Sirena Data Analyst Assessment

Instructions

Hello there! Thanks for your interest in our Data Analyst position here at Sirena. In order to understand if we are the right match for your skills, we would kindly ask you to complete the following assessment.

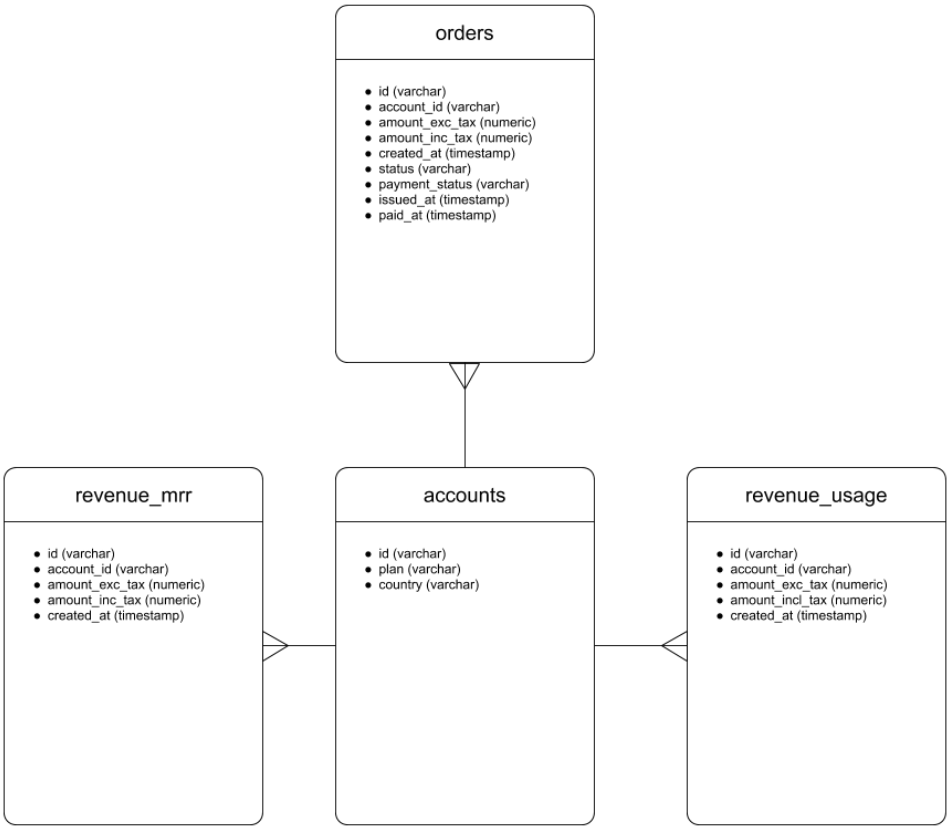
If at any point you have any questions regarding this assessment, feel free to contact your interviewer. They will be happy to answer any questions you may have.

Best of luck!

Databases

1 SQL (PostgreSQL)

For the following questions, please consider the following ERD:



1.1 Write a SQL query to get a list of accounts and include the total historical MRR, total historical usage revenue, and total historical paid (orders with payment status “success”). Your result should look something like this:

account	mrr	usage	paid
Demo Account 1	18432	1093	19025
Demo Account 2	18432	1093	19025
...	...	...	...

1.2 Write a SQL query to get the monthly revenue (MRR and usage) coming from each plan. Your result should look something like this:

month	plan	revenue_type	amount
2020-01-01	PRO	mrr	1453
2020-01-01	PRO	usage	187
2020-01-01	BASIC	mrr	987
2020-01-01	BASIC	usage	10
2020-02-01	PRO	mrr	1453
2020-02-01	PRO	usage	187
2020-02-01	BASIC	mrr	987
2020-02-01	BASIC	usage	10
...	...	...	...

1.3 Write a SQL query to get accounts that have churned (accounts that stopped having a revenune_mrr at some point in time) with their churn date and last MRR. Your result should look something like this:

month	account	amount
2020-01-01	Demo 1	1453
2020-01-01	Demo 2	1453
2020-02-01	Demo 3	1453
...	...	...

Python

2 Python Classes

2.1 In the context of a bigger project, we are in need of creating a tool for logging certain actions, events, and outputs. Usually, we use Python's built-in function **print** to log progress and important variables. However, as the project scaled and increased in complexity, we see the need to be able to have more complex logging functionality. You are asked to build a logging **class** for our Python project following these requirements:

- We must be able to define a "logging level". We would use these logging levels to decide what things to log in development, staging, and production. Additionally, the logging level should be visible in the log output. Here are a few suggestions: DEBUG, INFO, WARNING, ERROR. Think of ways these could be easily extended/modified. The current logging level should be configured through an environment variable and levels should have an order of relevance (if the env variable is set to show INFO logs, DEBUG logs should not be displayed).
- Logging should be as easy as using the **print** function (no complicated function with multiple parameters).
- We should be able to define a file and function where the logging took place. These should be printed in the logs.

This is an example of an expected output of the logs (but feel free to modify if you see fit):

```
[INFO] [file1.py] [function_1] Started to do something
[DEBUG] [file1.py] [function_1] Value: 4
...
```

Build a Python class to support this use case without using any external libraries. Provide examples on how you would use the logging class in a normal workflow (import, any setup you have to do, and log events and variables).

Bonus:

- Write tests
- Write a Jupyter Notebook to show how it works
- Write logs to an output file

Data Analysis

3 Reporting

3.1 In light of recent events, it has become critical to understand the development of the 2019 Novel Coronavirus COVID-19 (2019-nCoV). There is a lot of confusion, misinformation, and general lack of rigorous data analysis in the media about the Coronavirus. In this context, we would like to understand what is going on, what are the risks and what are the consequences of the outbreak. The Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE) has published a Github repository where they share data about the outbreak from a variety of different sources. At the same time, people have started collaborating and building APIs upon this data to make it easy to analyze it.

Using one of these APIs, generate a report about the outbreak of the virus. Feel free to use any tools that you feel comfortable with, but focus on delivering clear and data-driven conclusions.

JHU CSSE dataset: https://github.com/CSSEGISandData/COVID-19
API released by user Kilo59: https://covid-tracker-us.herokuapp.com/

Some tips for the report:

- Feel free to add more data from other datasource (for example weather/temperature data)
- Provide general statistics and describe the data you are using
- Try to find trends over time, patterns, and correlations
- If you find correlations, try to run statistical tests
- Suggest next steps, for example, things you couldn't determine with this data but you would be some other data
- Suggest how you could draw more conclusions