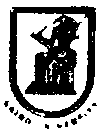
****Data Mining**

**Information Systems Department**

**Faculty of Computers and Artificial Intelligence**

###### **Cairo University**

Assignment 2

Clustering

**Instructions:**

* This assignment should be performed individually; copies will be graded -5.
* The assignment total grade is 5.
* The discussion for the assignment will be during the week starting with **Saturday April 23 until Wednesday April 27.**
* Select one problem only to apply K-means clustering algorithm using any programming language.
* The number of clusters (K) and the distance function (Manhattan or Euclidean) should be provided as an input from the user during runtime.
* Initial centroids should be chosen randomly.
* The output should show each cluster (K) with its final points (No need to show each iteration).

**Problem 1:**

* Consider the perfume dataset in the file named *Odor\_Meter\_Results.xlsx*, it contains the strength of the smell of 20 perfumes measured by an odor meter throughout 25 seconds, so each second (column) is a new measurement.
* Write a program to group these perfumes based on the similarity in the strength of their smell.
* You should use k-means clustering algorithm.
* The final output of your program should show k lists of perfumes, and show outlier perfume’s records (if exists).

**Problem 2:**

* Consider the household power consumption data set in the file named *Power\_Consumption.csv*, it contains the power consumption information of 200 households.
* Write a program to group these households based on the similarity in their power consumption.
* You should use k-means clustering algorithm.
* The final output of your program should show k lists of households’ IDs, and show outlier records (if exists).

Good Luck