

ADS Assignment 10.

Hi everyone. The assignment focuses on what you have learned in the Unsupervised Machine Learning notes. Use the following link to source your data for the exercise on Sales Conversion Optimization: <https://www.kaggle.com/loveall/clicks-conversion-tracking>.

Ensure you have read and understood the instructions about the data including what is in each column so as to correctly answer the questions below. You are expected to finish and submit your notebooks latest by Friday at 5 PM. All the best.

1. Import the libraries and load the data.
2. Dummy encode any categorical or object values in the data and save the resulting data frame to variable X.
3. Using a heat map to show the correlation in the data.
 - a. Drop the first 4 columns in the data frame X.
 - b. Basing your answer on what can be seen in the heat map, why did we drop these columns?
4. Using the elbow method:
 - a. Determine the best number of clusters for the data in the range of 2 to 20.
 - b. Also include the graphical plot for the elbow curve.
5. Based on the result above in 4b use the value at your elbow point to cluster the values in the data frame X.
6. Use the model to predict the labels from the data and save them to variable y_means.
7. Add the values in y_means to the original data frame (not X) as column 'Advert_Type'.
8. Using any form of distribution plot of your choice and the original data frame, plot 2 graphs that can be used to answer the following:
 - a. Which advert type lead to the highest and consistent amount of sales by customers of all the age brackets?
 - b. Does the company xyz have gender bias in terms of their ad spending? Are their products gender neutral?
9. Perform the same analysis using hierarchical clustering and compare the results in terms of the number of useable clusters.