

# Our charts rule!

## Here's our dataset

### Data Science Salaries 2023

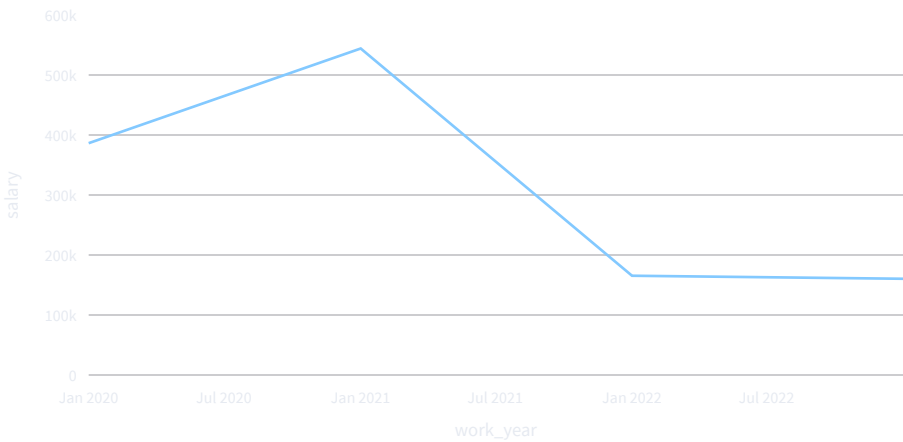
Salaries of different Data Science fields in the DS domain. Below we let you get familiar with the data table. Click on each column name to sort high/low:

	work_year	experience_level	employment_type	job_title	salary	salary_currency
5	2023-01-01 00:00:00	SE	FT	Applied Scientist	222,200	USD
6	2023-01-01 00:00:00	SE	FT	Applied Scientist	136,000	USD
7	2023-01-01 00:00:00	SE	FT	Data Scientist	219,000	USD
8	2023-01-01 00:00:00	SE	FT	Data Scientist	141,000	USD
9	2023-01-01 00:00:00	SE	FT	Data Scientist	147,100	USD
10	2023-01-01 00:00:00	SE	FT	Data Scientist	90,700	USD
11	2023-01-01 00:00:00	SE	FT	Data Analyst	130,000	USD
12	2023-01-01 00:00:00	SE	FT	Data Analyst	100,000	USD
13	2023-01-01 00:00:00	EN	FT	Applied Scientist	213,660	USD
14	2023-01-01 00:00:00	EN	FT	Applied Scientist	130,760	USD

## Next is a line chart

This line chart illustrates the average salary trends of Data Science professionals over the years. The x-axis represents the work years, and the y-axis displays the average salary amount. The chart provides insights into how salaries have evolved from 2020 to now, offering a glimpse of potential salary growth or fluctuations over time. The 'bi-annual' aspect indicates that the data is presented on a twice a year pulse, allowing viewers to observe salary changes over six-month intervals. By analyzing this chart, users can discern patterns in salary progression and make informed decisions related to career development and work experience.

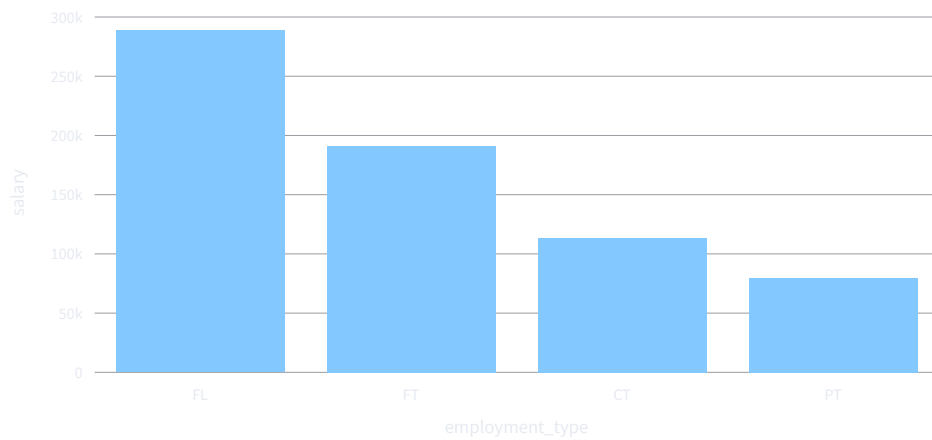
Salary Over Work Time (bi-annual)



## Here is a bar chart

Bars indicate the average salary of Data Science professionals grouped by their employment type. It provides an overview of salary differences between different types of employment, such as Full-Time,

Average Salary by employment type



## Time for interactive participation

This allows you to choose a specific year for the bar chart below, and the code calculates the average salary for FT and PT employee groups.

Select a year:

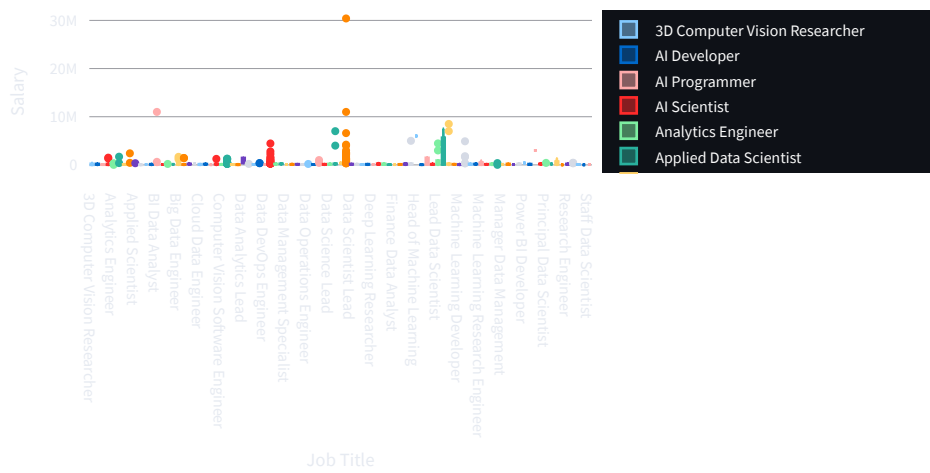
Average Salary by employment type



## Interactive Dot Plot

The result will be a job title plot that visually represents the distribution of average salaries for different job titles. Each bubble represents the salary distribution for a specific job title and allows for a quick visual comparison of salary distributions across different job titles.

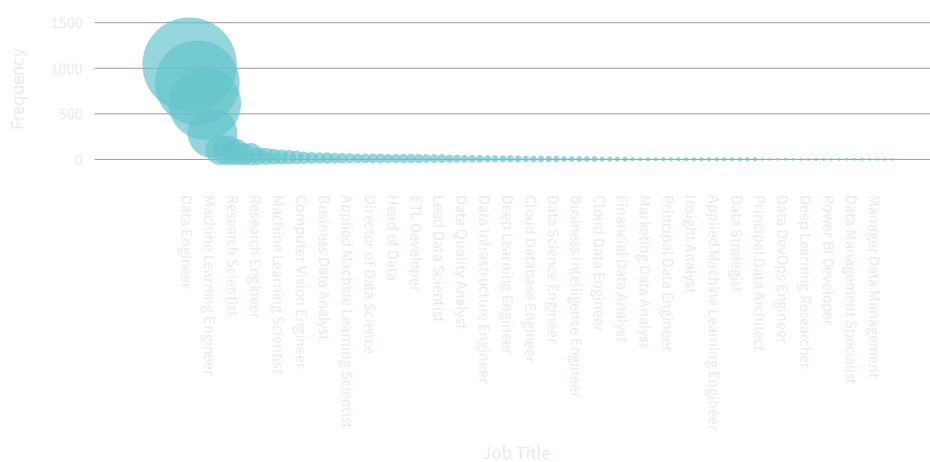
Average Salary Distribution by Job Title



## Specialization frequency (size of the bubble matters)

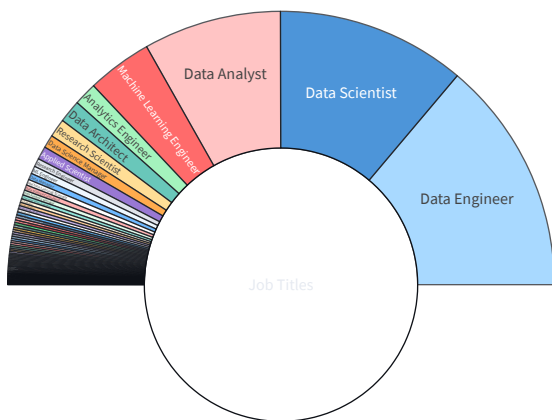
The chart below displays the frequency of each job title in the dataset. Data engineer being the most sought after. It provides a visual representation of the number of occurrences of each job title

Job Title Frequency (Bubble Chart)



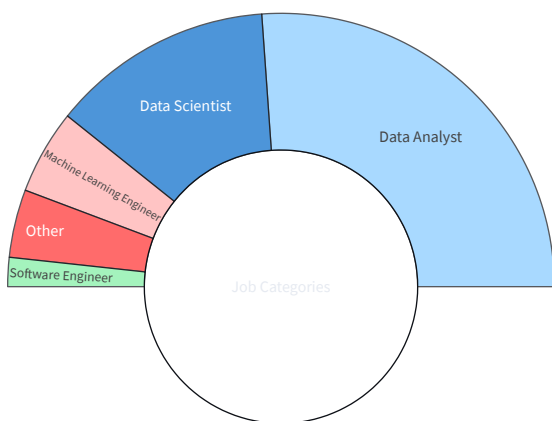
An example of great looking chart but useless as the proportions are unreadable

Job Title Frequency (Tree Chart)



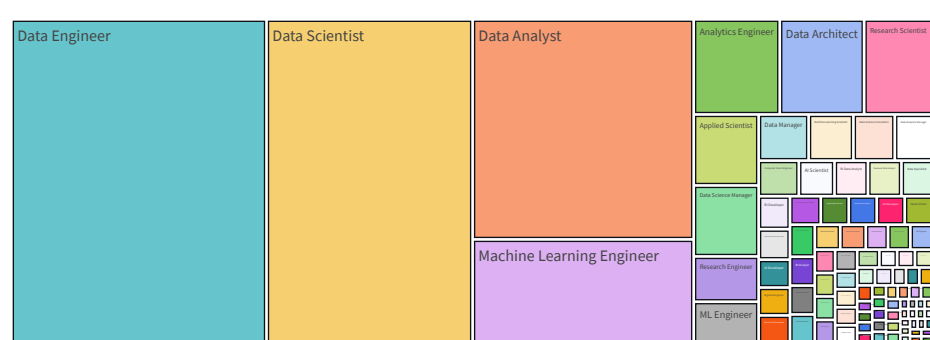
## Fixed with grouping to six categories

Job Category Frequency (Tree Chart)



## And Treemap is a winner (IMHO)

Job Title Frequency (Treemap Chart)



# Now, what about the relationship between salary and remote-work?

Surely, salary and remote-work have a positive correlation indicating a trend of DS professionals working from home.

**ValueError:** could not convert string to float: 'SE'

Traceback:

```
File "/Library/Python/3.8/site-packages/streamlit/runtime/scriptrunner/script_exec(code, module.__dict__)\n\nFile "/Users/nikita_epita/Desktop/EPITA/*S3/Data viz/Final project/Code_DSA_E0\n    correlation_matrix = df[['experience_level', 'salary_in_usd', 'salary', 'r\n\nFile "/Library/Python/3.8/site-packages/pandas/core/frame.py", line 10054, in\n    mat = data.to_numpy(dtype=float, na_value=np.nan, copy=False)\n\nFile "/Library/Python/3.8/site-packages/pandas/core/frame.py", line 1838, in t\n    result = self._mgr.as_array(dtype=dtype, copy=copy, na_value=na_value)\n\nFile "/Library/Python/3.8/site-packages/pandas/core/internals/managers.py", li\n    arr = self._interleave(dtype=dtype, na_value=na_value)\n\nFile "/Library/Python/3.8/site-packages/pandas/core/internals/managers.py", li\n    result[rl.indexer] = arr
```