Time-series analysis take-home assignment

In this assignment, you have to:

- analyze a collection of candidate feature time-series and assess which among them is/are most useful in forecasting a target time-series, for instance by being a leading indicator of the target series

- explain your methodology and reasoning, making the case that your results are reliable

You have 48 hours from the reception of the assignment to complete it and return the deliverable by email to the following address: hugo@delphia.com.

If you run out of time to complete the assignment, please explain what you would have done next and why.

Have fun!

The data

The archive contains two files:

- **features.parquet** contains different 1000 time-series, each of them consisting of 10,000 daily values
- **target.parquet** contains one time-series consisting of 10,000 daily values

All time-series share a same date index.

The data is entirely numeric and feature column names range from 0 to 999.

The task

This task simulates the work of finding out if some dataset contains one or more time-series that could be used to predict another time-series of interest.

Explain and apply a methodology aimed at identifying which of the candidate feature time-series are the most useful in forecasting the target time-series, then present the results of your analysis. The results should include the column identifier(s) of the time-series you identified.

The deliverable

The deliverable is an archive containing:

- a single Python 3 Jupyter notebook (in .ipynb format) containing all code and documentation of your final solution
- the two data files (**features.parquet** and **target.parquet**) that are included in the assignment archive

There is no need to provide the Python environment to run the notebook. Please just ensure that it runs properly if all 3rd party modules are installed!

Any questions?

You can write hugo@delphia.com if you need clarifications about the assignment.