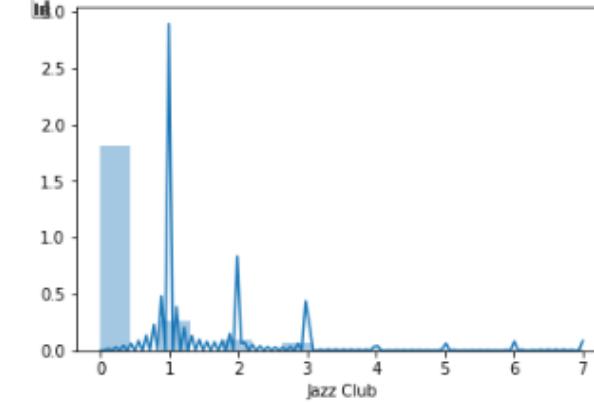
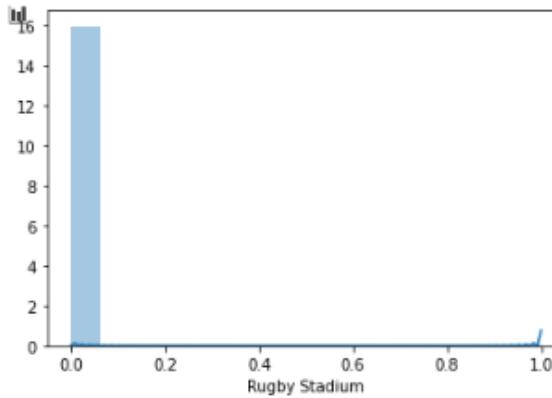
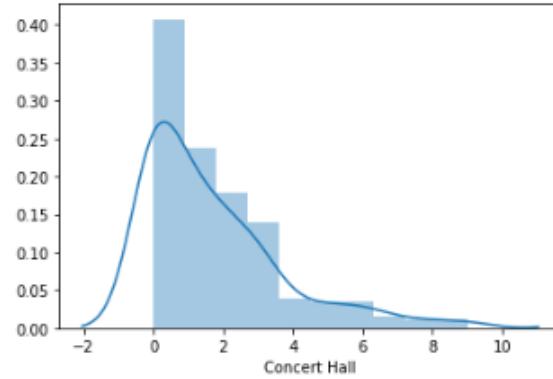
FOURSQUARE



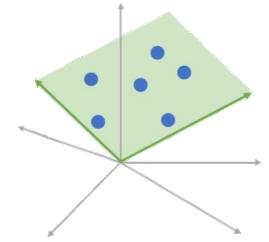
# FEATURES



## VENUES COUNT HISTOGRAM



A 42 dimensions dataset if all different venue categories are considered

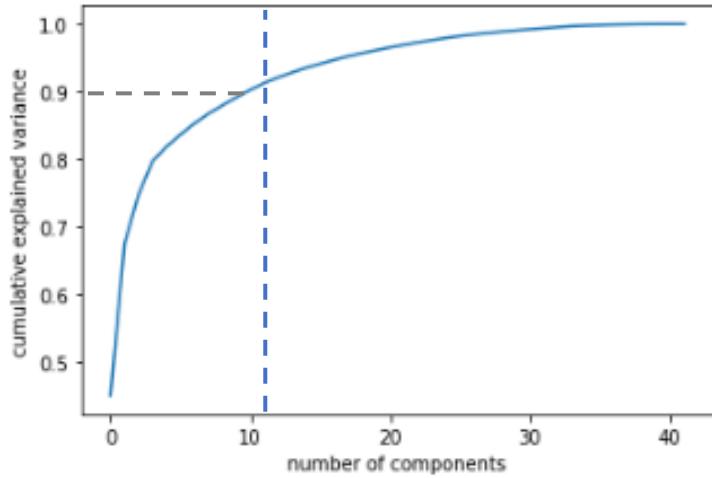


# NEW FEATURES

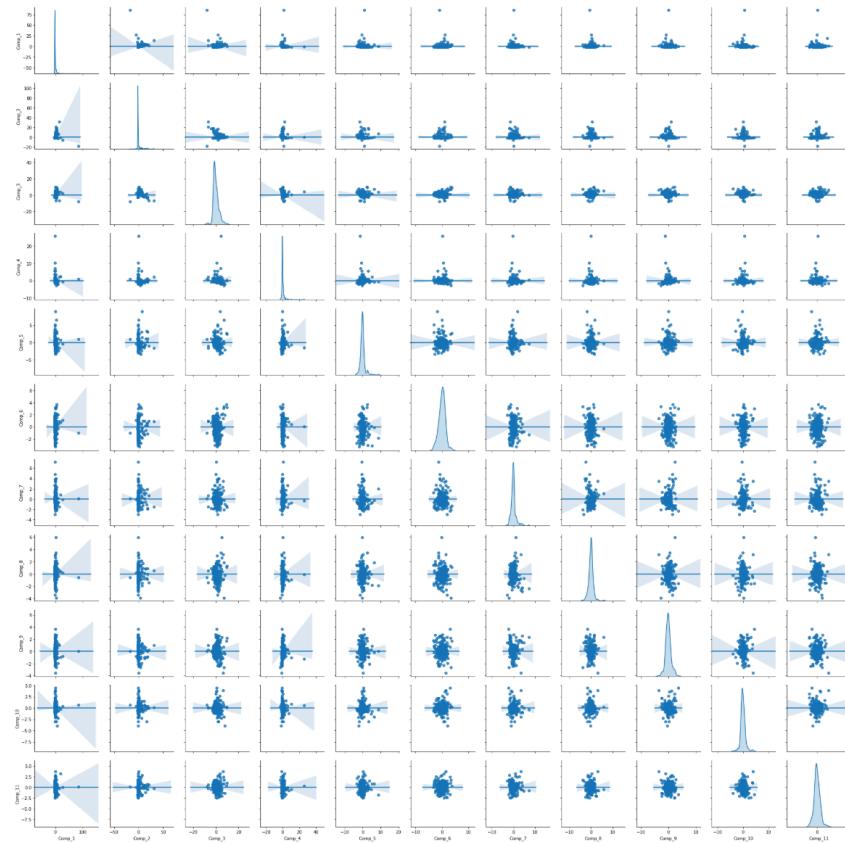


## DIMENSIONALITY REDUCTION WITH PCA

11 DIMENSIONS ARE ENOUGH TO EXPLAIN 90% OF THE VARIABILITY IN DATA



THE DIMENSIONS ARE **ORTHOGONAL**

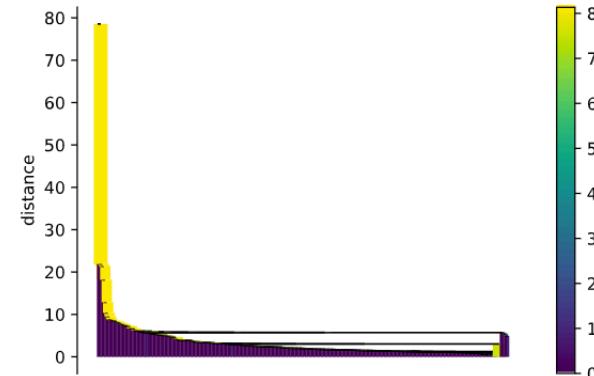
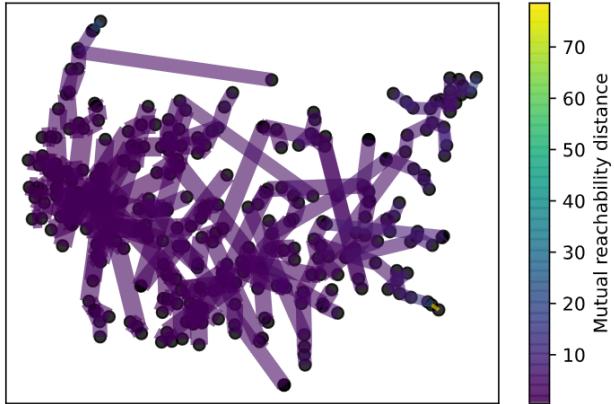


# CLUSTERS

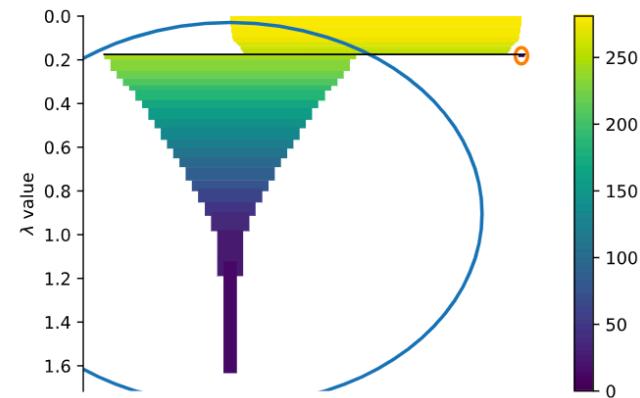
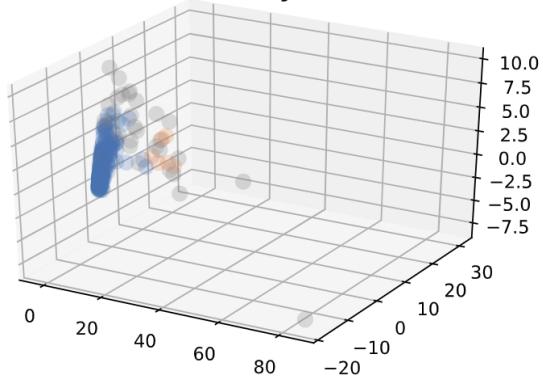


**HDBSCAN**

HIERARCHICAL DENSITY-BASED SPATIAL CLUSTERING OF APPLICATIONS WITH NOISE



Clusters found by HDBSCAN



# CLUSTERS

IN GEOGRAPHY

```
df_final.sort_values(by=['distances'])
```

	city	latitude	longitude	cluster	distances
15	Avenida Paulista, São Paulo, Brazil	-23.557000	-46.661248	-1	0.000000
135	Los Angeles, California	34.053692	-118.242767	-1	59.909923
41	Chicago, Illinois	41.875561	-87.624420	-1	69.486573
185	Phoenix, Arizona	33.448437	-112.074142	-1	76.993621
214	San Antonio, Texas	29.424601	-98.495140	-1	78.554055



LOS ANGELES IS THE CLOSEST TO SAO PAULO

