



보고서

1. 프로젝트 개요
2. 프로젝트 팀 구성 및 역할
3. 프로젝트 수행 절차 및 방법
4. 분석 내용 및 결과
 - 1) 의료 서비스 이용 만족도 분석
 - 2) 지역별/연령별 병원 미충족의료 경험 및 원인 파악
 - 3) 당뇨병 유병에 영향을 미치는 원인 파악 및 진단 예측
5. 프로젝트 회고 및 개선점
6. 부록

1. 프로젝트 개요

1-1. 분석 주제 및 목적



의료 패널 데이터 분석에 따른 개선안 도출

1. 의료 서비스 이용 만족도 분석

- 환자들의 의료 서비스 만족도 현황을 확인하고 어떤 측면에서 불만 사항이 발생하고 있는지 식별함
- 만족도 및 불만족도와 관련된 주요 요인을 파악하고 개선 사항을 도출함

3. 지역별/연령별 병원 미충족의료 경험 및 원인 파악

- 어떤 지역 혹은 연령층에서 의료 서비스 부족 현상이 심각한지 확인함
- 미충족 의료에 대한 불평등을 최소화할 수 있는 방안을 모색함



미충족 의료율이란?

- 지난 1년간 병원(치과 제외) 치료 또는 검사를 받아볼 필요가 있었으나, 받지 못한 적이 한 번이라도 있었다고 응답한 사람의 비율

2. 당뇨병 유병에 영향을 미치는 원인 파악 및 진단 예측

- 당뇨병 발병의 주요 원인을 파악하고 예방을 위한 개선 방안을 제시함
- 데이터 기반의 당뇨병 진단 예측 모델을 개발하여 조기 진단과 관리에 기여함



왜 당뇨병인가?

- 전 세계적으로 증가하는 만성질환 중 하나로, 당뇨병의 원인은 유전적 요인 및 환경적 요인들이 복합적으로 작용해서 생기는 경우가 많음
- 단순히 혈당만 상승하는 병이 아니라, 잘 조절하지 않으면 시간이 지날수록 많은 합병증을 유발함

1-2. 주제 선정 배경

30대 8% 제때 병원 못 갔다...건강 만족도 37% 그쳐

30세 이상 40세 미만 인구 8%가 제때 의료서비스를 이용하지 못하는 것으로 나타났다. 본인 건강 상태에 만족하는 국민은 37% 수준에 그쳤다. 경제 활동으로 의료서비스 접근이 제약된 젊은층이 고령층보다 만족도

 <http://www.docdocdoc.co.kr/news/articleView.html?idxno=3004470>



"당뇨병 유병자 75% '관리' 안 된다"...증가 추이도 '심각'

500만 당뇨인 시대는 지났다. 최근 대한당뇨병학회가 공개한 (DFS 2022)에 따르면 2020년 기준 국내 당뇨병 인구는 570만 1000명(30세 이상 526만 9000명)에 이른 것으로 나타났다. 게다가 당뇨병으로 이환될

 <http://www.doctorsnews.co.kr/news/articleView.html?idxno=146606>



- 대한민국의 의료 서비스와 환자 경험의 문제점을 파악하고 이를 다각적으로 이해하고 개선하고자 함
- 세가지의 분석 주제(의료 서비스, 의료 접근성, 만성질환 예방 및 관리)를 통하여 의료 서비스의 품질 개선과 의료 격차 해소에 기여하고자 함
- 패널 데이터의 장점을 활용하고자 **한국의료패널** 데이터로 선정하게 됨
 - 구체적이고 질 좋은 데이터로 인하여 분석 시 편향 감소

- 매년 동일한 7,000~8,000 가구를 대상으로 의료이용 행태 및 의료비 지출에 대해 조사

2. 프로젝트 팀 구성 및 역할



팀명 : 건강이 최고조

팀원 : 김예린(팀장), 김수민, 민윤홍, 황상엽

역할 분담

Aa Name	Tags	Description	Link
<u>아이디어 회의</u>	공통		
<u>자료조사</u>	공통		
<u>데이터 탐색 및 EDA(IND)</u>	김예린		
<u>일정 관리</u>	김예린		
<u>회의록 작성</u>	김예린		
<u>병의원 미충족의료 경험 분석</u>	김예린		
<u>데이터 전처리</u>	공통		
<u>기획서 작성</u>	공통		
<u>보고서 작성</u>	공통		
<u>발표 자료 제작</u>	공통		
<u>데이터 탐색 및 EDA(HH)</u>	김수민		
<u>데이터 수집</u>	김수민		
<u>의료 서비스 이용 만족도 분석</u>	김수민		
<u>데이터 탐색 및 EDA(MS)</u>	민윤홍		
<u>데이터 수집</u>	민윤홍		

Aa Name	Tags	Description	Link
<u>당뇨병 유병의 원인 파악 및 진단 예측 (회귀 분석)</u>	민윤홍		
<u>발표</u>	민윤홍		
<u>데이터 탐색 및 EDA (PHI)</u>	황상엽		
<u>의료 서비스 이용 만족도 분석</u>	황상엽		

3. 프로젝트 수행 절차 및 방법

3-1. 데이터 설명 (데이터 출처, 데이터 개요)



한국의료패널

- 매년 동일한 7,000~8,000 가구를 대상으로 의료이용 행태 및 의료비 지출에 대해 조사하는 국내 최초의 의료패널조사
 - 조사목적상 전국규모의 대표성을 유지하기 위해 2005년 인구주택총조사 90% 조사구중 섬조사구, 특수시설, 기숙시설을 제외한 조사구를 추출틀로 사용하고 있음
 - 한국보건사회연구원과 국민건강보험공단이 공동으로 수행하는 조사로 보건 의료비용과 의료비 지출 수준의 변화를 파악하여 보건의료정책 및 건강보험정책 수립의 기초자료로 활용함
- 급속한 고령화로 인한 인구구조를 포함한 인구사회경제적 환경변화, 질병구조의 변화, 의료 서비스의 수요 다양화 등으로 인해 우리나라 의료 이용 패턴 및 의료비 양상은 크게 변화하고 있음
- ⇒ 이러한 미래 환경변화 및 의료서비스 수요 변화에 적극적으로 대응하기 위해서는 **한국의료패널**과 같은 종단자료 분석이 필요함
- 의료이용에 영향을 미치는 요인과 의료이용/의료비를 포괄하는 자료원을 구축하여 이를 활용한 정책 평가 및 모니터링을 통해 보건의료체계 성과와 국민건강 증진에 기여함

3-2. 데이터 구성

DB명	단위	설문영역
ID	가구또는 가구원	ID 및 가구원 변동사항
가구 데이터(HH)	가구	가구 일반 사항, 소득2, 부채 및 자산, 지출, 일반의약품 및 의료 관련 지출, 가구 의료비 부담, 가구별 민간의료보험 가입
가구원 데이터 (IND)	가구원	가구원 일반사항, 경제활동 상태, 소득1, 만성질환 관리, 장기요양 및 기타 돌봄서비스 이용, 건강생활 습관(신체활동, 흡연, 음주 등), 건강수준, 의료접근성, 상용치료원, 민간의료보험2, 의료서비스 이용 및 민간의료보험 관련
의료서비스 이용 (MS)	의료이용건별	의료 이용 공통(A: 의료 이용 날짜 등), 응급 서비스 이용, 입원 서비스 이용(의과/치과/한방), 외래 서비스 이용(의과/치과/한방), 건강검진 이용, 의료이용 공통(B: 의료비 등)
민간의료보험 (PHI)	민간의료보험건별	민간의료보험
부가조사_건강 정보 이해능력 (HLIT)		

2019년 / 2020년 / 2021년 의료패널 데이터

```

files
├── 2019 의료패널 데이터
│   ├── a_id.csv
│   ├── a_hh.csv
│   ├── a_ind.csv
│   ├── a_ms.csv
│   └── a_phi.csv
├── 2020 의료패널 데이터
│   ├── b_hh.csv
│   ├── b_ind.csv
│   ├── b_ms.csv
│   ├── b_phi.csv
│   └── hlit.csv
└── 2021 의료패널 데이터
    ├── c_hh.csv
    ├── c_ind.csv
    ├── c_ms.csv
    ├── c_phi.csv
    └── hlit.csv

```

3-3. 데이터 샘플

▼ HH(가구 데이터)

```
a_hh = pd.read_excel('/content/a_hh.csv')
a_hh.head()
```

	HHID	H_WGC	DATE_Y	DATE_M	DATE_D	REGION1	REGION2	FAM_N	IND_N	DEATH_YN	...
0	112001011	4039.375590	2020	7	26	26	1	1	1	0	...
1	112002011	2261.090118	2020	6	22	26	1	2	1	0	...
2	112003011	2122.429913	2020	7	7	26	1	4	4	0	...
3	112010011	1508.932397	2020	6	19	26	1	2	2	0	...
4	112012011	2109.530794	2020	6	18	26	1	2	2	0	...

5 rows × 75 columns

`a_hh1.shape` (6748, 75)

▼ IND (가구원 데이터)

```
a_ind = pd.read_csv('/content/a_ind.csv')
a_ind.head()
```

	HHID	PIDWON	I_WGC	REGION1	H_INC_TOT	DEATH_I_YN	PRE_RES	HH_REL	SEX	BIRTH_Y	...
0	112001011	11200101	3066.669714	26	627	2	1	10	2	1938	...
1	112002011	11200201	1297.872486	26	979	2	1	10	2	1946	...
2	112002011	11200202	2728.104905								...
3	112003011	11200301	3422.030522	26	4497	2	1	10	1	1975	...
4	112003011	11200302	2898.667157	26	4497	2	1	20	2	1978	...

5 rows × 294 columns

`a_ind.shape` (16587, 294)

▼ MS (의료서비스 이용)

```
a_ms = pd.read_csv('/content/a_ms.csv')
a_ms.head()
```

	HHID	PIDWON	YEAR	MS1	DATE1_Y	DATE1_M	DATE1_D	DATE1	ER_DAY	IN_DIS	...
0	112001011	11200101	2020	3	2019	1	17	1/17/2019			...
1	112001011	11200101	2020	3	2019	1	17	1/17/2019			...
2	112001011	11200101	2020	3	2019	2	7	2/7/2019			...
3	112001011	11200101	2020	3	2019	2	7	2/7/2019			...
4	112001011	11200101	2020	3	2019	2	28	2/28/2019			...

5 rows × 196 columns

`a_ms.shape` (260989, 196)

3-4. 데이터 수집 및 전처리

<https://www.khp.re.kr:444/web/data/data.do>



데이터 수집 절차

1. 데이터 활용동의서 다운로드 및 작성
2. 담당자에게 데이터활용동의서 이메일 또는 팩스 전송
3. (이메일에서) 데이터 다운로드
4. SPSS 형식에서 csv 형식으로 변환

▼ 결측치 처리

```
#[_hh]
#지출비용 결측치 = 이용하지않음 => 0원처리
def trans_blank(param):
    df[param] = df[param].astype(str)
    df[param] = df[param].apply(lambda x: '0' if x.strip() == '' else x)
    df[param] = df[param].astype(int)
```

```
#[_ind]
a_ind1 = a_ind[a_ind['REGION1']!= ' ']
a_ind1.reset_index(drop=True, inplace=True)
```

```
# 수치형 변환
a_ind2 = a_ind1.copy()
label_encoder = LabelEncoder()
for i in a_ind2.columns[2:]:
    a_ind2[i] = label_encoder.fit_transform(a_ind2[i])
```

▼ 컬럼 이름 재설정 (한국어)

```
a_ind1 = a_ind1[['HHID', 'PIDWON', 'HT', 'WT', 'BIRTH_Y',
                 'REGION1', 'SEX', 'MARR', 'EDU',
                 'DISA_YN', 'ECO1', 'HEALTH_INS',
                 'CD', 'CD1_HTN', 'CD2_HTN', 'CD1_DM', 'CD2_DM',
                 'CD1_MD', 'CD2_MD', 'CD1_DMN', 'CD2_DMN',
                 'P1', 'P1_1', 'WTMG', 'S1', 'D1', 'HS1', 'HS_SRH', ]]

a_ind1.columns = ['가구식별번호', '가구원고유번호', '키', '몸무게', '출생년도',
                  '시도', '성별', '결혼상태', '교육수준',
                  '장애유무', '경제활동 참여 상태', '의료보장 형태',
                  '만성질환 유무', '고혈압 유무', '고혈압 진단시기', '당뇨병 유무', '당뇨병 진단시기',
                  '우울증/조울증 유무', '우울증/조울증 진단시기', '치매 유무', '치매 진단시기',
                  '규칙적 운동 실천 여부', '규칙적으로 운동하지 않는 이유', '체중 조절 시도 여부',
                  '평생 흡연 여부', '음주 빈도', '스트레스 인지 정도', '주관적 건강상태']
```

▼ 파생변수 생성

```
#[_ind]
# 키 & 몸무게 => BMI
df['HT'] = df['HT']/100
df['BMI'] = df['WT']/(df['HT']*df['HT'])
df['BMI']

def categorize_bmi(bmi):
    if bmi <= 18.5:
        return 1 #'저체중'
    elif 18.5 <= bmi <= 22.9:
        return 2 #'정상'
    elif 23.0 <= bmi <= 24.9:
        return 3 #'과체중'
    else:
        return 4 #'비만'

df['BMI_category'] = df['BMI'].apply(categorize_bmi)

# 출생년도 -> 연령
a_ind2['연령'] = 2020-a_ind1['출생년도'].astype('int')

# 보유한 총 만성질환 개수
```



```
selected_df['CD1_TOTAL'] = selected_df[selected_df.columns[selected_df.columns.str.starts
with('CD1')]].eq('1').sum(axis=1)
```

```
#[_phi]
# '가구식별번호' 열을 기준으로 중복된 행의 수를 계산
duplicate_counts = df_phi1.groupby('가구식별번호').size().reset_index(name='보험 가입 수')

#[_ind]
# 연령대를 나타내는 열을 추가
CD1['Age_Group'] = ''

# 숫자형으로 바꾸기
CD1['BIRTH_Y'] = pd.to_numeric(CD1['BIRTH_Y'], errors='coerce')

# 각 연령대에 해당하는 사람들을 나누어서 'Age_Group' 열에 값 설정.
CD1.loc[CD1['BIRTH_Y'] >= 2000, 'Age_Group'] = '10대'
CD1.loc[(CD1['BIRTH_Y'] >= 1990) & (CD1['BIRTH_Y'] < 2000), 'Age_Group'] = '20대'
CD1.loc[(CD1['BIRTH_Y'] >= 1980) & (CD1['BIRTH_Y'] < 1990), 'Age_Group'] = '30대'
CD1.loc[(CD1['BIRTH_Y'] >= 1970) & (CD1['BIRTH_Y'] < 1980), 'Age_Group'] = '40대'
CD1.loc[(CD1['BIRTH_Y'] >= 1960) & (CD1['BIRTH_Y'] < 1970), 'Age_Group'] = '50대'
CD1.loc[(CD1['BIRTH_Y'] >= 1950) & (CD1['BIRTH_Y'] < 1960), 'Age_Group'] = '60대'
CD1.loc[(CD1['BIRTH_Y'] >= 1940) & (CD1['BIRTH_Y'] < 1950), 'Age_Group'] = '70대'
CD1.loc[(CD1['BIRTH_Y'] >= 1930) & (CD1['BIRTH_Y'] < 1940), 'Age_Group'] = '80대'
CD1.loc[(CD1['BIRTH_Y'] < 1930), 'Age_Group'] = '90대 이상'
```

```
#재난적 의료비
merged_df['재난적의료비'] = merged_df['MEXP3_1'] / merged_df['H_INC_TOT'] *100
```

```
# 소득분위 산정
# 'inc_r' 열 초기화
hh_a_a['inc_r'] = np.nan

# 조건에 따라 inc_r 열 설정
hh_a_a.loc[(hh_a_a['FAM_N'] == 1) & (hh_a_a['H_INC_TOT'] <= 2040*0.3), 'inc_r'] = 1
hh_a_a.loc[(hh_a_a['FAM_N'] == 2) & (hh_a_a['H_INC_TOT'] <= 3480*0.3), 'inc_r'] = 1
hh_a_a.loc[(hh_a_a['FAM_N'] == 3) & (hh_a_a['H_INC_TOT'] <= 4512*0.3), 'inc_r'] = 1
hh_a_a.loc[(hh_a_a['FAM_N'] == 4) & (hh_a_a['H_INC_TOT'] <= 5532*0.3), 'inc_r'] = 1
hh_a_a.loc[(hh_a_a['FAM_N'] == 5) & (hh_a_a['H_INC_TOT'] <= 6552*0.3), 'inc_r'] = 1
hh_a_a.loc[(hh_a_a['FAM_N'] == 6) & (hh_a_a['H_INC_TOT'] <= 7584*0.3), 'inc_r'] = 1
hh_a_a.loc[(hh_a_a['FAM_N'] == 7) & (hh_a_a['H_INC_TOT'] <= 8604*0.3), 'inc_r'] = 1
hh_a_a.loc[(hh_a_a['FAM_N'] == 8) & (hh_a_a['H_INC_TOT'] <= 9564*0.3), 'inc_r'] = 1

hh_a_a.loc[(hh_a_a['FAM_N'] == 1) & (hh_a_a['H_INC_TOT'] > 2040*0.3) & (hh_a_a['H_INC_TO
T'] <= 2040*0.4), 'inc_r'] = 2
hh_a_a.loc[(hh_a_a['FAM_N'] == 2) & (hh_a_a['H_INC_TOT'] > 3480*0.3) & (hh_a_a['H_INC_TO
T'] <= 3480*0.4), 'inc_r'] = 2
hh_a_a.loc[(hh_a_a['FAM_N'] == 3) & (hh_a_a['H_INC_TOT'] > 4512*0.3) & (hh_a_a['H_INC_TO
```

```

T'] <= 4512*0.4), 'inc_r'] = 2
hh_a_a.loc[(hh_a_a['FAM_N'] == 4) & (hh_a_a['H_INC_TOT'] > 5532*0.3) & (hh_a_a['H_INC_TO
T'] <= 5532*0.4), 'inc_r'] = 2
hh_a_a.loc[(hh_a_a['FAM_N'] == 5) & (hh_a_a['H_INC_TOT'] > 6552*0.3) & (hh_a_a['H_INC_TO
T'] <= 6552*0.4), 'inc_r'] = 2
hh_a_a.loc[(hh_a_a['FAM_N'] == 6) & (hh_a_a['H_INC_TOT'] > 7584*0.3) & (hh_a_a['H_INC_TO
T'] <= 7584*0.4), 'inc_r'] = 2
hh_a_a.loc[(hh_a_a['FAM_N'] == 7) & (hh_a_a['H_INC_TOT'] > 8604*0.3) & (hh_a_a['H_INC_TO
T'] <= 8604*0.4), 'inc_r'] = 2
hh_a_a.loc[(hh_a_a['FAM_N'] == 8) & (hh_a_a['H_INC_TOT'] > 9564*0.3) & (hh_a_a['H_INC_TO
T'] <= 9564*0.4), 'inc_r'] = 2

hh_a_a.loc[(hh_a_a['FAM_N'] == 1) & (hh_a_a['H_INC_TOT'] > 2040*0.4) & (hh_a_a['H_INC_TO
T'] <= 2040*0.5), 'inc_r'] = 3
hh_a_a.loc[(hh_a_a['FAM_N'] == 2) & (hh_a_a['H_INC_TOT'] > 3480*0.4) & (hh_a_a['H_INC_TO
T'] <= 3480*0.5), 'inc_r'] = 3
hh_a_a.loc[(hh_a_a['FAM_N'] == 3) & (hh_a_a['H_INC_TOT'] > 4512*0.4) & (hh_a_a['H_INC_TO
T'] <= 4512*0.5), 'inc_r'] = 3
hh_a_a.loc[(hh_a_a['FAM_N'] == 4) & (hh_a_a['H_INC_TOT'] > 5532*0.4) & (hh_a_a['H_INC_TO
T'] <= 5532*0.5), 'inc_r'] = 3
hh_a_a.loc[(hh_a_a['FAM_N'] == 5) & (hh_a_a['H_INC_TOT'] > 6552*0.4) & (hh_a_a['H_INC_TO
T'] <= 6552*0.5), 'inc_r'] = 3
hh_a_a.loc[(hh_a_a['FAM_N'] == 6) & (hh_a_a['H_INC_TOT'] > 7584*0.4) & (hh_a_a['H_INC_TO
T'] <= 7584*0.5), 'inc_r'] = 3
hh_a_a.loc[(hh_a_a['FAM_N'] == 7) & (hh_a_a['H_INC_TOT'] > 8604*0.4) & (hh_a_a['H_INC_TO
T'] <= 8604*0.5), 'inc_r'] = 3
hh_a_a.loc[(hh_a_a['FAM_N'] == 8) & (hh_a_a['H_INC_TOT'] > 9564*0.4) & (hh_a_a['H_INC_TO
T'] <= 9564*0.5), 'inc_r'] = 3

hh_a_a.loc[(hh_a_a['FAM_N'] == 1) & (hh_a_a['H_INC_TOT'] > 2040*0.5) & (hh_a_a['H_INC_TO
T'] <= 2040), 'inc_r'] = 4
hh_a_a.loc[(hh_a_a['FAM_N'] == 2) & (hh_a_a['H_INC_TOT'] > 3480*0.5) & (hh_a_a['H_INC_TO
T'] <= 3480), 'inc_r'] = 4
hh_a_a.loc[(hh_a_a['FAM_N'] == 3) & (hh_a_a['H_INC_TOT'] > 4512*0.5) & (hh_a_a['H_INC_TO
T'] <= 4512), 'inc_r'] = 4
hh_a_a.loc[(hh_a_a['FAM_N'] == 4) & (hh_a_a['H_INC_TOT'] > 5532*0.5) & (hh_a_a['H_INC_TO
T'] <= 5532), 'inc_r'] = 4
hh_a_a.loc[(hh_a_a['FAM_N'] == 5) & (hh_a_a['H_INC_TOT'] > 6552*0.5) & (hh_a_a['H_INC_TO
T'] <= 6552), 'inc_r'] = 4
hh_a_a.loc[(hh_a_a['FAM_N'] == 6) & (hh_a_a['H_INC_TOT'] > 7584*0.5) & (hh_a_a['H_INC_TO
T'] <= 7584), 'inc_r'] = 4
hh_a_a.loc[(hh_a_a['FAM_N'] == 7) & (hh_a_a['H_INC_TOT'] > 8604*0.5) & (hh_a_a['H_INC_TO
T'] <= 8604), 'inc_r'] = 4
hh_a_a.loc[(hh_a_a['FAM_N'] == 8) & (hh_a_a['H_INC_TOT'] > 9564*0.5) & (hh_a_a['H_INC_TO
T'] <= 9564), 'inc_r'] = 4

hh_a_a.loc[(hh_a_a['FAM_N'] == 1) & (hh_a_a['H_INC_TOT'] > 2040), 'inc_r'] = 5
hh_a_a.loc[(hh_a_a['FAM_N'] == 2) & (hh_a_a['H_INC_TOT'] > 3480), 'inc_r'] = 5
hh_a_a.loc[(hh_a_a['FAM_N'] == 3) & (hh_a_a['H_INC_TOT'] > 4512), 'inc_r'] = 5
hh_a_a.loc[(hh_a_a['FAM_N'] == 4) & (hh_a_a['H_INC_TOT'] > 5532), 'inc_r'] = 5
hh_a_a.loc[(hh_a_a['FAM_N'] == 5) & (hh_a_a['H_INC_TOT'] > 6552), 'inc_r'] = 5
hh_a_a.loc[(hh_a_a['FAM_N'] == 6) & (hh_a_a['H_INC_TOT'] > 7584), 'inc_r'] = 5
hh_a_a.loc[(hh_a_a['FAM_N'] == 7) & (hh_a_a['H_INC_TOT'] > 8604), 'inc_r'] = 5
hh_a_a.loc[(hh_a_a['FAM_N'] == 8) & (hh_a_a['H_INC_TOT'] > 9564), 'inc_r'] = 5

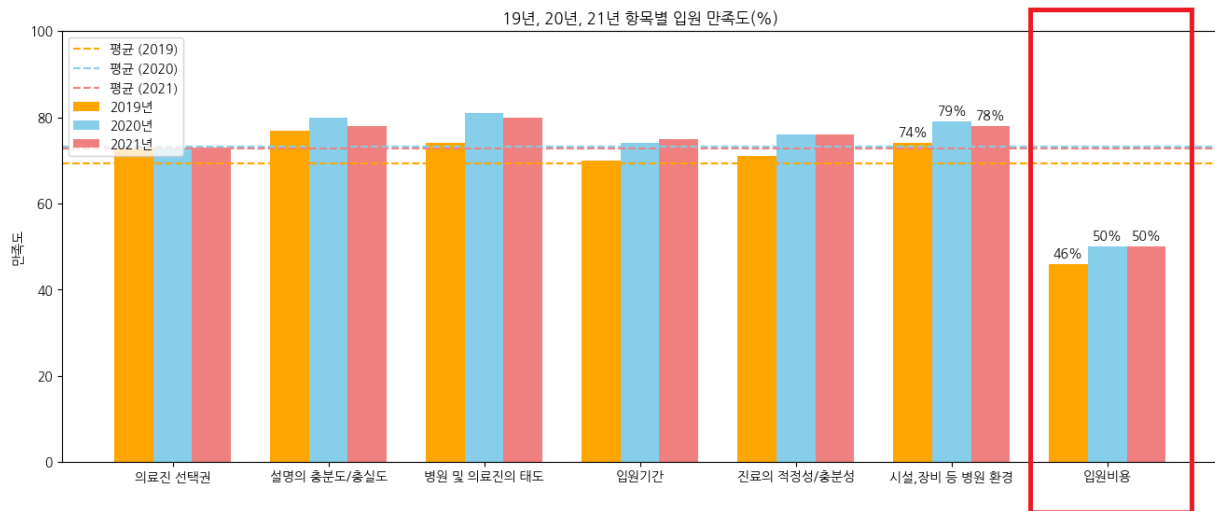
```

```
hh_a_a['inc_r'].value_counts()
```

4. 분석 내용 및 결과

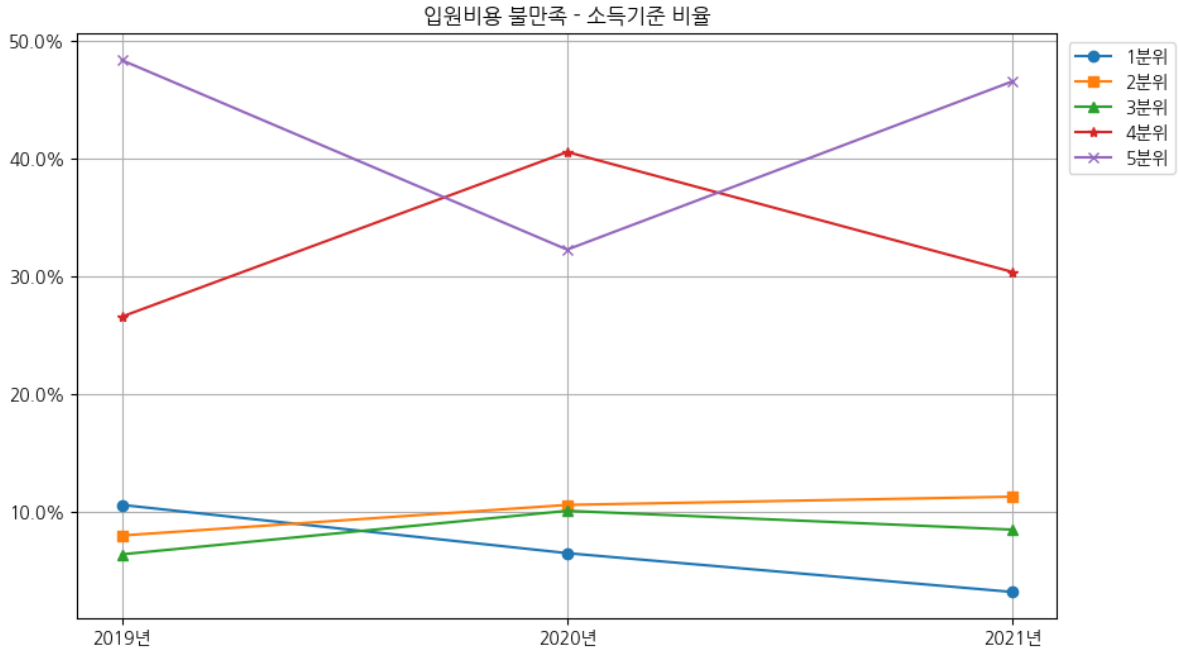
1) 의료 서비스 이용 만족도 분석

1. 유형별 입원 의료서비스 만족도

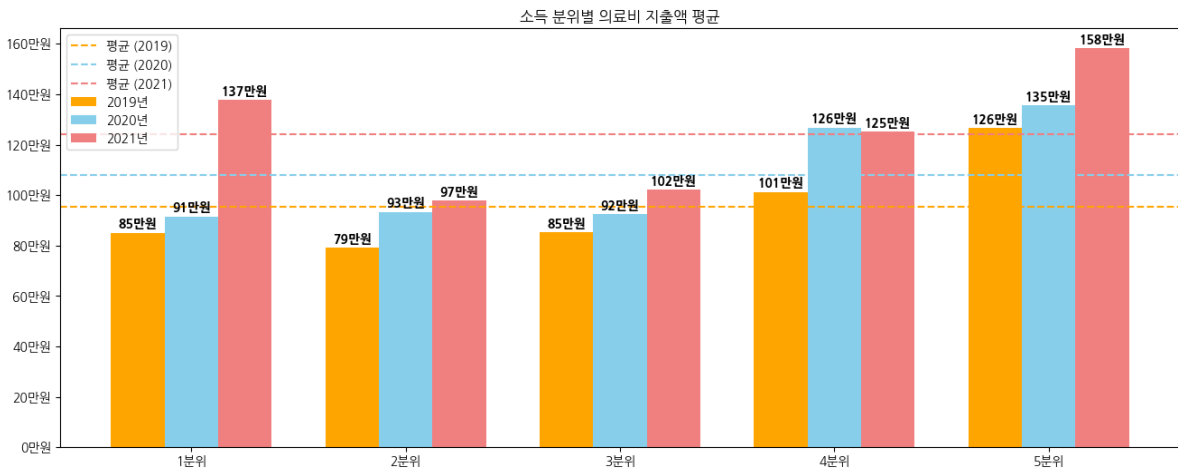


- 입원 관련 서비스가 평균 70% 만족도를 보이는 반면, 입원비용에 대한 만족도는 19년 46%, 20년과 21년 50%로 현저히 낮은 만족도를 보임

2. 입원비용 불만자 유형에 따른 분석



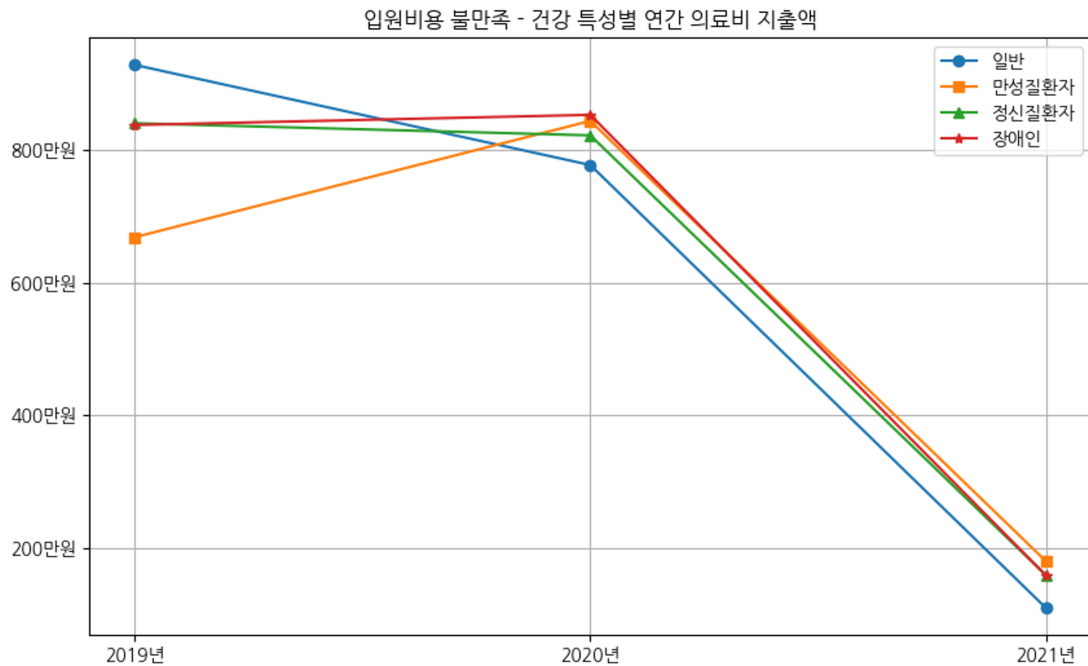
- 입원비에 대한 불만은 4, 5분위가 각 48%, 26%를 차지해 높은 소득분위의 불만족도가 높음을 확인



- 4,5분위 가구의 의료비 지출이 1,2,3분위 가구보다 높음

▼ 소득분위 산정방식

연도	1인가구	2인가구	3인가구	4인가구	5인가구	6인가구	7인가구	8인 이상 가구
2015	1,562,337	2,660,196	3,441,364	4,222,533	5,003,702	5,784,870	6,566,039	1인 증가시마다 781,169원씩 증가 (8인가구: 7,347,208원)
2016	1,624,831	2,766,603	3,579,019	4,391,434	5,203,849	6,016,265	6,828,680	1인 증가시마다 812,465원씩 증가 (8인가구: 7,641,005원)
2017	1,652,931	2,814,449	3,640,915	4,467,380	5,293,845	6,120,311	6,946,776	1인 증가시마다 826,465원씩 증가 (8인가구: 7,773,241원)
2018	1,672,105	2,847,097	3,683,150	4,519,202	5,355,254	6,191,307	7,027,359	1인 증가시마다 836,052원씩 증가 (8인가구: 7,863,411원)
2019	1,707,008	2,906,528	3,760,032	4,613,536	5,467,040	6,320,544	7,174,048	1인 증가시마다 853,504원씩 증가 (8인가구: 8,027,552원)
2020	1,757,194	2,991,980	3,870,577	4,749,174	5,627,771	6,506,368	7,389,715	1인 증가시마다 883,347원씩 증가 (8인가구: 8,273,062원)
2021	1,827,831	3,088,079	3,983,950	4,876,292	5,757,373	6,628,603	7,497,198	1인 증가시마다 868,595원씩 증가 (8인가구: 8,365,793원)



- 각 소득 분위별 의료비 평균 지출액

- 19년, 20년에 비해 21년의 비율 급감 → 코로나19로 인한 의료시설 이용 제한의 영향으로 보임

3. 재난적 의료비 경험률 추이



재난적 의료비란?

- 가구 소득 수준 대비 의료비 부담이 일정수준을 넘어서는 경우
- 소득수준에 비해 과도한 의료비를 부담하는 가구에 비급여를 포함한 의료비의 일부를 지원하여 가계파탄을 방지하기위한 목적으로 재난적의료비 지원 사업 시행
 - 소득기준 : 기준중위소득 100%이하 지원

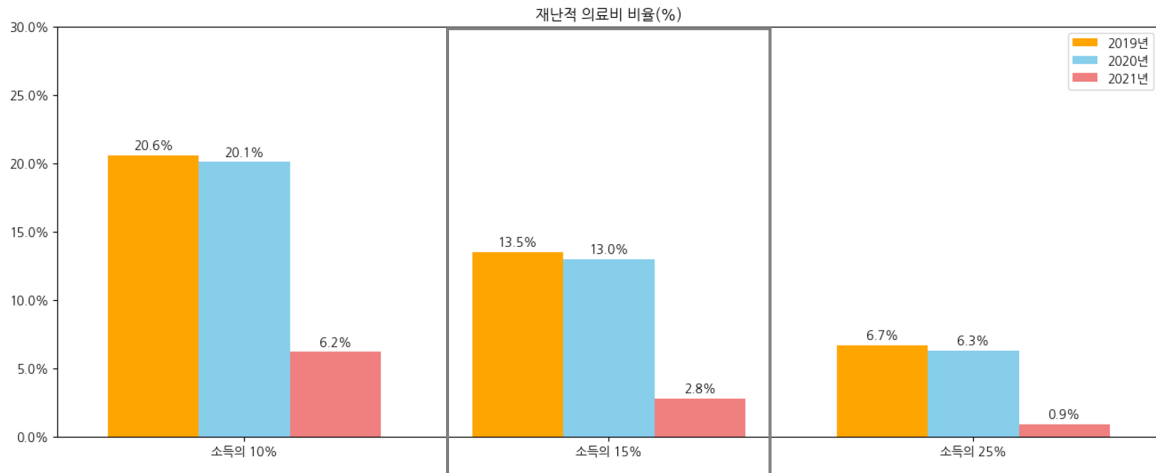
	2019년	2020년	2021년
소득기준	기준 중위소득 100%이하	기준 중위소득 100%이하	기준 중위소득 100%이하
재난적의료비	15%초과	15%초과	15%초과
질환기준	입원진료, 중증질환	입원진료, 중증질환	입원진료, 중증질환
지원비율		본인부담의료비 50%	소득분위별 차등 지급
지원한도	2천만원	2천만원	3천만원



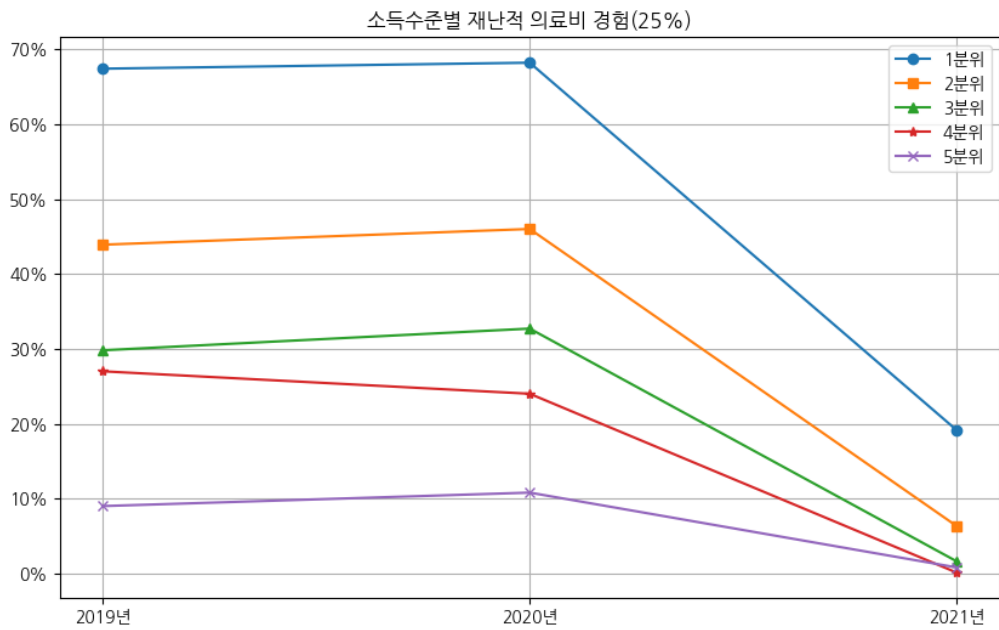
재난적 의료비 발생 정의

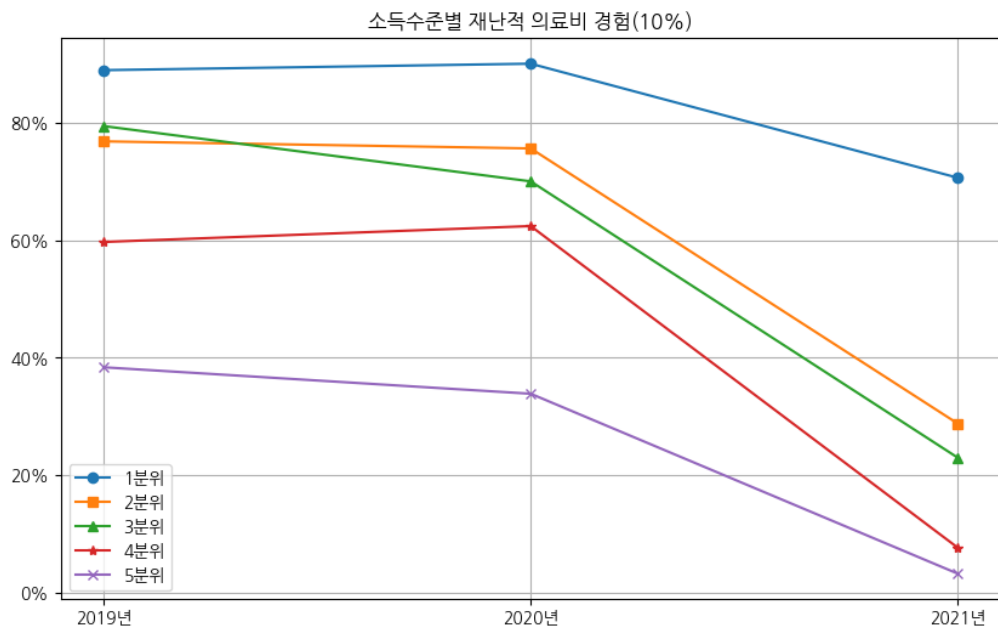
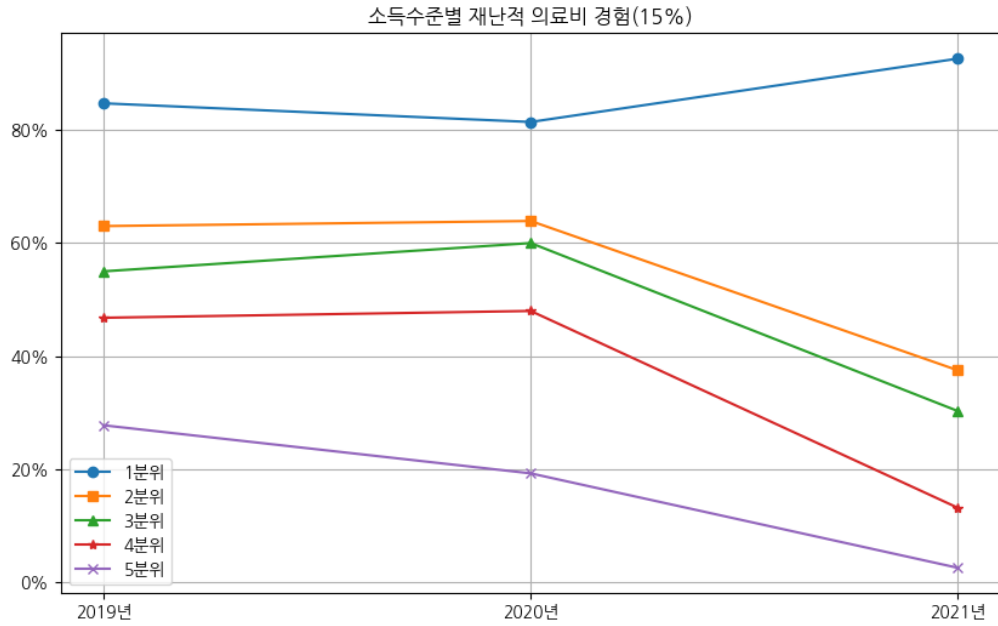
- 지불능력 : 가구 총 소득
- 의료비 : 응급,입원,외래 이용시 의료비 지출 비용(수납금액 MEXP3_1 + 처방약값 MEXP4_1)
- 재난적 의료비 비율 선정 : 10% 초과 , 15초과, 25% 초과

비율 기준점에 대해서는 합의된 바가 없고, 연구자별로 5~40%까지 다양한 기준을 다소 자의적으로 적용하고 있다. 이에 재난적 의료비 지표결과에 대한 편향된 해석의 위험을 줄이고 비교가능성을 높이기 위해 여러 수준의 기준점을 동시에 제시하기위해 조사 해당 년도 적용되었던 15%를 기준으로 10%, 25%를 추가 기준점으로 설정

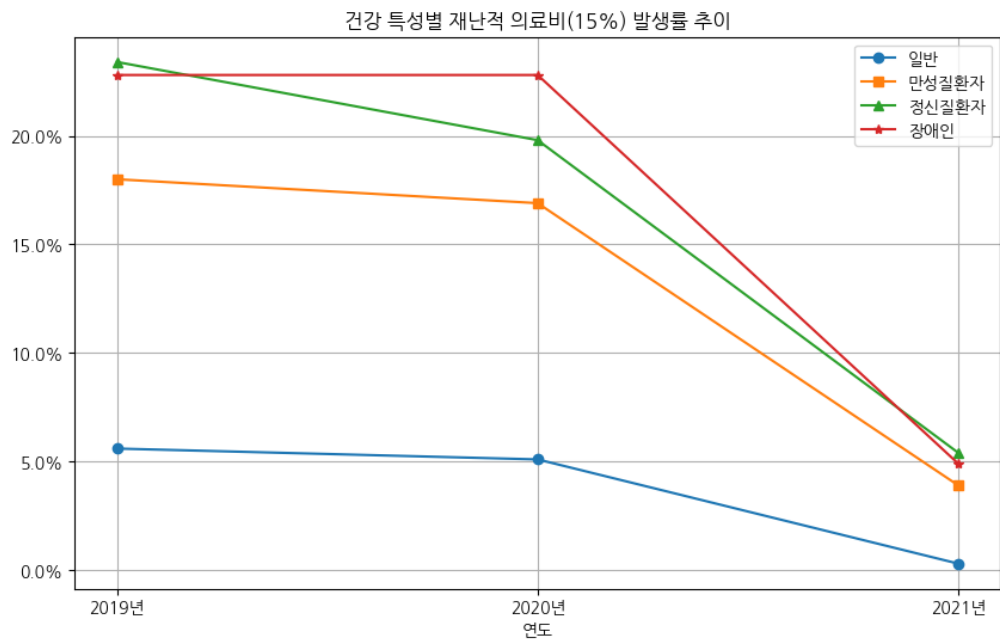
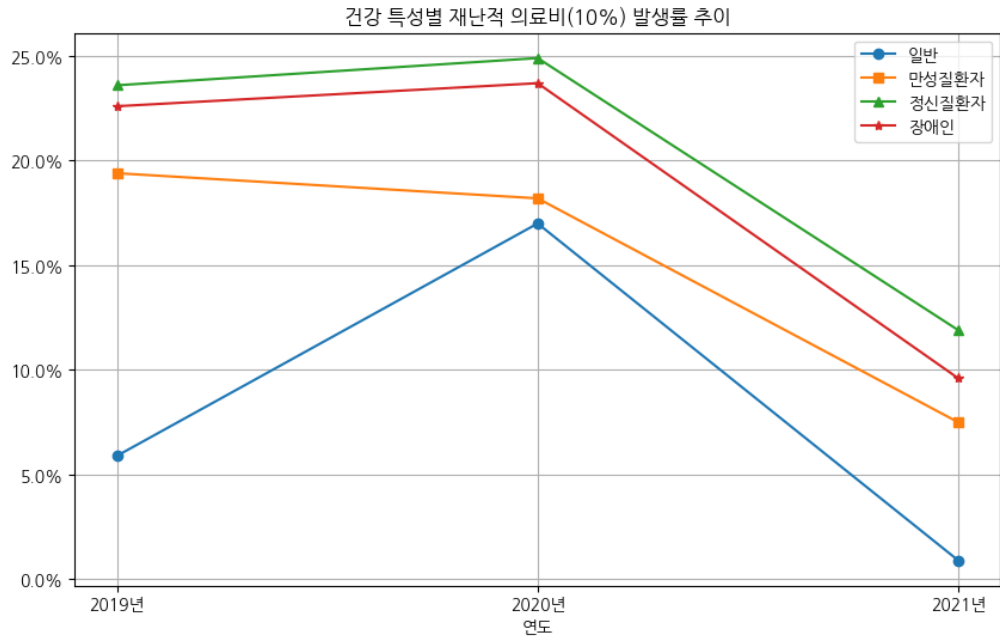


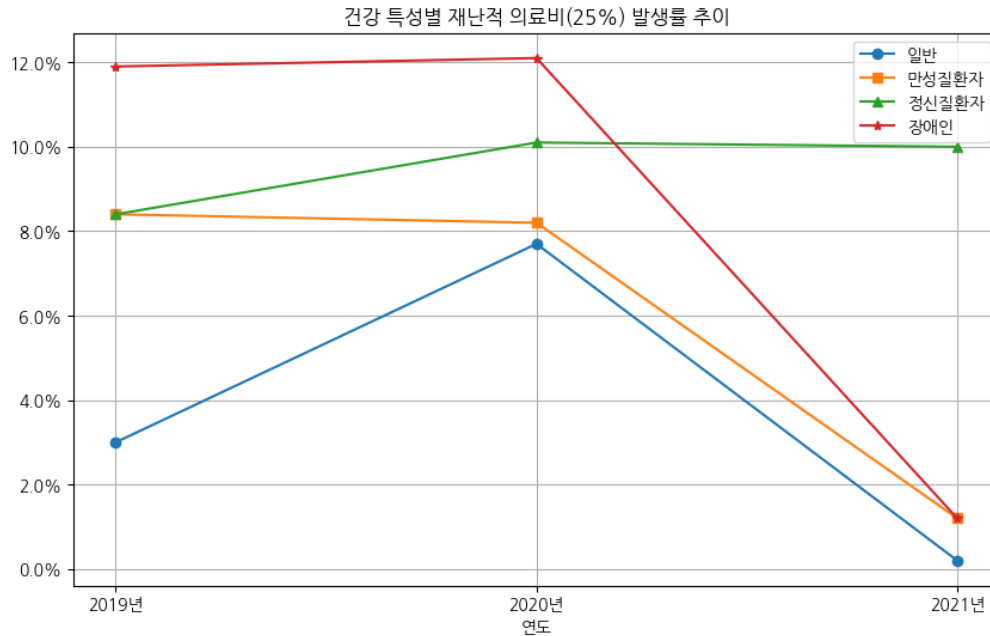
- 입원) 전 가구의 재난적 의료비 비율을 확인
 - 19년, 20년에 비해 21년의 비율이 급감
 - 현행 기준인 15%로 지원혜택을 받는 가구는 13%내외로 나타났고 지원 기준을 10%로 완화한 경우 혜택 대상 가구가 15%대비 최소 52.6% 증가함





- 소득수준별로 재난적 의료비 경험률
 - 재난적 의료비 발생률을 1분위 가구에서 가장 높게 나타남
 - 15%기준 1분위 가구를 제외한 모든 분위 2021년에 전반적인 하락률을 보임





- 건강 특성별(일반/만성질환/정신질환/장애) 재난적 의료비 발생
 - 정신질환, 장애인 > 만성질환 > 일반 순으로 재난적 의료비 발생률이 높은 것으로 확인
 - 재난적의료비 25% 발생률에서 2021년 다른 특성의 비율은 급감하는 반면, 정신질환자의 재난적의료비 발생률은 유지됨
 - 정신질환자 유형은 소득 대비 의료비 지출이 높은 경향을 나타냄

인사이트 및 결과 정리

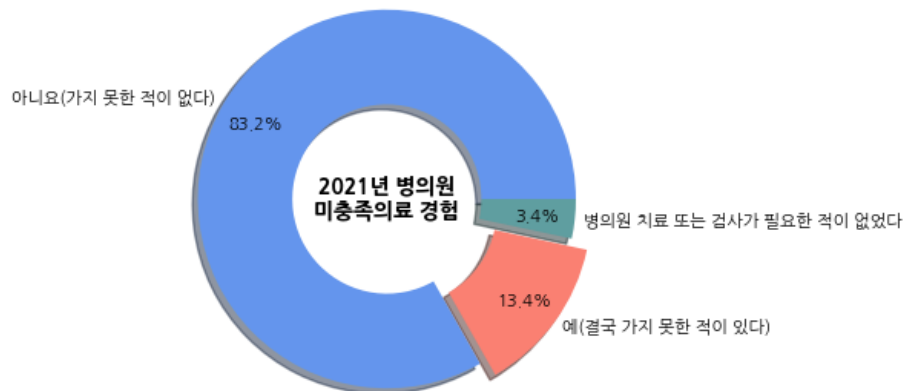
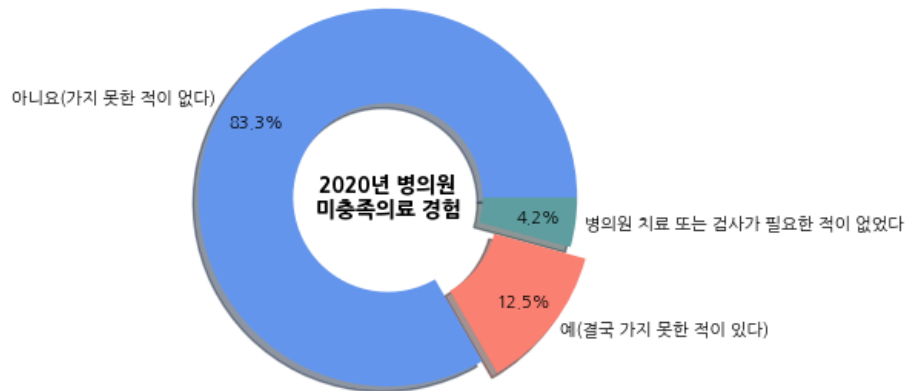


환자의 특성별 의료 지원 제도에 대한 개선점 제안

1. 지불 능력이 낮은 계층의 경제적 부담 경감을 위한 현제도 완화
 - 재난적 의료비 지원한도 상향을 통한 의료비 부담 완화
2. 현재 시행 중인 지원사업의 질환기준 확대
 - 정신질환 지원 범위 확대를 위한 정신질환자의 정보 수집

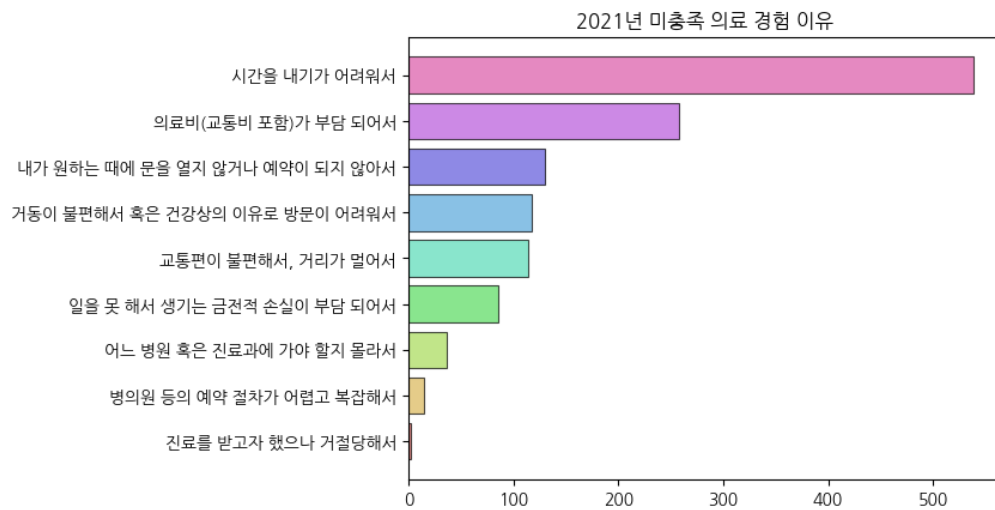
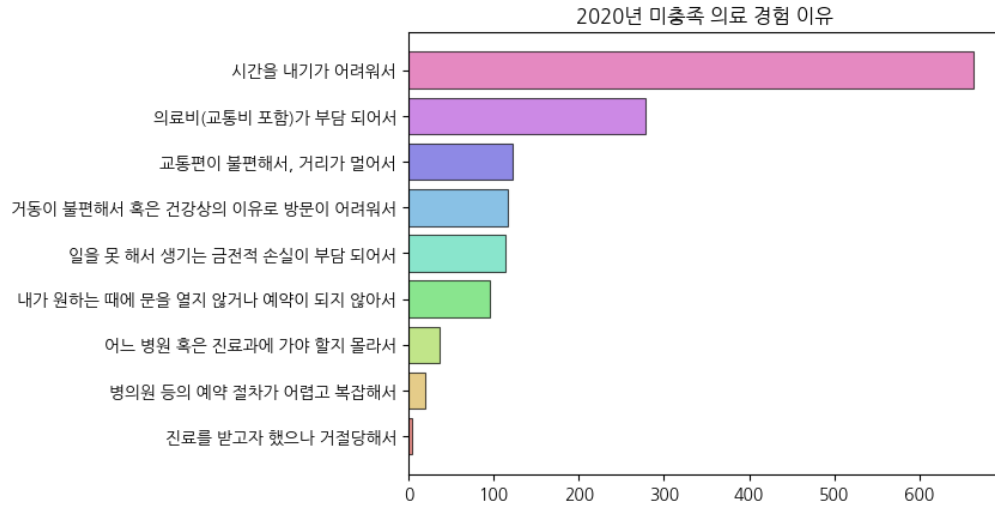
2) 지역별/연령별 병원 미충족의료 경험 및 원인 파악

1. 미충족 의료 경험률 현황(단위 : %)



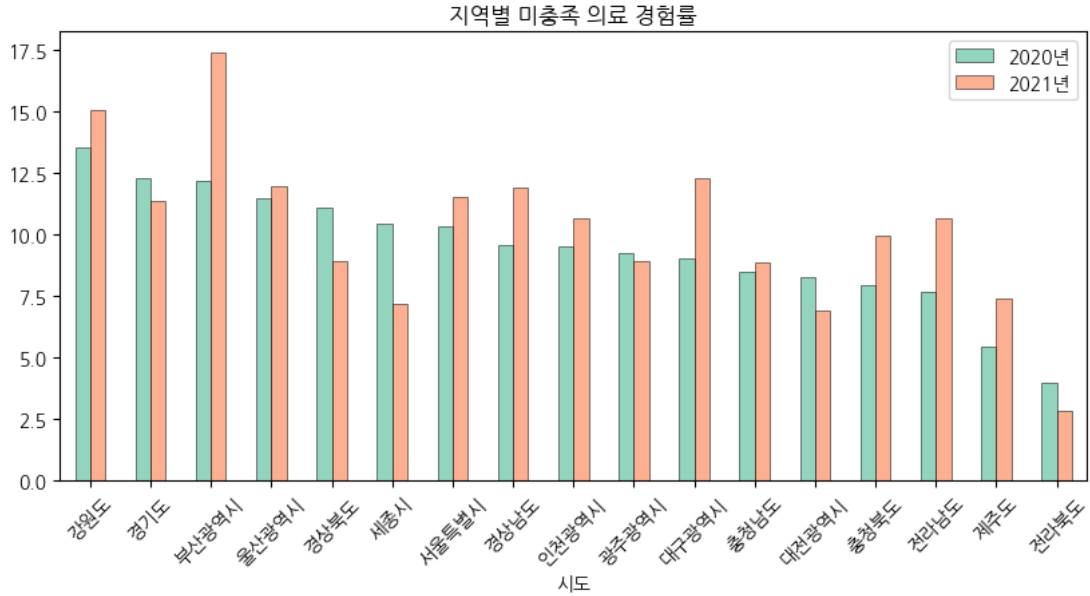
- 미충족 의료를 경험한 사람의 비율은 2020년 12.5% 에서 2021년 13.4% 로 조금 증가함
- 전 국민 건강보장제도와 의료서비스 제공과 관련된 자원 규모 또한 지속적으로 확대되어 왔음
- 이러한 발전이 의료 접근성 개선에 기여하였지만, 여전히 소득계층 및 연령층 간의 불균형 해결이 필요함

2. 미충족 의료 경험 이유



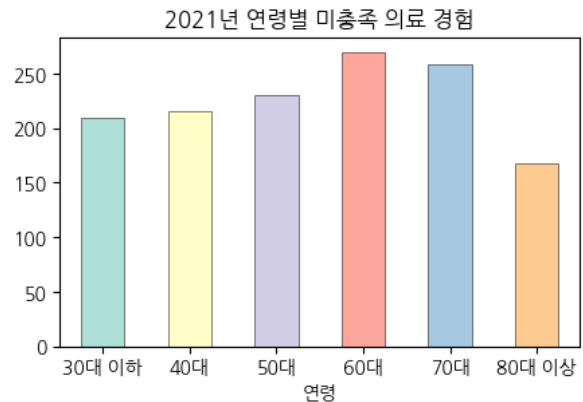
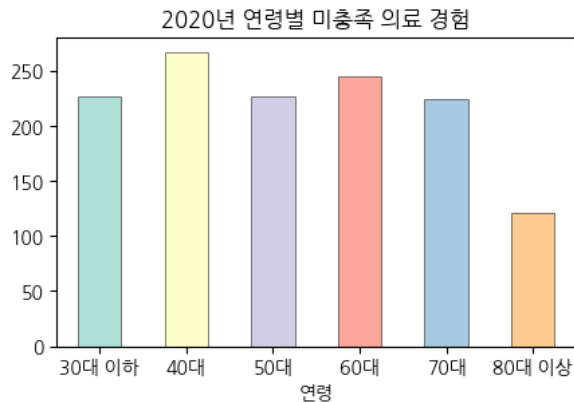
- 시간을 내기가 어려워서 절반 이상 차지, 이어서 의료비가 부담되어서 가장 많음
- 의료 서비스 접근성을 저해하는 가장 주요한 요인은 시간 부족과 경제적 부담인 것으로 보임

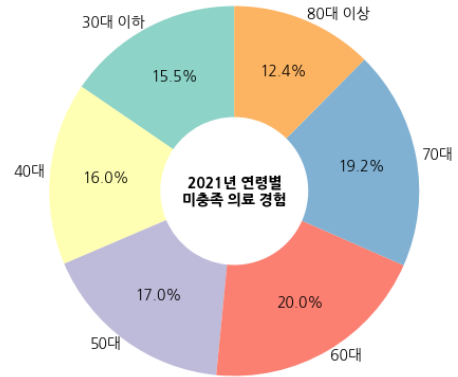
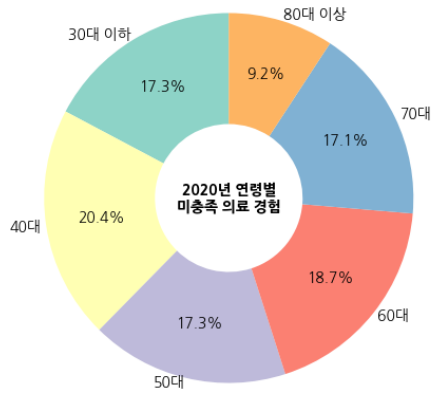
3. 지역별 미충족 의료 경험률



- 지역별로 미충족률의 규모가 상이함
 - 미충족 의료에 미치는 영향이 개인적인 측면뿐만 아니라 개인이 속해있는 지역의 특성에 따라서도 다르게 나타날 수 있음을 보임
 - 특히 현재 한국에서는 의료자원이 수도권을 중심으로 분포하고 수요와 공급의 불균형을 나타내고 있음
- 부산, 대구, 전남 → 급격하게 증가
 - 코로나의 영향이 큰 것으로 보임

3. 연령별 미충족 의료 경험률

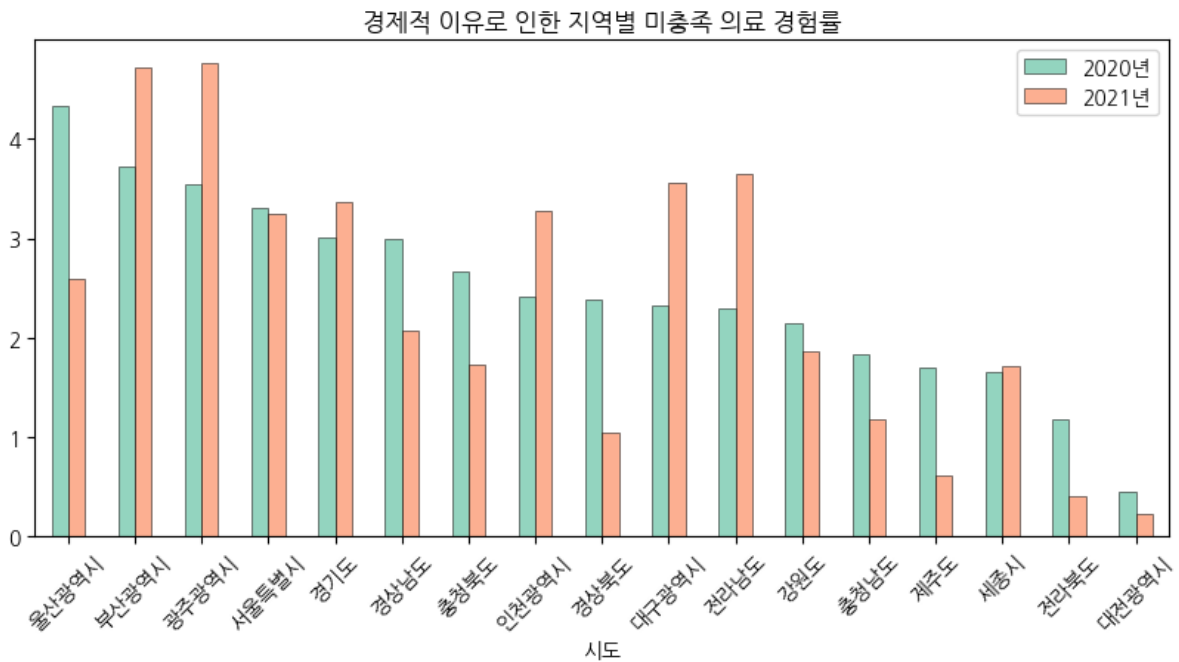




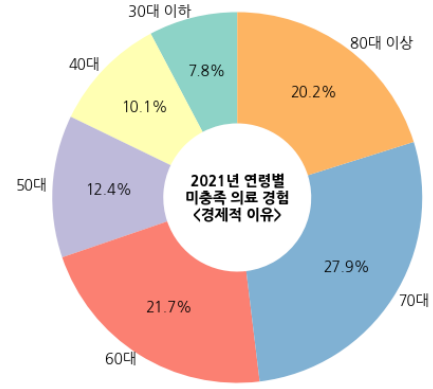
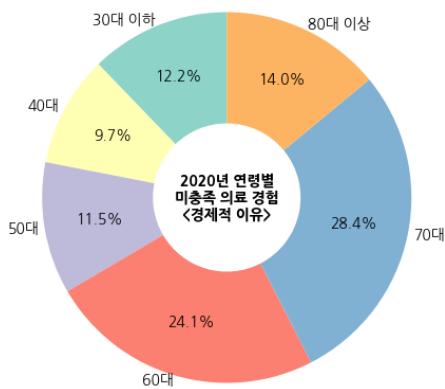
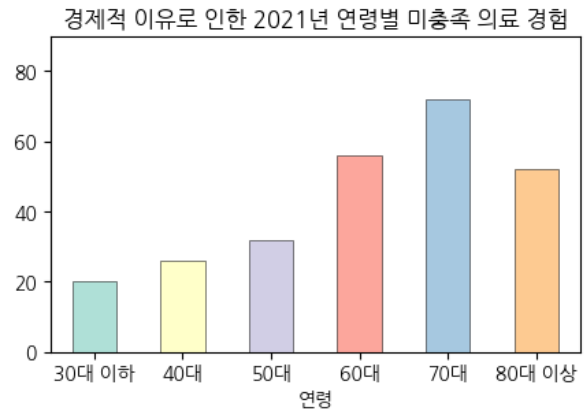
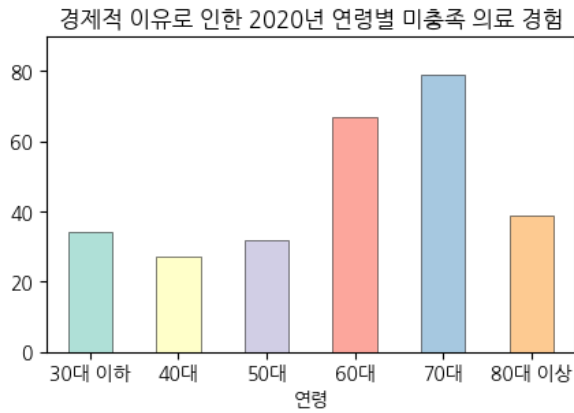
- 연령별로 골고루 분포되어 있는 것을 볼 수 있음
- 2021년에는 2020년에 비해 60대 이상 고령층의 비율이 많아짐

5. 경제적 이유로 인한 미충족 의료 경험률

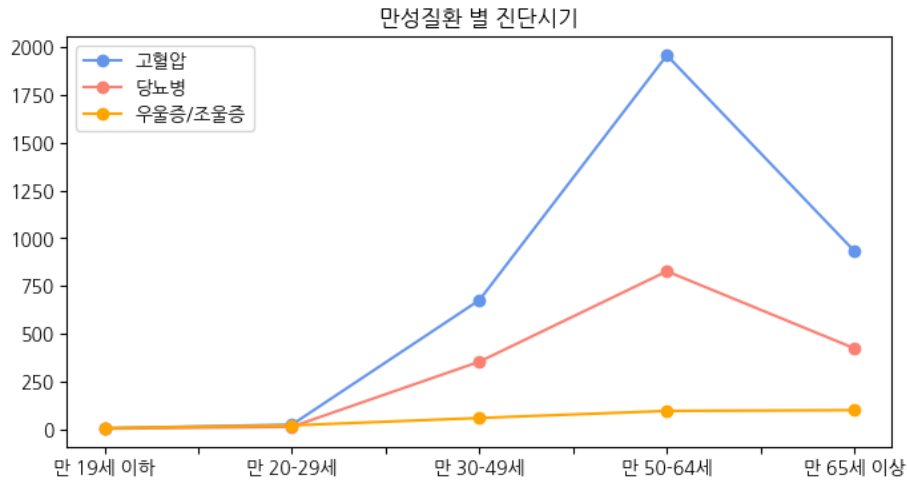
→ **의료비가 부담되어서** 치료나 검사를 못한 적이 한 번이라도 있었다고 응답한 사람의 비율



- 부산, 광주, 대구, 전남 → 급격하게 증가
- **광주**는 전체 미충족 의료율은 감소하였지만, 경제적 이유로 인한 미충족 의료율은 증가하였으므로 경제적 취약 지역이라고 볼 수 있음



- 경제적 이유로 한정 지어보니 70대 이상의 연령층에서의 미충족 의료 경험률과 젊은 연령층들과의 미충족 의료 경험률의 편차가 확연히 나타남
- 령층에서 의료비에 대해 경제적 부담을 더 많이 겪고 있는 것으로 보임
- OECD에 의하면 우리나라의 고령 인구는 2050 년에 45% 로 세계의 가장 높은 수준이 될 것으로 예측됨
→ 인구 구조의 빠른 고령화로 인한 미충족 의료 경험률 증가가 예상되어 이를 기반으로 한 대안 및 탐색이 필요함



인사이트 및 결과 정리



지역별/연령별 의료 서비스 격차에 대한 개선점 제안

1. 디지털 헬스케어 산업 고도화

- 원격 의료 서비스로 인한 의료 접근성 향상
- 진료 프로세스와 효율성 향상

2. 고령 연령층을 위한 특별한 관리 방안 도입

- 가격 정보와 보험 혜택에 대한 명확한 정보를 제공하고 환자들이 비용을 미리 이해하도록 도움
- 만성 질환(고혈압, 당뇨병)과 같은 고령 연령층에서 주로 발생하는 의료 문제에 대한 관심을 기울임

3) 당뇨병 유병에 영향을 미치는 원인 파악 및 진단 예측

2019년도 가구원 데이터(IND) 를 이용하여 당뇨병 예측 모형 개발

1. 데이터 전처리

- 예측을 위한 독립변수와 종속변수 선택
 - 독립변수 : 성별, 출생년도, 교육수준, 경제활동 참여 상태, 의료보장 형태,

고혈압 유무, 규칙적 운동 실천 여부, 체중 조절 시도 여부, 평생 흡연 여부,

음주 빈도, 스트레스 인지 정도, 주관적 건강상태

◦ 종속변수 : 당뇨병 유무

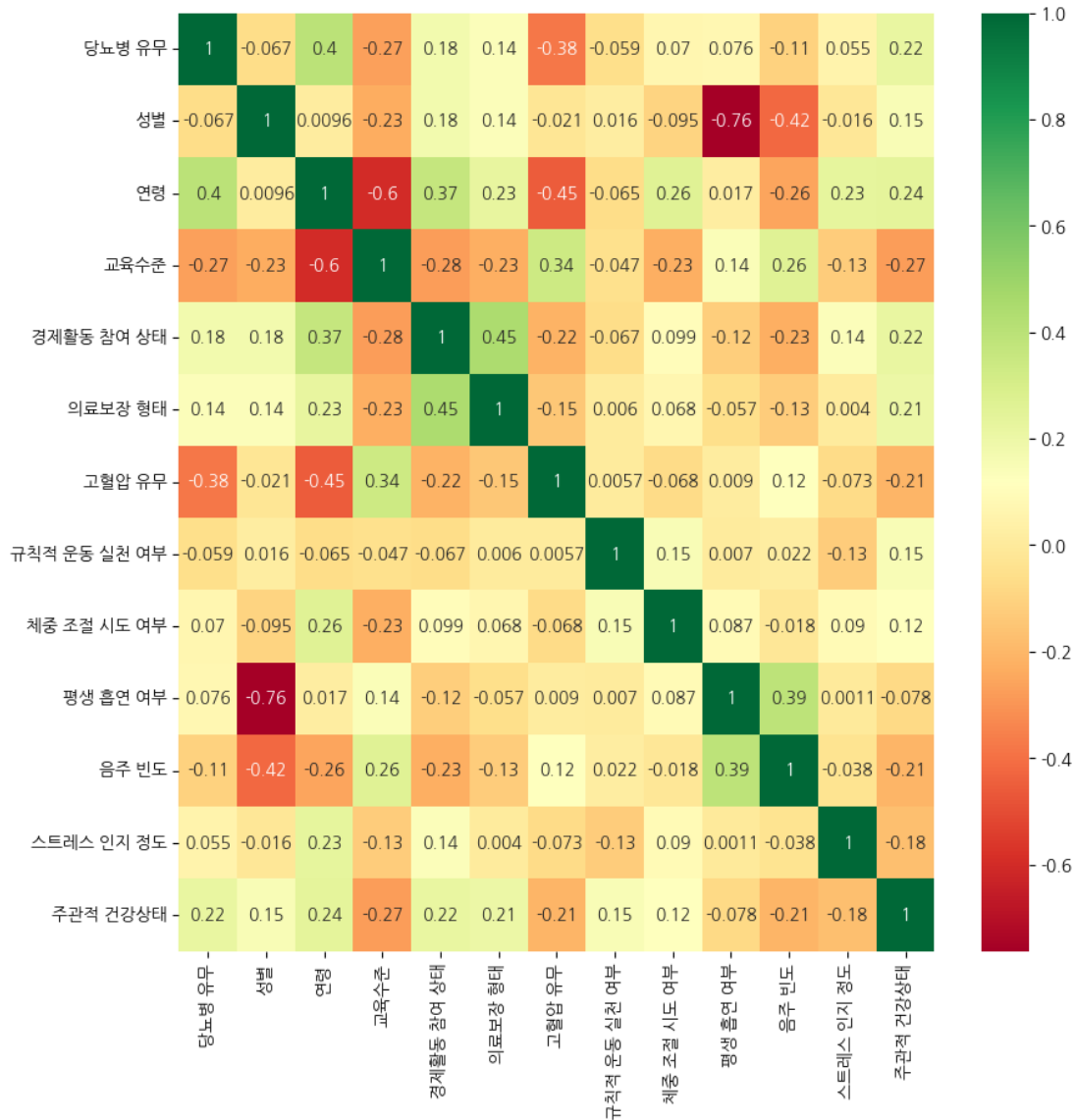
- 파생변수 '연령' 추가
- '음주 빈도' 변수에서 결측값 처리
- '평생 흡연 여부' 변수에서 값 바꾸기
- object 형 변수들 int형으로 전환
- '의료 보장 형태' 변수에서 결측값 제거
- '규칙적 운동 여부' 변수에서 결측값 제거
- '의료 보장 형태' 변수에서 값이 7,8인 행 제거
- '음주 빈도' 변수에서 값이 0인 행 제거
- 당뇨병 예측 분석을 위한 데이터 추출
 - '당뇨병 유무' 변수에 값인 1인 행 추출, 값인 2인 행을 값이 1인 행 개수만큼 추출
 - '당뇨병 유무' 변수에 2인 값을 0으로 바꾸기

2. 회귀 분석

추출 데이터에서의 당뇨병 유무 변수 값

당뇨병 유무	사람 수
0 (당뇨병 X)	1567
1 (당뇨병 O)	1567

당뇨병 유무(종속 변수) 와 다른 변수(독립 변수)들의 상관관계



- 연령이 0.4, 고혈압 유무가 -0.38 로 당뇨병 유무와 상대적으로 높은 상관관계를 가지고 있습니다.

train 데이터와 test 데이터를 8:2로 나누어 모델링 진행

- 기본 모델

Model	Accuracy
Support Vector Machine	0.6778

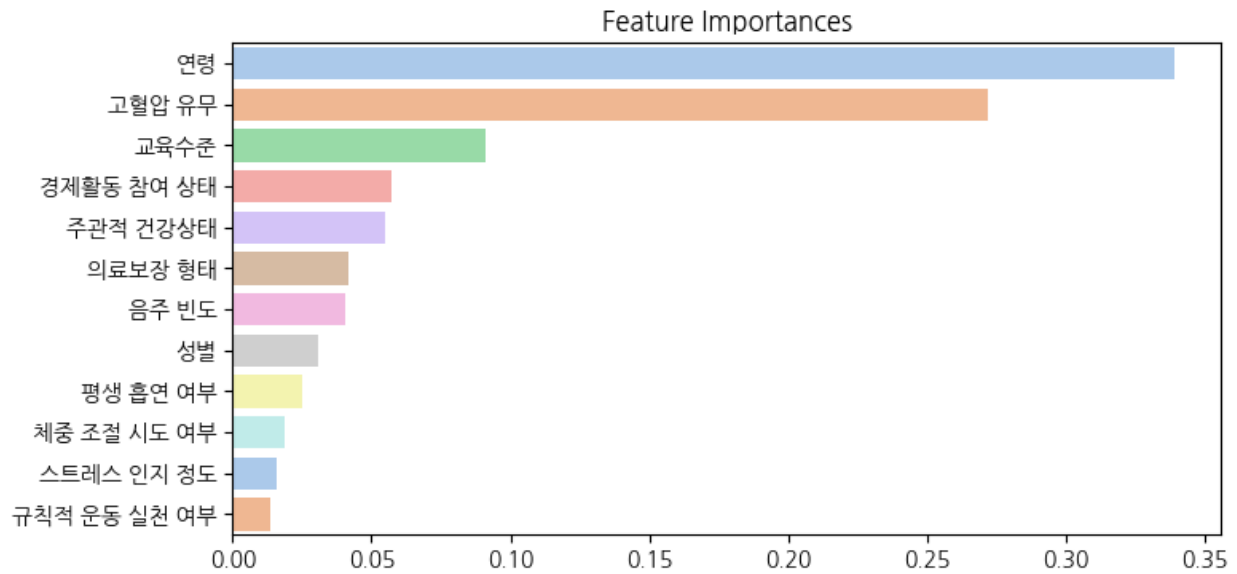
- 하이퍼 파라미터 튜닝(그리드 서치)

Model	Accuracy
Support Vector Machine	0.6905

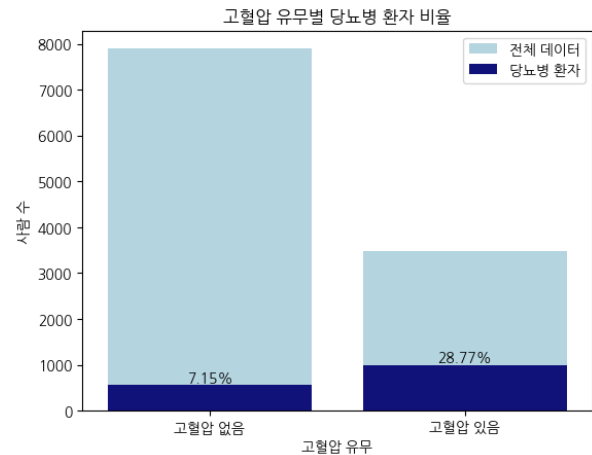
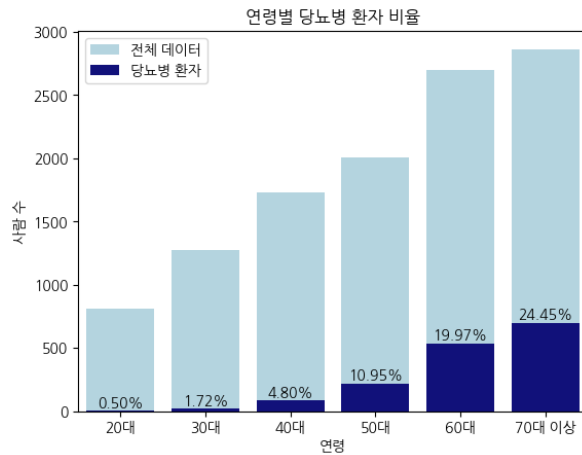
Random Forest	0.7002
XGBOOST	0.6986

Random Forest	0.7240
XGBOOST	0.6769

변수 중요도 그래프 (랜덤 포레스트 모델)



- 연령, 고혈압 유무, 교육 수준 등의 변수 중요도가 상대적으로 높게 나옵니다.



- 연령대가 올라갈 수록 당뇨병 환자 비율이 올라갑니다.
- 고혈압이 있는 사람들이 당뇨병을 갖고 있는 비율이 약 29%로 고혈압이 없는 사람보다 상대적으로 높다.

인사이트 및 결과 정리



당뇨병 유병 예측을 통한 병의 원인 파악 및 관리

1. 연령이 올라갈수록 당뇨병의 유병 확률이 증가
 - 고령화, 노인인구의 증가 등을 고려하여 향후 당뇨병의 발생이 더 증가할 것으로 예측됨
 - 연령 고위험군에 대한 예방 및 적절한 관리가 필요
2. 고혈압이 있는 사람들이 없는 사람들에 비해 당뇨병을 갖고 있는 비율 높음
 - 생활습관 교정(식습관 개선, 운동, 금연) 등으로 주기적인 혈압 관리 필요

5. 프로젝트 회고 및 개선점

5-1. 현업자 피드백 + 각종 질문

5-2. 회고(개인의 문제인식~극복과정)

이름	회고 내용
김예린	2019~2021년 데이터로만 분석을 진행하여 최신 데이터를 사용하지 못하였고, 분석 결과가 현재 상황을 정확하게 반영하지 못한 점이 아쉬움. 또한 19년도에서 21년도가 코로나 19로 인하여 의료 환경이 엄청난 과도기를 겪은 시기인 것을 파악했을 때, 일시적인 현상과 장기적인 추세를 구분하기가 어려웠음.
김수민	의료 데이터 수집이 쉽지않아 데이터 확보에 어려움을 겪었음. 분석 방향의 변경으로 인해 시간관리에서 아쉬움이 남았음
민윤홍	패널 데이터라 당뇨병 예측 관련 변수에 한계점이 있었음. 세 가지의 분석 주제가 아닌 한 가지의 주제를 다 같이 깊게 분석했으면 더 좋지 않았을까 하는 아쉬움이 있음.
황상엽	정해진 주제와 맞는 분석을 진행하기까지 많은 시간을 소비하였음. 코로나에 대한 데이터가 포함되어 있지 않아 분석이나 결과 해석에 어려움이 있었음

5-3. 개선점 (팀 내에서 논의 및 합의된 개선 방향)

5-4. 추후 개선 계획

6. 부록

6-1. 참고자료 및 출처


<http://media.guidebook.com.s3.amazonaws.com/upload/TKcoBa16TXxQwJud7zuzotmfiOkLf7comSNoaQGJ/ZOGNChDvjCvAYW6elkVPwuTBPHR5WoTuxbrP.pdf>

SVM과 meta-learning algorithm을 이용한 고지혈증 유병 예측모형 개발과 활용

<https://ir.ymlib.yonsei.ac.kr/bitstream/22282913/176048/1/T202001221.pdf>

2018 미충족의료율과 추이

지역의 의료자원 분포와 미충족 의료의 현황 및 영향 관계

 <https://s-space.snu.ac.kr/handle/10371/128416>

국민 80% 이상 디지털 헬스케어 건강상태 개선 도움

국민의 80% 이상이 디지털 헬스케어가 개인의 건강상태 개선이 도움이 되는 것으로 평가했다. 반면 원격의료에 대해선 기대와 우려가 함께 존재하는 것으로 나타났다.이같은 결과는 KDI(한국개발연구원) 경제정보센

<https://www.khanews.com/news/articleView.html?idxno=212307>



<https://repository.kihasa.re.kr/bitstream/201002/36665/1/2020.11.No.289.04.pdf>

건강보험 보장성 강화에 따른 가구 의료비 부담 변화와 시사점

<https://www.129.go.kr/common/downFile.jsp?filename=2019%B3%E2%20BB%F3%B9%DD%B1%E2%20C0%E7%B3%AD%C0%FB%C0%C7%B7%E1%BA%F1%20C1%F6%BF%F8%BB%E7%BE%F7%20BE%C8%B3%B.B.pdf&dataDir=/board/data/>

2019년 재난적의료비지원사업 안내

https://www.nhis.or.kr/static/html/wbda/h/2020_book.pdf

2020년 재난적의료비지원사업 안내

2021년 재난적의료비 지원사업 안내책자 | 국민건강보험

2021년 재난적의료비 지원사업 안내책자

https://www.nhis.or.kr/_custom/nhis/_common/board/index/725.do?mode=view&articleNo=10811634&title=2021년+재난적의료비+지원사업+안내책자



2021년 재난적의료비 지원사업 안내

재난적 의료비 발생의 추이와 합의