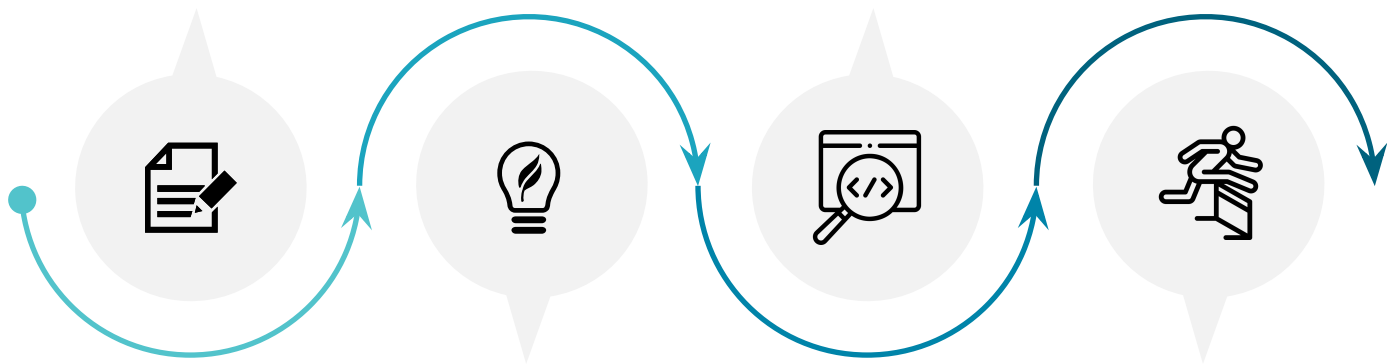


스마트폰 과의존 실태조사 통계



목차



Step 01

- ▶ 수행 기간
- ▶ 시스템 환경
- ▶ 라이브러리

Step 02

- ▶ 분석 배경
- ▶ 필요성

Step 03

- ▶ 가설 설정
- ▶ 데이터 분석

Step 04

- ▶ 한계점
- ▶ 보완점

Step1



시스템 환경 & 라이브러리

▶기간 2023.05.16 ~ 2023.05.25



▶ Window 10



▶ Python 3.8.10



▶Jupyter notebook 6.5.4



▶ Pandas 2.0.1



▶ Matplotlib 3.7.1



▶ Scipy 1.10.1

Step2



분석 배경 & 필요성

정보통신산업의 발달로 인해 다양한 콘텐츠 및 전자기기들의 이용률이 높아지고 있다. 그 중에서도 특히 스마트폰은 현대인들에게 있어 필수품으로 자리잡았으나, 한편에는 이로 인한 부작용에 대한 우려의 목소리도 점차 커지고 있다.

따라서 스마트폰 이용자들의 스마트폰 과의존에 관한 설문조사 데이터를 바탕으로 하여 그 중 스마트폰 과의존성과 가장 연관이 높다고 판단한 일부 변수들을 선택하여 변수 간 연관성과 그로 인해 나타나는 차이점을 파악하는 것에 중점을 두었다.

설문조사 응답을 바탕으로 스마트폰 과의존성을 완화하거나 해소하는데 적합한 제도나 방향에 대한 단초를 제공할 수 있을 것이다.

[데이터 출처]

- 공공데이터 포털 사이트 (<https://www.data.go.kr/>)
- NIA한국지능정보사회진흥원 (https://www.nia.or.kr/site/nia_kor/main.do)

Step3



가설 설정

▶ **스마트폰 과의존 예방교육을 경험한 사람이 경험하지 않은 사람보다**

1. **본인의 하루 평균 스마트폰 이용시간이 더 과도하다고 평가했을 것이다.**
2. **주변인과 비교했을 때 자신의 스마트폰 의존도 평가에 있어서 의존도가 높다고 답했을 것이다.**
3. **우리 사회의 스마트폰 과의존 문제가 심각하다고 답했을 것이다.**
4. **일상생활 속 본인의 스마트폰 이용 적절성이 부적절하다고 답했을 것이다.**

▶ **스마트폰 이용 구성비 중 가장 많은 비중을 차지한 것은 '여가'일 것이다.**

▶ **남성과 여성의 스마트폰 이용 구성비는 5%p 넘는 차이가 있을 것이다.**

▶ **동영상 서비스 평균 이용 시간은**

1. **코로나 격리를 경험한 사람들이 아닌 사람들에 비해 높게 나타날 것이다.**
2. **나이가 10대에 가까울 수록 하루평균이용시간이 높게 나타날 것이다.**
3. **맞벌이 가족인 자녀(유아동, 청소년)들이 아닌 자녀들에 비해 하루평균이용시간이 높게 나타날 것이다.**

* 굵게 기울임 표시한 가설에 대해서만 설명하였으며, 그 외 가설들에 대해서는 첨부된 주피터 노트북 파일 등을 통해 자세한 확인이 가능합니다.

Step3

주요 가설 분석

01

	예방교육	시간과도성	의존도_본인	의존도_사회	이용적절성
0	2	3	3	3.0	3.0
1	2	2	3	3.0	3.0
2	2	3	3	3.0	3.0
3	2	3	3	3.0	3.0
4	2	2	2	3.0	3.0

02

```
edu_y = data.groupby('예방교육')[['시간과도성']].value_counts().sort_index()[1]  
edu_n = data.groupby('예방교육')[['시간과도성']].value_counts().sort_index()[2]
```

03

```
fig, ax = plt.subplots(1,2)  
ax[0].pie(x=edu_y)  
ax[1].pie(x=edu_n)  
  
plt.show()
```

데이터 분석

[주요 가설 1]

▶ 스마트폰 과의존 예방교육을 경험한 사람이 경험하지 않은 사람보다 본인의 하루 평균 스마트폰 이용시간이 더 과도하다고 평가했을 것이다.

01

▶ 설문조사 결과를 바탕으로 1차 추출한 정제된 데이터

02

▶ 예방교육 경험 여부를 그룹화 기준으로 삼고, 본인의 스마트폰 이용 시간 과도성을 평가한 컬럼을 뽑은 뒤 문항 번호별 개수를 계산한 결과를 각각 저장한다.

03

▶ 저장한 값을 matplotlib를 활용하여 시각화한다.

*데이터 전처리 과정과 시각화를 꾸며주는 요소들의 경우는 다수 생략하였으며, 원본 코드는 동봉된 주피터 노트북 파일들을 통해 확인 가능합니다. [주요가설 1]의 시각화 커스터마이징 코드는 뒤의 페이지에 별도 표시하였습니다.

Step3

03

시각화 - 상세 코드

```
fig, ax = plt.subplots(1,2)
fig.set_size_inches(9, 6.5)

label = ['전혀 그렇지 않다', '그렇지 않다', '그렇다', '매우 그렇다']
color = ['#006f95', '#1ba4be', 'Salmon', 'Crimson']
explode = [0, 0, 0, 0.1]
wedgeprops={'width':0.6, 'edgecolor': 'silver'}
textprops={'weight':'bold'}

ax[0].pie(x=edu_y,
          autopct='%1.f%%',
          labels=label,
          colors=color,
          shadow=True,
          startangle=90,
          wedgeprops=wedgeprops,
          textprops=textprops,
          explode=explode,
          )
ax[0].set_title('경험 있음',
               fontdict={'size':15, 'weight':'bold'},
               )
```

```
ax[1].pie(x=edu_n,
          autopct='%1.f%%',
          labels=label,
          colors=color,
          shadow=True,
          startangle=90,
          wedgeprops=wedgeprops,
          textprops=textprops,
          explode=explode,
          )
ax[1].set_title('경험 없음',
               fontdict={'size':15, 'weight':'bold'},
               )

plt.suptitle(f"<본인의 하루 스마트폰 사용시간이 과도하다고 느끼십니까?>\n스마트폰 과의존 예방교육 경험 여부에 따른 답변 비율",
            size=15,
            weight='bold',
            x=0.5,
            y=0.9,
            )
plt.legend(ncols=2,
          loc=(0, -0.2),
          labels=label,
          shadow=True,
          )

answer_y = data.groupby('예방교육')[['시간과도성']].value_counts().sort_index()[1].sum()
answer_n = data.groupby('예방교육')[['시간과도성']].value_counts().sort_index()[2].sum()

plt.text(-1.2, -1.9, f"('있음' 응답자: {answer_y}명, '없음' 응답자: {answer_n}명)", bbox={'boxstyle':'round', 'color':'w'})

plt.savefig('edu_1.png')
plt.show()
```

Step3

독립성 검정 - 카이제곱검정

01

```
df_chi = data.groupby('예방교육')[['시간과도성']].value_counts().sort_index()\
.unstack().rename(index=[1:'경험 있음', 2:'경험 없음'],\
                    columns=[1:'전혀 그렇지 않다', 2:'그렇지 않다', 3:'그렇다', 4:'매우 그렇다'])
```

02

시간과도성	전혀 그렇지 않다	그렇지 않다	그렇다	매우 그렇다
예방교육				
경험 있음	56	969	1189	254
경험 없음	1531	11954	7707	1538

03

```
chiresult = chi2_contingency(df_chi)\
print(f'카이제곱 : {chiresult.statistic}')\
print(f'P값 : {chiresult.pvalue}')
```

카이제곱 : 314.9522575375269
P값 : 5.773259545409229e-68

01

▶ 두 변수 간의 독립성 여부를 검증하기 위해 필요한 데이터프레임을 생성하여 df_chi 에 저장한다.

02

▶ 각 항목별 빈도를 표로 나타낸 빈도표

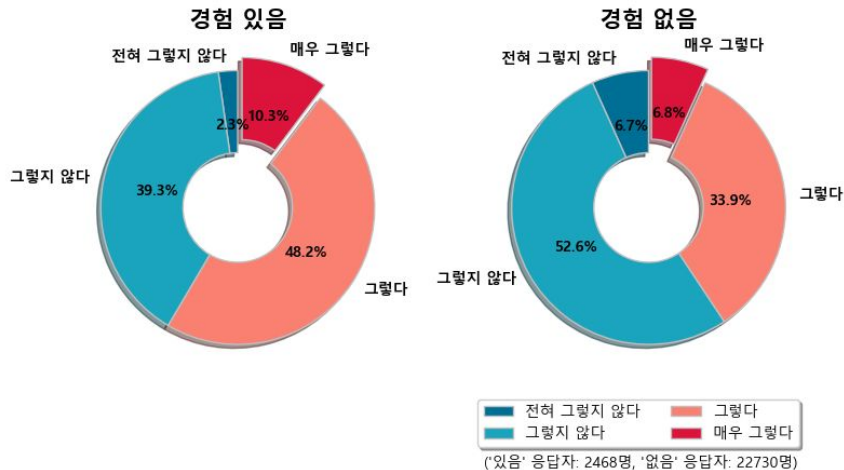
03

▶ scipy 라이브러리의 chi2_contingency 메서드에 df_chi 를 인수로 입력하여, 변수들간의 독립성 여부를 판별하는 주요 지표를 계산한다.

Step3

시각화 & 가설 검증

<본인의 하루 스마트폰 사용시간이 과도하다고 느끼십니까?>
스마트폰 과의존 예방교육 경험 여부에 따른 답변 비율



[변수간 독립성 검정]

- ▶ H0(귀무가설): '스마트폰 과의존 예방교육 경험 여부'와 '본인의 하루 평균 스마트폰 이용 시간의 과도함 평가'는 독립적이다.
- ▶ H1(대립가설): H0이 아니다.
- ▶ 유의수준 (0.05)보다 P값 (5.77e-68)이 낮기 때문에 영가설을 기각한다. 즉, 스마트폰 과의존 예방교육 경험 여부와 본인의 하루 평균 스마트폰 이용 시간의 과도함 평가에는 연관성이 있다.

[결론]

- ▶ 스마트폰 과의존 예방교육을 경험한 사람 중 자신의 스마트폰 사용시간이 '과도하지 않다'라고 응답한 비율은 41.6%이고, '과도하다'라고 응답한 비율은 58.5%이다.
- ▶ 스마트폰 과의존 예방교육을 경험하지 않은 사람 중 자신의 스마트폰 사용시간이 '과도하지 않다'라고 응답한 비율은 59.3%이고, '과도하다'라고 응답한 비율은 40.7%이다.
- ▶ 따라서 스마트폰 과의존 예방교육을 경험한 사람이 그렇지 않은 사람에 비해서 자신의 스마트폰 사용시간이 과도하다고 평가하였으며, 과도함의 정도도 '매우 그렇다'라고 평가한 비율이 더 높았다.

Step3

주요 가설 부서

01

```
d_3_b = data[['나이', '동영상_서비스', '동영상_하루평균이용시간']] \
[data['동영상_서비스'] != 2].drop(columns='동영상_서비스').astype('int8')
```

	나이	동영상_하루평균이용시간
0	33	4
1	40	2
2	59	2
3	49	2
4	55	3

02

```
d_3_b.groupby('나이').mean()
```

03

```
d_3_b.groupby('나이').mean().plot(legend='',
                                figsize=(10,5),
                                )
plt.title(f'나이별 동영상 서비스 1일 평균 시청시간 지표 추이\n(지표상세설명 참고)')
plt.ylabel('')

plt.savefig('나이별 동영상 서비스 1일 평균 시청시간 지표 추이.jpg')
plt.show()
```

데이터 분석

[주요 가설 2]

▶ 나이가 10대에 가까울 수록 하루 평균 이용 시간이 높게 나타날 것이다.

01

▶ 1차 정제된 데이터에서 동영상 서비스를 이용하는 사람들의 나이와 동영상 서비스 일일 평균 이용시간을 추출하였다.

02

▶ 나이에 따른 동영상 서비스의 하루 평균 이용시간의 변화 추이만 비교하기 위해 각 나이에 응답한 번호들의 합을 평균 계산하였다.

▶ 설문조사 문항에 대한 보기는 아래와 같다.

1:'10분 미만', 2:'10분 이상~30분 미만', 3:'30분 이상~1시간 미만',
4:'1시간 이상~2시간 미만', 5:'2시간 이상~6시간 미만', 6:'6시간 이상'

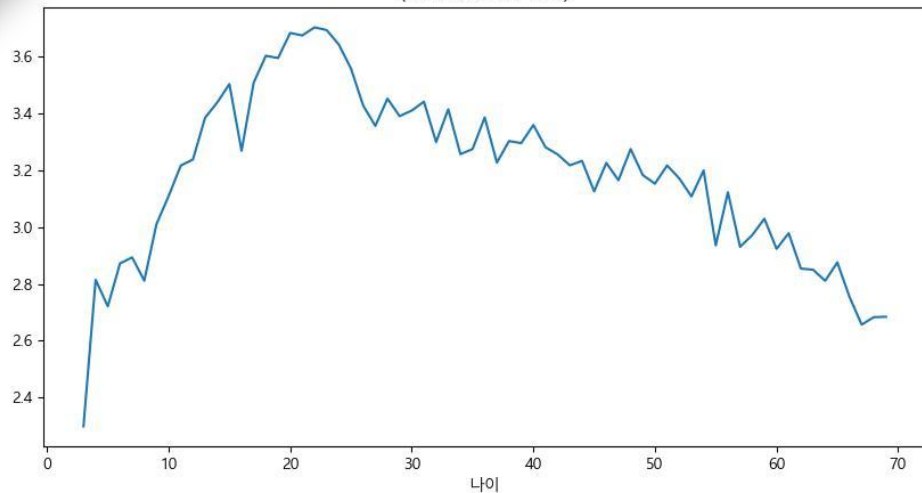
03

▶ 저장한 값을 matplotlib를 활용하여 시각화한다.

Step3

시각화 & 가설 검증

나이별 동영상 서비스 1일 평균 시청시간 지표 추이
(지표상세설명 참고)



【결론】

- ▶ 20대 초반에 가까워질 수록 일일 평균 시청시간 지표 추이가 높아지는 것으로 나타났다.
- ▶ 1일 평균 시청시간 지표에서 나타난 상위 5개 나이는 20~24세로 나타났으며, 22세가 3.70으로 가장 높은 지표를 보였다.
- ▶ 따라서 나이가 10대에 가까울 수록 일일 평균 이용 시간이 높게 나타나는 것이 아니라 22세(20대 초반)에 가까울 수록 하루 평균 이용시간이 높게 나타났다.

Step3

주요 가설 분석

01

```
d_3_c = data[(data.가구원_관계.isin([3, 4])) & (data.동영상_서비스 == 1)]\n[['동영상_이용빈도', '동영상_하루평균이용시간', '맞벌이']].astype('int8')
```

동영상_이용빈도	동영상_하루평균이용시간	맞벌이
10850	1	1
11001	2	4
11139	1	6
11257	1	4
11283	1	4

02

```
dt_freq_video = d_3_c.groupby(['맞벌이']).mean().동영상_이용빈도.values\n dt_mt_video = d_3_c.groupby(['맞벌이']).mean().동영상_하루평균이용시간.values
```

데이터 분석

[주요 가설 3]

▶ 맞벌이 가족인 자녀(유아동, 청소년)들이 아닌 자녀들에 비해 동영상 서비스 이용 빈도와 일일 평균 이용 시간이 높게 나타날 것이다.

01

▶ 자녀(유아동, 청소년)에 해당하는 행들의 동영상 서비스 이용 빈도, 일일 평균 이용 시간 및 맞벌이 가족 여부에 해당하는 데이터를 추출한다.

02

▶ 맞벌이 가족 여부를 기준으로 동영상 서비스 이용 빈도, 일일 평균 이용 시간 항목의 평균값을 리스트 자료형으로 각각 저장한다.

03

▶ 저장한 값을 matplotlib를 활용하여 시각화한다.

(뒷 페이지에 코드설명)

Step3

03

시각화 - 상세 코드

```
dt_freq_video = d_3_c.groupby(['맞벌이']).mean().동영상_이용빈도.values
dt_mt_video = d_3_c.groupby(['맞벌이']).mean().동영상_하루평균이용시간.values
group_dt_video_x = ('맞벌이', '외벌이')
group_dt_video_y = {
    '동영상 서비스 이용 빈도 지표':dt_freq_video,
    '동영상 서비스 일일 평균 이용 시간 지표':dt_mt_video
}

x = np.arange(len(group_dt_video_x)) # the label locations
width = 0.4 # the width of the bars
multiplier = 0
color=['#1ba4be', 'Salmon']
i = 0

fig, ax = plt.subplots(layout='constrained')
# fig, ax = plt.subplots()
fig.set_size_inches(4, 5)
```

```
for attribute, measurement in group_dt_video_y.items():
    offset = width * multiplier
    rects = ax.bar(x + offset, measurement, width*0.9, label=attribute, align='edge', color=color[i])
    ax.bar_label(rects, padding=3, fmt='%.2f')
    multiplier += 1
    i += 1

# Add some text for labels, title and custom x-axis tick labels, etc.
ax.set_ylabel('')
ax.set_title('맞벌이 여부에 따른 자녀들의 동영상 서비스 이용 지표', pad=20)
ax.set_xticks(x + width, edu)
ax.legend(loc='upper left', ncols=1, shadow=True)
ax.set_ylim(0, 4)
plt.grid(axis='y', alpha=0.5, ls=':')

plt.savefig('맞벌이 여부에 따른 자녀들의 동영상 서비스 이용 지표.jpg')
plt.show()
```

Step3

독립성 검정 - 카이제곱검정

01

```
df_chi_1 = d_3_c.groupby('맛별이')['동영상_이용빈도'].value_counts().unstack()\n.rename(index={1:'맛별이', 2:'외별이'},\n        columns={1:'매일', 2:'일주일에 5~6일', 3:'일주일에 3~4일', 4:'일주일에 1~2일', 5:'한달에 1~3일'})
```

02

동영상_이용빈도	매일	일주일에 5~6일	일주일에 3~4일	일주일에 1~2일	한달에 1~3일
맛별이	1087	507	371	115	13
외별이	638	276	241	77	8

03

```
chi_result_1 = chi2_contingency(df_chi_1)\nprint(f'카이제곱값 : {chi_result_1[0]}')\nprint(f'p-value : {chi_result_1[1]}')
```

```
카이제곱값 : 3.2538486533851256\np-value : 0.5162776264762328
```

01

▶ 두 변수 간의 독립성 여부를 검증하기 위해 필요한 데이터프레임을 생성하여 df_chi_1에 저장한다.

02

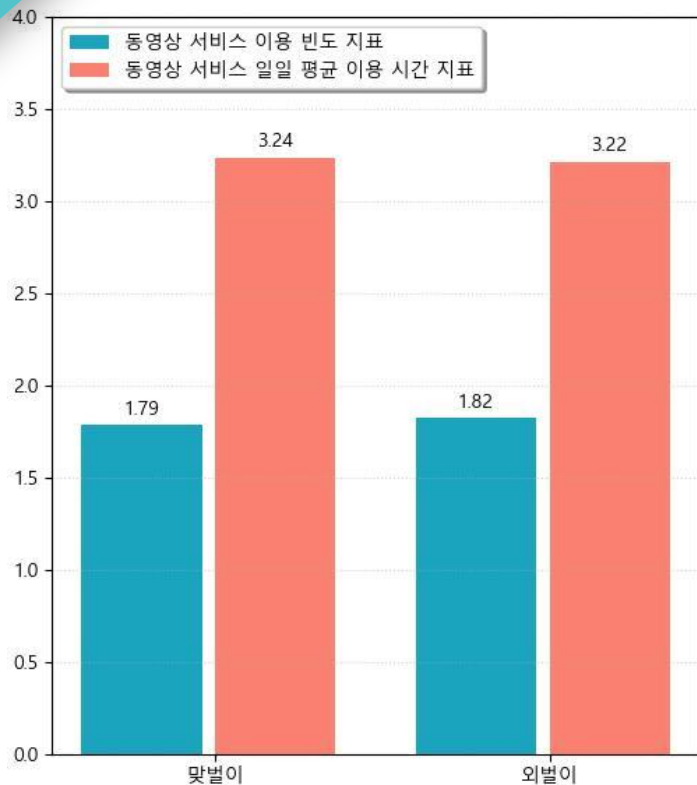
▶ 각 항목별 빈도를 표로 나타낸 빈도표

03

▶ scipy 라이브러리의 chi2_contingency 메서드에 df_chi_1 를 인수로 입력하여, 변수들간의 독립성 여부를 판별하는 주요 지표를 계산한다.

Step3

맞벌이 여부에 따른 자녀들의 동영상 서비스 이용 지표



시각화 & 가설 검증

【변수간 독립성 검증】

- ▶ H0(귀무가설) : '자녀(유아동, 청소년)들의 '맞벌이 가족 여부'와 '동영상 서비스 일일 평균 이용 시간'은 독립적이다.
- ▶ H1(대립가설) : H0이 아니다.
- ▶ 유의수준 (0.05)보다 P값(0.64)이 낮기 때문에 영가설을 기각한다. 즉, 자녀들의 맞벌이 가족 여부와 동영상 서비스 일일 평균 이용 시간에는 연관성이 없다.

【결론】

- ▶ 맞벌이 부모의 자녀들의 동영상 서비스의 이용빈도 지표와 평균 이용 시간 지표는 각각 1.79, 3.24으로 나타났으며, 외벌이 부모의 자녀들의 경우는 각각 1.82, 3.22으로 나타났다.
- ▶ 따라서 동영상 서비스 이용 빈도 지표의 경우는 외벌이 부모의 자녀들이 1.82로 맞벌이 부모의 자녀들보다 0.03 높게 나온 반면, 동영상 서비스 일일 평균 이용 시간 지표에서는 맞벌이 부모의 자녀들이 3.24로 외벌이 부모의 자녀들보다 0.02 높게 나왔다.

Step4



한계점

▶ [주요가설 1]은 개인에 따라 평가가 달라지는 문항이기 때문에 실제 스마트폰 과의존도와 개인의 평가 사이에 괴리가 존재할 가능성이 있다.

▶ [주요가설 2]는 범주형 데이터로 표현되었으나, 실제로는 연속형 데이터이다. 그러나 설문조사 설계의 문제로 인해 범주형 데이터로 표현되었으며, 그 범주마저도 특정한 간격을 두고 나뉘어지지 않기 때문에 정확한 평균 시간을 계산할 수 없는 한계가 있다.

보기를 구성하는 범위도 보기의 번호가 커질 수록 매우 커지도록 구성이 되어 있어 높은 번호의 보기들의 경우는 같은 보기에 응답하였더라도 실제 값들이 최대 4시간 이상의 차이가 존재하는 경우도 있을 수 있다.

특히 나이대별 상대비교를 위해 보기의 번호들을 해당 숫자로 치환하여 평균을 구하였기 때문에 실제 연령별 일일 스마트폰 사용 평균 시간과 차이가 클 수 있으며, 지표값이 비슷한 경우에는 순위가 정확하지 않을 가능성이 있다.

Step4



보완점

- ▶ 선형 회귀분석을 통해 각 문항들과 스마트폰 과의존 정도 간의 상관계수를 구하여 스마트폰 과의존에 악영향을 미치는 요소가 무엇인지 파악하는 것이 필요하다.
- ▶ 동영상 플랫폼의 이용 빈도와 일일 사용 시간만을 갖고 분석하였으나, 동영상 플랫폼 내에서 가장 많이 시청하는 분야들에 대한 데이터 항목이 존재하기 때문에 해당 데이터를 바탕으로 동영상 플랫폼이 주로 오락을 목적으로 이용되는지, 그 외 정보 수집 등의 목적으로 이용되는지를 분석함으로써 스마트폰을 통한 동영상 플랫폼의 장단점의 영향을 보다 자세하게 살펴볼 수 있을 것이다.
- ▶ [주요 가설 외] 차후년도의 설문조사 데이터에도 코로나 자가격리 경험 여부에 관한 항목이 존재한다면, 코로나 자가격리 경험에 따라 스마트폰 사용량의 차이가 나타나는 것이 지속적으로 영향을 미치는지 분석할 수 있고, 자가격리 시점에 관한 더 자세한 데이터가 있다면 보다 더 정밀한 분석이 가능할 것으로 예상된다.