Errors

ValueError: Could not interpret value `level` for `x`. Value is a string, but `data` was not passed.

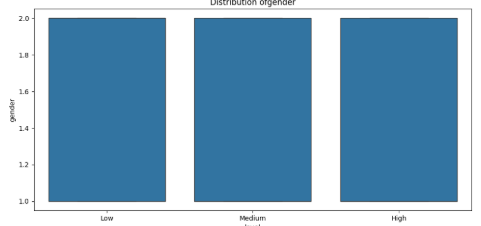
       sns.boxplot(x='level',y=column,order=['Low','Medium','High'])

because the data wasnt passed data=df

3.Learning how to analyze boxplot [Box Plot - GeeksforGeeks](https://www.geeksforgeeks.org/machine-learning/box-plot/)

2 . we cannot deaw any insight from this boz boxplot is not right we can draw insight from countplot

To know which gender has more high level



3. df.corr['level']=df.corr['level'].map({'Low':0,'Medium':1,'High':2})

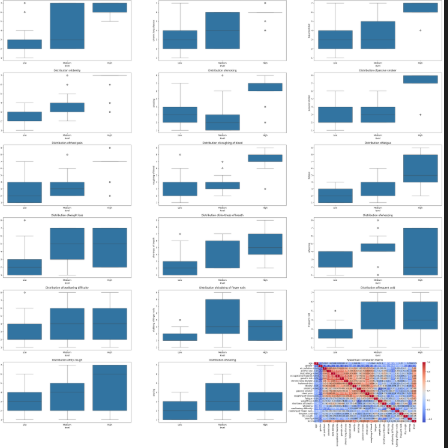
TypeError: 'method' object is not subscriptable

result[rl.indexer] = arr

~~~~~~^^^^^^^^^^^^

ValueError: could not convert string to float: 'Low'

* Approach:So for this tried mapping technique didnt work so did ordinal encoding

4 . mixing of the graphs

* Approach:wanted to separate function for different analysis
* Problem:Matplotlib has memory so until you use plt.close it would jus tscak it with other figure

1. What all approaches you have considered and why you are opting for certain approach and not considering other
2. What are you planning to do in case you don’t have any clarity on different approaches
3. Anything relevant to the decision-making process for working out the solution

EDA Insights:

why you are opting for certain approach and not considering other

1.The data is balanced so no need of Oversampling and Undersampling techniques

2. What are you planning to do in case you don’t have any clarity on different approaches

Ans: there are many features basically overwhelming to identify what is exactly causing

Or to check it individually all through 22 features

So, divided it in three steps:

* Univariate :one feature vs target(Level)
* Multivariate : focusing on corelation and which one to use as there is spearman,pearson.

Answer :Spearman because Pearson works with continuous interval or ratio data. Spearman is suitable for ordinal, ranked, interval, or ratio data.(geeks for geeks)

Boxplot :because we can also know

* From Univariate:

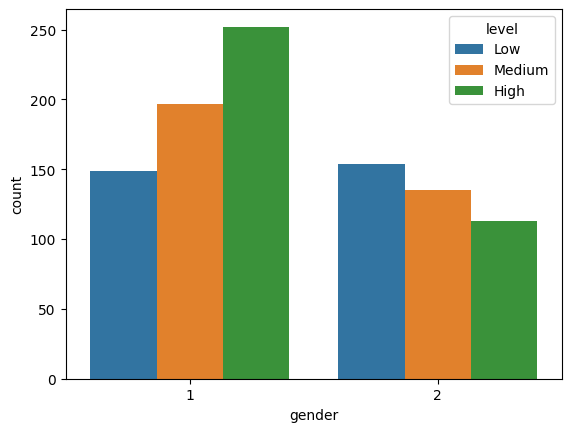
i.e level distribution: level:

High 365

Medium 332

Low 303

2. Which gender is more prone to Lung Cancer



3.From feature distribution we can say that most the features are a bimodal and normally distributed

4.Preprocessing : not really required for certain models because it categorical data

[Python error: Nonetype object not subscriptable, when loading JSON into variables - Stack Overflow](https://stackoverflow.com/questions/54581297/python-error-nonetype-object-not-subscriptable-when-loading-json-into-variable?utm_source=chatgpt.com)

raise InvalidParameterError(

...<2 lines>...

)

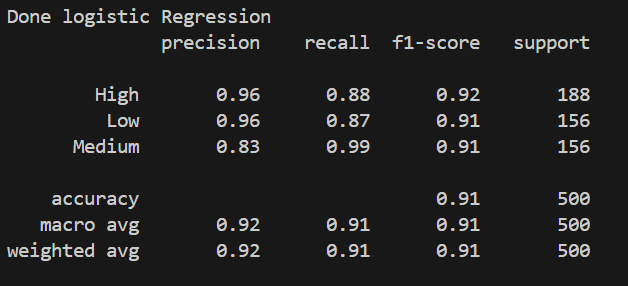
sklearn.utils.\_param\_validation.InvalidParameterError: The 'target\_names' parameter of classification\_report must be an array-like or None. Got OrdinalEncoder() instead.

5.Training:

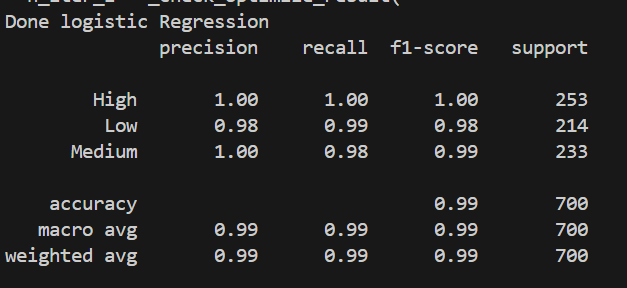
Why did choose this model:

Logistic,Random forest

With train test spilt as 0.3 and 0,7



With train test split as 0.3 and 0.7



With iter 200

