My work can be divided into four parts.

I. The first step was data cleaning. I observed and filtered out some fields that were not meaningful for data analysis (host information, etc.). Data cleaning was done by python: read the csv file in, and then keep the useful fields:

(roomtype: indicates the overall size of the room)

(bedrooms: Number of bedrooms, which is important for hotel pricing)

(Amenities: indicates the equipment information of the room. The more the number of devices, the more advanced the room is.)

(Number of reviews: indicates how hot the room is on Airbnb)

(Scores of reviews: auxiliary indicates the popularity of the room)

(Price: final goal)

II. Then I defined all the fields locally, so that the original data could be completely converted to local data, and each field used enum to simplify the complexity of the data.

III. Use the C4.5 algorithm to calculate the read data step by step, including the calculation of information entropy, information gain, information gain rate and so on.

IV. Finally, the decision tree is constructed by calculating the data, and the tree nodes are labeled by the attribute classification standard. The original rule set was divided step by step through recursion, and the construction was finally completed.