Assignment

You are given a data set of recommendations from a social network in csv format. Please do the following operations using python libraries on Jupyter Notebook. The file that you will upload for marking should have both the source code and some visuals. The data on the data set represents 1) The recomendation id- which is an auto generated number representing a recording of each recommendation which was generated by the recommendation algorithm, 2)Subject-user, represents the user which received a recommendation ,3) Recommendation-time , represents the time that the subject-user gets a recommendation , 4) clicked-user, represents the user which was recommended and clicked, 5) clicked-time, represents the time that the clicked-user was clicked.

1. Find and display how many recommendations where generated by the algorithm from the data set [2]

2. How many people respond to the recommendations given [3]

3. Among the subject-user how many unique users received recommendations [5]

4. What is the time range that the algorithm generated a lot of recommendations [5]

5. Given that a session took about 10 mins , find how many total unique recommendations where generated and out of these how many responded by the subject-user [10]

6. Among the clicked users how many of these where clicked once by a particular user (or how many of the clicked users where unique or new to the subject user) [5]

7. What is the relationship between the subject-user and the recommended time(You can draw a graph showing this relationship), describe the results shown on the graph [10]

8. What is the relationship between subject-user, clicked-user and clicked time (use a graph to show this relationship),describe the results shown on the graph [10]

9. Show the correlation between each variable/feature [5]

10. Fill the missing values under clicked-user and clicked time with random data [10]

11. Repeat question 6 and 7 with the full data set using scatter graphs [20]

12. What is the accurate and recall rate of the algorithm [10]

13. Using a machine learning model of your choice , train model and then create a platform such that given a subject-user, recommendation time , clicked-time , your model can predict the clicked-user [20]