

SQL Problem

We have a new product - the Property Finder chat function. It enables users to connect with an agent to enquire about a property, and also agents to connect with each other about potential leads.

To use the chat function you create a user and 'connect' with another user - a bit like a social network. We'd like to reward our most active users, and understand the connection network better.

Given a table structure like the below table 'user_connections'

1. Please aggregate the number of connections that each user has
2. How do we see how many mutual connections each pair of users have?
3. Any other comments on this problem - are there better ways to store the data?

user 1	user 2
1	2
1	3
1	6
2	1
2	6
2	12
2	5
3	1
3	7
3	6
3	11
6	1
7	3
12	2

Note that if user A is connected to user B then we have an entry (A,B) and (B,A)

Modelling Problem

1. Please refer to the sample data file 'DS Assessment - Segmentation'. This is a sample of *some* of the data we collect

Given the structure of this data, please

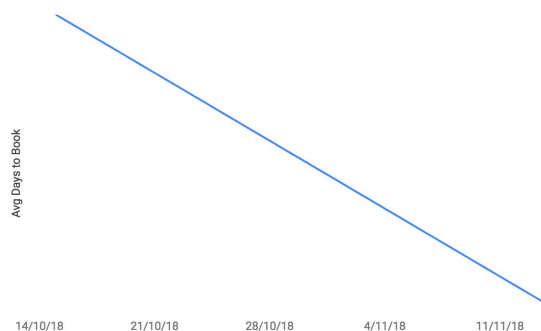
- a. Provide a script which creates features that could be descriptive in understanding our users
- b. Include an initial segmentation approach in this script and justify your logic
- c. Suggest some potential segments that might arise, and show how you would communicate these to the marketing manager to help her communications planning
- d. Please share any other observations

Please note: we like Python and R. Notebooks are fine but not required, just please make sure your thought process is clearly communicated - tell us a story!

2. Now that you have created your segmentation algorithm, please describe how you might go about putting this into production so that it's regularly updated
3. Aside from the features you create from the provided data, what else do you think could support an effective segmentation project from a tracking/features point of view?

Metrics Problem

1. We recently changed our search function, and implemented a new ranking algorithm for our properties. Please suggest some metrics that may be useful for assessing whether a change to our new ranking algorithm was successful. Why are these useful metrics?
2. One metric we check is the average time to lead after a first page view. Looking back from today (14th Nov) we get a chart that looks like this:



Please comment on this approach, and suggest alternative approaches.