Cafeyn Technical Test

As a data engineer, you will often be tasked with putting together scripts that will go through large amounts of data that need to be processed or cleaned or transformed into different blocks of data. In the following you will be asked to solve two exercises, the first will be a coding one and the second will be a general sql exercise.

Part I: Searching for an article

Users of the Cafeyn App are always searching for the name of a magazine and we would like to help them by adding an auto-complete feature that gives them 3 alphabetically sorted choices from the list of all the magazines that are offered by our service.
Write a simple script that would initially take in a list of magazine names from a file, and then write a function that uses the user's search input to return up to 3 alphabetically sorted choices from that list.

For example: If the user types "ma" we should return:

- Mademoiselle Blok
- Magicmaman Famili
- Marianne

If the user types "der":

- Derby Tiercé

If the user types "gard" we return nothing

NB: try to write well documented & clean code as much as possible and explain the complexity O(n) of your code as a whole.

Bonus: Can you describe the limitations(if it has any) of your code? Is it scalable? How could we improve our autocomplete service?

Part II: SQL
We have 3 tables :

issue_consumption:

ID_USER	ID_ISSUE	ID_MAGAZINE

- an id_issue is a serial number identifier of a magazine, i.e: the magazine 88134 has 23 issues ranging from 1 to 23… the magazine 652273 has 100 issues ranging from 1 to 100 etc…)
- an id_magazine is a unique identifier

article_consumption:

ID_USER	ID_ARTICLE	ID_ISSUE	ID_MAGAZINE

- an id_article is a unique identifier

user_session:

ID_USER	SESSION_LOGIN	SESSION_LOGOUT

- session login and logout are timestamps in the “yyyy-mm-DD hh:mm:ss” format
- an id_user is a unique integer identifier that can take the value -1 to represent anonymous users(users that are not logged in to their account)

After studying these tables write the sql queries that would best answer these questions:

- Describe in a few words what you have understood about these tables
- On average, how much time does a user session last?
- Calculate The duration of longest session ever
- Extract the list of unique non anonymous users that have read an issue but have never read any articles.
- How many articles does, a non anonymous user, read on average
- In total, how many issues have been read by anonymous users and logged users respectively
- The maximum number of articles read by a user, the maximum number of times a user has read the same article.
- **Bonus I:**Which is the least read magazine, and which is the magazine that has the least read number of articles
- **Bonus II:** extract the list of (user,magazine) where the user has read at least one issue from that magazine but has never read one article from that same magazine.