

# Test – Part A – Non-Technical

## *i) What raw input data is required to value the crypto assets?*

As we know, in token economics, according to different needs and different types of cryptocurrencies, there are different valuation methods and techniques which would influence the choice of data. The PoS implementation (e.g., delegated), and the ecosystem (number of validators, reward system, slashing risk) are information that could eventually improve the valuation in certain cases. However, those metrics are not easily available/computable.

In the limited amount of time I have for this task, since I'm currently involved in several recruitment processes, I decided to run the extraction and validation processes of my pipeline, by focusing on three metrics that will allow to operate the valuation of a great range of crypto assets: 1) last price in \$USD, 2) volume of the transactions that happened during a day, and 3) market capitalization at the end of a day. All metrics will be computed per asset and per day.

## *ii) Discuss the logic of your data validation procedures and how you validate input data.*

The purpose of my validation procedure is to check:

- a) the existence of the required data, by first creating a table that contains the required dates and assets, then joining the extracted data to this table. All irrelevant data will be discarded, while missing data will show up in the final table as Null values
- b) the type and range of the information. The idea is to ensure that all gathered data correspond to a defined range (for example, if we want to examine the value of an asset in the last 30 days, the check will guarantee that the data is correctly within the period that we want to study. Or we could also ensure that the metrics are greater than 0). In order to make such range check, I will use Pandera, before the deployment of the dashboard. Once the check is done, the deployed dashboard will be used to perform a final qualitative assessment of the data quality.

## *iii) How would you assess the accuracy of your data?*

I would like to assess the accuracy by comparing the metrics that were computed based on different data sources, which means to operate a cross-validation of the extracted data with information offered by other APIs and exchanges. Flags could be thus generated automatically for mismatching information and potentially resolved automatically, by discarding anomalies.

Such said cross-validation could have different results: 1) if all data provided by different APIs were identical/similar, we could presume the information we got was accurate; 2) if some APIs' data varied from some others', we could choose to trust the figure of the data that mostly appeared in different sources, or continue to execute other analytical exams.

It could also be beneficial to extract data at a higher rate with smaller temporal granularity, and compute aggregates at the semantic layer of a data warehouse, populated with this fine-grained information.