Test Task for Data Science Engineer at Open Mineral

Dear candidate, the goal of the current task is to check your way of thought, analytical thinking and skills, along with the ability to adapt to the new provided data and practical usage of your Data Science skills. **Please, do not spend more than 3 hours of your time working on this project.** We will evaluate your result mainly not on the accuracy of your models but on the approach you choose to take for this task and your Python programming skills.

Input

You are provided with the file called "data.csv". That file contains a sample of real-world data. The structure of the file is the following:

- Date – the date in YYYY-MM-DD format, stating the day of recording the feature's values;
- Columns 1..4 - feature columns;
- Column "target" – target.

Process

We are asking you to:

1. Perform the data cleansing and handling of the missing values.
2. Build an EDA.
3. Apply the statistical, multicollinear, correlational, or lag analysis between the columns to find the possible relationships.
4. Using the information above and your knowledge of algorithms, design or build a system for forecasting the "target" variable 10 consecutive days ahead. The accuracy of the built model will not be evaluated as much as your selected approach for the task. You are not obligated to use all of the data provided.

Kindly reminding not to surpass the time limit of the test assignment. If you have some idea or additional work that should be done, describe it as the next steps after your main code.

Output

We are asking you to build your analysis in the Jupyter Notebook file with Python 3.7 or higher. Please, add comments to your code along with Markdown cells explaining your way of thought. After you finish, create a .zip folder with data, the current file, and the .ipynb notebook file. Then, send it for evaluation.

Thank you,

Open Mineral