DATA SCIENCE TEST

Instructions

1. The candidate must send us the code(s) via github, as well as all additional bases used so that we can reproduce the results

2. The complete reproduction of the code is part of the evaluation. We must be able to reproduce any results you want to show us

3. The visual quality of the generated document, as well as the general format of the presentation of the codes, will also be taken into account.

4. Input files submitted by us cannot contain any changes. All modifications have to be made in code. Thus, scripts that operate on a different basis from the version sent to the candidate.

5. The code has to be self-contained, that is, when executed it must reach the results what was proposed

6. Give preference to deliver the results in Rmarkdown or Jupyter (via git), but please Feel free to develop the work in any language.

7. The test must be done individually, without the assistance of third parties.

8. Feel free to exercise your creativity! Going further and surprising us will be considered a differential!

9. The selected evidence must be presented in the technical interview, when we will evaluate the ability to synthesise, organize and communicate

Question 1.

Descriptive Analysis: Attached, you have received a database (Bases Final ADS Jun2021) with the residential, commercial and industrial energy consumption of each Brazilian region. Do a descriptive analysis of the variables and, eventually, of the relationship between them.

Question 2.

Modeling: Using the variables provided in the Databases Final ADS Jun2021.xlsx, provide a model that projects, as accurately as possible, industrial energy consumption the Southeast region for the next 24 months

1. Explain the method and reason for using the chosen approach in your projection. Which are Can insights be gained from modeling?
2. Provide measures to assess the quality of the model's projection
3. Justify the choice of explanatory variables and assess their explanatory power.

Question 3.

Taking into account the modeling presented above, choose the 5 best models in terms of accuracy and argue why you chose them.

Question 4.

What can you draw conclusions from exercises 1, 2, and 3?