사이버 침해 및 유형 분석

Python 데이터 분석 및 시각화 프로젝트

Bisang

팀장 :
팀원1

팀원 :
팀원2
팀원3
정석진

목차

프로젝트 목표

1 Python 활용 능력 및 문제 해결 능력 향상

2 데이터 가공, 데이터 시각화 방법 학습

3 팀 프로젝트를 통한 협업 능력 향상

데이터 가공

데이터 출처

1. KISA 한국인터넷진흥원 - 2023년 하반기 사이버 위협 동향 보고서
https://www.boho.or.kr

2. KISA 한국인터넷진흥원 - 2024 사이버 보안 위협 분석과 전망 보고서
https://www.boho.or.kr

3. KISA 한국인터넷진흥원 - 랜섬웨어 대응 가이드(23년 개정본) 보고서
https://www.boho.or.kr

4. KISIA 한국정보보호산업협회 - 2024 사이버 보안 위협 분석과 전망 보고서
https://www.kisia.or.kr

5. 조선일보 - “제조기업 해킹 늘어... 한국, 조심해야”
https://www.chosun.com

6. 조선일보 - 사이버 침해사고 신고 4년간 2배 증가
https://www.chosun.com

7. 보안뉴스 - 정부가 예측하는 2024년 사이버 보안 위협 4대 키워드는?
https://m.boannews.com

8. DATANET - “기업화되는 사이버 범죄, 인텔리전스로 대응한다”
https://www.datanet.co.kr

데이터 가공

자료 조사 과정

데이터 수집

랜섬웨어 공격을 받은 뒤,
랜섬웨어에 대한 관심을 갖고
최근 사이버 공격율과
공격 유형을 파악하기 위해
보안 관련 보고서를 통해
데이터를 수집

데이터 탐색

수집한 데이터를 기간별,
유형별로 신고 건수를
다양한 방법으로 시각화

데이터 분석

1.기간별 사이버 공격
신고 빈도의 상관관계 분석
2.랜섬웨어의 공격 유형 분석
3.사이버 침해 발생 신고와
보안 업체 증가율의
상관관계 분석

인사이트 발견

사이버 침해가 발생하더라도
업무가 중단이 되지 않도록
백업 체계를 마련하고,
신속한 복구 프로세스를
반복적으로 점검하고
강화해야 한다는 결과 도출

데이터 가공

자료 조사 과정

침해사고 신고 통계

과학기술정보통신부(한국인터넷진흥원)은 「정보통신망 이용촉진 및 정보보호 등에 관한 법률」제48조의 3(침해사고 신고 등)에 따라 민간분야의 정보통신서비스 제공자로부터 침해사고 신고를 받고 있다. 2023년 침해사고 신고 통계를 살펴보면 2022년 1,142건에서 2023년 1,277건으로 전년대비 약 12% 증가하였으며, 2021년부터 2023년까지 반기별 침해사고 신고 현황을 살펴보면 2021년 상반기 298건/ 하반기 342건, 2022년 상반기 473건/ 하반기 669건이며, 2023년 상반기 664건/ 하반기 613건의 침해사고 신고가 있었다. 2023년 상반기 침해사고 신고 건수는 664건으로 전년대비 40% 증가하였다.

[단위 : 건수]

구분 \ 연도	2021년		2022년		2023년	
	상반기	하반기	상반기	하반기	상반기	하반기
건수	298	342	473	669	664	613
합계	640		1,142		1,277	

표 1-1 침해사고 신고 현황

	A	B	C	D	E	F	G	H	I	J	K
1	연도	firsrHalf	secondHalf								
2	2021	298	342								
3	2022	473	669								
4	2023	664	613								
5											
6	통계표명:	사이버 침해사고 신고 현황									
7	단위:	건수									
8	출처:	과학기술정보통신부(한국인터넷진흥원) - 사이버 위협 동향 보고서(2023년 하반기)									
9		https://www.boho.or.kr/kr/bbs/view.do?searchCnd=&bbsId=B0000127&searchWrd=&menuNo=205021&p									

사이버 침해사고 보고서를 바탕으로 CSV 파일 직접 작성

데이터 가공

에러 원인 및 해결 방안

```
ax1.bar(li_x_year,li_y_revenue, alpha=0.3,  label='revenue')
ax2 = ax1.twinx()
ax2.plot(li_x_year, li_y_growthRate, 'o--', label='growthRate', color='red')
```

라인 그래프를 그리는 과정에서 x, y 축 개수가 맞지 않아 에러 발생

```
ValueError: x and y must have same first dimension, but have shapes (3,) and (2,)
```

연도	정보보안 [	정보보안
2020	3921387	-
2021	4549734	16
2022	5617174	23.5

막대그래프의 경우 x, y 축의 개수가 동일하지만, 라인 그래프의 경우 y 축 개수가 1개 부족

```
ax2.plot(li_x_year[1:3]
```

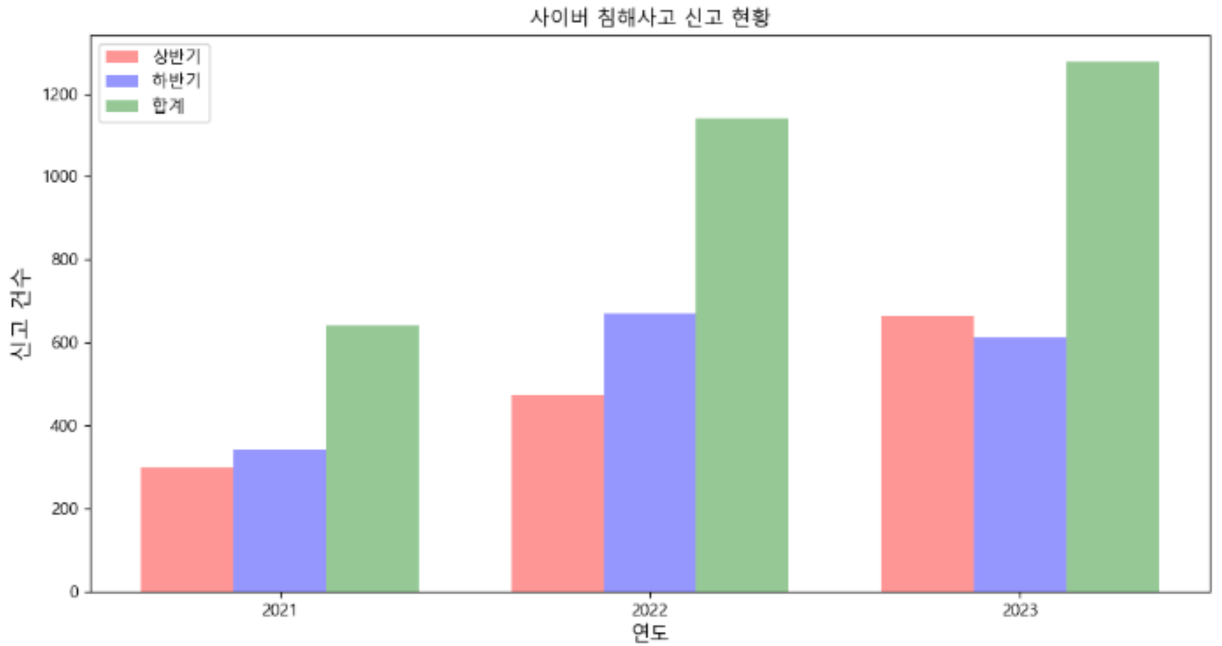
x, y 축 개수가 일치하도록 코드를 수정해 해결

데이터 시각화

데이터 시각화 종류와 선택 이유

다중 막대그래프

2021년 ~ 2023년 사이버 침해사고 신고 현황



그래프 특징

다중 막대그래프는 여러 범주의 데이터를 비교하는 데 유용한 그래프로,
데이터를 직관적으로 이해할 수 있도록 도와줌
각 범주 내에서 여러 항목의 값을 비교할 수 있도록 도와주며, 비교 항목이 많을 때 유용함

선택 이유

연도별 상반기, 하반기 신고 건수와 연간 신고 건수를 같이 비교하기 위해 사용

데이터 시각화

데이터 시각화 코드 ▶ 다중 막대그래프

```
year = []          # 연도 데이터를 저장할 리스트
firstHalf = []     # 상반기 데이터를 저장할 리스트
secondHalf = []    # 하반기 데이터를 저장할 리스트
total = []         # 상반기와 하반기 합계를 저장할 리스트
```

그래프로 출력하기 위한
데이터를 저장해줄
데이터 리스트 선언

```
    year.append(int(row[0]))                    # 연도 데이터를 연도 리스트에 추가
    firstHalf.append(int(row[1]))               # 상반기 데이터를 상반기 리스트에 추가
    secondHalf.append(int(row[2]))              # 하반기 데이터를 하반기 리스트에 추가
    total.append(int(row[1]) + int(row[2])) # 상반기와 하반기의 합계를 합계 리스트에 추가

# 데이터 프레임 생성 (연도를 인덱스로 설정)
df = pd.DataFrame({'firstHalf' : firstHalf, 'secondHalf' : secondHalf, 'total' : total}, index = year)
```

CSV 파일에서
필요한 데이터만
추출 및 연산 후
DataFrame 에 저장

```
# x축 인덱스 생성 (연도 수에 따라 동적으로 생성)
index = np.arange(len(year))

# 각 연도별로 3개의 바(bar)를 순서대로 나타내는 과정, 각 그래프는 0.25의 간격을 두고 그림
# 첫 번째 바(상반기)의 위치와 속성을 설정
b1 = plt.bar(index, df['firstHalf'], bar_width, alpha=0.4, color='red', label='firstHalf')

# 두 번째 바(하반기)의 위치와 속성을 설정, 첫 번째 바의 오른쪽으로 0.25만큼 이동
b2 = plt.bar(index + bar_width, df['secondHalf'], bar_width, alpha=0.4, color='blue', label='secondHalf')

# 세 번째 바(합계)의 위치와 속성을 설정, 두 번째 바의 오른쪽으로 0.2만큼 더 이동
b3 = plt.bar(index + 2 * bar_width, df['total'], bar_width, alpha=0.4, color='green', label='total')
```

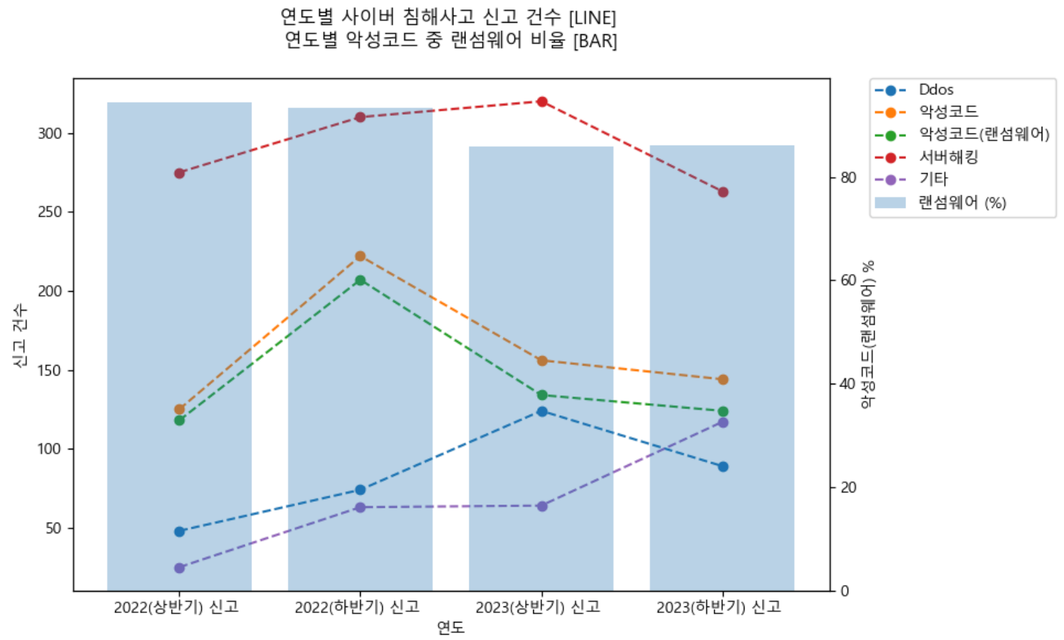
matplotlib 을 사용해
막대그래프를
그리기 위한 코드

데이터 시각화

데이터 시각화 종류와 선택 이유

라인그래프 & 막대그래프

유형별 사이버 침해사고 신고 현황



그래프 특징

라인그래프는 데이터 포인트를 선으로 연결해 데이터 변화를 시각적으로 나타내는 그래프로,
주로 시간 경과에 따른 데이터 변화(주식 가격, 날씨 데이터 등)를 보여줄 때 사용
막대그래프는 데이터값을 막대로 나타내는 그래프 형태로, 주로 범주형 데이터를 비교할 때 사용

선택 이유

라인그래프는 사이버 침해사고를 유형별 발생 추이 및 변화를 파악하기 쉽도록 사용
막대그래프는 악성코드에서 랜섬웨어가 차지하는 비율을 한눈에 파악하기 쉽도록 사용
서로 다른 데이터를 동시에 분석하기 위해 두 개의 그래프를 같이 사용

데이터 시각화

데이터 시각화 코드 ▶ 라인그래프 & 막대그래프

```
# 그래프에 사용할 데이터 리스트들 선언
li_x = []  # x축 데이터 (연도)
li_y_Ddos = []  # DdoS 공격 건수
li_y_Malware = []  # 악성코드 건수
li_y_Malransomware = []  # 랜섬웨어(악성코드) 건수
li_y_ServerHacking = []  # 서버 해킹 건수
li_y_ETC = []  # 기타 사이버 범죄 건수
li_y_RansomwarePercent = []  # 랜섬웨어가 전체 악성코드에서 차지하는 비율
```

그래프로 출력하기 위한
데이터를 저장해줄
데이터 리스트 선언

```
for i in range(len(header)):
    if i % 2 == 1:  # 홀수 번째 열(연도 데이터)을 가져옴
        li_x.append(header[i][0:13])  # 연도 데이터의 첫 13자를 리스트에 추가
#print(li_x)
for row in data:  # 나머지 데이터를 행 단위로 읽음
    if row[-1] == '':  # 마지막 열이 빈 경우 반복을 종료
        break
    category = row[0]  # 첫 번째 열의 데이터(카테고리)를 저장
    #print(category)  # 카테고리를 출력
    if category == 'DdoS공격':  # 카테고리가 'DdoS공격'인 경우
        for i in range(len(row)):
            if i % 2 == 1:  # 홀수 번째 열의 데이터를 정수로 변환하여 리스트에 추가
                li_y_Ddos.append(int(row[i]))
```

CSV 파일에서
필요한 데이터만
추출해 저장

```
# 두 개의 y축을 갖는 그래프 생성
fig, ax1 = plt.subplots()
# 각 데이터 시리즈를 선 그래프로 그림
ax1.plot(li_x, li_y_Ddos, 'o--', label='Ddos')
ax1.plot(li_x, li_y_Malware, 'o--', label='Malware')
ax1.plot(li_x, li_y_Malransomware, 'o--', label='Malware(Ransomware)')
ax1.plot(li_x, li_y_ServerHacking, 'o--', label='ServerHacking')
ax1.plot(li_x, li_y_ETC, 'o--', label='ETC')
# 두 번째 y축에 랜섬웨어 비율을 막대 그래프로 그림
ax2 = ax1.twinx()
ax2.bar(li_x, li_y_RansomwarePercent, alpha=0.3, label='Ransomware %')
```

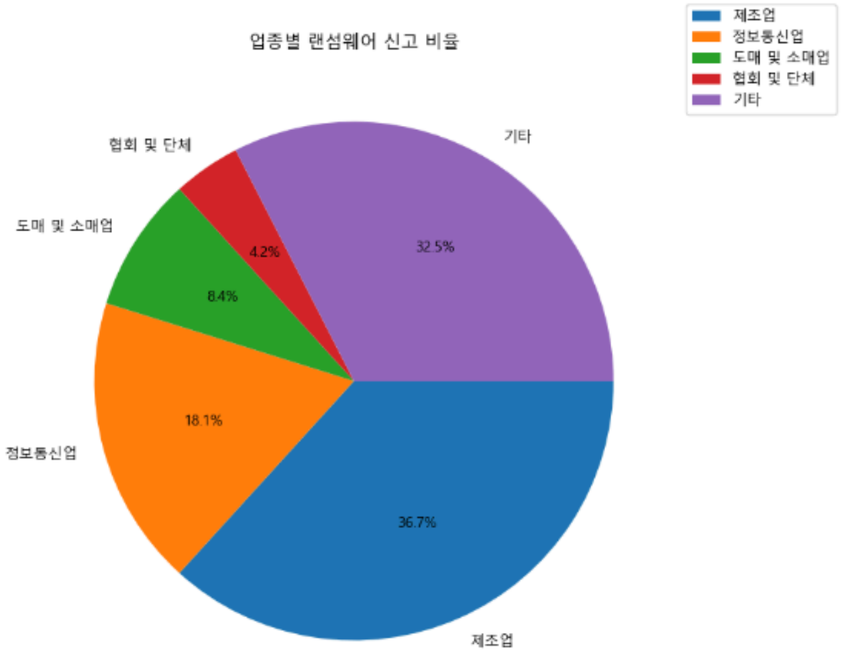
matplotlib 를 사용해
y 축의 범위가 다른 두가지 그래프를
라인과 막대그래프로 그리기 위한 코드

데이터 시각화

데이터 시각화 종류와 선택 이유

파이그래프

업종별 랜섬웨어 신고 비율



그래프 특징

파이그래프는 원형으로 나뉜 부분들의 비율을 시각적으로 보여주는 그래프로,
각 부분이 전체에서 차지하는 비율을 쉽게 이해할 수 있도록 도와줌
주로 비율을 중시하는 데이터(시장 점유율, 예산 배분 등)를 시각화 할 때 사용

선택 이유

업종별 랜섬웨어 신고 비율을 효과적으로 보여주기 위해 사용

데이터 시각화

데이터 시각화 코드 ▶ 파이그래프

```
import pandas as pd
import matplotlib.pyplot as plt
import csv

# 데이터 추출
reportPer = []
sectors = []
with open('bisang5.csv','r',encoding='UTF-8') as file:
    data=csv.reader(file)
    header=next(data)

    for row in data:
        if row[-1]=='':
            break
        reportPer.append(float(row[-1]))
        sectors.append(row[-2])

# 파이 차트를 위한 explode 설정
explode = [0.05] * len(reportPer)  # 데이터 포인트 수에 맞게 explode 리스트의 길이 설정

# 파이 차트 그리기
plt.title('Ransomware report rate by industry')
plt.pie(reportPer, labels=sectors, autopct='%.1f%%', counterclock=False) # 원형 그래프 설정
plt.legend(loc='lower left',bbox_to_anchor=(1,0.9)) # 범례 위치 설정
plt.show()
```

그래프로 출력하기 위한
데이터를 저장해줄
데이터 리스트 선언

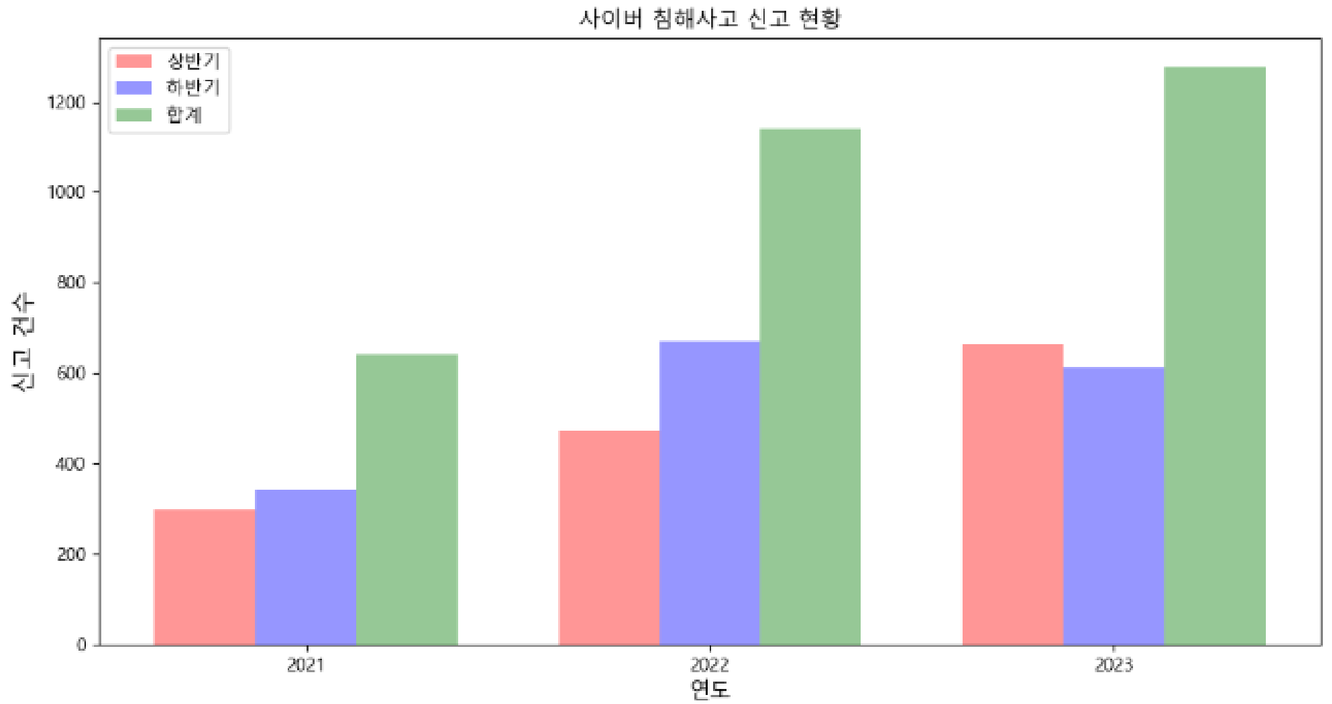
CSV 파일에서
필요한 데이터만 추출해
리스트에 저장

matplotlib 를 사용해
파이 그래프를
그리기 위한 코드

결과 보고

사이버 침해사고 신고 현황

2021년 ~ 2023년 사이버 침해사고 신고 현황

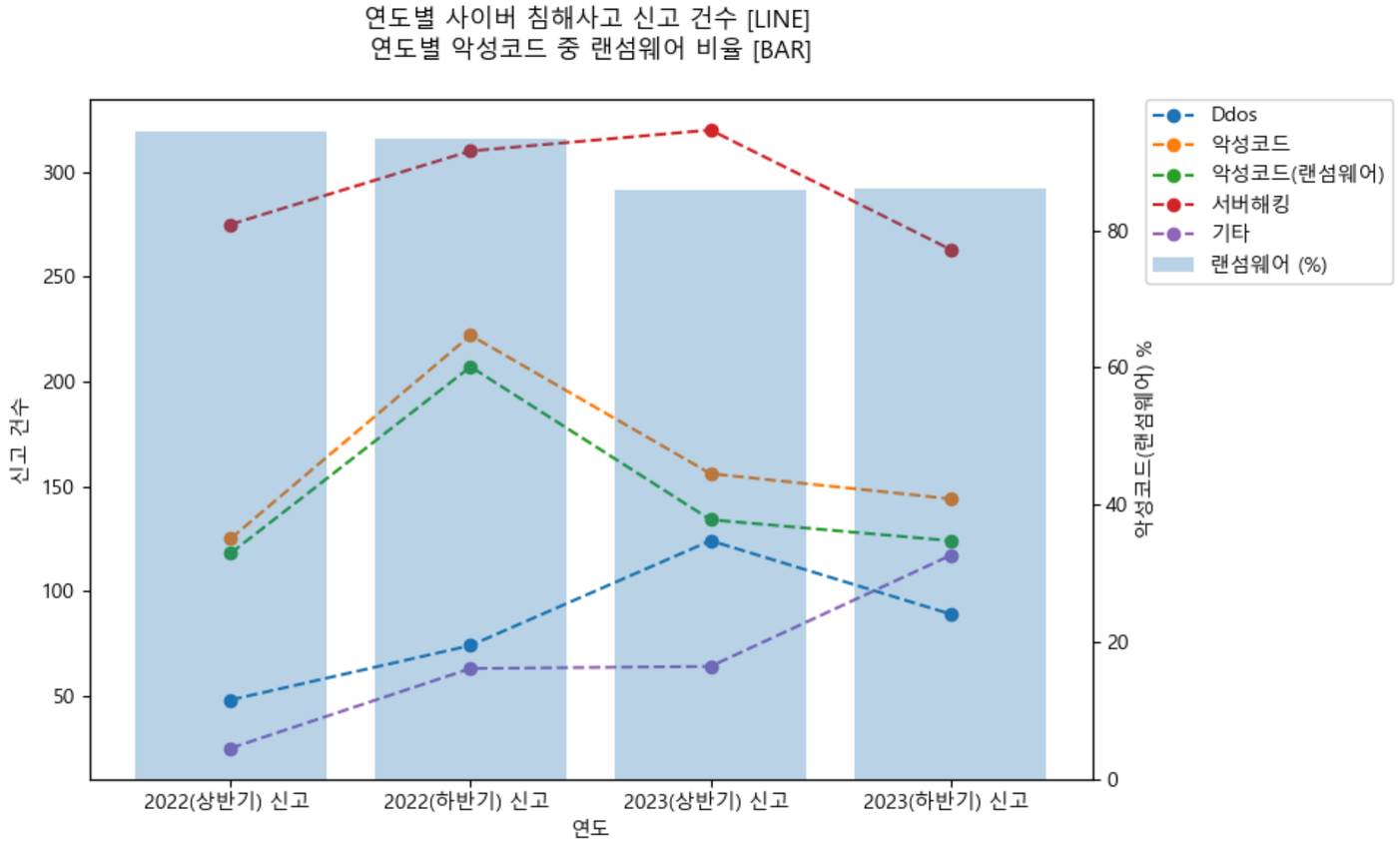


1. 2022년 1,142건에서 2023년 1,277건으로 **전년 대비 약 12% 증가**
2. 2023년 상반기 신고 건수는 664건으로 **전년 대비 40% 증가**

결과 보고

유형별 사이버 침해사고 신고 현황

유형별 사이버 침해사고 신고 현황



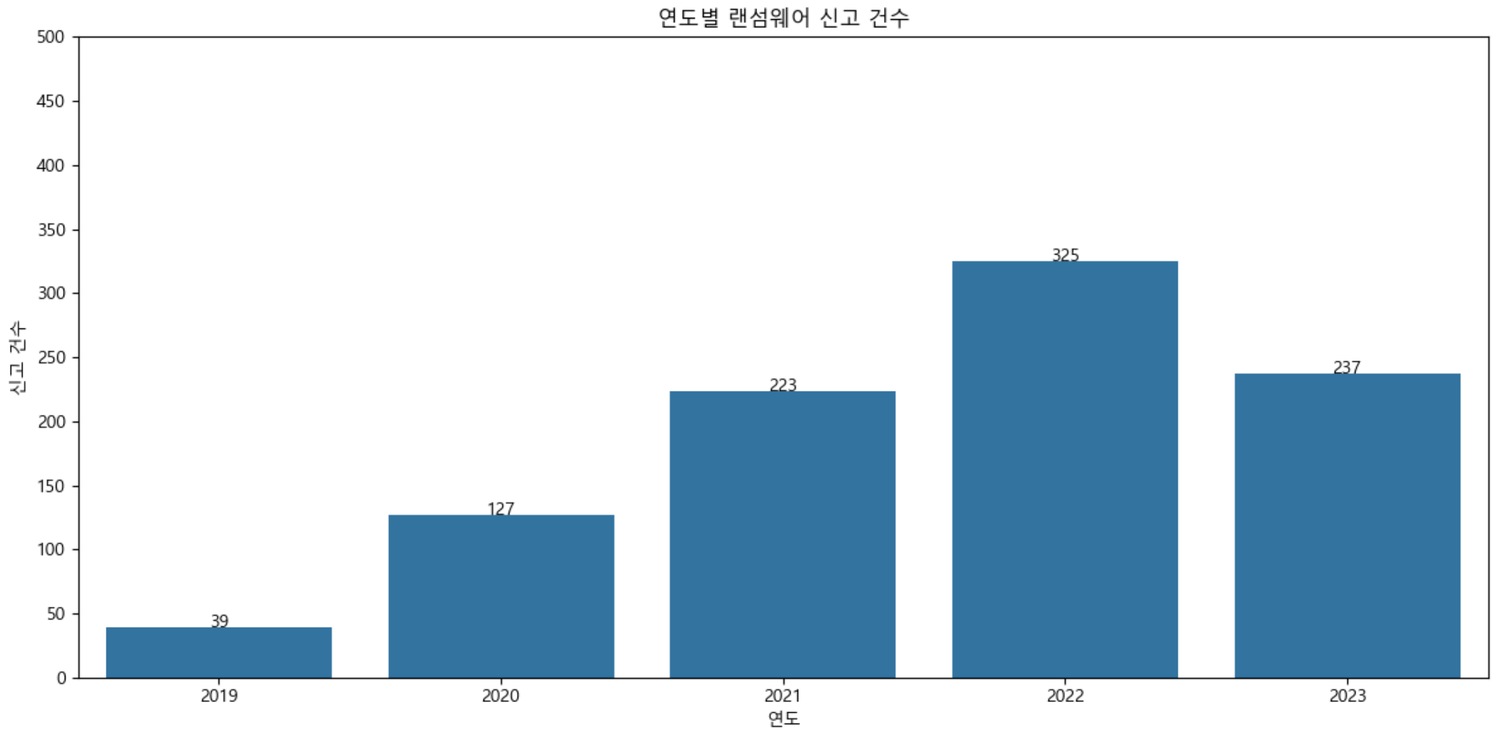
* 기타 분류에는 정보 유출, 스팸 문자 및 메일 발송 등의 침해사고 신고 건이 포함되어 있음

1. 전체 유형별 비중으로는 서버 해킹이 45.7%로 가장 높음
   그다음으로는 악성코드 감염이 23.5%, DDoS 공격이 16.7%, 기타 14.2%인 것으로 나타남
2. **악성코드 감염은 90% 이상의 비중을 랜섬웨어 신고가 차지하고 있으며,**
   2022년 랜섬웨어 신고는 325건으로 지난 4년간 8.3배로 급속히 증가

결과 보고

연도별 랜섬웨어 신고 건수

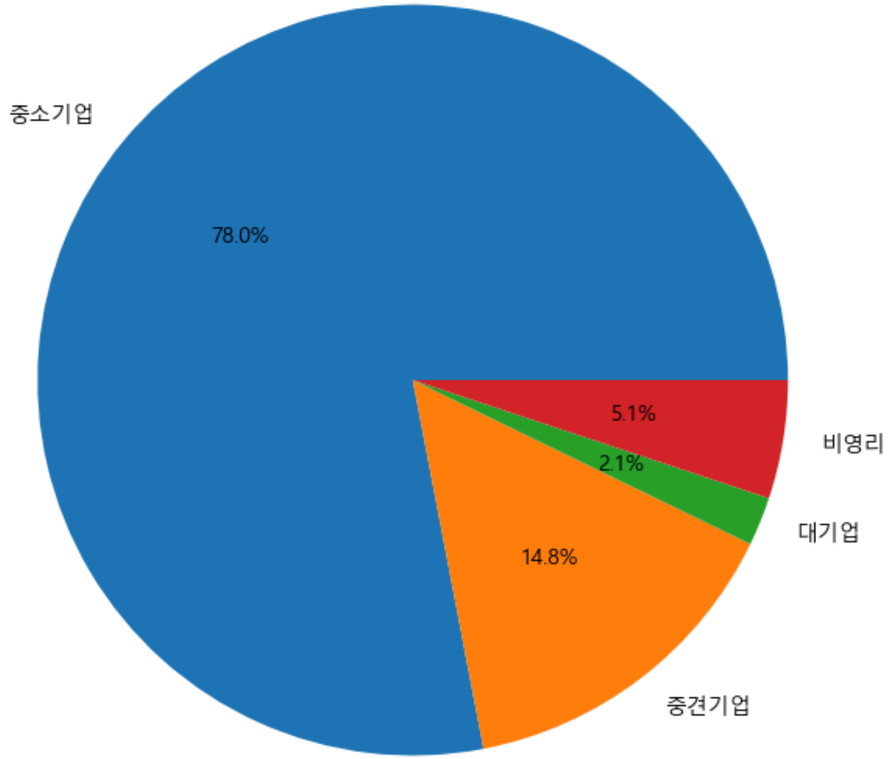
2019년 ~ 2023년 랜섬웨어 신고 건수



1. 사이버 보안 위협은 매년 지속해서 증가했지만,
   랜섬웨어 신고 건수는 작년 대비 27% 감소해 발생률이 줄어든 것처럼 보임

결과 보고

규모별 랜섬웨어 신고 건수

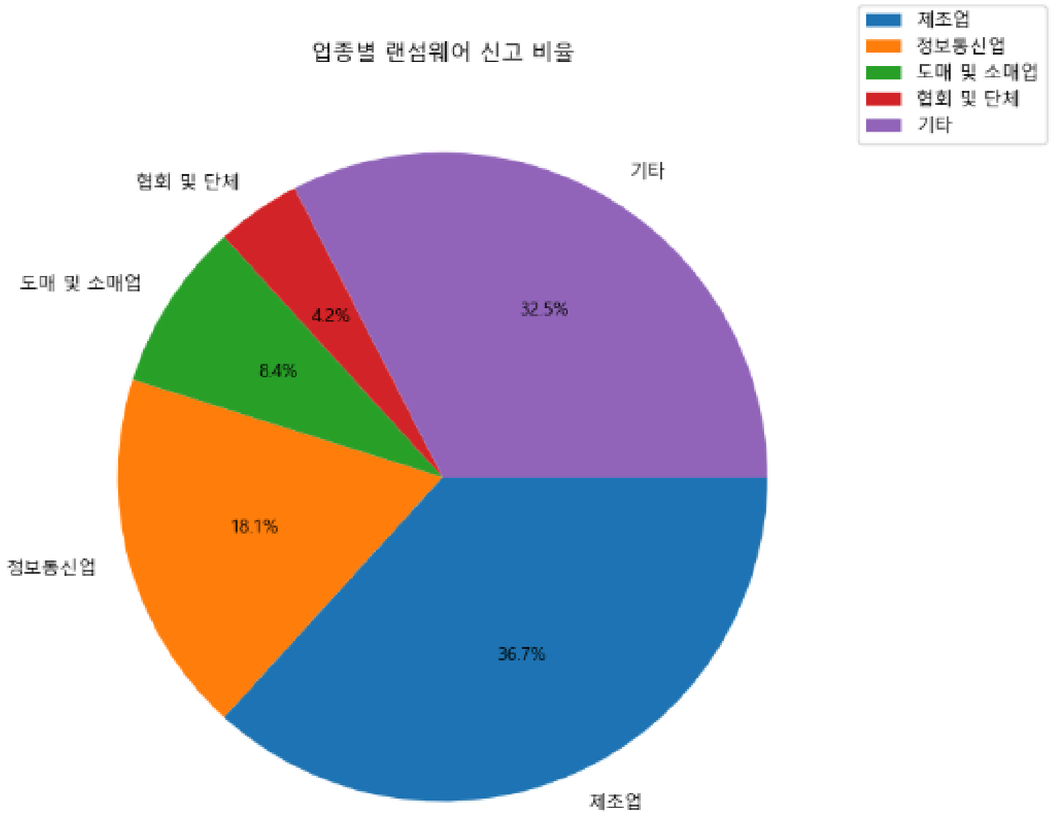


1. 최근 랜섬웨어 공격은 주로 **중소기업(78.1%)을 대상**으로 이루어짐
2. **기업의 정보를 빼내고, 운영 서버와 백업 서버 자료까지 찾아 암호화해 금전을 요구하는 복합적인 방식(Multi Exortion,다중협박)으로 이루어짐**

결과 보고

업종별 랜섬웨어 신고 건수

업종별 랜섬웨어 신고 비율

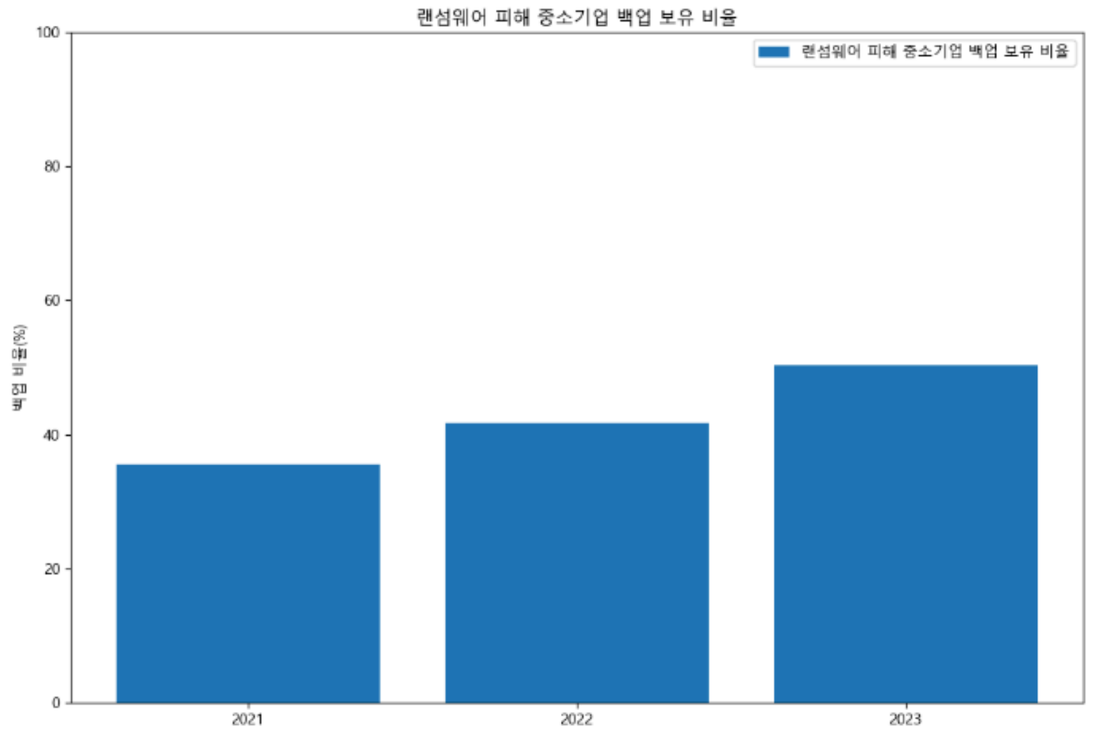


1. 업종별 공격 유형은 **제조업종(36.7%)이 가장 높음**
2. **기업의 정보를 빼내고, 운영 서버와 백업 서버 자료까지 찾아 암호화해 금전을 요구하는 복합적인 방식(Multi Exortion,다중협박)으로 이루어짐**

결과 보고

랜섬웨어 피해 중소기업 백업 보유 비율

랜섬웨어 피해 중소기업 백업 보유 비율

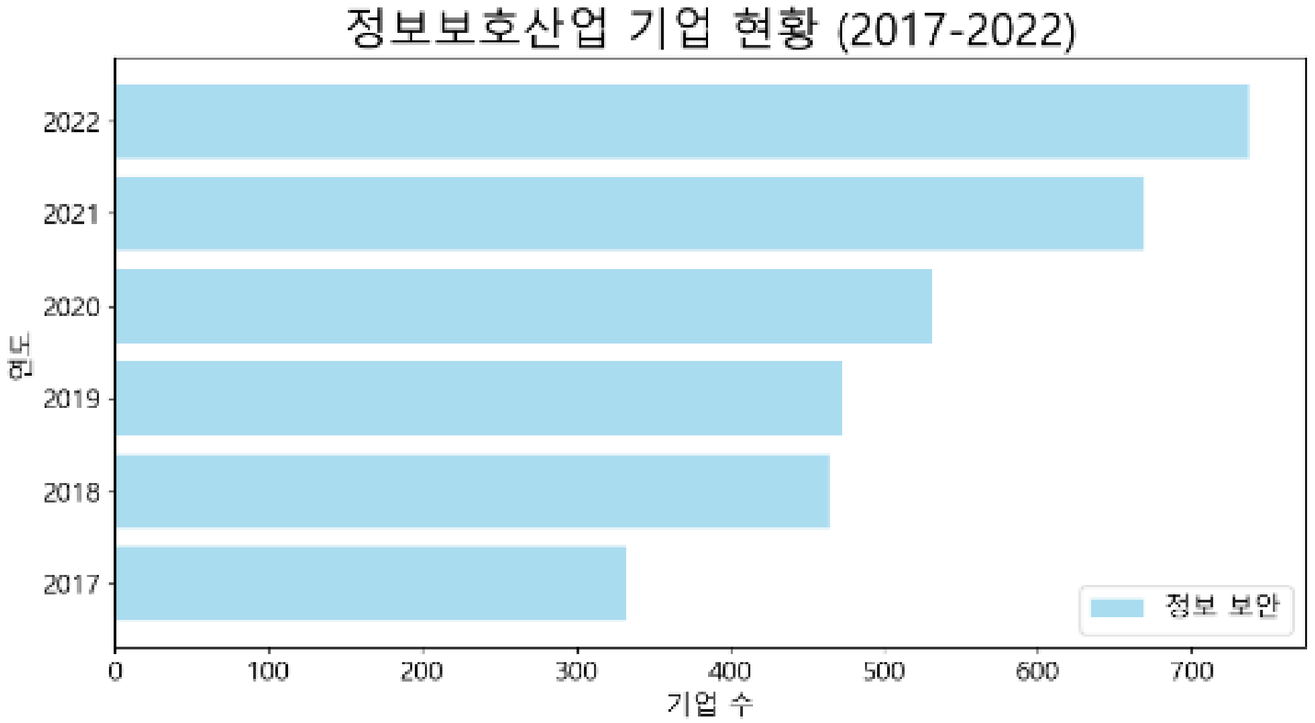


1. 과학기술정보통신부와 KISA의 중소기업 백업체계 구축 지원사업과 안내 등을 통해
랜섬웨어 피해를 신고한 중소기업의 약 50.3% (21년,35.6%)는
데이터 백업체계를 구축해 피해가 그나마 최소화 되었지만,
나머지 기업들을 여전히 데이터 복구에 어려움을 겪고 있음

결과 보고

정보보호산업 기업 현황

정보보호산업 기업 현황

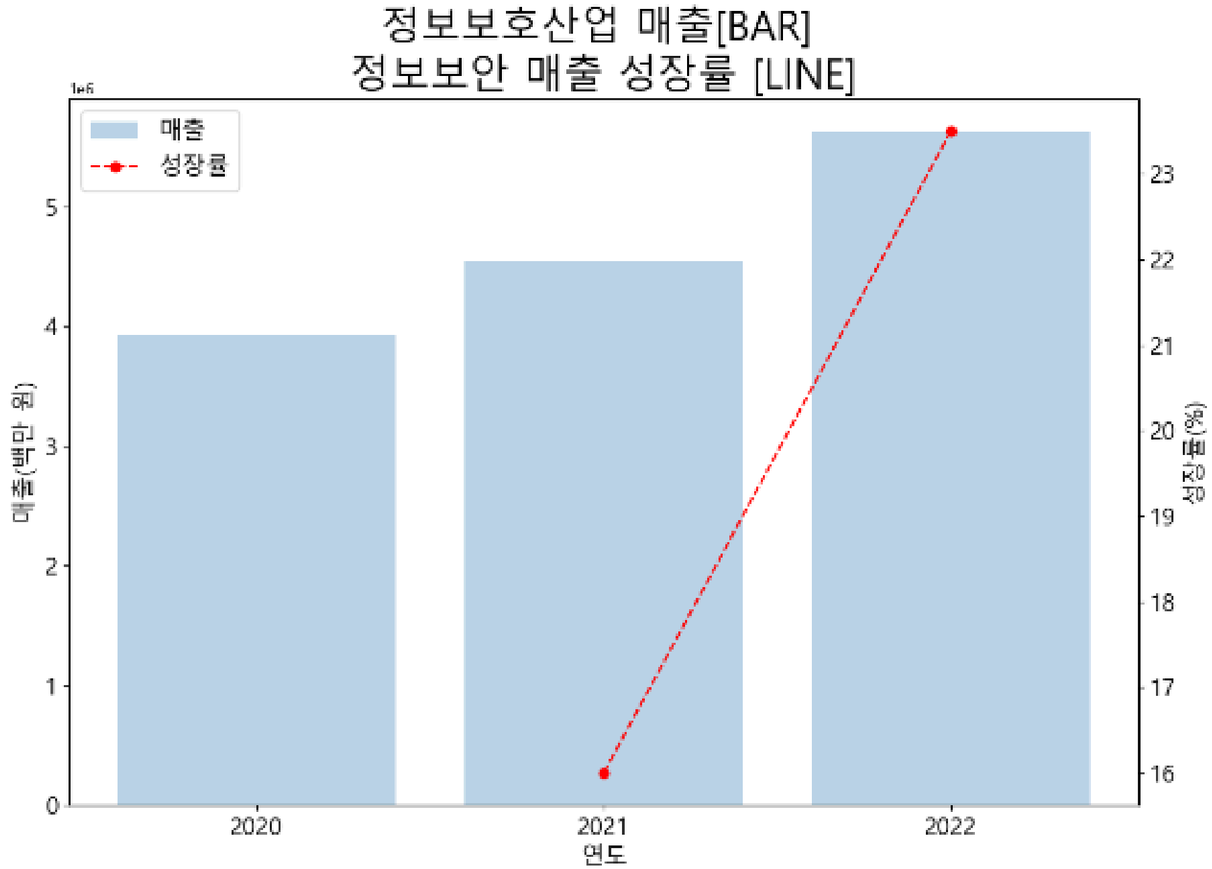


1. 국내 소재 정보보호 기업은 정보 보안 업체만 737개 기업이 있음
2. **정보보호 기업은 꾸준히 증가**

결과 보고

정보보호산업 매출 현황

정보보호산업 매출 현황



1. 정보보안 매출액은 전년 대비 23.5% 증가
2. 정보보호 기업은 매출 또한 꾸준히 증가

결과 보고

프로젝트 결과 보고

분 석

사이버 보안 위협은 매년 지속적으로 증가
그러나 2023년 기준으로 랜섬웨어 공격 건수는 작년 대비 27.1% 로 공격 위협 감소

하지만, **최근의 랜섬웨어 공격은 주로 중소기업(78.1%)과 제조업종(36.8%)을 대상으로 함**
이러한 공격은 기업의 기밀 정보를 탈취하고 금전을 요구하는 복합적인 방식으로 진화

사이버 방어체계의 고도화 노력에도 불구하고,
범죄 조직은 기업화되고 있으며, 공격자들 역시 새로운 취약점을 찾아 지속적으로 진화

2021년 기준으로 랜섬웨어 피해를 신고한 중소기업 중 50.3%는 데이터 백업 체계를 구축해
해를 최소화 했지만, 나머지 기업들은 백업 체계를 구축하지 않아 피해가 더 큼

결 과

따라서, 기관 및 기업 등 조직은 보안 시스템을 도입해 운영한다고 해서 안심할 수 없음
사이버 침해가 발생하더라도 업무 중단이 되지 않도록 백업 체계를 마련하고,
신속한 복구 프로세스를 반복적으로 점검하고 강화해야 함

소 감

단순히 많은 데이터를 보유하는 것만으로는 활용 가능한 정보가 되지 않음을 깨달음
데이터를 수집 후 유의미하게 분석해 활용 가능한 정보를 도출하는 것이 중요하다는 점을 배움

소감

팀원1

데이터 분석 과정에서 데이터가 예상과 다른 형식으로 제공되거나,
시각화가 원하는 대로 나오지 않아 이를 해결하기 위한 해결법을 찾아보며
문제 해결 능력이 향상 되었고 파이썬에 대해 깊게 공부할 수 있었습니다.
또한, 방대한 데이터 중 유의미한 데이터를 분석해 인사이트를 도출하는 것의
중요성을 깨닫게 되었습니다.

팀원2

파이썬으로 데이터 시각화를 하면서 다양한 라이브러리를 통해 데이터를
쉽게 시각화할 수 있었고, 특히, matplotlib을 사용해 시각적으로 표현하는
능력을 기를 수 있습니다.

팀원3

파이썬의 다양한 라이브러리를 사용하며 데이터를 여러 그래프로 시각화 할
수 있었습니다. 이번 경험을 통해 파이썬을 활용하는 여러 방법을 배웠고,
더 나아가 파이썬으로 웹 프로젝트를 진행해 보고 싶습니다.

정석진

데이터 분석 과정에서 다양한 문제에 직면하게 됩니다. 예를 들어, 데이터가
예상과 다른 형식으로 제공되거나, 시각화가 원하는 대로 나오지 않는 경우
등입니다. 이러한 문제를 해결하면서 문제 해결 능력이 향상되고, 스스로
로운 방법을 찾아내는 능력을 기를 수 있게 되었습니다.

팀 원 소 개

팀원1

팀원2

팀원3

정석진

Tel. 010-6622-1689
Email. ddoljin@gmail.com
Blog. https://varietyofit.tistory.com

감사합니다

Thank you