Laboratory Activities for Week 8: Clustering

SC310005 Artificial Intelligence
Khon Kaen Business School

(10 Points) K-Means Clustering on COVID-19 Dataset

Dataset Overview and Objectives

Dataset Description:

Dataset:
https://raw.githubusercontent.com/kaopanboonyuen/SC310005_ArtificialIntelligence_2023s1/main/dataset/covid19_vaccination_dataset.csv

The dataset you will be working with contains information on COVID-19 cases from Kaggle. The data includes various features related to the spread and impact of the virus across different regions.

Motivation:

Understanding the patterns and clusters within the COVID-19 data can provide valuable insights into how the virus has affected different areas. By applying k-means clustering, you can identify groups with similar characteristics, which may help devise targeted strategies for different regions.

Assignment Objectives:

- Apply k-means clustering to group regions based on selected COVID-19 features.
- Visualize and interpret the clusters.
- Explain the characteristics of each cluster.

Objective:

The main objective of this assignment is to explore patterns in COVID-19 data using k-means clustering and interpret the results.

Assignment Problem:

You are tasked with applying k-means clustering on the provided COVID-19 dataset. The goal is to identify distinct clusters of regions based on selected features related to COVID-19 cases.

Task for Students:

- □ Import the COVID-19 dataset.
- □ Explore the dataset and understand the available features.
- □ Select relevant features for clustering (e.g., people_fully_vaccinated, daily_vaccinations).
- □ Handle missing values, if any.
- □ Normalize or Standardize the selected features.
- □ Use the k-means algorithm to cluster the regions based on the selected features.
- □ Determine the optimal number of clusters (k).
- □ Visualize the clusters in 2D or 3D space using a scatter plot. (Differentiate the clusters with distinct colors.)
- □ Explain the characteristics of each cluster.
- □ Explore the patterns and trends within each cluster.

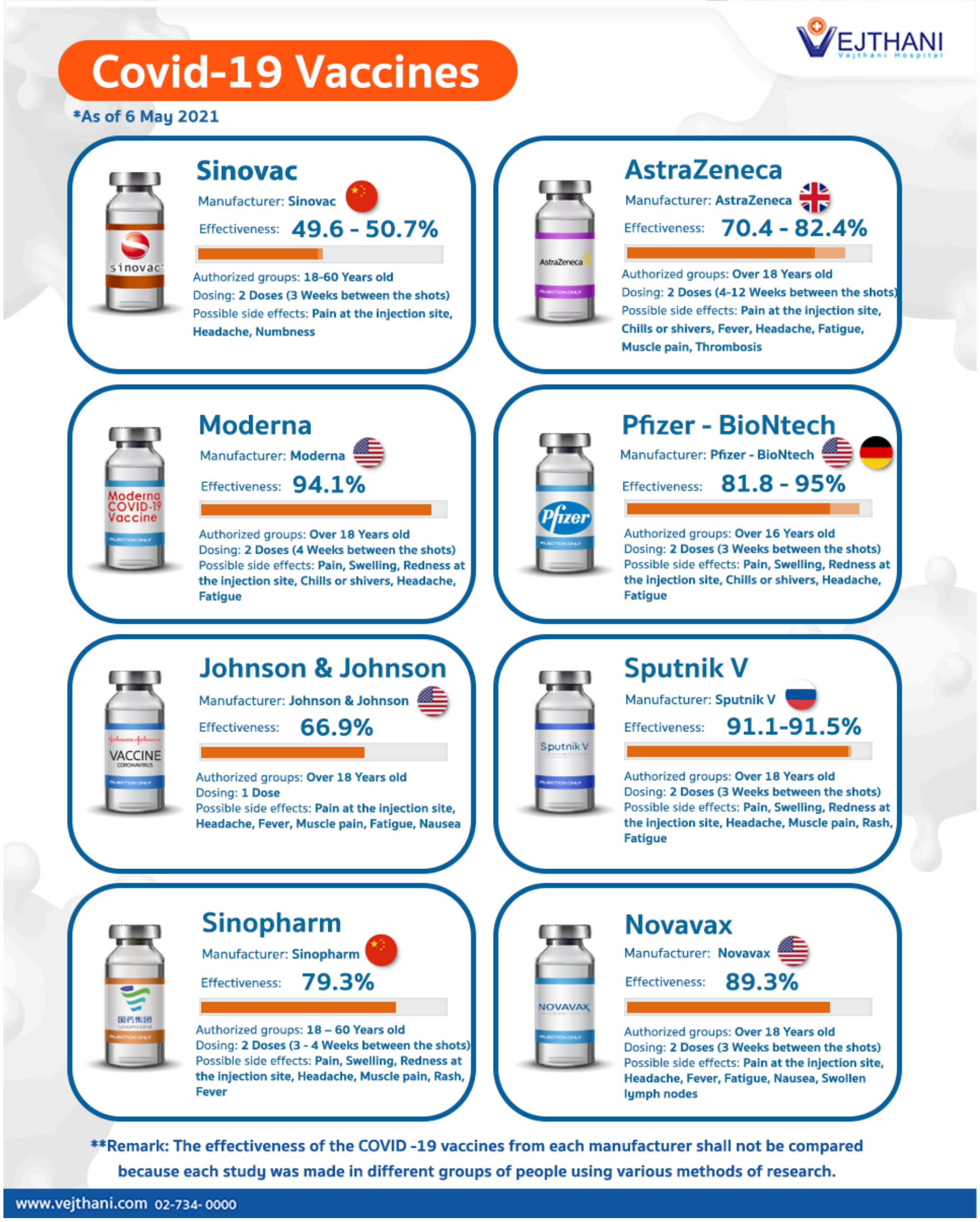
Data Dictionary:

The data (country vaccinations) contains the following information:

- **Country**- this is the country for which the vaccination information is provided;
- **Country ISO Code** - ISO code for the country;
- **Date** - date for the data entry; for some of the dates we have only the daily vaccinations, for others, only the (cumulative) total;
- **Total number of vaccinations** - this is the absolute number of total immunizations in the country;
- **Total number of people vaccinated** - a person, depending on the immunization scheme, will receive one or more (typically 2) vaccines; at a certain moment, the number of vaccination might be larger than the number of people;
- **Total number of people fully vaccinated** - this is the number of people that received the entire set of immunization according to the immunization scheme (typically 2); at a certain moment in time, there might be a certain number of people that received one vaccine and another number (smaller) of people that received all vaccines in the scheme;
- **Daily vaccinations (raw)** - for a certain data entry, the number of vaccination for that date/country;
- **Daily vaccinations** - for a certain data entry, the number of vaccination for that date/country;
- **Total vaccinations per hundred** - ratio (in percent) between vaccination number and total population up to the date in the country;
- **Total number of people vaccinated per hundred** - ratio (in percent) between the population immunized and total population up to the date in the country;
- **Total number of people fully vaccinated per hundred** - ratio (in percent) between population fully immunized and total population up to the date in the country;
- **Number of vaccinations per day** - number of daily vaccination for that day and country;
- **Daily vaccinations per million** - ratio (in ppm) between vaccination number and total population for the current date in the country;
- **Vaccines used in the country** - total number of vaccines used in the country (up to date);
- **Source name** - source of the information (national authority, international organization, local organization etc.);
- **Source website** - website of the source of information;

There is a second file added recently (country vaccinations by manufacturer), with the following columns:

- **Location** - country;
- **Date** - date;
- **Vaccine** - vaccine type;
- **Total number of vaccinations** - total number of vaccinations / current time and vaccine type.



Covid-19 Vaccines

*As of 6 May 2021

Sinovac

Manufacturer: **Sinovac**
Effectiveness: **49.6 - 50.7%**
Authorized groups: **18-60 Years old**
Dosing: **2 Doses (3 Weeks between the shots)**
Possible side effects: **Pain at the injection site, Headache, Numbness**

AstraZeneca

Manufacturer: **AstraZeneca**
Effectiveness: **70.4 - 82.4%**
Authorized groups: **Over 18 Years old**
Dosing: **2 Doses (4-12 Weeks between the shots)**
Possible side effects: **Pain at the injection site, Chills or shivers, Fever, Headache, Fatigue, Muscle pain, Thrombosis**

Moderna

Manufacturer: **Moderna**
Effectiveness: **94.1%**
Authorized groups: **Over 18 Years old**
Dosing: **2 Doses (4 Weeks between the shots)**
Possible side effects: **Pain, Swelling, Redness at the injection site, Chills or shivers, Headache, Fatigue**

Pfizer - BioNtech

Manufacturer: **Pfizer - BioNtech**
Effectiveness: **81.8 - 95%**
Authorized groups: **Over 16 Years old**
Dosing: **2 Doses (3 Weeks between the shots)**
Possible side effects: **Pain, Swelling, Redness at the injection site, Chills or shivers, Headache, Fatigue**

Johnson & Johnson

Manufacturer: **Johnson & Johnson**
Effectiveness: **66.9%**
Authorized groups: **Over 18 Years old**
Dosing: **1 Dose**
Possible side effects: **Pain at the injection site, Headache, Fever, Muscle pain, Fatigue, Nausea**

Sputnik V

Manufacturer: **Sputnik V**
Effectiveness: **91.1-91.5%**
Authorized groups: **Over 18 Years old**
Dosing: **2 Doses (3 Weeks between the shots)**
Possible side effects: **Pain, Swelling, Redness at the injection site, Headache, Muscle pain, Rash, Fatigue**

Sinopharm

Manufacturer: **Sinopharm**
Effectiveness: **79.3%**
Authorized groups: **18 – 60 Years old**
Dosing: **2 Doses (3 - 4 Weeks between the shots)**
Possible side effects: **Pain, Swelling, Redness at the injection site, Headache, Muscle pain, Rash, Fever**

Novavax

Manufacturer: **Novavax**
Effectiveness: **89.3%**
Authorized groups: **Over 18 Years old**
Dosing: **2 Doses (3 Weeks between the shots)**
Possible side effects: **Pain at the injection site, Headache, Fever, Fatigue, Nausea, Swollen lymph nodes**

**Remark: The effectiveness of the COVID -19 vaccines from each manufacturer shall not be compared because each study was made in different groups of people using various methods of research.

www.vejthani.com 02-734- 0000