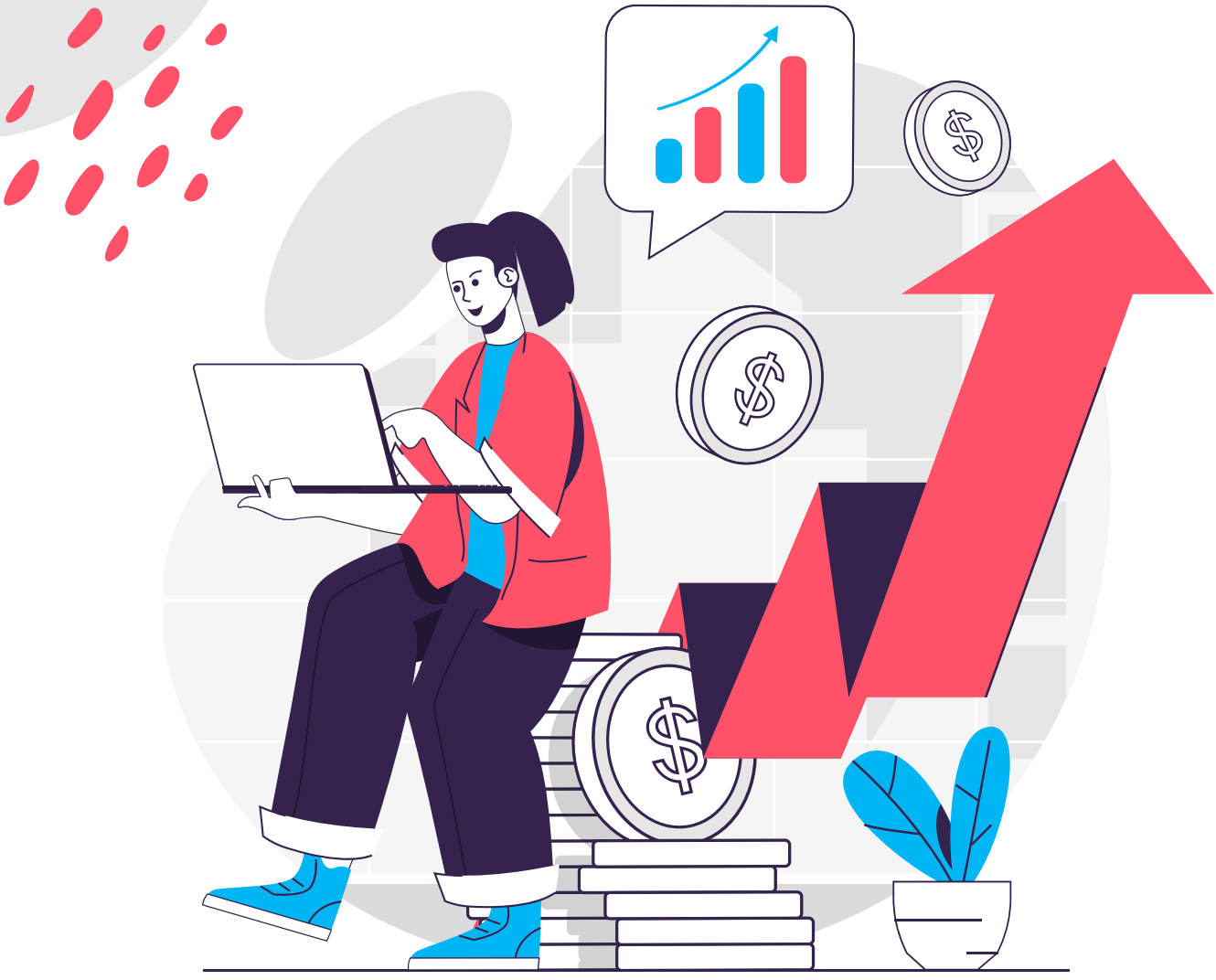
CAIRO
UNIVERSITY

DATA SCIENCE PROJECT

PROPOSAL



TEAM MEMBERS

Donia Abdelfattah
Sec: 01 BN: 28

Raghad Khaled
Sec: 01 BN: 30

Menna Alla Ahmed
Sec: 02 BN: 29

Nada El-Sayed
Sec: 02 BN: 32

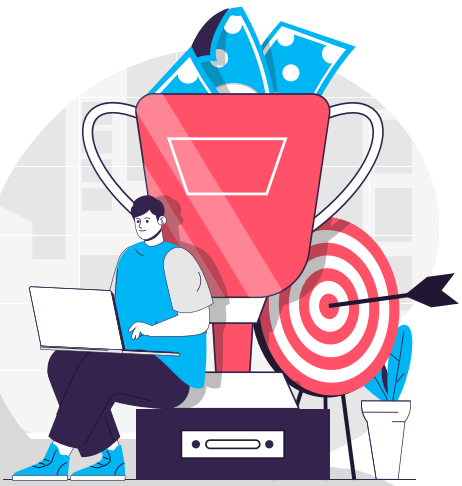
Our Customer

Our client, Saluslab, is a security company that has developed a tool to protect networks and companies from potential attackers.

SALUSLAB

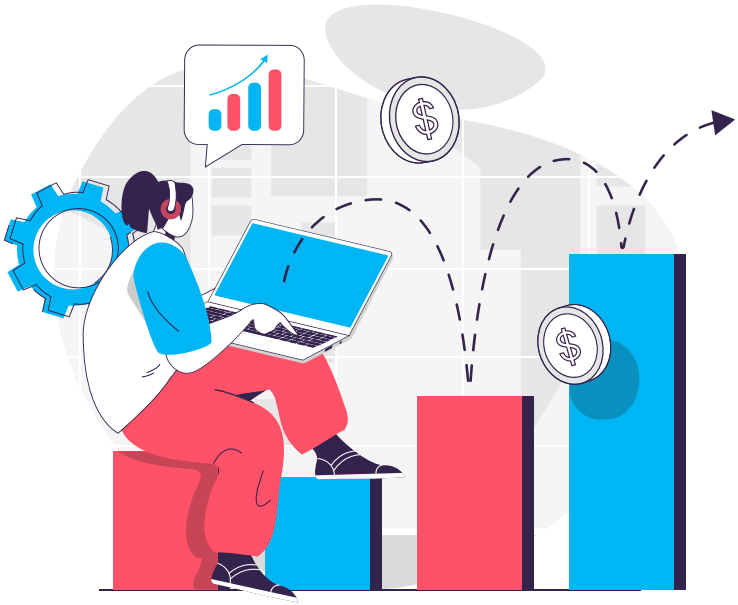
What We Do

As part of our marketing strategy, we are considering various approaches, including gathering data on past attacks to identify patterns and predict the likelihood of future attacks. By leveraging this information, we can target companies that may benefit from Saluslab's security tool and offer them a solution to mitigate potential security risks.



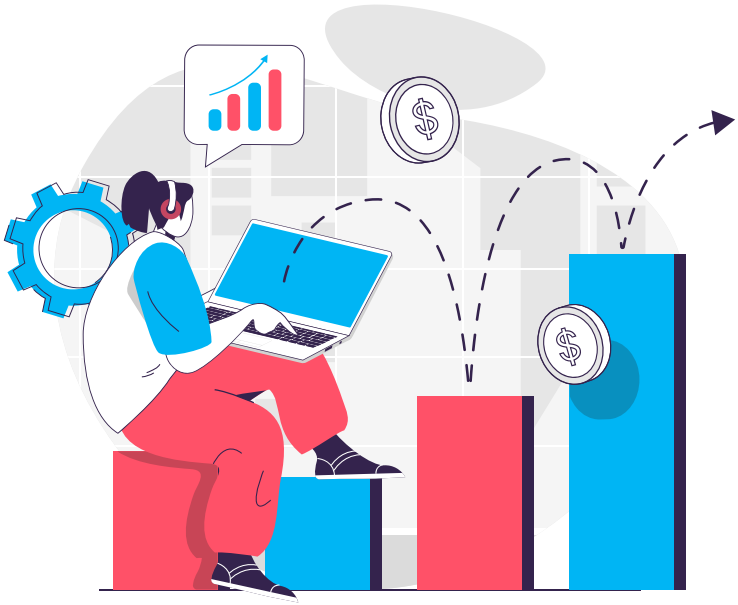
Main Questions

1. What are the sectors which attackers target regularly?
2. Do attackers target companies that lie in specific geographical regions?
3. Do the attackers target the companies with higher revenue more than those with lower revenue?
4. What is the relation between the gang and the company's year of establishment?
5. What affects the ransomware amount? Company revenue, data criticality, or neither?
6. What is the probability of the data being published given information about the company?
7. What is the interesting region for each gang?



More Questions (Optional)

1. If the company had already been attacked once, what is the probability for it to get attacked again?
2. What are the gangs' patterns in attacking the companies? How often do they repeat the attacks?
3. What types of data are most commonly leaked on the dark web, and how do these trends change over time?
4. What payment methods do these sites accept, and how have these methods changed over time?
5. In which sectors and geographical regions do the companies that have the most viewers of their leaked data exist?



Our Work Plan

Data Collection

First of all, we will collect data using web scraping techniques to gather data from ransomware websites on the deep web and companies' websites and also using finance APIs.

Answering Questions

1. What are the sectors which attackers target regularly?

- To answer this question we need to collect attacks data then from the company description we should extract its sector (maybe using NLP techniques or using ChatGPT-API, or BingChat-API, etc), then compute the most sectors that had been attacked.

2. Do attackers target companies that lie in specific geographical regions?

- In this question, we will follow the same approach as the first question but to find the region of the company.

3. Do the attackers target the companies with higher revenue more than those with lower revenue?

- In this question, we will collect data about revenue for the companies using finance API, then we will compute the relation between the company revenue and the number of attacks.

Answering Questions (cont.)

4. What is the relation between the gang and the company's year of establishment?
   - In this question, we will collect data about gangs and find out if gangs are interested in the startups.
5. What affects the ransomware amount? Company revenue, data criticality, or neither?
   - This question will need to build a machine learning model to determine the relation between ransom amount and other factors
6. What is the probability of the data being published given information about the company?
   - This question needs to build a machine learning model to determine the probability of publishing the date and company revenue, year of establishment, sector, region, and ransomware amount.
7. What is the interesting region for each gang?
   - In this question, we will collect data about each gang and the region to get the relationship between the gang and the region of interest for its attack
8. If the company had already been attacked once, what is the probability for it to get attacked again?
   - This will be a probabilistic model that will represent how the company learns from its previous mistakes

Answering Questions (cont.)

9. What are the gangs' patterns in attacking the companies? How often do they repeat the attacks?

- we will focus on this question on the year slot when gangs repeat attacks and when they have almost zero attacks and we will try to analyze the output with the time of known holidays

10. What types of data are most commonly leaked on the dark web, and how do these trends change over time?

- We scrape data from various dark websites that contain data leaks.
- We then identify the types of data that have been leaked and categorize them into groups such as personal data, financial data, etc.
- Finally, we analyze the frequency of each category of data leak over time to identify trends and patterns.

11. What payment methods do these sites accept, and how have these methods changed over time?

- First, we collect data about payment methods that are advertised on ransomware sites, including cryptocurrency methods.
- Then, we analyze these data to identify trends in these payment methods and if the popularity of the payment methods has changed over time

Answering Questions (cont.)

12. In which sectors and geographical regions do the companies that have the most viewers of their leaked data exist?

- We scrape data from various dark web sites that contain data leaks and get the number of views of the data of each attacked website
- we get the companies with the highest number of viewers and then see if they have a common sector/region or if there exists any other common pattern between them.