Thank you for your time.

Following our interview and our conversation regarding the position of **Data Scientist**, below you can find a number of challenges (5) related to our scope of work where we would appreciate to have your feedback within **3 days**.

The answers should be in proper document/format, accompanied with any relative material (i.e. image, script file etc.) you think it is appropriate for us to understand your way of thinking.

I remain at your disposal in case you need any further information or clarification.

- How could you apply your knowledge of SQL in order to store data acquired from a python script to local or remote database?
- How could a python script take input and provide corresponding output (messages initially printed to *std* or processed data via csv/txt or image output) via webpage?
- How would you approach a task of scraping data from a forum using python?
  - In example, save per entry: username, post message, URL (if any), quoted post (if any), posted time from: **http://www.capital.gr/forum/thread/5491339?page=1**
- How would you approach a problem of tag (class) assignment to texts where there are *n* available tags and each text can get *m* tags?
  - Where: $1 < m < n/2$
  - Already tagged text $N \sim 10n$
  - Average numbers of words per text: 10
  - 70% of texts contain also image (jpg or gif) or video
- How would you approach a task of:
  - Finding similar images to a given one:
    - i. From a local collection - database
    - ii. From the internet
  - Finding websites:
    - i. Similar to a given one
    - ii. Relative to a content (i.e. health) and the main traffic originated from a specific country (i.e. Thailand)