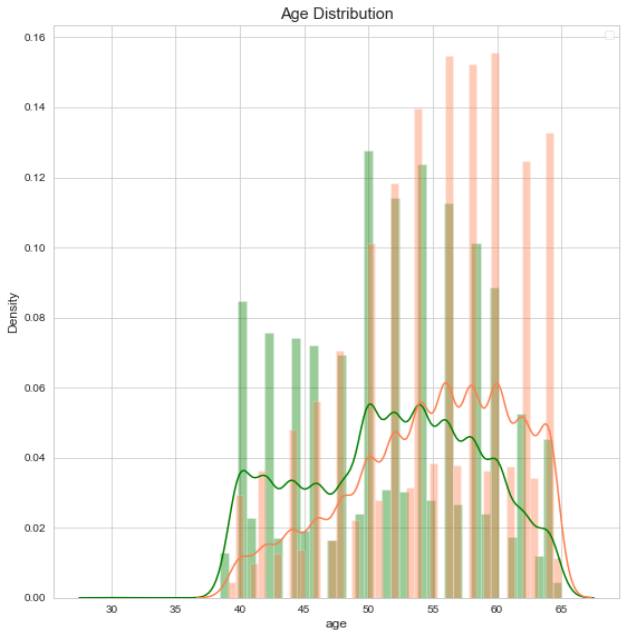
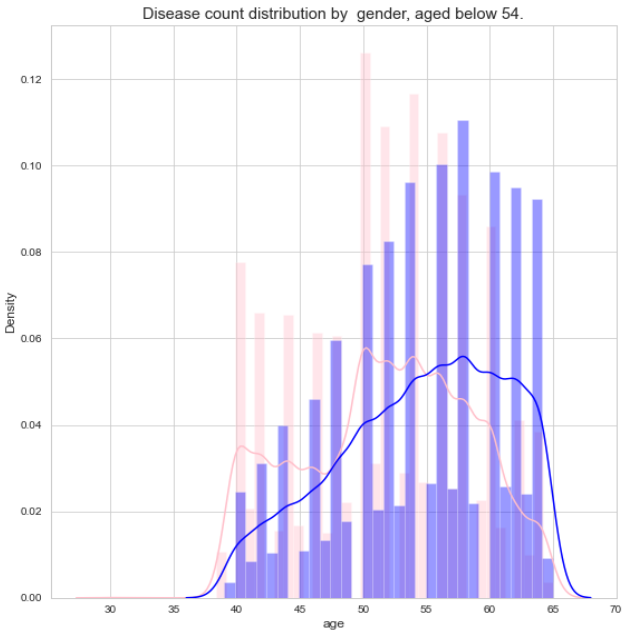
THE GOAL:

diagnosing the cardiovascular disease based on several features and symptoms given by the client I will use the features to determine if the disease exists or not in order to be able to warn the client and notify him either way. If he has the disease so he should see a doctor. In the other case, if he doesn't have the disease, he would know the most variable that could cause it in the future and try to avoid that cause.

When I start work on data I wanted to know if there is a relationship between age, gender and cardiovascular disease. I tried using EDA in two different ways as shown in the figure below.





('age',' disease not exists ') color='green'

('age',' disease exists '),color='coral'

I found that with age, the incidence of the disease increases

(('age','gender='F'" & ' disease exists ')color='pink')

('age', "gender='M'") & ('disease exists '),color='blue')

I found that with age, the incidence of the disease increases for male .

Next step:

Create a body mass variable using the height and weight column.

Understand the relationship between disease and other variables.

Using a logistic regression model.

Applying the cluster model to find the most vulnerable age group to infection with the disease.