

Data challenge

- Revo Insurance-

Context

The folder “data” contains data concerning financial indicators of listed companies of the Visegrad Group (Czech Republic, Hungary, Poland, Slovakia) from Q2 2017 to Q1 2021 from 6 different sectors.

Each row includes the country of the company, the sector in which it operates, plus 82 financial indicators (expressed as economic ratios) regarding the company.

The datasets stored in the folder “dimension” are metadata allowing to map the columns contained in dataset.

Tasks

First task

Use a Python Jupyter Notebook to retrieve the data. Then, perform some preprocessing steps: clean the data and do some exploratory analysis to get some evidence and insights.

Hence, answer to the following questions:

- Find the financial indicators that statistically changed between pre and after covid beginning period (pre \leq Q1 2020, after \geq Q2 2020).
- Find, for each sector, the financial indicators that statistically changed between pre and after covid beginning period (pre \leq Q1 2020, after \geq Q2 2020). Rank these financial indicators based on the number of sectors in which they significantly changed.

Second task

Implement a classification machine learning algorithm to predict the sector of the companies (column “S”(“sectors”).

Third task

Make your code production ready! Rewrite all your Python code in a dedicated code base organized by modules, classes, and methods. The code should replicate all the steps you did in the previous tasks.

Also, write some unit tests to be sure your code work as expected. As a bonus point, encapsulate your whole pipeline in a Docker container.

In solving the tasks, try to be as clear as possible in justifying your choices and methods. We are going to evaluate the readability, the simplicity and the correctness of your work.

Good luck!

The REVO DS&AI team!