PRDT Data Engineer (FT) Test Task

The purpose of this test task is to test your programming skills. We're looking for (a) clean, readable and organized code, that is (b) scalable and (c) follows best practices. Please submit all the deliverables on the checklist below in a .zip file to michelle@prioridata.com. Good luck!

Part A: SQL task

Task Description

1. Using the publicly accessible data base for StackOverflow (https://bigquery.cloud.google.com/dataset/bigquery-public-data:stackoverflow).
   Please write a query to find the 12 StackOverflow users who match the following criteria.
   a. He or she posted a question with tags including "java" or "python" (see the *tags* field in *stackoverflow_posts*; note that the tags field can include more than one tag) that was last active in 2016 (see the *last_activity_date* field).
   b. He or she has a reputation score (see field *reputation*) of at least 400,000.
2. For each of the users identified in *1.*, please calculate the percentage share of favorites (see field *favorite_count*) for each of three tiers defined as follows.
   a. Tier 1: Top 3 questions posted by user ranked in terms of favorite count (*favorite_count*)
   b. Tier 2: Questions 4 to 10 in terms of favorite count
   c. Tier 3: Remaining questions posted by user, if applicable

Deliverables Checklist

- Query for A1
- Query for A2

Example Output for A2:

userid	tier	share
1	Tier 1	0.50
1	Tier 2	0.20
1	Tier 3	0.30
2	Tier 1	0.80

Part B: Python coding task

Task Description

Please write a Python cli command line job to execute the queries in Part A and dump the results of A1 and A2 as tables back into the provided BigQuery project called "candidate-evaluation" under a new dataset. The tables should be called "a1_results" and "a2_results", respectively.

Deliverables checklist

- Python code
- Link to Github
- Link to the dataset holding your results in BigQuery

Things to keep in mind

- Use https://cloud.google.com/bigquery/docs/reference/libraries#client-libraries-usage-python to interact with BigQuery.
- You should set up and test authentication for the provided Google project with id = 'candidate-evaluation' using the gcloud utility and following the Python client documentation. You have been given the roles/bigquery.user access role on the above project.
- Apply best coding practices around testing, code convention, dependency management, naming and failover scenarios.
- Commit small and often to help us follow your thought process from the commit history.