**LEARNING FROM DATA ASSIGMENT**

**Please find link to my github account here:**

**https://github.com/gedeon809/learning\_from\_data\_assignment\_two**

**Descriptive statistics of the software developer salaries dataset**

From the output, we can see that the dataset contains information on 1,000 software developers, with a mean salary of $107,088.14 and a standard deviation of $38,759.01. The median salary is $100,000.00. The developers in the dataset range in age from 20 to 70, with a mean age of 37.96 years and a standard deviation of 10.61 years. They also have between 0 and 45 years of experience, with a mean of 14.09 years and a standard deviation of 6.81 years.

Through this assignment, I have gained a better understanding of how to use descriptive statistics in Python using the panda’s library. Compared to Excel, which is a widely used tool for analysing data, Python provides more flexibility, customization, and reproducibility. In Python, I can write code that can be executed repeatedly on any dataset, whereas in Excel, the analysis is often static and limited by the software's capabilities. Additionally, Python allows for more sophisticated statistical analysis and modelling compared to Excel, which can be limited in its ability to perform complex calculations.

In terms of when to use Python versus Excel, I believe it depends on the nature and size of the dataset and the complexity of the analysis. For smaller and simpler datasets, Excel may be sufficient for performing basic descriptive statistics, while for larger and more complex datasets, Python may be necessary. Additionally, Python may be preferred when the analysis requires statistical modelling or machine learning.

If I could do this assignment over, I would try to incorporate more advanced statistical techniques, such as hypothesis testing and regression analysis, to gain a deeper understanding of the data. Additionally, I would try to visualize the data in more interactive and informative ways, such as using interactive plots and dashboards.

As for the gaps in my programming or stats knowledge, I recognize that there is always room for improvement and growth. I would like to further develop my skills in statistical modeling and machine learning, as well as in data visualization and storytelling. Additionally, I believe there is always room for improvement in programming skills, especially in writing efficient and optimized code.

In conclusion, through this assignment, I have gained a deeper understanding of how to use Python and pandas for descriptive statistics. While Python offers more flexibility and sophistication than Excel, both tools have their strengths and weaknesses and should be chosen based on the needs of the analysis. Moving forward, I hope to continue developing my programming and statistics knowledge to become a more skilled data analyst.

# References

Beheshti, N., 2022. The Power of SQL Aggregate Window Functions with Disney Data. 21 January.