JOBS AND SALARIES IN DATA SCIENCE

SOURCE:

The source of this data is from Kaggle, which is an online community of data scientists and machine learning practitioners. It is the world's largest data science community with more than 10 million machine learners. Kaggle allows members to explore and find the datasets they want to use, publish datasets, work and network with other data scientists and machine learning engineers around the world.

While this data was originally compiled and collected from the ai-jobs website, transformations were made on the data for better analysis purpose.

CONTENT:

This dataset mainly explains the trends in the employees salaries of different jobs in the data science field from years 2020 – 2023. The variables that were used for defining these trends are the different job titles and the categories that job titles fall into, the size of the organization and location, salary, and country of residence of the employees and also their experience.

DATA PROFILE:

Column Names/ Variables	**Description**	**Time-variant/ invariant**	**Structured/ Unstructured**	**Qualitative/ Quantitative**	**Continuous /Discrete**
work_year	Data recorded year	variant	structured	quantitative	discrete
job_title	specific title of job role	invariant	structured	qualitative	nominal
job_ category	Classification of the job roles	invariant	structured	qualitative	nominal

salary_ currency	Salary paid currency	invariant	structured	qualitative	nominal
salary	Annual gross salary	variant	structured	quantitative	continuous
salary_in_ usd	Salary converted to usd	variant	structured	quantitative	continuous
employee_ residence	country of residence of employee	invariant	structured	qualitative	nominal
experience _level	Experience level of employee	variant	structured	qualitative	ordinal
employment_typ e	Type of employ-ment	variant	structured	qualitative	nominal
work_setting	Workplace/ enviro-nment	variant	structured	qualitative	nominal
company_ location	Country where company is located	invariant	structured	qualitative	nominal
company_size	Size of the employer company	invariant	structured	qualitative	nominal

LIMITATIONS AND ETHICS:

The data was compiled by an independent party and the data was collected from the ai-jobs website, which was collected by individuals, based on internal data such as survey submissions and jobs with open salaries. There are chances for it to have errors when interpreting the results, hence looking into each and every column and reviewing it for data quality and consistency checks would be required before performing further analysis.

As the data that I am going to be using is open-sourced, I don't think there would be an issue regarding data ethics.

QUESTIONS TO EXPLORE:

This dataset could give a clear understanding of the different roles of jobs that are available in the data science field and how the salaries have differed for each and every role, and in different countries across the world from 2020 – 2023.

In order to derive insights and looking for trends and patterns, I would be doing the analysis by looking into the following questions.

- Which are the roles that has seen an increase in the years?
- Which are the roles that has been consistent over the years?
- Is there a difference in salary for a same role in two different countries?
- Would the company location affect the salary structures?
- Based on what factors, there has been an increase in salary for a particular role from 2020 to 2023?
- Which are the most popular and the least popular jobs in the data field?
- Which are the countries with the most number of data science jobs?
- How do the residence of the employee, their experience and work setting make a difference in their salary trend?
- Which are the countries with the highest paid jobs in the data field?