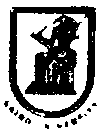
****Data Mining**

**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 assignment should be submitted before 26/12/2020 at 11pm on blackboard.
* Submission after 26/12/2020 and until 30/12/2020 will be considered a late submission and 50% of the grade will be lost.
* Select one problem only to apply K-means clustering algorithm using any programming language.
* The number of clusters (K) 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 sales dataset in the attached file, it contains weekly purchased quantities of 200 products over 31 weeks.

<Sales.xlsx>

* Write a program to group the products based on the similarity of their purchased quantity over all weeks.
* You should use k-means algorithm to cluster the products to k clusters.
* You should use Manhattan distance as your distance function.
* You should detect outlier data (if exists).
* The final output of your program should show k lists of products and show outlier product’s records.

**Problem 2:**

* Consider the course evaluation dataset in the attached file. It contains course evaluation scores provided by 150 students each row represents scores provided by one of the students for each question in the survey.

[Course Evaluation .xlsx](Course%20Evaluation%20.xlsx)

* Write a program to group the students based on the similarity of their answers on the survey.
* You should use k-means algorithm to cluster the students to k clusters.
* You should use Euclidean distance as your distance function.
* You should detect outlier data (if exists).
* The final output of your program should show k lists of students and show outlier student’s records.