K - Mean Clustering - Home Exercise:

The **Mall_Customers.csv** is a well known dataset that is often used for customer segmentation. Thus it can be used in unsupervised learning models to perform clustering tasks.
The dataset includes various features about customers of a mall.

Customer segmentation is a very important task for any company because it can help with targeted marketing, improving customer experience, and increasing sales.

Below is an explanation of typical columns you might find in this dataset:

- **CustomerID** → A unique identifier assigned to each customer.
- **Gender** → The gender of the customer (categorical: Male/Female).
- **Age** → The age of the customer in years.
- **Annual Income (k$)** → The annual income of the customer in thousands of dollars.
- **Spending Score (1-100)** → A score assigned by the mall based on customer behavior and spending patterns on a scale from 1 to 100, where a higher score indicates higher spending patterns.

Exercise instructions:
Use the **Mall_Customers.csv** file for this exercise.

Data preparation:

1. Check for missing values in the dataset.
2. Check for duplicate rows in the dataset.
3. In case you found any of those, remove them from the df.
4. Convert categorical features to numerical form using the get_dummies() method.

K-Mean Clustering Machine Learning:

1. Use your preprocessing dataset.
2. Ensure categorical features are converted to numerical.
3. Make sure to use feature scaling in case needed.
4. Run a simple K-Mean Clustering model with K = 4.
5. Print the cluster center values and print the prediction values.
6. Generate a simple chart that will show for each customer what cluster it belongs to.
7. Run elbow method for K values between 2 and 10 and find what is the optional K value. Provide an explanation why you chose this K value.
8. Train your K-Mean Clustering model again using the optimal K value you selected.
9. Add the clustering prediction to each data point in the dataset by adding a new column called cluster.
10. Generate a simple chart that will show for each customer what cluster it belongs to.

Model Deployment:

1. Export your final model into a joblib file.
2. Ensure that you also export other relevant preprocessing instances such as the standard scaler and one-hot encoder.
3. Import and Load → Import your final model and the preprocessing instances from the joblib files back to your working area when needed for deployment.

Model New Predictions:

Use your trained model and the following 3 new customer details and predict the cluster each customer is associating.

CustomerID	Genre	Age	Annual Income (k$)	Spending Score (1-100)
201	Male	45	20	45
202	Female	28	22	92
203	Male	33	19	20

ECOM SCHOOL

המכללה למקצועות הדיגיטל וההייטק