Evolution of the Salary Average in the Data Science

MOUSSET Maxime - 231xi07x

February 7, 2024

1 Introduction

The goal of this work is to visualize the impact of the experience level on the salary in the data science field. For this, the dataset "Data Science Salaries 2023" was used. This dataset is complete with no missing values and contains a lot of various data about the subject. A lot of different visualization can be done with the variety of data offered. In this paper, the focus is put on the salary average by experience level and year. Experience levels are classified in only 4 categories which is easier to work with than for example job title where there is clearly way more different values to work with.

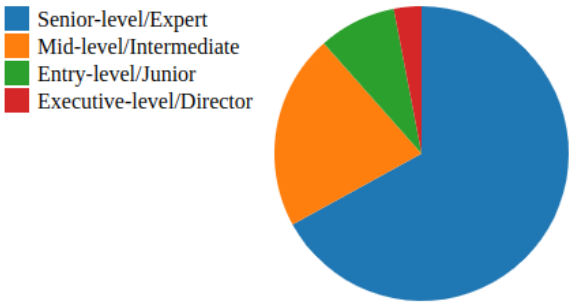
2 Method

The first step to this visualization was to show the distribution of each experience level, the 4 different type that can be found in the dataset are shown in the table below:

Table 1: Experience Level Values

Short Name	Full Name
EN	Entry-level/Junior
EX	Executive-level/Director
MI	Mid-level/Intermediate
SE	Senior-level/Expert

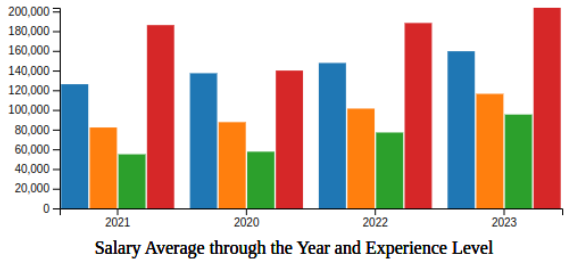
Using a pie chart instead of a bar chart was preferred because the main purpose here was to show the distribution and not really focus on the exact number of occurrence for each experience level.



Distribution of Experience Level in the dataset

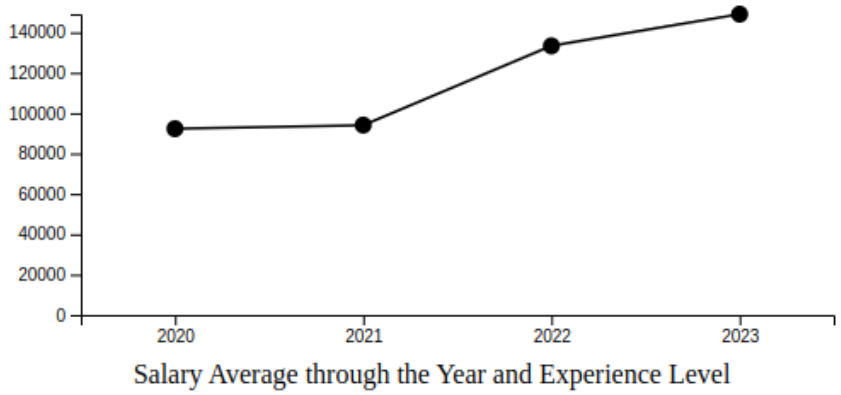
The second step was to visualize the actual salary average in the data science field by experience level and year. The data was first aggregated by year and the by experience level to obtained. The aggregated data was then plotted in a multi bar chart with 4 bar chart for each year to easily analyse the evolution between the years and experience level. The bar chart was preferred because here the value of the average salary is important and can be easily read in this type of chart.

A more precise visualization of the average salary by experience level through the year is also possible when interacting with the pie chart. A click induce changes on the multi bar chart by isolating



only one experience level. A second click on the same part of the pie allow to go back to the initial state.

The last step was to visualize the variation of mean salary for all experience level to get an idea of the trend in general. For this the most suitable seemed to be the line chart. This allow to understand the general trend which is harder by only looking at the multi bar chart.



3 Result

The first chart show that the dataset is mainly composed of senior-level/expert and mid-level/intermediate data. This is surprising that the entry-level/junior is represented this little but as the dataset doesn't provide the way the data collection as been done it's difficult to draw further conclusions.

The second chart shows that, as expected, higher position lead to higher salary. We can see that through the years the gap of salary between each experience level is roughly the same.

The last chart shows that the general trend of salaries in the data science field is increasing.

4 Discussion

Concerning the analyse of the evolution of the salary average in the data science, this visualization is just a first approach by focusing only on the experience level and the year. Even tho it shows the global trend more criterion needs to be analysed to have a more complete understanding of the influence of each criterion.

Concerning the dataset in general, a first improvement would be to strive for more balance between the experience levels, other criterion like employee residence and company location could also benefit

from more various origin. On the opposite side job title for example contains to much different values making it difficult to use.

5 Conclusion

This visualization offer a first approach to the analysis of the evolution of the salary average in data science. We can see that the experience level as of course an effect on the salary but the gap doesn't seem to vary through the years indicating that the salary of the entire field is on a rise, which is confirmed by the global trend from the last chart. We can also conclude that no inequalities based only on the experience level seems to emerge. This page offer more views using the dataset.

6 Reference

Kaggle: https://www.kaggle.com/
Data Science Salaries 2023: https://www.kaggle.com/datasets/arnabchaki/data-science-salaries-2023
EDA on Data Science Salaries: https://www.kaggle.com/code/arnabchaki/eda-on-data-science-salaries4.-Experience-Level-Analysis