Introduction

For this practical session, we use Kmeans to cluster data based on their locations to find all attractive regions. We use python to process our data, train our model and predict. Our data set came from Flick which is a Yahoo photo sharing system and consists of locations, upload time, user id and description, etc.. But we just used locations we wanted to find acctractive regions. We found that we have to set the size of different clusterings to find the clearest results.

At first we got our results and just ploted it without map. We just wanted to make sure it works. The first time we converted many colums to one colum "date upload time" and useed category to represent it. And we trained based on location, user id, picture id and upload time. The point cloud we got was really fuzzy, there was no cluster actually. Finally, we found that that was because we used so many non relevant colums. After we droped some other columns except locations, it was very clear. We could get something explicitly through out plotting.

After we finished this, we added our map as background. And we grouped them based on the same predict result and ploted group by group. We found that tete park and bellecour were really attractive.