# Glassnode BI Challenge

Our BI Challenge is an opportunity to demonstrate proficiency with problem solving as well as basic data and SQL queries we would expect you to occasionally run in order to generate and report insights to help us make data-driven business decisions.

# How to complete the challenge

### *Submitting a solution*

We do not want to put restrictions on how you approach the challenge, and leave it up to you to make use of the tools that you prefer and are most comfortable with. That being said, we do love Python and give out bonus points if you choose to use it as well 🙂

Please send everything that you want to include in your solution (e.g. answers, results, code, queries, visualizations, etc.) in your preferred format to daria.malanowska@glassnode.com

### *Time*

We respect your time and the challenge is designed in such a way as not to take more than 1-2 hours. We just want to get a sense of your thought process and the way you approach the task at hand. If there are things you don't have time to complete, feel free to simply describe your thoughts and the intended approach.

### *What we care about*

- **Presentation of results**: Does the solution clearly explain the approach, thought process, and solution? How do you communicate your results? Your audience can range from technical to non-technical, can you present in a way that everyone can understand?
- **Correctness**: Does the submission accomplish what was asked? If there is anything missing, does the submission explain why it is missing?
- **Technical aspects**: Does the solution contain code and/or SQL queries? Are those clean and understandable?
- **Choices and creativity**: What was your approach to tackling the questions? What tools did you use? What is the level of creativity/ingenuity of the solution?

### *Questions?*

*Should you have any questions while working on the challenge, please do not hesitate to drop us a line!*

---

# The Task

Please download the following CSV with pseudo user access logs to our platform. Each entry (row) corresponds to a user's access request to one of our many metrics on the platform. In particular, the file contains the following columns:

- **id :** The customer ID, associated with a single customer
- **timestamp:** The time at which the access log was recorded
- **metric:** The path of the metric
- **params**: The query parameters associated with the metric request
  - **a** : The requested asset
  - **c:** The requested currency (USD or native)
  - **i:** The requested interval/frequency
- **studio:** If True, the metric was accessed trough Studio (our web application). If false, the metric was accessed directly via our API
- **plan** : The product plan the user is on (can be A or B describing two different plans)

Given the above data we would like to obtain insights/answers to the following questions:

- **Studio vs API:** Is there a general user preference for accessing our data through Studio vs. API? Is there a difference for plan A users vs plan B users?
- **Top metrics:** What are the **top 5 metrics** for plan A users and the top 5 for plan B users ? Why is simply counting the number of access logs for a particular metric not an ideal approach to quantify this? What would you do instead?
- **API usage:** How many users use the API regularly? How would you define regularity here and apply this to your query/solution? Does API usage generally tend to show particular usage trends with respect to the parameters used, e.g. interval?
- **User churn:** Based on the above data, how many users come to our platform, perform any set of actions within a short time interval, and never come back?

**We are looking forward to receiving your submission!** 🙂