For Even Roll Numbers:

Use the file "Shared for Clustering.xlsx". The file has two sheets: Training and Validation. Data in these sheets are stream data. Each row represents values measured by 300 sensors. This means that every row is a distinct observation. Please note that row order in the sheet does not necessarily represent the order in which these data have arrived. Data in the Training sheet is to be used to learn the clustering model while data in the Validation sheet (there are 60 observations in the validation set where the first 30 are from cluster 1 while the remaining are from cluster 2) is provided to validate the model. In no case, data in the Validation sheet should be used for learning the model. Use The Stream Clustering Algorithm to cluster the data. You may use chunk size to be any value between 4 and 10. You may use any algorithm as the base algorithm of the Stream Clustering Algorithm. Submit YOUROLLNUMBER.ipynb. Your program when run shall allow me to upload a .xlsx file containing 140 test observations. It should then generate a YOUROLLNUMBER.csv file with prediction (1 or 2). This file should have only one column without header and 140 predictions/rows. I will compute the accuracy by comparing the predictions with the actual cluster label. Please note, if I will find a batch implementation, 0 marks would be awarded.

Deadline for Submission: 12:45 pm. I will share a mechanism (most probably, a Google form) for the submission before the deadline.