

DSA 17기 SIXCAN팀

환자의 재발가능성과 'Care 서비스' 필요 군집 분류 모델 기획을 통한 병원 재방문률 증대

목차 구성

1. 추진배경

2. 현황 및 개선

3. 프로젝트 목표

4. 분석계획

5. 분석결과 및 인사이트

6. 모델링 과정

7. 개선안

8. 프로젝트 소감

추진배경 : 비즈니스 안건

척추 질환의 치료 수요 증가에 따른 수익 증대 기회

- 국내에서 척추와 관련된 허리 디스크 환자가 매년 3만명 이상 증가하고 있음
- 컴퓨터와 스마트폰 사용 급증에 따라 젊은 층에서도 척추 질환 발병이 늘어나고 있음
- 본 병원에서도 척추 관련 수술 환자가 매년 증가하는 추세를 보임

수익에 영향을 미치는 수술 후 환자의 지속적인 방문 및 Care활동이 활발하지 않은 상황

- 수술뿐만 아니라 수술 후 환자의 지속적인 방문 및 Care활동도 수익성에 영향을 미침
- 그러나, 환자들의 수술 후 관리가 적절하게 이뤄지지 않는 상황임
- 따라서 수익 증대를 위한 전략의 일환으로 지속적인 Care서비스를 기획하고자 함

추진배경 : 전략연계



환자의 만족도 증가

기존의 획일적인 사후 관리 방식과 달리, 개별 환자의 연령, 성별, 질환 유형 등을 고려한 맞춤형 Care활동을 제공하면 환자의 치료 효과 극대화 가능함



병원의 신뢰도 증가

병원의 신뢰도는 단순히 치료 성공률뿐만 아니라, 환자에 대한 지속적인 관심과 관리 여부에 의해 결정됨



재방문을 증가로 인한 수익성 증대

병원 재방문율이 증가 시, 외래 진료, 검사, 물리치료 등에서 발생하는 수익이 증가하여, 병원 전체적인 수익성 향상 가능



치료 분야 활용을 통한 인건비 절감

맞춤형 Care서비스 추천 어플을 통해 병원의 사후 관리 시스템을 보다 효율적으로 운영할 수 있으며, 이에 따라 인건비 절감 효과를 기대 가능

추진배경 : 시장 & 고객니즈

시장 경쟁 상황

사후관리 분야 시장 활성화

사후관리 서비스 수요가 증가하여 관련 시장이 확대

care 활동 중요성에 대한 관심 증가

환자와 의료진 모두 사후 관리의 필요성을 인식하게 되어 치료의 연속성과 효과성을 높이는 문화가 형성

고객니즈

수술 후 재발 방지

환자별 최적화된 Care활동을 제공하여 재발률을 낮추고 치료 효과를 장기적으로 유지

병원 방문 번거로움 및 케어 활동 지속성 문제

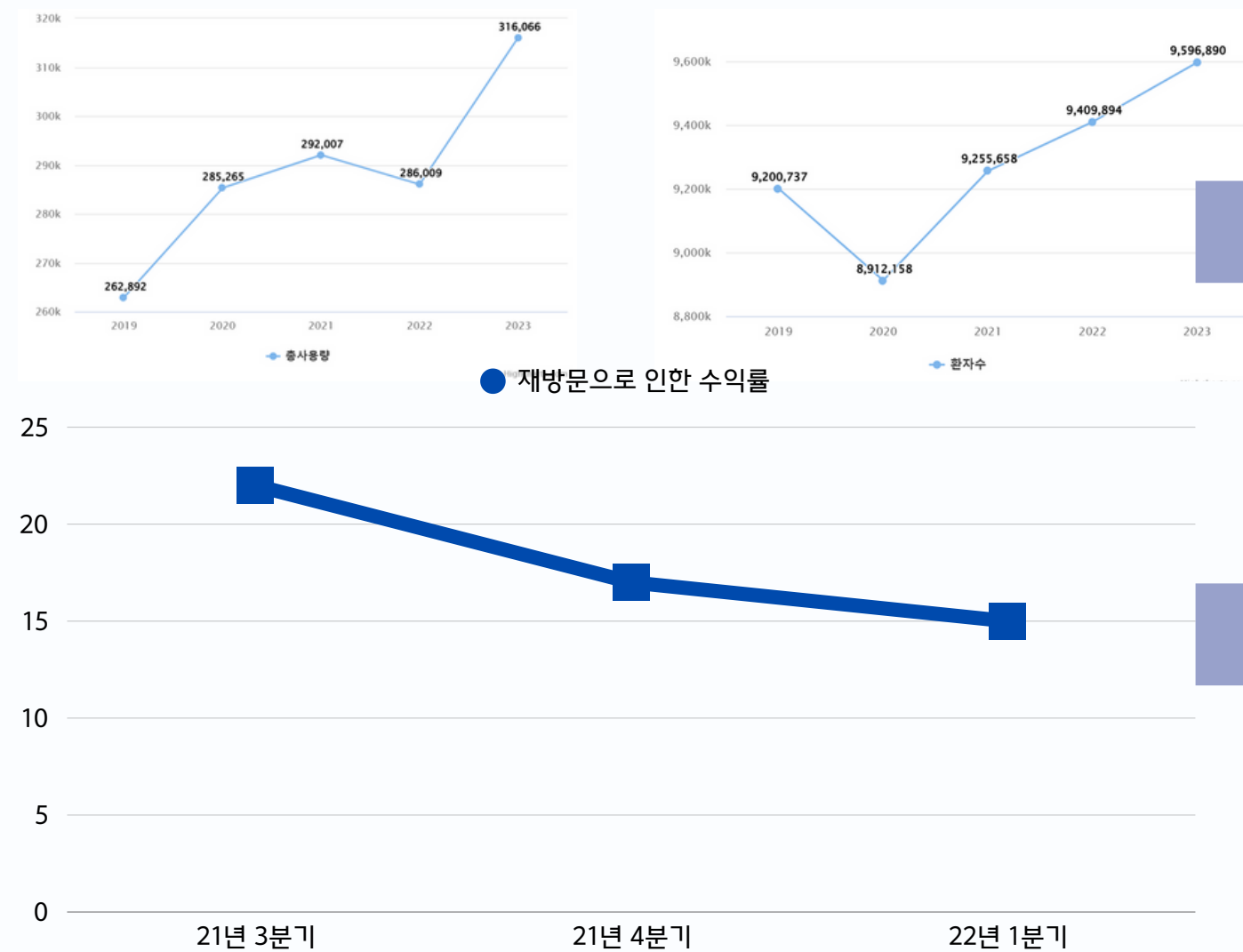
환자가 자주 병원을 방문하지 않아도 지속적인 Care활동을 받을 수 있는 체계를 마련하여 편의성 향상

맞춤형 케어 서비스 및 인화된 치료 솔루션 필요

환자 개개인의 특성을 반영한 맞춤형 치료 및 관리 서비스를 통해 만족도를 높이고 의료 서비스의 질을 향상

현황 및 개선

병원	문제	환자
수익 하락 으로 병원 운영 어려움	척추질환 환자 수는 매년 증가 함에도 사후 케어를 위한 병원 재방문률은 꾸준히 감소	병원 방문 번거로움
차후 팬데믹 대비	지속적인 적절한 케어를 받지 못함으로 질환의 재발률은 13.4%에 달함	비대면 선호도 증가
	코로나 이후 재방문율이 감소 함으로 병원의 수익률도 덩달 아 하락	



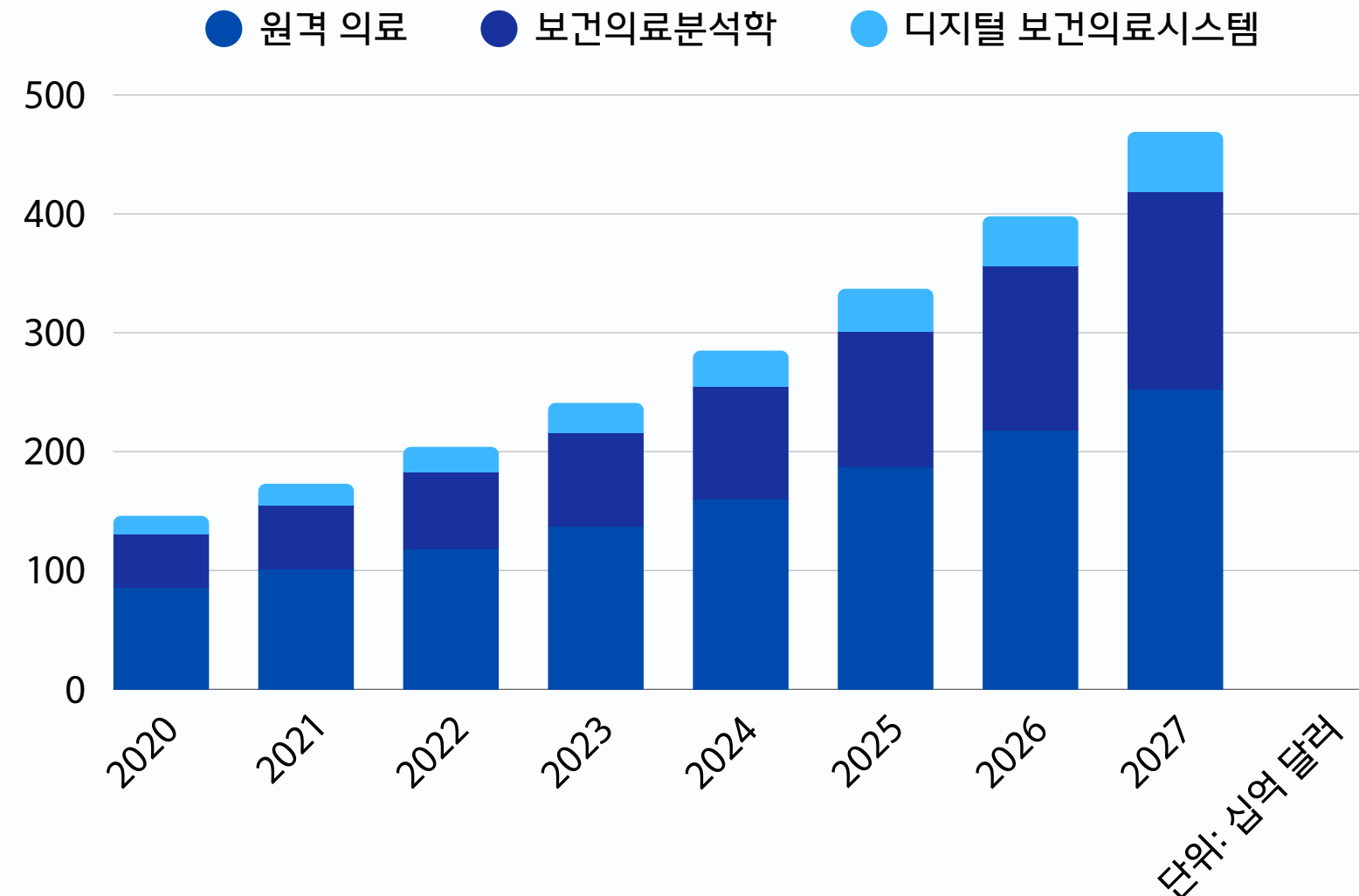
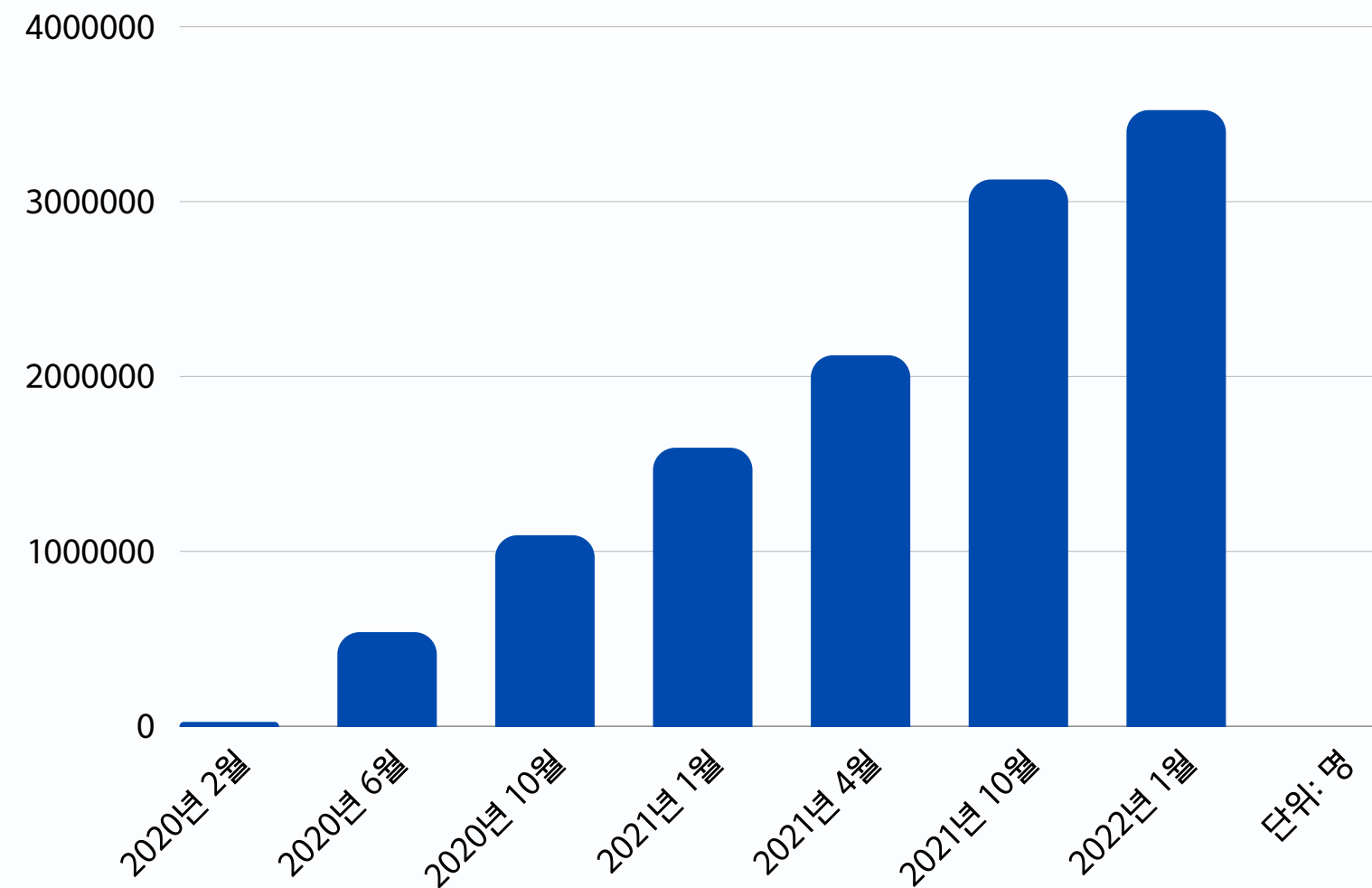
척추질환으로 인한 진료 수와
환자의 수가 꾸준히 증가함

병원 또한 치료를 받는 환자
수는 꾸준히 늘고 있지만 재
활을 위한 재방문률의 감소로
인한 수익이 꾸준히 감소함

현황 및 개선

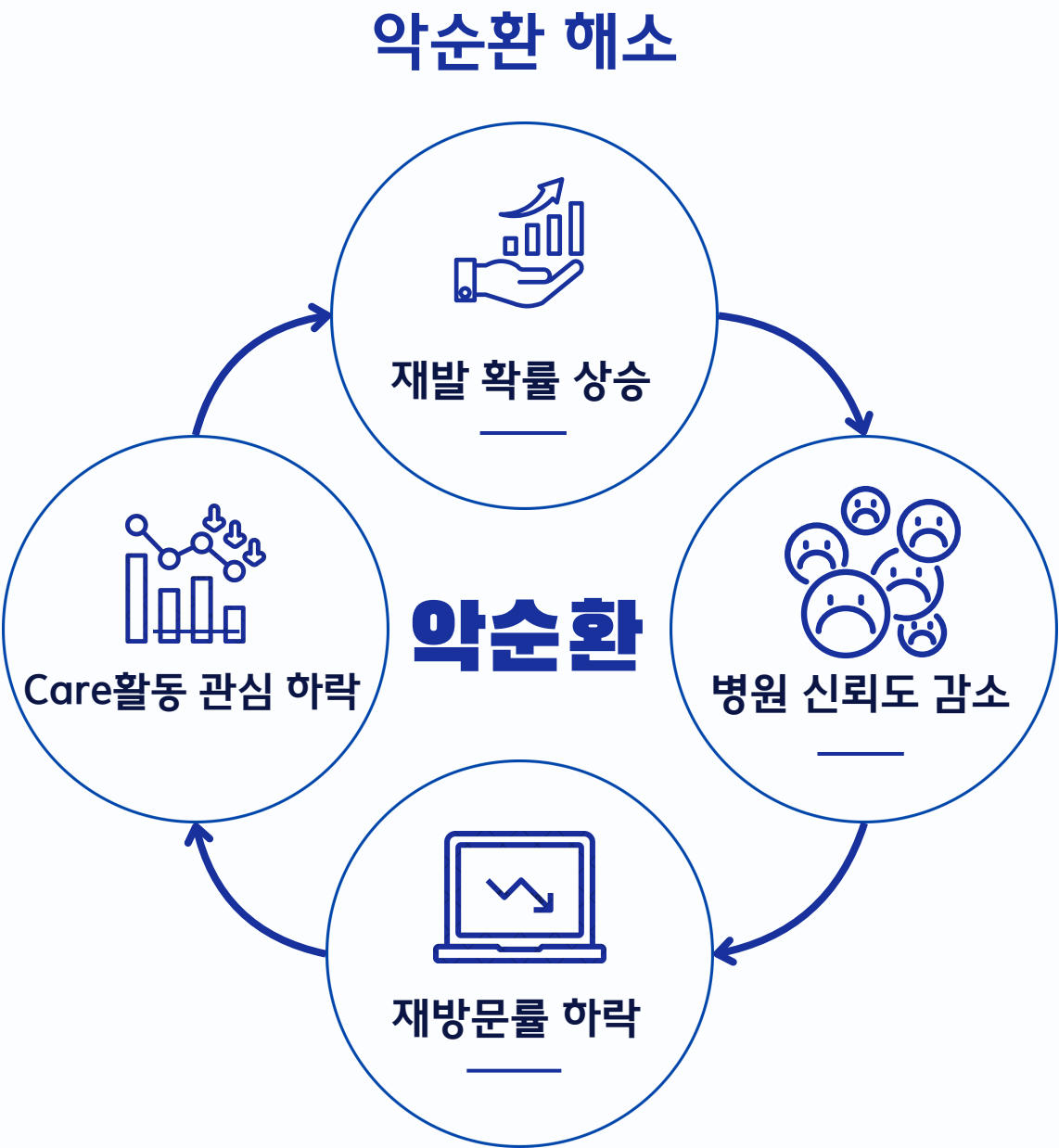
코로나 19여파로 인해 수술 이후 재방문률이 감소함에 따라
병원 수익과 직결된 환자들의 재방문을 도모할 수 있는 전략이 필요

● 비대면 의료서비스 이용 환자 누적수



코로나 19 이후, 국내외에서 비대면 사후 케어 서비스에 대한 수요와 시장이 확대되고 있으므로 이러한 환경에서 병원의 수익성과 재방문율을 높이기 위해서는, 변화하는 시장 수요와 니즈를 반영한 새로운 사후 케어 서비스 도입은 필수적인 전략

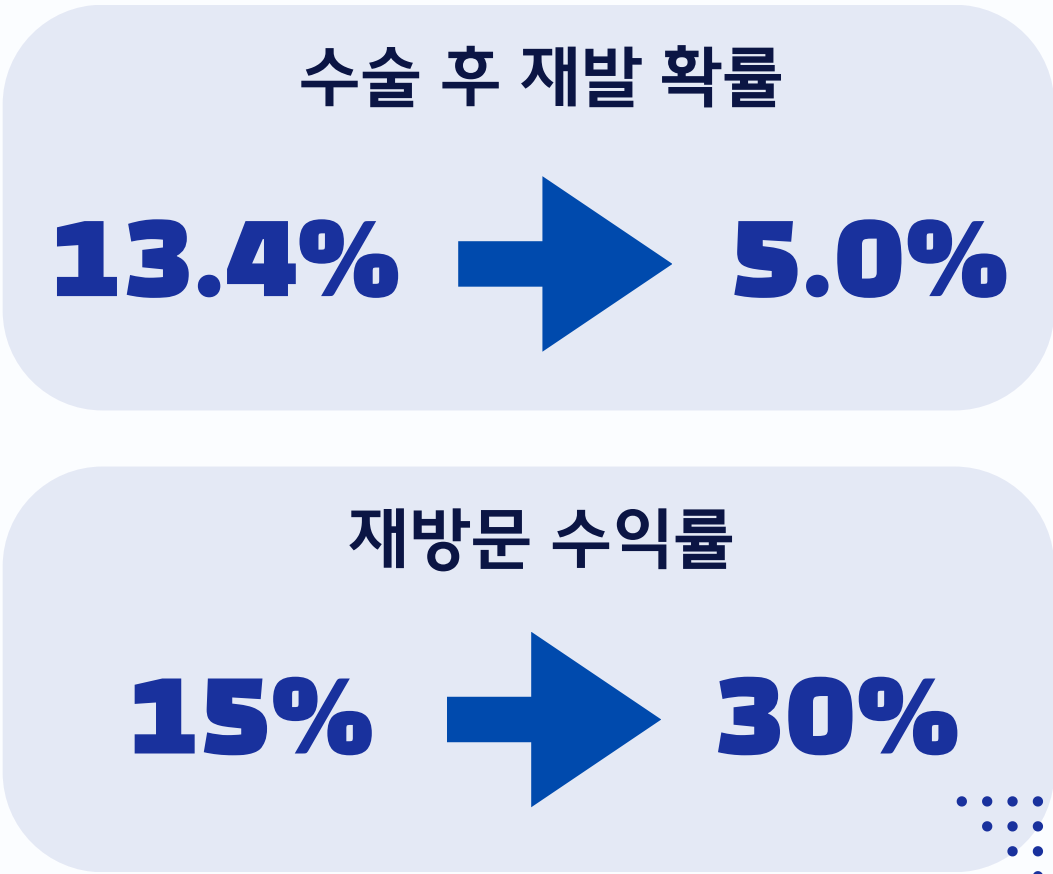
프로젝트 목표



어플 유입률과
Care활동 참여도 증대

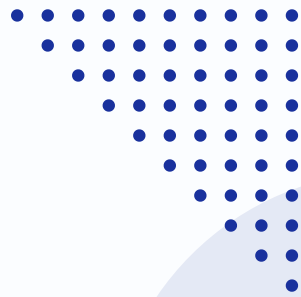


재발 확률 감소 및
재방문 수익률 증대



분석 계획

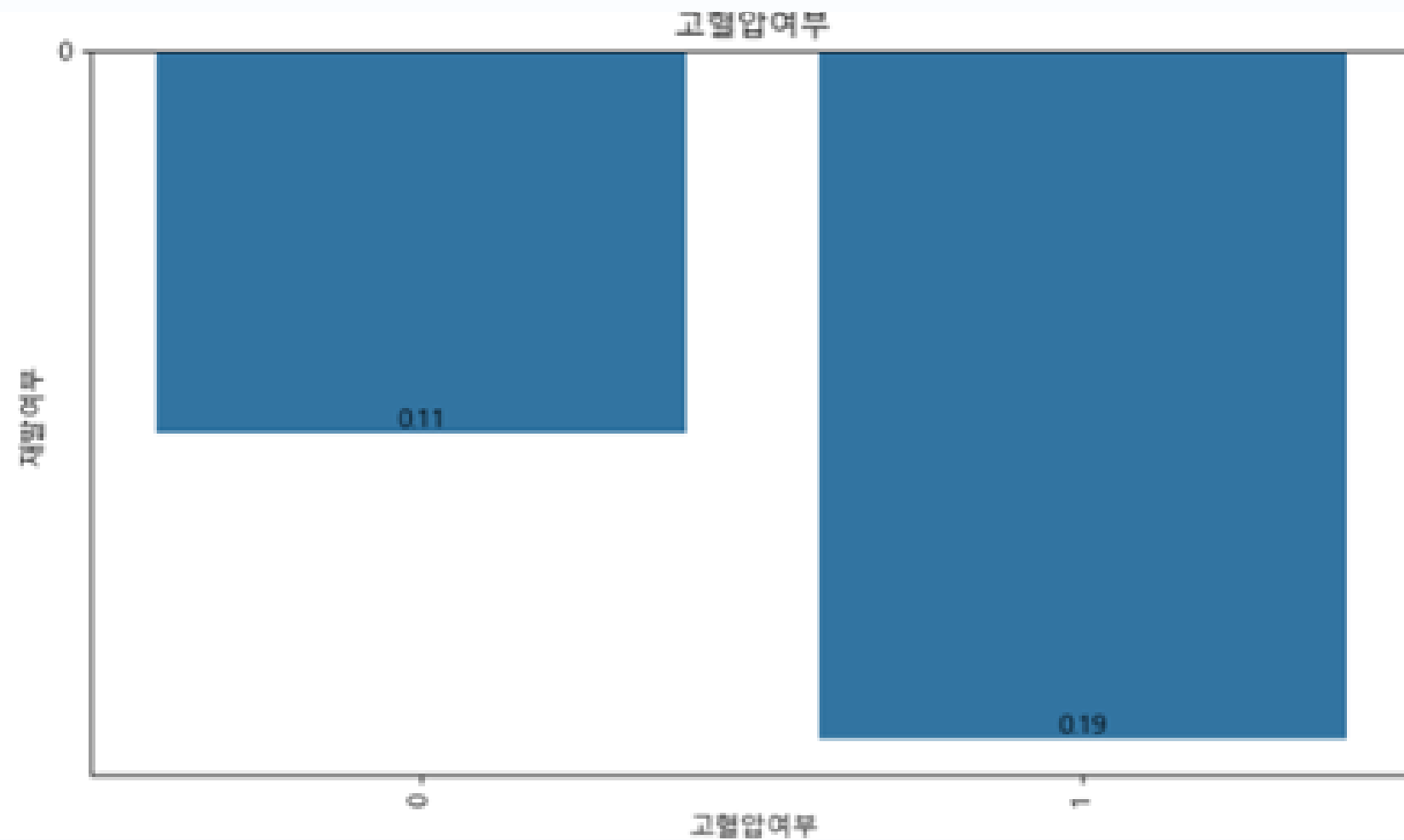
목적	분석 방법	분석 내용
전체 데이터 분포 및 이상치 확인	<ul style="list-style-type: none">Histogram, Box Plot막대그래프	<ul style="list-style-type: none">연속형 변수들의 분포 및 이상치 확인재발여부에 따른 체중의 변화 등 범주형 변수들의 빈도수 확인
군집 파악	<ul style="list-style-type: none">KNNK Means, K-Modes	<ul style="list-style-type: none">연속형 변수들에 대한 군집분석으로, 군집 분리 및 군집 별 특징 파악
각 변수 간의 주요 특성, 관계파악 및 가설검정	<ul style="list-style-type: none">상관계수분석Chi2 Test	<ul style="list-style-type: none">연령대, 수술시간대, 환자통증정도에 따른 재발 비율/직업군 비율 확인연속형 변수 간 상관계수분석 수행수술실패여부, 재발여부, 성별, 가족력, 직업 등 범주형 변수에 대한 연속형 변수와의 관계 파악 후 가설검정 수행이진 범주형 변수 간 가설검정 수행
재발 가능성 예측 모델 개발	<ul style="list-style-type: none">Logistic RegressionRandom Forest	<ul style="list-style-type: none">환자의 특성에 따라 ‘재발여부’를 예측하여 재발 가능성 확인이 가능하고, 재방문률 증대를 위한 전략에 도움을 줄 수 있는 분류/예측 모델 개발Train/Test accuracy, roc-auc 등의 평가지표를 반영해 최우수 모델 선정



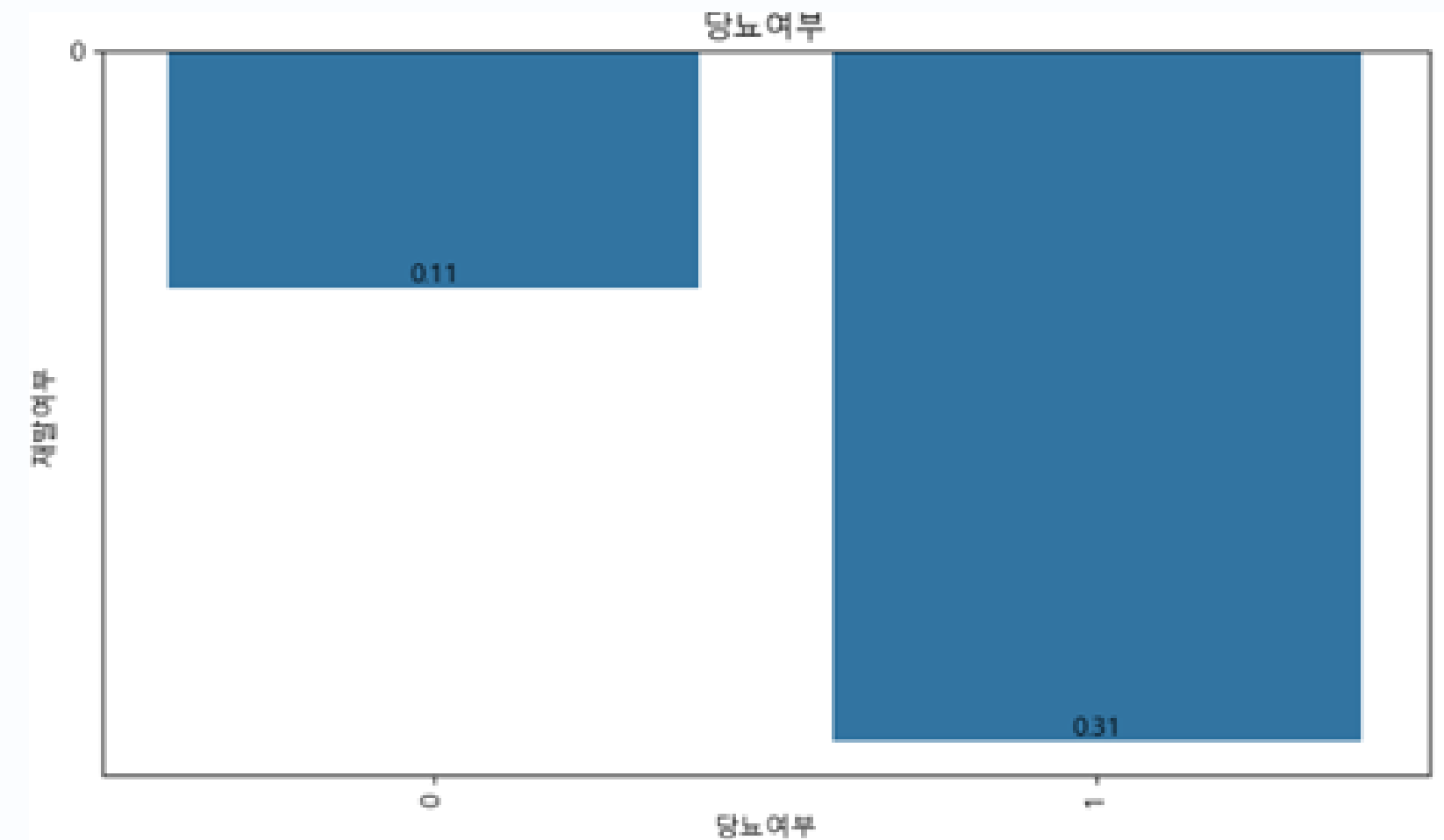
분석결과 및 인사이트

범주형 변수별 재발환자 비율 시각화 & 카이제곱검정

chi2 test
P = 0.00086



chi2 test
P = 1.151521604329538e-08

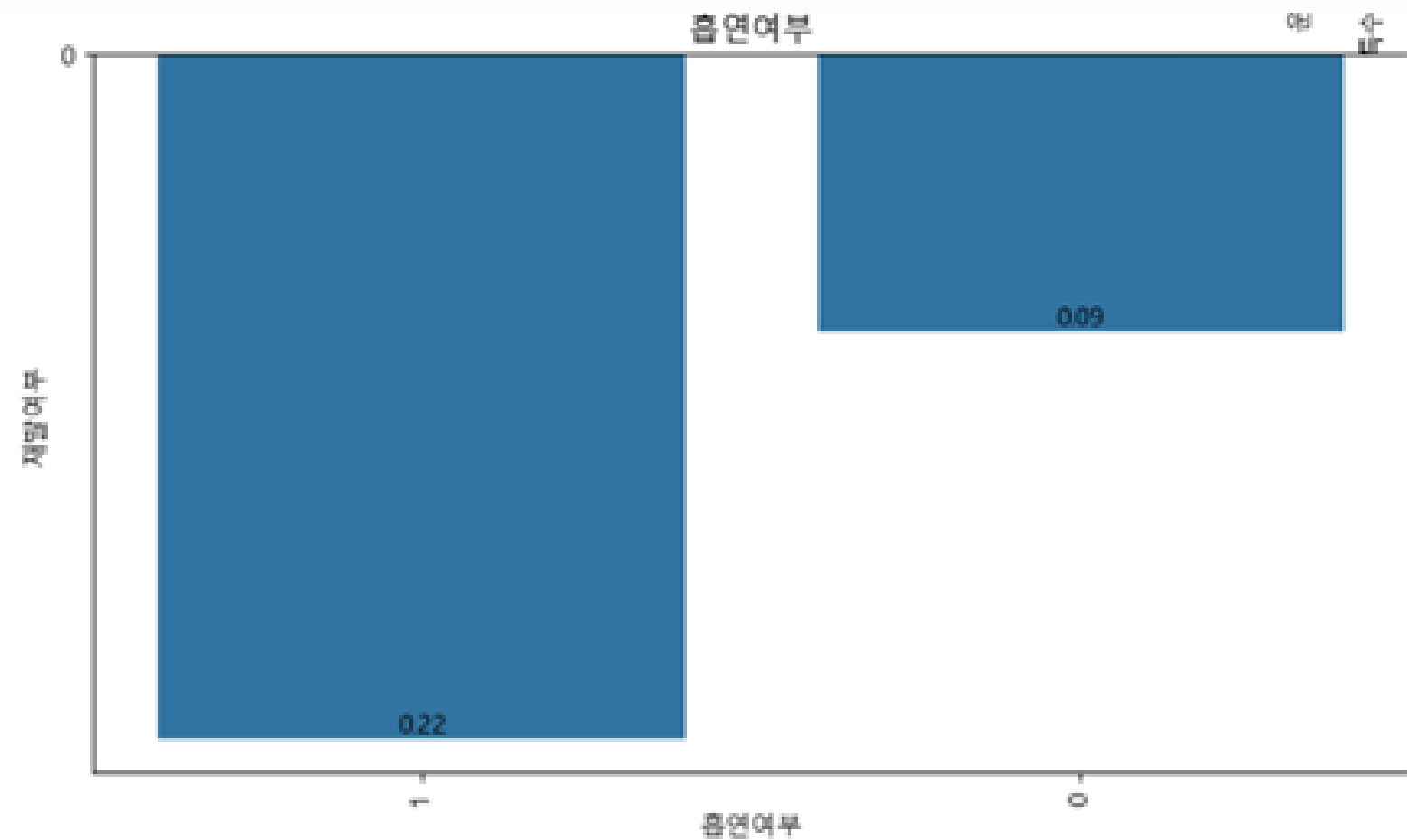


고혈압, 당뇨 기저질환 환자의 재발 확률이 높은것으로 확인되었다.
해당 그룹의 환자들의 경우 Care 서비스의 필요성이 두드러짐.

분석결과 및 인사이트

chi2 test

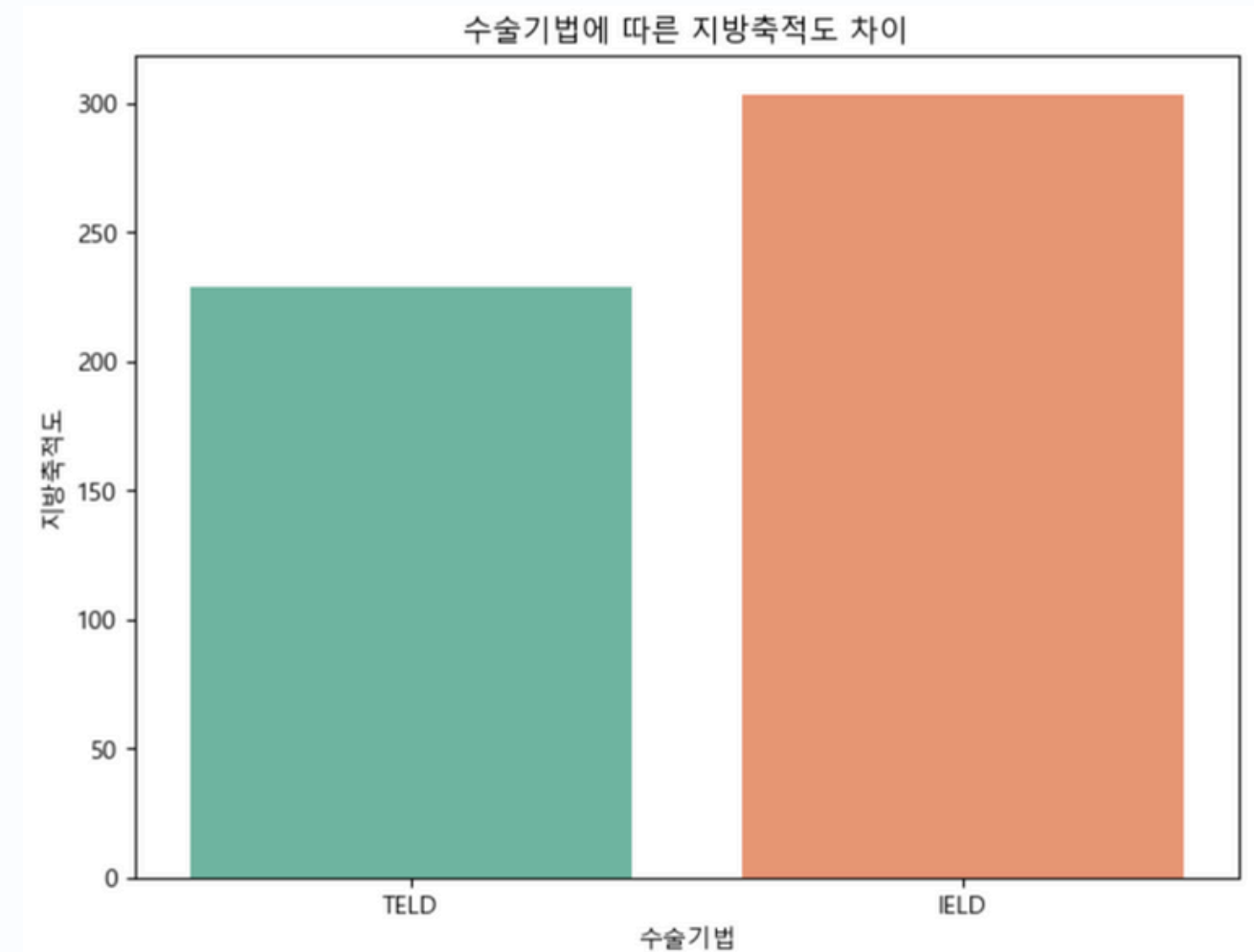
$P = 3.908867127781989e-10$



흡연자의 재발 확률이 높다.
해당 환자들은 케어 서비스가 필요하다.

ANOVA test

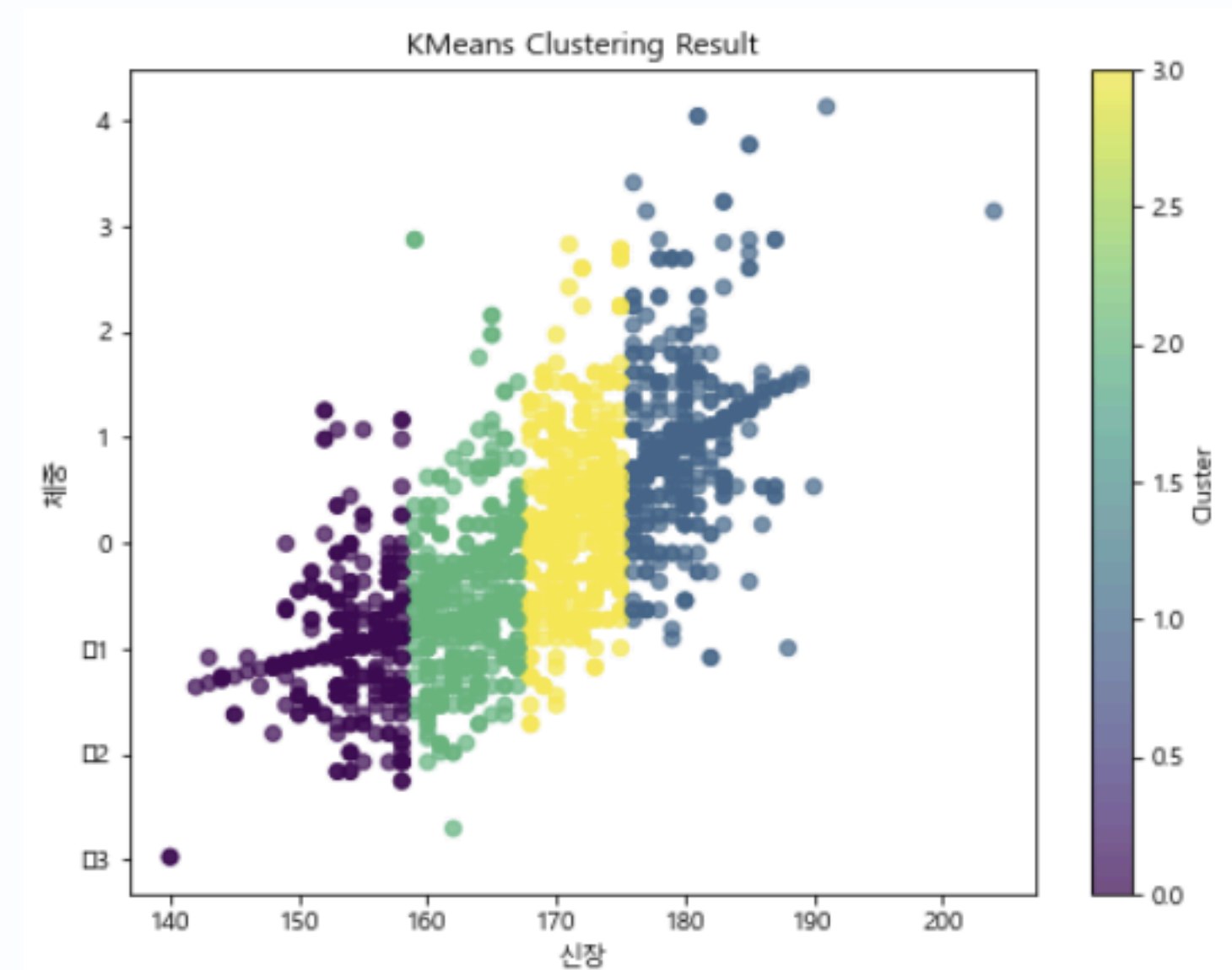
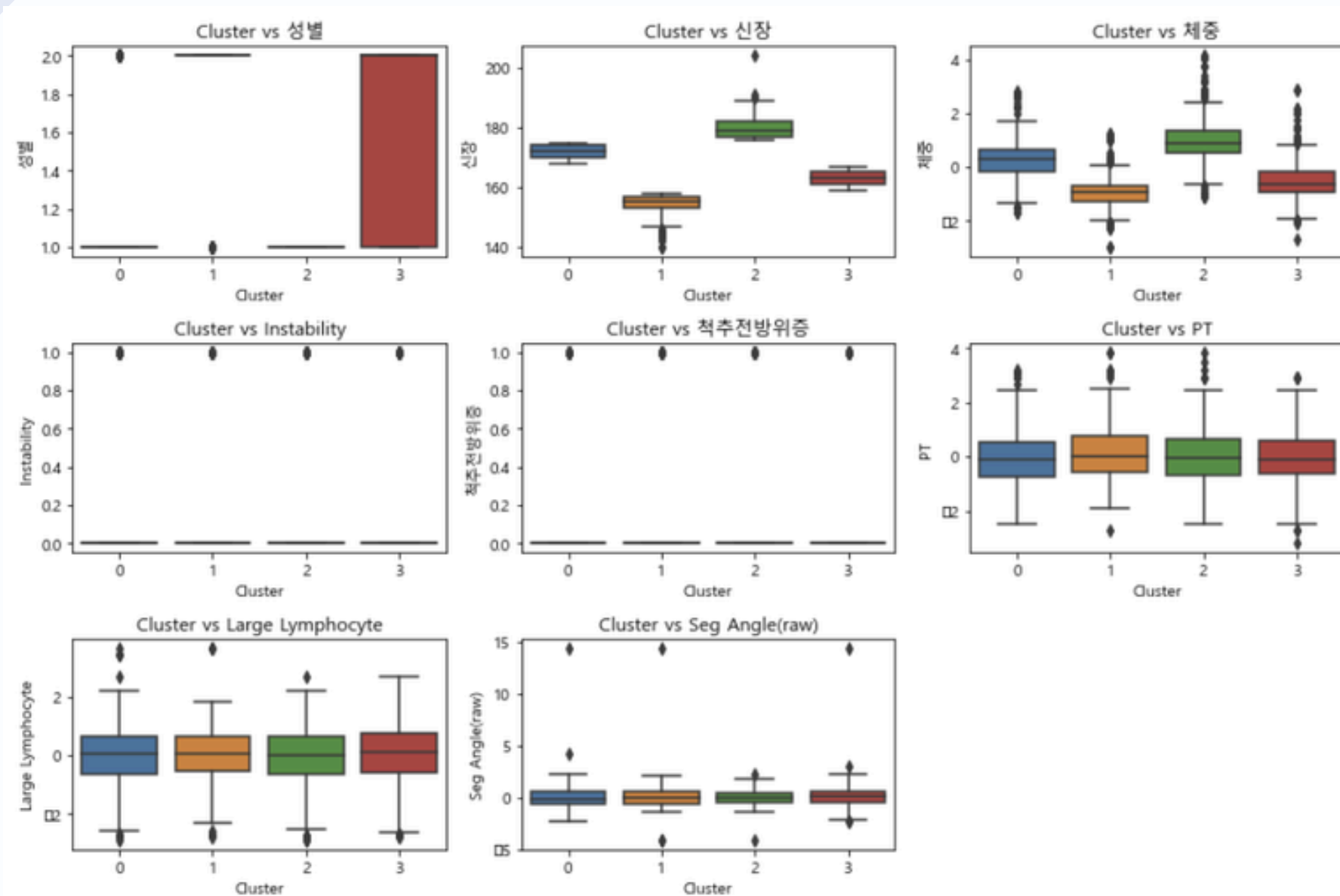
$P = 0.0186$



비만 환자일수록 특정 수술법(IELD)을 따른다.

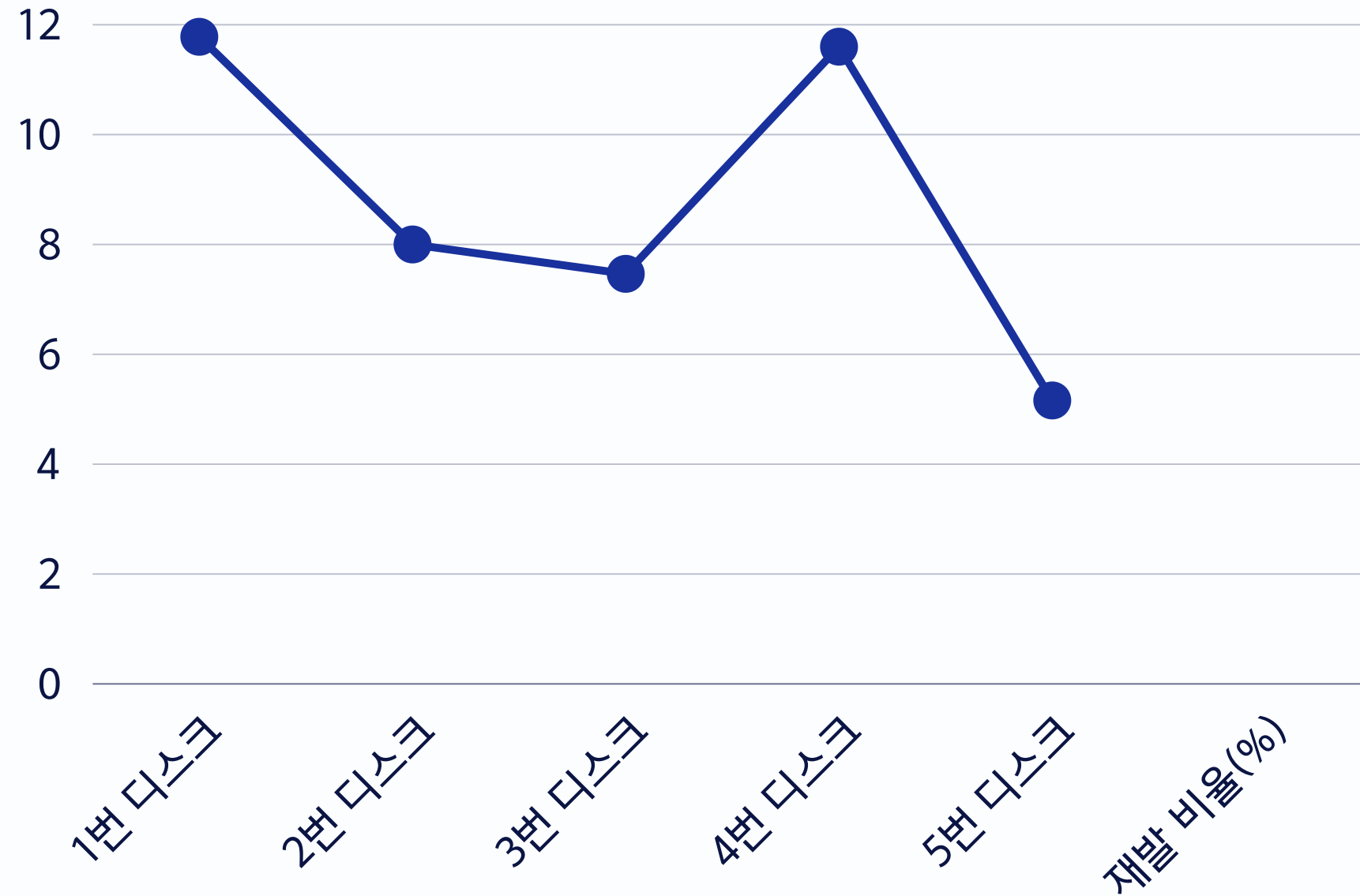
분석결과 및 인사이트

- 신장과 체중에 따라 환자 그룹이 네 그룹으로 두드러지게 나타남



분석결과 및 인사이트

- 디스크 탈출이 1번과 4번에서 일어난 환자의 재발 비율이 높게 나타남에 따라, 1번/4번 디스크 탈출 환자들을 대상으로 사전에 물리치료 등 밀착 Care서비스 필요



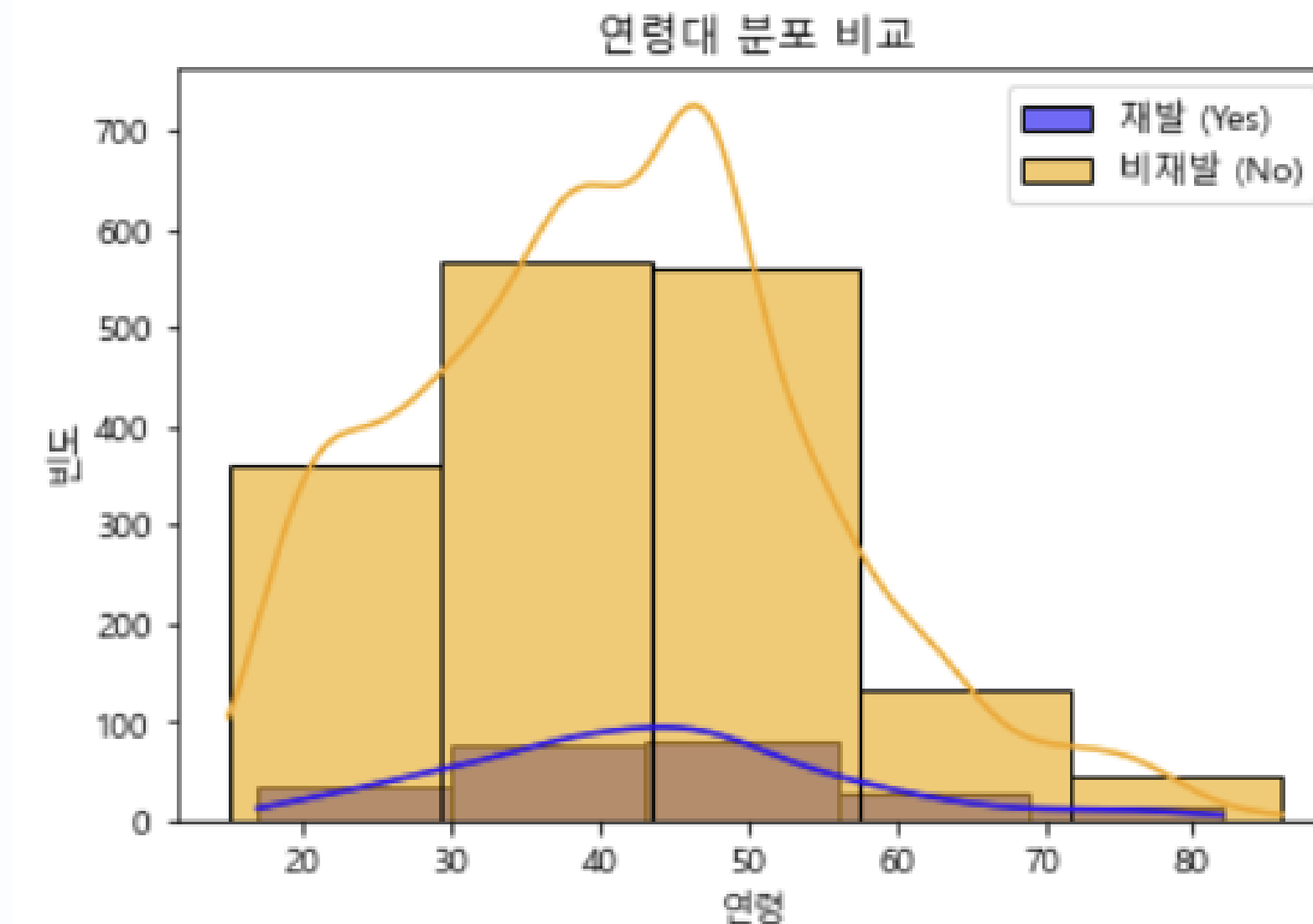
구분	1번 디스크	2번 디스크	3번 디스크	4번 디스크	5번 디스크
재발 여부 있음	410명	467명	556명	128명	25명
재발 여부 없음	38명	66명	86명	12명	6명
비율	11.789%	8.07%	7.546%	11.6%	5.16%

chi2 test

P = 0.046590371812853394

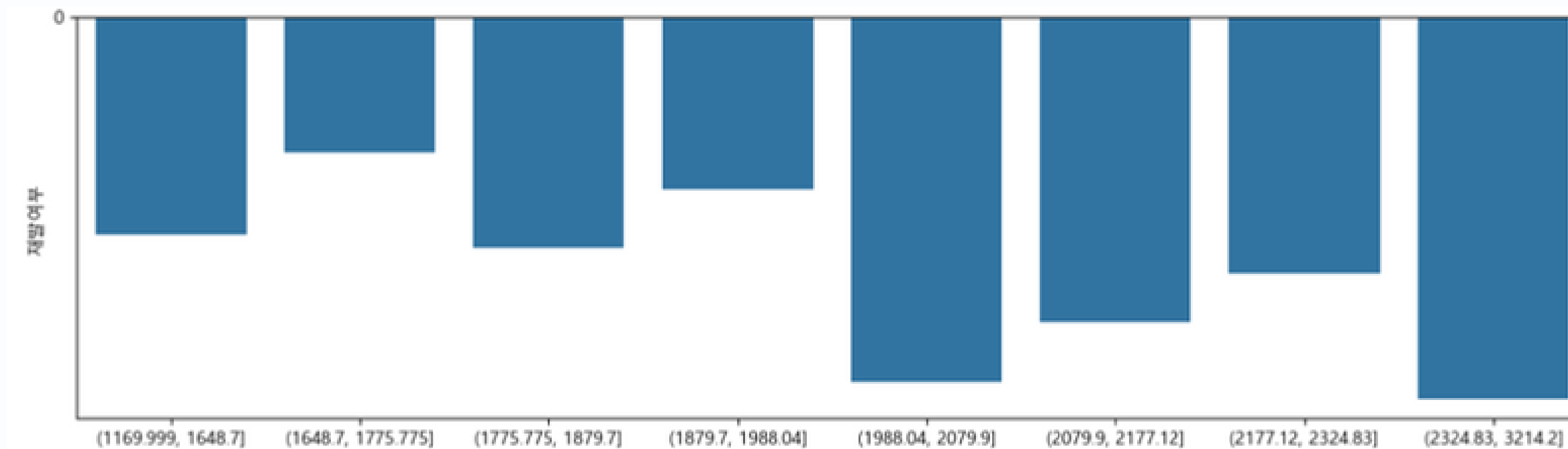
분석결과 및 인사이트

- 60대 이상 환자보다 30-50대 환자 재발율이 더 높은 것으로 확인했다.
- 30-50대 환자를 수술할때 기존 데이터를 가지고 더 면밀한 수술이 필요하다.



분석결과 및 인사이트

- 디스크 면적이 2000이상인 그룹이 미만인 그룹보다 재발여부가 크다 .
- 디스크 면적이 2000이상인 환자들에게 밀착케어 서비스가 필요하다

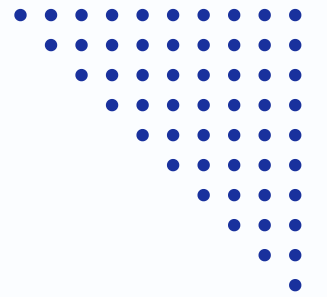


chi2 test
P = 0.007195498719155322



chi2 test
P = 0.0002883987302966692

모델링 과정 : 데이터 특성



- 환자 정보 Data set (image data , 진단 데이터, 수술 데이터) → merge_data

Data columns (total 41 columns):

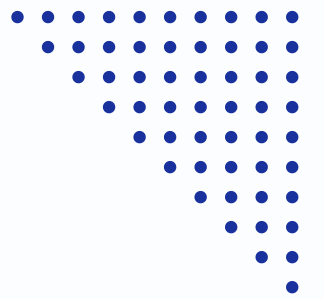
#	Column	Non-Null Count	Dtype
0	전방디스크높이(mm)	1889 non-null	float64
1	후방디스크높이(mm)	1889 non-null	float64
2	지방축적도	1889 non-null	float64
3	Instability	1889 non-null	int64
4	MF + ES	1889 non-null	float64
5	Modic change	1889 non-null	int64
6	PI	1889 non-null	float64
7	PT	1889 non-null	float64
8	Seg Angle(raw)	1889 non-null	float64
9	Vaccum disc	1889 non-null	int64
10	골밀도	1889 non-null	float64
11	디스크단면적	1889 non-null	float64
12	척추전방위증	1889 non-null	int64
13	Large Lymphocyte	1889 non-null	float64

○ ...

14	Location of herniation	1889 non-null	int64
15	ODI	1889 non-null	float64
16	가족력	1889 non-null	float64
17	간질성폐질환	1889 non-null	int64
18	고혈압여부	1889 non-null	int64
19	과거수술횟수	1889 non-null	int64
20	당뇨여부	1889 non-null	int64
21	말초동맥질환여부	1889 non-null	int64
22	빈혈여부	1889 non-null	int64
23	성별	1889 non-null	int64
24	스테로이드치료	1889 non-null	int64
25	신부전여부	1889 non-null	int64
26	신장	1889 non-null	int64
27	심혈관질환	1889 non-null	int64
28	암발병여부	1889 non-null	int64
29	연령	1889 non-null	int64
30	우울증여부	1889 non-null	int64
31	종양진행여부	1889 non-null	int64
32	체중	1889 non-null	float64

- 총 row 수 : 1889 개
- column : 41 개

모델링 과정 : 분류 모델 학습 위한 데이터 전처리



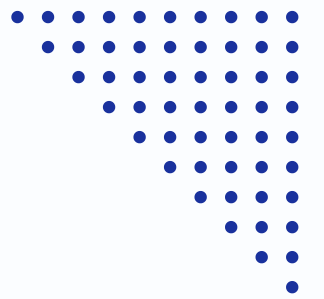
- feature 중요도 분석을 통해 높은 순으로 18개 선정
 - 총 row 수가 약 1800개이기 때문에 10% 인 18개를 설명변수로 가져감

	Feature	Importance
32	체중	0.816733
36	흡연여부	0.503758
23	성별	0.460247
20	당뇨여부	0.339087
37	통증기간(월)	0.190789
13	Large Lymphocyte	0.188461
31	종양진행여부	0.186087
6	PI	0.186058
39	수술실패여부	0.161296
12	척추전방위증	0.155649
14	Location of herniation	0.150558
0	전방디스크높이(mm)	0.143841
11	디스크단면적	0.130747
38	수술시간	0.130190
29	연령	0.129051
8	Seg Angle(raw)	0.129013
17	간질성폐질환	0.121007
26	신장	0.120210

```
features = [  
    "체중", "흡연여부", "성별", "당뇨여부", "통증기간(월)", "Large Lymphocyte",  
    "종양진행여부", "PI", "수술실패여부", "척추전방위증", "Location of herniation",  
    "전방디스크높이(mm)", "디스크단면적", "수술시간", "연령", "Seg Angle(raw)",  
    "간질성폐질환", "신장"  
]  
target = "재발여부"
```

- 연속형 변수 : 표준화 or 정규화
- 범주형 변수 : one - hot encoding

모델링 과정 : 모델링 및 평가 후 선정



로지스틱 회귀분석

- ✓ 정확도 (Accuracy): 0.8942
- ✓ ROC-AUC Score: 0.6896

◆ 분류 보고서 (Classification Report):

	precision	recall	f1-score	support
0	0.89	1.00	0.94	333
1	0.86	0.13	0.23	45
accuracy			0.89	378
macro avg	0.88	0.57	0.59	378
weighted avg	0.89	0.89	0.86	378

랜덤포레스트

- ✓ RandomForest 정확도 (Accuracy): 0.9074
- ✓ RandomForest ROC-AUC Score: 0.8209

◆ 분류 보고서 (Classification Report):

	precision	recall	f1-score	support
0	0.90	1.00	0.95	333
1	1.00	0.22	0.36	45
accuracy			0.91	378
macro avg	0.95	0.61	0.66	378
weighted avg	0.92	0.91	0.88	378

정확도가 0.9074로 더 높은 랜덤포레스트 선정

개선안 : 분류 모델 활용, 재발 예측

재발율에 따른 등급 분류 모델을 통해
새로운 환자 데이터를 입력 후 -> 재발 예측 등급 도출

```
def categorize_risk(prob):  
    if prob < 0.3:  
        return "Low Risk (낮음)"  
    elif prob < 0.7:  
        return "Medium Risk (중간)"  
    else:  
        return "High Risk (높음)"
```

환자 예시 DATA



성별 : 남성

체중 : 70

연령 : 55

당뇨여부 : no

디스크단면적: 15.0

PI : 1.2

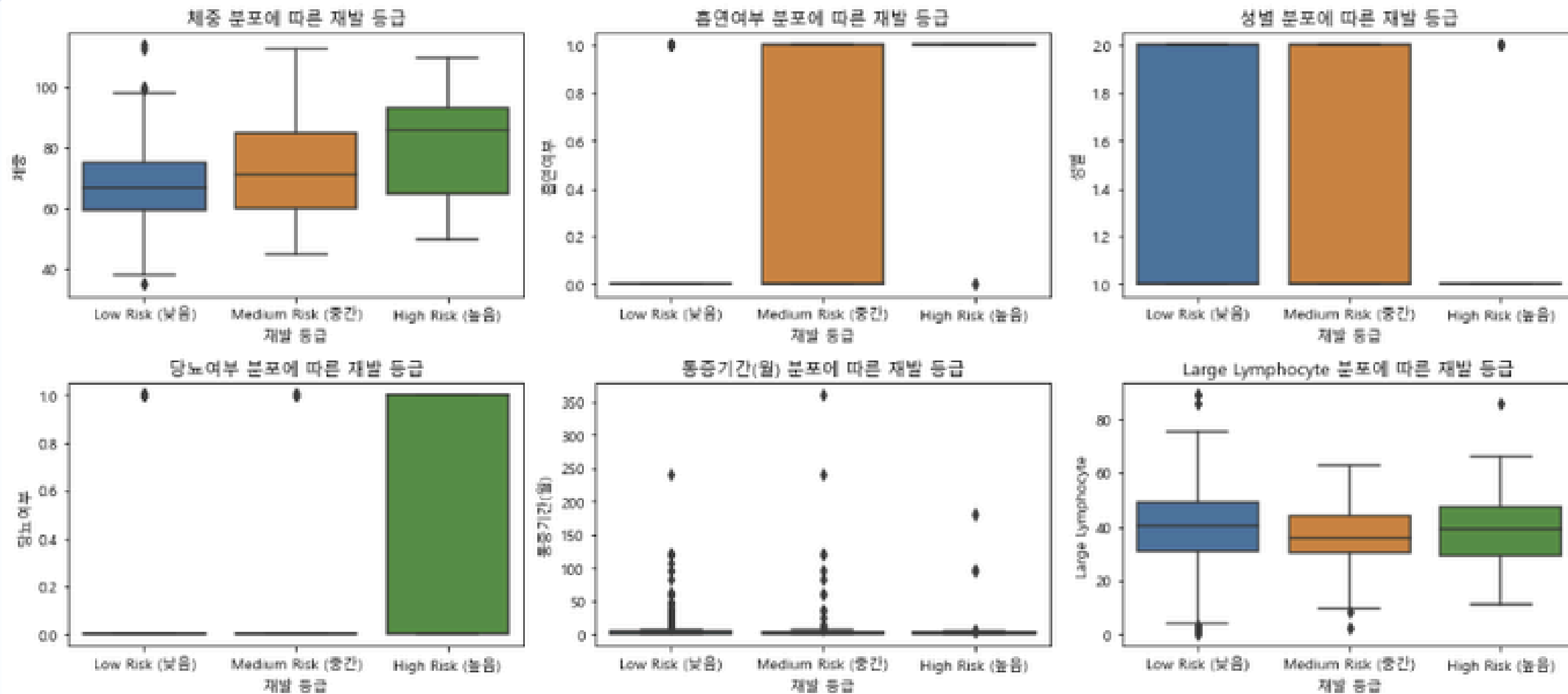
+ 그 외 데이터

결과 : 환자의 재발 확률 21.17%

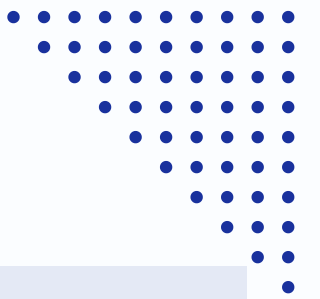
★ 재발 등급: Low Risk (낮음)

개선안 : 군집 분류 모델 활용, 체중관리 케어 서비스

- 재발 위험도 등급에 따라 체중분포가 달라지기 때문에 케어서비스는 '체중관리'에 초점



개선안



고혈압, 당뇨 기저질환 환자들을 위한
care서비스 필요

혈압관리, 혈당 관리등 맞춤형 프로그램
신규 개설

흡연자 케어 서비스 필요

흡연자 수술 전후 금연 프로그램 개설

30-50대 환자를 위한 케어 서비스 필요

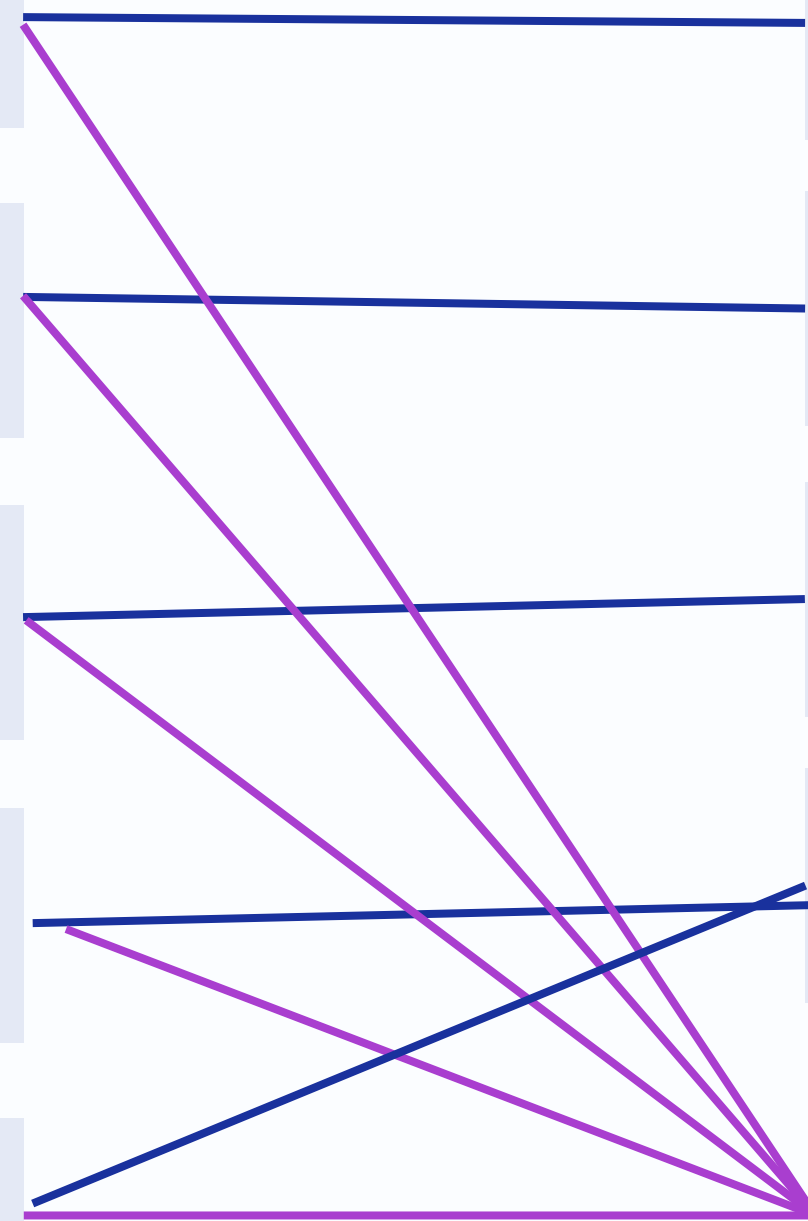
직장인을 고려한 진료 서비스

1번/4번 디스크 탈출 환자
대상 집중적 관리 및 Care 서비스 필요

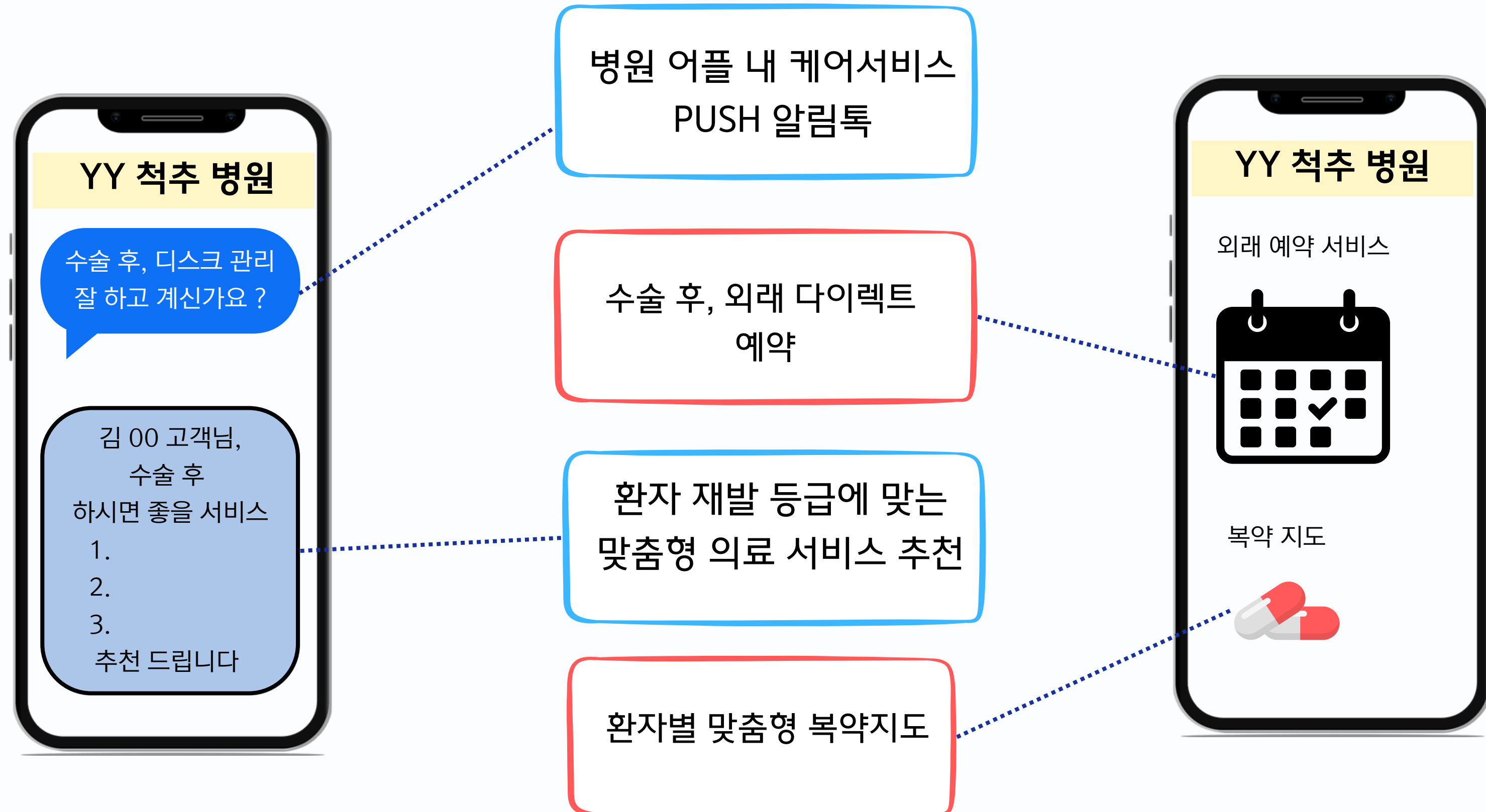
정기 검진 및 예방 교육 진행

디스크단면적이 2000mm 이상인
환자를 위한 밀착케어 서비스

재발 위험 환자 대상
APP PUSH 알림 서비스 시행



개선안 : APP PUSH 알림 서비스



프로젝트 소감



1. 비대면으로 진행되어서 팀원들과의 소통이 아쉬웠지만 직접 데이터를 분석하고 모델링하여 인사이트까지 경험할 수 있어서 의미있는 프로젝트였습니다. (이예진)
2. 실제 현장의 데이터를 분석해보며 기대했던 변수들과의 뚜렷한 상관관계가 나타나지 않아 어려움이 많았지만, 보다 정교한 모델 설계와 서비스 기획의 필요성을 다시금 느낄 수 있었던 값진 경험이었습니다. (임지웅)
3. 프로젝트가 비대면으로 진행되어 아쉬운 부분이 있었지만, 인사이트 도출을 위해 데이터를 다방면으로 분석하는 좋은 경험을 쌓은 것 같아 만족스럽습니다. (윤택한)
4. 평소에 접하지 못했던 ‘의료도메인’ 관련 데이터를 다뤄볼 수 있는점이 좋았습니다. 또한 실제 현업과 비슷하게 깔끔하지 않은 데이터를 어떻게 처리하고 인사이트를 낼 수있을지 고민하는 과정도 좋았습니다.(고유경)
5. 프로젝트 시작 전에 데이터 분석 방법에 대한 교육 진행으로 이전에 부족했던 부분들을 많이 개선할 수 있게 되어서 좋았고 평소에 보기 힘들었던 실제 현장의 데이터를 접함으로 데이터를 다루는 통찰력을 키울 수 있었던 좋은 경험이었습니다. (최영환)
6. 실제와 유사한 비즈니스 문제를 직접 해결해 볼 수 있다는 점이 가장 좋았습니다. 그저 성능이 좋은 모델을 만드는 것이 아니라 비즈니스적으로 도움이 되는 방향으로 분석과 모델링을 진행해야 하는 것이 중요하다는 것을 느낄 수 있었습니다. (조은나라)

THANK YOU

감사합니다