Kingdom of Saudi Arabia
Ministry of Education
UMM AL-QURA University

Data Mining
Second Assignment
First semester 1442

Course name	Data mining - 14016165-3 / Data mining - 14016313-3
Assignment tittle	Assignment 2: Clustering problem
due date	03-12-2020 (Week 14).
Assignment weight	25%

Please perform the following tasks:

Stage	Task 1
Downloading the Dataset	1. Choose any dataset -for clustering problem- from the UCI Machine Learning Repository or from sklearn datasets. 2. Download the selected dataset. 3. Describe the dataset and the clustering task.
Data Exploration	4. Display the number of instances. 5. Display the number of attributes. 6. Display a statistical summary for all the attributes.
Data Preprocessing	7. Check whether the selected dataset has any data quality issues and choose suitable strategies to deal with any issue (if exists).
Appling clustering algorithms	8. Build clustering models using K-Means algorithm 9. Find the best K value using the elbow method, use a Python package, kneed, to identify the elbow point programmatically.

Stage	Task 2
Downloading the Dataset	1. Choose another labeled dataset from the UCI Machine Learning Repository or from sklearn datasets. 2. Download the selected dataset. 3. Describe the dataset and the clustering task.
Data Exploration	4. Display the number of instances. 5. Display the number of attributes. 6. Display a statistical summary for all the attributes.
Data Preprocessing	7. Check whether the selected dataset has any data quality issues and choose suitable strategies to deal with any issue (if exists).
Appling clustering algorithms	8. Build two clustering sets using the following Clustering Algorithms: a. Hierarchical clustering b. Density based clustering 9. Evaluate the two clustering sets using Entropy or Purity.

Important notes:

- . This an individual assignment
- . Youneedto use the Jupyter Notebook t o perform the all the required tasks
- Theipynd file name should be in the following format: (first n ame)_(last name)_clustering.ipynb for example Majed_Farrash_classification.ipynb
- . Bythe due date, you must submit your ipynb file using the blackboard