**ECONOMIC OF DATA CAREERS:**

Which data you have selected?

# Data Science Job Salaries (Economic of data careers)

It is the perfect dataset to demonstrate some ANOVA tests and learn more about the differences between the three major data careers, Data Science, Engineering, and Analytics groups.

WHAT ANYLYSIS YOU HAVE DONE IN STARTER CODE?

I do anylysis in the form of questions.

* Am I being underpaid?
* Are there data careers that are more lucrative than others?
* What effect does remote work have on salary?
* Is there a difference between the US and the UK?
* What opportunities for advancement should I seek as a data scientist?

WHAT INFORMATION YOU GOT?

Lets start with the list of questions I had in the beginning:

* Am I being underpaid?

.However, the distribution of data science salaries had a standard deviation of approximately 64k.Could I use this as leverage to make more money? Well I also showed that the data science and engineering careers are influenced by seniority level. Am I being underpaid as a data scientist in general? Not necessarily. But as a Sr. Data Scientist? I am currently below the 25th percentile. This could be good evidence for a request for a raise!

* Are there data careers that are more lucrative than others?

This was one of my biggest questions. Are Data Science/Data Engineering/Data Analytics similar in salary? The answer that was consistent throughout the tests was a definite no. The one way anova test showed that data anlytics underperforms the other two by quite a bit. The two way anova showed that experience level had almost no effect on analytics salaries. The interaction plot from the remote vs job title anova test also showed that data science and machine learning engineering are at the top of the pyramid with salary level.

* What effect does remote work have on salary?

It was clear to see that remote work became much more popular in 2021 and 2022. The interaction plots also showed that going remote resulted in higher salaries for data science/engineering/analytics but not necessarily for machine learning engineering. But we also saw from the one way tests that salary levels between remote and in-person were not significantly different in the distributions for both the 2022 isolated and all us data.

* Is there a difference between the US and the UK?

There is not enough data in 2022 to really jump to any conclusions about the differences between the two countries.

* What opportunities for advancement should I seek as a data scientist?

Considering myself a data scientist should definitely not switch to data anlytics. I could probably jump to data engineering and have a similar salary. But by going into either management or machine learning engineering I have the best chance for advancement in pay.

* What comes next?

One thing to keep in mind here is companies can have very wide ranging definitions of what a 'data scientist' or 'data analyst' are. Hence so many unique career names in this data set.see how different clustering algorithms can yield some insight here. Are there attributes in this small dataset that can separate the different careers? Could be interesting!

It is the perfect dataset to demonstrate some ANOVA tests and learn more about the differences between the three major data careers, Data Science, Engineering, and Analytics groups.

LIBRARY:

Some of the library which were used in the coding process.

NumPy :

It is a Python library used for working with arrays. It also has functions for working in domain of linear algebra, fourier transform, and matrices

Matplotlib :

Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python.

Pandas :

It is an open source Python package that is most widely used for data science/data analysis and machine learning tasks.

SciPy :

It is a scientific computation library that uses NumPy underneath. SciPy stands for Scientific Python. It provides more utility functions for optimization, stats and signal processing.

Statsmodels :

It is a Python package that allows users to explore data, estimate statistical models, and perform statistical tests.

Seaborn :

It is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.