



Mohammad Ali Jinnah University Karachi

Introduction to DataScience Assignment 02

Q1. Examine the given dataset to formulate k-mean clusters and show calculation at all iterations in Python. **Please use cosine similarity and find the most probable together, you can define a threshold from 0.60-0.90 together.** Monkeypox is an infectious disease caused by the monkeypox virus that can occur in certain animals, including humans. Symptoms begin with fever, headache, muscle pains, swollen lymph nodes, and tiredness.

An ongoing outbreak of monkeypox was confirmed on 6 May 2022, beginning with a British resident who, after traveling to Nigeria (where the disease is endemic), presented symptoms consistent with monkeypox on 29 April 2022. The resident returned to the United Kingdom on 4 May, creating the country's index case of the outbreak. The dataset contains the timeline for confirmed cases, in which The trends show the effect of Travel_History on each country in two different perspectives of MonkeyPox- Confirmed_Cases, and Suspected_Cases.

Take these two variables **Travel_History_Yes X1 and Travel_History_No X2. and Confirmed Cases Y** and find the similarity among Countries.

- A.** Find out the similarity among Countries based on $X1 = \text{Travel_History_Yes}$, $X2 = \text{Travel_History_No}$, and $Y = \text{Confirmed Cases}$. **[8 marks]**
- B.** Summarize the Country from highest confirmed to lowest in a separate table based on clusters similarity among countries. Please name Cluster 1 as "A" for a set of countries and Cluster 2 as a country "B" **[4 marks]**
- C.** Please take two clusters and find the similarity among countries. **[3 marks]**

Note: The dataset is uploaded to the classroom with this assignment.