**Course Project**

**Milestone 1**

**A Deep Dive into Cybersecurity Salary Dynamics**

**Research goal**

As of writing this, the current focus is to conduct Exploratory data Analysis on the given [Cyber\_salaries.scv data](https://www.kaggle.com/datasets/whenamancodes/infoseccyber-security-salaries/data) obtained from kaggle. so as to provide insights into Cybersecurity Salary Dynamics

**Abstract**

This project aims to explore and analyze the salary distribution of cybersecurity professionals across different demographics and job-related factors. Using a dataset containing information about cybersecurity salaries, including job titles, experience levels, employment types, company sizes, and geographical locations, we investigate various research questions related to predicting salary levels, assessing the impact of remote work and company size, analyzing currency conversion effects, and identifying geographical variations in salaries.

The dataset was first loaded and explored to understand its structure and characteristics. Missing values were checked and handled appropriately. Several visualizations were created to provide insights into the data. The aim is to employ machine learning methodologies so as to uncover underlying patterns and discern predictive models that illuminate the interplay between various factors and cybersecurity salaries.

By harnessing the power of data-driven insights, the findings from this project contribute to a better understanding of the factors influencing cybersecurity salaries and provide valuable insights for both professionals and organizations in the cybersecurity industry.

**Research questions**

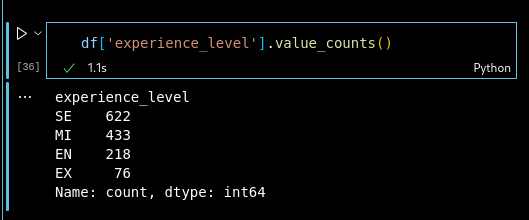
1. **Can machine learning algorithms accurately predict cybersecurity professionals' salary levels based on their experience level, job title, and employment type?**
2. **Is there a discernible difference in the prediction accuracy of remote work impact on salaries when using machine learning models, and does the model perform better for specific employment types?**
3. **Can machine learning effectively identify patterns in the influence of company size on cybersecurity salaries, and is the prediction more accurate for certain company size categories?**
4. **Is there a reliable machine learning approach to determining the impact of currency conversion on cybersecurity salaries, and does the model perform differently for various salary currencies?**
5. **Can machine learning models accurately predict geographical variations in cybersecurity salaries based on employee residence and company location, and is the prediction more precise for certain regions or countries?**

****Introduction****

Cybersecurity professionals play a crucial role in safeguarding digital assets and protecting against cyber threats. Understanding the factors influencing cybersecurity salaries is essential for both professionals seeking career advancement and organizations aiming to attract and retain top talent. In this project, we analyze a dataset containing information about cybersecurity salaries to explore various aspects of salary distribution and identify key factors influencing salary levels.

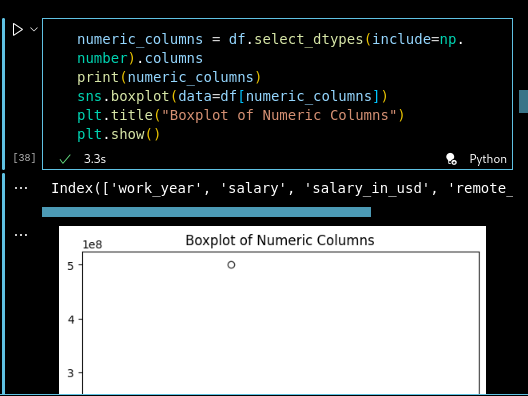
The code snippets to gain more insight on the data were referenced from [Cybersecurity EDA notebook](https://www.kaggle.com/code/nmd2104/cyber-security-eda/notebook) in kaggle.

**Filtering data**

****

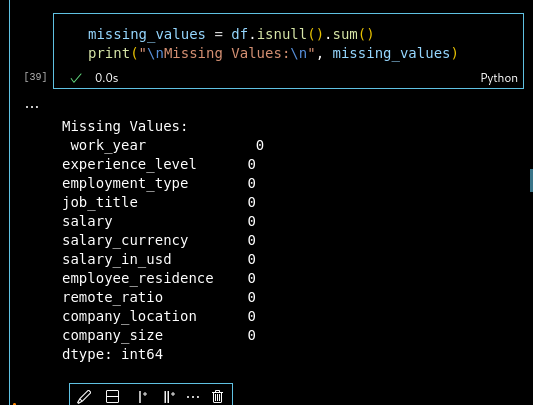
A series of data filtering was performed on various columns so as to understand it fully. The above is just a single snippet of how the data was filtered.

**Checking for outliers**



While trying to plot the outliers in the dataset, the plot came out clean meaning that there were none.

**Checking for Missing values**

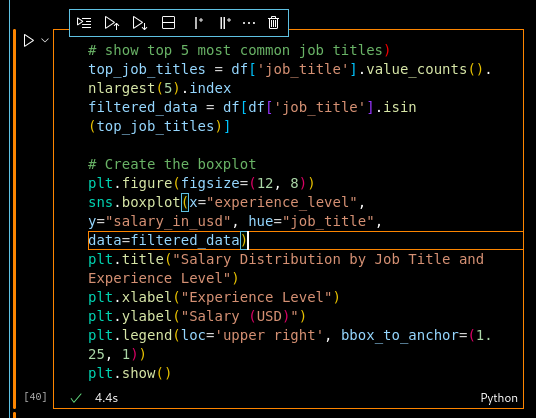
****

Luckily the data didn’t have any missing values so no further cleaning was necessary. Since this milestone involved only Exploring the data and try to understand it, a couple of visualizations were plotted as below.

**Visualizations**

**Salary Distribution by Job Title and Experience Level**

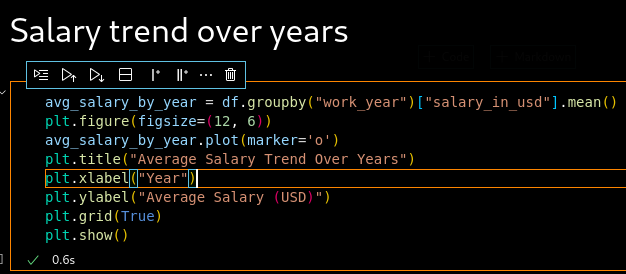
This visualization examines the distribution of salaries across different job titles and experience levels within the cybersecurity domain. By analyzing salary variations based on job roles and career stages, it provides insights into the impact of experience level on salary levels and highlights potential career progression paths within the field.

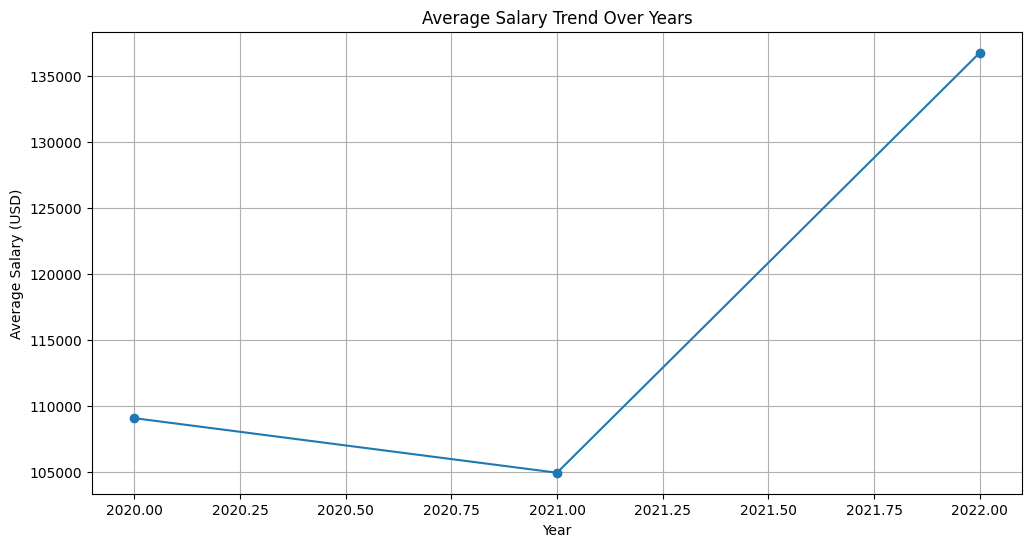




**Average Salary Trend Over Years**

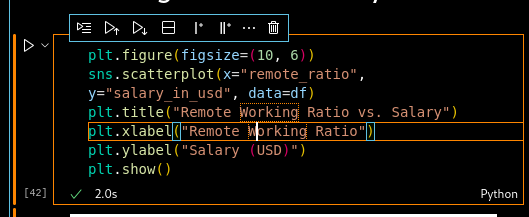
This plot showcases the trend of average cybersecurity salaries over the years, offering insights into salary fluctuations and industry trends. By identifying any notable changes or patterns in salary levels over time, it helps stakeholders understand the evolving dynamics of the cybersecurity job market.

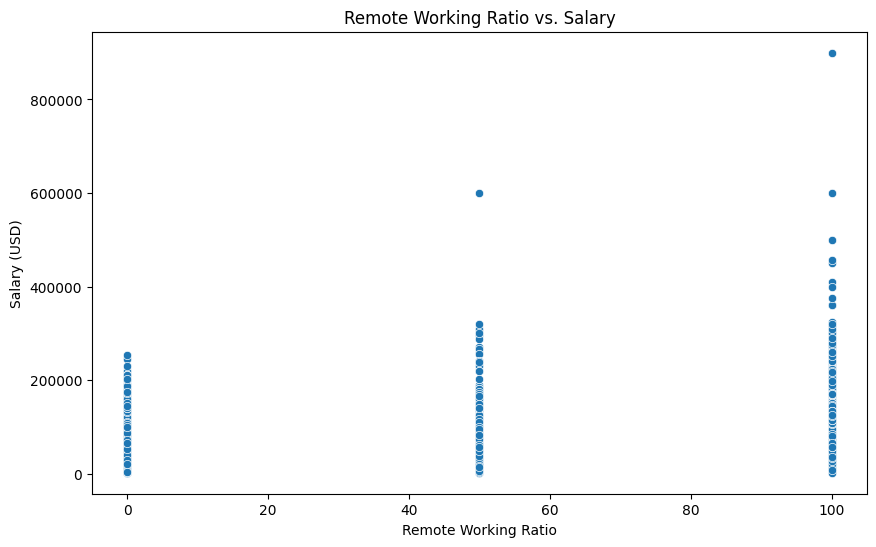




**Remote Working Ratio vs. Salary**

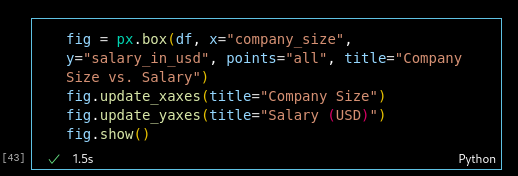
This scatter plot investigates the relationship between remote working ratios and cybersecurity salaries. By examining how remote work flexibility impacts salary levels, it provides insights into the importance of work-life balance and remote work arrangements in shaping salary dynamics within the cybersecurity industry.

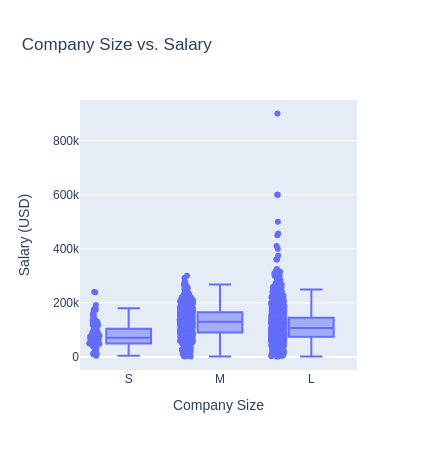




**Company Size vs. Salary**

This box plot compares cybersecurity salaries across different company sizes, shedding light on the influence of organizational scale on salary levels. By analyzing salary distributions within small, medium, and large companies, it offers insights into the role of company size in determining salary competitiveness and career opportunities.





**Geographical Distribution of Salaries**

This visualization presents the geographical distribution of cybersecurity salaries across different countries. By mapping average salaries by country, it highlights regional variations in salary levels and provides insights into the global landscape of cybersecurity salaries.

