
Task

by Dr. Emre Akanak

You are a data scientist who works for ICL¹ and you have request to analyze employee salaries and provide a salary analysis and recommendation. You have received a dataset from company which belongs to Monsanto² which was acquired by Bayer.³

The company investigates the closest competitors' salary policy. You have received direct order from board. The consideration is understanding if there is any discrimination in any group or if any other department or proficiency makes more money.

Your task includes,

- a) What influence salaries and what is the main dominator in salaries?
- b) Is there any difference between departments in salary and budget. The company tries to understand the salary distribution. What is the role of departments in salaries?
- c) As a data scientist what is your analysis, and recommendation?

You are given datasets below:

- employees_data.csv

id: ID of the employee

Birth Year: Birth year of employee

Marital Status: Current marital status of employee

Sex: Sex of employee

Degree: Degree of employee

Profession: Profession of employee

Job: Current job of employee in company

Department: Current department of employee

Years of Education: Total years of education of employee

Years in the Company: Total years in the company

Salary: Salary of employee in 2025

¹ <https://www.icl-group.com>

² <https://www.monsantotechnology.com>

³ <https://www.bayer.com/en/>

Your Task:

As a data scientist you are asked to analyze given data and provide a recommendation regarding your analysis.

Related additional data:

There are several big companies in the industry such as like Nutrien, CF, Yara, Gubretas, Fertiglobe, Haifa Group, EuroChem, K+S AG, BASF, Mosaic, Koch Industries, Helena, etc. You need to understand in the given dataset if the salaries provide any information for employee salary strategy. You need to analyze also department, professions and jobs. You need to analyze your competitor and produce as much as information you can.

Please pay attention to;

1. You have 2 weeks. Task should be sent until 23:59 Wednesday, May 21, 2025,
2. You should use only Python. Any other tool, language and software will not be accepted. You should send the study only in .ipynb file. Any other format (including .py files) will not be considered, (it should be as name_surname.ipynb)
3. You should not send analysis, statement and/or observation via any .docx files. Everything, all your analysis and all your studies should be one solid Jupyter Notebook as .ipynb file. Your analytics and statement should be consistent,
4. You should ignore operational costs (transportation, subcontracting, warehouse, production, material, financial costs, etc.), You have only employees data,
5. It is 250000 row data not very big, but not small and it has been designed to make you familiar to work with industry data. You are also responsible to catch logical issues and if anything does not make sense, you need to add it to report. Drill down into details, look at the bigger picture and drill down again. Try to relate local and global information and try to understand the competitor.

Good luck...

Attachments;

employees_data.csv⁴

⁴ Provided confidentially by company marketing and industry research and strategy department