**Applications of Explainable Machine Learning in Prediction Tasks**

1. **Project Team Members:**

Yuqi He yh3653

Donglizhen Shi ds4224

Helena Hu xh2626

Wendy Wang cw3539

1. **Data**

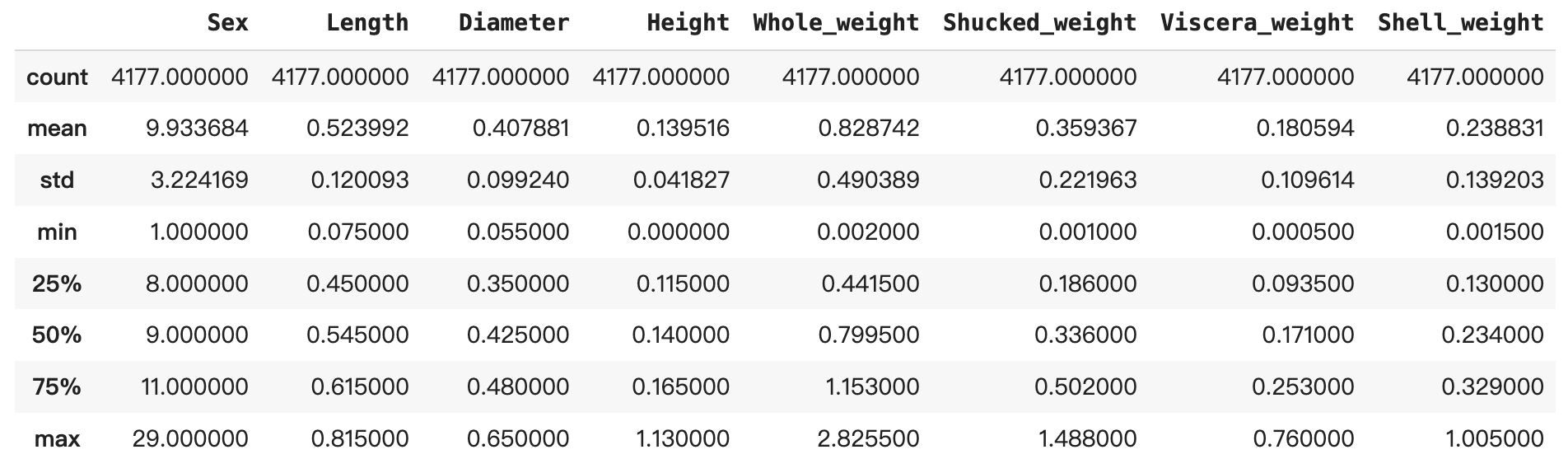
We want to first briefly address our project objective to make sense of our data choices. We plan to discover the applicability of Explainable Machine Learning tools on different tasks, especially ones that involve prediction (accuracy matters).

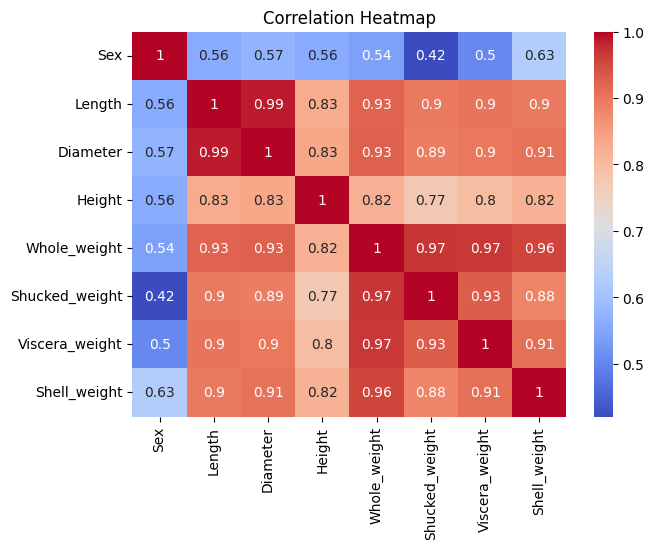
Several datasets will be used for prediction tasks. We are still updating the list, but we think the 3 datasets we have now is a good baseline (meaning we will probably only add to this list).

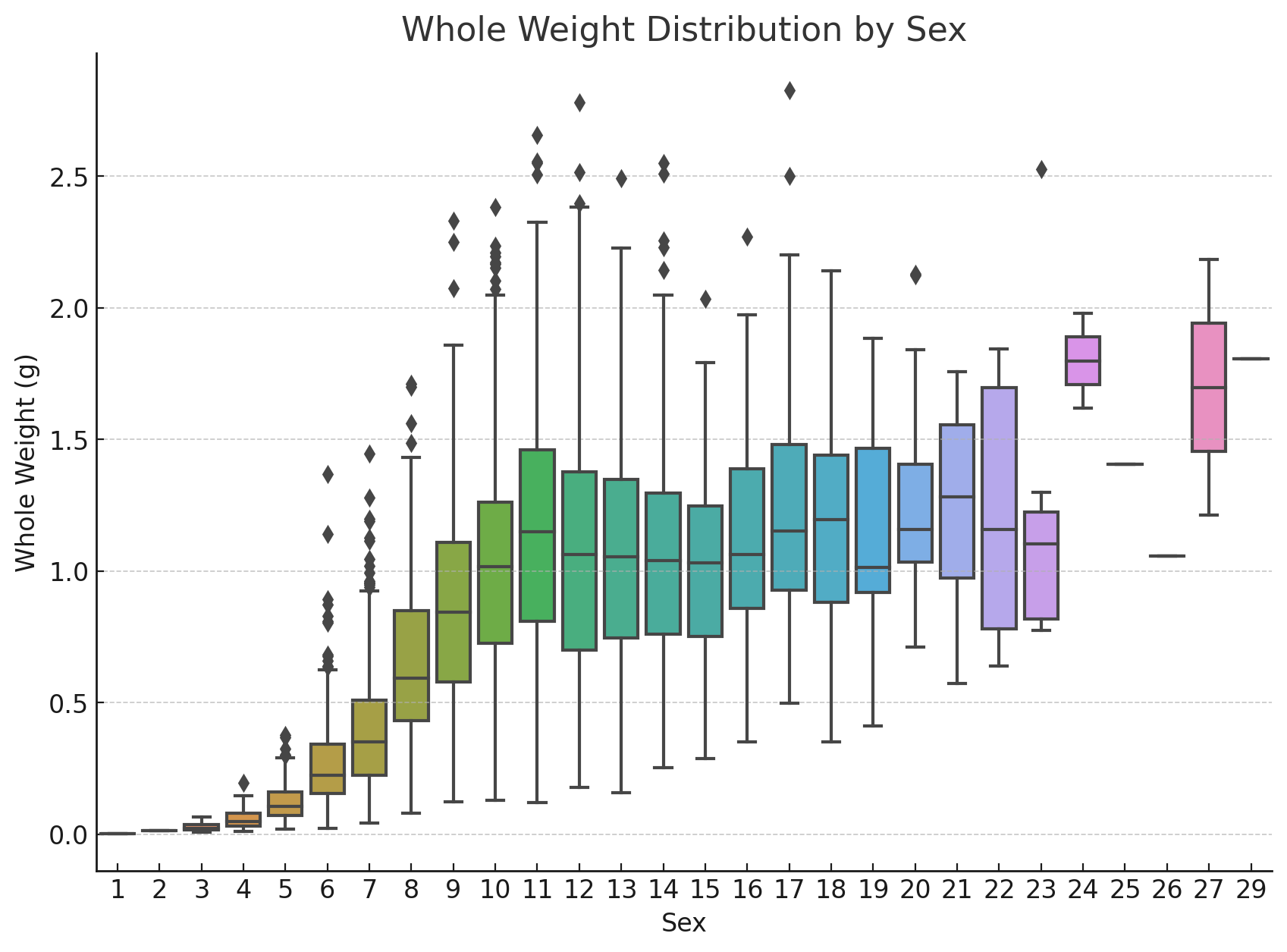
The first dataset is a baseline dataset that already was analyzed with our chosen Explainable ML tools. Abalone.csv contains 8 variables for 4176 abalones. The task itself is to distinguish their sex (Male, Female, Intersex).

The second dataset is Iris species. Iris.csv includes three species of 150 iris: Setosa, Versicolor, and Virginica. The models predict the species of each flower by analyzing four features—sepal length, sepal width, petal length, and petal width.

The third dataset is credit classes. Credit.csv predicts credit ratings categorized as "good" or "bad" using 20 features.

1. **Descriptive Statistics and Graphics**
   1. Abalone

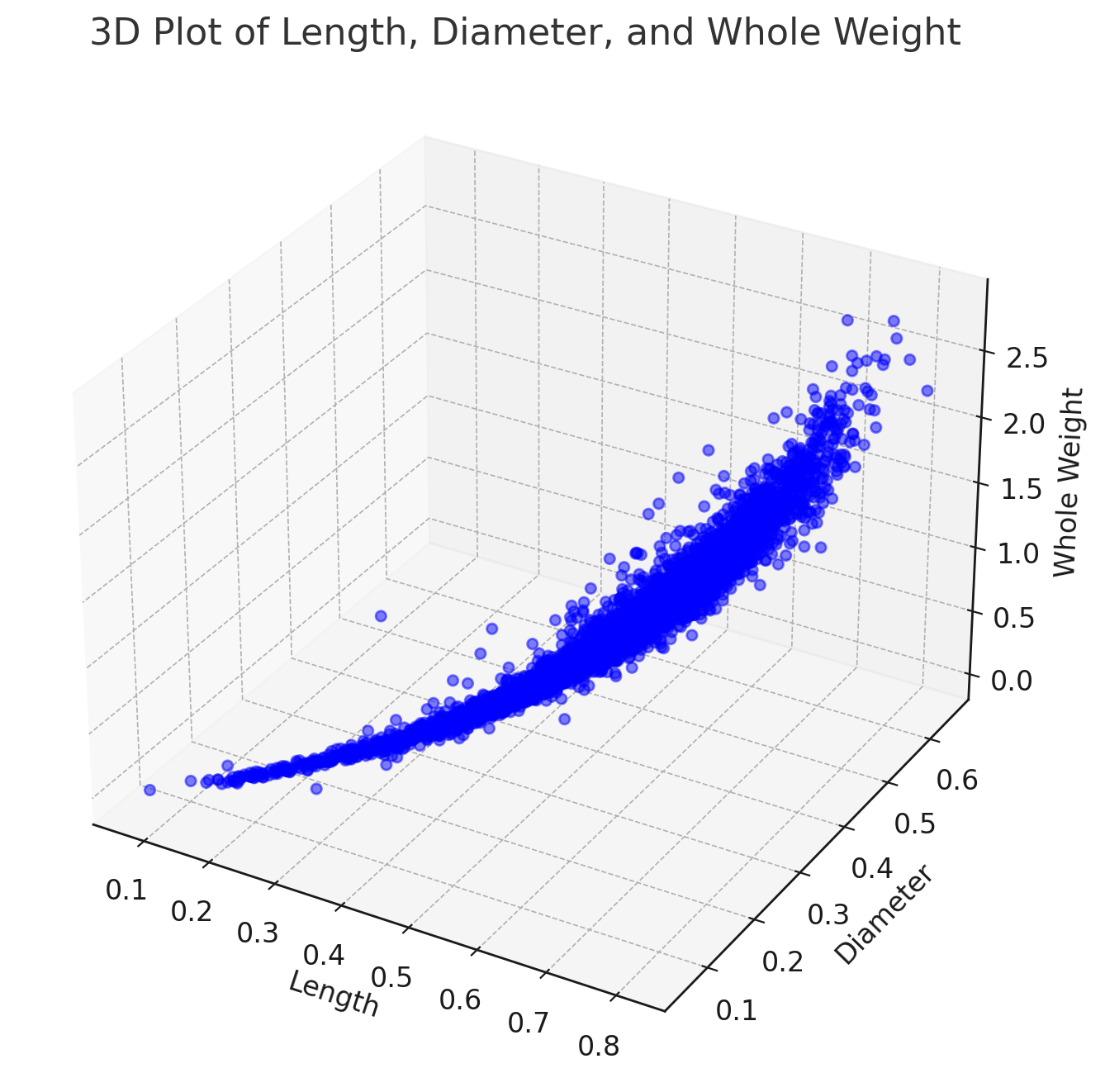


The correlation heatmap shows there are strong positive relationships between all variables except sex, especially among whole weight and other variables. We need to be aware that strong correlations might lead to multicollinearity.

The boxplot overall shows whole weights increase as the number that represents sex becomes larger. We would get more information after mapping the sex.



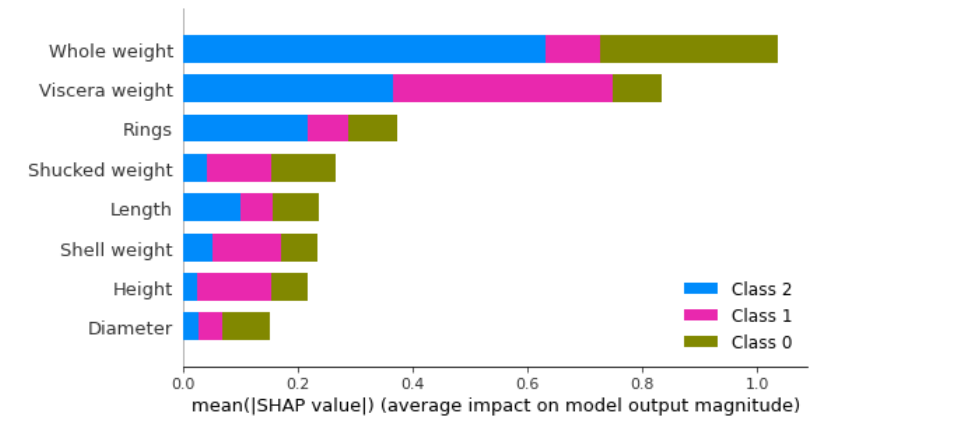
The scatterplot shows a positive relationship between shucked weight and viscera weight, but there isn't a distinct cluster of sex, suggesting there could be overlapping characteristics between sexes.



The trend illustrates a positive relationship between size (length and diameter) and weight.

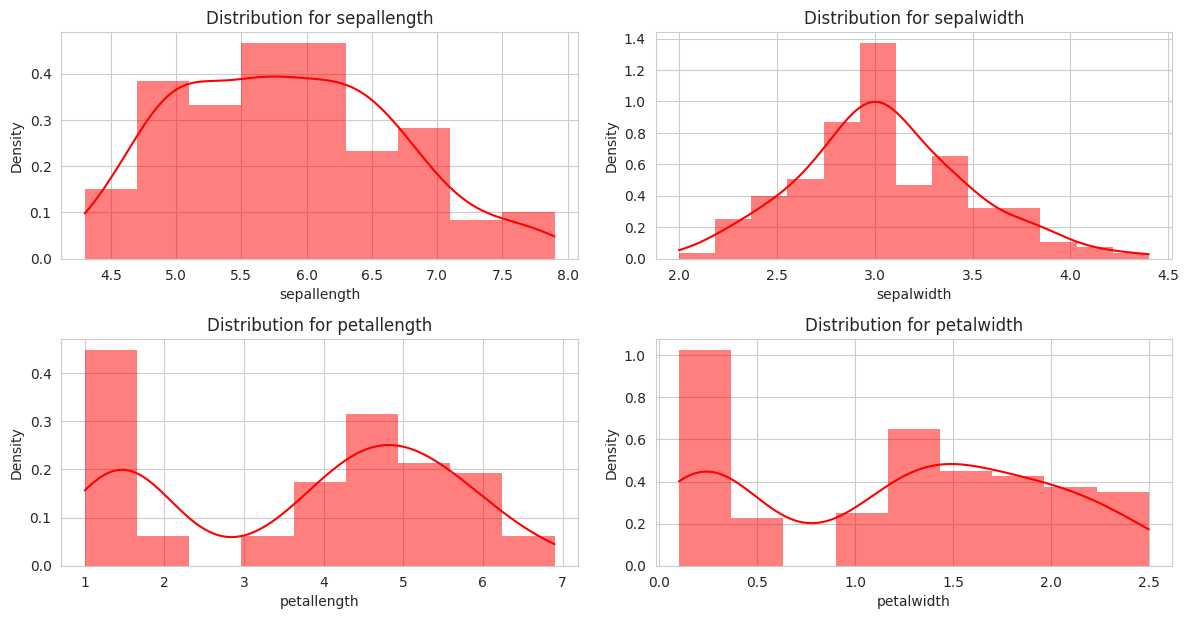


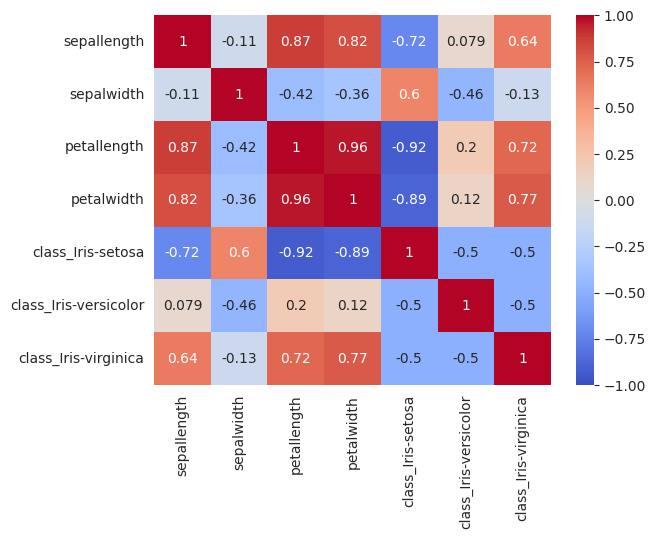
LIME output of one of the instances



SHAP feature importance

* 1. Iris

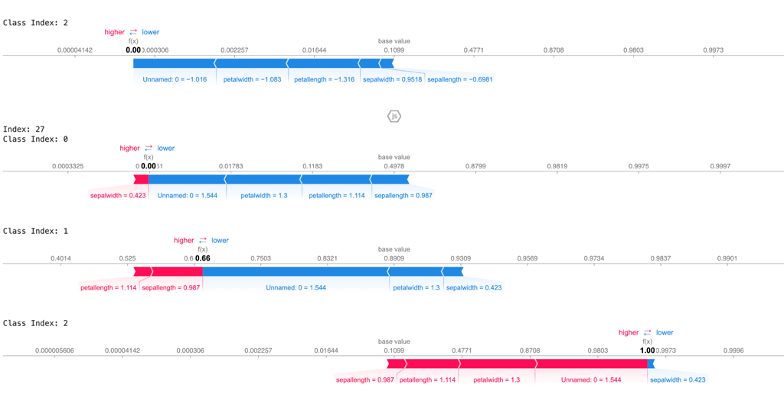


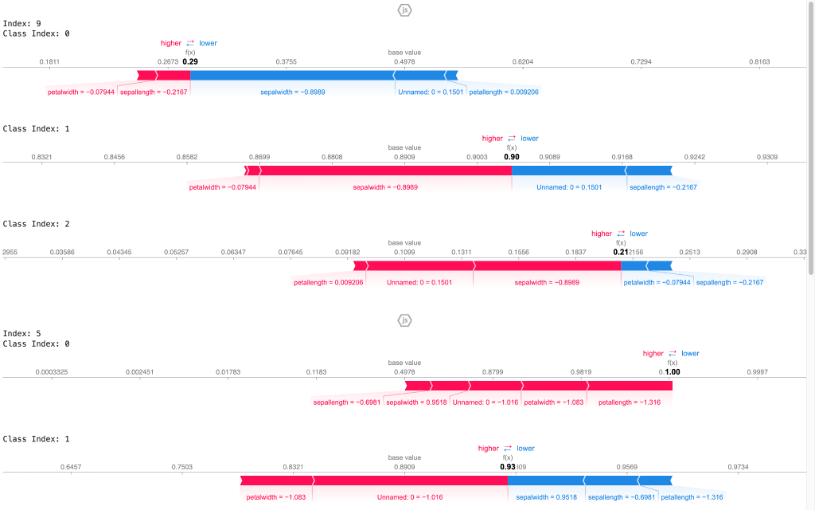


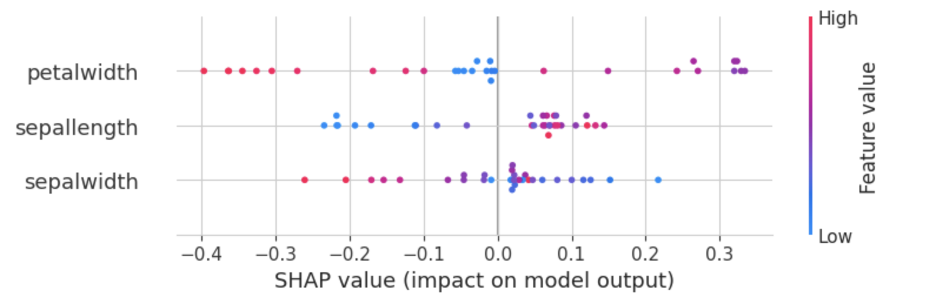
The heatmap illustrates that there is a significant correlation between the physical dimensions, such as the size and weight variables, of the Iris flowers. Conversely, the 'Species' of the Iris does not exhibit a substantial linear association with these physical measurements.



Lime Output

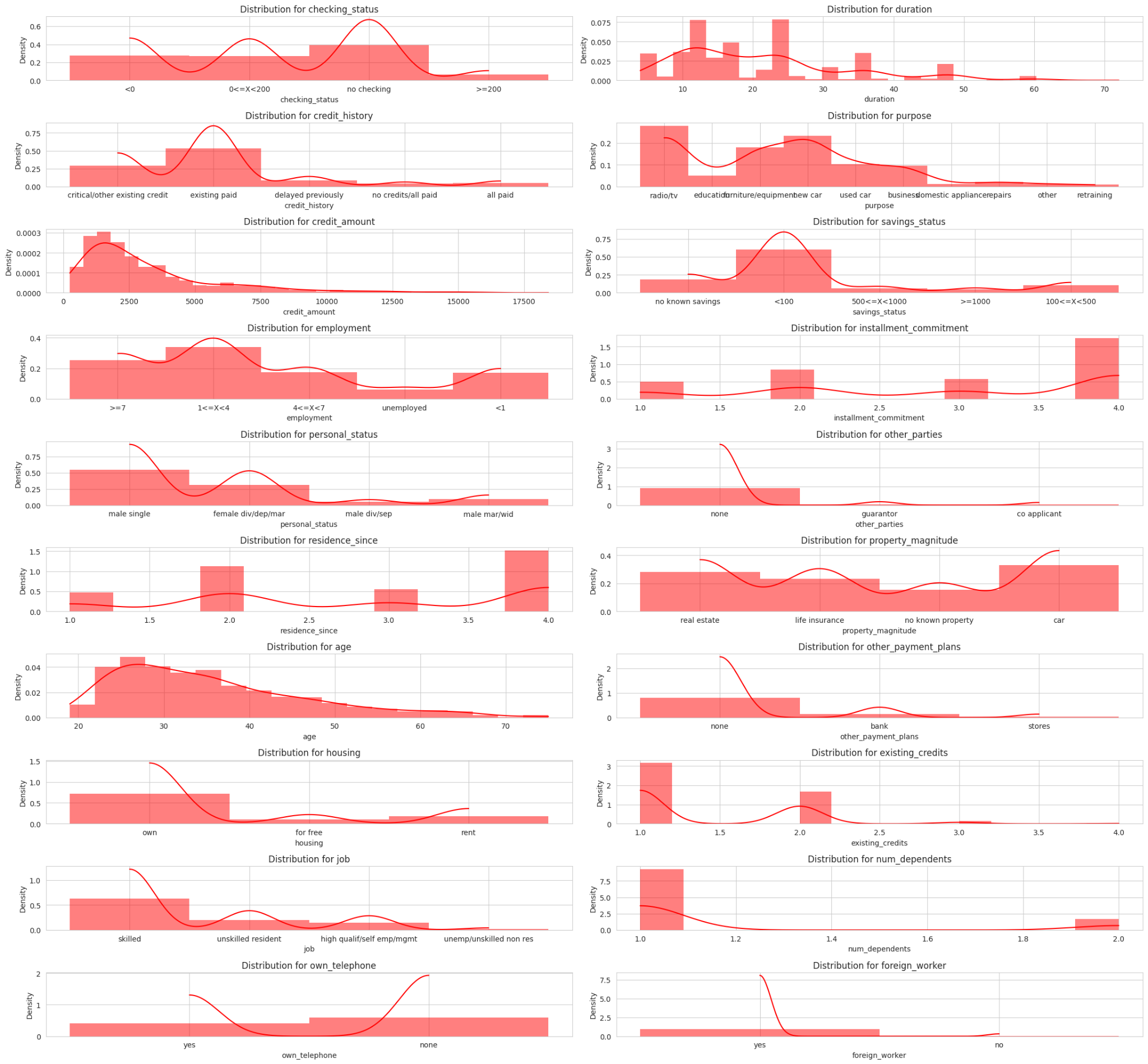


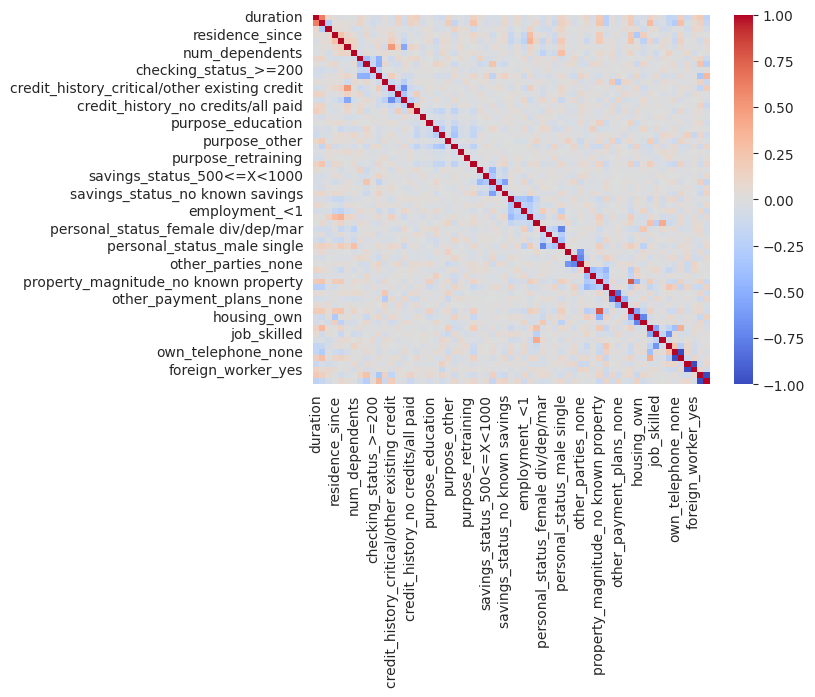




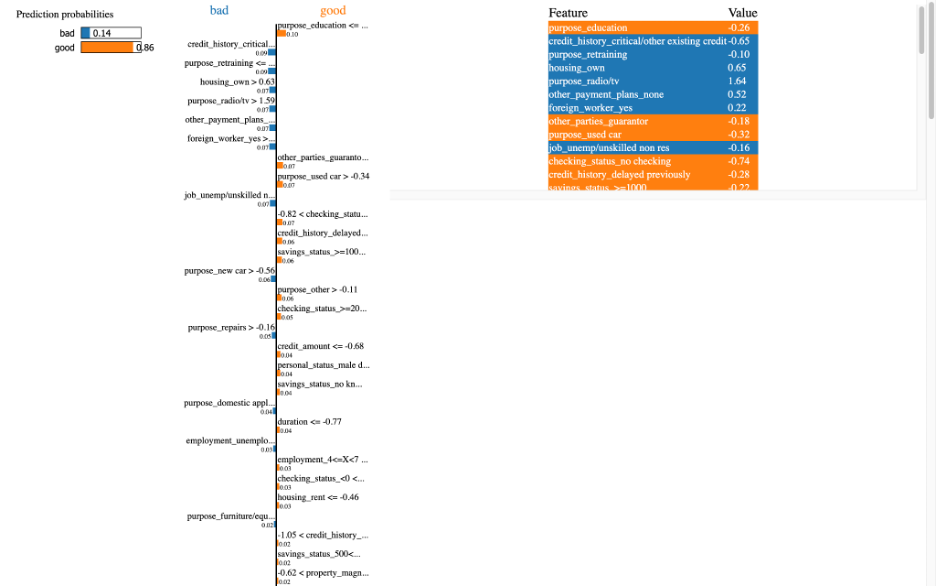
SHAP output

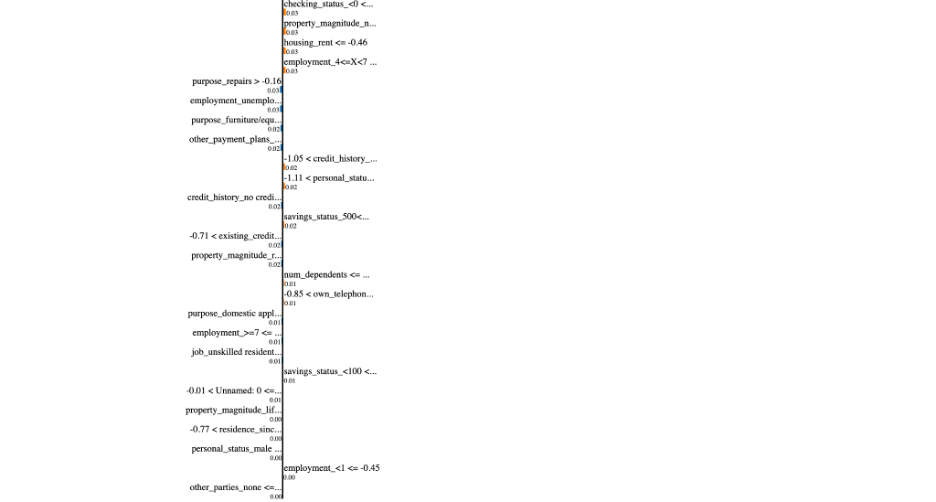
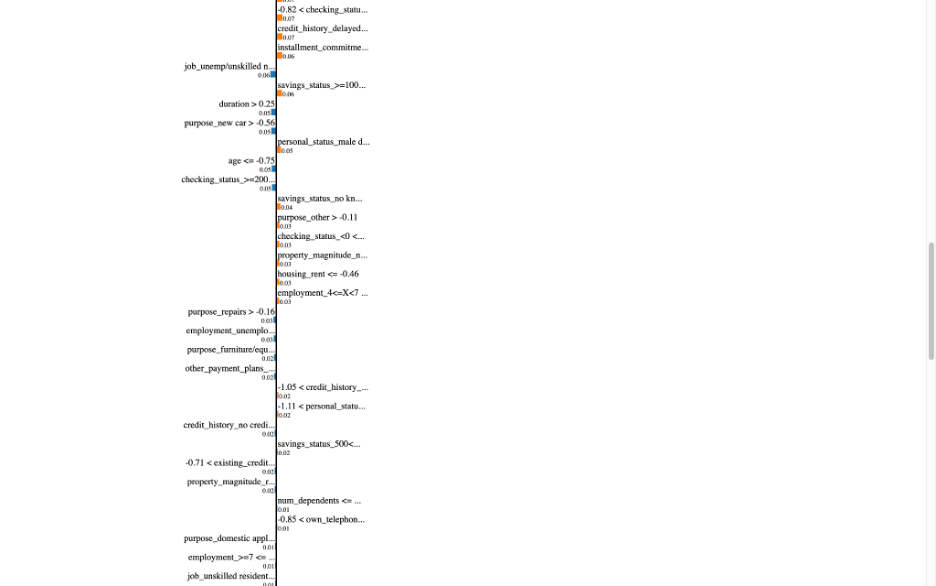
c.Credit





The heatmap indicates that there is a considerable correlation among various credit-related attributes, such as credit history and savings status. In contrast, variables such as 'foreign\_worker' do not show a strong linear relationship with other financial indicators in the dataset.



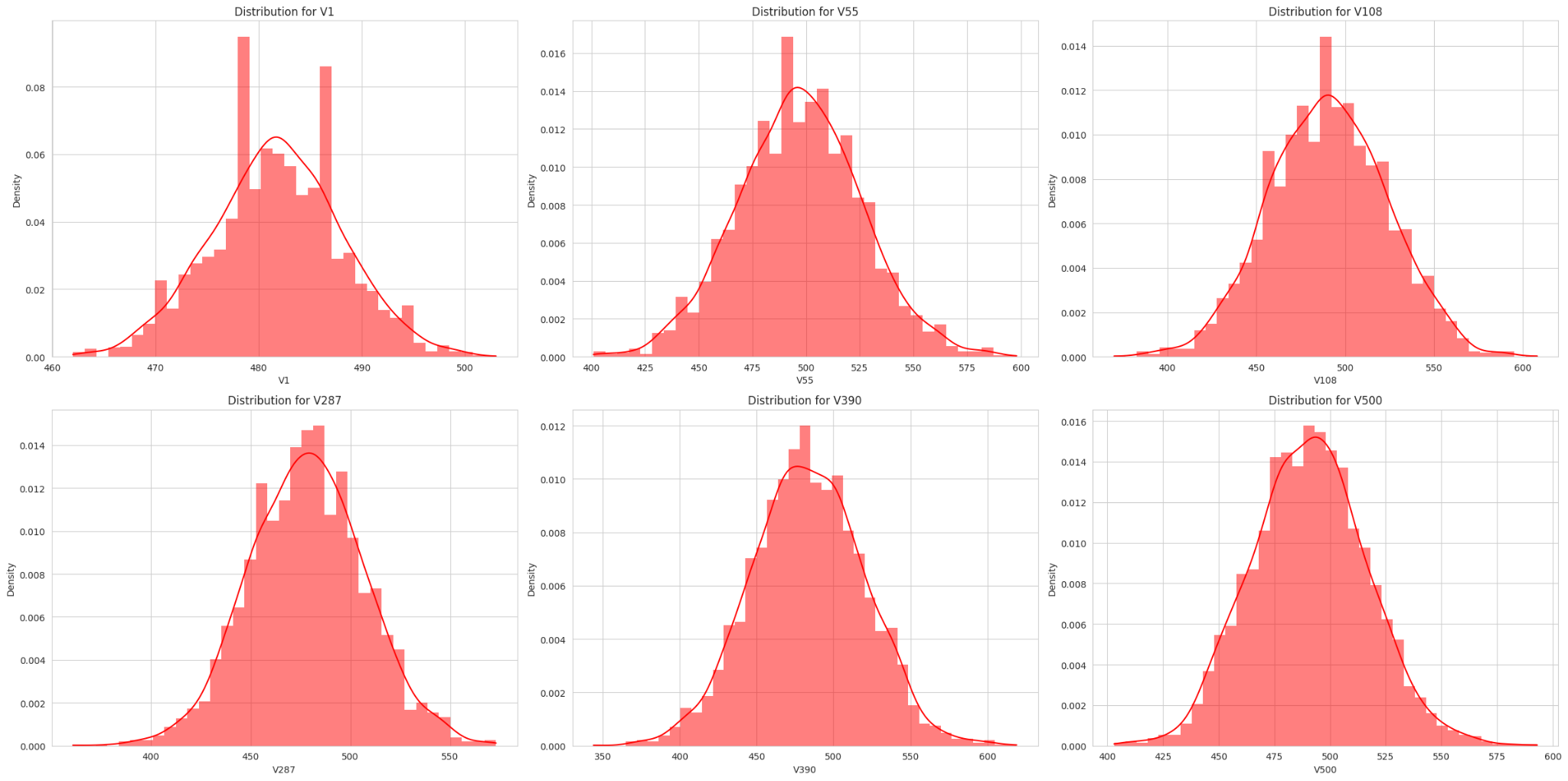


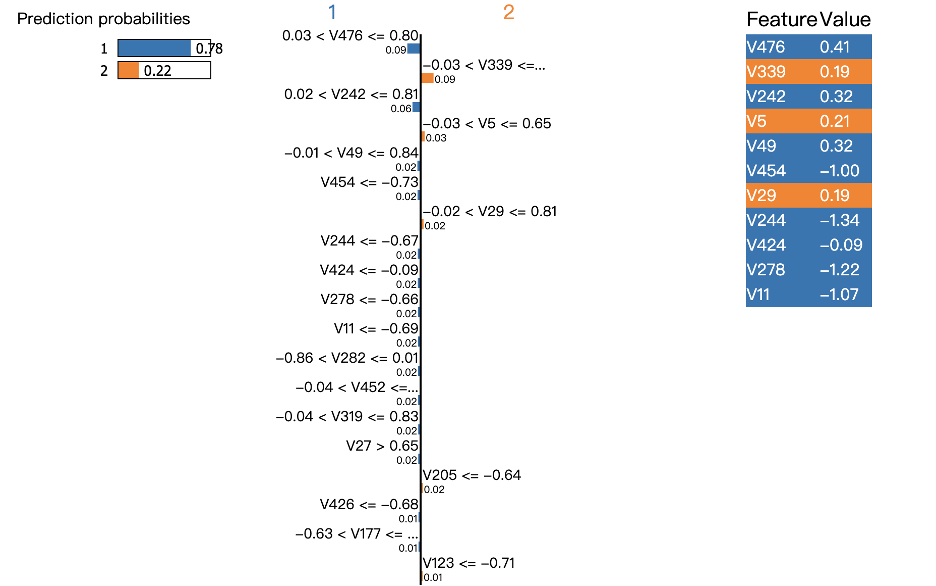


Lime

SHAP

d. Madelon



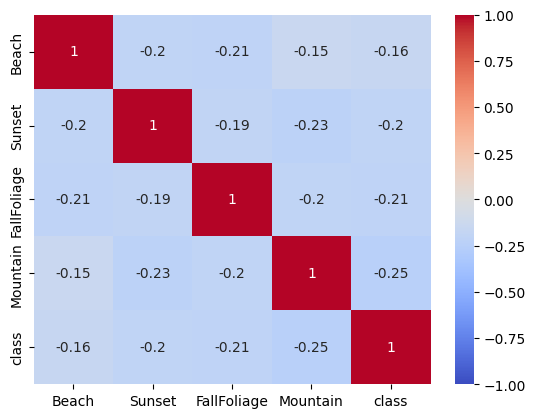


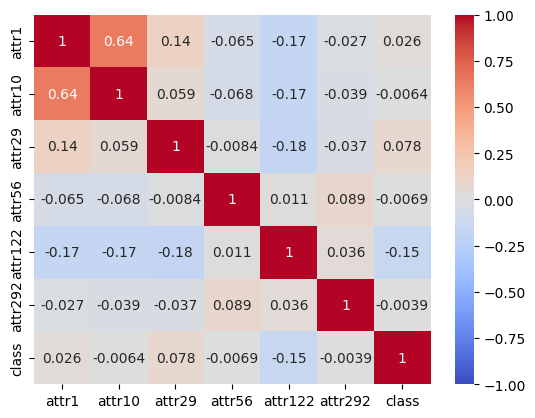
Lime

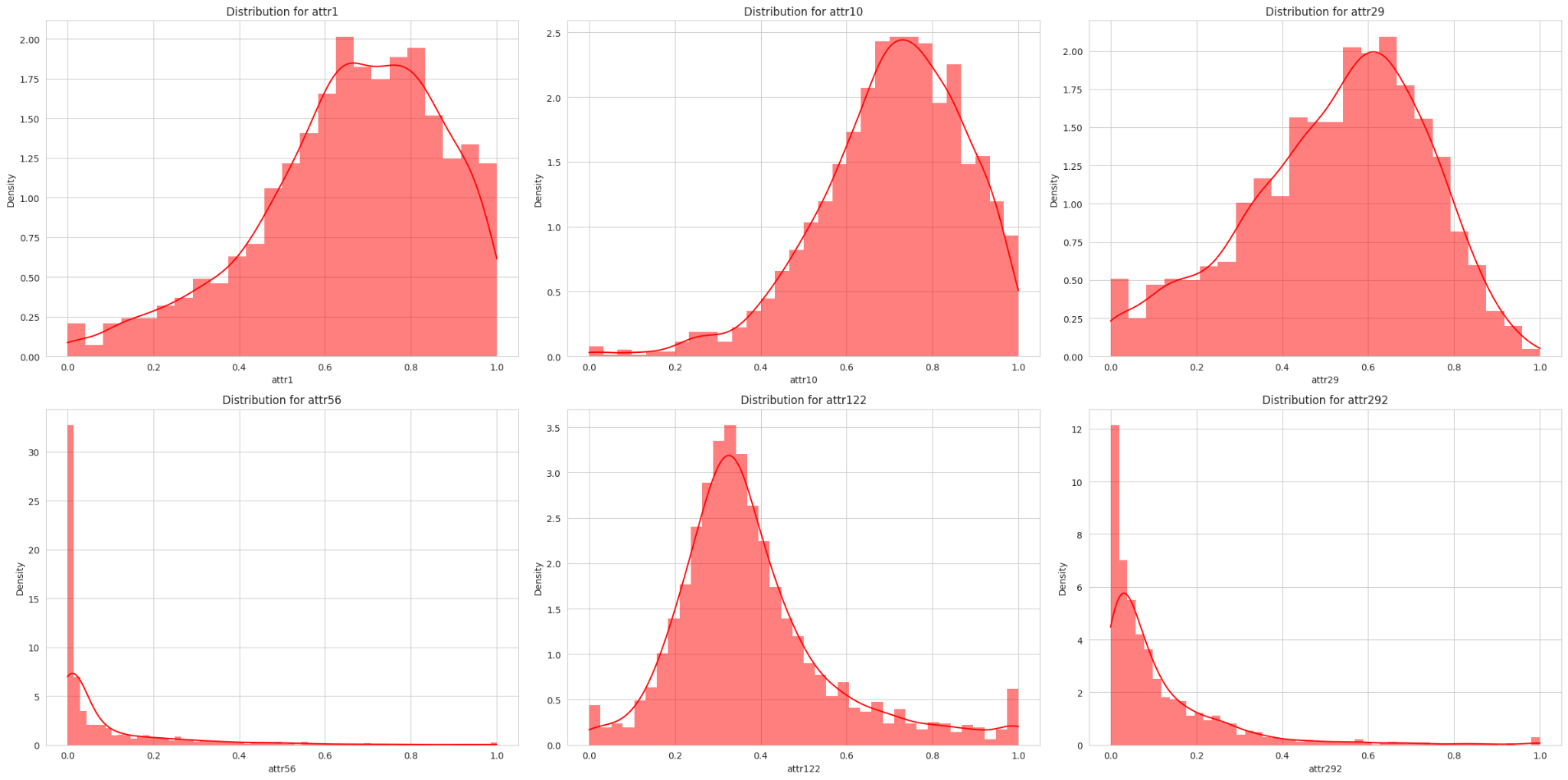


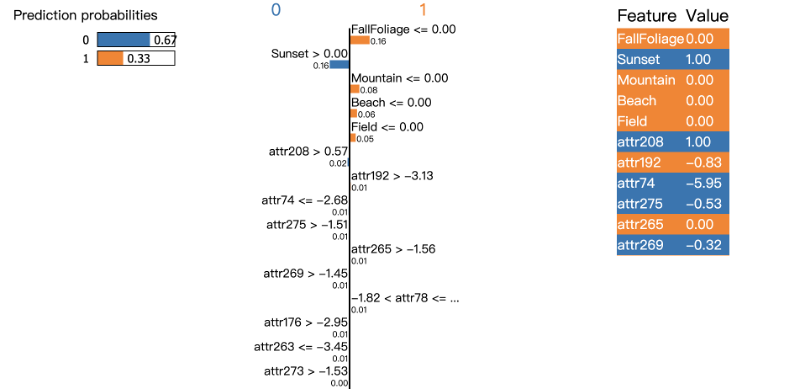
SHAP

e. Scene

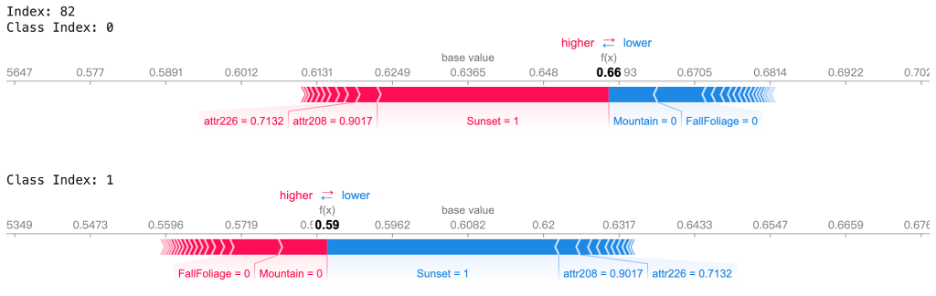


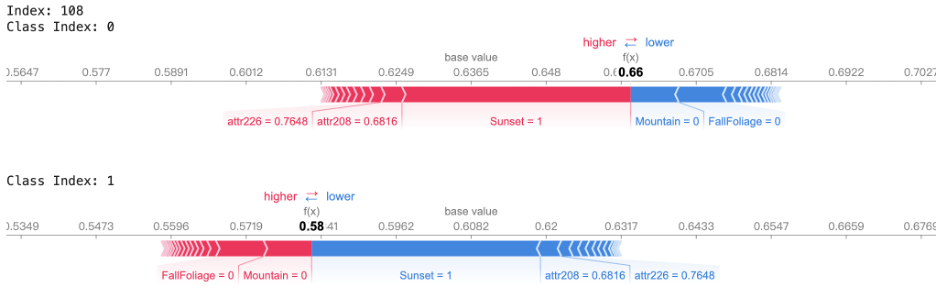


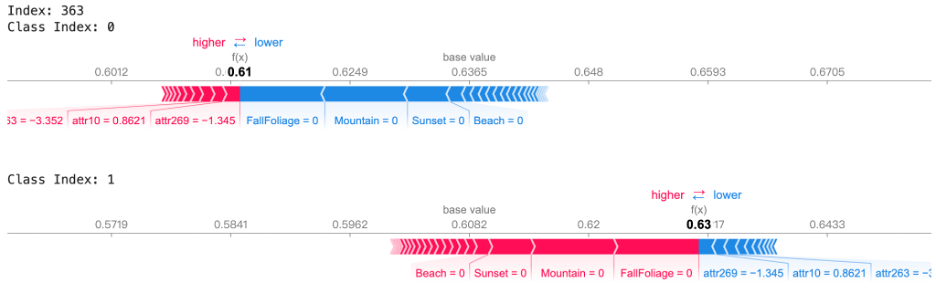




Lime





SHAP

1. **Some comments on Graphs and statistics**

In the abalone dataset, We think the most interesting thing is that we have so many highly correlating predictors (whole weight and shell weight, etc.) and there could be a more extensive autocorrelation analysis on this. We are excited to see if SHAP or LIME can capture these and perhaps improve upon a full model. ( So an attempt to validate the model as well as understanding variables using external tools instead of our traditional F tests and t tests, at the same time see if these inferences can be capitalized for a more accurate prediction).

The SHAP and LIME outputs are pretty intuitive, but we want to first understand what SHAP and LIME did to the model to know what they are exactly explaining. We also want to find some ways to improve our model using them.

1. **Objectives**

Our general research question is: What are some applications of explainable ML tools (specifically SHAP and LIME, maybe Fanova or others depending on time constraints) for different prediction tasks. We chose a wide range of task for a reason, we want to see how applicable these tools are in terms of:

1. Ability in finding out faults and errors in assumptions or falsely used variables in the models.
2. Increasing accuracy of prediction tasks by giving some inference of the model.
3. Robustness in terms of doing tasks a and b in different types of predictions ( models).