dadosfera
by datasprints

Technical test in Data Science

Hello candidate. Congratulations on reaching this stage! We look forward to seeing more of your ability to analyze, transform, and process data. Any questions that arise during the process, please feel free to contact recrutamento@dadosfera.ai. Shall we?

Goals

The purpose of this test is to assess your proficiency in the **minimum requirements** for our job opportunity:

- Basic SQL programming;
- Basic Python programming;
- Cloud Computing experience;
- Linux experience;
- Statistics, Machine Learning, Optimization and Data Mining fluency;

Also, **differentiation skills**:

- Ability in analyzing large volumes of data;
- Knowledge in building data pipelines;

Therefore, focus on showing that you **meet our criteria** by answering the minimal requirements using a mix between *Python and SQL*. Leave to cover the differentials in the points raised in bonus items.

Deadline

You will have **5(five) days** to submit this test **from the date this email was sent**. We consider that you have at least 1(one) weekend of effort to complete it. If you want more time, please contact us to set a deadline that best fits us all. :)

About the data

As an community-driven company, we appreciate using Open Data on our analysis. *Therefore, for this test, we want that you download and be creative on your analysis about Open Ocean Data from the Ireland's Digital Ocean Institute*. So you should:

- Download data, with **all the available parameters**, with 1 year period.
    - E.g. Like, if today is Aug/22, you have to get data from Aug/21 to Aug/22, if it is available.

There are TWO different sets of files:

Dataset	***Description***
Tides	Measures about **Tides** collected by several bouys at the Ireland Sea
Waves	Measures about **Waves** collected by several bouys at the Ireland Sea

Minimum requirements

From the data provided, we want you to answer, preferably with graphs, the following questions:

1. What is the lowest temperature of each one of the Bouys?
    a. Which usually month it occurs?
2. Where (lat/long) do we have the biggest water level?
    a. Which usually month it occurs?
3. How the Wave Lenghts correlates with Sea Temperature?
    a. It is possible to predict with accuracy the Wave Lenght, based on the Sea Temperature and the Bouy location?

Bonus items

These are things that we will be glad if you do, and they will definitely set you apart from the other candidates:

- Build a Time Series model that can predict the sea temperature throughout the year
- Build a Data App, using Streamlit or R Shiny showing your results in the best way for a business user (non-tech person)

How to deliver

You must deliver to us 3(three) essential packages:

● ***README.md*** **-** file containing instructions for reproducing your analysis; ● **Analysis.html** - file with the answers to each question from the minimum requirements section and with the additional reviews from the bonus section. When possible, use storytelling techniques to make your thinking clear to us;

- Source code: The codes and queries you used to build your analysis. (extra points for deliveries in repositories such as ***Github*** **or** ***Gitlab***).