Intern Assignment

Consider you are working for a game analytics company and have been tasked to create a web service where data analysts can upload a csv file and then run analysis on it. As an enterprising intern at that company you decide to solve this problem on your own. Your mission if you chose to accept is to implement the below three parts.

- Api
  - Api to upload csv data
  - Api to analyse data based on filters
- DB
  - Store the csv data.
  - Sqlite or clickhouse
- Deployment
  - Create a docker image so that code/db can be deployed anywhere.
  - Deploy the system to a **free tier of any cloud provider** for analysts to try.

You can use any online resource at your disposal like google, chatgpt for completing the assignment but make sure you understand what you are doing. We would ask you to explain your choices in next rounds.

Judging criteria

You will be judged on 4 main elements in order.

- Completeness of solution.
- Completeness of instructions and documentation you write.
- Code readability and maintainability.
- Performance of the solution.
- Cost effectiveness.

Submission

- Submit **link to the github project**. No zip files in google drive or email would be accepted. **Add shobhit@segwise.ai as collaborator** to your github project. Any documentation or instructions you want to provide should be in a README in the repo itself.
- Also submit the **link to the deployed solution**.

- Please also submit sample api request/response for the apis. Either in README or separately in the email.
- Along with this, also submit the cost of running your system in production 24x7 for 30 days, assuming one file upload and 100 queries a day. You can send your calculations/estimates on the submission email.
- Submission needs to be made by **11 am IST on 8th July.** Sorry but because we want to hire ASAP, we would not be able to extend deadline.

API

- Uploading csv endpoint
  - This endpoint should accept a single csv file link as input. The link will be open to public and does not require any authentication. The file could be as big as 150 MB in size.
  - You should save this file in some format as all future queries will be querying data from this file.
  - The exact api design and how you want to take input params and values is your choice.
  - Columns as well as data types and format in final csv will remain same as the sample csv. Sample csv has 100 rows to give you an idea about the data.
  - sample file: https://docs.google.com/spreadsheets/d/1ShbFMzRUuIJY8amTA58UuEHwsc3UmAnd_LzduBwcBhE/edit?usp=sharing
  - BONUS: if you can add some simple authentication so not everybody can make a call.
- Query endpoint
  - This endpoint should accept parameters with values to query the data from csv. The service should support query by any field in the csv. For numerical field, exact match should work. For string fields, response should contain anything that matches the input as a substring. For eg:
    - For a age, age=20 would mean return all records where age is exactly 20.
    - But for string field like a name, substring match should work. So name=Raj should match Raj, Rajesh, Rajan …. and so on.
    - Date fields should have exact match like numbers.
  - The exact api design and how you want to take input params and values is your choice.
  - BONUS: if you can have a simple UI to query data.
  - BONUS: if you can allow aggregate searches like records where total (x) > 100 or get max/min/mean value of a numerical column etc.
  - BONUS: if you can allow greater than and less than searches for dates.
  - BONUS: if you can add some simple authentication so not everybody can make a call.

DB

- Save data in a sqlite or clickhouse. Clickhouse might help with performance but might need more resources to run. Based on the free limits of the cloud provider you chose, you should decide the DB type.

- You are allowed to use the cloud version of these DBs if you can get a free trial anywhere but remember to cancel it after the assignment review is done.
- You are encouraged to decide and come up with creative data model for best performance and insights given the data.

Deployment

- Create a docker image for your submission. Commit the dockerfile in the repo with rest of the code.
- Deploy your solution on a free tier of any cloud provider. Make sure you are not getting charged for it. There will be no reimbursement or cloud credits provided.

Time is of the essence. Happy building.