Capstone – project:

AUSTRIAN CITIES

„Analysis of Foursquare Profiles“

21. JUNI

Datahouse.at

Verfasst von: Albert from the countryside

DataHouse.at

Austrian Cities

QUESTION 2: Description of the data and how it will be used to solve the problem

After a short research it turned out that there are hardly any API`s available for retrieving coordinates of Austrian cities.
It seems straightforward to use the geopy() package to address this problem. A table of Austrian cities is available on a Wikipedia website.

de.wikipedia.org/wiki/Liste_der_Städte_in_Österreich#FN_1

Stadt	Bezirk	Bundesland	Markt seit	Stadt seit	VZ 2001	RZ 2011	POPREG 2019
Allentsteig	Zwettl	Niederösterreich		1380	2.163	1.998	1.808
Altheim	Braunau am Inn	Oberösterreich	1581	2003	4.875	4.784	4.933
Althofen	St. Veit an der Glan	Kärnten	ca. 1230	1993	4.732	4.656	4.678
Amstetten	Amstetten	Niederösterreich	1276	1897	22.595	22.847	23.727
Ansfelden	Linz-Land	Oberösterreich		1988	14.789	15.716	16.654
Attnang-Puchheim	Vöcklabruck	Oberösterreich	1955	1990	8.756	8.862	9.042
Bad Aussee	Liezen	Steiermark	1285 (?)	1994	5.086	4.879	4.862
Bad Hall	Steyr-Land	Oberösterreich	1287	2001	4.752	4.773	5.375
Bad Ischl	Gmunden	Oberösterreich	1466	1940	14.081	13.876	14.126
Bad Leonfelden	Urfahr-Umgebung	Oberösterreich	1356	2001	3.847	4.086	4.245
Bad Radkersburg	Südoststeiermark	Steiermark	1265/67	1299	1.599	1.332	3.154

Figure 1: https://de.wikipedia.org/wiki/Liste_der_St%C3%A4dte_in_%C3%96sterreich#FN_1

As the cities are relatively stable it seems not useful to use a scraper for this special task.
Instead of this the basic list will be formatted as a .csv file and hosted on github in order for everyone to get this data.
In a next step the coordinates will be added to the dataset by using the geopy package. A major assumption that could be avoided in this stage by some efforts in manual support is the usage of the retrieved geopy latitude and longitude values as “center” of the cities.
The coordinates will be used to identify venues within the foursquare API and determine the “density” of locations per city. This should lead to “high”,”medium” and “low” densitiy cities (large/medium/small) that shouldn`t be compared or clustered identically.

This step requires some additional analysis of the data, but it seems to be quite important as in Austria we have only a few larger cities with a high number of expected venues. Therefore it wouldn`t be appropriate to compare the percentage of pubs in Vienna (which might be 100+) to the percentage of pubs in e.g. Litschau which might be incidentially one pub that makes up more than 25% of overall restaurants in this small town.
In summary the goal is to cluster the cities after completing the first segmentation into larger and smaller cities. The “radius” parameter will be of great importance for doing this as the “ultimate” center of any town might retrieve many venues and result in a high density whereas this should change with increasing the radius.