Horse Colic Analysis Report

Prepared by:

Motaz hashhoush

Horse Colic Analysis Report....0
Business Problem....2
Research Goal:-....2
Preprocessing....3
Step 1: Data Understanding....3
Step 2: Merge Features....3
Step 3: Handling Missing Values....3
Step 4: Identifying and Removing Outliers....3
Exploratory Data Analysis (EDA)....5
Visualization Categorical features....5
Age....5
Surgery....6
Temperature of extremities....6
Capillary Refill Time....7
Pain....8
Peristalsis....8
Abdominal Distension....9
Rectal Exam Feces....9
Abdomen....10
Surgical Lesion....10
Analysis, Visualization of the Numerical Features....11
Pulse....11
Packed cell volume....11
Total protein....12

Business Problem

A farmer noticed that the death rate of horses is increasing for reasons that are not clear, and he gave us a list of all the horses on the farm, followed by a number of characteristics.

Research Goal:-

Reasons that increase the possibility of horse death.

Preprocessing

Step 1: Data Understanding

Before beginning the preprocessing stage, it is important to analyze the data to gain a better understanding of its characteristics. In this initial step, we examined the data type for each column, identified any null values, and identified potential outliers. Based on this analysis, we developed a plan for addressing any issues we identified. This may include dropping columns with a high percentage of missing values or using filling methods for columns with a lower percentage of missing values.

Step 2: Merge Features

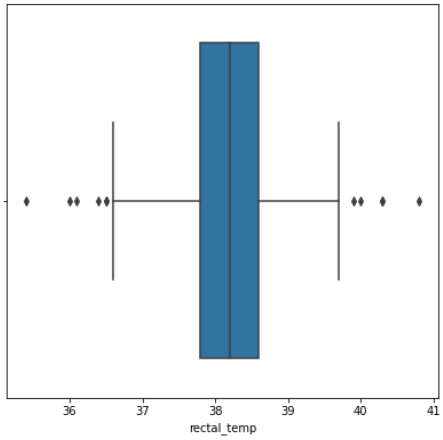
The features lesion 1, 2, and 3 will give us trouble because the way they are formatted doesn't make any sense. Therefore, we will create a new feature called **num_lesions** that counts the number of lesions the horse has to replace them.

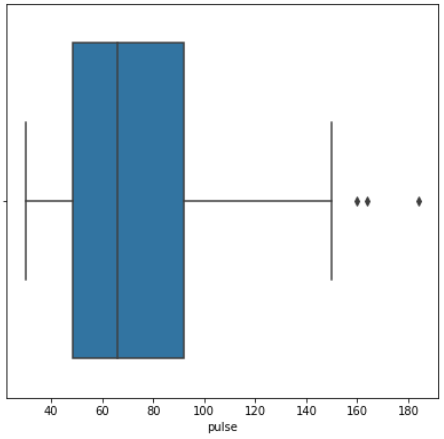
Step 3: Handling Missing Values

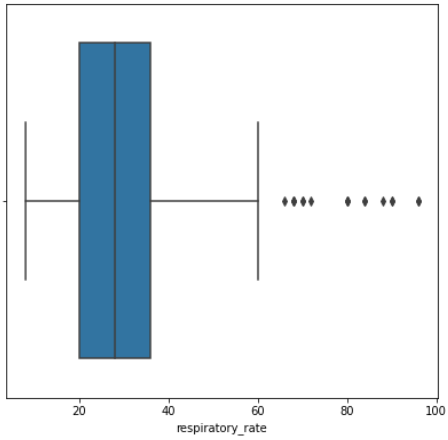
To address missing values in the data, we separated the columns into two categories: object (categorical data) and int64, float64 (numeric data). Each type required a different approach for handling missing values. For numeric data, we used the k-nearest neighbor algorithm to fill in missing values. This machine learning algorithm is able to provide estimates that are likely to be close to the true values. For categorical data, we used the **mode**, which fills missing values with the most common value in the column.

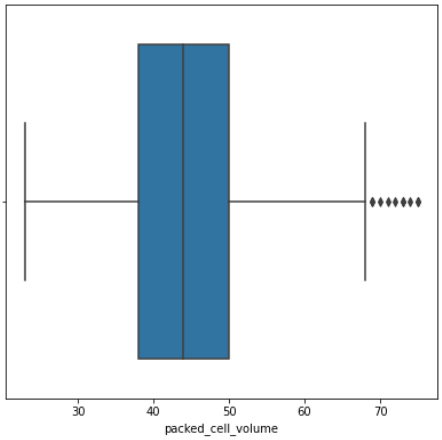
Step 4: Identifying and Removing Outliers

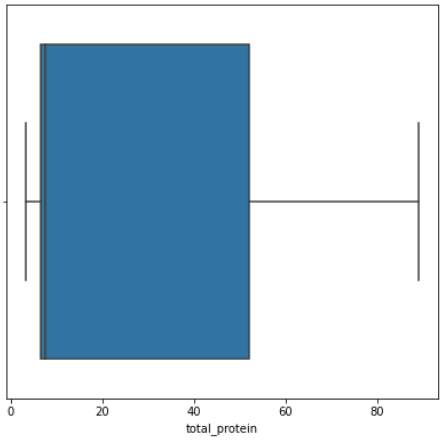
In order to identify and address outliers in our numeric data, we utilized box plots. After carefully reviewing the data, we determined which columns were suitable for outlier removal and which were not. To remove any identified outliers, we employed the quantile technique. This allowed us to effectively address any outlying values that may impact our analysis.

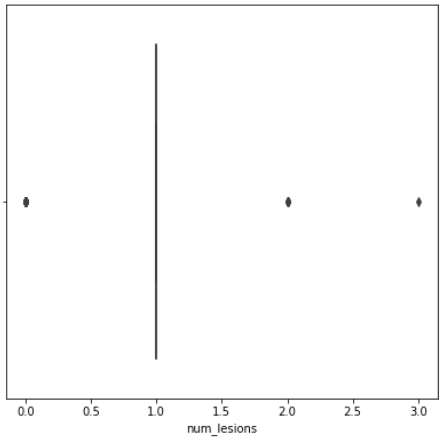












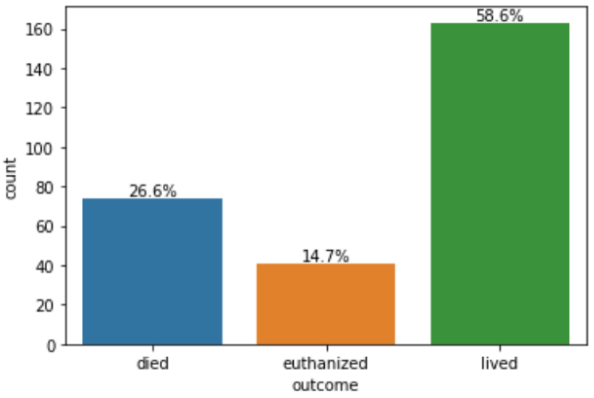
we took care not to remove outliers from certain columns, such as:

- rectal_tem Because: Because:
  - An elevated temp may occur due to infection.
  - temperature may be reduced when the animal is in late shock
- pulse; Because:
  - animals with painful lesions or suffering from circulatory shock may have an elevated heart rate
- packed_cell_volume; Because:
  - the number of red cells by volume in the blood
  - normal range is 30 to 50. The level rises as the circulation becomes compromised or as the animal becomes dehydrated.
- num_lesions; Because:
  - number of lesions between 0 and 3

However, we remove outliers from certain columns, such as

- respiratory_rate; because:
  - normal rate is 8 to 10
  - usefulness is doubtful due to the great fluctuations

Exploratory Data Analysis (EDA)

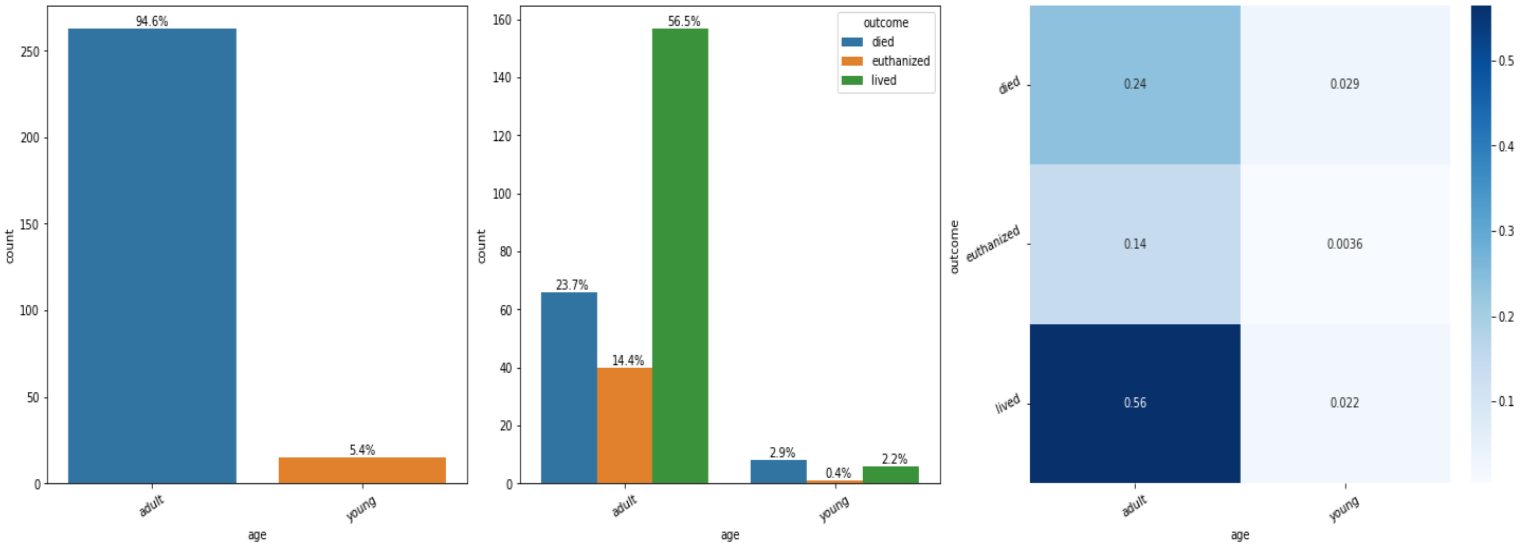


in general, we have 58.6% of horses lived, 14.7% of horses were euthanized, and 26.6% of horses died.

Visualization Categorical features

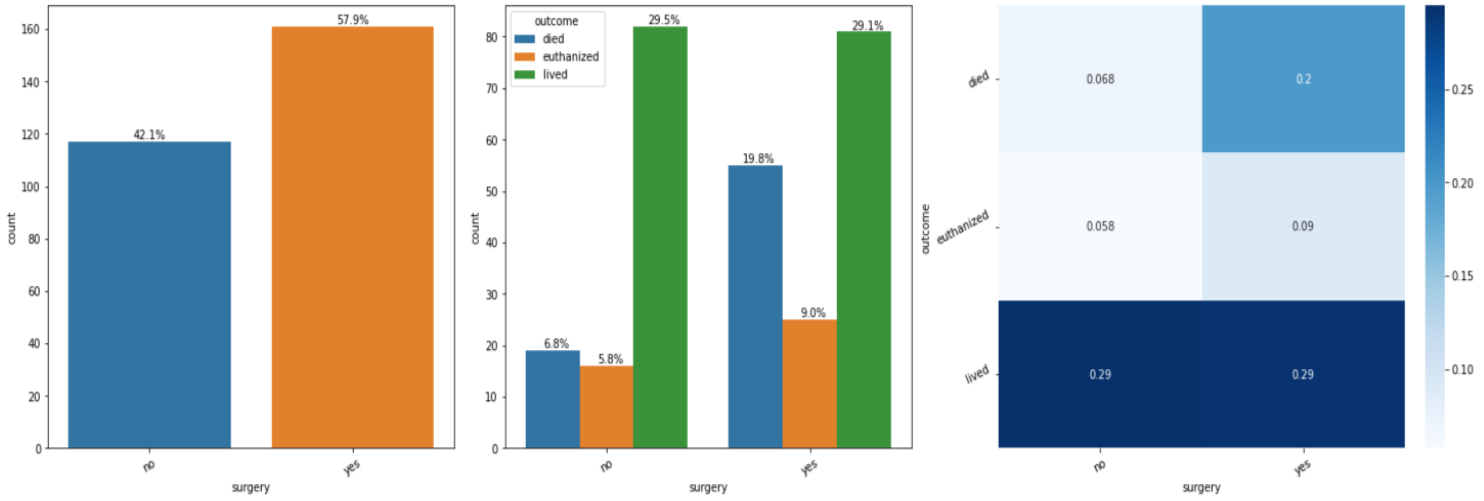
In this visualization, we focus on the increase in horses' deaths. By examining the bar chart, we can see the locations where horses lived, were euthanized, and died. From this information, we have drawn the following conclusions:

Age



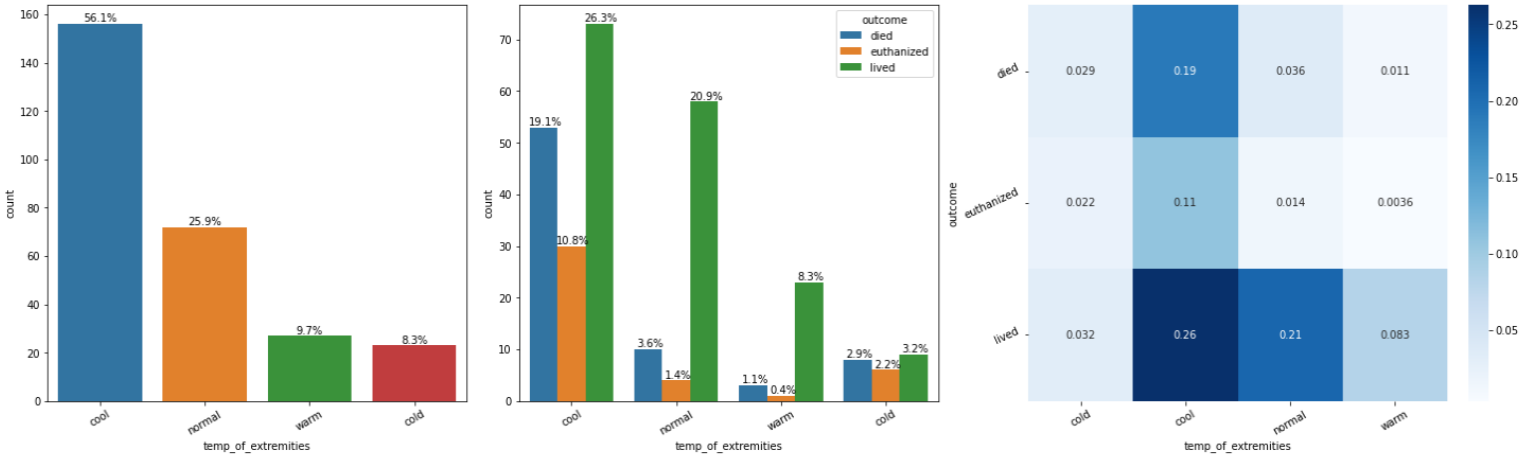
The age of the horse is an important factor in the possibility of its death, But as we can see from our sample, it seems unbalanced so no actual decision can be made on it

Surgery



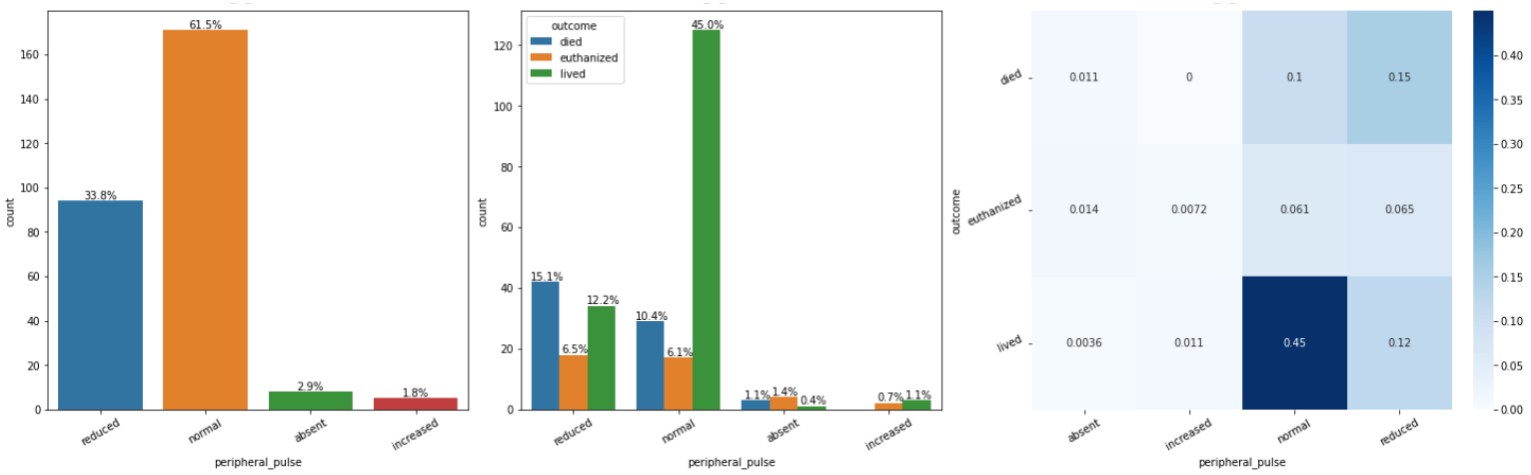
Surgery has a significant impact on whether a horse lives or dies. Both percentages were close in terms of living probability, but we can see that the chances of a horse death increased when a surgery was performed. In cases where surgery was not applied, the chance of horses death was very low.

Temperature of extremities



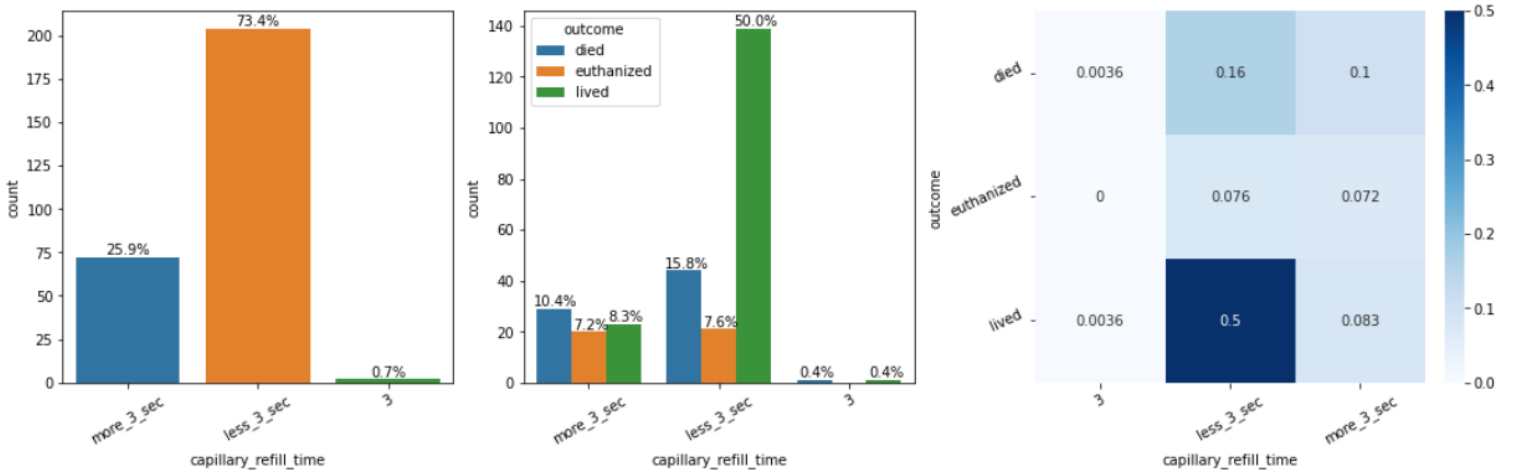
Temperature also appears to have an effect on a horse's likelihood of dying. Colder temperatures have a higher probability of causing death,

Peripheral pulse



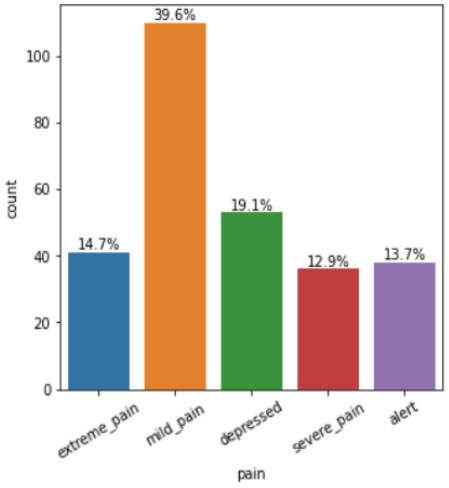
Pulse also affects a horse's chances of living or dying. We can see that more horses died at **reduced pulse**,

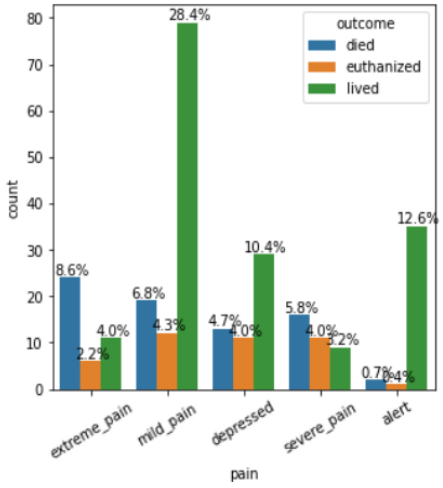
Capillary Refill Time

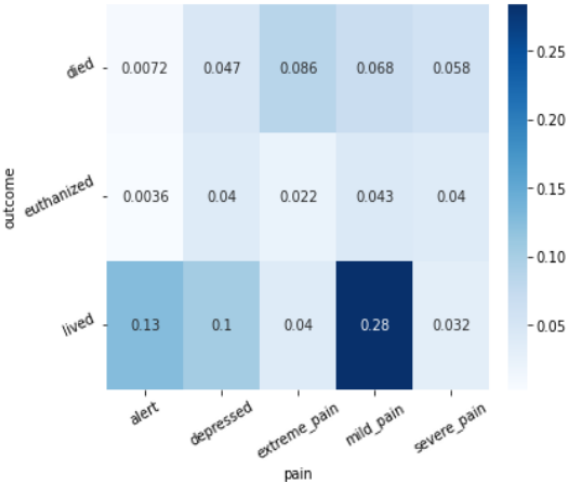


The capillary refill time, which is the time it takes for blood to flow through the blood vessels, also plays a role in a horse's likelihood of survival. If it takes more than 3 seconds, the horse has a higher chance of dying, as it indicates a slow blood flow (Cardiac weakness).

Pain

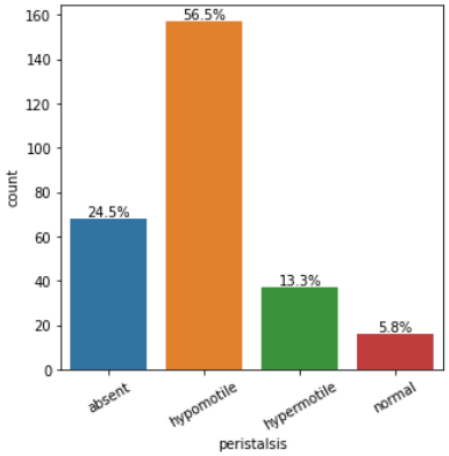


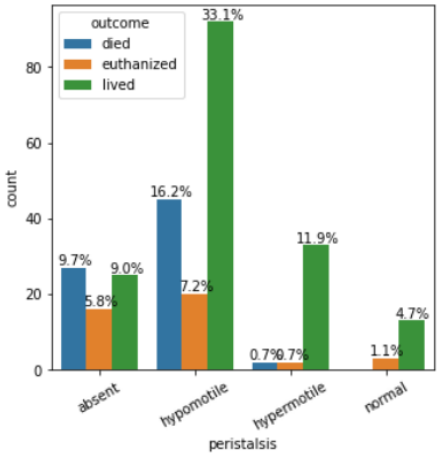


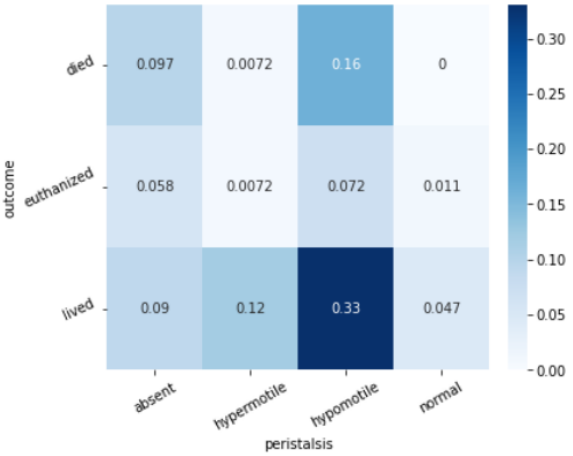


Pain levels also affect a horse's chances of survival. If a horse is in a state of **alert** or **mid-pain**, it has a higher chance of surviving, but if it is in **extreme-pain**, it has a higher chance of dying.

Peristalsis

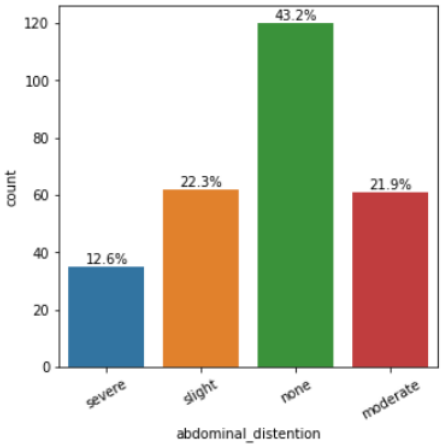


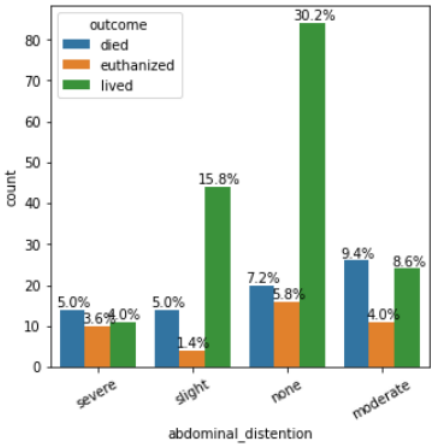


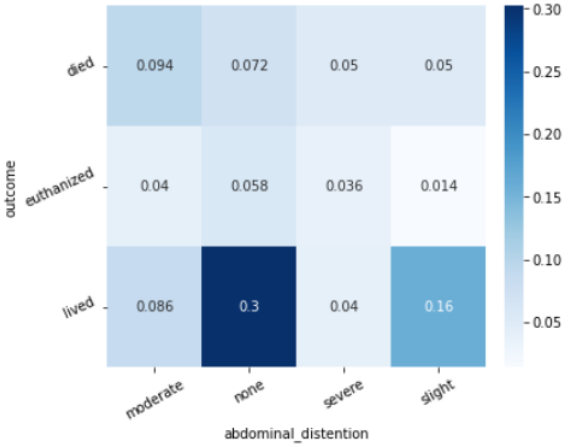


Peristalsis, which refers to the activity in a horse's gut, has an impact on survival rates if it is **absent**, the horse has the highest likelihood of dying, while a **normal** level of peristalsis almost guarantees survival.

Abdominal Distension

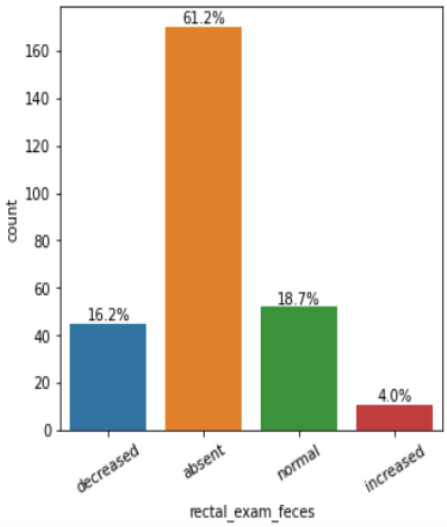


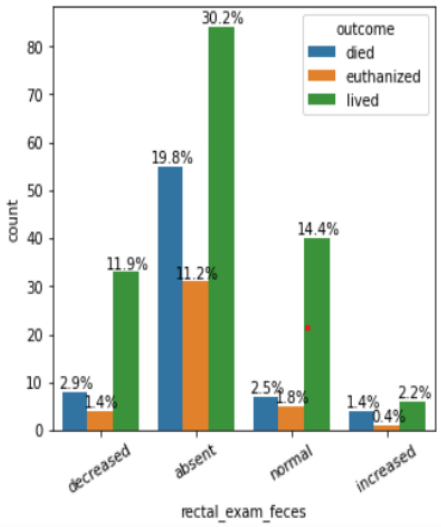


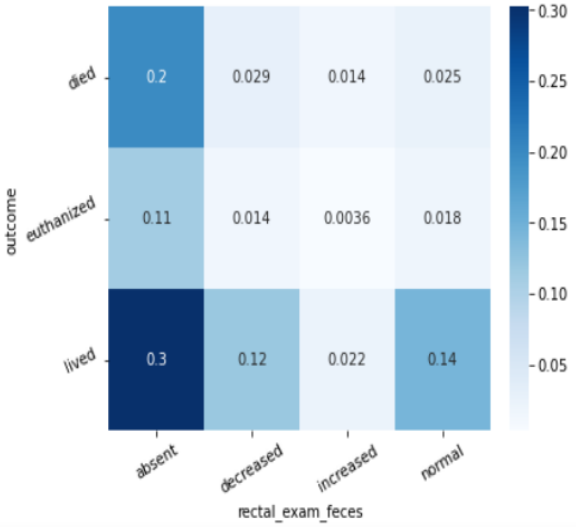


Abdominal distension, or the swelling of the abdomen, also plays a role in a horse's survival. If it is **severe** or **moderate**, the chances of the horse dying are higher than surviving. However, if the distension is none or slight, the chances of survival are higher.

Rectal Exam Feces

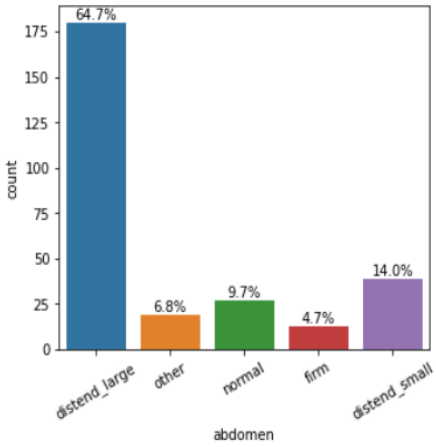


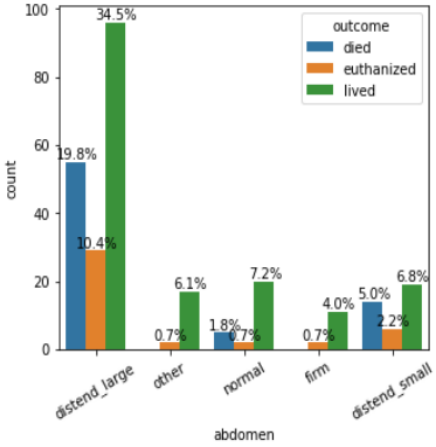


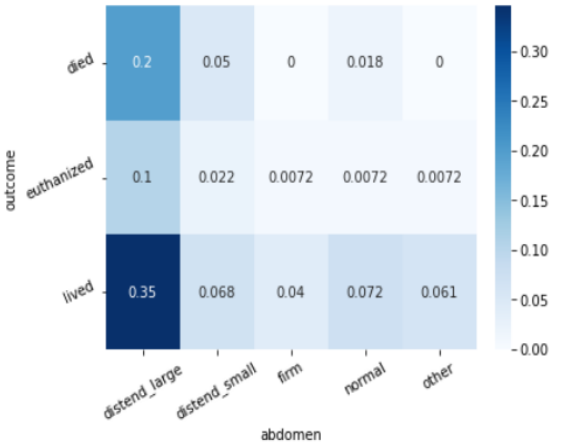


The amount of feces present during a rectal exam also affects a horse's chances of survival. If feces are increased or absent, the horse has a higher chance of dying, with the highest chance belonging to the category of increased feces. On the other hand if the feces are decreased or normal the chance of surviving are increased.

Abdomen

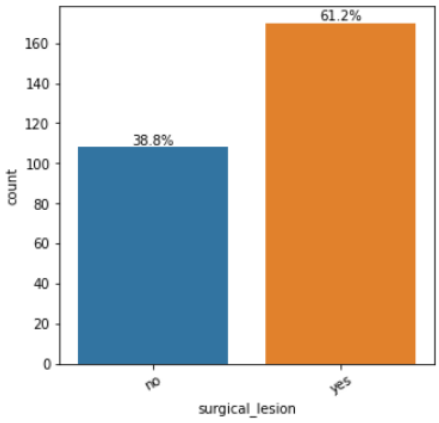


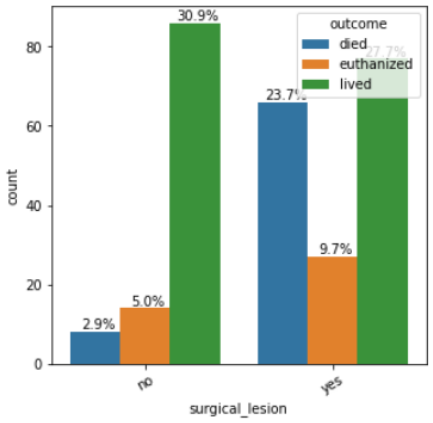


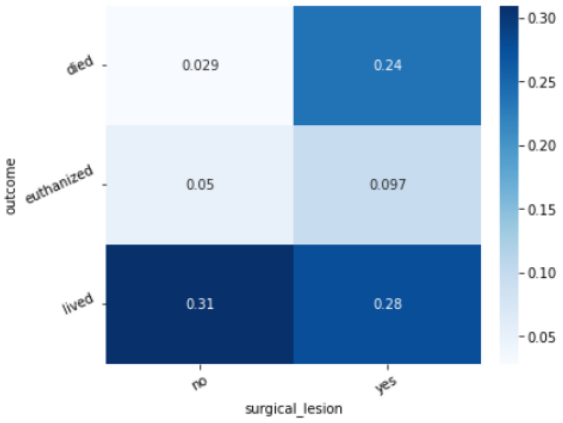


The condition of the abdomen, which refers to the area of the body containing the pancreas, stomach, intestines, liver, gallbladder, and other organs, also plays a role in survival. In our data, a distended small abdomen indicates that the horse has the highest chance of dying, even though the overall chance of horses living in that area is high.

Surgical Lesion



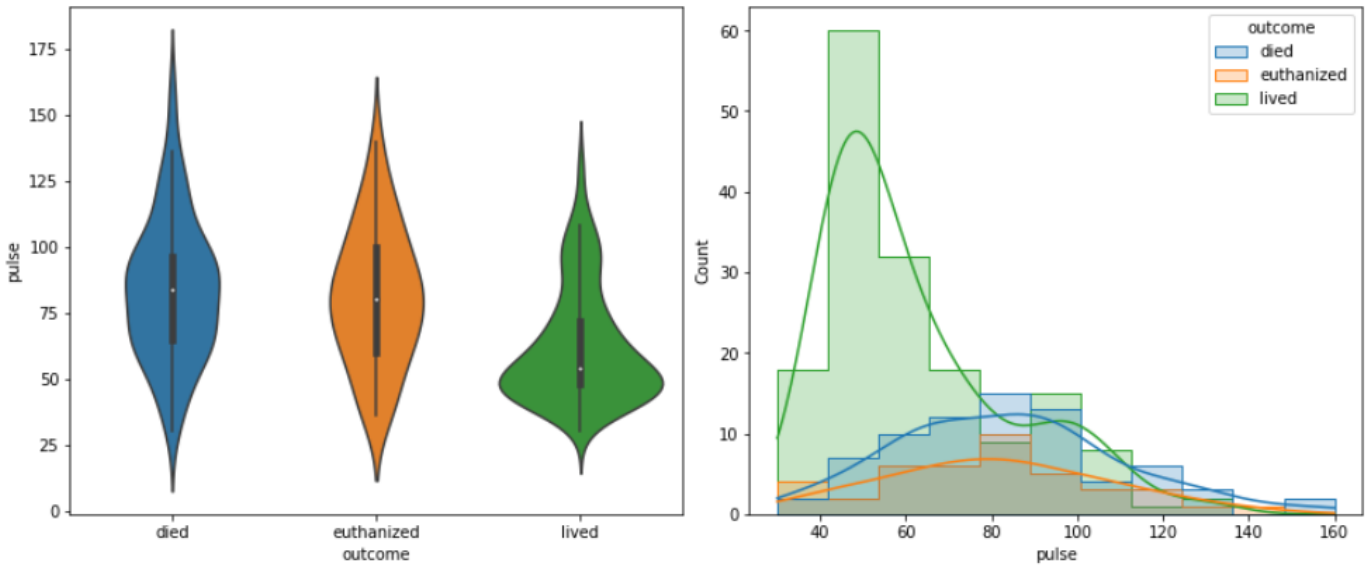




The presence of a surgical lesion refers to whether the problem is surgical or not. In this case, if the answer is "Yes," the horse has the highest chance of dying in that field,

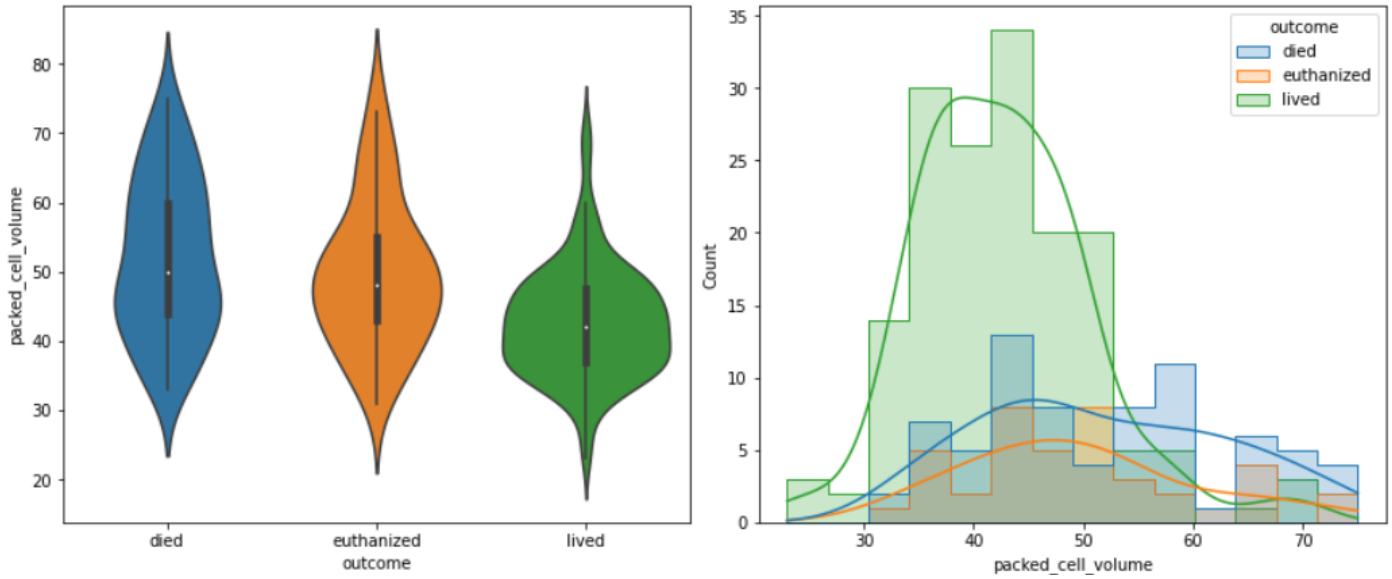
Analysis, Visualization of the Numerical Features

Pulse



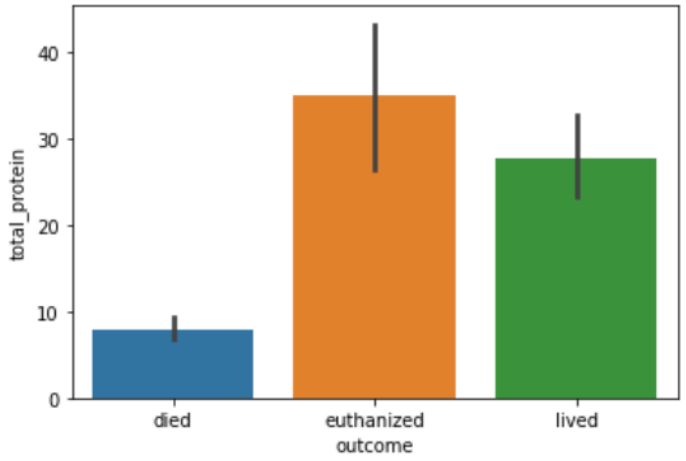
It seems that after approximately 70 of pulse, the horse is likely to die (died or be euthanized).
my advice: Follow the pulse process constantly because it is an indication that the horse has problems or is normal

Packed cell volume



It seems that after approximately 50 of **Packed cell volume**, horses begin to die (died or be euthanized).

Total protein



It seems to us that the horses that died took little **total protein**, and the horses that were euthanized got a lot of **total protein**, while the horses that lived had a moderate **total protein**.

The End