Assignment 8
Clusterization

Suppose some of your foreign friends wants to visit Almaty. He researched a bit and found more than 40 interesting places he would like to visit in your city. He has sent you a file ‘almatyPlaces.txt’ with place names and addresses and he want you to help him with his sightseeing within one day. You understand that it is impossible to visit all places by taxi in terms of time and money as well, so you decide to clusterize these places and map them on a google map to see the results and optimize your tour.

1) Your first task is to parse the file provided by your friend and find the exact locations of the places on map. In a geo systems the locations are labeled by longitude and latitude. So read file ‘almatyPlaces.txt’ and extract addresses from lines, each line correspond to a separate address. They come in a following format: Name\tAdress. ‘\t’ means tab. Now, use geopy library to extract latitudes and longitudes of these addresses. Write the results into ‘places.txt’ file. You have to install geopy library first and look through it’s documentation.

2) Now having exact coordinates of visiting places you have to build a map. You are going to use pygmaps library in order to do that. Install it and look through it’s documentation. Pygmaps can build a map using Google Maps API and generate an appropriate html file. Name it ‘mymap.html’. After you have your pygmaps installed, you should read all addresses and their appropriate coordinates from your previously prepared ‘places.txt’ file and map them to a map using pygmaps addpoint() function.

3) This is the time when you get your hands dirty with clusterization. Import scikit k_means library (I used cluster.k_means() algorithm) and clusterize all these locations. Try using different number of clusters (e.g. 3,4,5,...). Add the calculated centroids to your map as well and label them with different color so that you can easily find them.
Now if everything was done correct, you can open ‘mymap.html’ and see the results. Go and meet your friend!