For this assignment I have chosen the Data Science Salaries 2023 dataset, available at: <https://www.kaggle.com/datasets/arnabchaki/data-science-salaries-2023>. I consider this relevant because it is important at any time during one’s career to know if the salary is right. Since this can be quite complicated, I consider that FCA might help in understanding the current market.

As such, I chosen a subset from the dataset to contain only the job title, position (junior, mid, senior and director), the company size, if remote working is an option, and the salary split in 3 categories: low, medium, high.

The dataset is also available here: <https://fca-tools-bundle.com/view-context/64809a94ef7188e866e399ac> under the name “Data Science Salaries 2023”.

After some more processing, I have generated the context.csv file which I then uploaded on the website. The idea is to obtain jobs titles as objects, company size and salary, and remote status as attributes and job level from junior to senior as the conditions. In this way we can analyze jobs at various levels of experience.

For instance, from the following we can tell that junior ML scientists have low salaries in small companies but can work remote.

A screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated

