Data

Data Sources

For this analysis the following data sources will be used:

- **List of Amsterdam Neighbourhoods**: https://nl.wikipedia.org/wiki/Buurten_en_wijken_in_Amsterdam#Buurten
- **Geo coordinates** of these Neighbourhoods. These will be obtained with the help of the geocoder tool.
- **Top venues** in each Neighbourhoods gathered from Foursquare by the API.

Data Usage

These data sources will be combined, to cluster the different neighbourhoods.

Steps to take:

- For each neighbourhood, the geo-location will be determined with the geocoder.
- With this geo-location, the Foursquare-api will be used to gather popular venues for each neighbourhood. By sorting these an categories, a description of the neighbourhood is gathered (top 1 category, top 2 category, etc).
- Because the sorted categories are available for all neighbourhoods this can be used for clustering the neighbourhoods, using K-means machine learning.
- By looking into the characteristics of the clusters, the best fitting cluster can be found, based on the categories related to tourists attractions and restaurants.