심장병 조기 경보 모델 개발

발등튀김

강다훈 · 김태연 · 유현 · 이지민

목차

심장 질환?

2023년 대한민국 사망원인통계 2위, **심장질환**
심장병을 조기에 예측하지 못할 경우 막대한 사회·경제적 비용을 초래하지만,
적절한 예측 및 조기 개입을 통해 충분히 예방할 수 있다.

분석 목적

심장병(HeartDisease) 발생 여부를
예측하는 분류 모델 개발

↓

의미 있는 임상적 예측성을 위해,
소수 클래스(positive case)를
잘 예측하는 것에 집중

문제 정의

소수 클래스(심장병 O)를 잘 분류+해석 용이한 모델을 개발하고자 함.

[주요 평가지표]

- **Recall (재현율)**: 실제 질환자 중 맞춘 비율
  → False Negative 방지
- Precision (정밀도): 질환이라고 예측한 것 중 실제 질환자 비율
- F1 Score: 정밀도와 재현율의 조화 평균
- ROC AUC: 전체 이진 분류 성능 평가

데이터 선정

데이터 출처

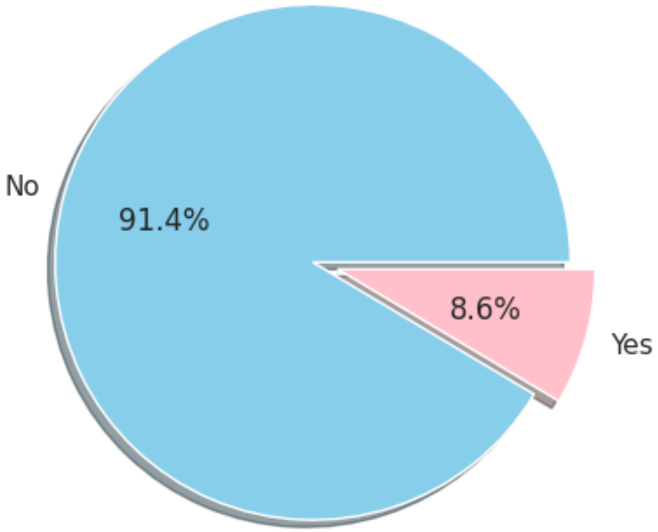
Kaggle **‘Indicators of Heart Disease’**

2020년 미국 질병통제예방센터(CDC)가 실시한
미국 성인 대상 연간 건강 설문조사 데이터

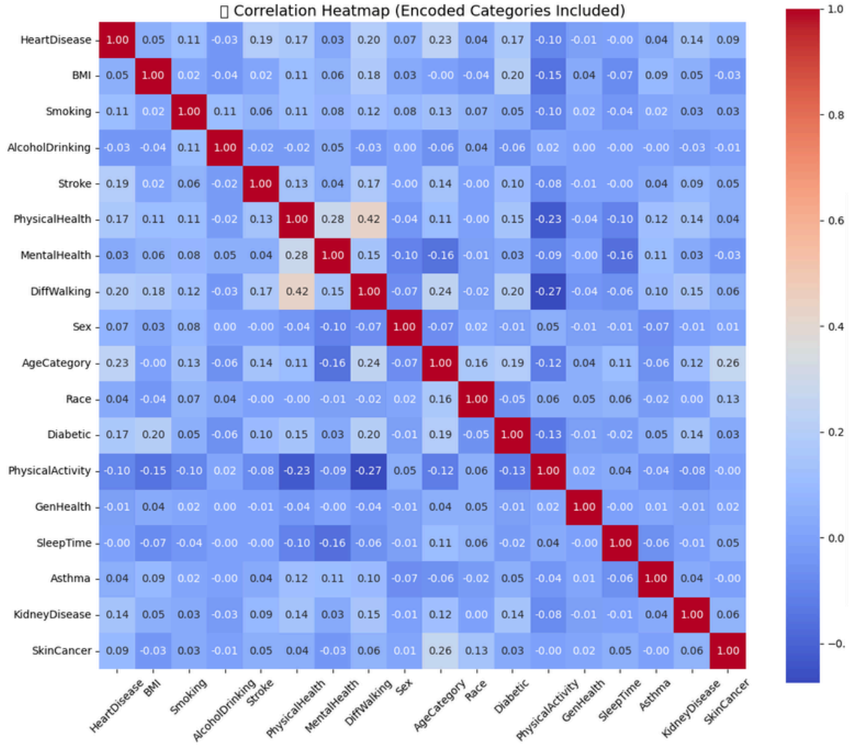
변수 설명 315252 rows × 18 columns

분류	변수명	설명
타겟 변수	HeartDisease	심장병 여부 (관상동맥질환 또는 심근경색 병력 여부)
인구통계학적 정보	Sex	성별
	AgeCategory	나이 구간 (14단계 구간)
	Race	인종
건강 지표	BMI	체질량지수 (Body Mass Index)
	PhysicalHealth	최근 30일 중 신체적으로 건강하지 않았던 날 수
	MentalHealth	최근 30일 중 정신적으로 건강하지 않았던 날 수
	GenHealth	본인의 일반적인 건강 상태에 대한 자가 평가
습관 및 활동	Smoking	평생 100개비 이상의 담배를 피운 적 있는가?
	AlcoholDrinking	고위험 음주 여부 (남성 주 14잔 초과, 여성 주 7잔 초과)
	PhysicalActivity	최근 30일 이내 여가 시간 중 운동 또는 신체 활동을 한 적 있는가?
	DiffWalking	계단 오르기 또는 걷기 어려움 존재 여부
	SleepTime	하루 평균 수면 시간(1~24)
병력	Diabetic	당뇨 진단 여부 (임신성 당뇨 포함 여부는 데이터에 따라 다름)
	Stroke	뇌졸중 진단 여부 (병원에서 받은 적 있는가?)
	Asthma	천식 진단 여부
	KidneyDisease	신장 질환 진단 여부 (요로결석, 방광염 등 제외)
	SkinCancer	피부암 진단 여부

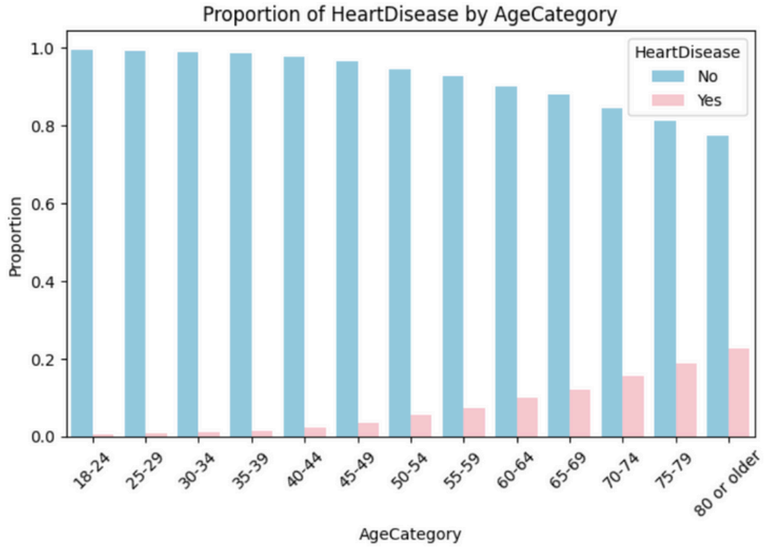
EDA 및 시각화



심장병 유무 (Yes/No) 비율
: 클래스 불균형 존재



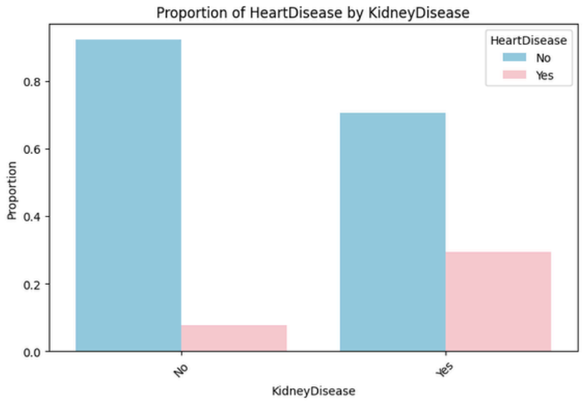
전체 변수 히트맵



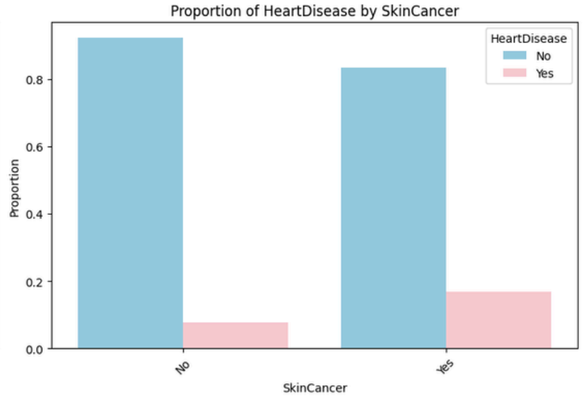
연령대 별 심장병 비율

EDA 및 시각화

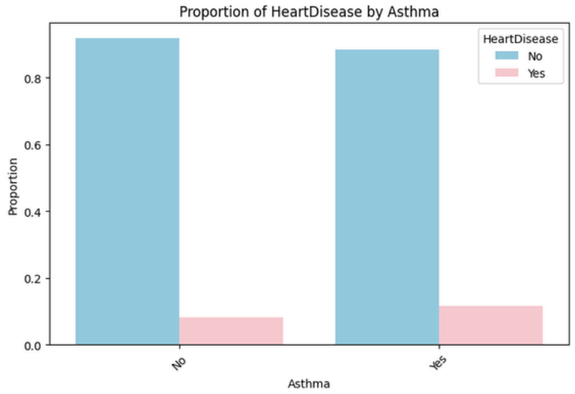
병력별 심장병 발병 비율



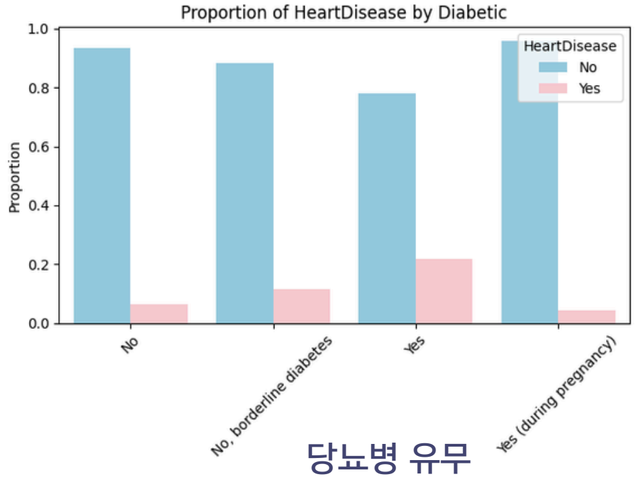
신장질환 유무



피부암 유무

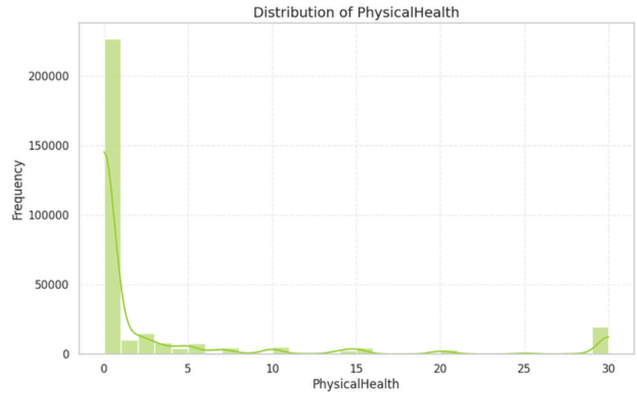


천식 유무

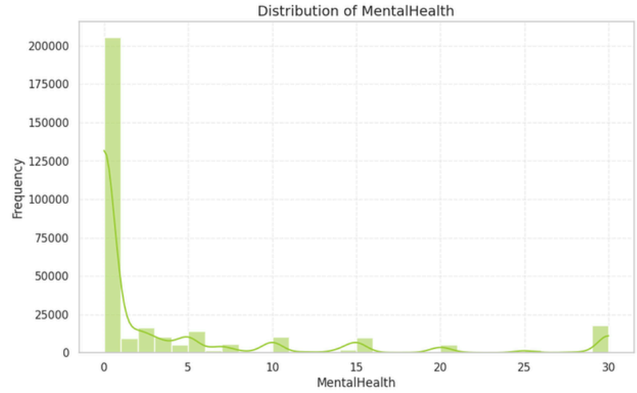


당뇨병 유무

심한 우측 왜도 & 0 근처로 쏠린 값

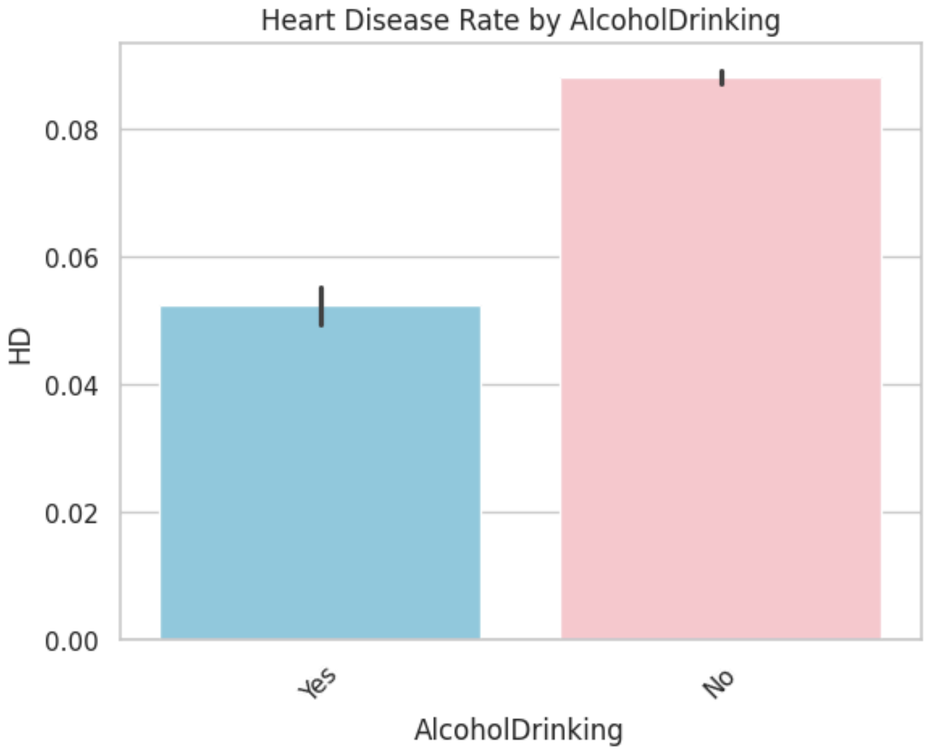


신체 불건강 일수

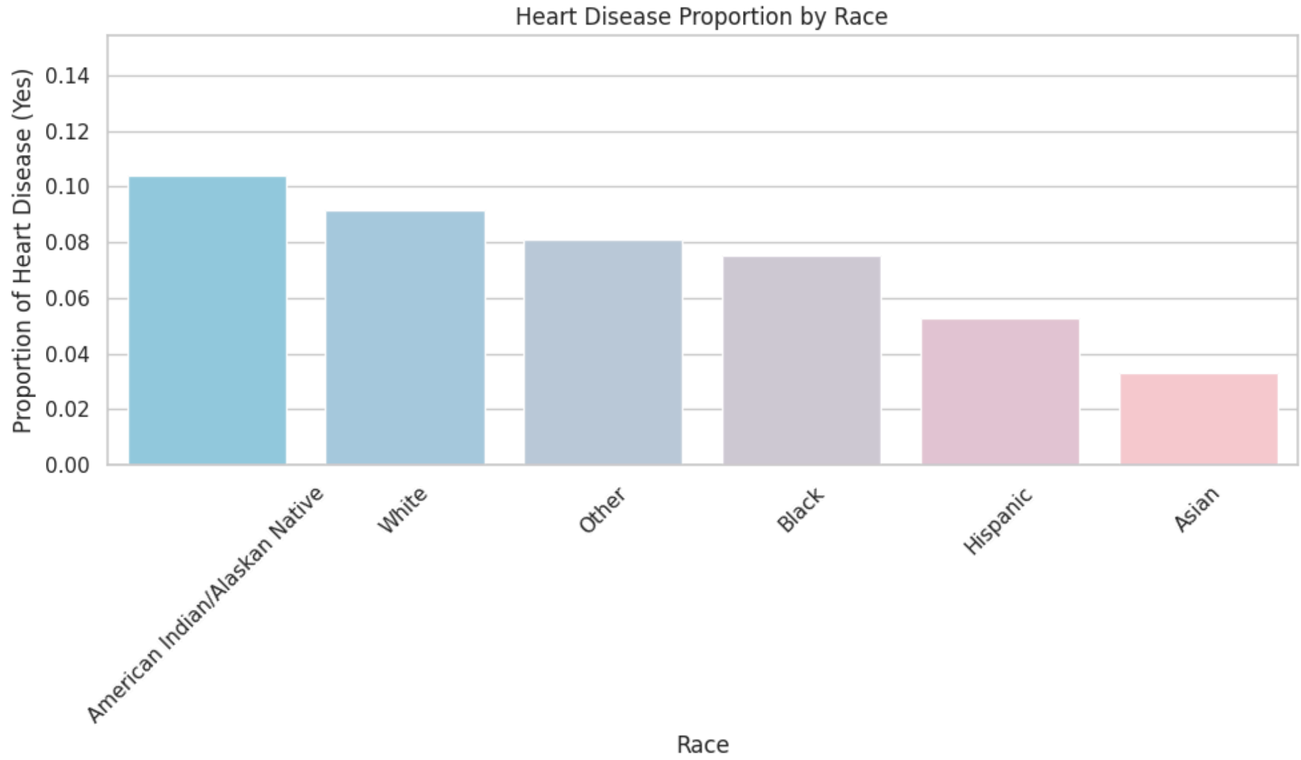


정신 불건강 일수

EDA 및 시각화

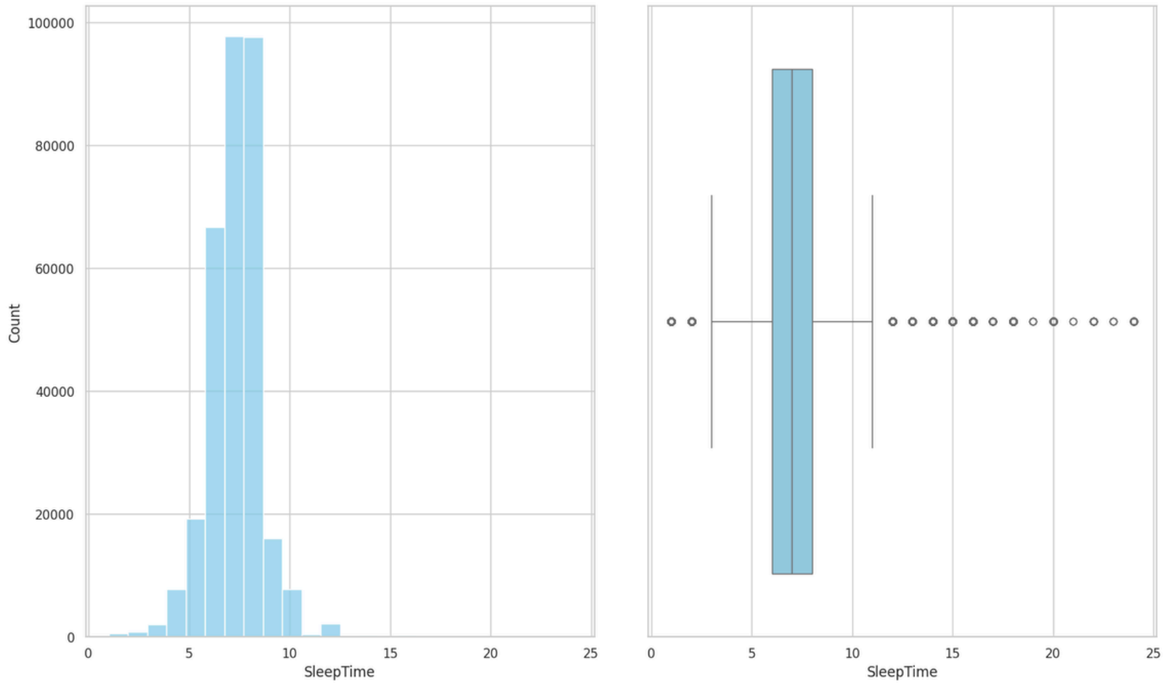


비과음자 집단이 과음자보다
높은 심장병 발생률을 보임

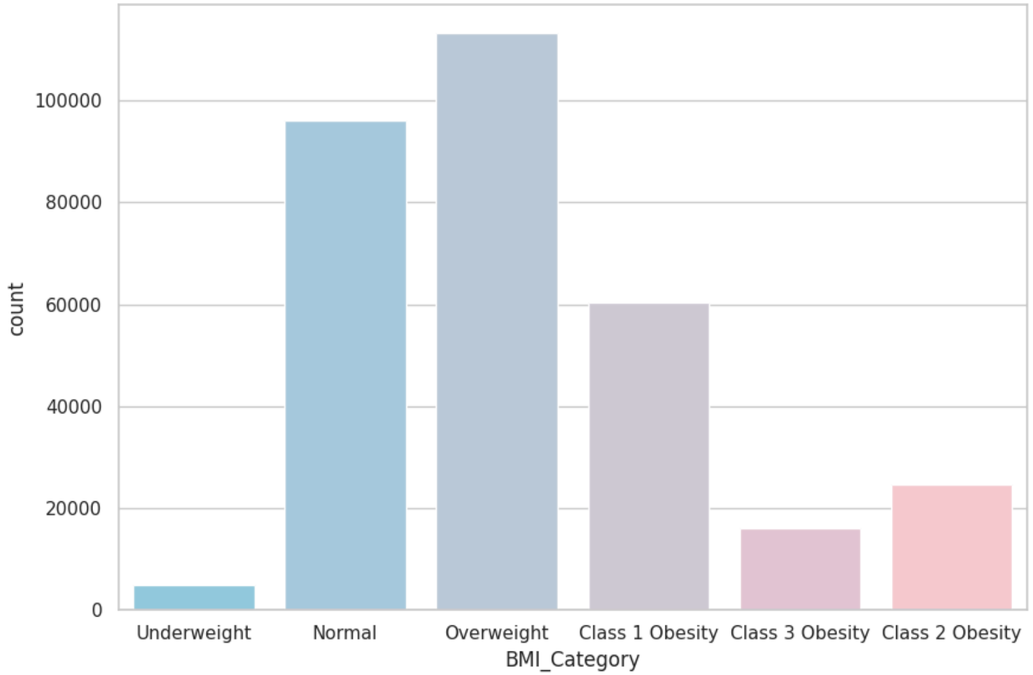


인종별로 상이한 심장 발병 양상

전처리



SleepTime 이상치 제거
: IQR 방식



BMI 범주화(BMI_Category)
: 미 질병통제예방센터(CDC) 기준

전처리

범주형 변수 인코딩

순서형: Label Encoding

```
'GenHealth'    : ['Poor', 'Fair', 'Good', 'Very good', 'Excellent'], '
AgeCategory': ['18-24', '25-29', '30-34', '35-39', '40-44',
               '45-49', '50-54', '55-59', '60-64', '65-69',
               '70-74', '75-79', '80 or older'], '
BMI_Category': ['Underweight', 'Normal', 'Overweight', '
               Class 1 Obesity', 'Class 2 Obesity', 'Class 3 Obesity']
```

명목형: One-Hot Encoding

```
 'Smoking', 'AlcoholDrinking', 'Stroke', 'DiffWalking', 'Sex', 'Diabetic',
'PhysicalActivity', 'Asthma', 'KidneyDisease', 'SkinCancer', 'Race'
```

왜도 심한 수치형 변수 (PhysicalHealth, MentalHealth) 로그 변환

```
df['PhysicalHealth'] = np.log1p(df['PhysicalHealth'])
df['MentalHealth'] = np.log1p(df['MentalHealth'])
```

가설검정 0

모든 변수에 대한 유의성 검정
: 독립변수와 종속변수가 각각 통계적으로 유의한 관계에 있는가?

범주형 변수

카이제곱검정

```
categorical_cols = ['Smoking', 'AlcoholDrinking', 'Stroke',
        'DiffWalking', 'Sex', 'AgeCategory', 'Race',
        'Diabetic', 'PhysicalActivity', 'GenHealth',
        'Asthma', 'KidneyDisease', 'SkinCancer', 'BMI_Category']
```

```
Smoking p-value : 0.0000000000
AlcoholDrinking p-value : 0.0000000000
Stroke p-value : 0.0000000000
DiffWalking p-value : 0.0000000000
Sex p-value : 0.0000000000
AgeCategory p-value : 0.0000000000
Race p-value : 0.0000000000
Diabetic p-value : 0.0000000000
PhysicalActivity p-value : 0.0000000000
GenHealth p-value : 0.0000000000
Asthma p-value : 0.0000000000
KidneyDisease p-value : 0.0000000000
SkinCancer p-value : 0.0000000000
BMI_Category p-value : 0.0000000000
```

수치형 변수

로지스틱회귀

```
numeric_cols = ['PhysicalHealth', 'MentalHealth', 'SleepDeviation']
```

```
PhysicalHealth
  Coefficient      : 0.0519
  Raw p-value      : 0.0000
  Adjusted p-value: 0.0000
  Odds Ratio       : 1.0533
MentalHealth
  Coefficient      : 0.0106
  Raw p-value      : 0.0000
  Adjusted p-value: 0.0000
  Odds Ratio       : 1.0107
SleepDeviation
  Coefficient      : 0.2953
  Raw p-value      : 0.0000
  Adjusted p-value: 0.0000
  Odds Ratio       : 1.3435
```

*이때, 수면시간의 비선형적 영향은 다음과 같이 반영

```
df['SleepDeviation'] = abs(df['SleepTime'] - 7)
```

가설검정 1

과음하는 사람의 심장병 발생률이 더 낮다?!

2표본 비율 z-검정

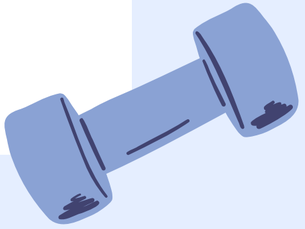
H0: 과음자와 비과음자의 심장병 발생률은 같거나, 과음자가 더 많다.

H1: 과음자가 비과음자보다 심장병에 더 적게 걸린다.

Z-statistic: -18.2527
p-value < 0.05
→ H0 기각, H1 채택

통계적으로는 과음자의 심장병 발생률이 낮다.
: 관찰된 상관관계일 뿐, 인과관계가 아님.

→ **잠재적 교란 변수**(confounding) 또는
선택 편향 가능성 존재

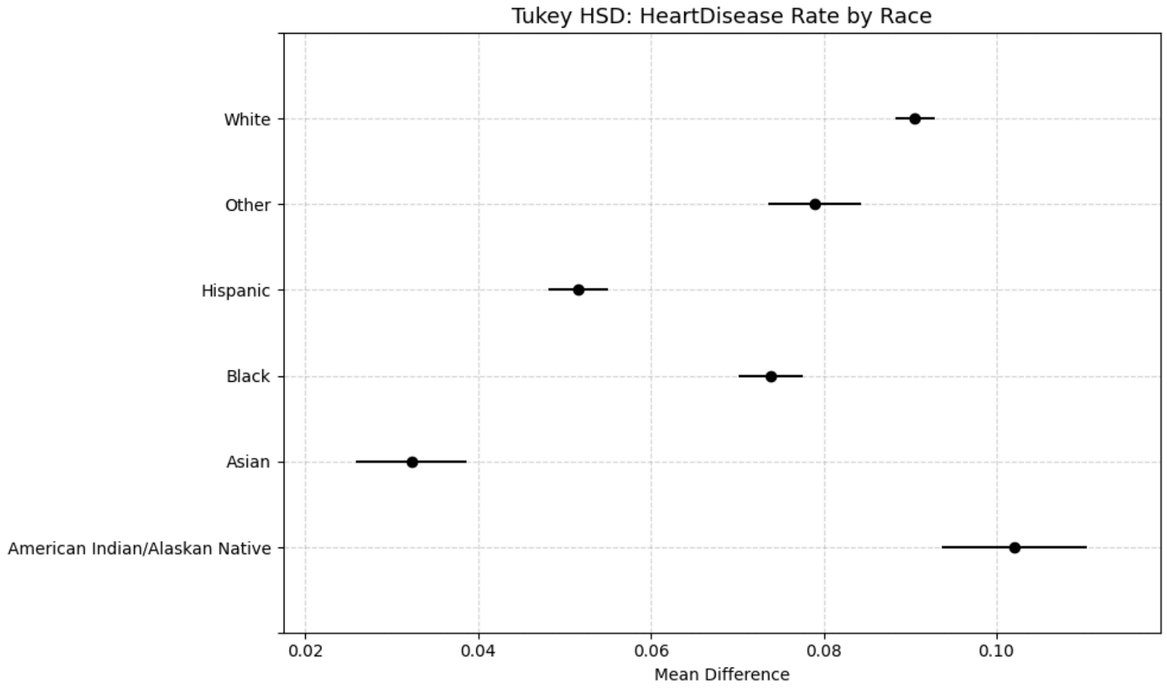


가설검정 2

인종에 따라 심장병 발생에 유의한 차이가 있다?

ANOVA 검정

H0 : 모든 인종 그룹에서 HeartDisease의 평균이 동일하다.
H1 : 적어도 하나의 그룹은 평균이 다르다.



PR(>F)

5.941246e-178

=> H0 기각, H1 채택

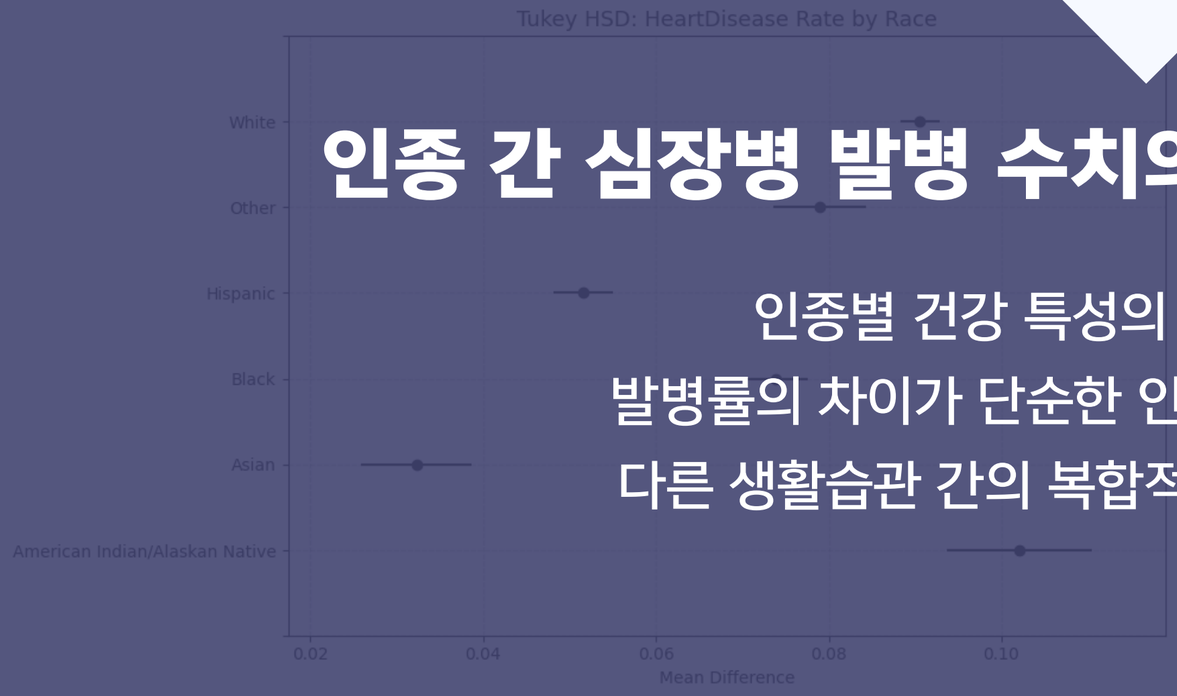
=> 적어도 하나의 그룹은 평균이 다름을 검증

가설검정 2

인종에 따라 심장병 발생에 유의한 차이가 있다?

ANOVA 검정 결과,

인종에 따라 심장 발병률에 유의미한 차이가 있음



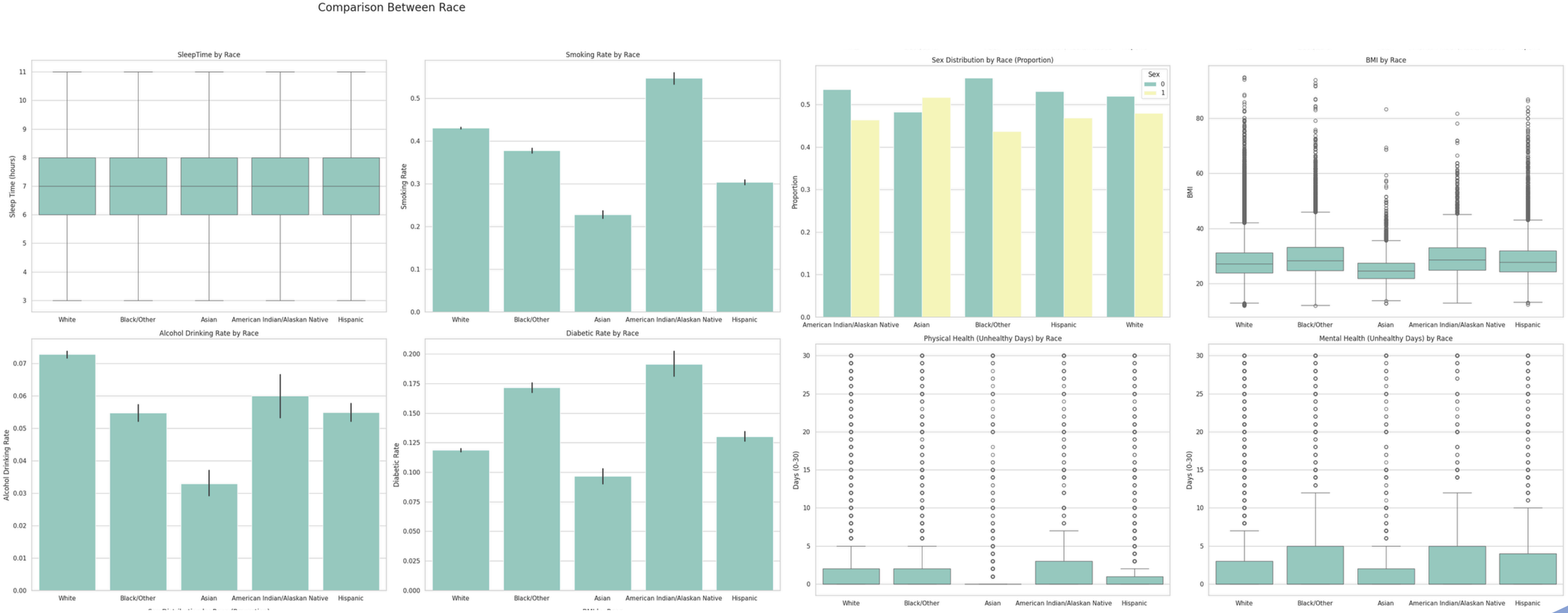
인종 간 심장병 발병 수치의 차이, 어떤 영향인가?

인종별 건강 특성의 차이를 검정하여
발병률의 차이가 단순한 인종별 특성의 차이인지,
다른 생활습관 간의 복합적 관계 때문인지를 탐색

가설검정 2

인종에 따라 심장병 발생에 유의한 차이가 있다?

Race에 따른 생활 습관 관련 특성 시각화



가설검정 2

인종에 따라 심장병 발생에 유의한 차이가 있다?

시각화 결과 의미있는 6개 변수로 ANOVA 검정 진행

(Smoking, AlcoholDrinking, Diabetic, BMI, PhysicalHealth, MentalHealth)

H0 : 모든 인종 그룹에서 각 [컬럼]의 평균이 동일하다

H1 : 적어도 하나의 그룹은 각 [컬럼]의 평균이 다르다

```
=== ANOVA: Smoking by Race ===
                sum_sq        df           F  PR(>F)
C(Race)     804.190556       4.0  839.239989     0.0
Residual  75520.311096  315247.0         NaN     NaN

=== ANOVA: AlcoholDrinking by Race ===
                sum_sq        df           F        PR(>F)
C(Race)      26.130187       4.0  102.933706  9.255117e-88
Residual  20006.718921  315247.0         NaN           NaN

=== ANOVA: Diabetic_Binary by Race ===
                sum_sq        df           F         PR(>F)
C(Race)     110.107759       4.0  250.926692  1.242052e-215
Residual  34582.949795  315247.0         NaN            NaN
```

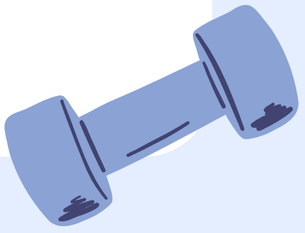
```
=== ANOVA: BMI by Race ===
                sum_sq        df            F  PR(>F)
C(Race)   1.619253e+05       4.0  1022.234282     0.0
Residual  1.248404e+07  315247.0          NaN     NaN

=== ANOVA: PhysicalHealth by Race ===
                sum_sq        df           F         PR(>F)
C(Race)   3.886656e+04       4.0  158.135827  1.919319e-135
Residual  1.937032e+07  315247.0         NaN            NaN

=== ANOVA: MentalHealth by Race ===
                sum_sq        df           F         PR(>F)
C(Race)   3.493483e+04       4.0  140.871157  1.587486e-120
Residual  1.954463e+07  315247.0         NaN            NaN
```

모든 변수에서 H0 기각, H1 채택

=> 인종 별로 생활습관 관련 특성이 유의하게 다르다



가설검정 2 인종에 따라 심장병 발생에 유의한 차이가 있다?

인종별 특징

1. Asian:
- 매우 낮은 흡연율, 낮은 BMI, 가장 좋은 Physical·Mental Health
- → 심장병 위험 요인이 전반적으로 적은 집단

2. American Indian/Alaskan Native (AI/AN):
- 가장 높은 흡연율, 높은 BMI, 가장 나쁜 정신·신체 건강
- 당뇨 비율도 전체 중 가장 높음
- → 다수 위험 요인이 중첩된 고위험 집단

3. Black/Other:
- 흡연·BMI·당뇨 모두 중간 이상, 음주율 낮음
- 정신·신체 건강 지표는 AI/AN보다는 낫지만 평균보다는 나쁨 → 잠재적 고위험군

4. Hispanic:
- 흡연율 낮고, BMI·당뇨·정신건강 지표 모두 평균 수준
- 전반적으로 위험 요인은 적당한 수준 → 중간 위험군

5. White:
- 흡연율 높음, 음주율 가장 높음
- BMI는 중간 수준, 당뇨는 가장 낮은 편
- → 생활습관 요인(흡연·음주)은 나쁘지만, 당뇨 측면에선 위험이 적은 이질적인 패턴

결론

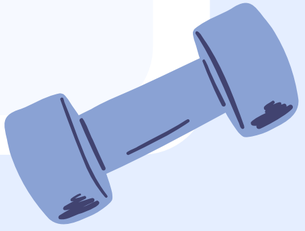
인종 간 생활습관 특성 차이 확인

→ 유전적 원인보다는 인종별

생활습관/패턴 요인 가능성

→ 단일 변수 수준에서 설명에 한계 존재

⇒ **복합 요인 탐색 필요**



“정확도와 해석력을 동시에 갖춘 심장병 예측 모델 구축”

단순한 예측 정확도 향상을 넘어,
→ **의미 있는 변수 해석**이 가능한 모델을 지향

모든 변수를 그대로 투입할 경우
예측력은 향상될 수 있으나 **개별 변수의 영향력 해석이 어려움**

차원을 줄이고 **핵심 정보를 유지**하는 **파생 변수 설계** 전략을 수립

변수 생성

단일 로지스틱 회귀

H_0 (귀무가설): 각 개별 변수의 계수(β)가 0이다. → "단일 변수만으로는 심장병 발병에 유의한 영향을 미치지 않는다."

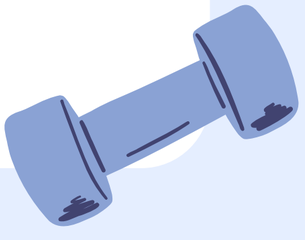
H_1 (대립가설): 특정 변수의 계수(β)가 0이 아니다. → "예를 들어, Smoking_Yes는 심장병 발병 확률을 높인다(β > 0), 반면 어떤 변수는 오히려 낮출 수도 있다

심장병 발병확률을 높이는 변수

['AgeCategory', 'DiffWalking_Yes', 'Diabetic_Yes', 'Stroke_Yes', 'PhysicalHealth', 'KidneyDisease_Yes', 'Smoking_Yes', 'SkinCancer_Yes', 'Sex_Male', 'BMI_Category', 'BMI', 'Race_White', 'Asthma_Yes', 'Diabetic_No, borderline diabetes', 'MentalHealth']

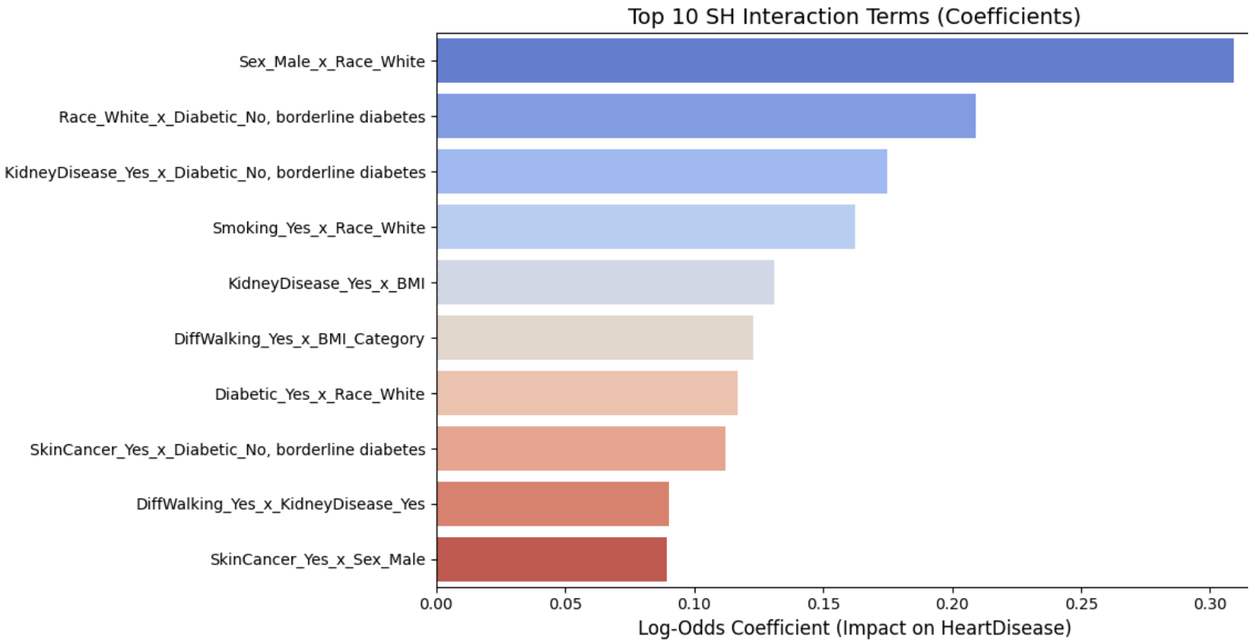
심장병 발병확률을 낮추는 변수

['GenHealth', 'PhysicalActivity_Yes', 'Race_Hispanic', 'AlcoholDrinking_yes', 'Race_Asian', 'Race_Black', 'SleepTime', 'Race_other']



변수 생성

다중 로지스틱 회귀 – 심장병 발병률을 올리는 조합

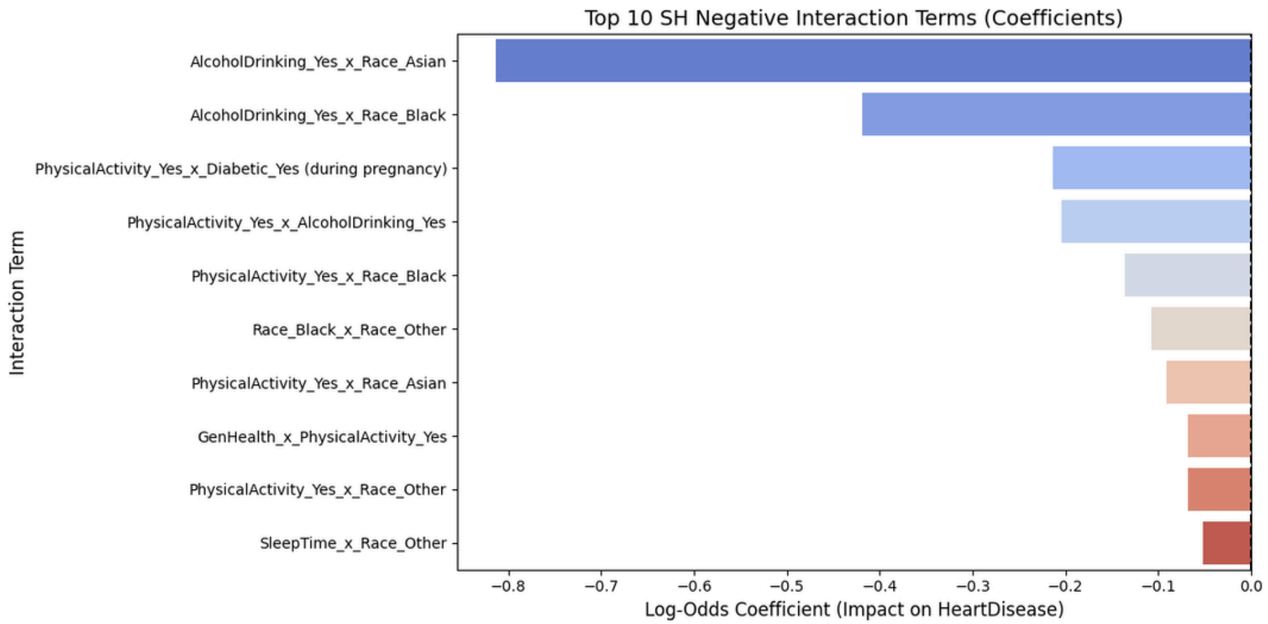


유의한 조합 필터링 및 확률 변환

조합	의미	Odds 증가율	확률 변화
Sex_Male x Race_White	백인 남성	+36.2%	+5.41%p
Smoking_Yes x Race_White	백인 흡연자	+17.6%	+2.73%p
DiffWalking_Yes x BMI_Category	보행 어려움 + 높은 BMI	+13.1%	+2.04%p
Diabetic_Yes x Race_White	백인 당뇨병환자	+12.4%	+1.93%p

변수 생성

다중 로지스틱 회귀 – 심장병 발병률을 낮추는 조합



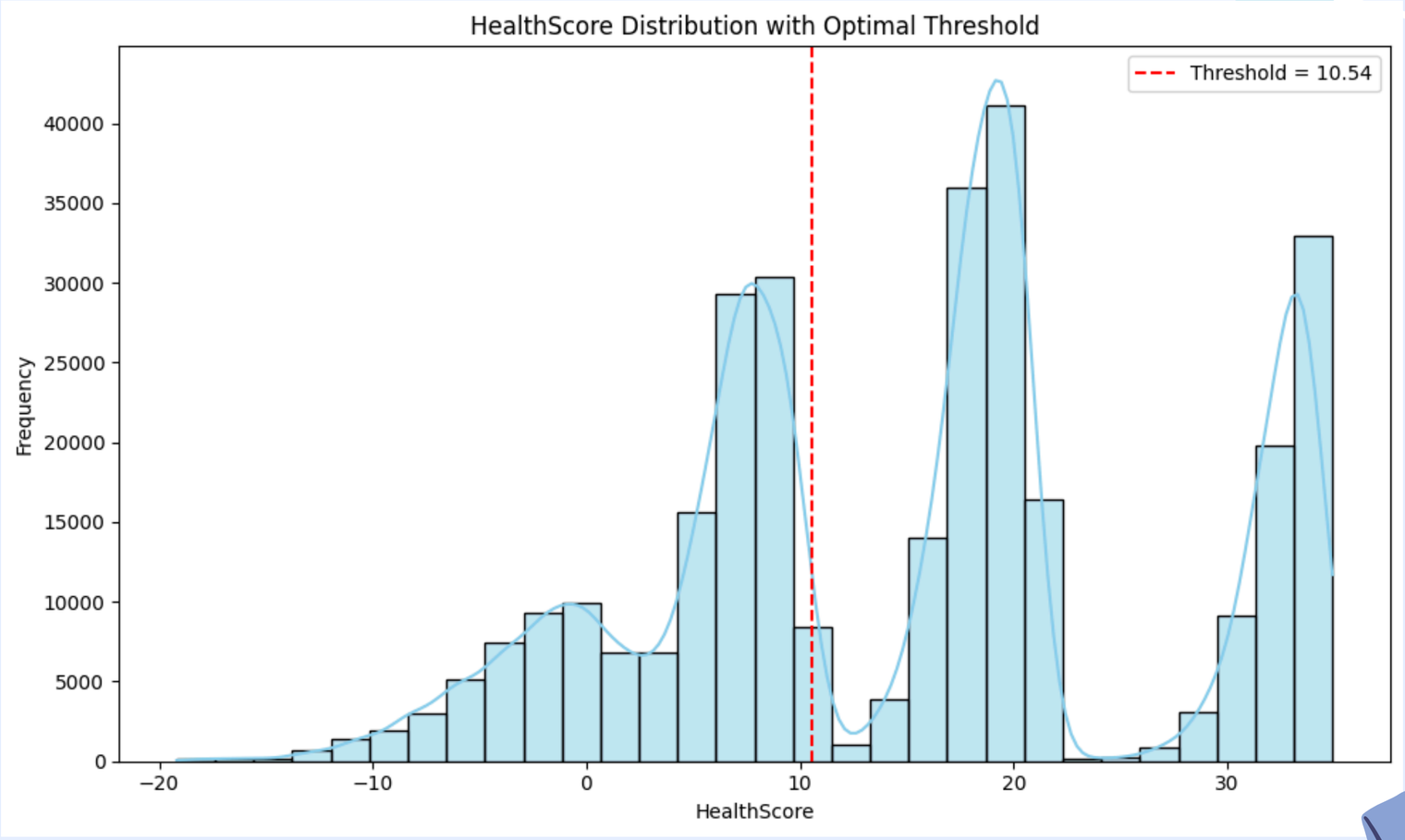
유의한 조합 필터링 및 확률 변환

조합	의미	Odds 증가율	확률 변화
PhysicalActivity_Yes x AlcoholDrinking_Yes	음주하더라도 활동적인 경우	−18.5%	−3.07%p
PhysicalActivity_Yes x Race_Black	흑인 + 활동적	−12.7%	−2.09%p
GenHealth x PhysicalActivity_Yes	건강 상태가 좋고 활동적인 경우	−6.6%	−1.07%p

변수 생성

종합 건강 점수 (HealthScore)

BMI, PhysicalHealth, MentalHealth, SleepTime, GenHealth를 기반으로
건강 전반을 반영한 종합 스코어
(p.35 참고)



건강위험여부 (HealthFlag)

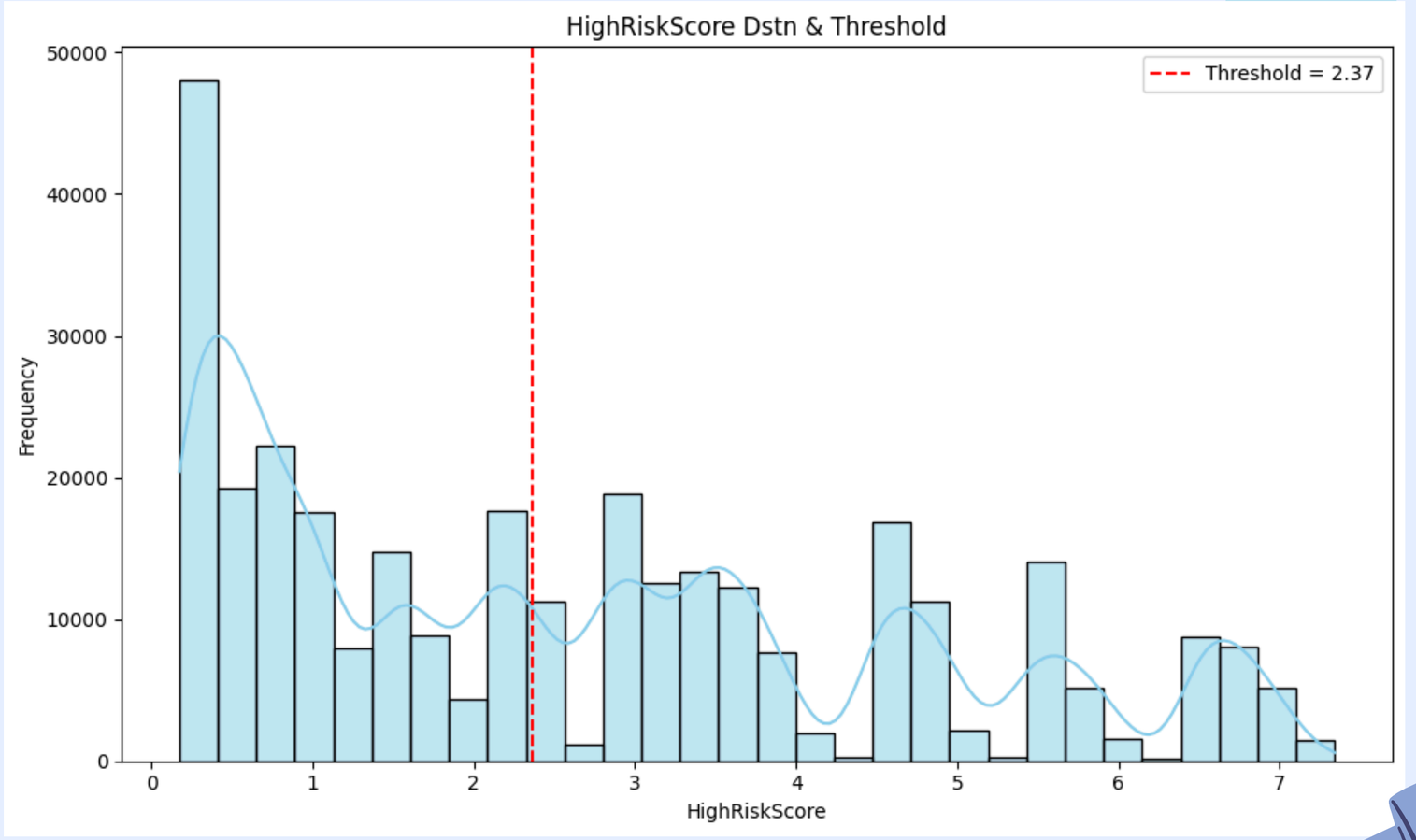
HealthFlag
TPR(재현율) – FPR(위양성률)을
최대화하는 최적 임계값을 설정.

최적 임계값 이하인 경우 위험 판정
건강 상태가 나쁨 → 심장병 위험군으로
HealthFlag 기준 1=위험, 0=비위험

변수 생성

위험요인점수 (HighRiskScore)

심장병 발병을 올리는 위험 변수 조합 + 연령에 따른 위험도(AgeScore)를 고려한 **연속형 위험 점수**
상호작용효과(Odds Ratio)를 초기 가중치로 설정(p.36 참고)



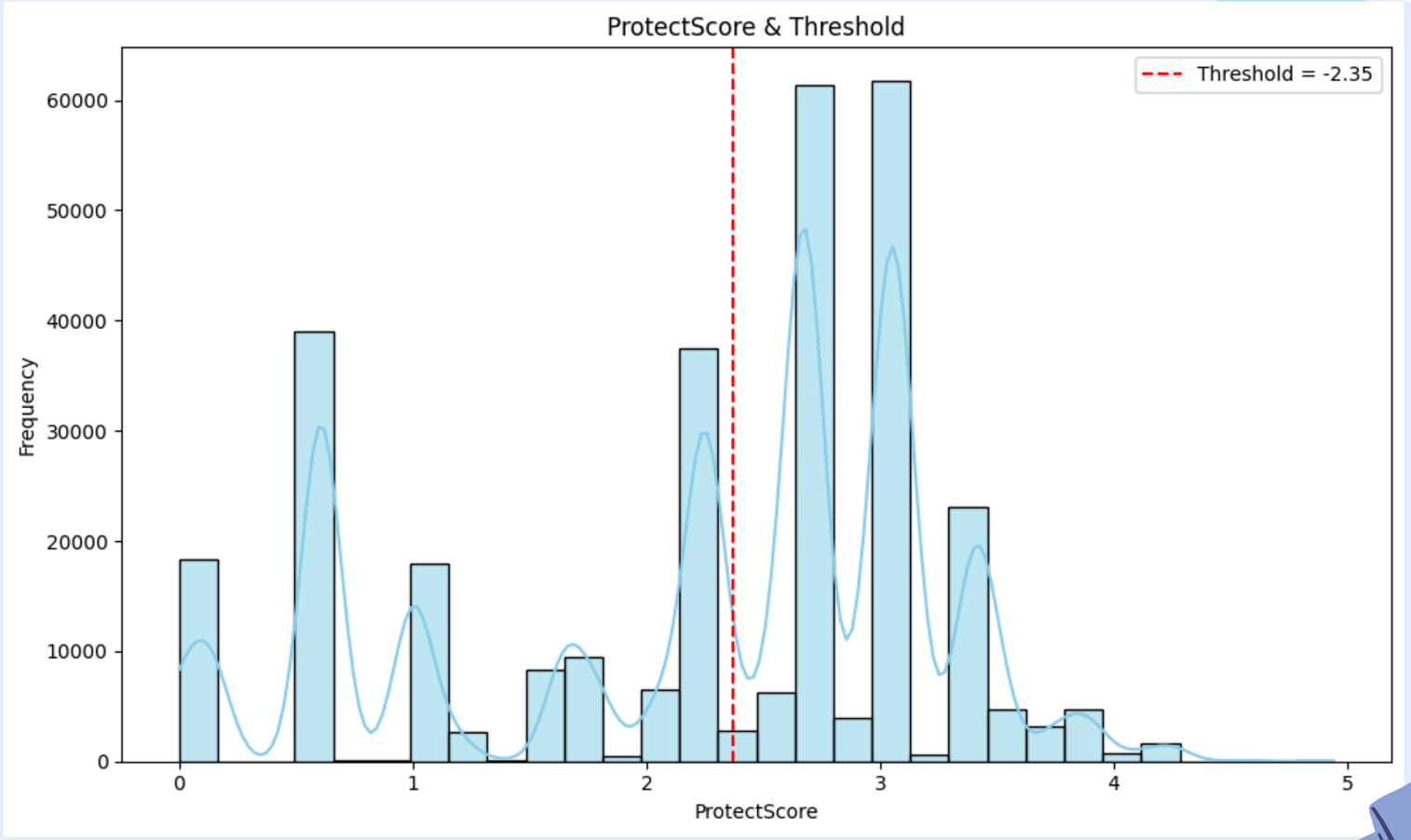
고위험군여부 (HighRiskFlag)

최적 임계값 이하인 경우 위험 판정
구조적 위험 요인(질병 이력 + 사회 인구학적 요인)이 높아 심장병 발생 가능성이 큼
점수 기준 1=위험, 0=비위험

변수 생성

보호점수 (ProtectScore)

심장병 발생 위험을 낮추는 조합(보호 요인) + 건강한 수면 습관 (SleepTime ≈ 7) 을 활용해
심장병 보호 효과를 반영한 점수
(p.37 참고)



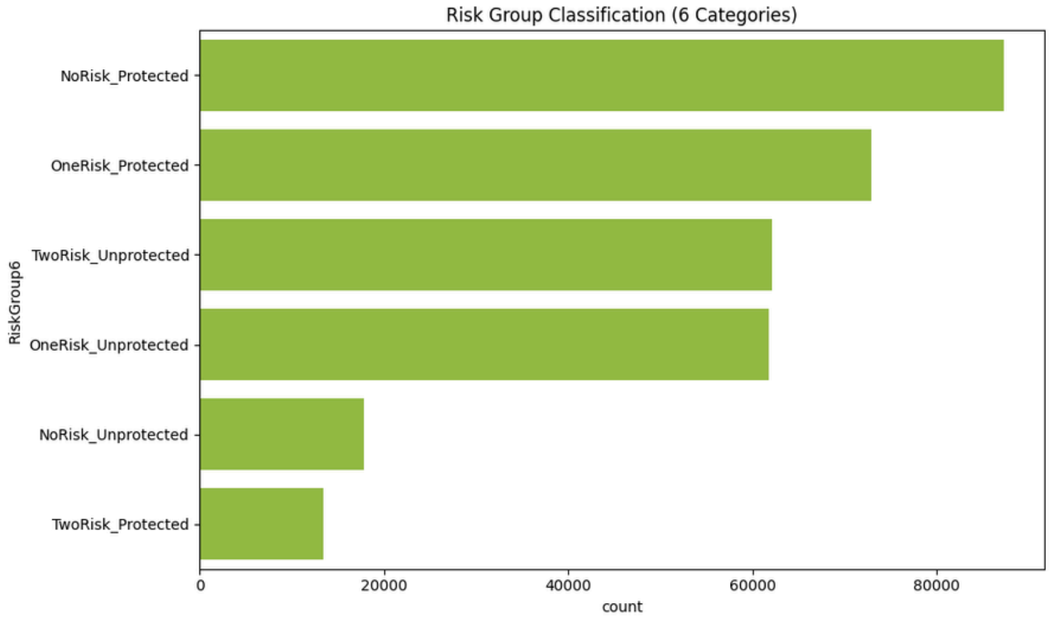
보호취약여부 (LowProtectFlag)

최적 임계값 이하인 경우 위험 판정
보호 요인이 부족한 상태인지 여부
보호요인부족=1, 충분=0

변수 생성

6단계위험군분류 (RiskGroup6)

위험 요인 2가지 (Health_Flag, HighRiskFlag)와 보호 요인 (LowProtectFlag)의 조합에 따라 총 6가지 위험 그룹으로 분류한 변수.
이후 실수 인코딩(p.38 참고)



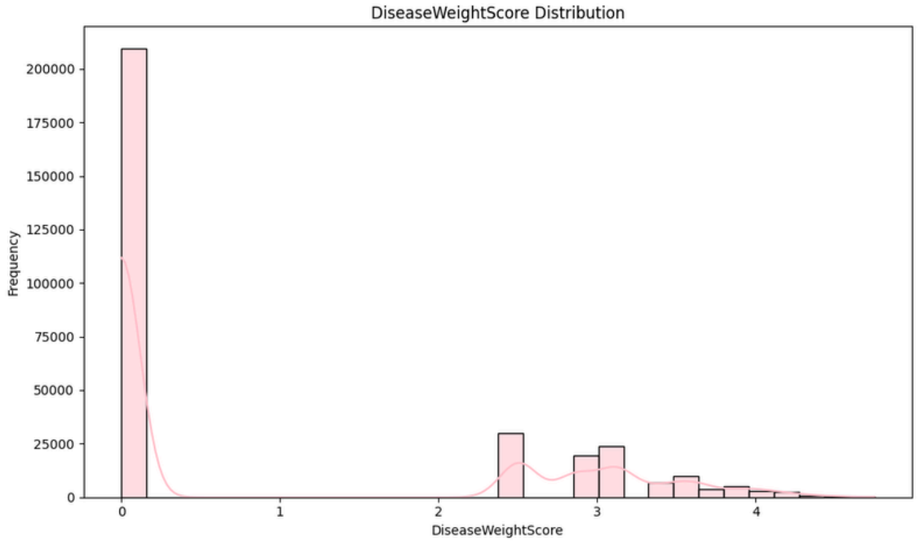
그룹 이름	발병률	해석 요약
TwoRisk_Unprotected	23.5%	가장 위험한 그룹 – 위험 요인 2개, 보호도 없음
TwoRisk_Protected	15.3%	위험은 많지만 보호 요인이 있는 경우 – 완충 효과
OneRisk_Unprotected	7.0%	위험 요인 1개, 보호 없음 – 중간 수준의 위험
OneRisk_Protected	6.1%	위험은 있지만 보호 요인도 있는 균형 상태
NoRisk_Unprotected	1.7%	위험 요인은 없지만 보호도 없음 – 잠재적 노출군
NoRisk_Protected	1.0%	가장 안전한 그룹 – 위험 요인도 없고 보호 있음

변수 생성

질병요약점수 (DiseaseWeightScore)

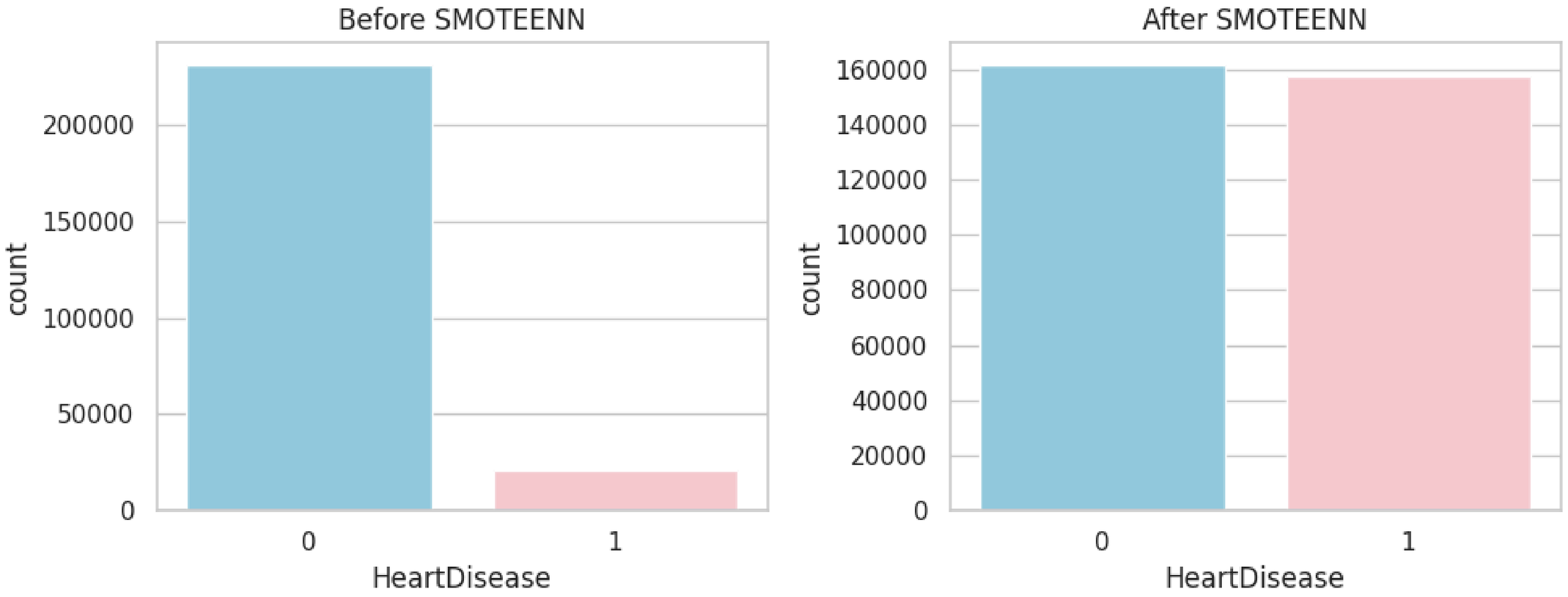
질병 병력이 존재할 때, 그 위험성을 하나의 수치형 변수로 요약한 지표

질병 (Stroke, Diabetic, Asthma, KidneyDisease, SkinCancer)이
있을 때 심장병 발병률(가중치)의 합으로 총 위험 점수 산출(p.38 참고)




```
질병별 심장병 발생률 (가중치)
Stroke_Yes: 0.35820251854407453
Diabetic_Yes: 0.21735953640715547
Asthma_Yes: 0.11259160559626916
KidneyDisease_Yes: 0.29051799824407376
SkinCancer_Yes: 0.16532299214205531
```

예측 모델 개발



Standard Scaler 적용 -> SMOTEENN

예측 모델 개발

Logistic Regression

Recall: 0.7756

Precision: 0.2079

F1-score: 0.3279

AUC: 0.7516

Random Forest

Recall : 0.5914

Precision: 0.2321

F1-score : 0.3334

AUC : 0.7055

XGBoost

Recall : 0.722

Precision: 0.2305

F1-score : 0.3495

AUC : 0.75

LightGBM

Recall : 0.7314

Precision: 0.2368

F1-score : 0.3578

AUC : 0.7571

Neural Network

Recall : 0.7459

Precision: 0.2206

F1-score : 0.3405

AUC : 0.8311

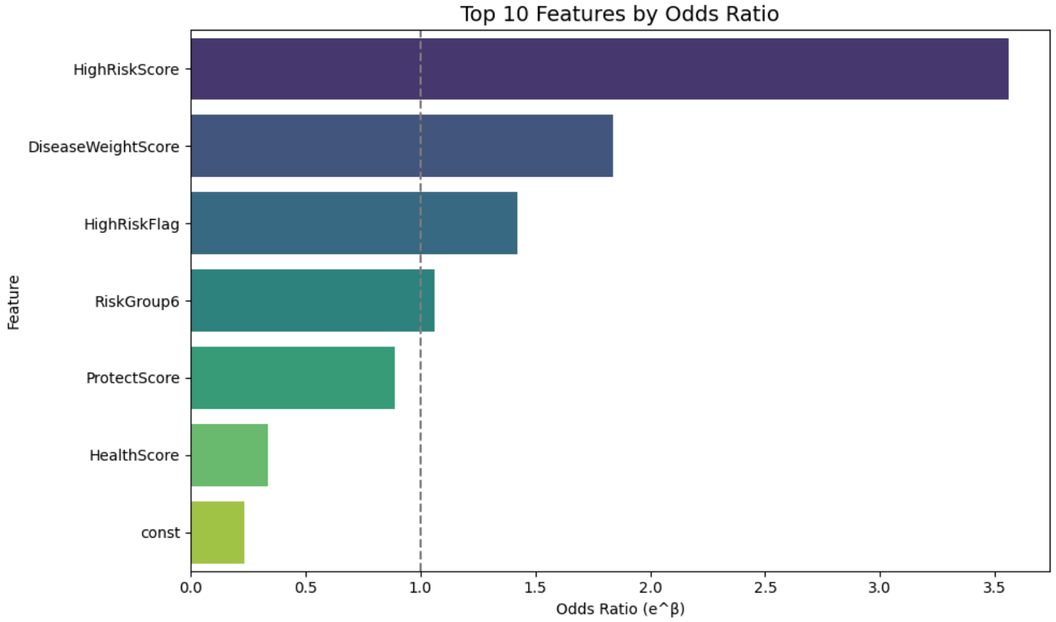
결론

건강 위험 여부(Health Flag), 보호 취약여부(LowProtect Flag)를 제외하고 모두 유의함을 확인

Logit Regression Results

Dep. Variable:	HeartDisease	No. Observations:	319078
Model:	Logit	Df Residuals:	319069
Method:	MLE	Df Model:	8
Date:	Tue, 01 Apr 2025	Pseudo R-squ.:	0.5038
Time:	08:35:58	Log-Likelihood:	-1.0973e+05
converged:	True	LL-Null:	-2.2114e+05
Covariance Type:	nonrobust	LLR p-value:	0.000

	coef	std err	z	P>|z|	[0.025	0.975]
const	-1.4593	0.029	-50.683	0.000	-1.516	-1.403
HighRiskScore	1.2697	0.010	127.868	0.000	1.250	1.289
Health_Flag	-0.0220	0.021	-1.023	0.306	-0.064	0.020
ProtectScore	-0.1177	0.009	-12.412	0.000	-0.136	-0.099
LowProtectFlag	-0.0178	0.021	-0.854	0.393	-0.059	0.023
HealthScore	-1.0944	0.012	-88.923	0.000	-1.119	-1.070
HighRiskFlag	0.3530	0.019	18.685	0.000	0.316	0.390
DiseaseWeightScore	0.6078	0.005	122.180	0.000	0.598	0.618
RiskGroup6	0.0617	0.010	6.280	0.000	0.042	0.081



결론

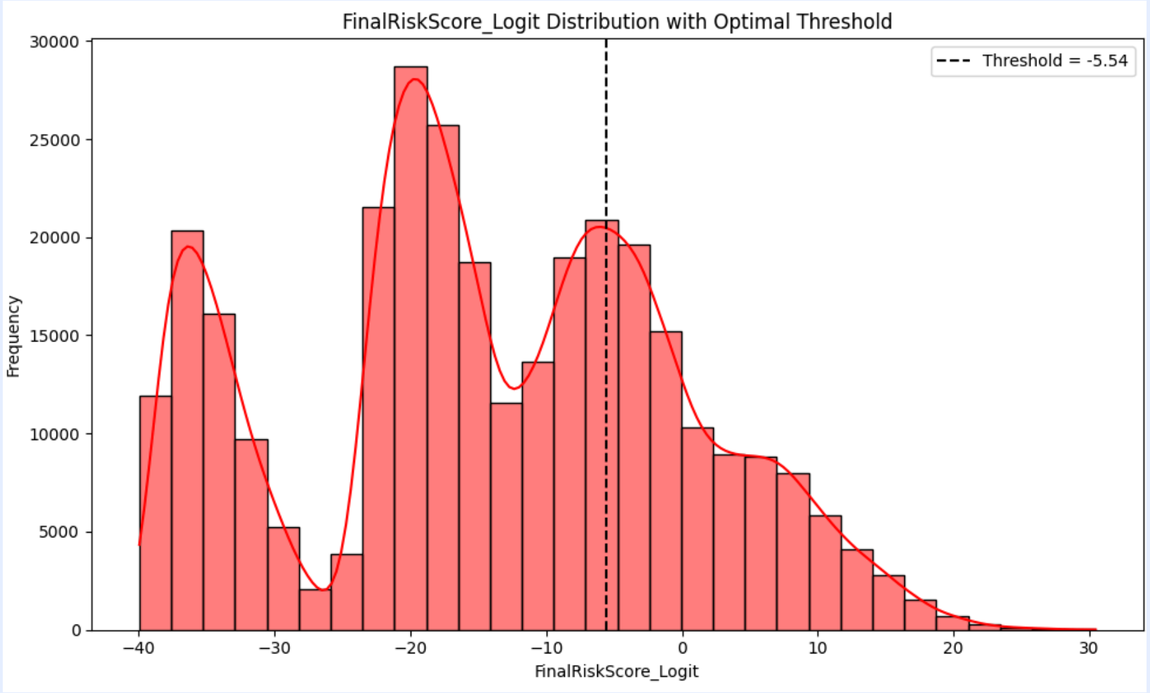
FinalRiskScore

(1.2717×위험요인점수) + (0.6089×질병요약변수) + (0.4031×고위험군여부)
– (0.0495×6단계 위험군 분류) – (0.1029×보호점수) – (1.1077×건강요약점수) – 1.2421(절편)

최종 위험 여부 (FinalRiskFlag)

최적 임계값 이하인 경우
심장병 위험 판정
(심장병 위험 =1, 심장병 안전=0)

Recall = 0.69로, 이 기준을 통해
간단하게 심장 질환 여부 확인 가능



Thank You

Appendix

변수 계산식

종합건강변수(HealthScore)

위험요인점수(HighRiskScore)

보호점수(ProtectScore)

6단계위험군분류(RiskGroup6)

질병요약점수(DiseaseWeightScore)

종합 건강 변수 계산식

HealthScore

{− log(BMI + 1) × (1 + 0.5 if BMI > 30)

− (PhysicalHealth)^(1/2) × (1 + 0.2 if PhysicalHealth > 15)

− (MentalHealth)^(1/2) × (1 + 0.2 if MentalHealth > 15)

− |SleepTime − 7|^1.5 × (1 + 0.3 if |SleepTime − 7| > 3)

+ (GenHealth + 1)^2 × 1.5}

항목	적용 조건 / 변환 방식	해석
BMI	BMI가 30 초과인 경우, 추가 감점 가중치 적용	고도비만일수록 심장병 위험이 높아지므로 감점을 크게 부여
PhysicalHealth	신체 건강 이상 일수 > 15일일 경우, 추가 감점 가중치 적용	장기적인 신체 이상 상태는 건강 악화를 의미하므로 감점 증가
MentalHealth	정신 건강 이상 일수 > 15일일 경우, 추가 감점 가중치 적용	지속적인 정신적 불안정 상태 역시 위험 요인으로 간주하여 감점
SleepTime	7시간에서 ±3시간 이상 벗어난 경우, 추가 감점 가중치 적용. 편차는 제곱승으로 반영됨	수면 시간은 7시간이 가장 이상적이며, 벗어날수록 건강에 악영향을 미치므로 감점. 특히 편차가 클수록 급격한 감점 적용
GenHealth	주관적 건강 상태 점수가 높을수록 가산점 부여	스스로 건강하다고 인식하는 상태는 실제 건강과도 연관이 있어 긍정적으로 반영

위험 요인 점수 계산식

HighRiskScore

$$w1 \cdot R1 + w2 \cdot R2 + w3 \cdot R3 + w4 \cdot R4 + w5 \cdot AgeScore + b$$

- R1 : 백인 남성 여부
- R2 : 백인 흡연자 여부
- R3 : 고도비만(3단계 이상) + 보행장애 여부
- R4 : 백인 당뇨병 여부
- AgeScore :연령 기반 심장병 위험 가중치
- b : 절편

보호 점수 계산식

ProtectScore

$$w1 \cdot \log(1+P1) + w2 \cdot \log(1+P2) + w3 \cdot \log(1+P3) + w4 \cdot \exp\{-(SleepTime-7)^2/2\sigma^2\} + b$$

항목	적용 조건 / 변환 방식	해석
P1	신체활동 있음 × 음주 있음 → 조합 변수 생성→ log(1 + P1)로 변환	음주가 있더라도 활동적인 경우 심장병 위험이 낮아지는 보호 효과 존재
P2	신체활동 있음 × 흑인 여부 → 조합 변수 생성→ log(1 + P2)로 변환	흑인이면서 활동적인 경우, 다른 인종 대비 보호적 상호작용 관찰됨
P3	주관적 건강 상태 × 신체활동 있음 → 조합 변수 생성→ log(1 + P3)로 변환	건강 상태가 좋고 활동적인 경우, 위험을 낮추는 조합 효과 발생
SleepTime	평균 수면 시간인 7시간을 기준으로 정규분포(가우시안 함수) 형태 적용	7시간에서 멀어질수록 보호 효과가 줄어듦. 정상 수면 패턴은 건강 보호 요인으로 작용하며, 편차가 클수록 급격히 감소
Bias (b)	상수항	모델 기준선 조정 용도

6단계 위험군 분류 계산식

RiskGroup6

$$\log\{1+1000 \cdot (\text{i 그룹의 빈도} \times \text{i 그룹의 심장병 발병률})\}$$

질병 요약 점수 계산식

DiseaseWeightScore

$$\log[1 + \Sigma\{1(\text{i번째 질병 O} = 1) \times 100 \times \text{i번째 질병당 심장병 발병률}\}]$$

다중검정

각 검정에서의 p-value를 보정하여 다중검정으로 인한 오류 줄임

범주형-FDR 보정

Smoking (FDR 보정 후 p-value) : 0.00000
AlcoholDrinking (FDR 보정 후 p-value) : 0.00000
Stroke (FDR 보정 후 p-value) : 0.00000
DiffWalking (FDR 보정 후 p-value) : 0.00000
Sex (FDR 보정 후 p-value) : 0.00000
AgeCategory (FDR 보정 후 p-value) : 0.00000
Race (FDR 보정 후 p-value) : 0.00000
Diabetic (FDR 보정 후 p-value) : 0.00000
PhysicalActivity (FDR 보정 후 p-value) : 0.00000
GenHealth (FDR 보정 후 p-value) : 0.00000
Asthma (FDR 보정 후 p-value) : 0.00000
KidneyDisease (FDR 보정 후 p-value) : 0.00000
SkinCancer (FDR 보정 후 p-value) : 0.00000
BMI_Category (FDR 보정 후 p-value) : 0.00000

수치형-FDR 보정

PhysicalHealth
  Coefficient : 0.0519
  Raw p-value : 0.0000
  Adjusted p-value: 0.0000
  Odds Ratio : 1.0533
MentalHealth
  Coefficient : 0.0106
  Raw p-value : 0.0000
  Adjusted p-value: 0.0000
  Odds Ratio : 1.0107
SleepDeviation
  Coefficient : 0.2953
  Raw p-value : 0.0000
  Adjusted p-value: 0.0000
  Odds Ratio : 1.3435

불균형 보완

집단 간 표본 수 차이(불균형) 존재.
발생률 차이의 범위에 대한 해석적 안정성을 보완

2표본 비율 z검정 CI

Z-statistic: -18.2528
p-value: 0.0000

과음자 심장병 비율: 0.0510
비과음자 심장병 비율: 0.0868
비율 차이 (과음 - 비과음): -0.0358
95% 신뢰구간 (비율 차이)
  : (-0.0390, -0.0327)