Task 3 Clustering with k-means clustering AI-lab 3

Student 1: Joachim Johnson

Student 2: Jonas Åsander

Course: Artificial Intelligence (**DT112G**)

franziska.klugl@oru.se.

Introduction:

In this task shall we learn more about clustering as were presented in lecture. in this task shall we use k-means clustering and implement it in code. The code that shall we build is the the case based code that's where developed in task 2 where you compare an input of attributes from an user or possibly another system and an database of restaurants that stored in our system.

In this task shall we continue to develop this code to cluster the database using different averages of the attributes we did chose to use in task 2.

One critical point is that we shall calculate the Centroid of the clusters and later print them out. Here can it be several solution then it is several "symbolic" values that we most find an ccenteroid on

When it is all the restaurants are sortert in clusters should we print out the clusters and how many restaurants is in them and the Centroid of them. How many cluster it should be is or print out how many is not specified in task description.

Language uset Python
Built upon task 2 Previous done in this course

Data structure design:
Its obvious we continued to use a lot of the structure we build in task 2 with classes and the database that we had and that is stored in csv files. We added more cases where added to store the clusters that were created and use instances and dictionary of these instances.

We have kept everything from task 2 and all functions still intact. We use the database class as an "center" of the program that the functions build structures from.

Overall processes description:
Here did we recycle from task 2 with the similarity function and how we calculated the similarity from it. We have functions that calculate the centroid of a set of restaurants that sent in and a function that's compare an restaurant to an set of centroids and says which is most similar to each other.
With these functions and the cluster class do we run everything in iterations in an loop to get the the clusters. The initial clusters are randomised to set out the early clusters that floats out in different ways as the loops running.
After this print them out with the information that the tasks is asking after.

Results and Discussion:

In discussion can we say that quite a lot where changes were we planned to continue develop and it where not too much focus on this in task 2.

Here did we do more tests with the database. With more variation on the number of restaurants to see the effects of the clustering that's where the point of this task. Oure database changed abit to see the clusters.

It could be discussed what kind of database we used and what restaurants were used and what are most similar to what. What kind of restaurant is most similar to which other type.

How we used our comparison algorithm in task 2 did we continue on the same way with the comparison and recycle the code and did the centroid calculation in simulare manor to keep consistency. There is a few way that it could be done with the "symbolic" values.