Customized Suggestions for Travelers

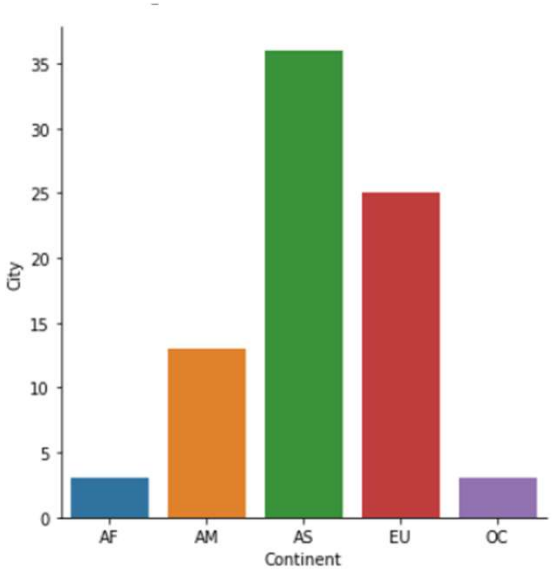
Customizing Suggestions is profitable to Travel Agencies

- Travelers have different preferences, it's based on their personalities and experiences
- Travel agencies can increase sales by offering right packages tours to right costumers
- Costumer fell comfortable to return for more enjoyable tours

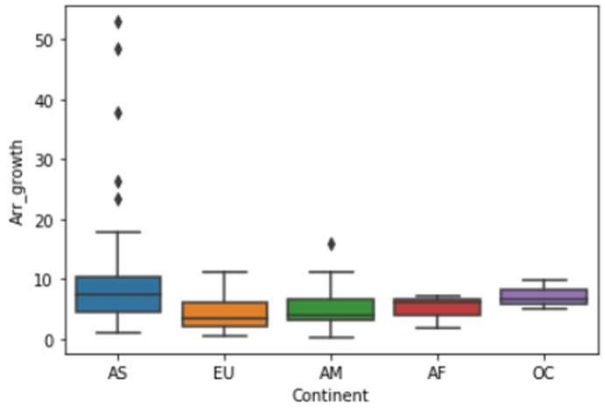
Data acquisition and cleaning

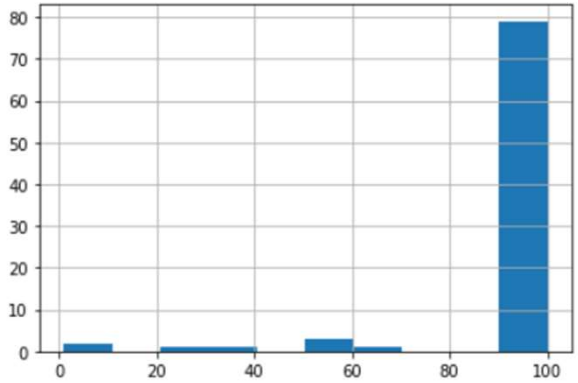
- Ranking of top 100 most visited cities in the world provided by Euromitor (scraped from Wikipedia page)
- Longitude and Latitudes obtained using Nominatim from Geopy library
- Venues information provided by Foursquare trough **explore** function
- CSV file containing tourism types for each Foursquare categories
- Cities with negative arrivals grow rate were excluded
- Asian cities with arrivals grow rate outliers were excluded
- Cities with less than 30 venues around city center were excluded
- Cleaned data contains 80 cities

Analyzed Cities



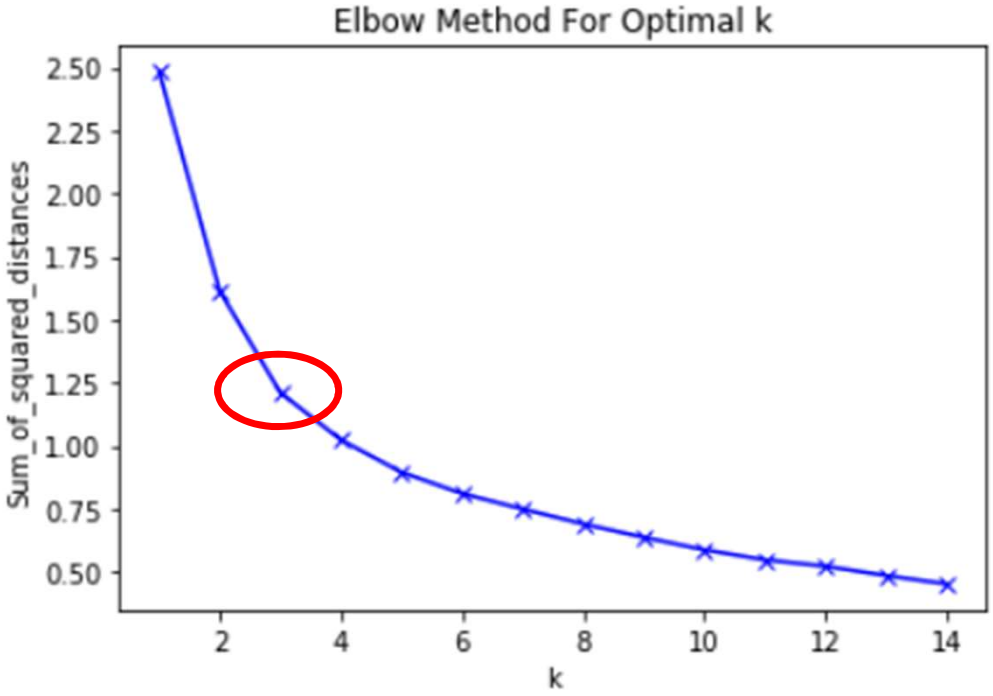
Outliers & Lack of Information





- 5 Asian cities with arrival growth rates above 17%
- 3 cities with less than 30 venues near their city centers

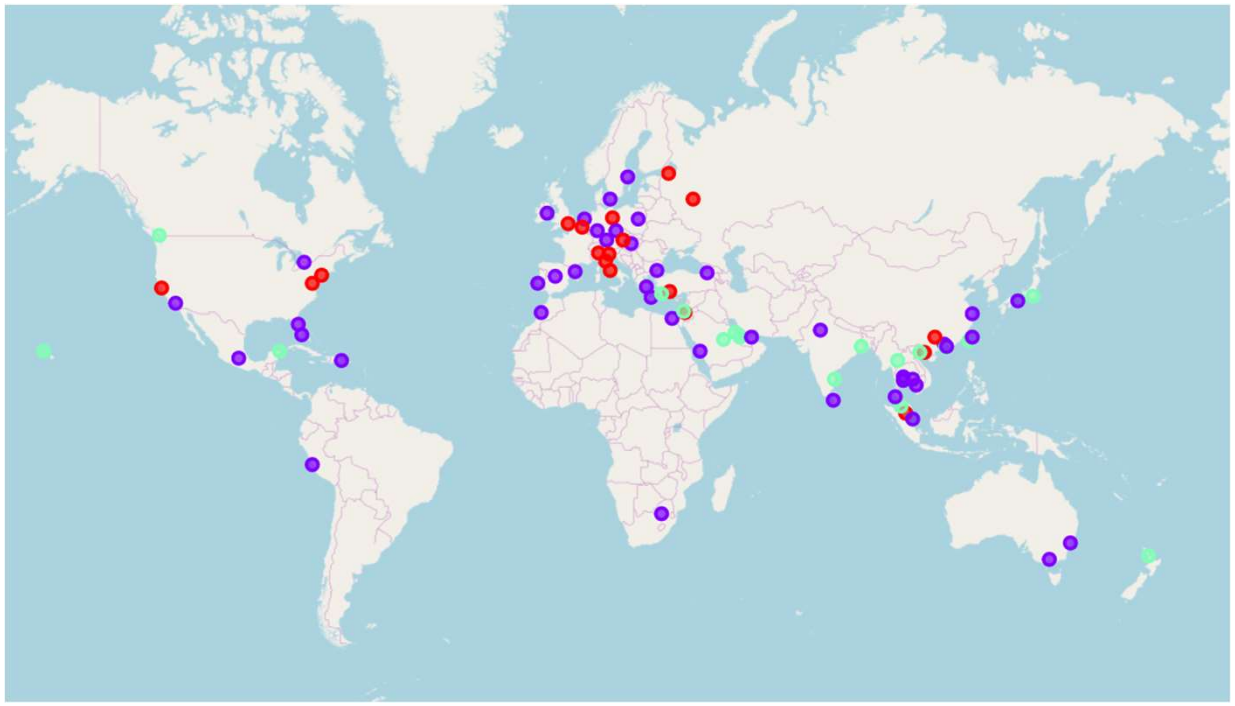
How to run K-Near Clustering Algorithm with best K



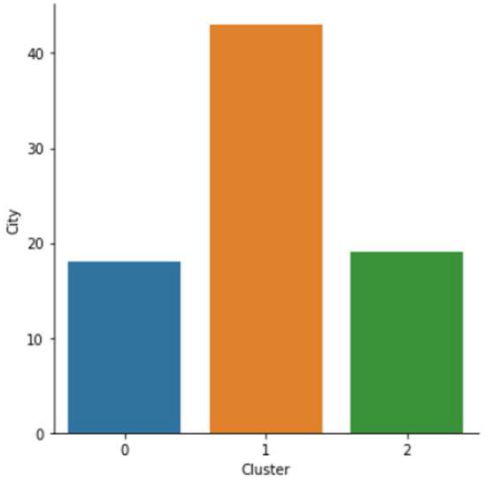
The idea of Elbow Method is that we want a small error, but that error tends to decrease toward 0 as we increase *K*.

So our goal is to choose a small value of *k* that still has a low error, and the elbow of chart usually represents where we start to have diminishing returns by increasing *K*

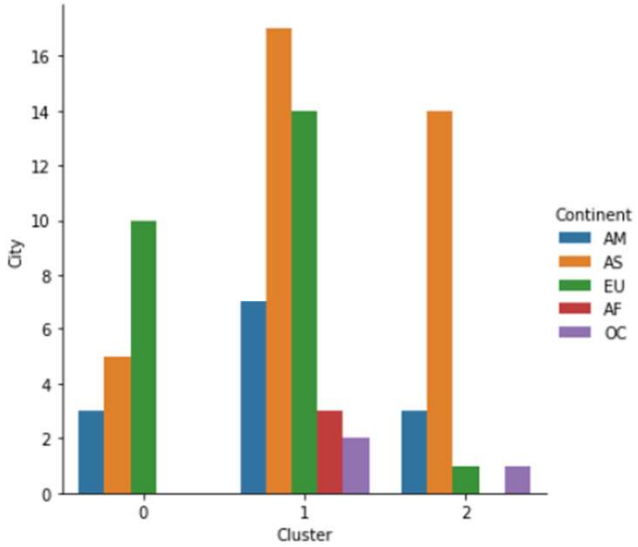
Clustering with K=3



- Red - Cluster 0
- Purple - Cluster 1
- Green - Cluster 2

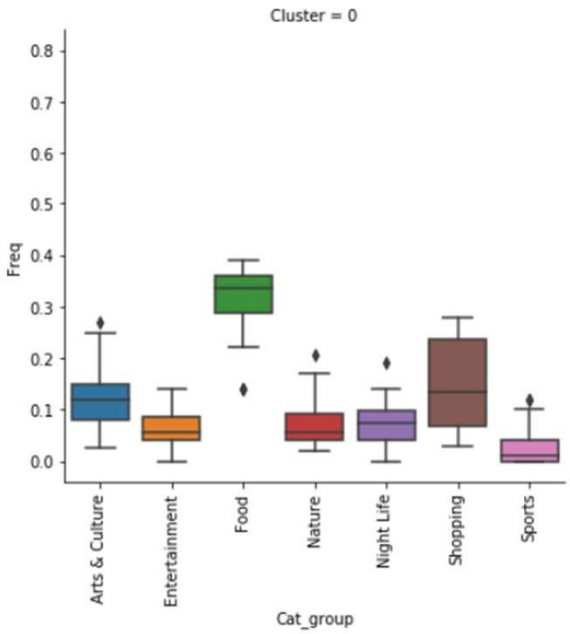


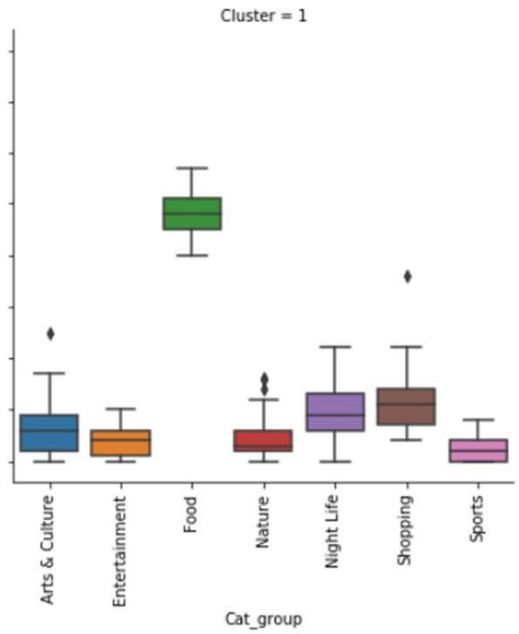
Cluster 1 is by far the most populated cluster. Clusters 0 and 2 have similar sizes.

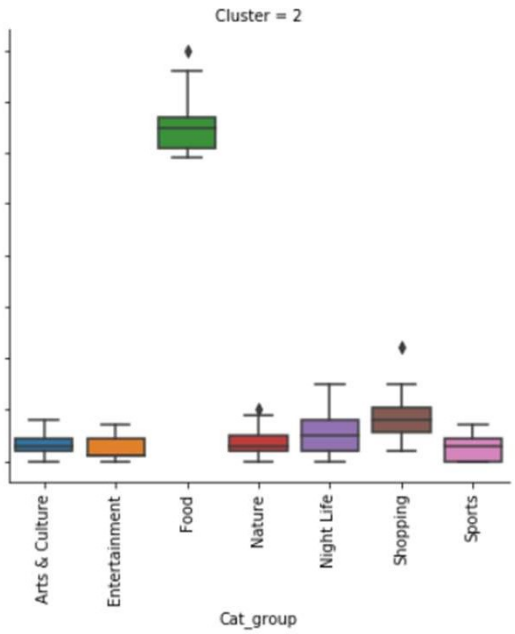


We can see that majority of cities in cluster 0 belongs to Europe. Cluster 1 has cities spread by all continents. However, we can see that cities in Cluster 2 are majority located in Asian.

But we can go deeper…







Cluster 0 holds best cities when visitors want to have a taste of each kind of experience. If visitors want to shopping, enjoy art and culture or see the nature definitely Cluster 0 should be their destiny.

Cluster 1 is the largest cluster. It means its cities can be likely found near the visitors, what can make trips cheapest. Places to enjoy night life and shopping can be found easily as well.

Cluster 2 is definitely recommended to visitor who prefer travel to enjoy food. Places to eat can be found everywhere. Despite that, we can't find a good number of places to enjoy other kind of experiences.

Conclusion & Improvements

- Clusters got from K-Mean Algorithm shows that we can split cities in three main clusters
- In general, places to have fun and enjoy or do sports figure similarly in all clusters
- First cluster holds cities indicated to visitors who wants to experience all kind of experiences. Cities in this cluster are mainly located in Europe and Asia
- Second cluster is indicated to visitors who wants to spend less money to reach their final destination, enjoy food or experience an intense night life
- Third cluster is heavily recommended to visitors who are looking for cities to enjoy food
- Algorithm can be improved if we search venues around entire city limits