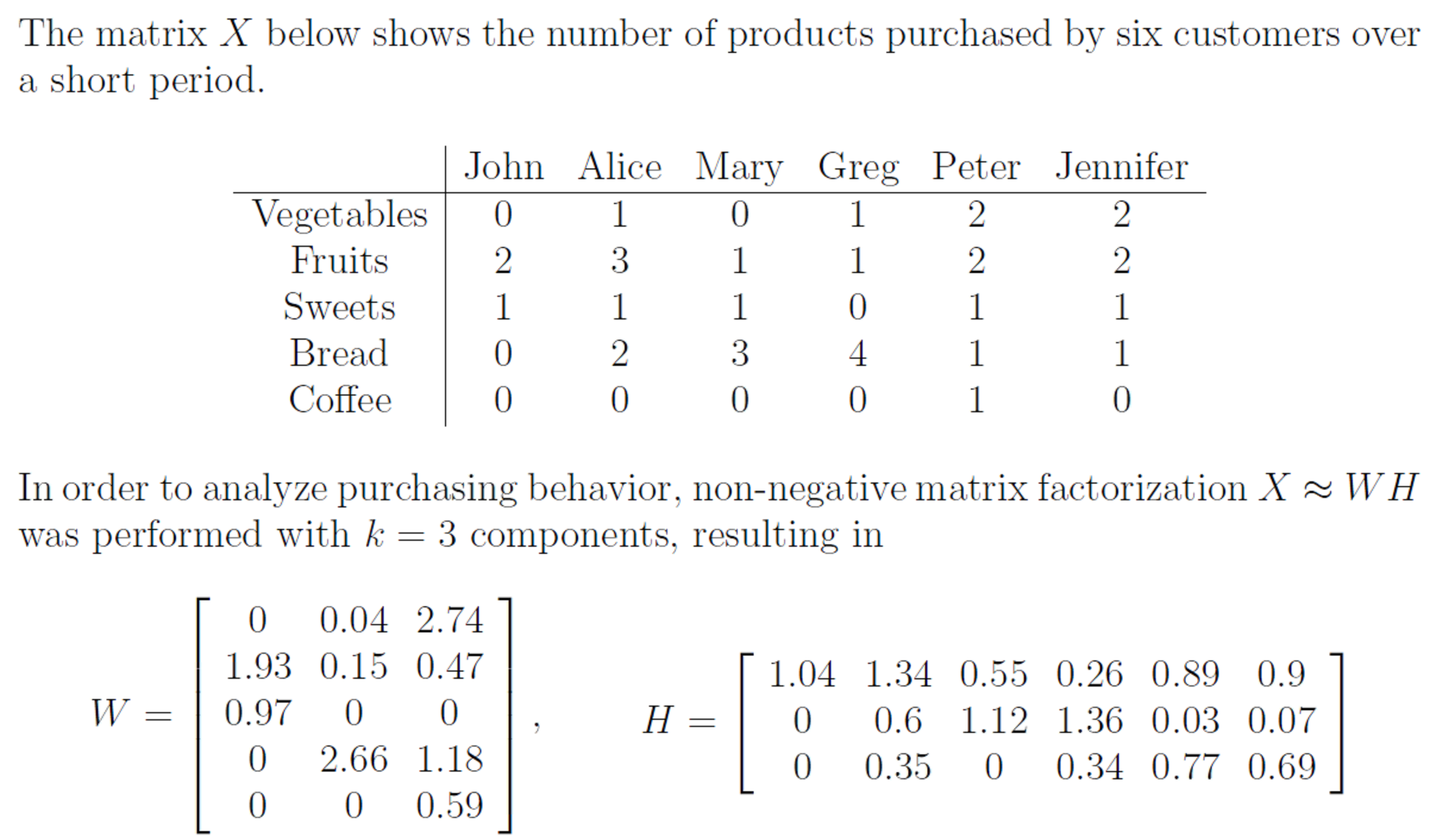
**REI602M Quiz 19.2.2021**

Hand in on Gradescope before 22:00 on Feb. 20 (Saturday). Each question will be given 1, 0.5 or 0 points as follows. If the question is more or less correct it gets 1 point. If it is partly correct it gets 0.5, and if it is missing or completely wrong it gets 0 points.

1) [Exam 2019]

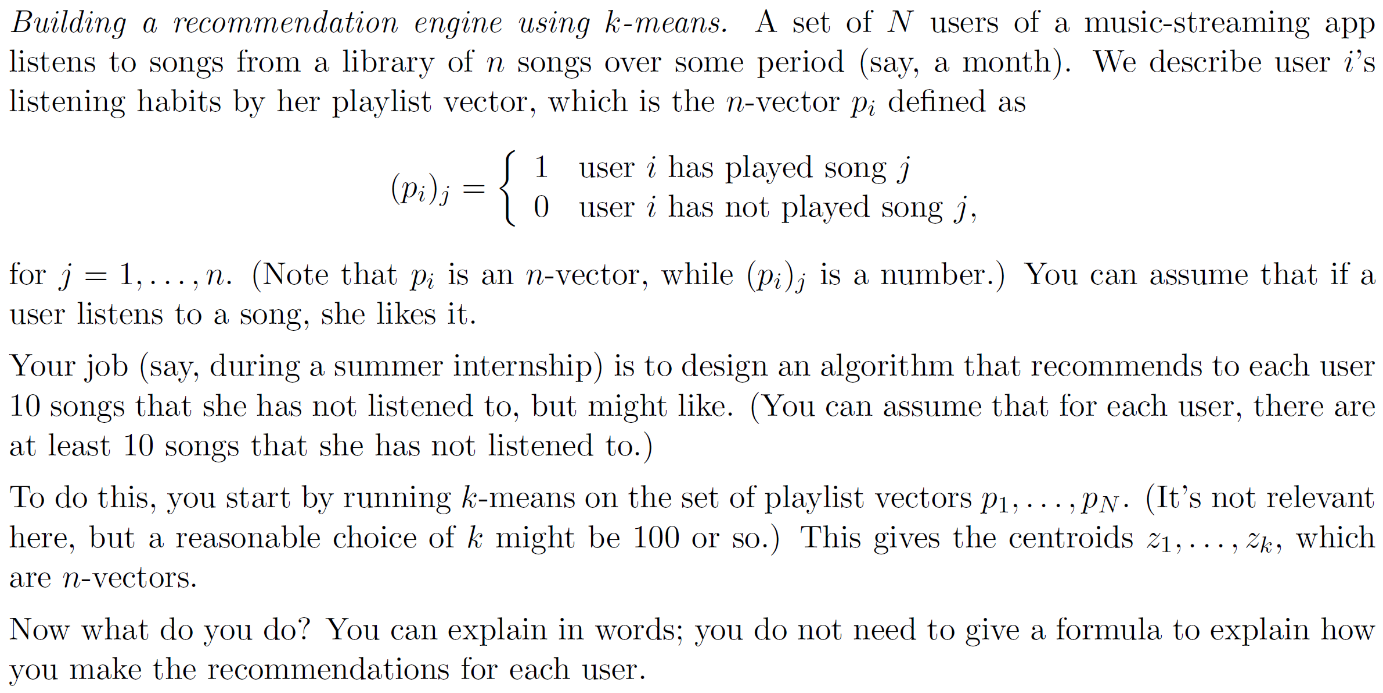
a) Provide a qualitative interpretation of the columns of W.

If we had K = 6 we would have each column for each person, Each line in each column in W is a representation of each food purchase.

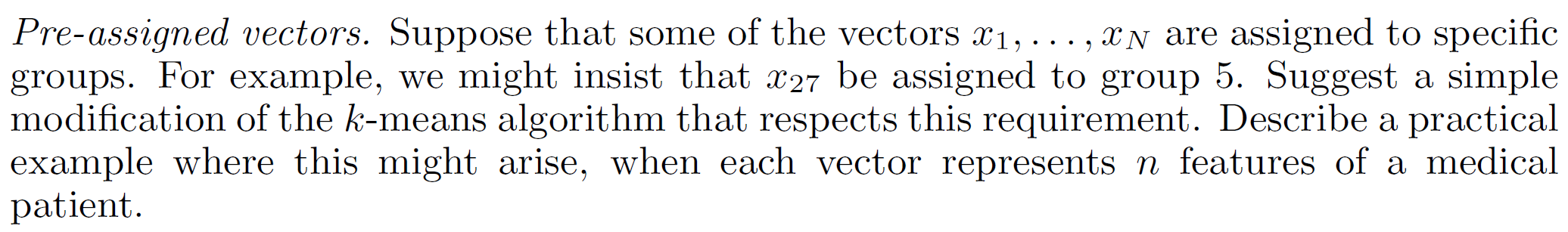
b) Provide a qualitative interpretation of the first two columns of H.

Column 1 is a reference to John and column 2 is a reference to Alice.

2)



First thing to note is that P(i)j = User likes song.  
(1)We start by placing each play’d song on a dataset. If we have k = 100, therefore 100 centroids. (2)We start by placing playlists in each cluster and (3)calculate the mean of all datapoints in each cluster and use the mean as a new centroid. Repeat from step 2 until the clusters do not change anymore.  
Now we can recommend songs to user according to his preference of songs.

3) 

We could use weighted k-means by using the pre-assigned points as a centroid or with more weight so the centroid of the group that it was defined in does not move away from the point in the iteration process.

Example for this is medication that works for most people but we need to place it in a “not useable for this patent cluster” because the medication contains a chemical that the patent this patent is allergic to.

4) Consider ratings data for a group of products in the form of like/dislike. Is non-negative matrix factorization (NMF) a suitable method for this type of data? Explain your answer briefly.

NMF should not be considered as a method for this kind of data, the data with two groups of

(like and dislike) are really binary data.