In the data preparation stage, we first import the CSV file and then rename and sort the columns in the data to make it more readable. For the pre-processing step of this project, we start by dealing with the numerical feature first. We convert all the float numbers to the integer, it will not influence the experiment a lot, but it can reduce the domain and make the process easier for us to run. Then we deal with some missing values in the attribute property.winner, since only this attribute occurs missing value. We fill in missing values with the number 0. Next, we extract attributes for price and the winner. We can assign values to the winner's attributes, such as carrying capacity (for which we need to add a function that introduces a new attribute), and adjust the bidder's capacity concerning the winner.

We are planning to run experiments that involve:

Training prediction models for different targets, including verification results, verification time, and allocation revenue.

Using models with varying complexities, specifically Random Forest models with different numbers of trees (1, 10, and 100).

We employ different cross-validation splitting methods such as k-fold, reverse k-fold, and Leave-One-Group-Out for capacity, product, and position.

We assess model performance using relevant evaluation metrics, such as the Matthews correlation coefficient for classification tasks and the R2 score for regression tasks.

We are analyzing feature importance to gain insights into the impact of different features on the prediction models.

We plan to design a series of functions for creating charts, data statistics, and performance metrics:

1. We will create functions for plotting various types of charts, including functions for plotting validation time distribution, predictive performance, and feature importance.
2. We will define an evaluation function that executes the complete evaluation process. This function will load the dataset from a specified directory and pre-process it to prepare it for subsequent prediction tasks. Afterward, the code will analyze the dataset, calculate various statistics, and output the results.
3. Using the pre-processed dataset, the code will run different prediction tasks, calculate various performance metrics, and plot the corresponding charts.