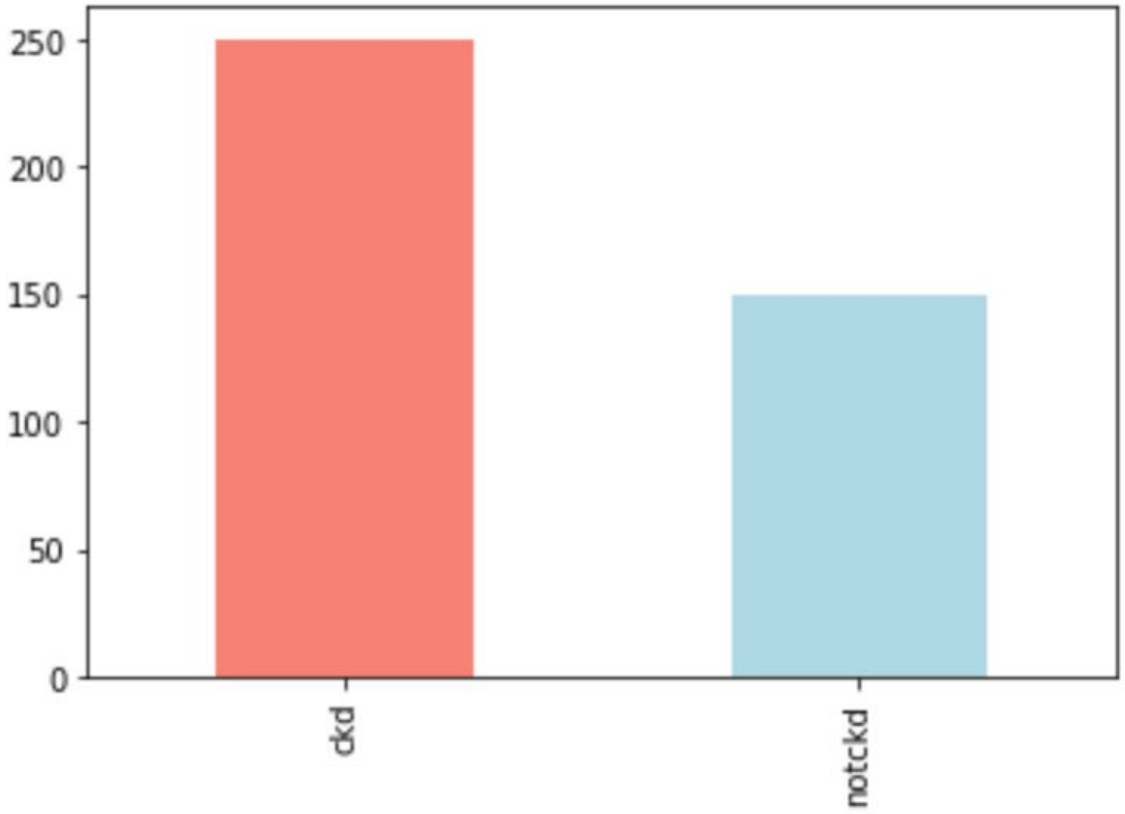
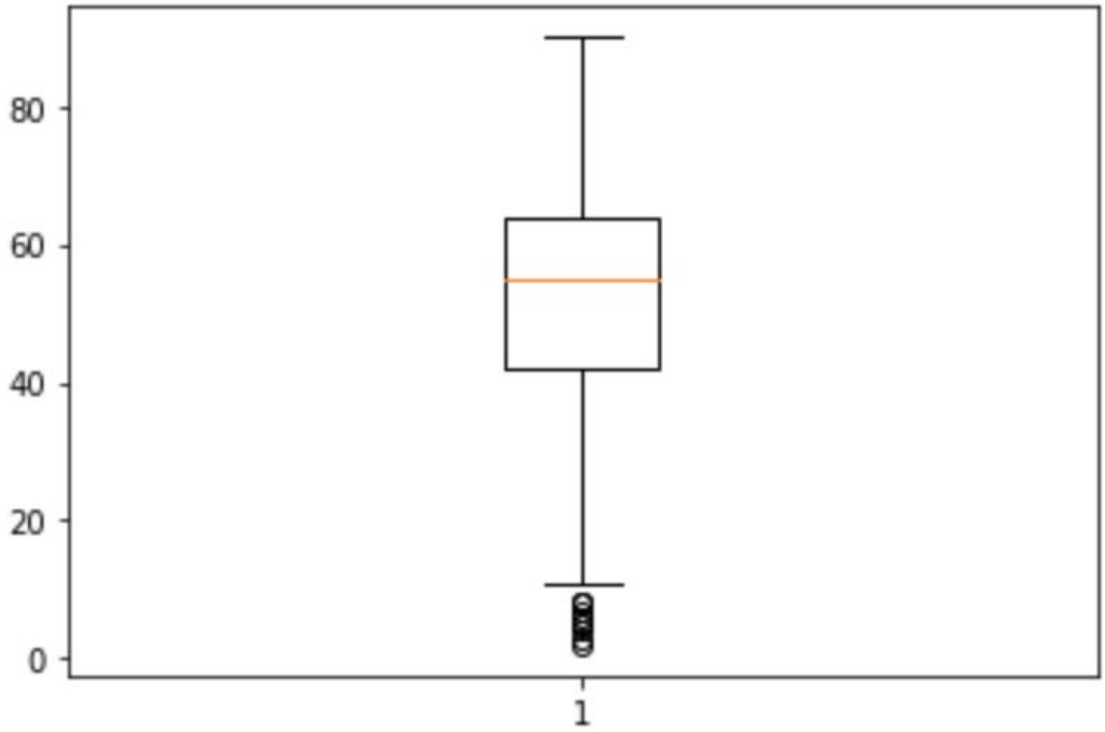
First of all we used a barplot to visualize the classification column:



With the number of patients with chronic kidney disease being 250, and the number of patients without being 150, we can clearly see that the dataset is imbalanced (the classes are not represented equally), which may make our classification model slightly biased towards patients with chronic kidney disease.

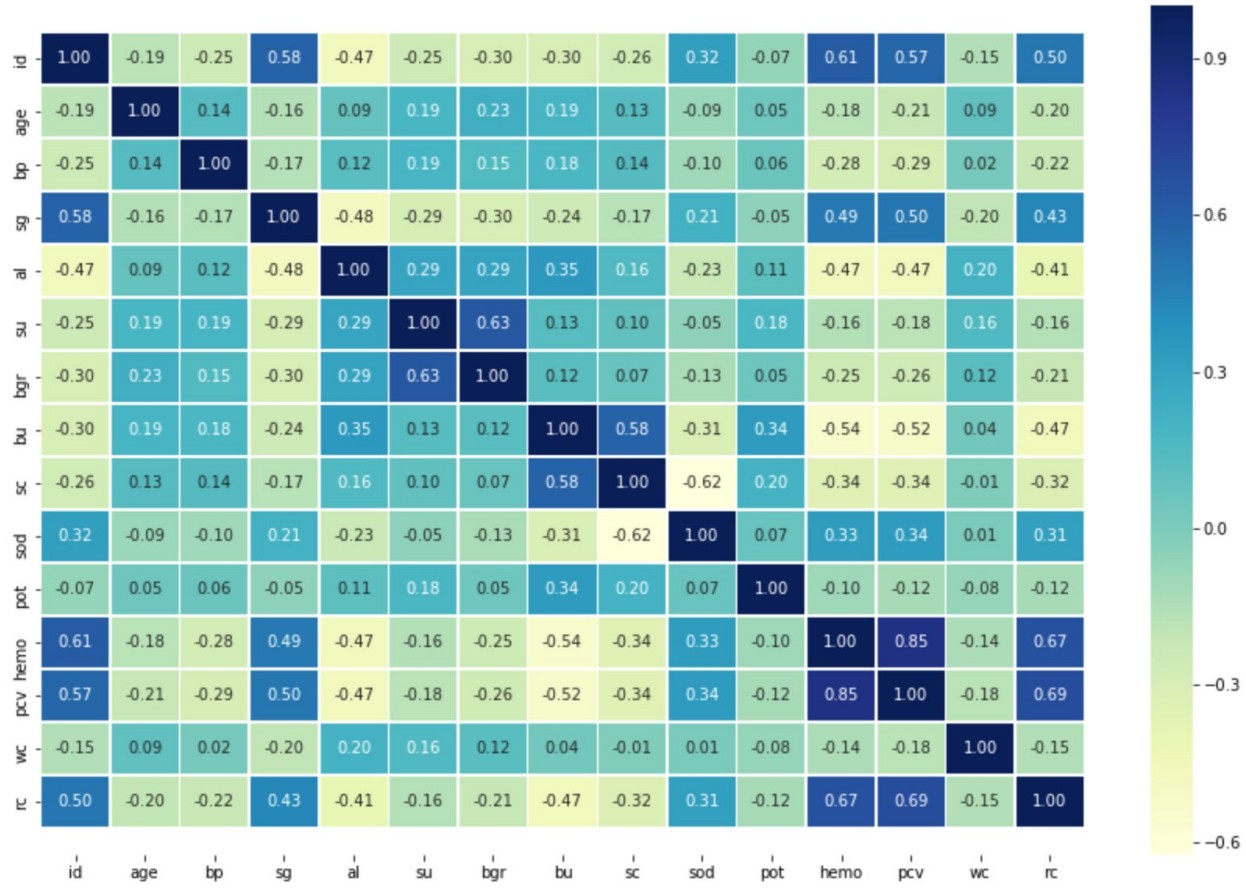
Then we used a boxplot to visualize the distribution of the ages of patients in the dataset:

```
Q0: 11.0
Q1: 42.0
Q2: 55.0
Q3: 64.0
Q4: 90.0
```



With the 0th percentile having a value of 11, it is clear that we have some outliers consisting of young children whose ages range from 2 to 10 years old.
With the first quartile having a value of 42, the second quartile 55, the third quartile 64, and the fourth quartile 90, we can conclude that most of the patients in the dataset (the middle 50%) are men between 42 and 64 years old, with the median age being 55, and the oldest patient in the dataset being 90 years old.

Then we displayed the correlation matrix of the dataset (the numerical columns):



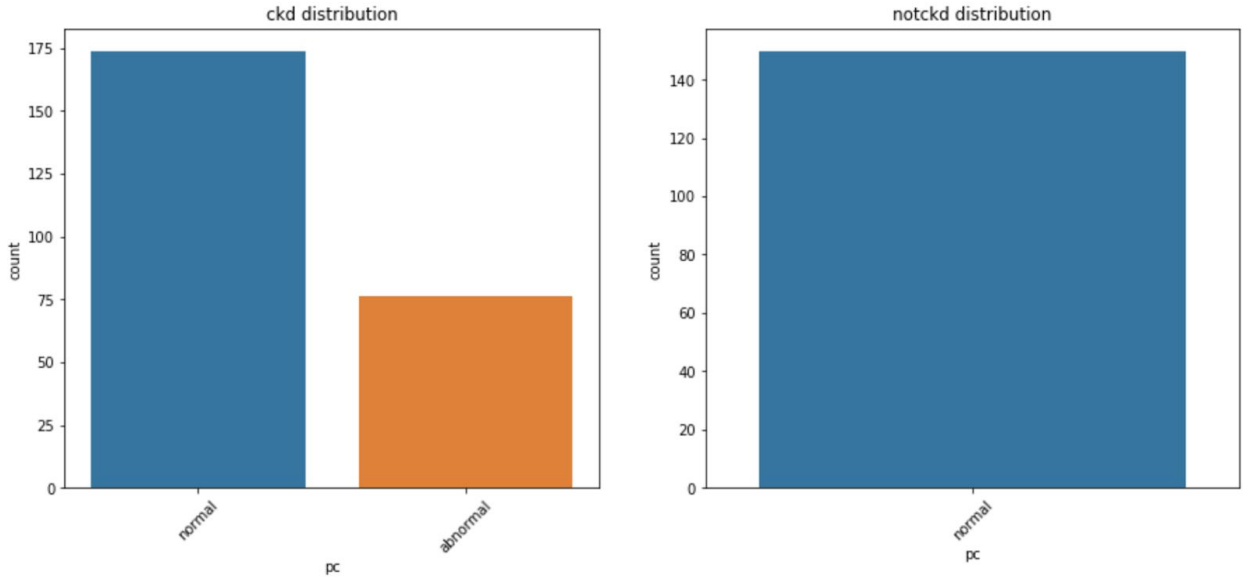
From which we can conclude the following:

1) There is a very high positive correlation (0.85) between the Packed Cell Volume (pcv) and the Hemoglobin (hemo), which makes perfect sense, and that's because a convention that has been adopted in medicine is to estimate haemoglobin concentration as a third of packed cell volume or vice versa.
2) There is a high positive correlation (0.69) between the Packed Cell Volume (pcv) and the Red Blood Cell Count (rc), which also makes sense, as red blood cells account for nearly all the cells in the blood, where the pcv rises when the number of red blood cells increases or when the total blood volume is reduced.

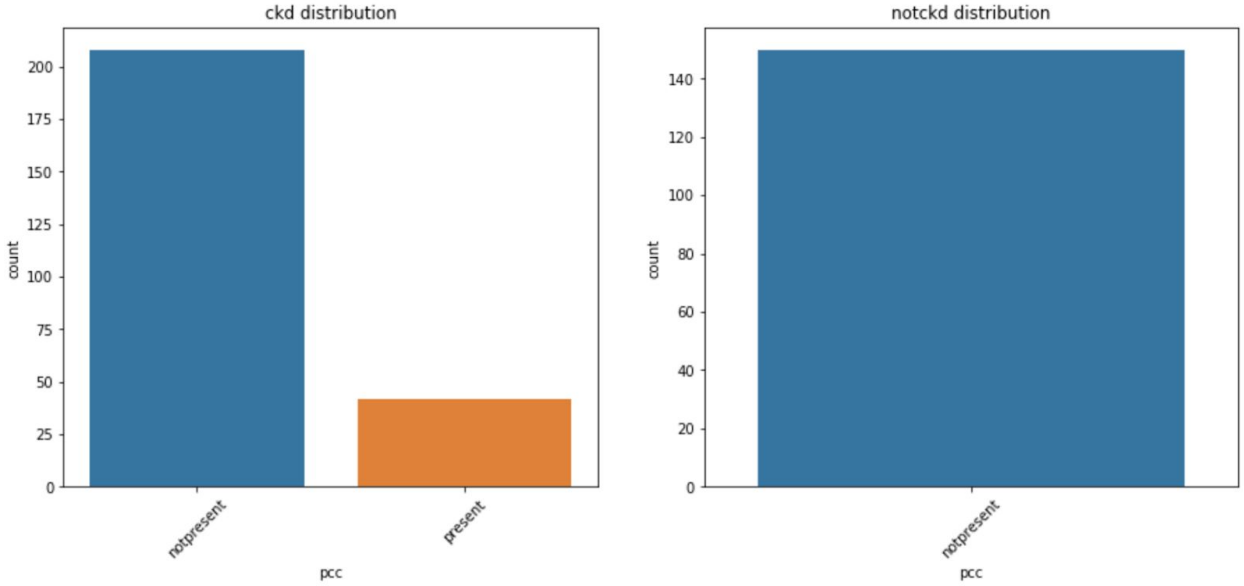
(The Packed Cell Volume (pcv) is highly dependent on the Hemoglobin (hemo) and the Red Blood Cell Count (rc))

Then we visualized all the nominal attributes for patients with chronic kidney disease (ckd) and patients without chronic kidney disease (notckd):

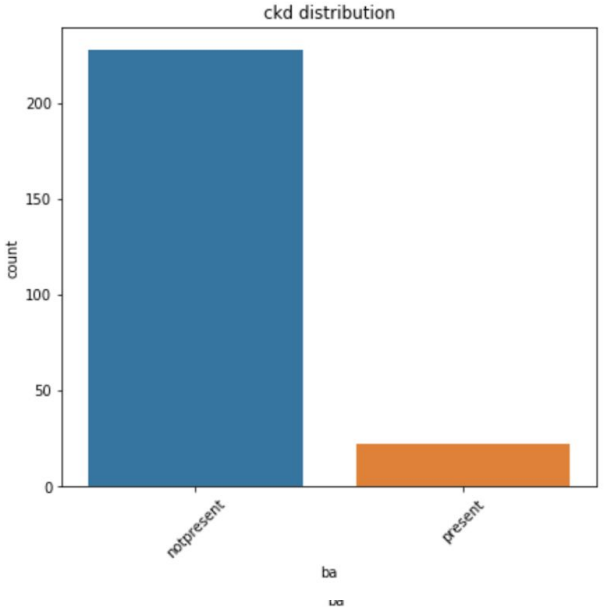
pc

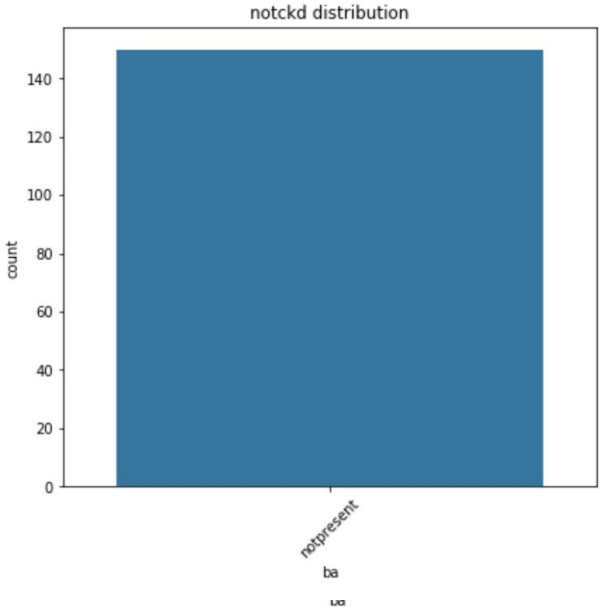

ckd distribution
count
pc
normal
abnormal
notckd distribution
count
pc
normal

pcc

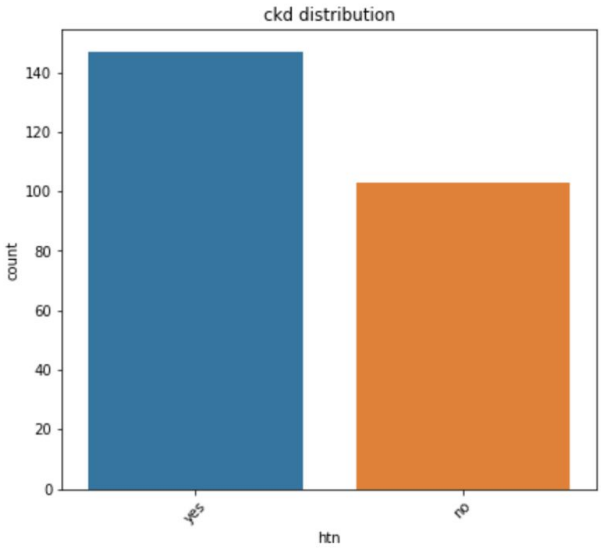

ckd distribution
count
pcc
notpresent
present
notckd distribution
count
pcc
notpresent

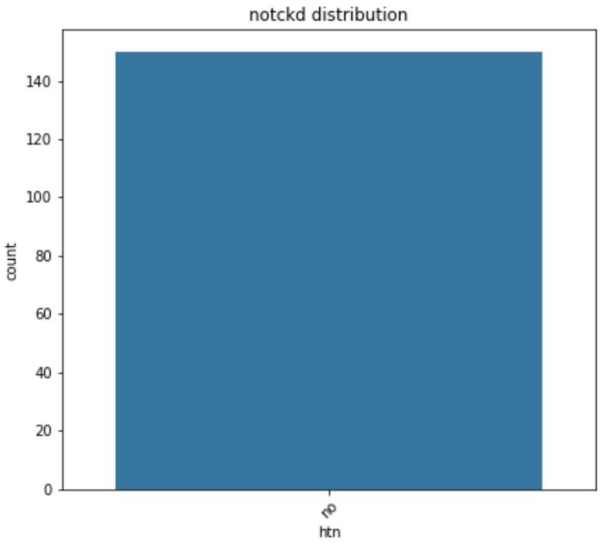
ba


ckd distribution
count
0
50
100
150
200
notpresent
present
ba

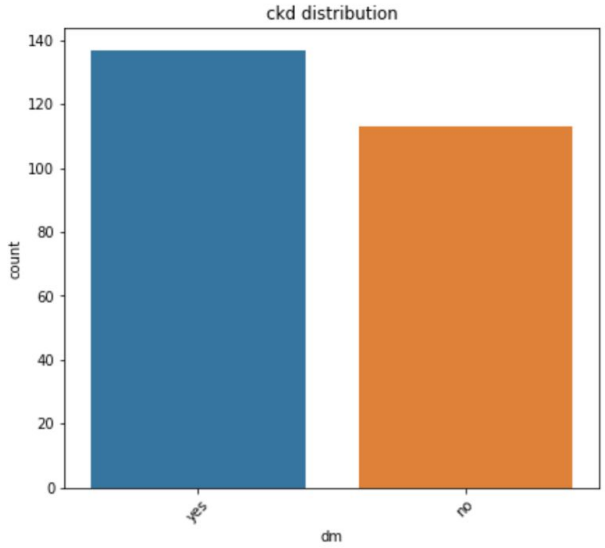

notckd distribution
count
0
20
40
60
80
100
120
140
notpresent
ba

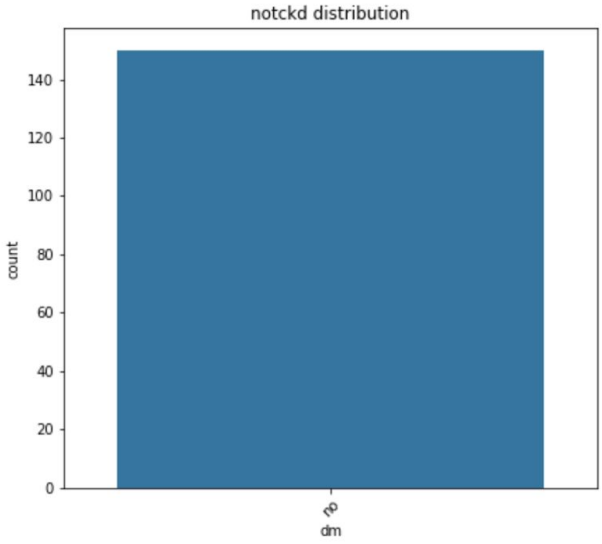
htn


ckd distribution
count
0
20
40
60
80
100
120
140
yes
no
htn

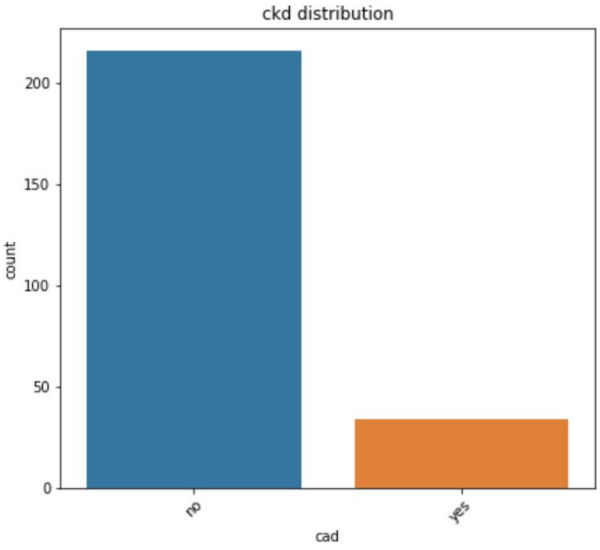

notckd distribution
count
0
20
40
60
80
100
120
140
no
htn

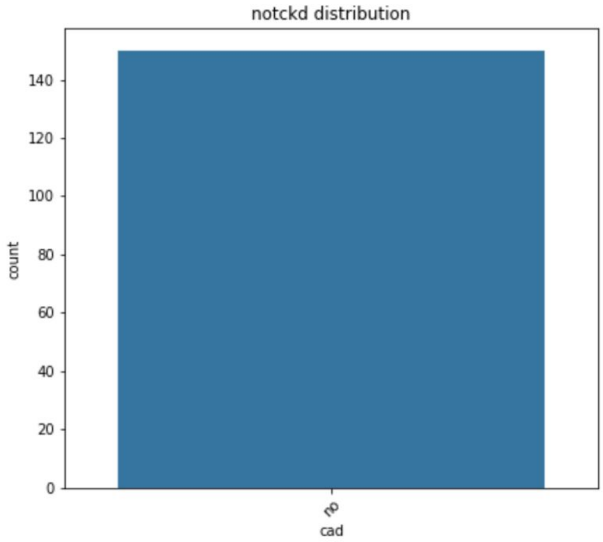
dm

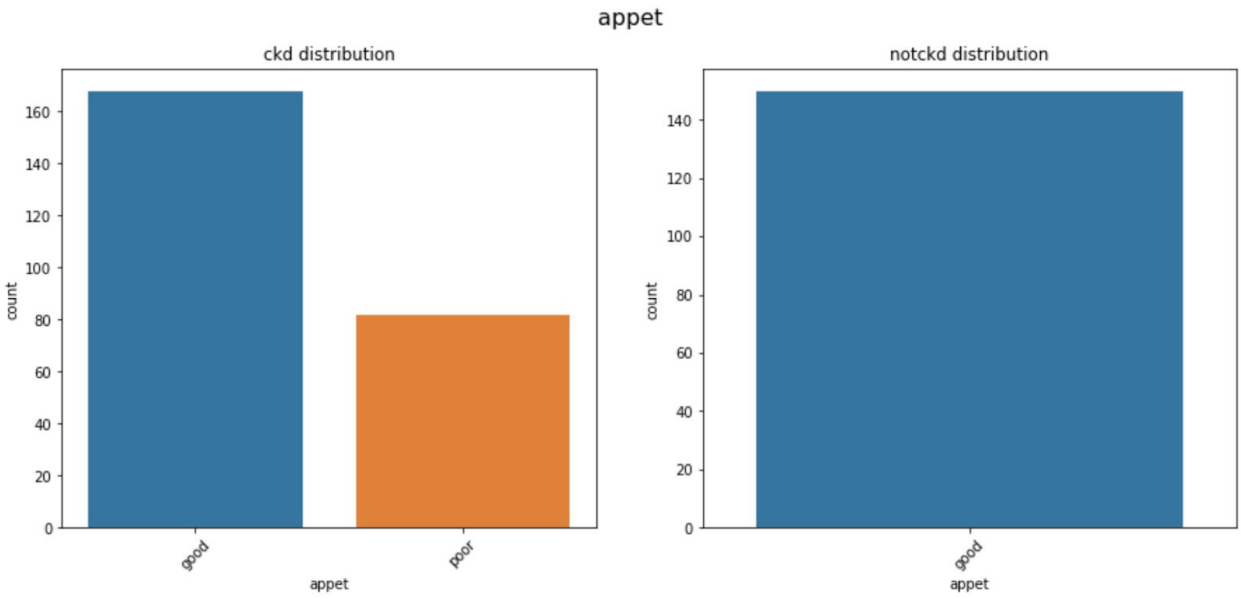

ckd distribution
count
dm
yes
no

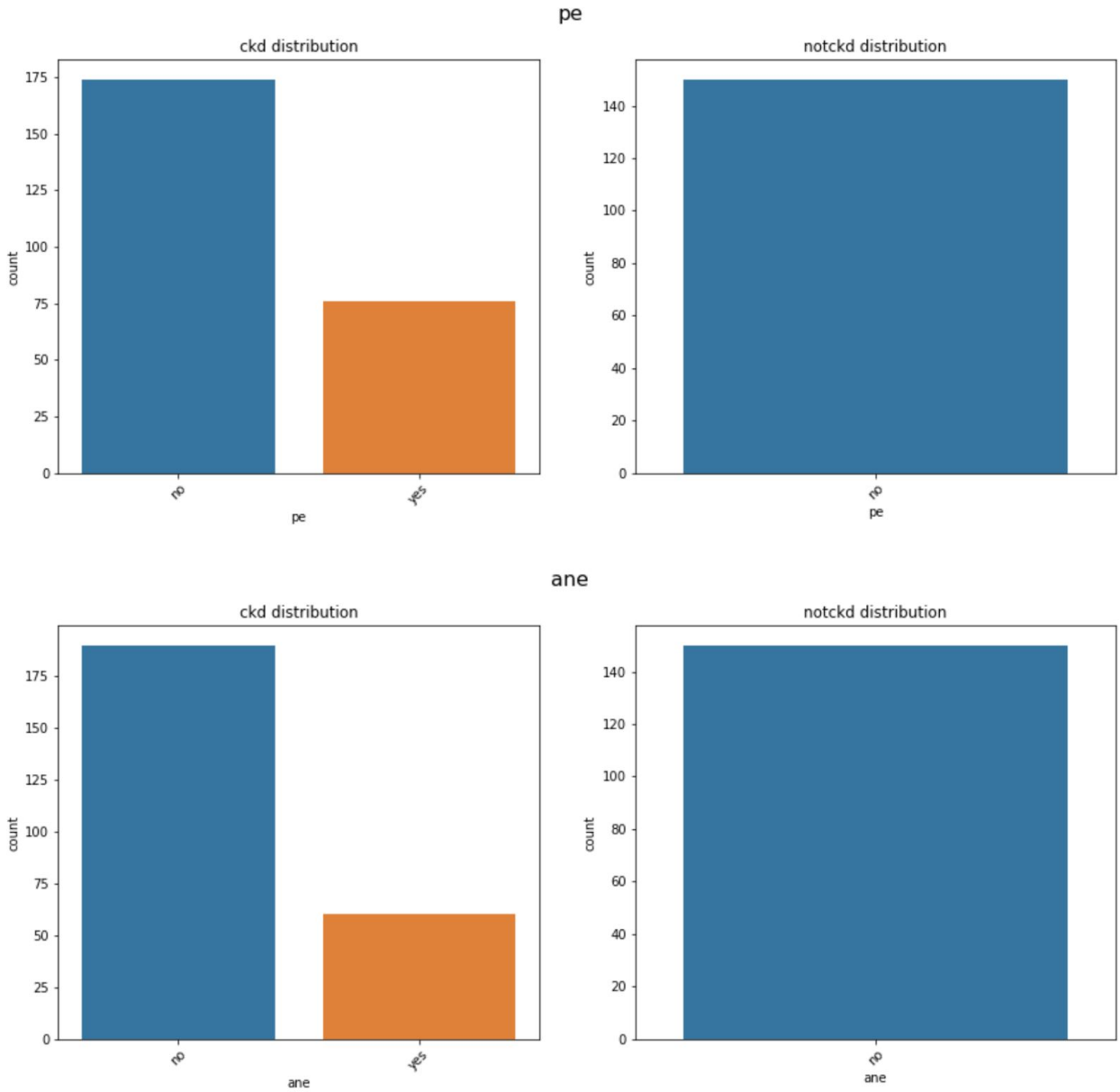

notckd distribution
count
dm
no

cad


ckd distribution
count
cad
no
yes


notckd distribution
count
cad
no


appet
ckd distribution
notckd distribution
count
appet
good
poor
0
20
40
60
80
100
120
140
160



From which we can conclude that if the patient has an abnormal Plus Cell (pc), or has Plus Cell Clumps (pcc), Bacteria (ba), Hypertension (htn), Diabetes Mellitus (dm), Coronary Artery Disease (cad), poor Appetite (appet), Pedal Edema (pe), or Anemia (ane), then he has Chronic Kidney Disease (ckd), which explains why we got perfect accuracy in two of our classification models.