Task 1:

You are working for an eCommerce company who has been able to collect information on the income of a small group of their customers using a survey. However, they have been able to collect much more customer data through their website and are hoping that they can use this to predict the incomes of the rest of their customers to help them make business decisions such as how best to price their products and target their advertising.

You have been provided with a small sample of the complete data (note: this data has already been cleaned, so there is no need to check for incorrect, duplicate, or missing data). This contains information such as a customer's salary, their activity on the website and various demographic attributes.

Data

To complete this task, you must use the data attached with the assignment in classroom.

Submissions based on other data will not be marked and will receive 0 marks.

Task

You have to make a new column based on the last salary column and assign 1 to all people with salary greater or equal to 35k and assign 0 to all with salary less than 35k.

Alter the data accordingly and apply a binary classification model using KNN.

Note: You have to write algorithm for KNN using only Numpy and Pandas. You are **not allowed** to use any other library like sklearn.

Task 2

K-means Clustering

You have to apply k means clustering algorithm on the provided speech dataset. Divide the data into 10 clusters and display those 10 clusters in a Dataframe as shown below**.**

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10
0_01_1	0_02_5	0_04_11	0_03_3	0_05_8	0_08_11	0_06_13	0_07_15	0_09_12	0_10_6
0_02_2	0_03_4	0_05_17	0_06_1	0_05_7	0_07_12	0_01_11	0_04_12	0_08_11	0_9_5

Note: You have to write algorithm for K-means Clustring using only Numpy and Pandas. You are **not allowed** to use any other library like sklearn for it.

Dataset:

https://drive.google.com/file/d/1zXm9os8XB4Tpqk6WpQEJaoG6vAwtpGjc/view?usp=sharing