Project Report

Project Title:

Cybersecurity Salary Dataset Analysis

Submitted By:

Name: Umed Ali
Student ID: BCB-24F-043
Course: Artificial Intelligence Lab
Instructor: Jatin Kumar
Department: Computer Science
Submission Date: [Date Here]

Cybersecurity Salary Analysis Report

1. Introduction

The purpose of this project is to analyze a real-world dataset related to **cybersecurity job salaries** across different job roles, experience levels, and geographical regions. Using Python, Pandas, and Matplotlib inside Google Colab, this project performs data cleaning, exploration, and visual analysis.
 The key goal is to observe salary trends and patterns that help understand how compensation varies across the cybersecurity field.

2. Dataset Description

The dataset used in this project contains structured information about cybersecurity professionals, including:

Column Name	Description
work_year	Year in which the salary was recorded.
experience_level	Level of experience (e.g., EN = Entry, MI = Mid, SE = Senior, EX = Executive).
employment_type	Full-time, Part-time, Contract, etc.
job_title	Name of the cybersecurity job role.
salary	Salary amount in original currency.
salary_currency	Currency format of the salary.
salary_in_usd	Salary converted to USD.
employee_residence	Country where the employee resides.
remote_ratio	Percentage of work done remotely.
company_location	Location of the hiring company.
company_size	Size of the company (S, M, L).

This dataset is suitable for analytics because it includes key variables that influence salary levels.

3. Tools and Environment

The project was executed in **Google Colab**, using the following Python libraries:

- **pandas** – data manipulation

- **matplotlib** – data visualization

- **numpy** – numeric operations

All dataset operations, cleaning, and graphs were generated inside Google Colab.

4. Data Cleaning & Preparation

4.1 Handling Missing Data

Missing values were checked using:

```
df.isnull().sum()
```

The dataset had **no significant missing values** in key columns such as `job_title`, `experience_level`, or `salary_in_usd`, so no imputation was required.

4.2 Salary Normalization

Because salaries were already provided in USD through the `salary_in_usd` column, this cleaned column was used for all analysis to avoid currency inconsistencies:
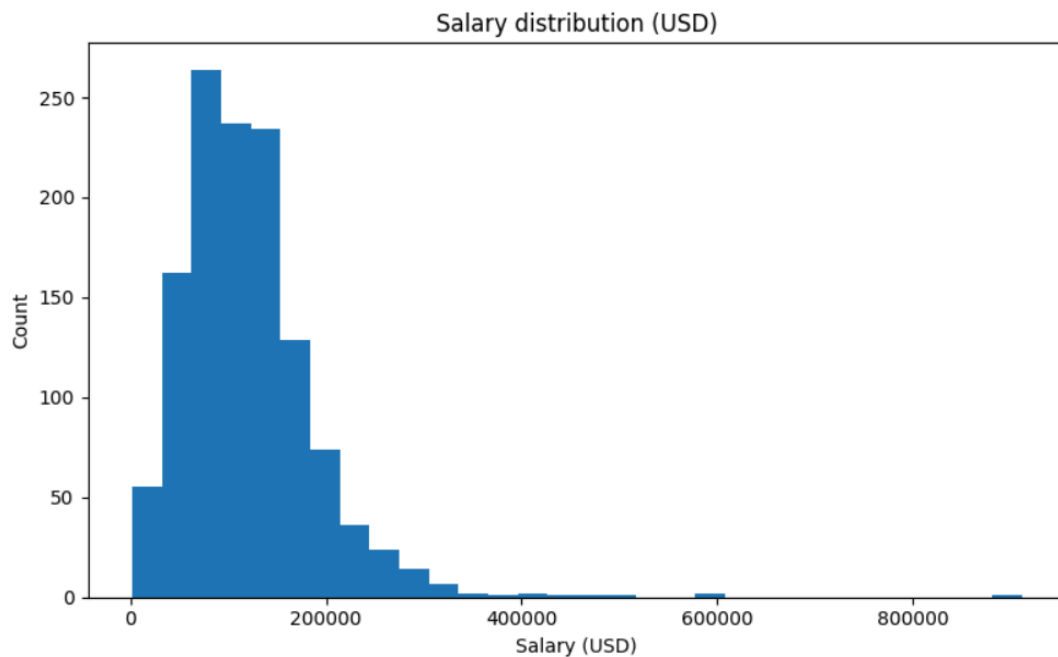
```
df['salary_usd'] = df['salary_in_usd']
```

5. Exploratory Data Analysis (EDA)

5.1 Salary Distribution

Understanding the overall distribution helps identify average earnings and outlier salaries in cybersecurity roles.

Graph 1: Salary Distribution (Histogram)

Salary distribution (USD)

Interpretation:

- The distribution is slightly right-skewed.

- The majority of salaries fall in a typical mid-range.

- A few high-paying roles create a long tail.

5.2 Average Salary by Job Title

This visualization highlights which cybersecurity roles are the highest paying on average.

Graph 2: Top 10 Job Titles by Average Salary


Average salary by job_title (top 10)
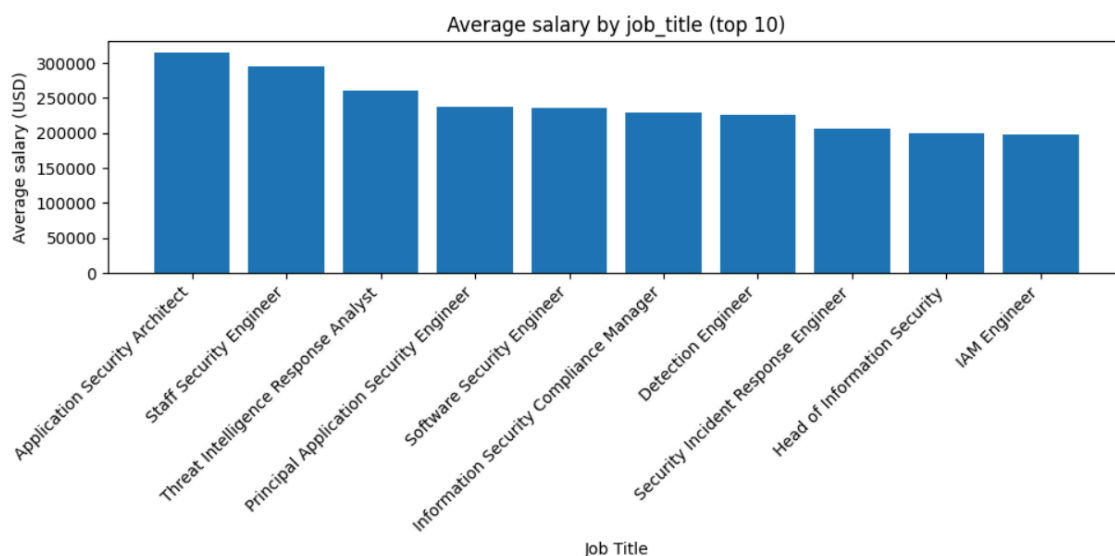
Interpretation:

- Senior and managerial positions appear at the top.

- Specialized technical roles (e.g., Security Engineer) also show strong compensation.

- Entry-level positions cluster toward lower salary ranges.

5.3 Salary by Experience Level

A boxplot clearly shows how salary grows with experience in cybersecurity.

Graph 3: Salary by Experience Level
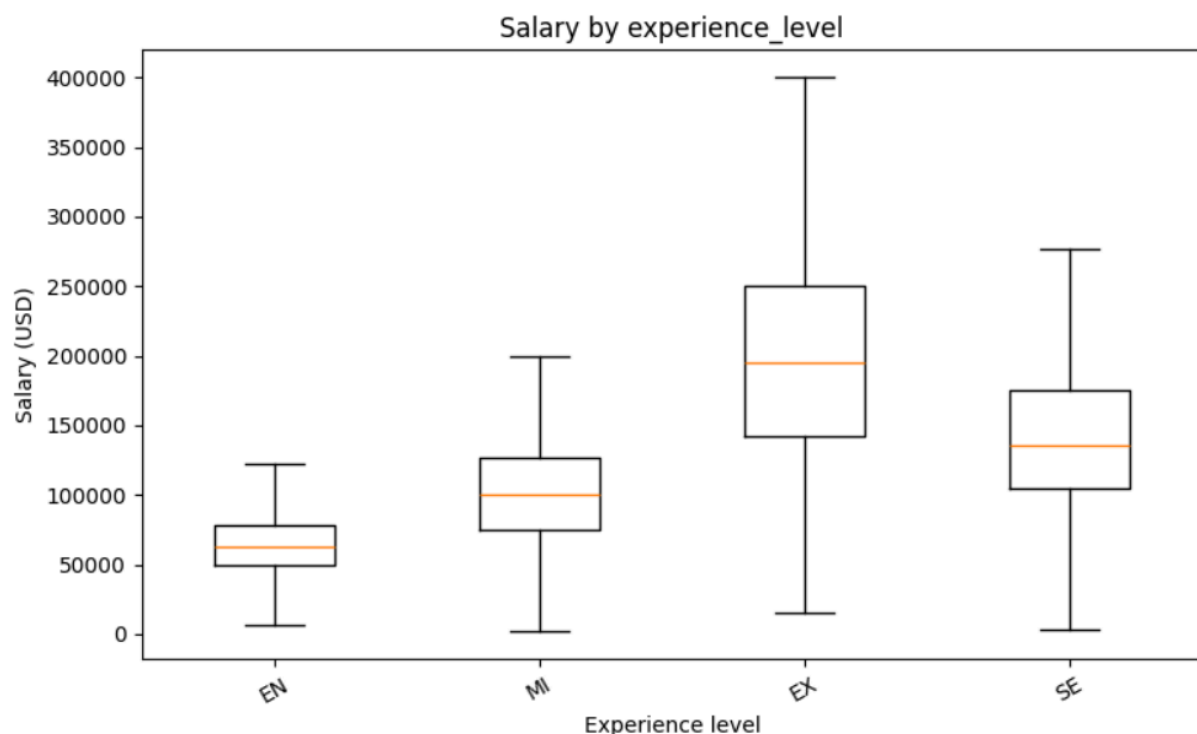
Interpretation:

- **Executive (EX)** and **Senior (SE)** levels have significantly higher medians.

- Entry-level roles show lower and more tightly grouped salaries.

- Higher experience levels show broader salary spread, indicating diverse compensation depending on role or employer.

5.4 Remote Ratio vs Salary

This optional scatter plot shows whether remote work has an impact on salary.

Graph 4: Remote Ratio vs Salary

Remote ratio vs Salary
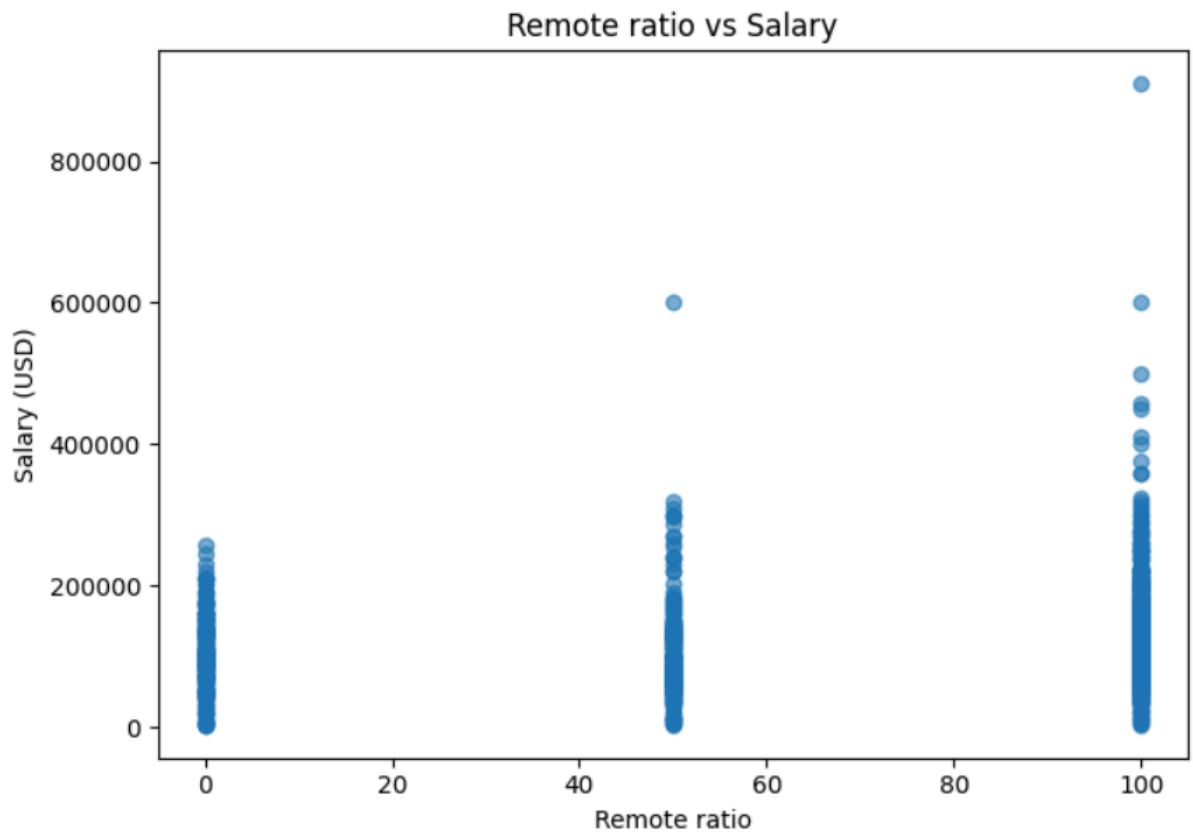
Interpretation:

- There is no strong correlation between remote ratio and salary.

- Both remote and on-site roles offer competitive salaries depending on job type.

6. Key Insights

Based on the analysis performed:

1. **Salary is strongly influenced by job title.**
   Technical leadership, engineering, and specialist roles make up the highest-earning categories.

2. **Experience level has a clear impact.**
   Senior and Executive levels earn significantly more than entry-level roles.

3. **Remote work does not strongly affect salary** (based on dataset trends).
   Cybersecurity allows flexibility, so many remote positions pay competitively.

4. **Salary distribution indicates a wide range of earnings** across roles, with some very high-paying outliers.

7. Conclusion

This analysis shows clear salary patterns in cybersecurity:

- High demand leads to competitive pay.

- Experience and job specialization greatly increase salary.

- Cybersecurity remains a strong, well-paying field globally.

This project demonstrates data cleaning, analysis, and visualization skills using Python and Google Colab. The insights gained can help students and professionals understand the salary landscape in cybersecurity roles.

8. References

- Dataset: Cybersecurity Salaries Dataset (keggle.com)

- Tools: Google Colab, Pandas, Matplotlib, Python 3