Batch 10–Data Mining 1–Small-team Case Assignment 1– Required (Honor code: 3N-b)

- This is a small-team (preferably pair) case assignment.
- Please upload your work on the LMS by the deadline as specified on LMS.
- You may use any software to work on your assignment as long as each team member is comfortable using the software.
- Please be precise while answering any question. Unnecessary details or superfluous answers will attract penalty even if it has the right answer embedded in it. Whenever appropriate you must include relevant output from the data analysis in support of your answer. An answer without a justification will not be awarded any credit.

Case: Mall of America gains insights from Wi-Fi data

Datasets included along with the case narrative: (1) MOA Case Data Dictionary.xlsx, (2) moa_analytics.csv, (3) moa_recurrence.csv, (4) moa_session_update.csv

Imagine you are part of the analytics team working with Mall of America as your client. The nature of your engagement is to derive value from the Wi-Fi data in ways that support MOA's business questions and goals. Ignore the questions that are provided on the last page of the case. Instead, answer the following questions:

1. Is there a business value to the MOA in giving free Wi-Fi to its consumers? Justify your answer in no more than 500 words.
2. How to design a pilot study that serves as a proof-of-concept to show the business value of individual level Wi-Fi usage data? Where could this go in the future? Please limit your answer to 500 words.
3. What kinds of use-cases exist from the mining of Wi-Fi-data? Please be very specific here. Do not enumerate all questions provided in the case on pages 6 and 7. Enumerate only those use cases that the data provided to you along with this case can support, and while doing so, clearly mention the connection between the data and the use cases. Please limit your answer to 1000 words.
4. Run segmentation analysis to better understand the behavioral patterns of mall visitors based on when and where people touch the Wi-Fi access points.
    i. Run a clustering algorithm and determine an optimal number of clusters. Please state your approach in data preparation, choice of the clustering algorithm, the attributes your clustering is based on, and eventually your decision on the number of clusters.

Recall the framework we discussed in class: What are you clustering? Each row in the final dataset should uniquely represent an entity. What are your clusters based on? You should only include meaningful attributes of the entities that you are clustering. Keep in mind the business objective while framing the cluster analysis problem.

ii. Identify, name and describe clusters of mall visits. Trying to answer the following questions against each segment may help you label these segments better. What is the average duration a visitor spends in the mall for these segments? Is there a segment that does not peak on Saturdays or Sundays? What percentage of people belonging to a segment is likely to be mall employees? At what times do entries and exits peak across these segments? Do any of these segments have a lot of quick visitors or a lot of long-duration visitors?

iii. Make meaningful recommendations to answer the use-cases you identified in Part 3 above.

After submission, team members may be asked to fill a peer evaluation report, if multiple team members indicate insufficient contribution of a member. An appropriate penalty will apply to concerned members in such a case.