

HEALTHCARE PREDICTIVE ANALYTICS



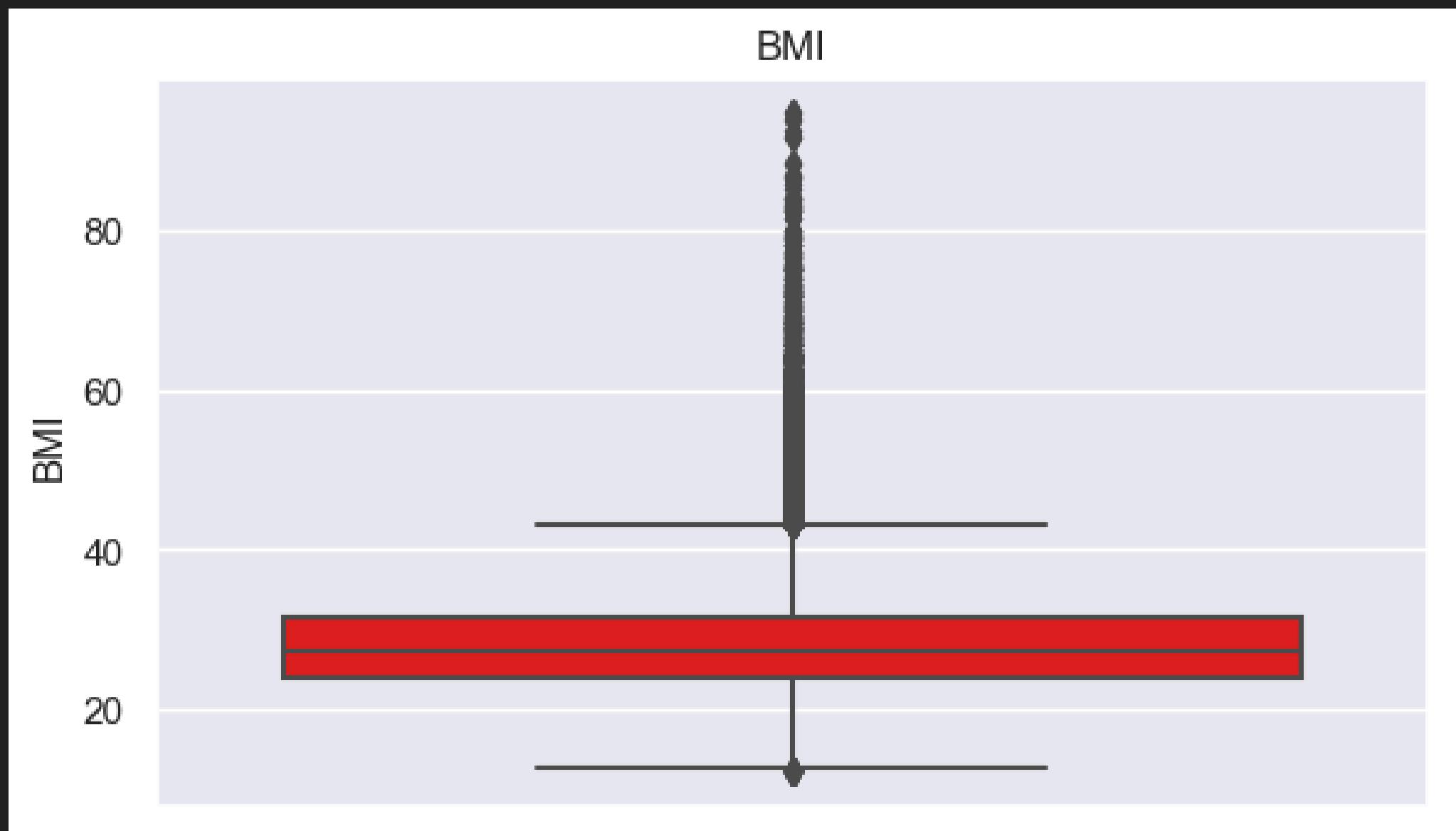
ABOUT THE PROJECT

In today's digital age, data is a critical asset. Effective data analysis enables organizations to identify trends, improve operations, so we try to do so for health .

Developing a predictive model for Heart Disease by providing data-driven insights.

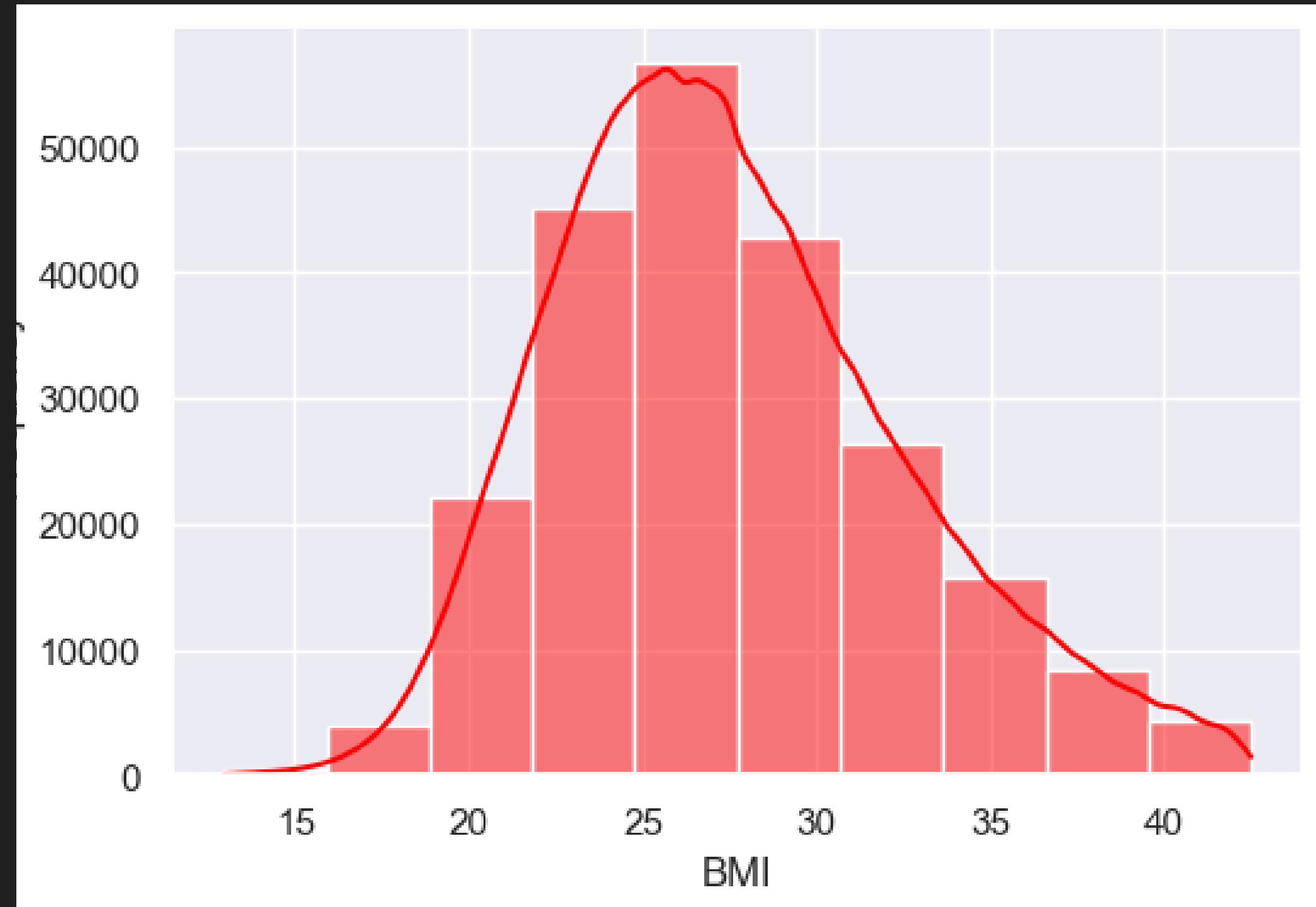
1) DATA EXPLORATION

- 1- Explore data by checking null values and inconsistent .
- 2- checking outliers in numeric features by making **Box plot**



BMI DISTRIBUTION

We noticed that most people have BMI values between **25:30**



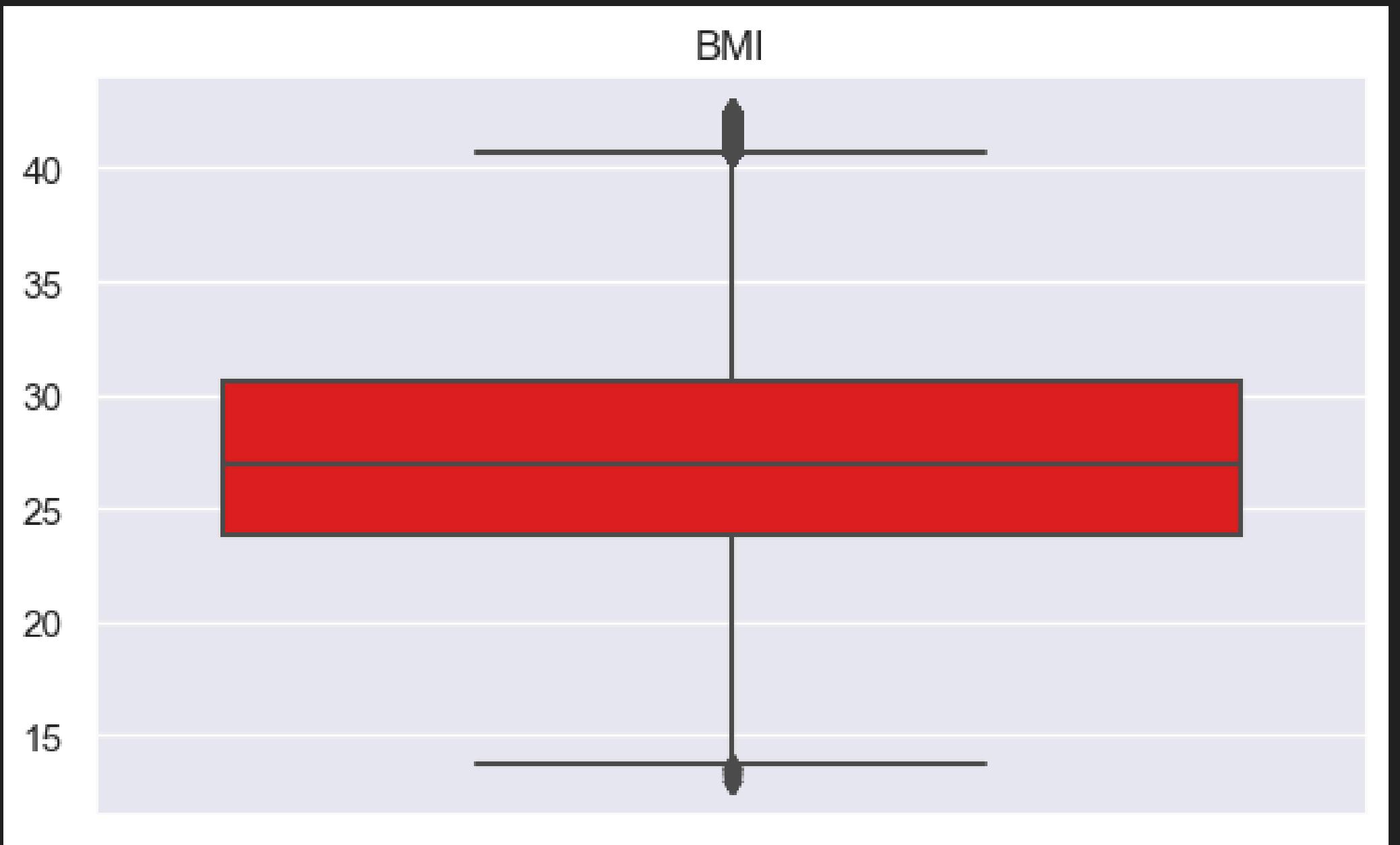
2) DATA PREPROCESSING

Fixing outliers by making
IQR.

BMI max value = **40**

instead of 80

Then, filter the DataFrame
to remove outliers.



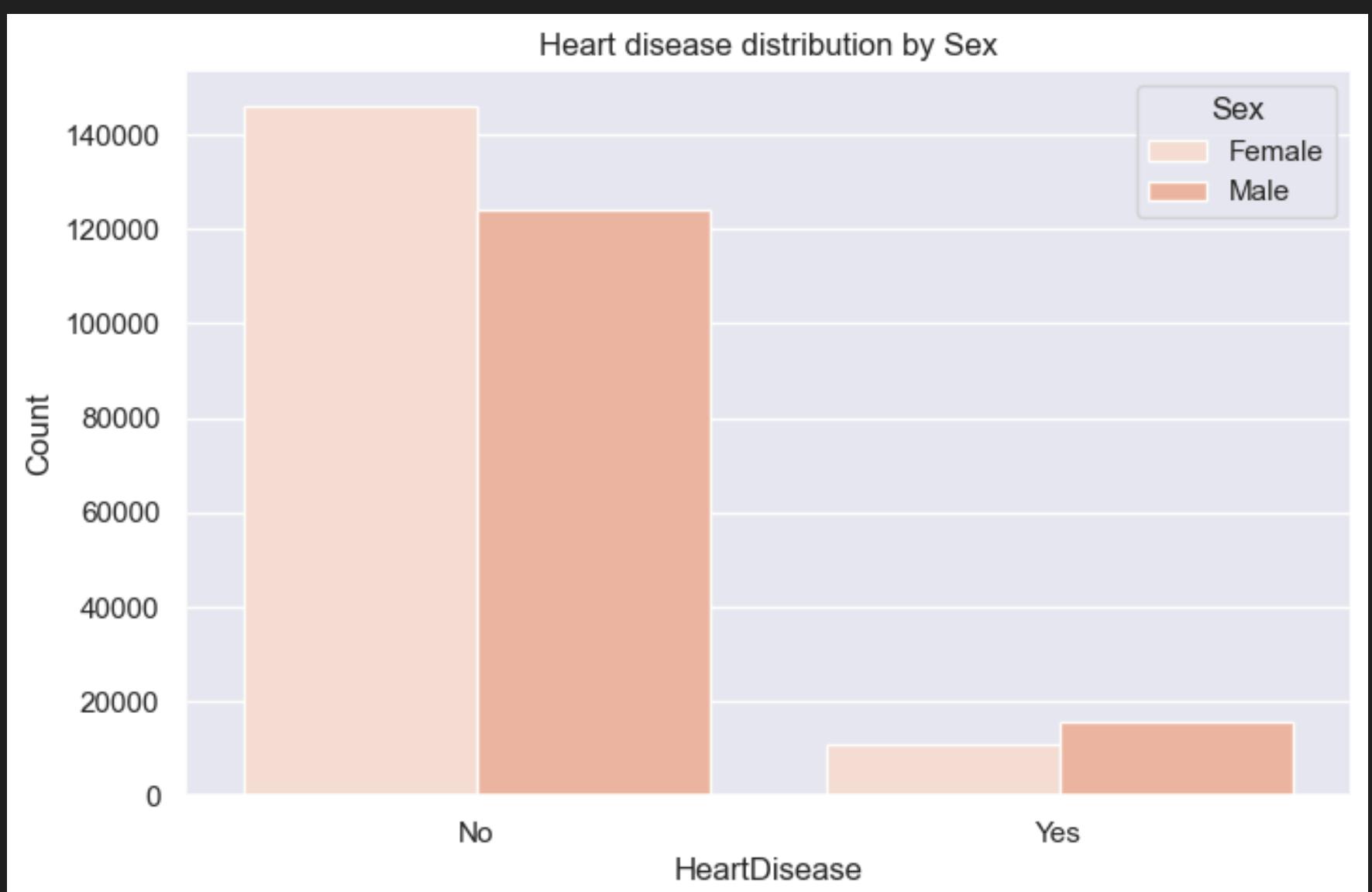
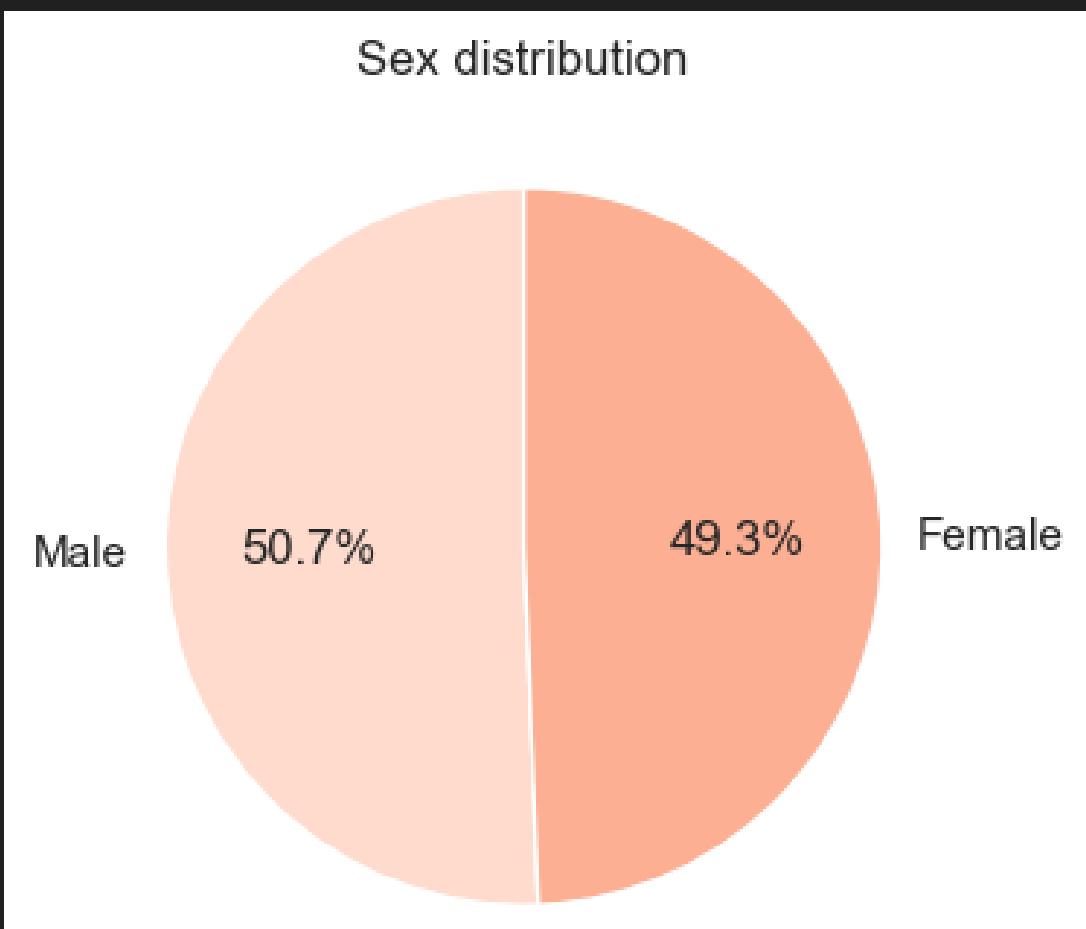
KEY FEATURES

As Heart Disease is the main feature,
Check distribution for
several features vs Heart Disease.

1) Sex distribution approximately the same.

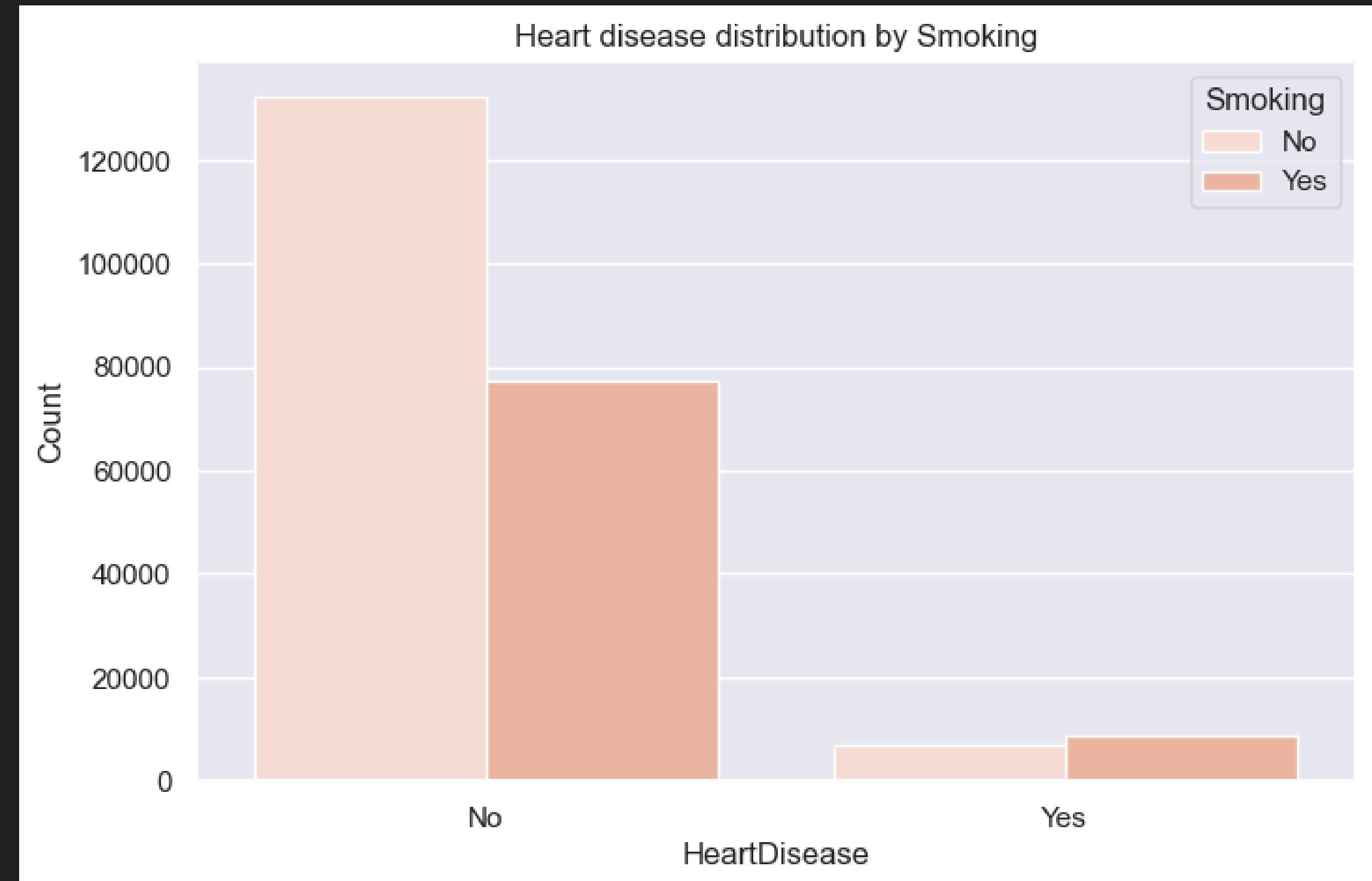
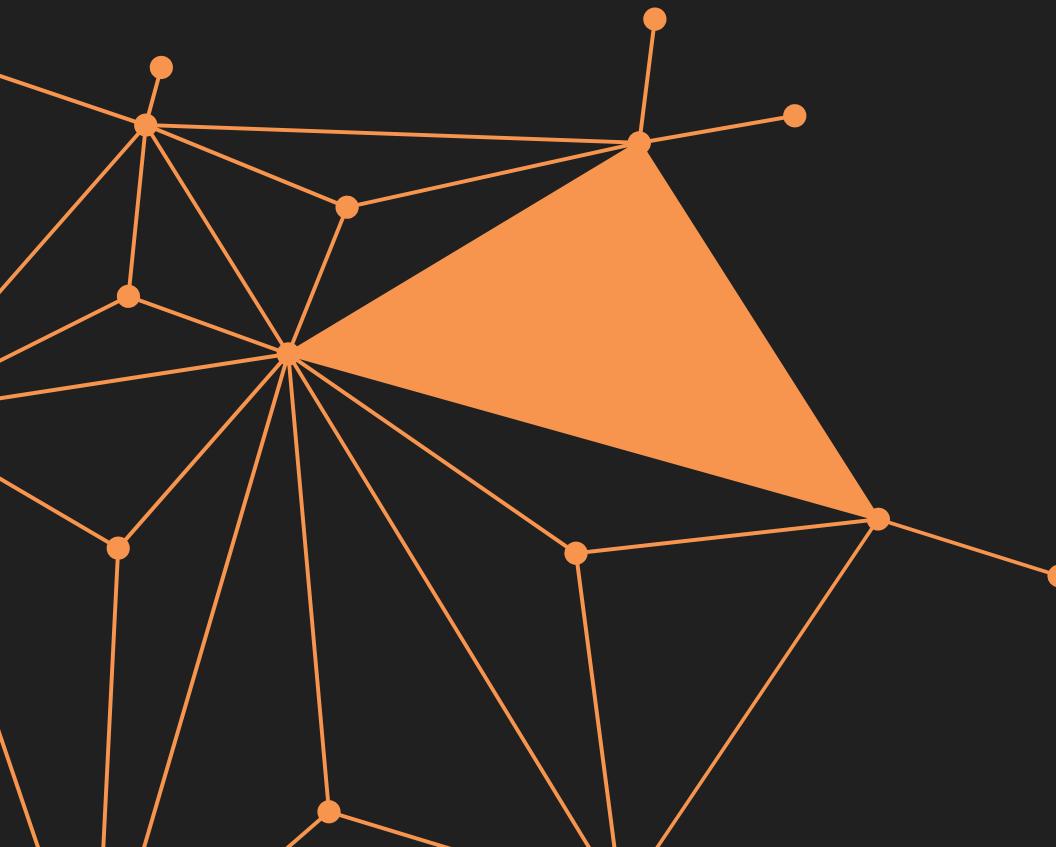
Males with disease = 18000,
females = 15000.

so we can say that **Males** are **more susceptible** to have heart disease than females.



NO strong relation between smoking vs heart disease.

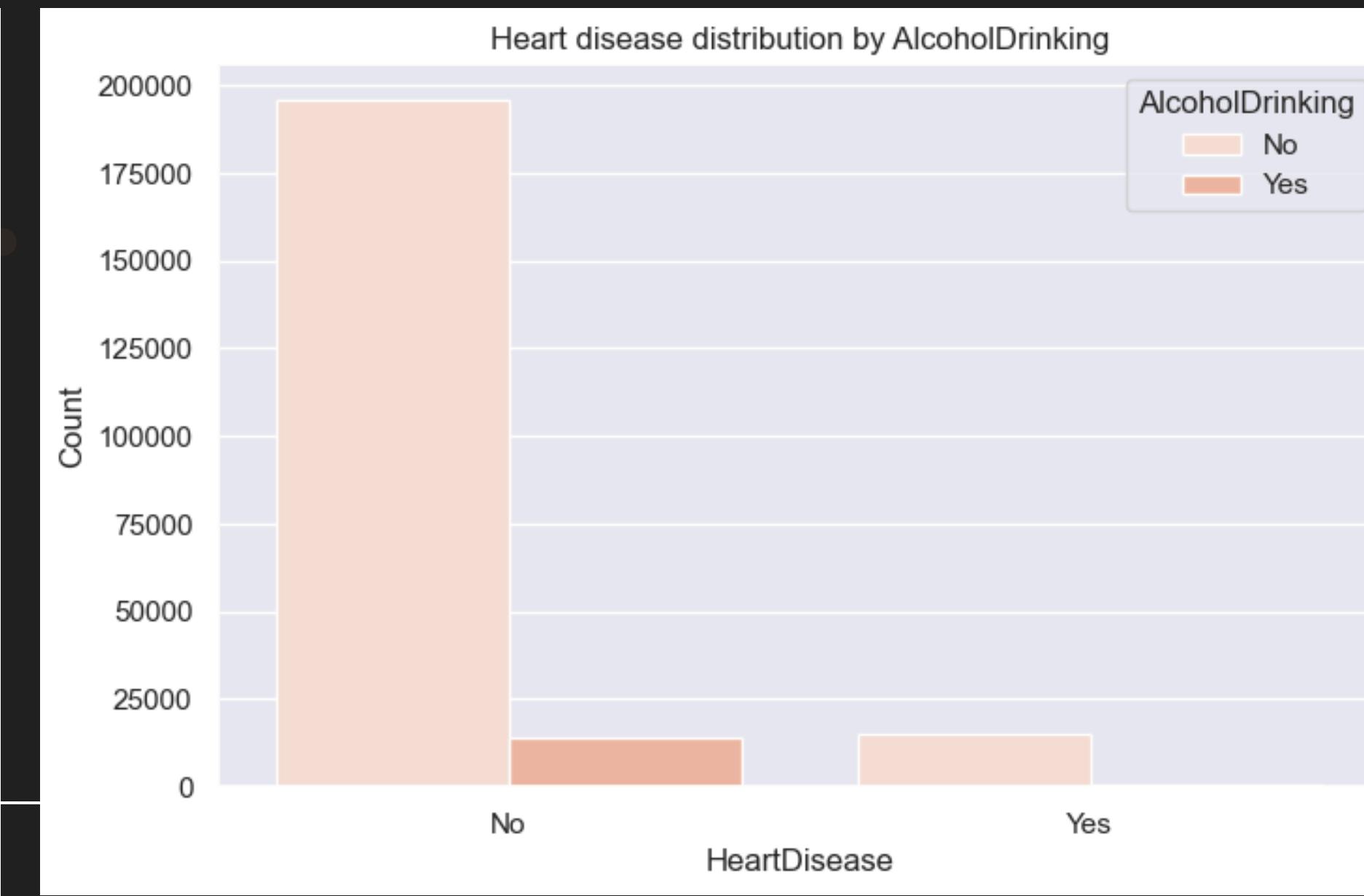
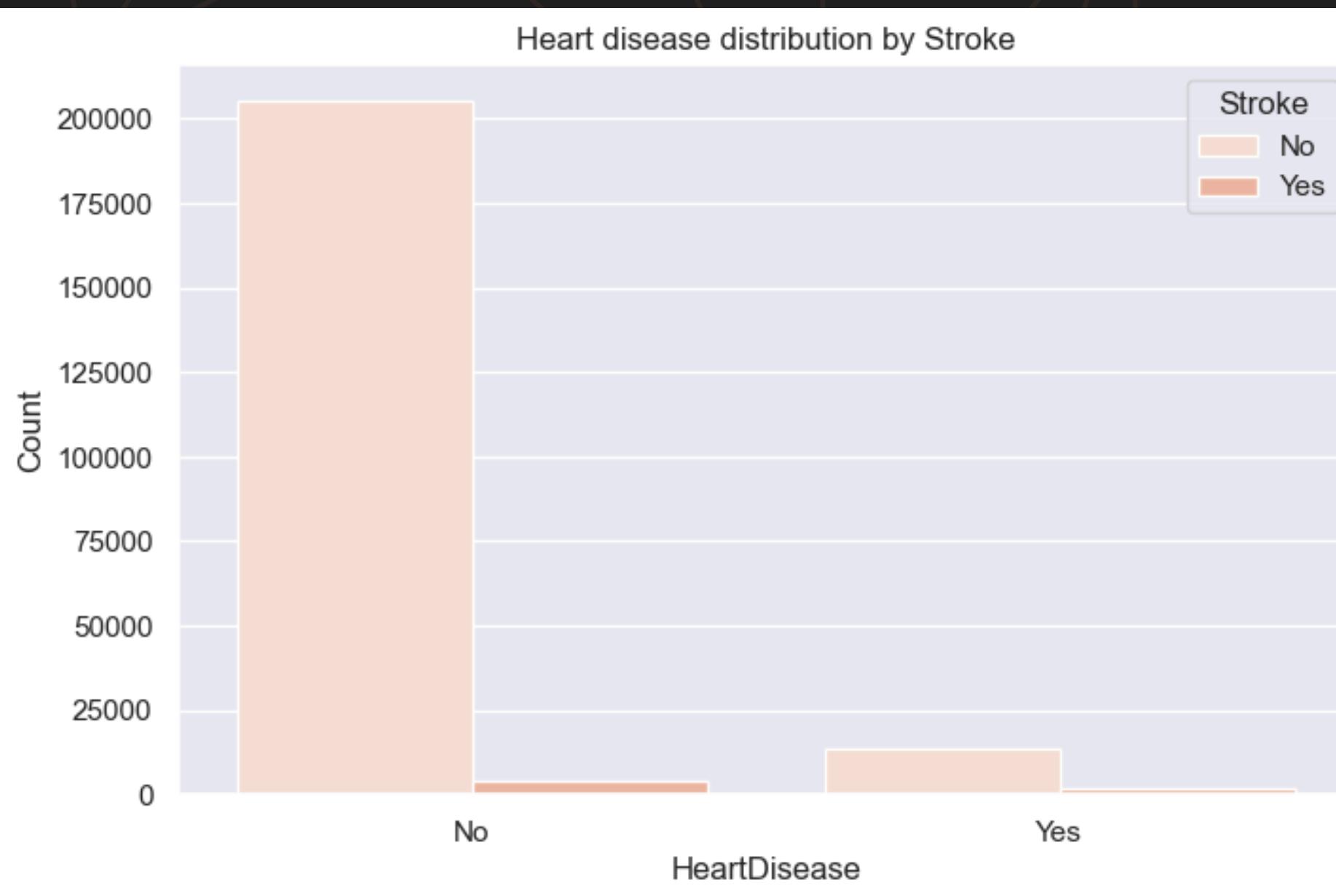
As percent of smoking & non smoking is very close



HeartDisease vs Stroke

There a **Strong** relationship between them .
as the higher percent for both have a '**Yes**' Heart Disease

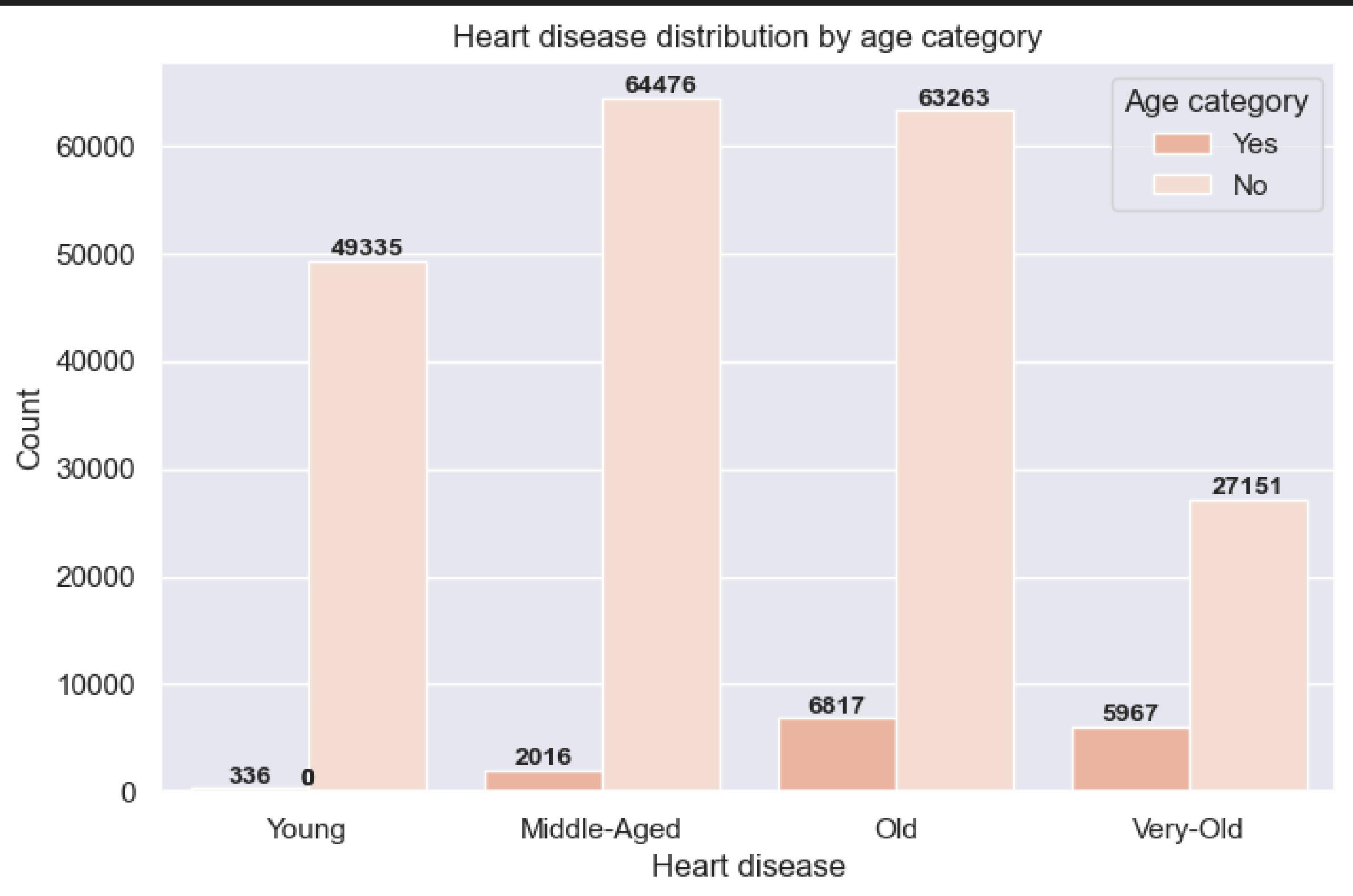
HeartDisease vs Alcohol



AGE DISTRIBUTION

Age Intervals	Age Category	Value Count
'18-24' '25-29' '30-34' '35-39'	'Young'	49671
'40-44' , '45-49' '50-54' '55-59'	'Middle-Age'	89638
'60-64' , '65-69' '70-74'	Old	46934
'75-or older'	Very old	33118

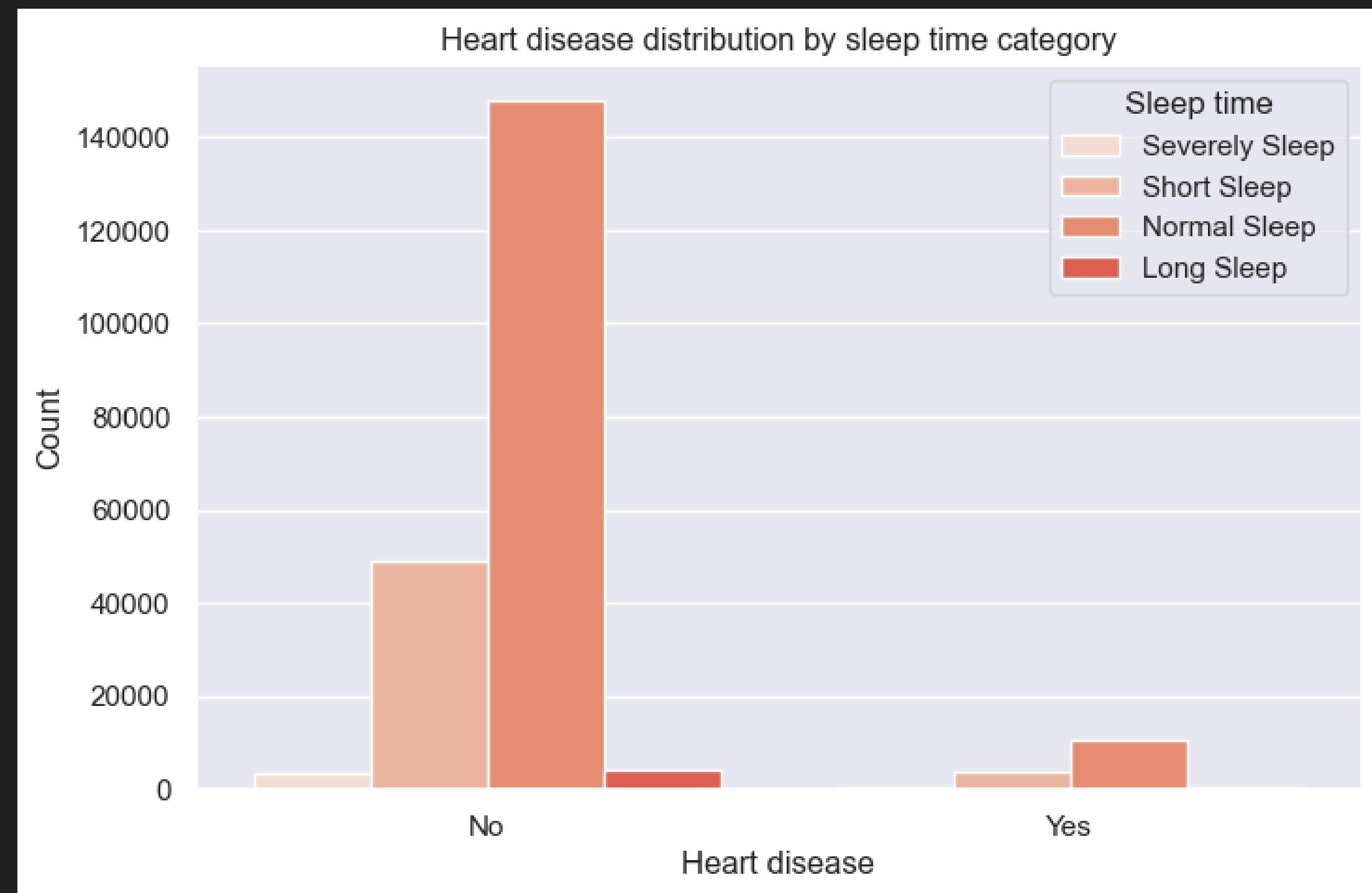
AGE DISTRIBUTION



Most people with heartDisease are of **Middle-Old** category '**75 or more'** aged.

SLEEP TIME VS HEARTDISEAS

People with **good sleep** quality have **more probability** to have **NO heart-disease**



HEAT MAP

Notice that,

- **HeartDisease** mostly related to **AgeCategory**
- **Diabetic** affected by **BMI**
- **Age** affect the **diffWalking**

Correlation Matrix																	
HeartDisease	1.00	0.04	0.10	-0.03	0.16	0.05	-0.04	0.14	0.08	0.22	0.15	-0.06	-0.19	0.02	0.02	0.11	0.09
BMI	0.04	1.00	0.03	-0.03	0.02	0.04	-0.02	0.11	0.09	0.02	0.18	-0.11	-0.21	-0.06	0.05	0.03	-0.03
Smoking	0.10	0.03	1.00	0.11	0.05	0.02	-0.02	0.07	0.10	0.14	0.05	-0.07	-0.14	-0.02	0.00	0.03	0.03
AlcoholDrinking	-0.03	-0.03	0.11	1.00	-0.02	-0.00	0.04	-0.02	0.01	-0.04	-0.05	0.01	0.02	0.00	-0.00	-0.02	-0.00
Stroke	0.16	0.02	0.05	-0.02	1.00	0.04	-0.02	0.13	0.00	0.13	0.08	-0.05	-0.11	0.02	0.02	0.06	0.04
PhysicalHealth	0.05	0.04	0.02	-0.00	0.04	1.00	0.13	0.11	-0.03	0.02	0.05	-0.03	-0.18	-0.03	0.06	0.04	0.02
MentalHealth	-0.04	-0.02	-0.02	0.04	-0.02	0.13	1.00	-0.02	-0.10	-0.19	-0.04	0.05	-0.03	-0.05	0.06	-0.01	-0.04
DiffWalking	0.14	0.11	0.07	-0.02	0.13	0.11	-0.02	1.00	-0.05	0.22	0.15	-0.18	-0.25	0.01	0.05	0.10	0.06
Sex	0.08	0.09	0.10	0.01	0.00	-0.03	-0.10	-0.05	1.00	-0.10	0.00	0.03	-0.01	-0.03	-0.05	-0.00	0.01
AgeCategory	0.22	0.02	0.14	-0.04	0.13	0.02	-0.19	0.22	-0.10	1.00	0.20	-0.09	-0.19	0.10	-0.05	0.12	0.26
Diabetic	0.15	0.18	0.05	-0.05	0.08	0.05	-0.04	0.15	0.00	0.20	1.00	-0.10	-0.25	0.01	0.03	0.12	0.04
PhysicalActivity	-0.06	-0.11	-0.07	0.01	-0.05	-0.03	0.05	-0.18	0.03	-0.09	-0.10	1.00	0.19	0.00	-0.01	-0.05	0.01
GenHealth	-0.19	-0.21	-0.14	0.02	-0.11	-0.18	-0.03	-0.25	-0.01	-0.19	-0.25	0.19	1.00	0.04	-0.08	-0.12	-0.04
SleepTime	0.02	-0.06	-0.02	0.00	0.02	-0.03	-0.05	0.01	-0.03	0.10	0.01	0.00	0.04	1.00	-0.03	0.01	0.05
Asthma	0.02	0.05	0.00	-0.00	0.02	0.06	0.06	0.05	-0.05	-0.05	0.03	-0.01	-0.08	-0.03	1.00	0.02	0.01
KidneyDisease	0.11	0.03	0.03	-0.02	0.06	0.04	-0.01	0.10	-0.00	0.12	0.12	-0.05	-0.12	0.01	0.02	1.00	0.06
SkinCancer	0.09	-0.03	0.03	-0.00	0.04	0.02	-0.04	0.06	0.01	0.26	0.04	0.01	-0.04	0.05	0.01	0.06	1.00
HeartDisease		BMI	Smoking	AlcoholDrinking	Stroke	PhysicalHealth	MentalHealth	DiffWalking	Sex	AgeCategory	Diabetic	PhysicalActivity	GenHealth	SleepTime	Asthma	KidneyDisease	SkinCancer

THANK YOU



MOHAMED
FARIDA

ZAHRAA
MOHAMED

MOAZ
MALAK