SLIDE 1 - INTRODUCTION

Hello, we are Team M and today we will be discussing our security data analysis project.

SLIDE 2 - INTIAL OBSERVATIONS

Our initial observations were to explore the given VERIS dataset, to do so, we used a python package called VERISPY. This allowed us to extract VERIS JSON objects and put them into a Pandas DataFrame structure.

We used the validated dataset from the VERIS community database, which contains over 8000 cyber incidents from different organizations. The data we decided to use from the dataset is mainly from the victim demographics, as we found that it was data-rich and we were quite interested in the attacked organizations, as well as this, the timeline column provided us with a unit that we could use to compare the different organizations. This also helped us form our question.

SLIDE 3 - DATABASE

The investigation was within areas thought to be essential for organizations, this included revenue, number of employees and the country of origin. We looked at the average discovery time in each of these areas, but one important thing to note here is that some data was excluded. The data excluded was the larger units, this was because the rounding of larger units skewed the data. Not only this, the lower units: seconds and minutes, these were also excluded since only a few instances had values while most did not.

SLIDE 4 – OUR QUESTION

We have decided to investigate factors that affect the response time of cyber incidents within different organizations. As a result of our observation, the question we have chosen to analyze is **"What factors best predict an organisation's response time to a cyber incident?"**

SLIDE 5 - SIZE

Firstly, we started off by looking organization sizes, we looked at eight different sizes ranging from only 1-10 employees to organizations with over 100,000. The best performing was size2, with a range of 11-100 employees that had an average discovery time of 44 days, while the worst performing was size8 with over 100,000 employees who had an average discovery time of 89 days.

SLIDE 6 - REVENUE

Moving onto revenue we looked at 3 different sizes, here the best performance came in at an average of 19 days from high revenue organizations, and the worst performance came from medium revenue organizations.

SLIDE 7 - INDUSTRY

When looking into the different industries we categorized the industries by the first 2 numbers in the NAICS code. One thing that should be noted is the performance of the healthcare sector, as it is somewhat misrepresented and is improved drastically when excluding years. Looking over the data, we can see that the information sector had the best average performance of 26 days while the retail sector had the worst average performance at 70 days.

SLIDE 8 – COUNTRY ANALYSIS

Finally, we looked at the country of origin, here we can see 6 of these (referring to the slides). We can clearly see that the slowest was Germany with an average of 111 days discovery time and the fastest was Japan with an average of 17 days

SLIDE 9 – STANDARD DEVIATION ANALYSIS

With the data collected, we next had to compare the different factors. We decided to look at the spread of the data. The factor with the highest spread would be the one which most influenced Time to Discovery. We used the averages for each factor to calculate the standard deviation.
To accurately compare the values, we first had to calculate the mean of the averages for each factor, and then use this to calculate the proportional standard deviation.
From these calculations, we can see that the standard deviation of Countries is by far the largest. It can be determined then, the Country an organization is in, has more impact on the Time to Discovery than its Industry, Revenue or Size.

Slide 10 – FUTHER ANALYSIS

To establish a deeper understanding of the standard deviation results we investigated further factors in the best and the worst performing countries to see what underlying factors might be causing the difference in performance.
We looked at the actions, employee count and industry columns as they were data rich. The countries we looked at was Japan, USA, and Germany, where Japan was the best performing, USA having average performance and Germany the worst.
Looking at the actions, both Japan and Germany had hacking as their most common incident, with 42% and 46% respectively, while the US had 19%, so the actions are unlikely to be an affecting cause.

SLIDE 11 – FUTHER ANALYSIS EMPLOYEE COUNT

Further analyzing this, the number of employees is something that we thought could be an important factor, as people tend to be a security threat towards organizations. However, we could see that Germany had 53% of their organizations with an employee count of 1000 employees or less, while Japan only had 9%.

SLIDE 12 – FURTHER ANALYSIS INDUSTRY

Finally, we looked at the percentage of industries within the different countries, once again Japan and Germany were quite similar with the most common industry being the information services sector, with 43 % and 50 % respectively. Since they are very similar, another high % industry will unlikely have a big effect on the response time.

SLIDE 13 – CONCLUSION

Although, there are no direct subareas within countries that could be connected to as why certain countries performed better than others, it should be noted that there are more subjective areas such as business cultures that could potentially play a part in the performance difference.

SLIDE 14 – MATTHIAS SCHULZE

This is something that Matthias Schulze, at the German Institute for International and Security Affairs said to raconteur. He said: “When it comes to the manufacturing sector, Germany is very different to the United States or UK. Its businesses are very hierarchical and steeped in tradition. This means many of them have until recently been skeptical towards digital innovation. It’s also very difficult to integrate cybersecurity awareness and training in this rigid structure, and therefore the German manufacturing sector, which includes many small and medium-sized businesses (SMEs), is exceptionally vulnerable to attack by opportunist cybercriminals.”

Another thing that should be highlighted is legislation, this could potentially be a factor why certain countries do better than others, as the there is a difference in how countries handles their cyber security legislation.

In conclusion, even though it is likely that there is more than one factor that contributes to the response time of an organization, our investigation has shown that the country seems to be the biggest factor.