



2021 데이터 분석 경진대회

우리나라 농작업 손상의 특성에 대한 다양한 고찰

이소희, 조수연, 김도훈, 장수진

목차

대회 공통 분석 주제

우리나라 농작업 손상의
특성에 대한 다양한 고찰



개인별 분석 주제

1

성별, 연령별로 본 농업인의 손상 요인 분석
- 이소희

2

안전사고 예방 물품 구비 여부와 구비한 물품
개수에 따른 농작업 관련 사고에서의 일하지 못한
일수에 대한 분포 분석 - 조수연

3

농가주민의 농업인식에 대한 분석
-김도훈

4

부상을 중심으로 상관관계를 분석
-장수진

성별, 연령별로 본 농업인의 손상요인 분석



데이터 분석 과정



문제 정의

분석 이유, 목적
설정 및 분석 계획
구체화

데이터 처리

원하는 컬럼 추출
등의 데이터 전처리

데이터 분석

지표정의, 탐색적
데이터 분석, 통계분석

리포팅

분석 결과 및
인사이트를 설득력
있게 정리

분석 주제



분석 주제

연령 및 성별로 본 농업인의 업무중 손상요인 분석



주제 선정 이유

대부분의 농업인 손상조사 보고서에서는 연령별 손상 요인에 대한 분석은 필수도 등장한다. 그러나 성별까지 함께 분석한 보고서는 찾을 수 없었다. 연령 별로 손상 요인이 다르다면 성별로도 손상 요인이 다를 것으로 예상되어 연령, 성별 요인분석을 진행하였다.

농업인의 연령 성별 별 업무상 손상 수

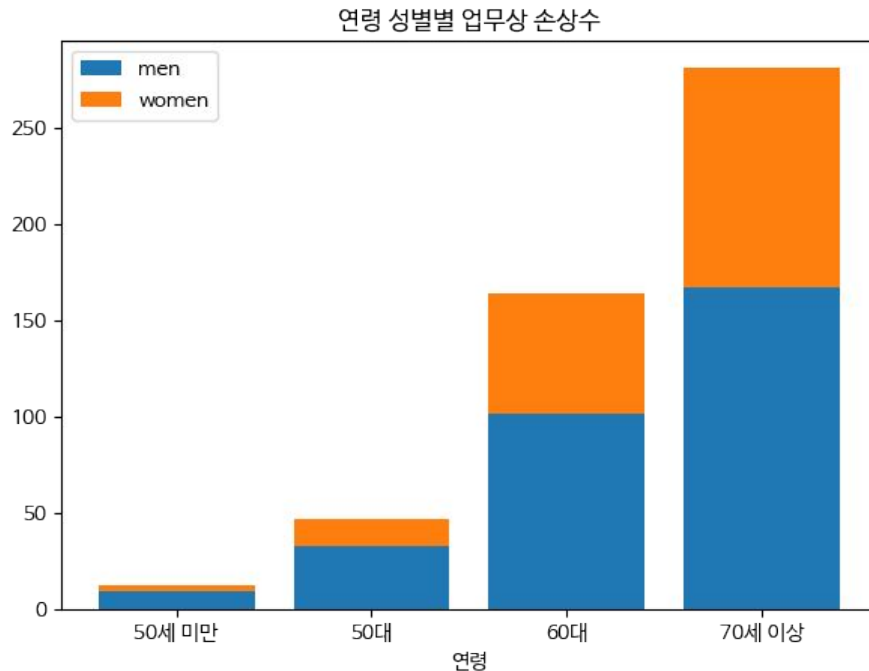


- 연령별 손상 수

연령이 증가할수록 업무상 손상 인구가 증가하는 것을 볼 수 있습니다.

- 성별 별 손상 수

인구가 늘어남에 따라 남성과 여성의 수가 일정하게 늘어나는 것을 볼 수 있다. 60대~70대 이상 구간에서 여성 인구가 급격히 증가한 것을 볼 수 있다.



농업인의 연령별 업무상 질병 유병률



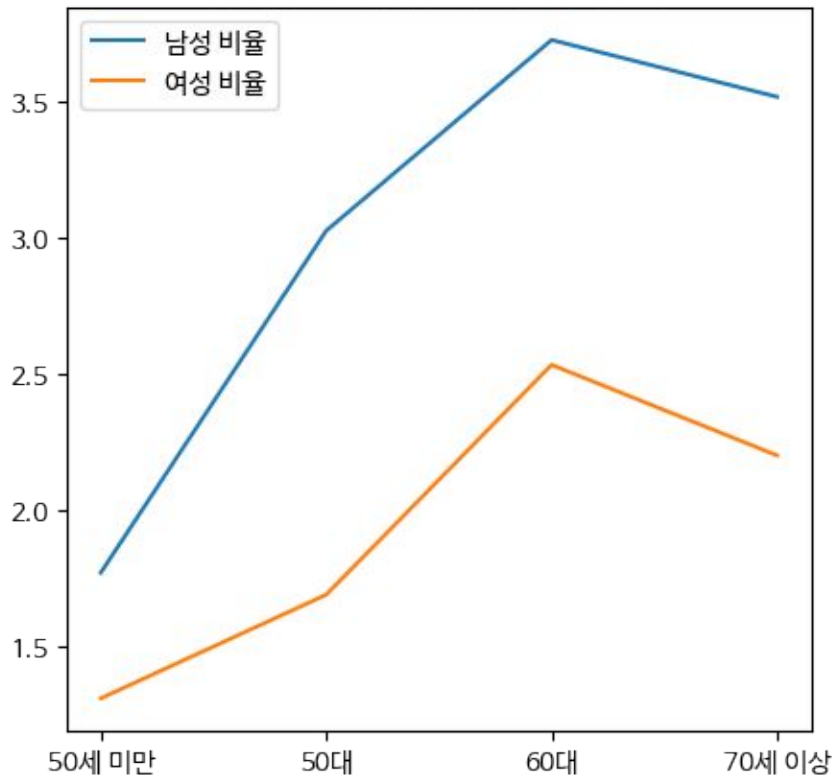
- **남성의 업무상 질병 유병률**

- 50세 미만 인구에서 50대로 넘어갈 때 급격하게 질병 유병 인구가 늘었다.
- 70세 이상 인구에서 질병 유병률이 낮아진다.

- **여성의 업무상 질병 유병률**

- 50세 미만 ~ 50대 구간의 증가 추세는 완만하다.
- 50대 ~ 60대 구간에서 급격하게 질병 유병률이 늘어나는 경향이 있다.
- 60대 ~ 70세 이상 구간에서 가파르게 질병 유병률이 감소한다.

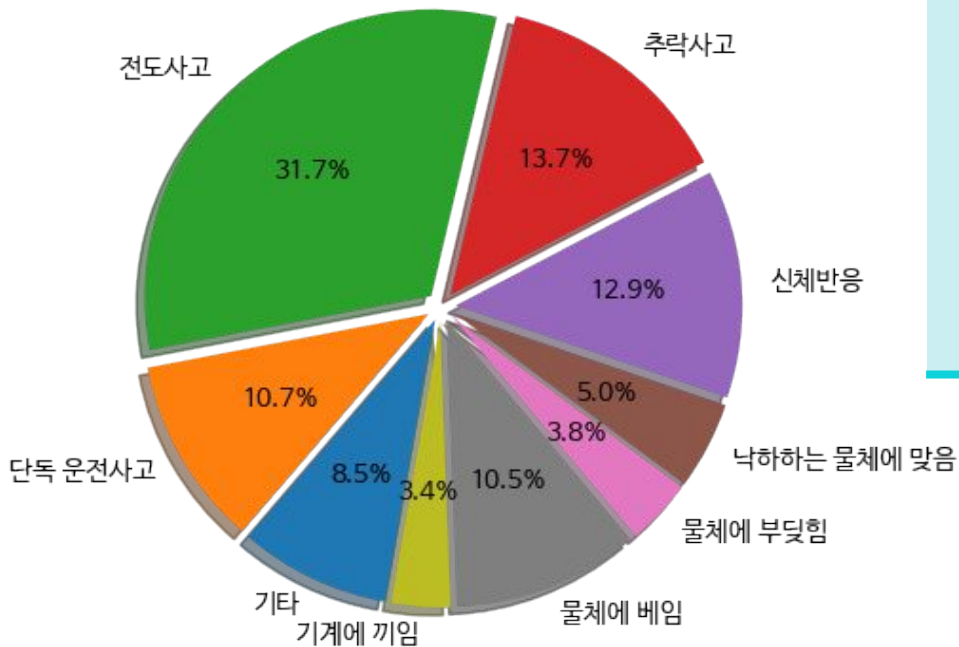
농업인의 연령별 업무상 질병 유병률



농작업별 사고 원인 비율



농작업별 사고 원인 비율



순위	농작업 종류	%	순위	농작업 종류	%
1	전도사고	31.7%	6	기타	8.5%
2	추락사고	13.7%	7	낙하하는 물체	5%
3	신체반응	12.9%	8	물체에 부딪힘	3.8%
4	단독 운전사고	10.7%	9	기계에 끼임	3.5%
5	물체에 베임	10.5%			

농작업별 사고 원인

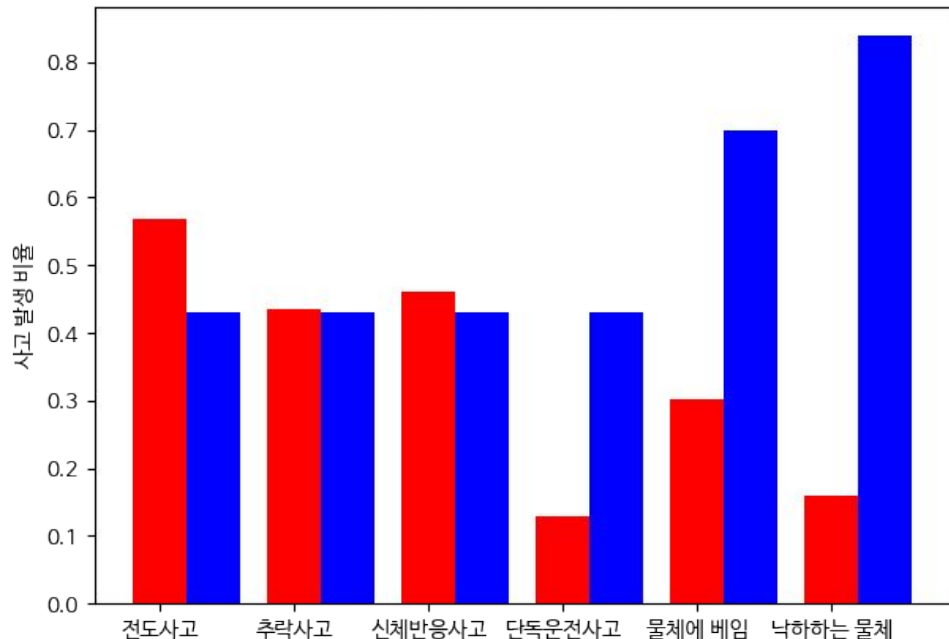
농작업 사고 원인 14개 중 3% 미만은 모두 '기타'로 분류하였다.

위의표는 1위부터 9위까지 농작업 종류와 비율을 나타낸다.

농업인의 성별 사고 발생율



농업인의 성별 사고 발생율



- **남성 사고 발생율**

- 낙하하는 물체에 부딪히는 경우가 1위, 물체에 베이는 경우가 2위, 나머지는 같은 비율이다.
- 남성의 경우 낙하하는 물체에 맞아 다치는 경우가 가장 많았고 대부분 비슷한 비율로 손상을 입었다.

- **여성 사고 발생율**

- 전도사고가 1위, 신체반응사고가 2위 추락사고가 3위, 물체에 베임, 낙하하는 물체에 맞음, 단독 운전사고가 각각 4,5,6 위다.
- 여성의 경우 전도사고로 가장 많이 다쳤고 단독 운전사고에서 가장 낮게 손상비율이 나타났다.

안전사고 예방 물품과 사고 발생의 연관 분석

- 안전 사고 예방 물품 구비/물품 구비 개수에 따른
농 작업 관련 사고에서의 일하지 못한 일수에 관한 분석



데이터 분석 과정



목표 설정하기
데이터를 보고
관심있는 부분을
찾아내어 분석
목표를 설정하고
방법과 절차
계획하기

데이터 나누기
분석 목표에
적합하도록 각 물품
구비 여부에 따라
데이터 나누어
분류하기

시각화하기
분류한 데이터를
각 응답에 대한
데이터 수가 잘
드러나게 그래프로
시각화 하기

추론, 결과 내기
그려진 그래프를
나열하고 분석
주제에 맞게 각
그래프의 차이점에
주목하여 추론하기

분석 주제

제가 설정한 분석
주제는 안전 사고 예방
물품의 구비 여부와
구비된 물품의 개수를
보고, 그것이 농작업에
관련된 사고와
어떻게 연관되었는지
보기 위한 분석입니다.



주제 설정 배경

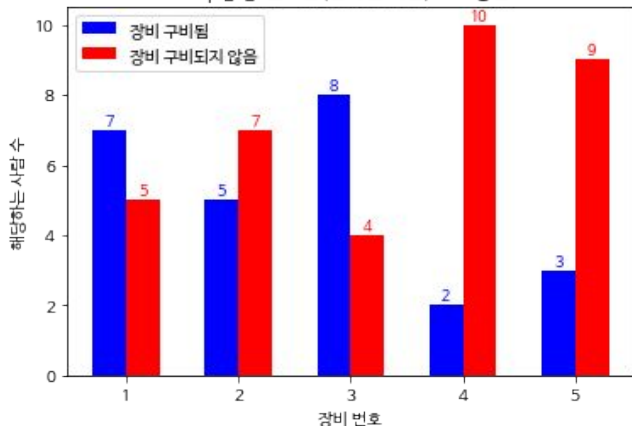
안전사고는 안전사고 예방
물품 구비와 안전의식이
있다면 충분히 예방할 수
있다고 할 만큼 중요한
요인입니다.

따라서 안전사고 예방
물품을 구비했는지에 따른
결과 분석을 보여줌으로써
안전사고 예방을 위한
노력은 중요함을 알리고
싶었습니다.

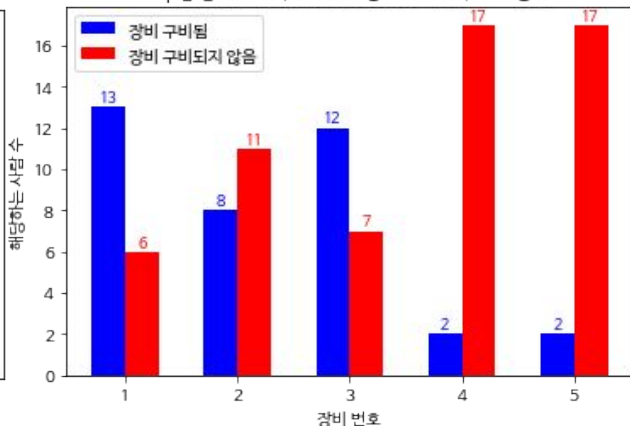
물품 구비 여부에 따른 그래프 비교



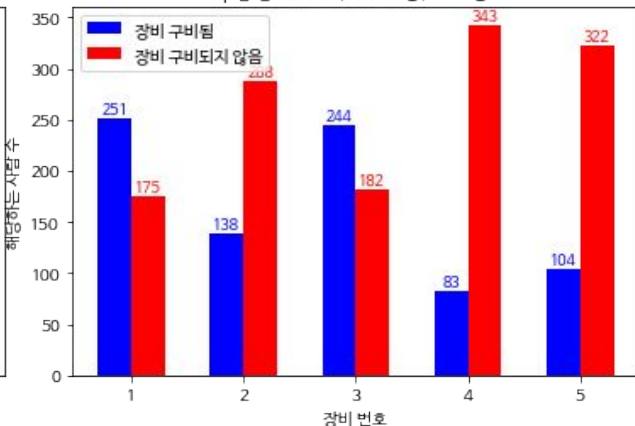
aq9_1_1을 1번(반나절 미만)으로 응답



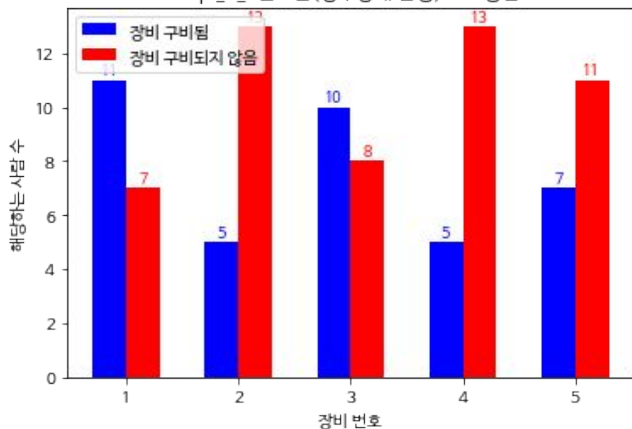
aq9_1_1을 2번(반나절 이상 하루 미만)으로 응답



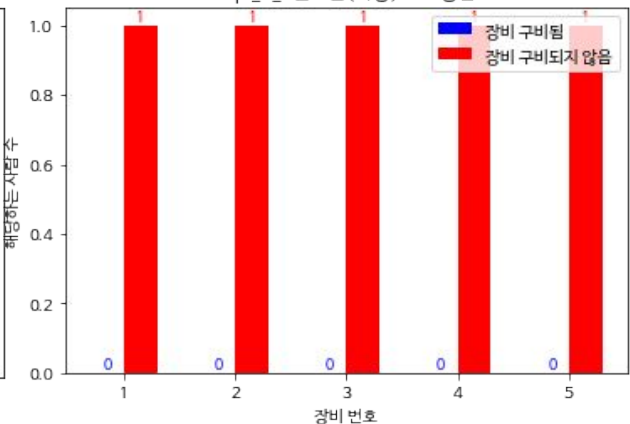
aq9_1_1을 3번(하루 이상)으로 응답



aq9_1_1을 4번(영구장애 발생)으로 응답



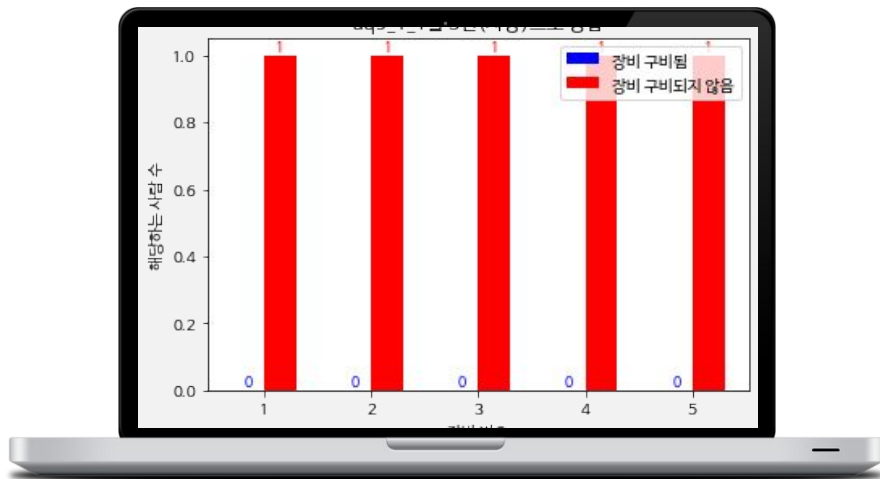
aq9_1_1을 5번(사망)으로 응답



-1번, 3번 장비는 구비되어도 구하지 않은 사람보다 손상된 경우가 많았습니다.

-4번, 5번 장비의 경우 구비하지 않았을 경우 손상을 입는 경우가 구비했을 때의 경우보다 월등히 많았습니다.

물품 구비여부 분석에 따른 문제점



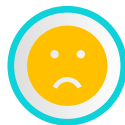
*극단적인 예시를 위해 앞선 페이지에서 하나의 그래프를 가져왔습니다.



각 장비 번호에 따른 구비 여부를 따졌기 때문에, 중복된 데이터가 발생할 수 있습니다.



왼쪽의 예시의 경우, 코드를 통해 각 장비 번호에 해당하는 사람은 공통된 한 사람임이 밝혀졌습니다.

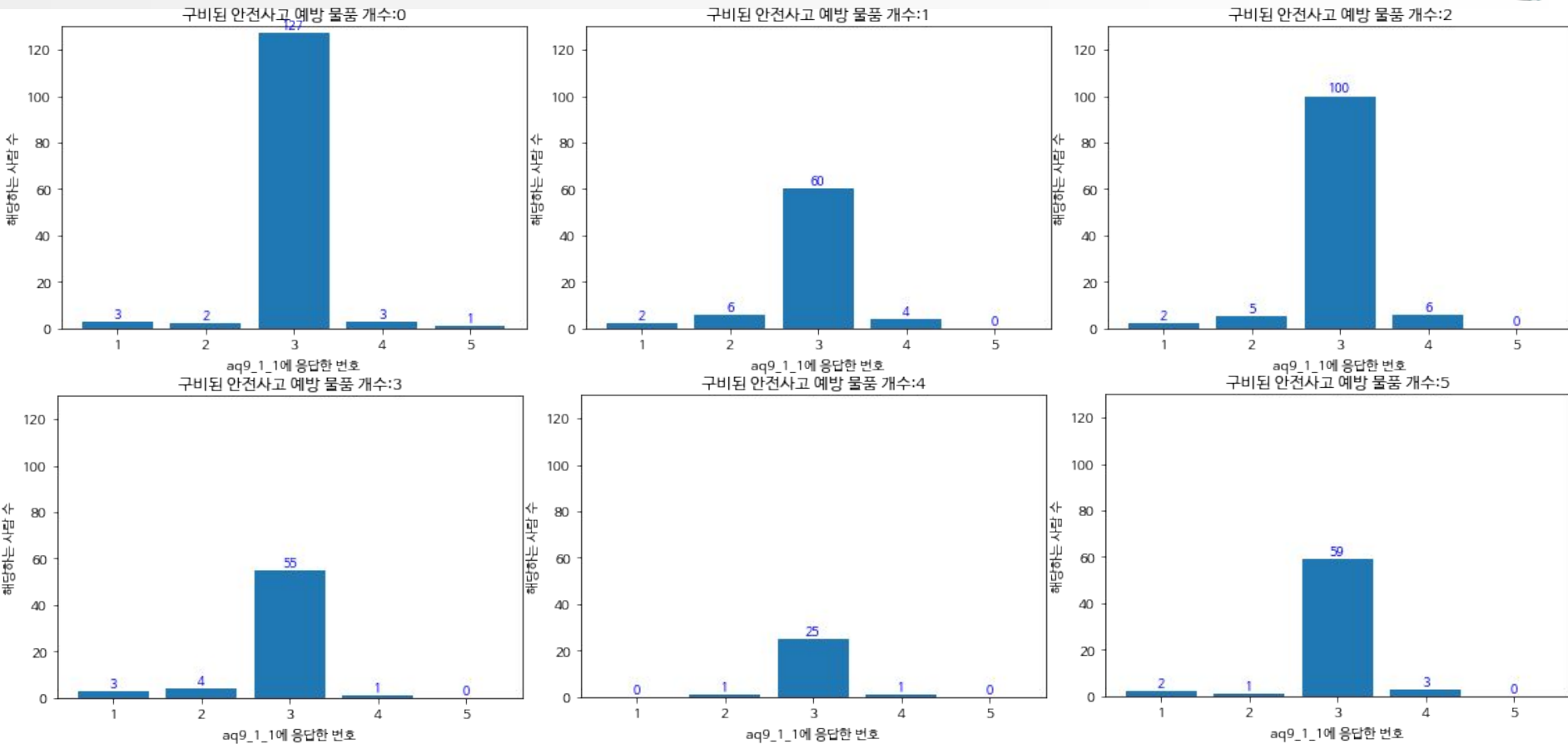


중복된 데이터로 인해, 그래프를 보고 정확한 판단을 하기가 힘들어질 수 있습니다.

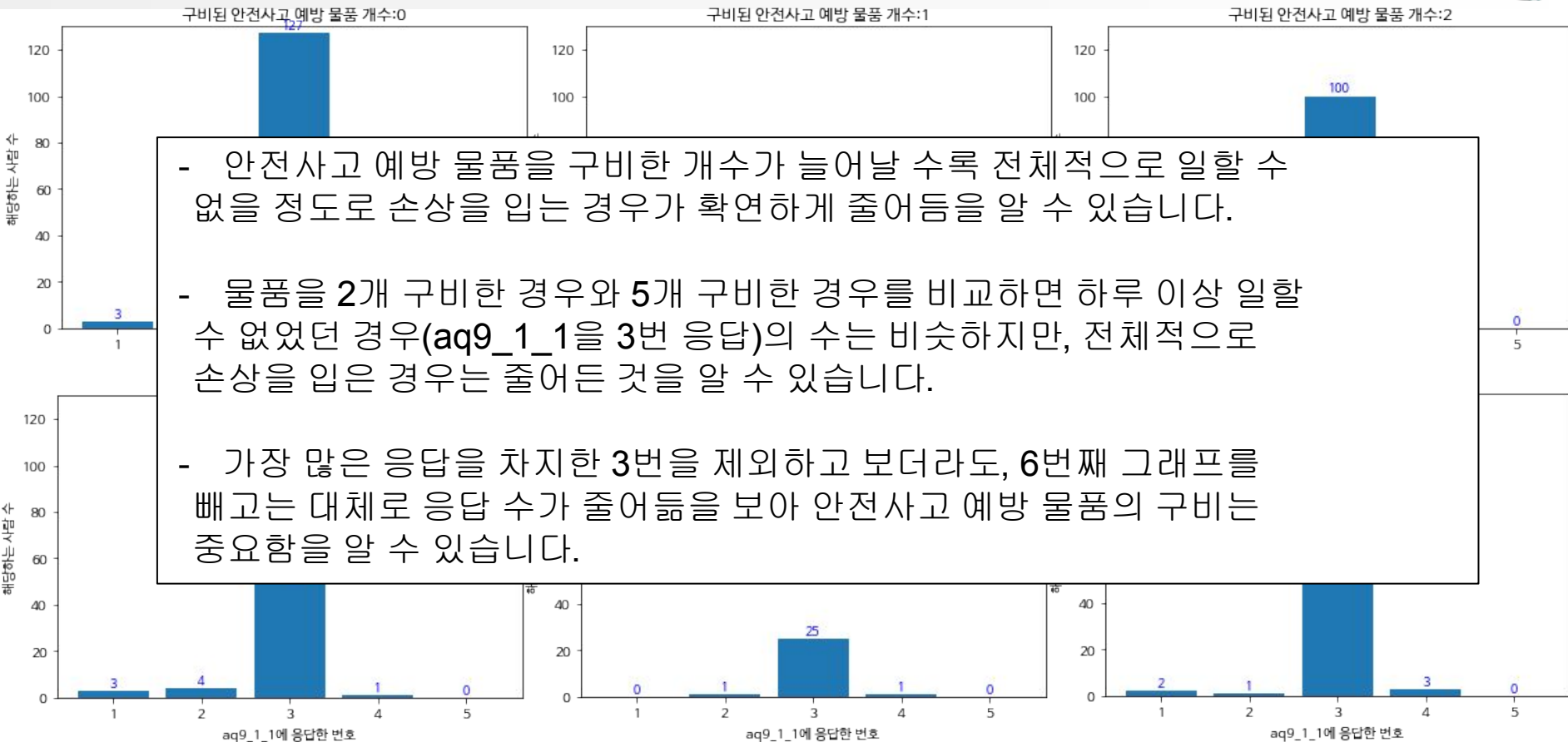


그렇다면 각 장비 번호에 따른 구비 여부가 아닌, 장비를 구비한 개수에 따른 그래프를 본다면 어떨까요?

물품 구비 개수에 따른 그래프 비교

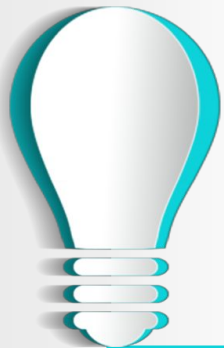


물품 구비 개수에 따른 그래프 비교



농가주민의 농업인식에 대한 분석

- 농가주민의 농가피해인식 간의 관련성 분석



분석 주제



분석 주제

농가주민의 농가위험인지수준에 대한 분석



주제 선정 이유

농가주민의 농가위험성인식에 대한 차별성과

농가주민의 농가일에 대해 부정적견해 유무에 관한 호기심

농가주민들의 작년의 농업활동기간이 많은가



```
df1 = pd.read_excel('C:/Users/LG/Desktop/경진대회/경진대회용데이터.xlsx')
pd.read_excel('C:/Users/LG/Desktop/경진대회/경진대회용데이터.xlsx')

##리스트화
farmPeriod = list(df1.iloc[:, 35])##q12

periodFirst = 0
periodSecond = 0
periodThird = 0
periodFourth = 0

for i in range(0, len(farmPeriod)):
    if(farmPeriod[i]== 1):
        periodFirst = periodFirst + 1
    if(farmPeriod[i]== 2):
        periodSecond = periodSecond + 1
    if(farmPeriod[i]== 3):
        periodThird = periodThird + 1
    if(farmPeriod[i]== 4):
        periodFourth = periodFourth + 1

x = [1,2,3,4]

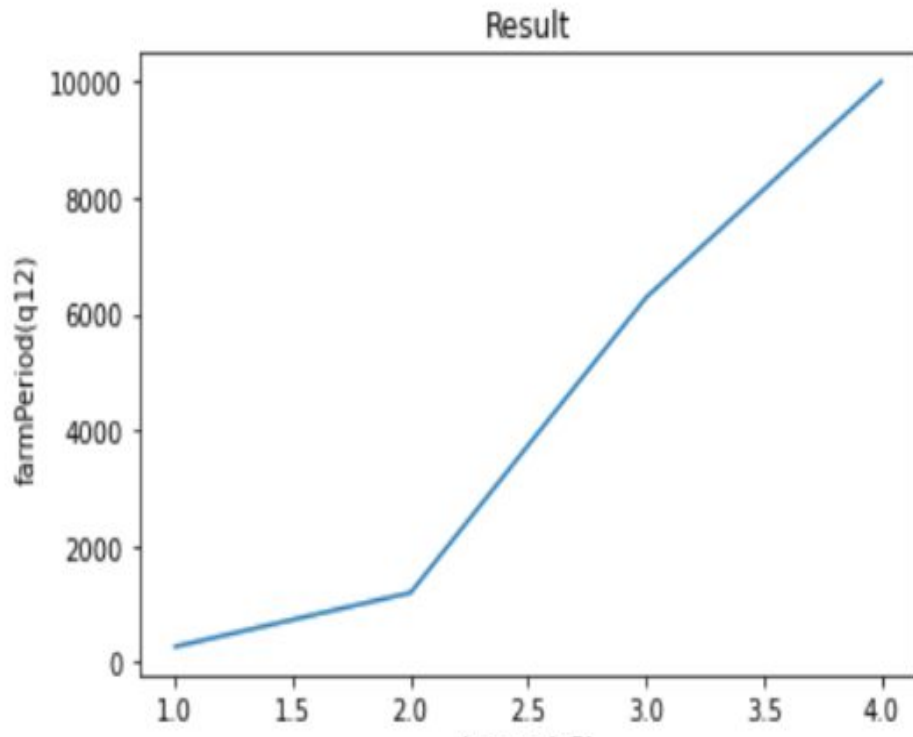
periodGraphY = [periodFirst, periodSecond, periodThird, periodFourth]
print(periodGraphY)

plt.plot(x,periodGraphY)
plt.ylabel("farmPeriod(q12)")
plt.title("Result")
##농가에 있는사람들은 대부분 오랫동안 농업활동을 하신 분들이 많다.
```

원하는 칼럼을 리스트화 한 후
1번 2번 3번 4번 선택의 빈도수를 코드로 작성후

농가주민들의 농업기간(q12)에 대해
그래프로 시각화를 한다.

농가주민들의 작년의 농업활동기간이 많은가



농가주민들의 농업활동기간이 대부분 많음을 알수가있다.

농가주민들은 작년 농업활동에 제한을 받지 않았는가



```
df1 = pd.read_excel('C:/Users/LG/Desktop/경진대회/경진대회용데이터.xlsx')
pd.read_excel('C:/Users/LG/Desktop/경진대회/경진대회용데이터.xlsx')
```

```
##리스트화
```

```
farmRisk = list(df1.iloc[:, 36]) ##q13
```

```
RiskFirst = 0
```

```
RiskSecond = 0
```

```
RiskThird = 0
```

```
RiskFourth = 0
```

```
for i in range(0, len(farmRisk)):
```

```
    if(farmRisk[i]== 1):
```

```
        RiskFirst = RiskFirst + 1
```

```
    if(farmRisk[i]== 2):
```

```
        RiskSecond = RiskSecond + 1
```

```
    if(farmRisk[i]== 3):
```

```
        RiskThird = RiskThird + 1
```

```
    if(farmRisk[i]== 4):
```

```
        RiskFourth = RiskFourth + 1
```

```
x = [1,2,3,4]
```

```
periodGraphY = [RiskFirst, RiskSecond, RiskThird, RiskFourth]
```

```
print(periodGraphY)
```

```
plt.ylabel("farmRisk(q13)")
```

```
plt.title("Result")
```

```
plt.plot(x, periodGraphY)
```

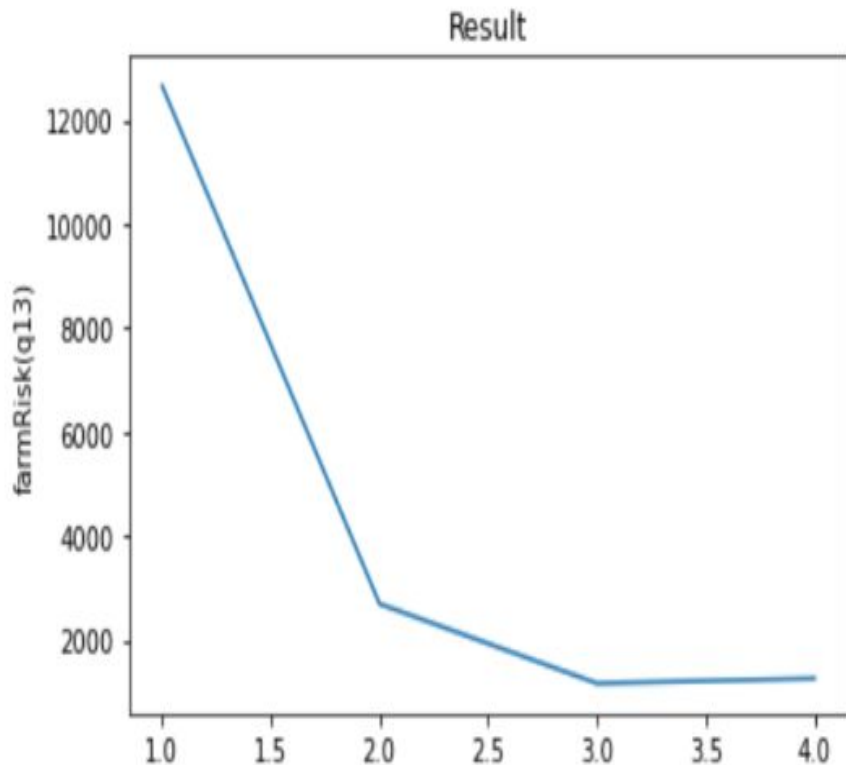
```
##농가주민들은 작년에 신체정신적어려움으로 농업활동에 제한을 받은사람이 적었다.
```

내가원하는 칼럼을 리스트화 한 후

1번 2번 3번 4번 선택의 빈도수를 코드로 작성후

농가주민들의 농업활동제한(q13)의 관계성에대해
그래프로 시각화 한다.

농가주민들은 작년에 농업활동에 제한을 받지않았는가



농가사람들은 농업활동에 제한을 받지않고
일하고 있다고 생각하고 있다는것을 알 수 있다.

농가사람들은 작년에 농약을 많이 썼는가



```
df1 = pd.read_excel('C:/Users/LG/Desktop/경진대회/경진대회용데이터.xlsx')
pd.read_excel('C:/Users/LG/Desktop/경진대회/경진대회용데이터.xlsx')
```

```
##리스트화
useInsecticide = list(df1.iloc[:, 38])##q17
```

```
useFirst = 0
useSecond = 0
useThird = 0
useFourth = 0
useFifth = 0
```

```
for i in range(0, len(useInsecticide)):
    if(useInsecticide[i]== 1):
        useFirst = useFirst + 1
    if(useInsecticide[i]== 2):
        useSecond = useSecond + 1
    if(useInsecticide[i]== 3):
        useThird = useThird + 1
    if(useInsecticide[i]== 4):
        useFourth = useFourth + 1
    if(useInsecticide[i]== 5):
        useFifth = useFifth + 1
```

```
x = [1,2,3,4,5]
```

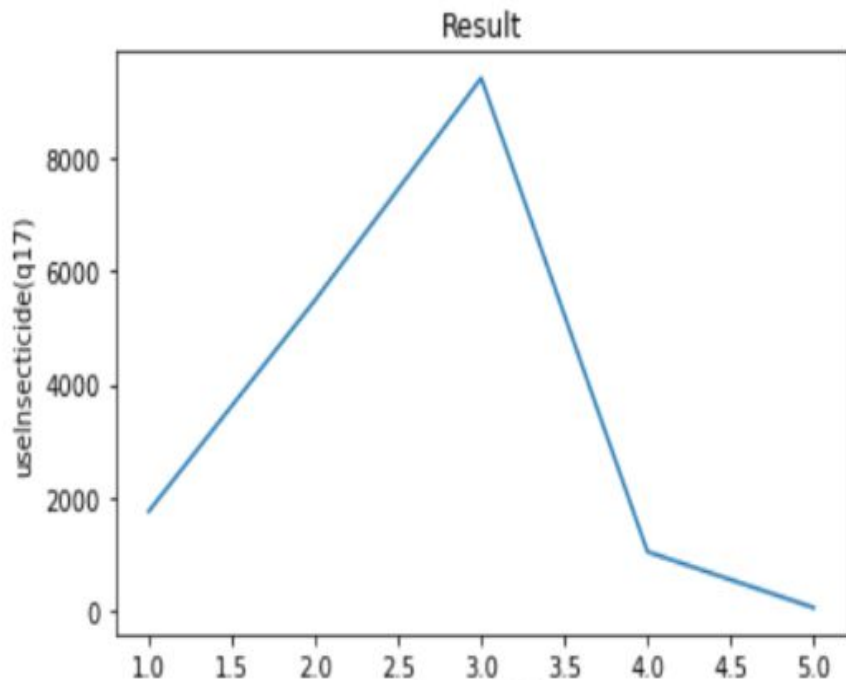
```
periodGraphY = [useFirst, useSecond, useThird, useFourth, useFifth]
print(periodGraphY)
```

```
plt.ylabel("useInsecticide(q17)")
plt.title("Result")
plt.plot(x, periodGraphY)
##농가주민들은 작년에 농약을 많이 썼다
```

내가 원하는 칼럼을 리스트화 한 후
1번 2번 3번 4번 선택의 빈도수를 코드로 작성 후

농가사람들의 농약사용(q17)의 관계성에 대해
그래프로 시각화 한다.

농가사람들은 작년에 농약을 많이 썼는가



농가사람들의 농약사용이 많다는것을
알 수 있다.

농가사람들은 안전을 중요하게 생각하는가



```
df1 = pd.read_excel('C:/Users/LG/Desktop/경진대회/경진대회용데이터.xlsx')
pd.read_excel('C:/Users/LG/Desktop/경진대회/경진대회용데이터.xlsx')
```

```
##리스트화
```

```
comfortList = list(df1.iloc[:, 40]) ##q22
```

```
comfotFirst = 0
```

```
comfotSecond = 0
```

```
comfotThird = 0
```

```
comfotFourth = 0
```

```
for i in range(0, len(comfortList)):
```

```
    if(comfortList[i]== 1):
```

```
        comfotFirst = comfotFirst + 1
```

```
    if(comfortList[i]== 2):
```

```
        comfotSecond = comfotSecond + 1
```

```
    if(comfortList[i]== 3):
```

```
        comfotThird = comfotThird + 1
```

```
    if(comfortList[i]== 4):
```

```
        comfotFourth = comfotFourth + 1
```

```
x = [1,2,3,4]
```

```
periodGraphY = [comfotFirst, comfotSecond, comfotThird, comfotFourth]
```

```
print(periodGraphY)
```

```
plt.ylabel("comfort(q22)")
```

```
plt.title("Result")
```

```
plt.plot(x, periodGraphY)
```

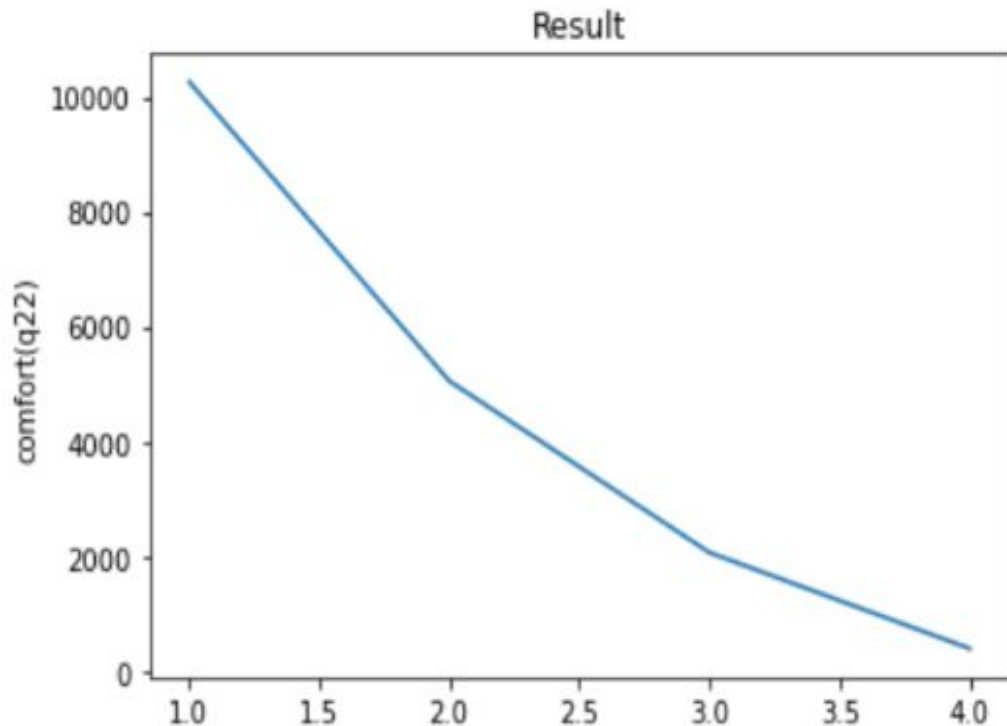
```
##농가주민들은 안전에 대해 매우 상당히 신경쓰는 편이다..
```

내가원하는 칼럼을 리스트화 한 후

1번 2번 3번 4번 선택의 빈도수를 코드로 작성후

농가사람들과 안전인지(q22) 간의 관계성에 대해
그래프로 시각화 한다.

수입이 많을수록 안전을 중요하게 생각하는가



농가사람들은 안전에 대해
신경을 많이 쓴다는 것을 알 수 있다.

농업활동중 부상과의 상관관계





분석 이유



주제 선정 이유

- 분석해야할 자료를 보고 무엇을 분석하면 좋을 지 고민하던 중, 부상입은 후 피해와 농업생활과의 연관성을 분석해보면 좋을 것 같았다.
- 연관성을 분석해 보면 부상으로 이끈 이유를 알 수 있을 것 같았다.
- 연관성을 분석해 보면 부상 이후, 농업생활에 영향을 알 수 있을 것 같았다.



분석주제

- 농업인들의 부상입은 것을 중심으로 부상입은 이유와의 상호작용여부 분석
- 부상과 부상 이후 농업활동에대한 영향 분석
- 부상과 부상 이후 일 못한 기간과의 상호작용여부 분석
- 부상과 부상 이후 치료여부와와의 상호작용 여부분석



분석 및 정리한 단계



자료분석 아이디어

자료를 어떤
중심으로 분석할
것인지 아이디어를
냅니다.



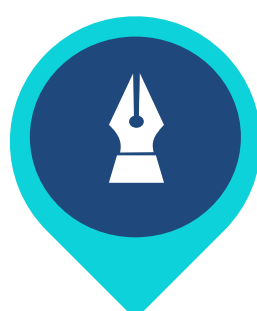
그래프 작성 및 자료분석

python3를 이용해
그래프를 그리고
R을 이용해 자료
분석을 합니다.



자료 정리하기

분석한 자료와
작성한 그래프를
고려하여 자료를
정리합니다.



ppt 작성

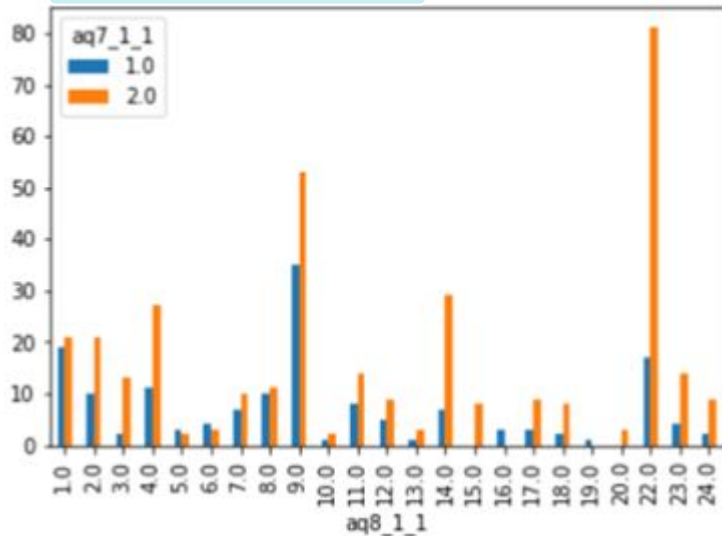
정리된 자료와
내용을 토대로
ppt를
작성합니다.



농기계 사용과 부상상관관계



부상부위와 농기구 사용



aq8_1_1=부상부위
aq7_1_1=농기구의 사용여부

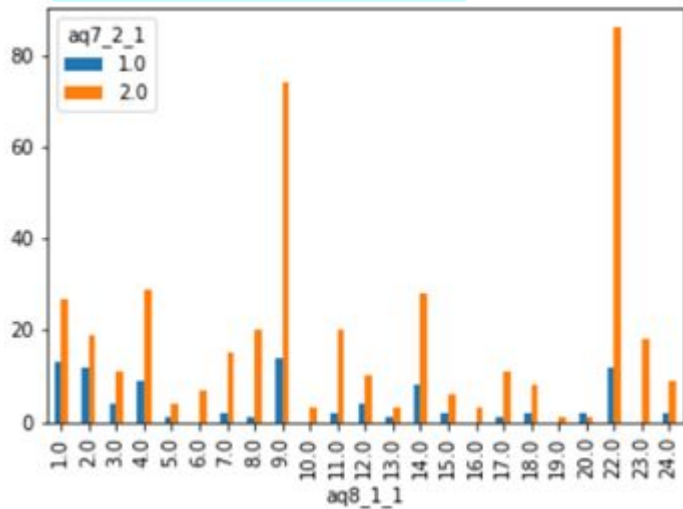
- 농기구를 사용하지 않았을 때 다리(9)와 척추(22)의 부상률이 높은 것으로 보인다.
- 사실상, 농기구를 사용하지 않았을 때 부상 비율이 높은 것을 알 수 있다.



농기계 사용과 부상상관관계



농기계 사용과 부상부위



aq8_1_1=부상부위

aq7_2_1=농기계 사용여부

- 농기계를 사용하지 않았을 때 척추(9)와 다리(22)에 부상을 입을 확률이 높은 것을 알 수 있다.
- 사실상, 농기계를 사용하지 않았을 때 부상확률이 더 높다는 것을 알 수 있다.



부상위치와 농기계사용& 농기구 사용의 상관관계 파악



R코드로 상관관계 분석

```

      Df Sum Sq Mean Sq F value    Pr(>F)
aq7_1_1    1    859    858.8    16.41 5.92e-05 ***
aq7_2_1    1   1221   1221.2    23.33 1.81e-06 ***
Residuals 502   26275     52.3
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
17265 observations deleted due to missingness
    
```

(가설) - [참고]

H1=농기구 사용여부와 부상부위는 상관없다,
H11= 농기구 사용여부와 부상부위는 상관이 있다.

H2=농기계 사용여부와 부상부위는 같다.
H12=농기계 사용여부와 부상부위는 다르다.

p값

$5.92e-05 < 0.05$, Reject H_0 ! 부상부위(aq8_1_1)는 농기구 사용여부(aq7_1_1)는 다르다.

$1.81e-06 < 0.05$, Reject H_0 ! 부상부위(aq8_1_1)는 농기계 사용여부(aq7_2_1)와는 다르다.

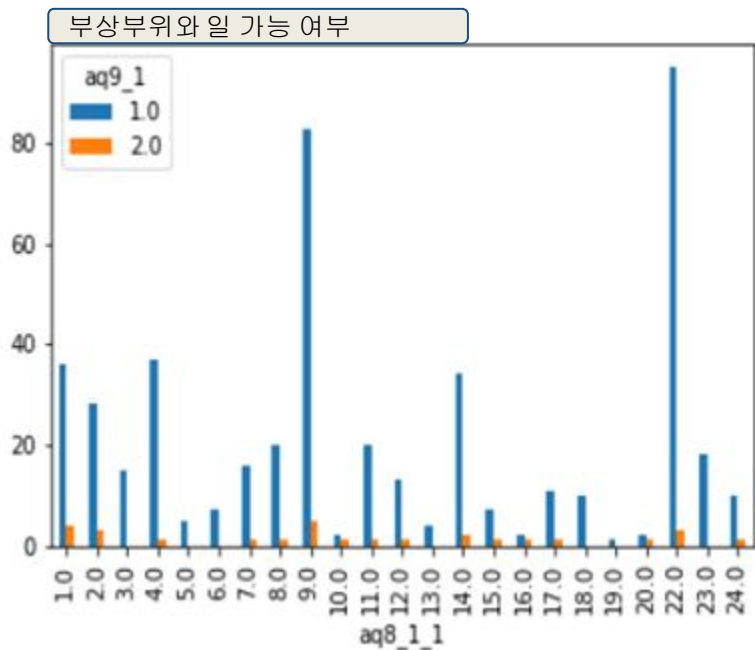
[결론]

부상부위

(aq8_1_1)는 농기구 사용여부(aq7_1_1)와 농기계사용여부(aq7_2_1)는 상호작용이 있다.



일 가능 여부-부상부위&종류



aq8_1_1=부상부위
aq9_1= 일 가능 여부

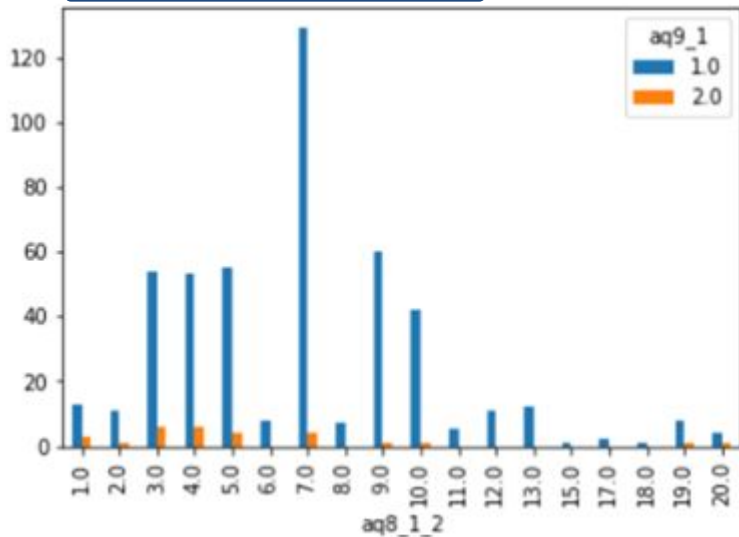
- 일 할 수 있었다고 답 했을 때, 척추(22번)와 다리(9번) 부상률과이 높았다는 것을 알 수 있다.
- 부상 직후, 일 할 수 있었음에도 부상률이 더 높았다는 것을 알 수 있다.



일 가능 여부-부상부위&종류



부상종류와 일 가능 여부



aq8_1_2=부상종류
aq9_1= 일 가능 여부

- 골절(7번) 부상률이 압도적으로 높았다는 것을 알 수 있다.
- 일을 할 수 있었음에도 부상확률이 더 높다는 것을 알 수 있다.



일 가능 여부-부상부위&종류(상관관계 분석)



R코드로 상관관계 분석

```
              Df Sum Sq Mean Sq F value Pr(>F)
aq8_1_1        1  0.042  0.04198   0.302 0.3708
aq8_1_2        1  0.194  0.19427   3.714 0.0545 .
Residuals    501 26.208  0.05231
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
17266 observations deleted due to missingness
```

(가설) - [참고]

H1=일 가능 여부와 부상종류는 다르다,

H11= 일 가능 여부와 부상종류는 같다.

H2=일 가능 여부와 부상부위는 같다.

H12=일 가능 여부와 부상부위는 다르다.

p값

$0.3708 > 0.05$, Reject H_0 !, 일 할 수 있었는지의 여부와 부상종류는 같다.

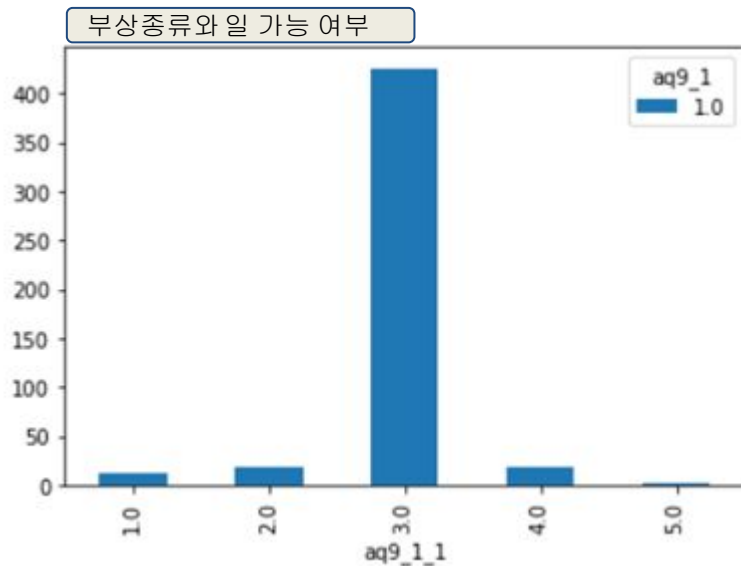
$0.0545 > 0.05$, Reject H_0 !, 일 할 수 있었는지의 여부와 부상부위는 같다.

[결론]

일 가능 여부(aq9_1)는 부상종류(aq8_1_1)와 부상부위(aq8_1_2)는 상호작용이 없다.



일 못한 기간(부상 이후)&일 못한다고 답한 사람



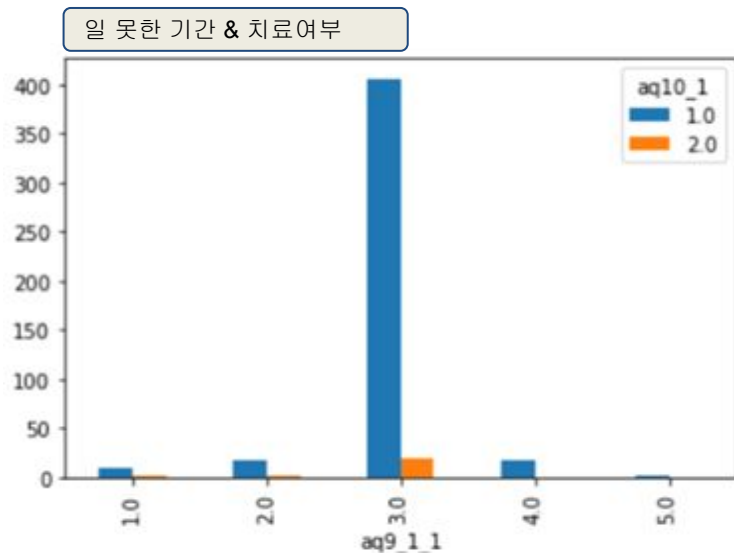
1=반나절 미만 일을 못함
2=반나절 이상, 1일 미만 일을 못함
3=1일 이상 일을 못함
4=영구장애 발생
5=사망

aq9_1_1=일 못한 기간
aq9_1= 일 가능 여부

부상 이후 일을 지속해 가는게 가능했다고 답 한 사람들 중 1일 이상(3번) 일을 못했다고 답한 사람들이 많았다는 것을 알 수 있다. 즉, 부상 이후 1일 이상 일을 못했던 사람이 많았다는 것을 알 수 있다.



치료여부와 일 수행의 영향



aq9_1_1 (일 못한 기간)

1=반나절 미만 일 못함

2=반나절 이상, 1일 미만 일 못함

3=1일 이상 일 못함

4=영구장애 발생

5=사망

aq10_1 (치료여부)

1=치료 함

2=치료 안함

aq9_1_1= 일 못한 기간

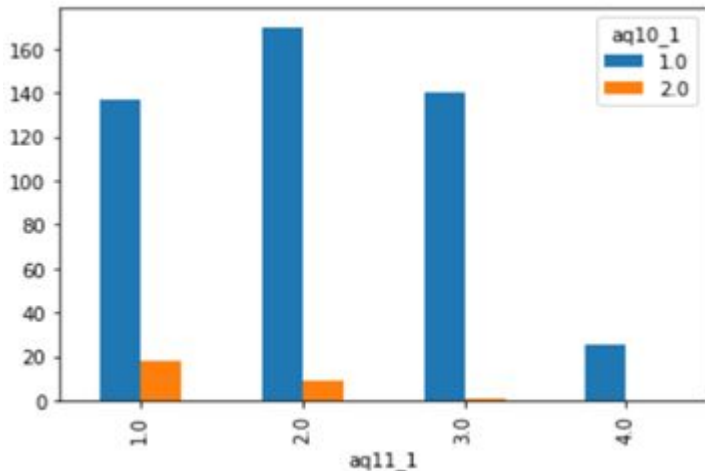
aq9_1= 일 가능 여부

1일 이상 일을 못했다고 답한 이들 중, 치료한 비율이 높았다.



치료여부와 일 수행의 영향

치료여부& 치료이후 상태



aq10_1= 치료여부
aq11_1= 치료 이후 상태

aq10_1(치료 여부)

1=치료 함

2=치료 안 함

aq11_1 (치료이후 상태)

1=농업활동 수행에 전혀 지장 없음

2=손상 전 농업활동 수행 능력이 약간
떨어짐

3=손상 전 농업활동 수행능력이 많이
떨어짐

4=농업활동을 전혀 수행하지 못함

치료한 이들 중, 수행능력이 약간 떨어진 것>수행능력 많이
떨어짐&농업활동에 전혀 지장없음(두 개 비슷하다.)
이러한 결과가 나온 것으로 보아, 많은 사람들이 수행능력이 약간
떨어졌다는 것을 알 수 있다.



치료여부와 일 수행의 영향(상호작용 분석)



R코드로 상관관계 분석

```
      Df Sum Sq Mean Sq F value    Pr(>F)
aq10_1    1   12.1   12.12   17.62 3.22e-05 ***
aq9_1_1    1   37.0   36.98   53.77 9.94e-13 ***
Residuals 470  323.2    0.69
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
17297 observations deleted due to missingness
```

(가설) - [참고]

Ho=치료여부와 치료이후 상태는 상관이 없다,
H1= 치료여부와 치료이후 상태는상관이 있다.

Ho=일 못한 기간과 치료이후 상태는 상관이 없다.
H1=일 못한 기간과 치료이후 상태는 상관이 있다.

[p값]

$3.22e-05 < 0.05$, Reject Ho! 치료여부(aq10_1)는 치료 이후 상태(aq11_1)와 일 못한 기간은 다르다.

$9.94e-13 < 0.05$, Reject Ho! 일 못한 기간(aq9_1_1)과 치료 이후 상태(aq11_1)는 다르다.

[결론]

치료 이후 상태는 치료여부와 일 못한 기간과 상호작용이 있음을 알 수 있다.

감사합니다.