

2021년도 1학기 빅데이터분석 프로젝트 보고서

학 번	2320	이 름	박진성
1. 개발 개요			
주 제	건강		
데이터	Heart Attack Analysis & Prediction Dataset https://www.kaggle.com/rashikrahmanpritom/heart-attack-analysis-prediction-dataset?select=o2Saturation.csv		
2. 프로젝트 내용			
문제정의	<p>의료 기술이 발전하면서 심각하지 않은 질병으로 죽게 되는 사람이 줄어들었다.</p> <p>하지만 심혈관 질환으로 죽는 사람은 암에 이어 2번째를 차지하고 있다.</p> <p>그래서 심혈관 질환 중 하나인 심장마비에 대해 분석을 해볼 것이다.</p> <p>가설</p> <ol style="list-style-type: none">1. 나이가 많을수록 심장마비의 확률이 높다.2. 남성이 여성보다 마비가 올 확률이 높다.3. 콜레스테롤 수치가 높을수록 마비가 올 확률이 높다.		

EDA

1. 성별 분석

성별 컬럼 (sex)을 value_counts()로 확인하면 증상을 보인 전체 303명 중 남자는 207명, 여자는 96명이다.

Countplot으로 보면 확연한 차이를 알 수 있다.

Pie chart로 백분율을 보면 남자가 약 68.3%, 여자가 31.7%로 남자가 여자보다 많다는 것을 알 수 있다.

2. 나이 분석

value_counts로 보면 50, 60, 40, 30, 70, 20 대순으로 많다는 것을 알 수 있다.

describe() 로 보면 평균은 54.4, 최저값은 29, 최댓값은 77세이다.

히스토그램, countplot으로 보면 4~60대 사이에 치중되어 있다는 것을 알 수 있다.

Pie chart를 보면 백분율을 알 수 있다.

3. 콜레스테롤 분석

Value_counts() 로 보면 수치가 너무 다양해서 제대로 파악할 수 없다.

수치를 단계별로 나누기 위해 인덱싱 한다.

분포를 알아보기 위해 scatter plot을 본다. 아직 확실하지는 않지만 수치가 높을수록 마비의 확률이 높을 것이라고 예측할 수 있다.

Countplot과 Pie chart로 대소 비교 및 백분율을 알 수 있다.

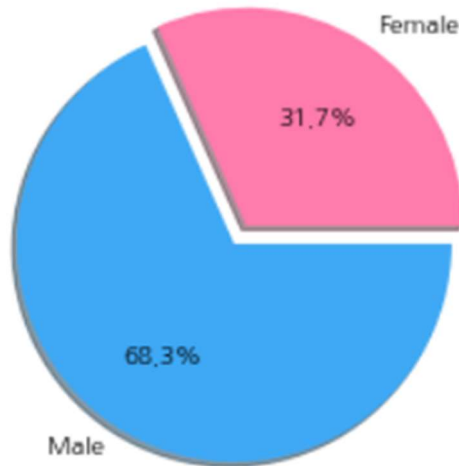
3. 분석 결과

시각화

1. 시각화 1

※ 시각화한 결과와 설명을 작성합니다.

(3개 이상: 15점, 2개: 10점, 1개: 5점)



1. 설명

labels로 어느 파이가 남성, 여성인지 알려준다.

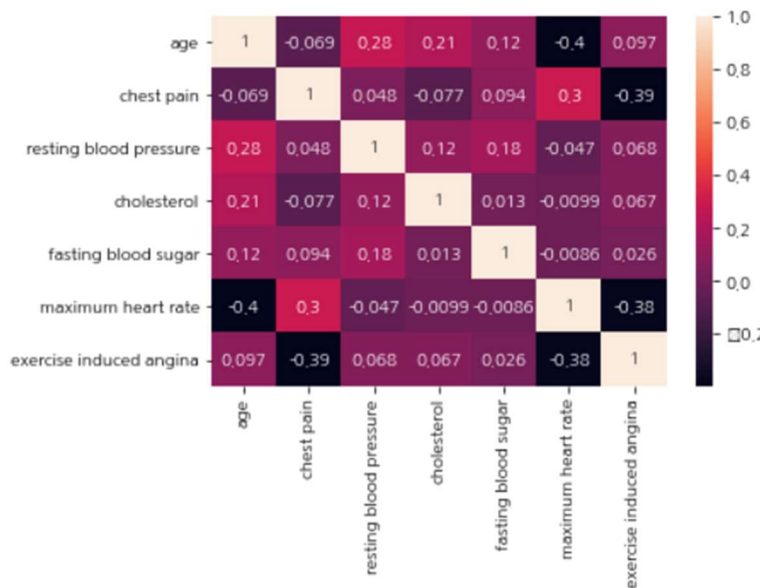
autopct를 이용해 소수점 2번째 자리에서 반올림한다.

Shadow 옵션으로 그림자를 넣는다.

Colors 옵션으로 Hexcode를 넣어 색을 입힌다.

Explode로 얼마나 중심에서 떼어낼 것인지 정한다.

2. 시각화 2

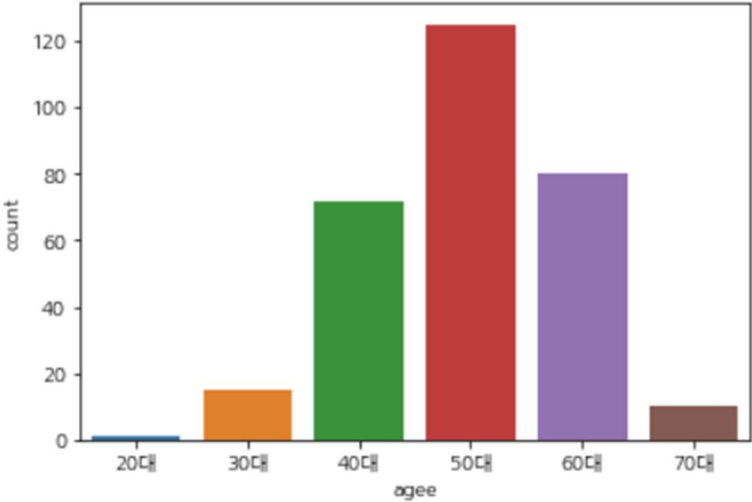


2. 설명

1. 두 컬럼이 정비례 관계이면 상관관계가 양수, 그 반대면 음수이다.

2. 양의 상관관계 중에는 흉통과 최고 심박수, 나이와 안정 시 혈압 사이의 관계가 있는 것으로 밝혀졌다.

3. 음의 상관관계 중에는 흉통과 운동 협심증, 나이와 최고 심박수 사이의 관계가 있는 것으로 밝혀졌다.

	<div>3. 시각화 3</div> <div></div>	<div>3. 설명</div> <div>나이대를 나누어 주는 age라는 함수를 작성, age함수의 return 값이 들어있는 age컬럼으로 countplot을 만들.</div> <div>심장마비에 걸릴 확률이 50대, 60대, 40대, 30대, 70대, 20대 순이라는 것을 보여줌.</div>
<div>분석 결과</div>	<div>나이가 많을수록, 여성보다 남성이, 콜레스테롤 수치가 높을수록 심장마비의 확률이 높음.</div> <div>심박수가 증가하면서 흉통이 생겼을 수도 있고, 흉통이 생기자 당황하여 심박수가 증가했을 수도 있다. 나이를 먹을수록 안정 시 혈압이 높아져 심장마비의 확률이 높아진다.</div> <div>즉, 나이를 많이 먹을수록 건강관리에 더 유의해야 한다.</div> <div>콜레스테롤 수치와 혈당을 잘 관리할 필요가 있음</div>	
<div>4. 평 가</div>		
<div>느낀점</div>	<div>여러 가지 요소를 사람이 분석하려면 적지 않은 노력이 필요한데 빅데이터 분석을 사용해서 빠르고, 편리하게, 쉽게 분석할 수 있다는 사실을 느꼈다.</div> <div>데이터셋의 크기가 별로 크지 않아서 이게 전체를 담을 수 있는 분석인지 의문이 드는 부분이 몇 가지 있었는데, 다른 분야에서 큰 데이터셋을 찾아 분석해봐야겠다고 생각했다.</div>	