

데이터분석 프로젝트 보고서

(주제: 당뇨병, 뇌졸중/심장질환, 심장질환 발생 시 나타나는 특징 및 이들 간의 연관성 분석)

이름: 김규산
날짜: 2024년 4월 8일

※ 목차

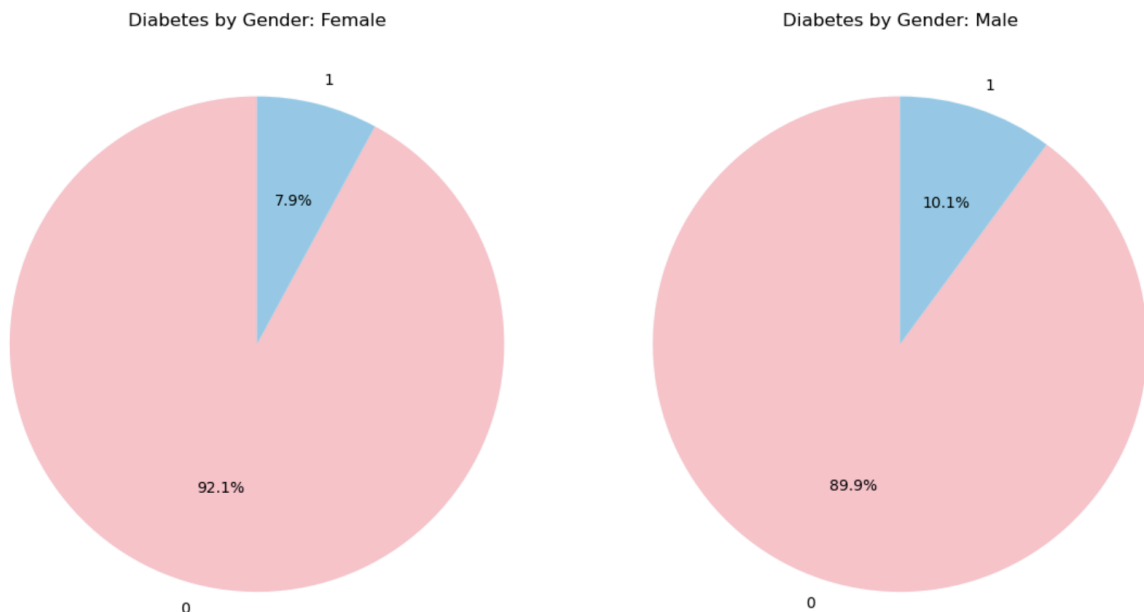
1. 분석의 목적
2. 당뇨 발생 유/무에 따른 시각화 그래프
 - 2-1. 당뇨 발생에 따른 요인들에 대한 결과 분석
3. 뇌졸중/심장질환 발생에 따른 시각화 그래프
 - 3-1. 뇌졸중/심장질환 발생에 따른 요인들에 대한 결과 분석
4. 심장질환 발생 유/무에 따른 시각화 그래프
 - 4-1. 심장질환 발생 유/무에 따른 요인들에 대한 결과 분석
5. 결론
6. Outro

1. 분석의 목적

가. 질병별 증상 발현 특징 및 유사성 조사: 각 질병이 발현될 때 나타나는 주요 특징을 조사하고, 이를 바탕으로 질병간 유사성을 분석합니다.

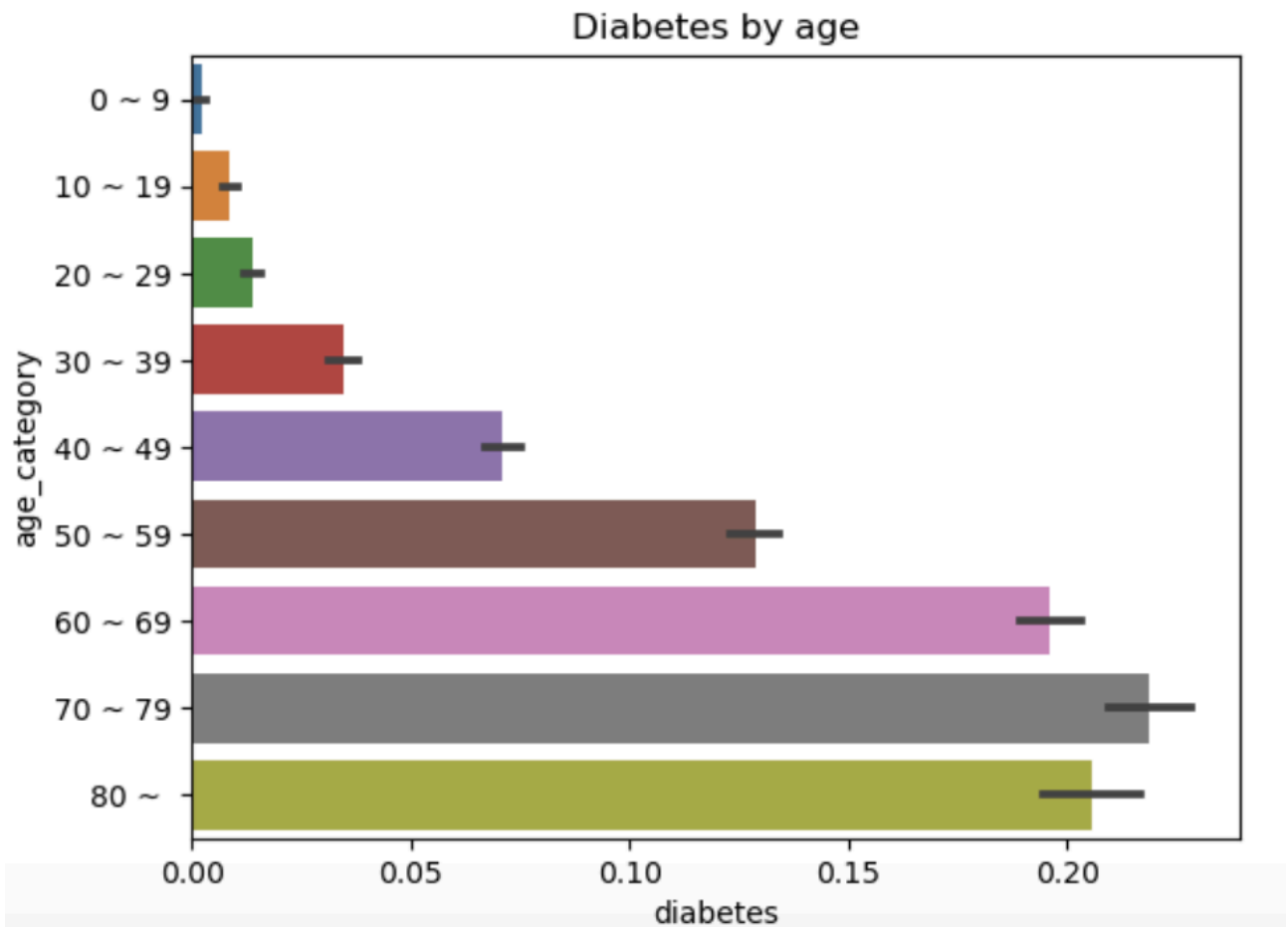
2. 당뇨 발생에 따른 시각화 그래프

가. 성별과 당뇨병 발병 관계



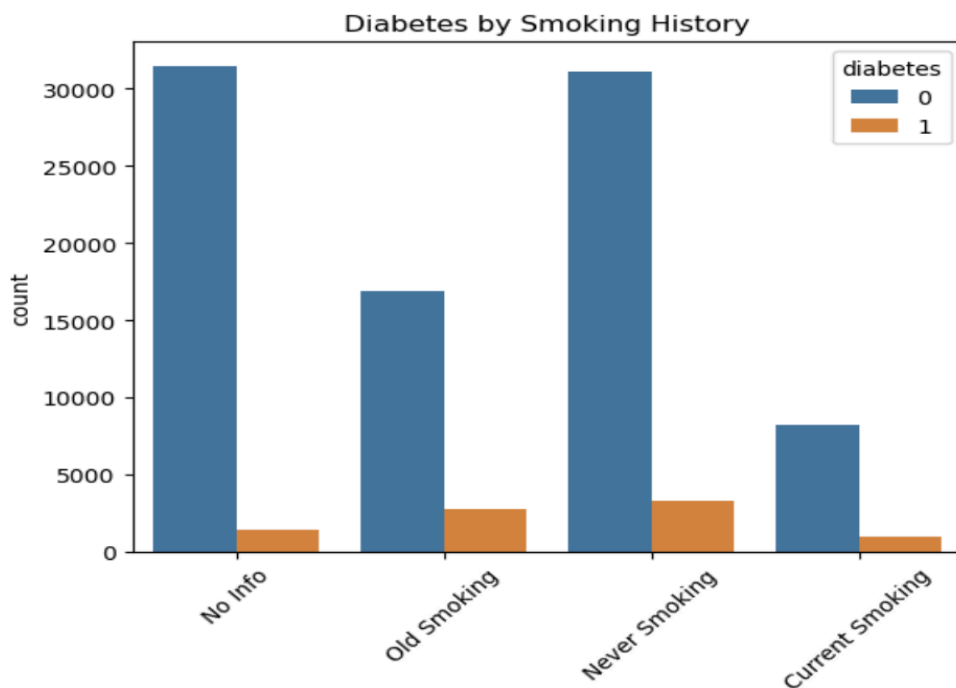
모집한 데이터 셋에서 성별에 따른 당뇨 발생 비율을 확인했을 때 여성보다 남성의 비율이 높은 것을 확인할 수 있다.

나. 연령대별 당뇨병 발병률



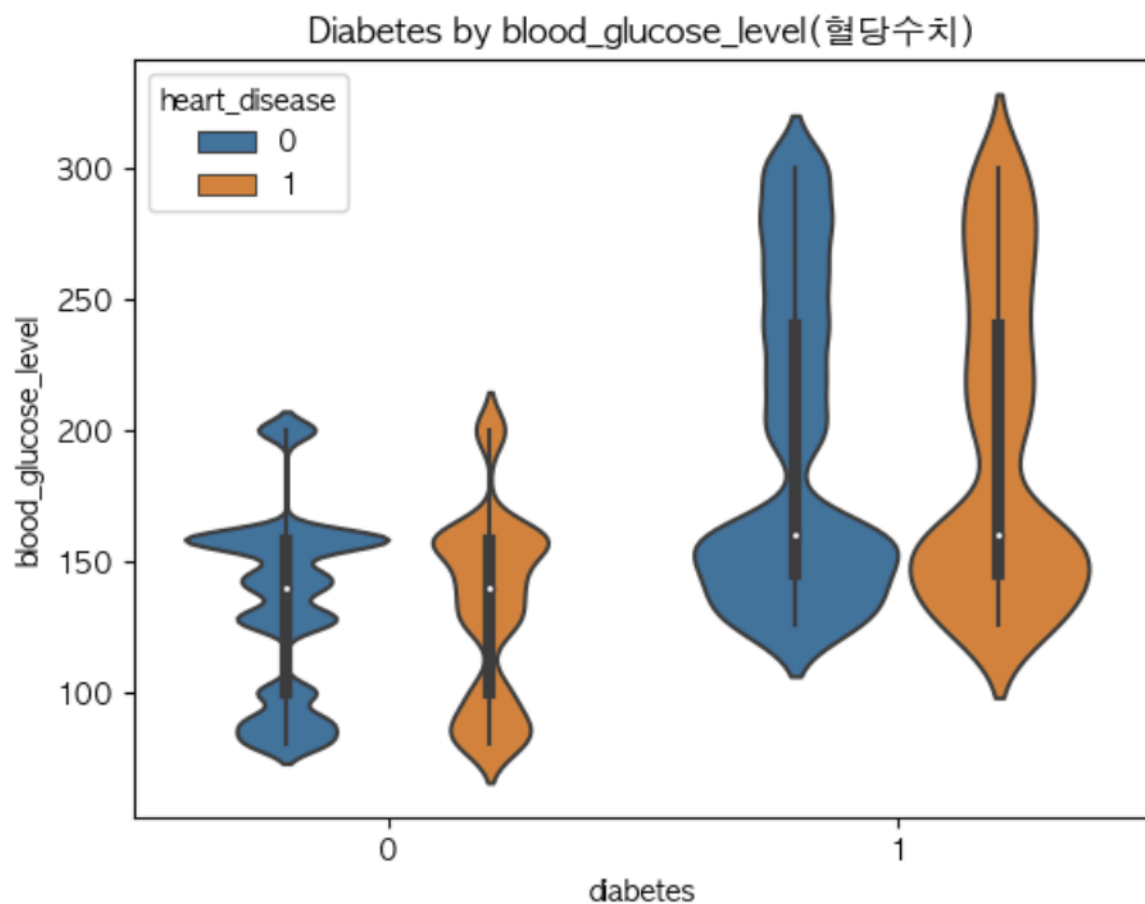
70대 연령대에서 당뇨병 발병률이 가장 높은 것으로 나타났으며, 이어 80세 이상, 그리고 60대 순으로 발병률이 낮아지는 경향을 보였습니다. 이러한 결과는 나이가 증가함에 따라 당뇨병에 걸릴 확률이 높아지지만, 80세 이상에서 발병률이 가장 높지 않은 것은 해당 연령대의 샘플 크기가 상대적으로 작기 때문일 수 있습니다. 이는 나이와 당뇨병 발생 간에 관련성이 있을 가능성을 시사합니다.

다. 흡연 상태와 당뇨병 발병의 상관성



위의 그래프에서는 한번도 담배를 비우지 않은 환자들과 현재 담배를 피고 있는 환자들끼리 당뇨 발생 여부에 큰 연관은 있는 것처럼 보이지 않으며 따라서 흡연과 당뇨 발생여부는 높은 상관 관계는 없다고 인지할 수 있습니다.

라. 혈당 수치와 당뇨병, 심장질환 발병 간의 분포

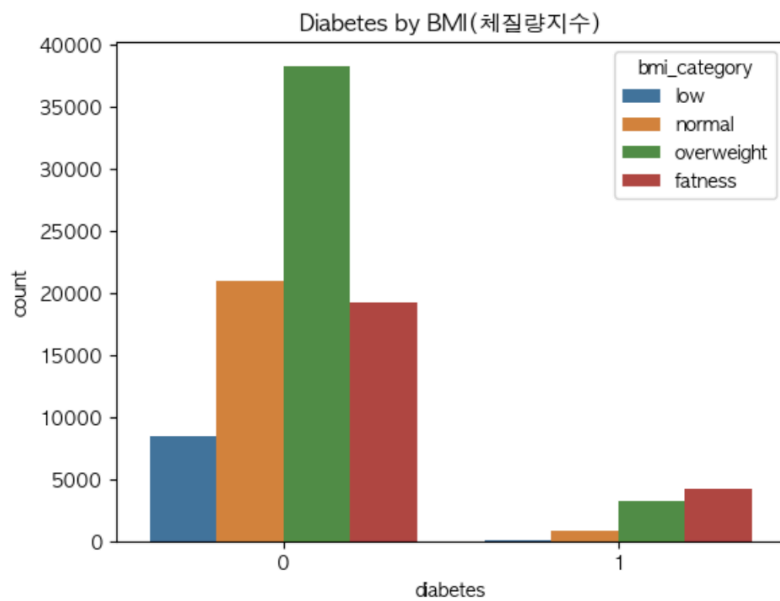


혈중 포도당 수치에서 정상치는 140mg/dL 이하의 값을 정상으로 간주합니다

위의 그래프에서 당뇨에 걸린 사람의 경우 걸리지 않은 사람보다 최대값이 높을 뿐더러 다수의 분포가 낮은 수치보다 높은 수치에 분포되어 있는 것을 확인할 수 있습니다.

더욱이 당뇨에 걸린 사람의 분포와 당뇨에 걸린 사람 중 심장 질환이 있는 사람의 그래프의 양상이 매우 유사하게 나타나고 있는 것을 확인할 수 있으며 이는 당뇨가 환자의 경우 심장질환이 있을 수 있는 확률이 높다는 사실을 나타낼 수 있습니다..

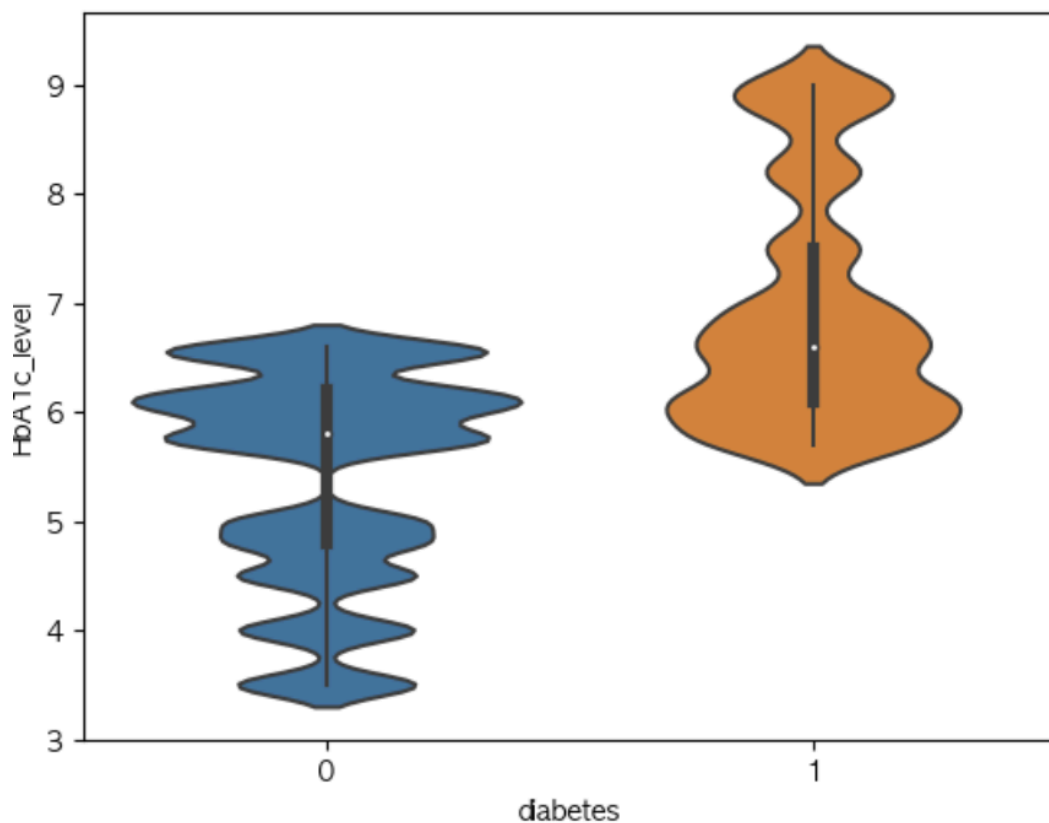
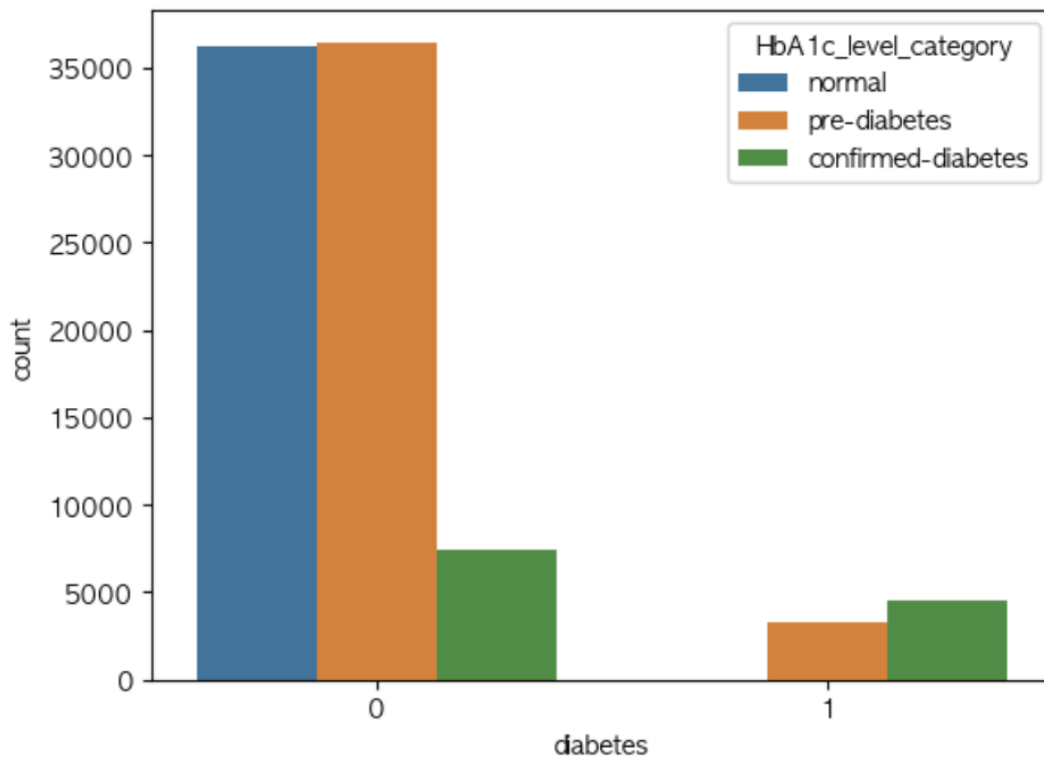
마. BMI 지수가 당뇨병 발병에 미치는 영향



bmi 범주를 기준으로 당뇨 발생 여부를 분석했을 때, 더 높은 bmi 범주에 속하는 환자들 사이에서 당뇨병 발생빈도가 더 많았습니다.

이러한 결과는 체질량지수(즉, 비만도)와 당뇨 발생 여부는 상당한 연관성이 있을 가능성을 나타냅니다.

바. 당화혈색소 수치에 따른 당뇨병 발병률 (평균 혈당 수준 반영)



정상적인 당화혈색소 수치를 가진 환자들에서는 당뇨병 발병 가능성이 낮지만, 수치가 높은 환자들은 당뇨병을 가질 확률이 있습니다. 특히, 당화혈색소 수치가 약 6.7%를 초과하는 경우에는 해당 환자들 당뇨병을 앓고 있음을 확인할 수 있습니다.

* 당화혈색소(HbA1c): 혈중 글루코스(포도당)가 헤모글로빈과 결합하여 형성된 혈액 구성요소

2-1. 당뇨 발생에 따른 요인들에 대한 결과 분석

가. 성별과 당뇨 발생 (!!)

- 데이터 분석 결과 남성이 여성에 비해 당뇨병에 걸린 비율이 높은 것으로 확인되었으며, 이는 성별에 따른 구분이 당뇨 발생률의 차이를 초래할 수 있음을 시사합니다.

나. 연령대별 당뇨 발생 (!!)

- 연령대가 증가할 수록 당뇨병에 걸린 비율이 높은 것으로 분석되었으며, 연령대가 증가함에 따라 당뇨병 발생확률도 증가하는 경향을 보였습니다. 이는 나이가 당뇨병 발생과 관련이 있음을 나타냅니다.

다. 흡연과 당뇨 발생 (!)

- 흡연 상태와 당뇨 발생간에는 높은 상관관계가 확인되지 않았으며, 흡연이 당뇨 발생에 직접적인 주요 요인은 아니며 상대적으로 적은 관계성을 나타냄을 시사합니다.

라. 혈중 포도당 수치와 당뇨발생 (!!!)

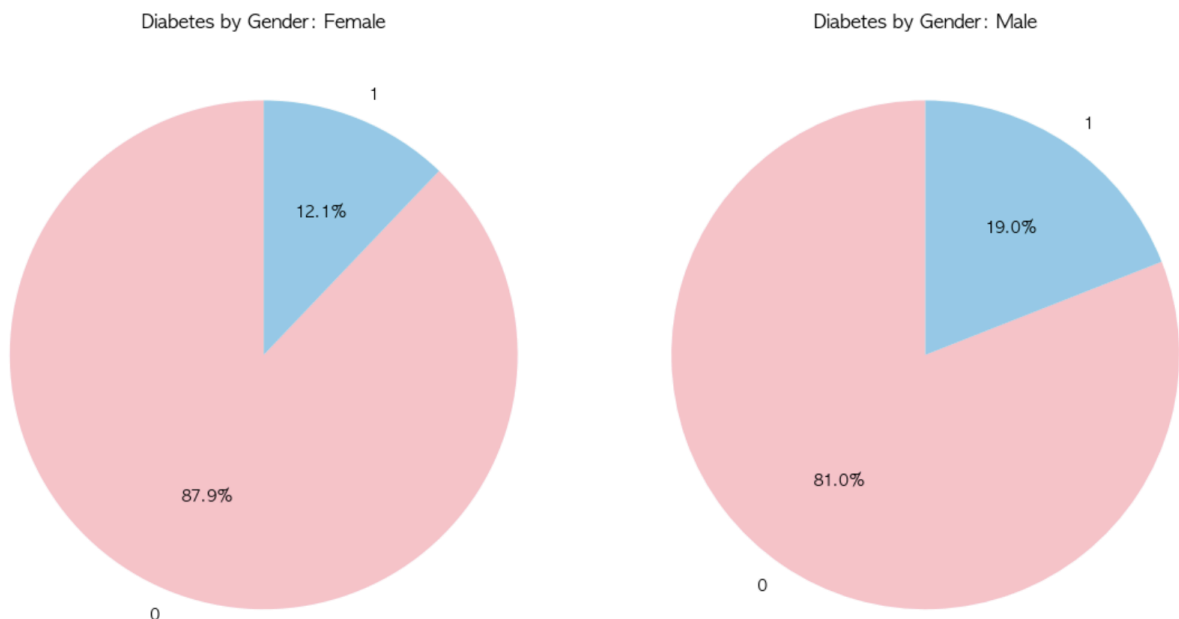
- 혈중 포도당 수치가 정상치를 초과하는 환자들 분포에서 당뇨병 발생 비율이 높았으며, 특히 당화혈색소가 6.7% 이상인 경우 당뇨병을 가지고 있음을 시사합니다. 따라서 혈당에 높은 비율이 당뇨병 발생과 직접적인 관련이 있음을 시사합니다.

마. BMI와 당뇨 발생 (!!!)

- 높은 체질량지수 범주에 속하는 개인들에게서 당뇨병 발생 빈도가 많았습니다. 이는 비만이 당뇨병 발생의 직접적인 요소임을 하나임을 시사합니다.
- 성별/연령대/ 포도당수치/ 체질량지수

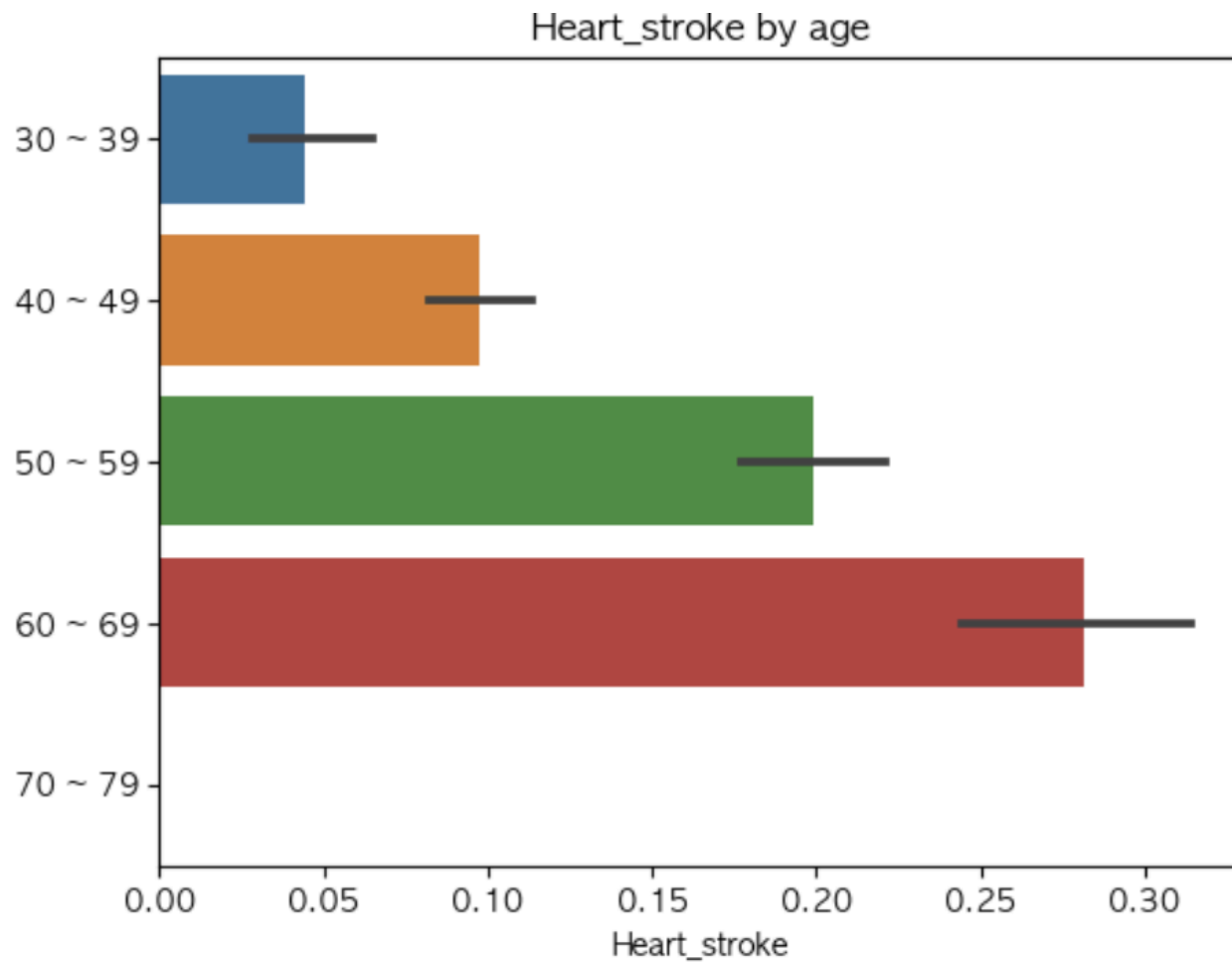
3. 뇌졸중/심장질환 발생에 따른 시각화 그래프

가. 성별과 뇌졸중/심장질환 발생 관계



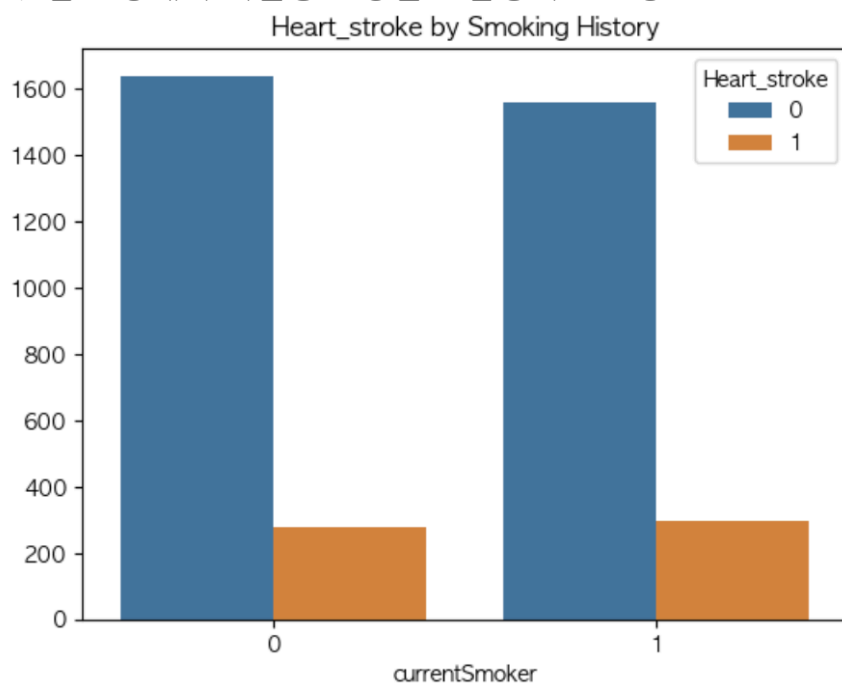
모집한 데이터 셋에서 성별에 따른 심장/뇌졸중 발생 비율을 확인했을 때 여상보다 남성의 비율이 높은 것을 확인할 수 있다.

나. 연령대별 뇌졸중/심장질환 발생 여부



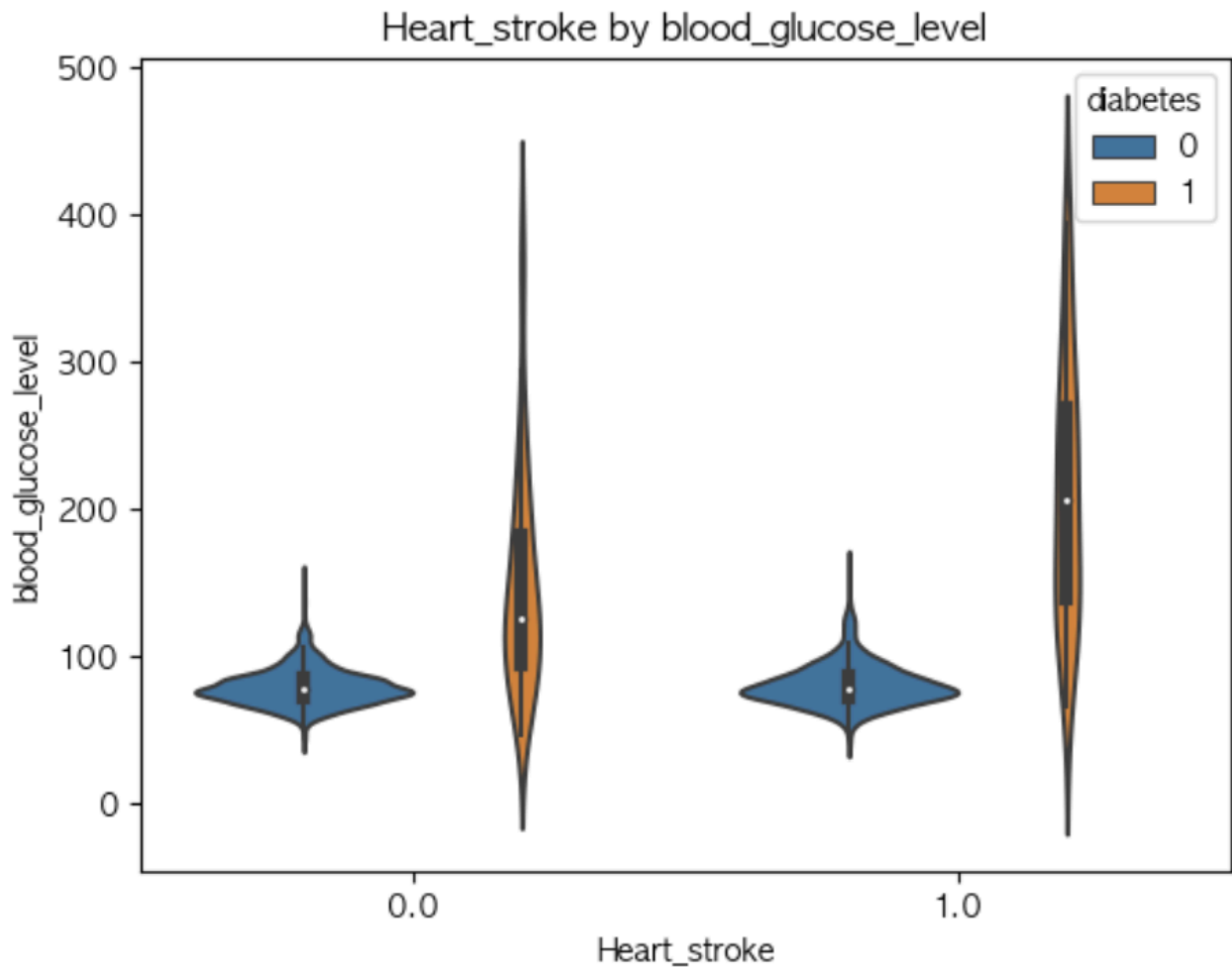
해당 시각화 그래프를 확인했을 때 나이가 증가할 수록 더 많은 수의 질환이 있다고 확인할 수 있습니다.
(70세 이상인 환자에 대해서는 표본의 수가 적으면서 해당 질환을 가진 환자가 없기 때문에 표현이 되지 않았습니다.)

다. 흡연 상태와 뇌졸중/심장질환 발생의 연관성



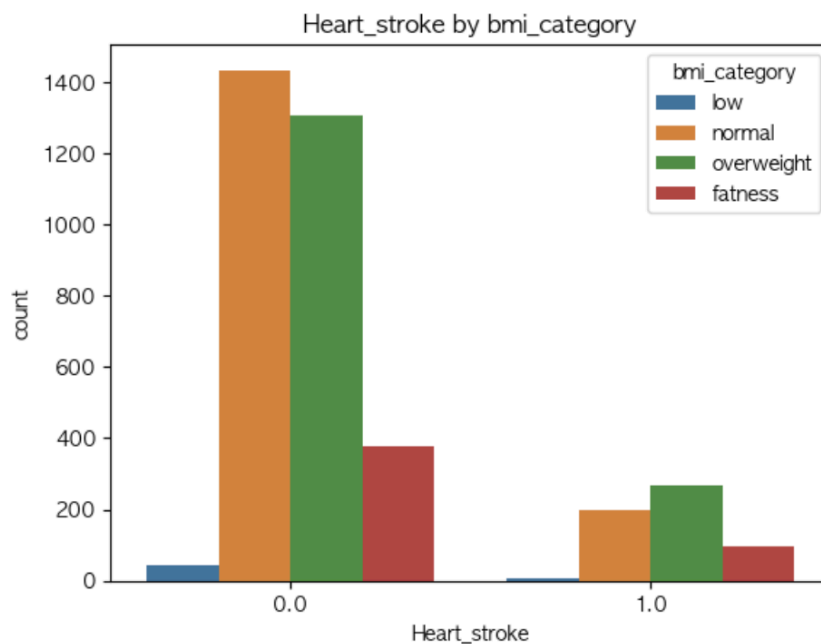
위의 그래프에서는 흡연 여부에 대한 심장/뇌졸중 질환 발생에 대하여 높은 연관성을 보이지 않으며, 따라서 흡연과 심장/뇌졸중 질환에 대하여 직접적인 관계는 없는 것으로 판단됩니다.

라. 혈당 수치와 뇌졸중/심장질환 발생 간의 관계



위의 그래프 분석 결과, 심장질환/뇌졸중을 가진 환자들 사이에서 혈중 포도당 수치와의 직접적인 연관성은 관찰되지 않았습니다. 또한, 당뇨병 발생과의 직접적인 관련도 없는 것으로 나타났습니다. 그러나 혈중 포도당 수치가 높아질수록 당뇨병 환자의 비율이 증가하는 경향은 확인되었습니다.

마. BMI 지수와 뇌졸중/심장질환 발생의 상관관계



현재 분석된 데이터를 통해 심장질환/뇌졸중 환자의 BMI 지수를 살펴보았을 때, 저체중 환자들에 비해 과체중과 비만 범주에 속하는 환자들에서 심장질환/뇌졸중 발생 빈도가 더 높은 것을 확인할 수 있습니다. 정상 체중에 비해 과체중이 더 높은 발생 빈도를 보이며, 비만 범주에 속하는 환자들 역시 모집단 대비 높은 빈도를 나타냅니다. 이는 BMI 지수가 심장질환/뇌졸중 발생과 관련이 있음을 시사합니다.

3-1. 뇌졸중/심장질환 발생에 따른 요인들에 대한 결과 분석

가. 성별에 따른 뇌졸중/심장질환 발생 (!!)

- 데이터 분석 결과 남성이 여성에 비해 뇌졸중과 심장질환 발생 비율이 높은 것으로 확인되었으며, 이는 성별에 따른 구분이 해당 질병 발생의 영향을 미칠 수 있음을 시사합니다.

나. 연령대별 뇌졸중/심장질환 발생(!!)

- 연령대가 증가할 수록 뇌졸중/심장질환의 발생 빈도가 증가되는 경향이 관련되었습니다. 연령대가 증가함에 따라 해당 질환의 발생 경향이 더욱 명확하게 나타나고 있으며, 이는 연령이 높을 수록 해당 질환의 발생빈도가 높아진다는 것을 시사합니다.

다. 흡연과 뇌졸중/심장질환 발생(!)

- 흡연 여부와 뇌졸중/심장질환 발생 사이에는 높은 연관성이 확인되지 않았습니다. 이는 흡연과 해당 질병 발생간의 직접적인 요인이 아닐 수 있음을 나타냅니다.

라. 혈중 포도당 수치와 뇌졸중/심장질환 발생(!!!)

- 심장질환/뇌졸중을 가진 환자들 사이에서 혈중 포도당 수치와 직접적인 연관성은 발견되지 않았습니다. 당뇨병 발생과도 직접적인 관련이 없는 것으로 나타났으나, 혈중 포도당 수치가 높아질수록 당뇨병 환자 비율이 증가하는 경향은 확인 되었습니다.

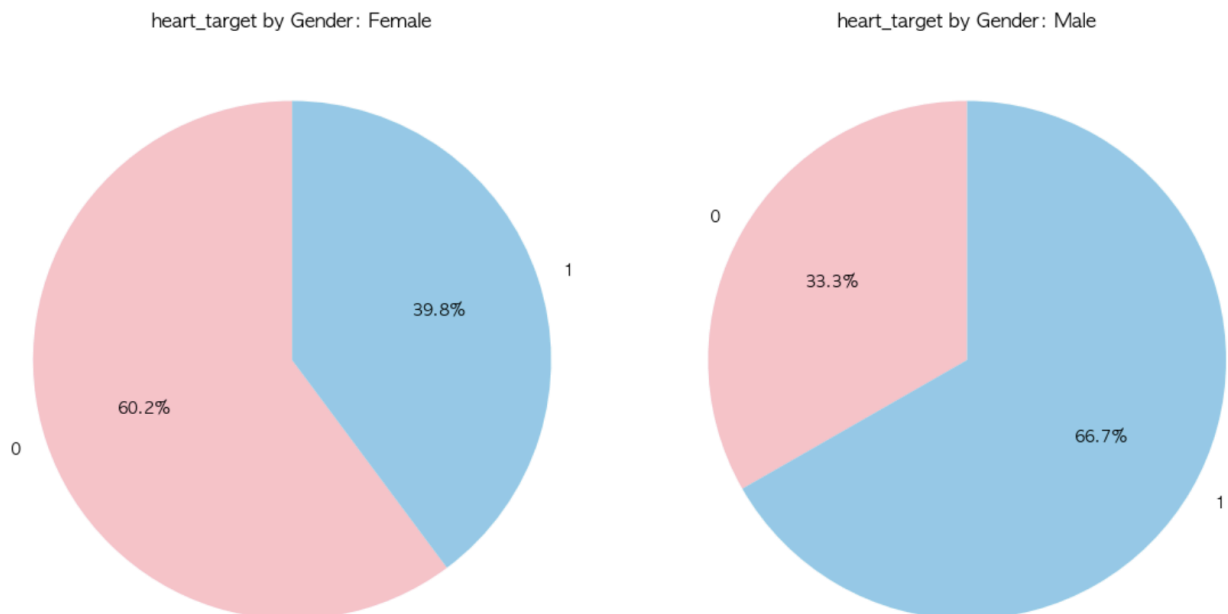
마. BMI와 당뇨 발생(!!!)

- 높은 체질량지수 범주에 속하는 개인들에게서 뇌졸중/심장질환 발생 빈도가 많았습니다. 이는 비만이 뇌졸중/심장질환 발생의 직접적인 요소임을 하나임을 시사합니다.

- 성별/연령대/ 체질량지수

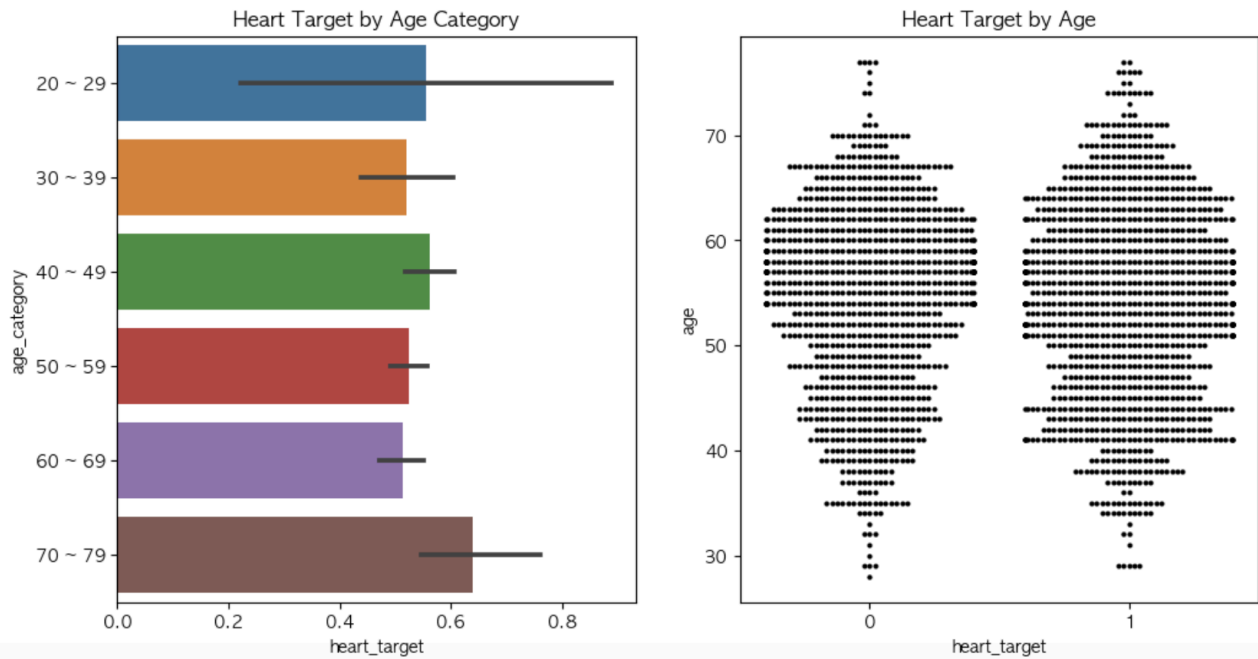
4. 심장질환 발생에 따른 시각화 그래프

1. 성별에 따른 심장질환 분포



모집한 데이터 셋에서 성별에 따른 심장질환 발생 비율을 확인했을 때 여성보다 남성의 비율이 높은 것을 확인할 수 있다.

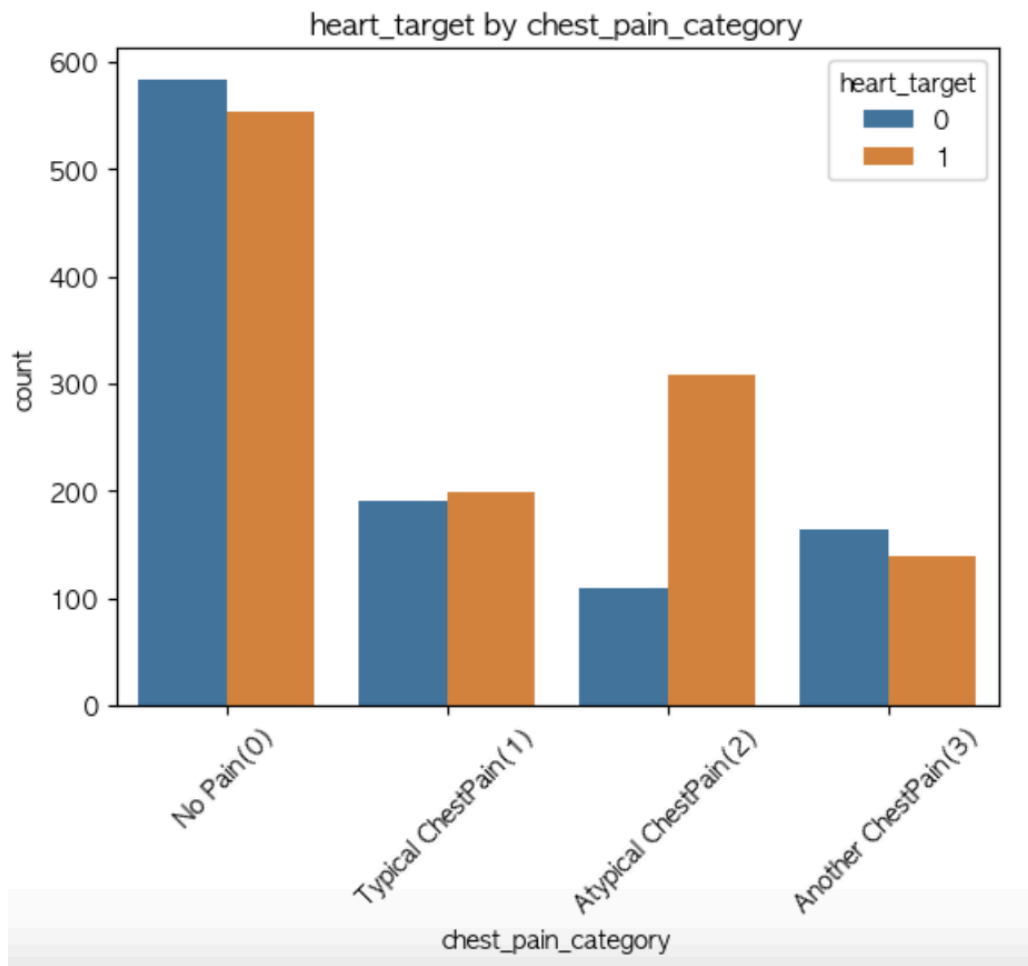
2. 나이 범주에 따른 심장 질환 유무



나이에 따른 심장질환 환자의 분포를 확인 했을 때 40대 이후로 심장질환 환자가 높아지는 부분을 확인할 수 있었습니다. 이와 관련하여 심장질환의 경우 절대적으로 나이가 높아질 수록 해당 질환이 발병되는 것이 아닌 특정 나이 이상이 되는 경우 심장 질환 발생 정도가 높아진다는 것을 보여주고 있습니다.

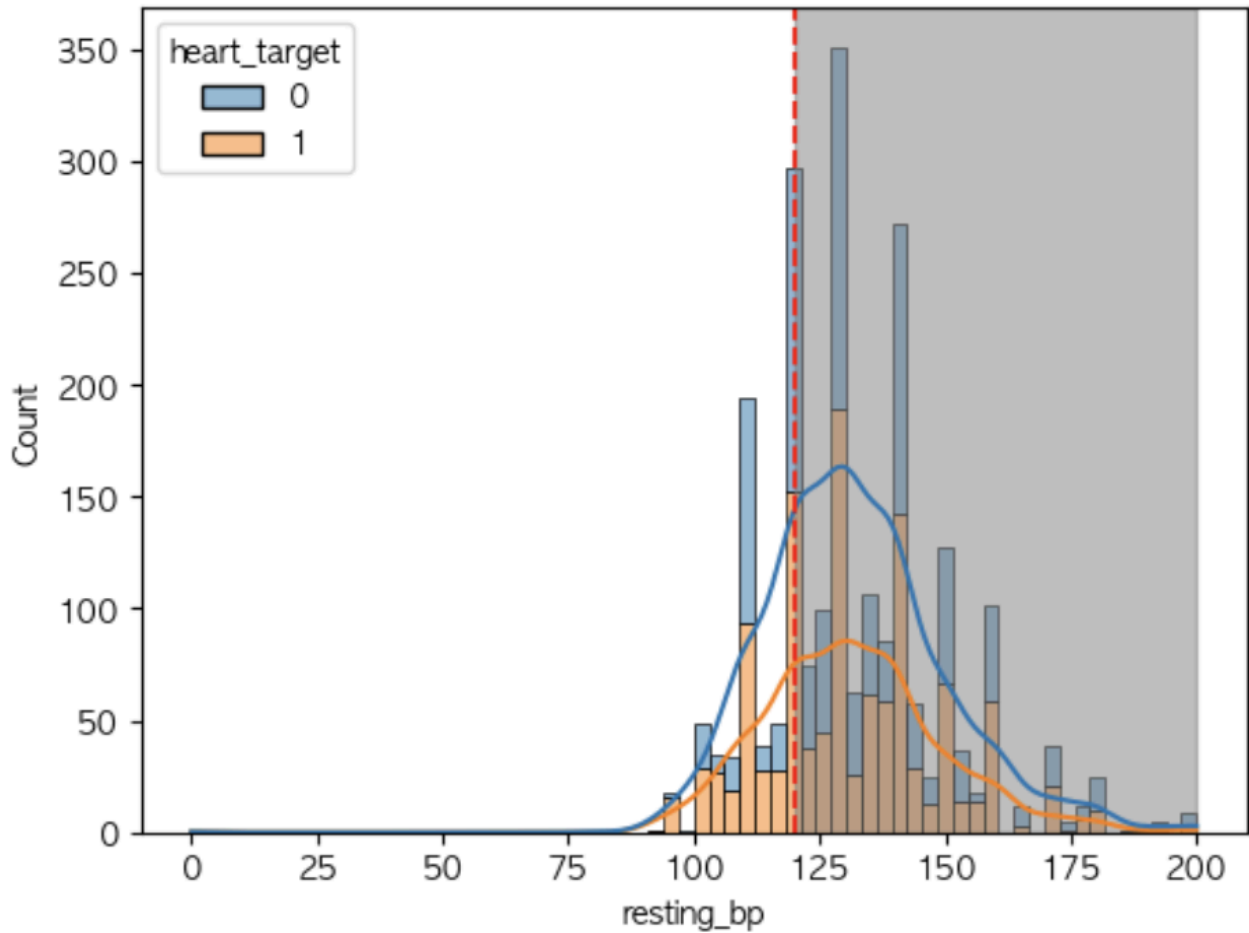
3. 통증 범주에 따른 심장 질환 유무

(0: 무통증, 1: 전형적인 심장협심증, 2: 불특정환 심장 협심증, 3: 심장이 아닌 다른 흉통)



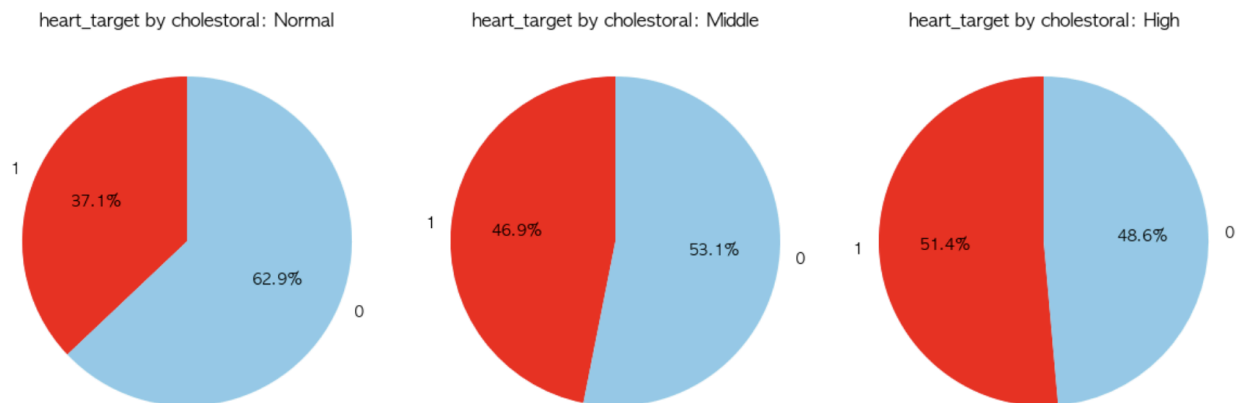
흉통의 종류에 따라 심장 질환 여부를 확인해봤을 때 심장 통증이 있는 경우 50% 이상의 경우로 심장 질환이 발생했으며, 무증상인 경우에도 45% 이상의 확률로 심장질환이 발생하고 있습니다. 따라서 심장에 통증이 없더라도 다른 증상에 의해서 심장질환이 발생할 수 있는 경우를 시사하고 있으며, 심장 질환의 경우 무조건 적인 통증이 동반되지 않음을 시사하고 있습니다.

4. 안정기 혈압에 따른 심장질환유무



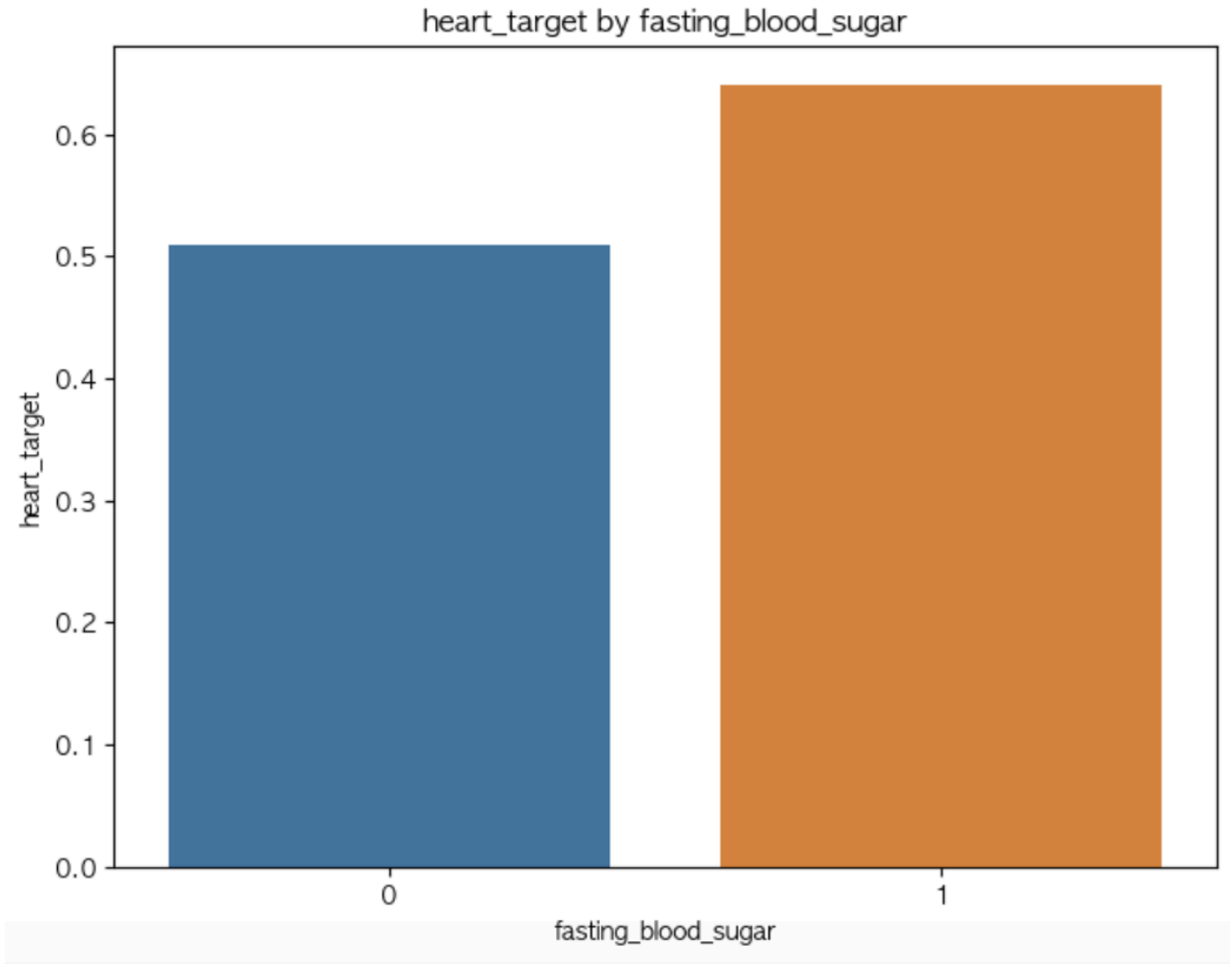
위에서 확인한 결과로 안정기 혈압이 120 이상이면 경계성 고혈압을 의심할 수 있습니다. bp 가 120 이상인 환자에 대하여 심장질환의 분포를 구분하였을 때 일반적으로 50% 가 넘긴 하지만 정상적인 (심장질환 없는) 환자의 경우도 유사한 분포를 나타내고 있어 해당 변수에 대해서는 심장질환과 연관성이 있다고 말할 수 없을 것 같습니다.

5. 콜레스테롤 수치에 따른 심장질환 유무



콜레스테롤 수치가 높아질 수록 심장 질환이 있을 확률이 점진적으로 높아지고 있는 양상을 보았을 때, 콜레스테롤 수치가 심장질환에 일부 영향을 줄 수 있음을 시사합니다.

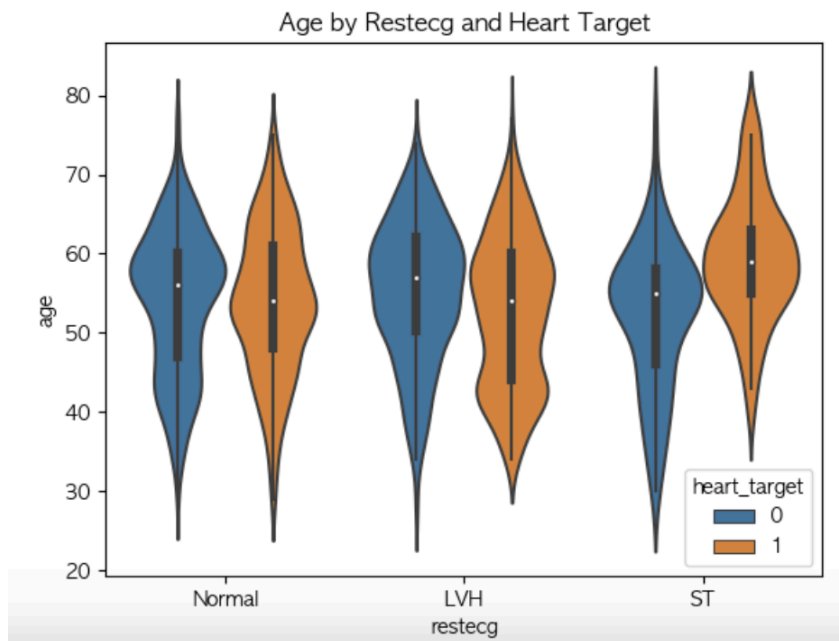
6. 공복혈당수치에 따른 심장질환 유무



공복혈당이 비정상인 경우에 심장질환 환자의 비율이 높은 것으로 나타났습니다.

하지만 실제 평균 비율을 확인했을 때 정상과 높은 경우 약 51%와 49% 수치로 비교 대상의 큰 차이가 없었으며, 이와 같은 결과를 확인했을 때 공복혈당이 정상치인 경우에도 심장질환의 위험성이 있다는 것을 시사하고 있고 해당 조건이 심장질환의 절대 조건이 없음을 시사합니다.

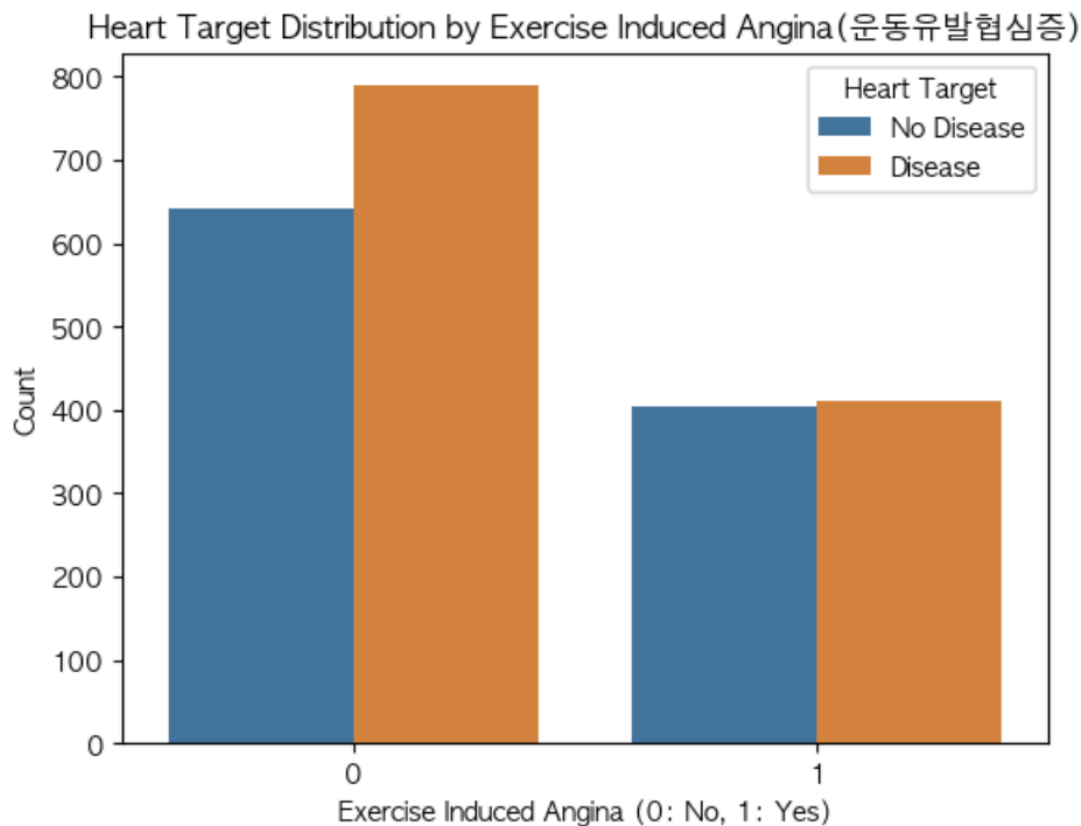
7. 심전도 이상에 따른 심장 질환유무



해당 분포 그래프로 보았을 때 심장 질환 발생과 관계없이 나이에 따라 유사한 분포 양상을 보이고 있습니다.

ST-T의 경우 관상동맥 질환, 또는 다양한 형태를 나타내고 있는 심전도 이상증상이며, 심장 질환이 있는 환자 중 높은 나이대에서 조금 더 많은 분포를 보이고 있습니다. 하지만 정상인 그래프와 비교해서 이상적인 특징은 나타나지 않았으며 해당 조건이 심장 질환 유,무를 판단할 수 있는 단편적인 조건이라고 판단하기는 어렵습니다.

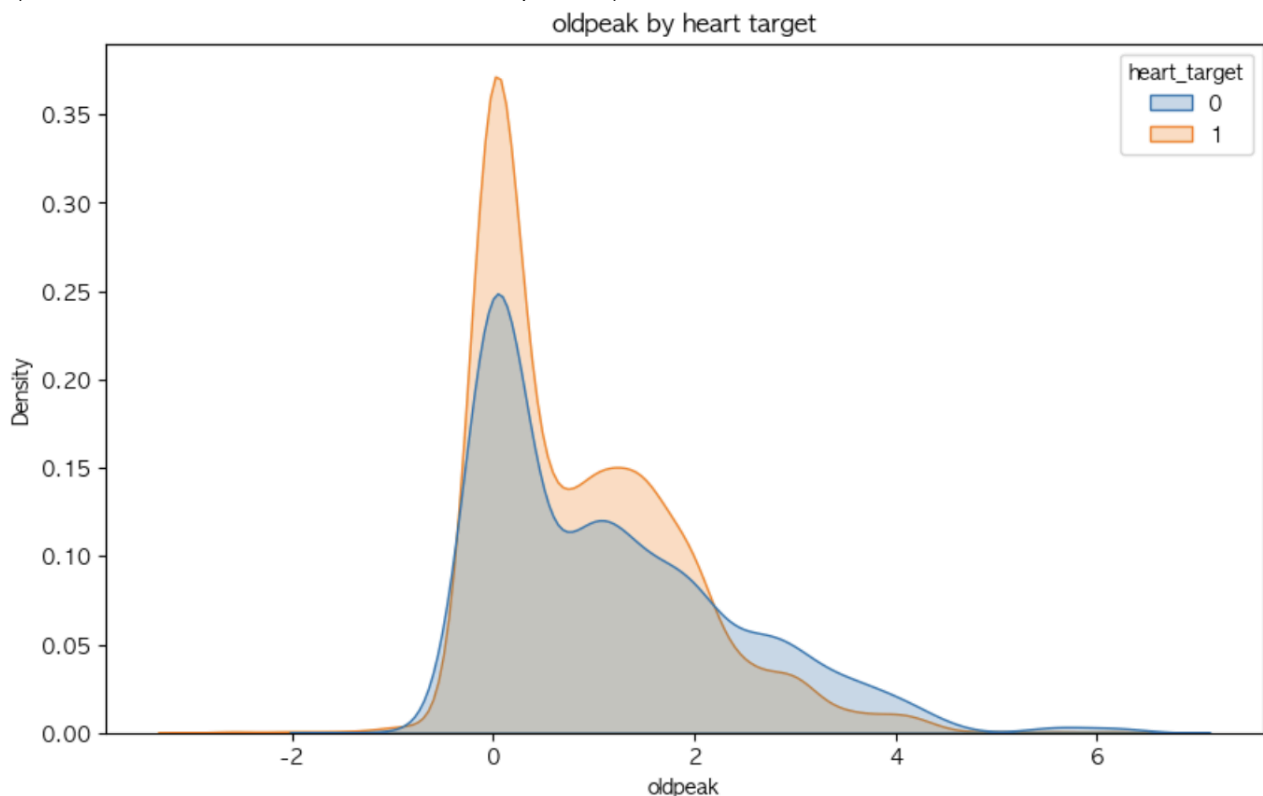
8. 운동 유발 협심증 관련 심장질환 유무



위의 그래프를 봤을 때 운동 유발 협심증을 인지한 사람들에 대해서는 거의 1:1 비율로 심장 질환이 있음을 시사할 수 있으며, 해당 증상을 느낀 사람에 대해서는 필수적으로 심장 질환 검사가 있는지 추가 검사를 진행해야 할 필요가 있어 보입니다.

단, 해당 검사를 통해 단일적으로 심장질환이 있을 수 있다고 판단하기 어려우며 해당 협심증을 느끼지 못한 환자일 경우라도 다른 검사와 복합적인 검사를 할 필요성이 있습니다.

9. 운동 이후 ST 분절 하강 정도에 따른 심장 질환 유무 (비교적 안정되기까지 운동으로 유발되는 ST depression)



일반적으로 분절하강 정도의 기울기가 0에 수렴할 수록 심장 질환에 이상이 없다는 가능성을 시사합니다.

위의 그래프를 봤을 때 **oldpeak**가 0~2 사이가 측정되는 경우 정상적인 환자에 비하여 심장질환이 있는 환자의 분포도가 높다는 것을 시사합니다. 이를 보았을 때 분절하강정도가 0에 측정되었더라도 심장 질환이 있는 환자가 존재할 수 있으며 해당 지표는 다양한 검사들에 대해서 일부임을 알 수 있습니다.따라서 해당 지표는 독립 변수로 볼 수 없으며 다수의 임상 지표와 연관지어서 생각할 필요가 있습니다.

4-1. 심장질환 발생에 따른 요인들에 대한 결과 분석

가. 성별과 심장질환 발생 비율(!!)

- 데이터셋 분석 결과, 남성이 여성에 비해 심장질환 발생 비율이 높음을 나타내고 있습니다.

나. 나이와 심장질환 발생의 상관관계(!!)

- 40대 이후부터 심장질환 환자 비율이 증가하는 경향 확인.
- 나이 증가와 심장질환 발생의 연관성 있음을 시사.

다. 흉통 유무와 심장질환 발생(!!!)

- 흉통이 있는 경우 약 50% 이상에서 심장질환이 발생했으나, 흉통이 없는 경우에도 45%에서 발생 확인.
- 흉통이 심장질환의 유일한 지표가 아님을 나타내며, 다른 증상들과 함께 고려 필요.

라. 안정기 혈압과 심장질환 발생(!)

- 혈압이 120 이상인 경우에서도 심장질환의 분포가 심장질환이 없는 그룹과 유사함을 확인.
- 혈압 수치만으로 심장질환의 위험을 판단하기 어려운 것으로 판단됨.

마. 콜레스테롤 수치와 심장질환 발생(!)

- 콜레스테롤 수치가 높을수록 심장질환 발생 확률이 점진적으로 증가하는 경향 확인.
- 콜레스테롤이 심장질환 발생에 영향을 줄 수 있는 요소로 판단됨.

바. 공복혈당 수치와 심장질환 발생(!)

- 공복혈당이 비정상인 경우 심장질환 환자의 비율이 다소 높게 나타났으나, 정상 수치를 가진 경우에도 심장질환 위험이 있음을 시사.
- 공복혈당 수치와 심장질환 발생은 직접적인 연관이 없음을 나타냄.

사. 심전도 이상증상과 심장질환 발생(!)

- ST-T 파형 이상이 있는 환자들 중 높은 나이대에서 더 많은 분포를 보이나, 심장질환 유무를 단독으로 판단하기 어려움.

아. 운동 유발 협심증 인지 여부와 심장질환 발생(!)

- 운동 유발 협심증을 인지한 사람들의 경우 거의 1:1 비율로 심장 질환과 관련이 있음을 나타냄.
- 협심증 증상이 없는 경우에도 다른 검사와 복합적인 접근 필요.

자. ST 분절하강 정도와 심장질환 발생(!)

- ST 분절하강 정도가 0에 수렴하는 것이 심장 질환 유무를 단독으로 결정하는 지표는 아님.
- 0~2 범위의 분절하강을 보이는 경우 심장질환 환자의 분포가 높아, 다양한 임상 지표와의 복합적인 분석 필요성을 시사.

5. 결론

A. 요약

가. 당뇨 발생 요인 분석 결과:

- 성별, 연령, 혈당 수치 및 BMI와 당뇨 발생 간의 상관관계는 명확하게 나타났습니다.
- 남성, 고령자, 비정상적 혈당 수치 및 높은 BMI 범주의 개인들이 당뇨병 발생 위험이 더 높은 경향을 보였습니다.

나. 심장질환/뇌졸중 발생 요인 분석 결과:

- 혈압과 콜레스테롤 수치가 높은 개인들이 심장질환 위험이 더 높지만, 혈당 수치와는 명확한 직접적 연관성이 확인되지 않았습니다.

다. 심장질환 발생 요인 분석 결과:

- 성별, 연령, 흡통 유무, 혈압 및 콜레스테롤 수치와 심장질환 발생 간의 상관관계는 일관적이지 않게 나타났습니다.

B. 각 질병에 대한 종합 평가

가. 당뇨와 심장질환 발생 요인 간에는 부분적인 연관성이 관찰됩니다.

특히, 비정상적인 혈당 수치는 두 조건 모두에서 발생 위험을 높이는 요소로 나타났습니다. 그러나 이것은 심장질환 발생의 유일한 결정 요인은 아니며, 추가적인 변수들과의 복합적인 상호작용을 통해 심장질환 발생 위험이 결정될 가능성이 있습니다.

나. 연령과 BMI는 두 조건 모두에서 중요한 요인으로 확인되었으나, 당뇨병의 경우 혈당 수치가 직접적인 지표로 작용하는 반면 심장질환에서는 여러 요인들이 상호작용을 하는 것으로 보입니다.

다. 당뇨병 환자가 심장질환을 개발할 위험이 높은 것으로 일반적으로 알려져 있으므로, 이 두 조건 간에는 상호 종속적인 관계가 있을 가능성이 높습니다.

라. 결론적으로, 성별, 연령, 혈당 수치 및 BMI와 같은 요인들은 당뇨와 심장질환 발생 모두에 영향을 미치는 것으로 나타났으나, 이러한 질병들 간의 연관성은 복잡하며 한 요인만으로 결정되지 않습니다.

심장질환과 뇌졸중의 연관성은 별도의 심층 분석을 필요로 하며, 복잡한 인과 관계를 이해하기 위해서는 더 많은 데이터와 다변량 분석이 필요합니다.

6. Outro

가. 질병 발생 여부의 확정을 위해서는 포괄적인 검사와 분석이 필수적입니다. 단일 검사 결과에서의 이상치가 반드시 질병을 의미하지는 않으며, 다양한 독립 변수들이 복합적으로 작용하여 질병의 가능성을 가늠하게 됩니다. 따라서, 어떠한 검사 결과도 질병 발생을 **100%** 확신할 수 없으며, 결과는 질병 발생의 가능성을 높이는 요소 중 하나로 해석되어야 합니다.