



## Seventh hands-on exercise (questions)

Hi, this is the last assignment session of the course. As I mentioned in the first class, you will eventually write an 'article' using MySQL other than a 'word' or a 'sentence' by considering MySQL as a language. Also, by writing an 'article', you will spend 90% of the time thinking and 10% of the time coding, such as in a data analytic project. In this assignment session, I try to use a new style of exercise to facilitate you to think before coding.

- Please offer your feedback on the exercise design.

### Suggestions:

1. While you may understand the code (or answer) to solve a problem in e.g., 10 minutes, developing such code may take you hours. Nonetheless, the process of developing such code will offer you more value/benefits than simply reading the answer. The knowledge obtained by reading code fades away easily in comparison to solid memory obtained by solving the problem yourself.
2. Code is a treasure, which can also be an art. Hope you will gain such feelings after completing the assignment.

### Task 1:

- **Problem statement:**

This is actually a problem that I got from a master student who is doing her master's thesis for a large Finnish online (auction) platform that facilitates people to sell/buy used staff (platform name is confidential, but for a possible similar site see Figure 1). Please try to understand the problem first and think about a solution as if you are the owner of the company.

A key problem of such an online marketplace that sells used staff is that **all the used staff is a kind of unique but also the quantity is limited** (See Figure 1). This differs greatly from conventional online shopping websites, such as Amazon, where the volume of products is large enough to generate product-level sale data, which makes it possible to use association analysis (or market basket analysis) to compute the association between different products and thus make recommendations. Nonetheless, for a second-hand product online store, the quantity for each product is limited, making it difficult to build an association based on sales records. As a result, a recommendation system similar to Amazon (see Figure 2) may not work for the platforms selling used staff.

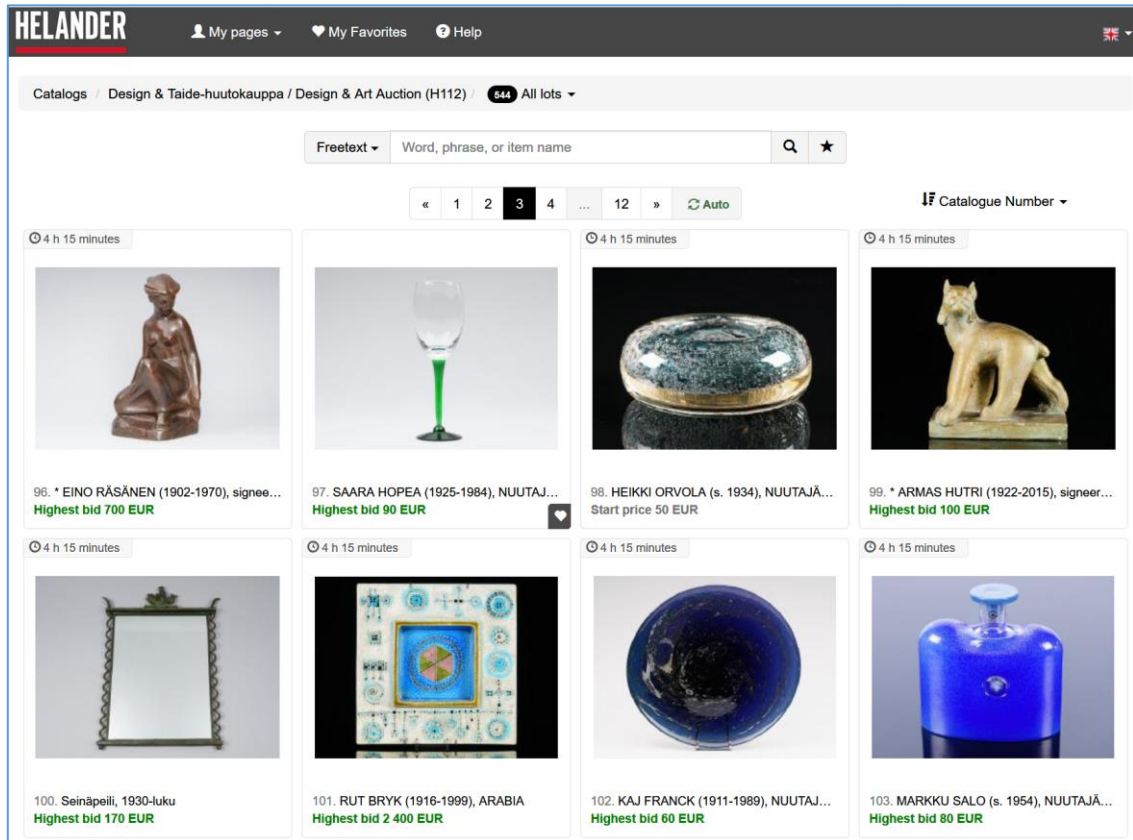


Figure 1. A screen of a Finnish online auction website selling used staff (helander.com)

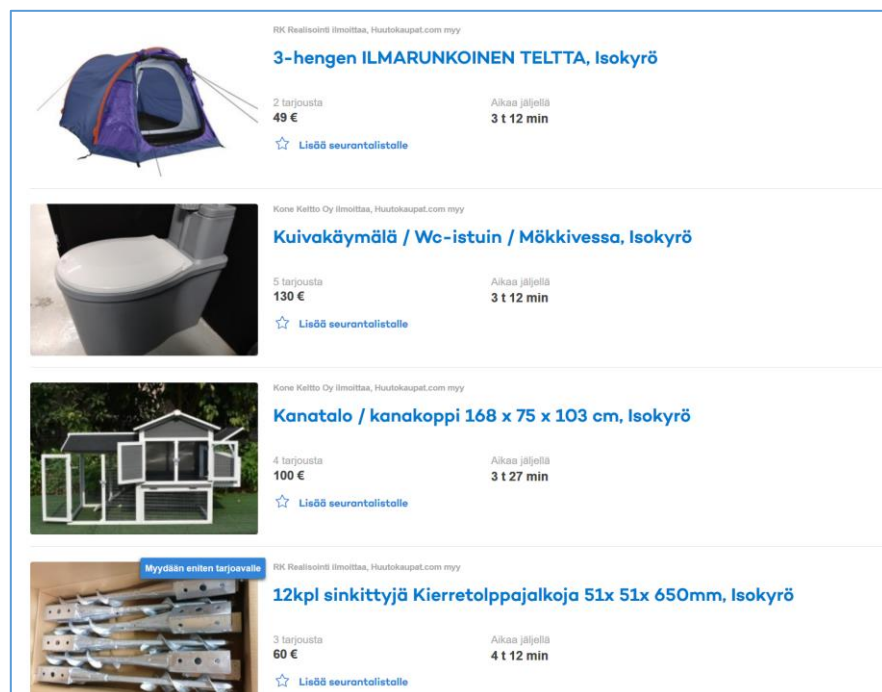


Figure 2. A screen of a Finnish online auction website selling used staff (huutokaupat.com)



Figure 3. Recommendation system of Amazon.

**Question 1:** If a recommendation system like Amazon will not work in this case, can you develop a system to offer recommendations for **registered** users? For instance, you are now recruited as a business analyst to develop a recommendation system for them.

After you developed some thoughts and obtained some preliminary answers, please read a possible answer to this question **at the beginning of the next page of this document**.

# THINK IS FUN

**Answer to Question 1:** The easiest solution is probably to recommend the products of the same category randomly. For instance, when a customer browses a product, e.g., a chair, you can recommend other chairs or furniture to the users. In this vein, when a user wants to sell his/her product on your site, you would request the user to specify the category of the product so that different products can be 'grouped' and then 'recommended'.

Evidently, such recommendations can be very ineffective, because products of the same category can differ a lot. Also, **this does not support cross-category selling** that customers explore other types of possibly interesting products. This is very important because many customers are just aimlessly wandering on the website trying to find something interesting.

**Question 2:** Can you figure out some other design of recommendation systems, if the first idea that you obtained is to recommend similar products of the same category?

- Please first think: For the answer to Question 2, please see Question 3.

**Question 3:** A possible solution to question 2 is to implement a recommendation system using a collaborative filtering method. To understand what the collaborative filtering method is, please read the following materials relevant to TikTok.

## **Collaborative-Filtering:**

This type of recommendation works by comparing you to others. So, if you are new to the app and start liking or rating a video highly (*in another use case*), you will start to see recommended videos. Those videos are the videos that other people like, who also liked the same past videos of you. This type of recommendation is technically referred to as alternating least squares — and matrix factorization.

- **Example:**

If person A likes and comments on videos 1, 2, 3, 4, and 5, and person B likes and comments on videos, 1, 2, and 3, they will tend to like videos 4 and 5. Another way to visualize this example is the following (*X being the recommended video similar to 4 and 5 from above*):

Person A: 1, 2, 3, 4, 5

Person B: 1, 2, 3, X, X

- You can find more information about collaborative filtering by reading from Wikipedia: [https://en.wikipedia.org/wiki/Collaborative\\_filtering](https://en.wikipedia.org/wiki/Collaborative_filtering) (simply a read of the overview section is sufficient).

After reading the materials above, we may obtain an understanding that the collaborative filtering method is **based on identifying users of similar interests**.

**Attention:** if the logic of association analysis (market basket analysis) for hotel analysis (introduced in the lecture) is to identify similar hotels pertinent to customer's interest, could you **use association analysis to implement the logic of the collaborative filtering method** by identifying users always having similar product (hotel) choice. In other words, the identified pair of customers show a strong similarity to the products that they like or purchased.

- Deep thinking is likely needed here.

## **Main task of Question 3 is to identify pairs of users with similar interests:**

- Please use TripAdvisor review data to identify pairs of users with similar interests in choosing a hotel.

**Question 4:** Think about how to apply the answer/results of question 3 for an online auction platform that sells used stuff. (Answer offered in "answer" document)

# THINK IS FUN

**Task 2:** Import “text mining.sql” to the database. Based on the title column of the table, please

- i) compute the association frequency of every possible two words in the title column.
- ii) compute the association frequency of every possible three words in the title column.
- iii) compute the positive and negative emotion scores of each title.
- iv) Your boss asks you a question: how do consumers’ sentiments (emotion or attitude) toward your hotel change with time? How will you answer such a question using data?
  - a. Would the use of the summed absolute value of emotion scores be a good result to report here? If not, what should be a good result?
- v) If your boss asks you how consumers’ attitude to the breakfast or service of the hotel changes with time, how will you answer that question with data?

**Note:**

- Answers to question i, ii, and iii are not provided here, which can be computed based on the codes offered in the lecture slides. Questions iv and v are examples of how such analysis can be extended to offer useful business insights, which can be easier done via “where title like ‘%breakfast%’” function. Thus, specific codes are not offered in the ‘answer’ document.
- Please try to code and answer the questions i, ii, and iii without checking the course materials. As mentioned before, a difficult data analytics project may take 90% of the time for thinking but only 10% of the time for coding. By reading the coding directly, there will be a loss of 90% of ‘thinking’, such ‘thinking’ is important for you being an independent coder.

**Task 3:** Following the lecture slide relevant to “Connect R to MySQL: Building Connection”, connect R to MySQL and test the relevant functions.