**Guidelines:**

At this stage, you have already got your access to REDASH; the tool which will enable you to create and test your queries directly on our test database to complete the test tasks

IMPORTANT:

Please change your domain from **51.178.128.212 to 51.178.128.212:8080**

This is a 2 parts test, divided on 2 pages. Please read carefully and good luck :)

**1st Part: Data Quality**

1. In this first part, the candidate is invited to inspect data and implement a dashboard with data quality kpis of their choice.

Interpretations and Explanations must be formatted in a report.

The submission of the 1st part must include the following:

a- the choice of the tool used.

b- The choice of different graphs used.

c- Explain your Kpis and the motivation behind your choice.

2. In case there’s some data quality issues encountered, how’d you correct it ? Please respond to this question with a short explanation and an example.

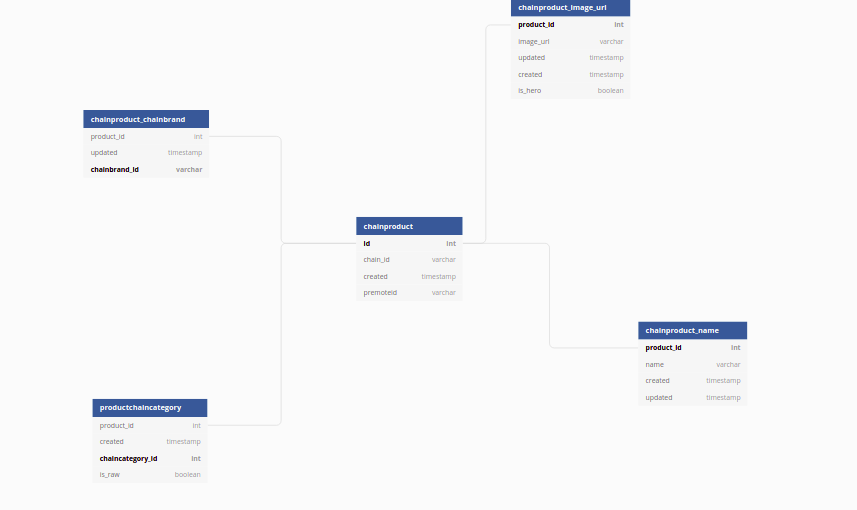
Choose an issue and show us how you'd proceed with solving the problem. **Please don’t copy/paste google answers if you don’t at least take time to rephrase. (We don’t appreciate plagiarism.)**

**--Note**: This part of the test is considered as an open space for the candidate to show their capacities of data extraction and manipulation, the accuracy of their choices and the ability to make concise and comprehensible interpretations.

**2nd Part: Oriented Data analysis.**

The candidates will be tested on their abilities to write SQL queries and their coding skills.

Write a python script to answer the following questions.

**--Database schema**

1 - How many unique categories is there in the table *productchaincategory*

*select count(distinct chaincategory\_id) from productchaincategory; (5386)*

2 - What are the top 5 categories by product ?

select chaincategory\_id, count(product\_id) as nbr from productchaincategory group by chaincategory\_id order by nbr desc limit 5;

3 - Calculate the percentage of products which don’t have an image

select(concat(cast(round(100.0 - (select count (distinct product\_id) from chainproduct\_chainbrand) \* 100.0 / (select count (distinct id) from chainproduct),2) as varchar(10)), '%')); (7.06%)

4 - Calculate the percentage of products which don’t have a brand

select(concat(cast(round(100.0 - (select count (distinct product\_id) from chainproduct\_chainbrand) \* 100.0 / (select count (distinct id) from chainproduct),2) as varchar(10)), '%')); (5.45%)

5 - Calculate the percentage of products which have a brand but don’t have a name

select(concat(cast(round(100.0 - ((select count(distinct chainproduct\_chainbrand.product\_id) from chainproduct\_chainbrand, productchaincategory where productchaincategory.product\_id=chainproduct\_chainbrand.product\_id) \* 100.0 / (select count (distinct product\_id) from chainproduct\_chainbrand)),2) as varchar(10)), '%')); (0.00%)

6 - Calculate the percentage of products which have a brand and which are not associated with any category.

select(concat(cast(round(100.0 - ((select count(distinct chainproduct\_chainbrand.product\_id) from chainproduct\_chainbrand, productchaincategory where productchaincategory.product\_id=chainproduct\_chainbrand.product\_id) \* 100.0 / (select count (distinct product\_id) from chainproduct\_chainbrand)),2) as varchar(10)), '%'));(36.65%)

**--Note**: Please don’t use Colab/Jupyter notebook.

**--Submission**: Once finished you can simply send us an email of your submission to the following email: [imen@datagram.ai](mailto:imen@datagram.ai) Good Luck :)