**Muhammad Ahmad 21L-5617 BS(DS) 6A**

**Question1**

1. What is the size of the training set?

The data set contain 14 rows

1. How many attributes exist in the training set?

There are 5 attributes in data set

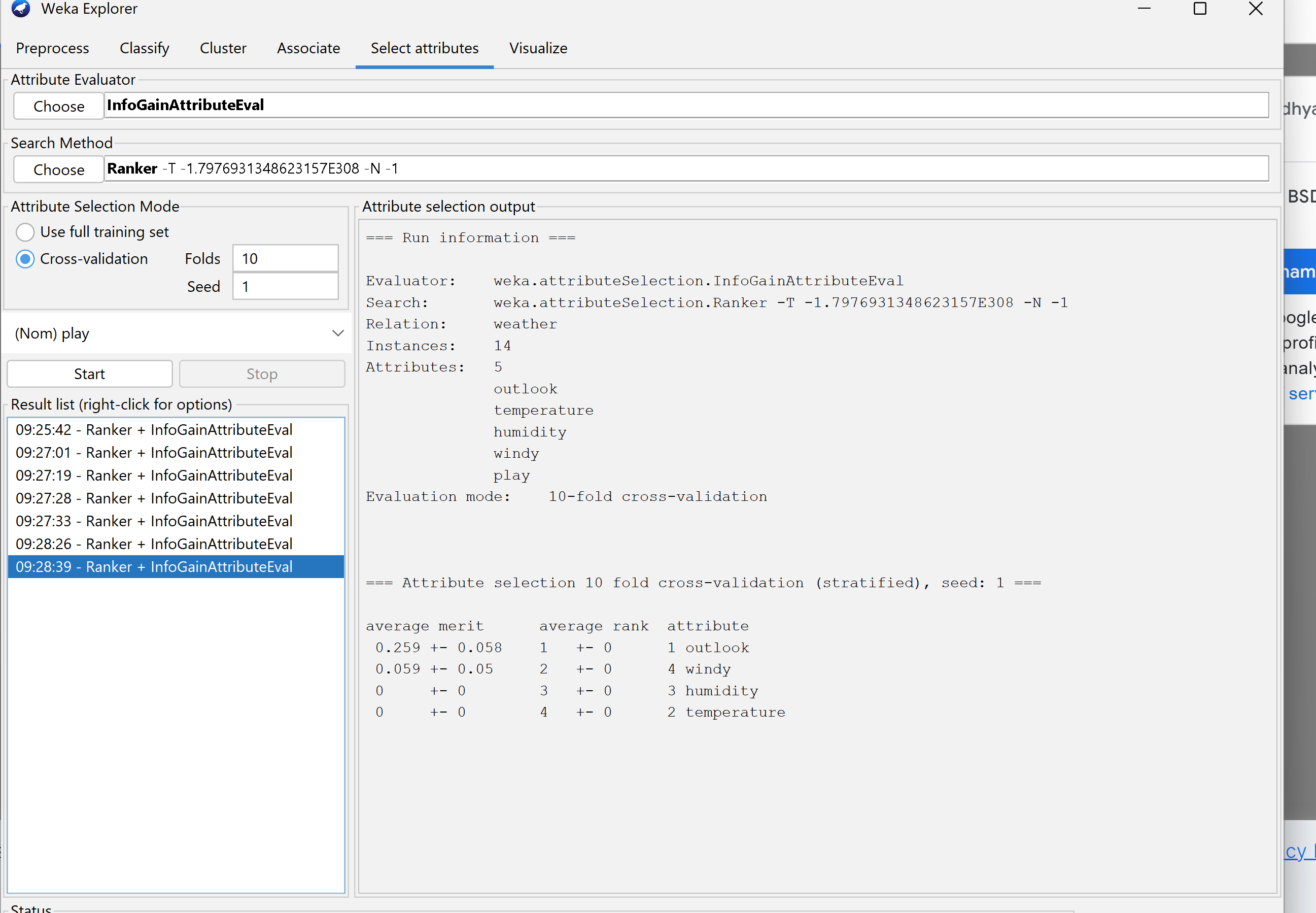
1. How many instances are positive (play = yes) and how many negative?

Yes=9

NO= 5

1. Which attribute best separates the data?

The information gain of outlook is best so it separate the data set in to two in a best foam

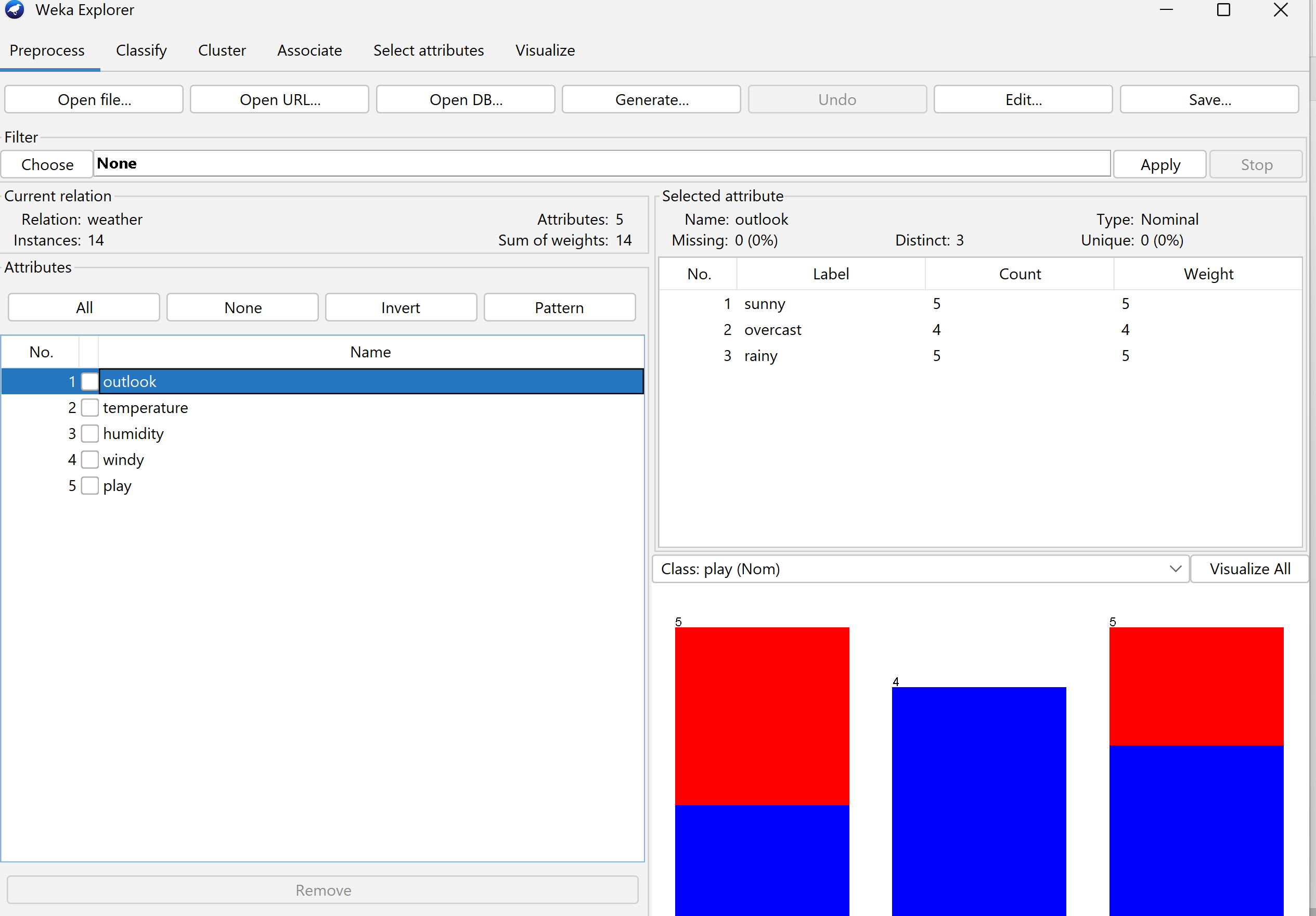


1. How many elements from the data set have the humidity attribute set as high?

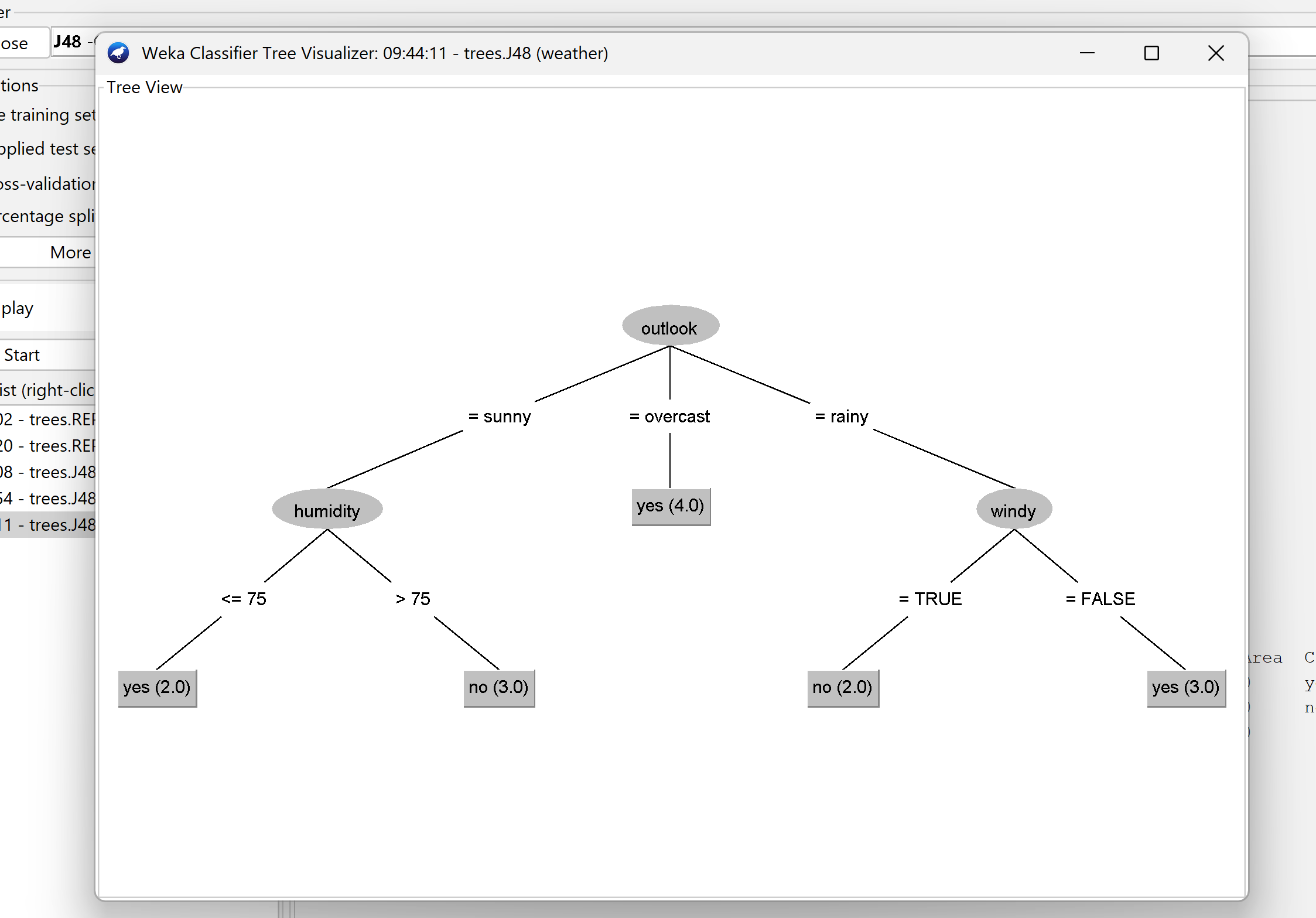
Sunny = 5

Overcast=4

Rainy=5



Question2 Load and analyse data



Question3

1. Run the algorithm. How many percent of cases were

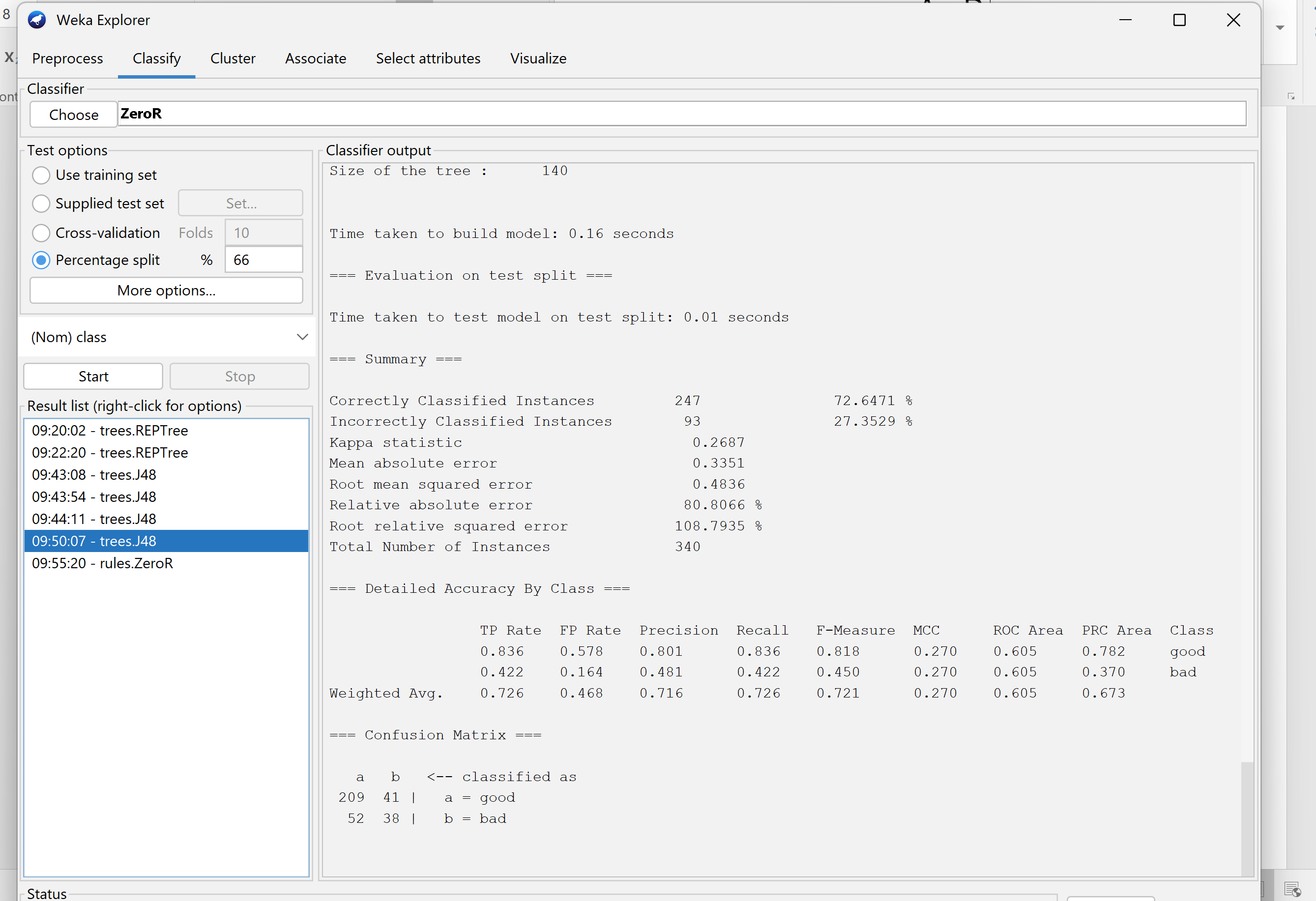
correctly classified? Is this a good result?

Solution :-

About 72% the classifier predicts correct result

And 28 % the classifier predicts wrong results

In my opinion it is good classifier

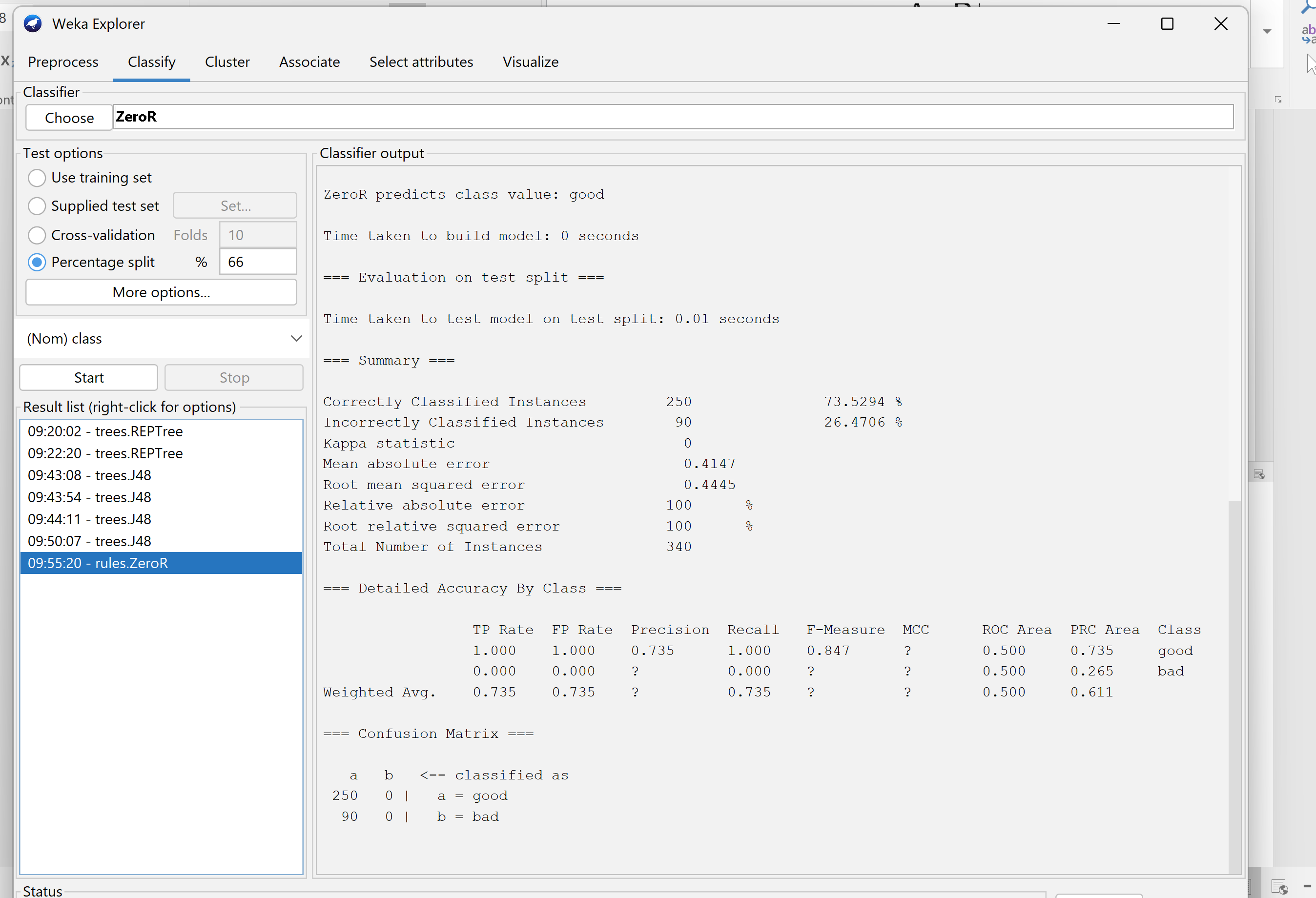


1. **Change the classifier to ZeroR from the rules branch. What is the obtained result?Better or worse than J48?**

About 73% the classifier predicts correct result

And 27 % the classifier predicts wrong results

it is better classifier then j48



1. **Try other classifiers. What are their results?**

The LMT is best across all other clasifers becaue

About 77% the classifier predicts correct result

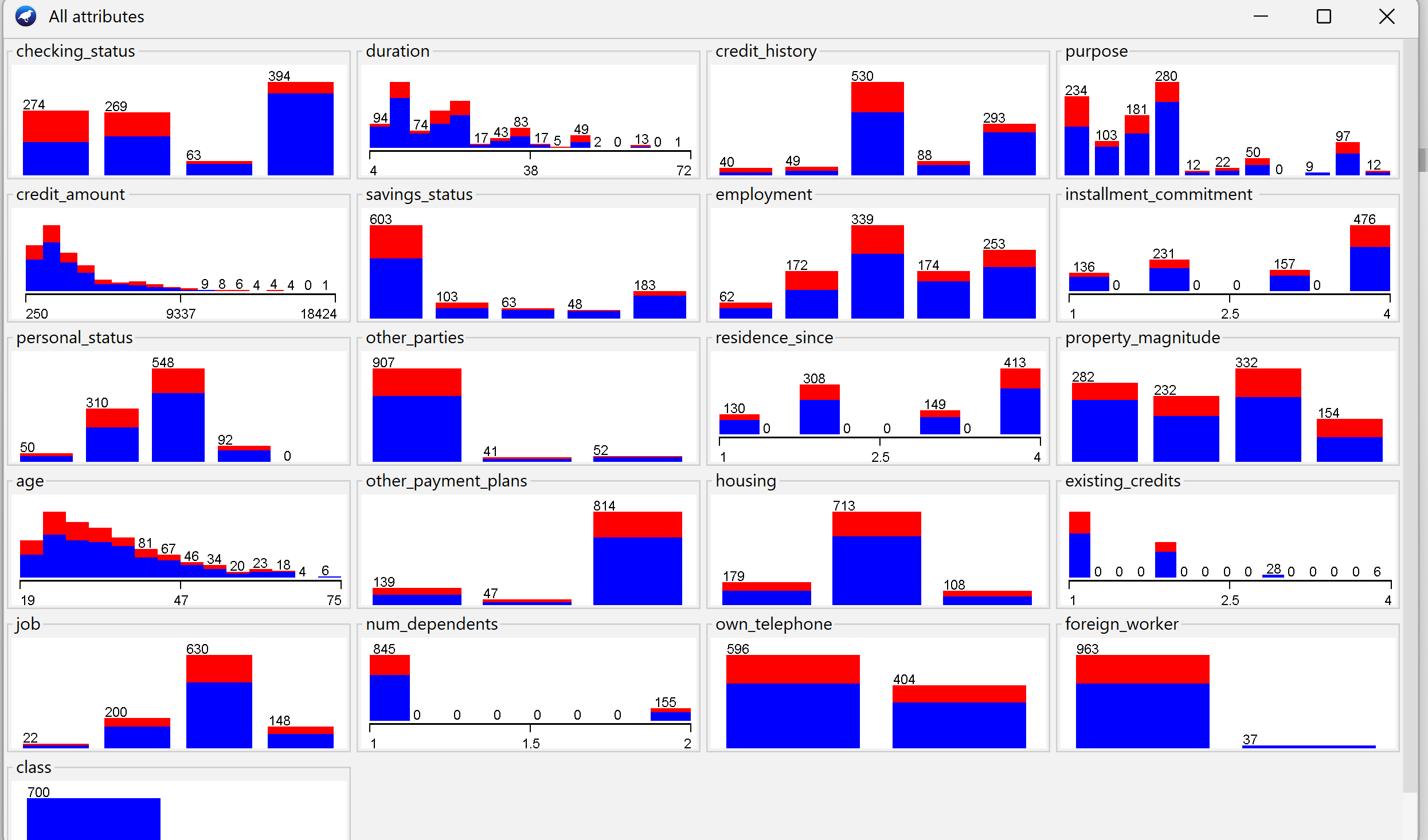
And 23 % the classifier predicts wrong results

1. **Go to the ‘Preprocess’ tab and see how the distribution of the**

**attribute defines whether the set is good or bad. What would**

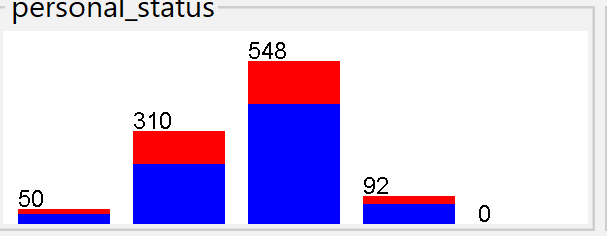
**be the effectiveness of an algorithm thatregardless of the**

**value of attributes would "shoot" that the user is reliab not?**

****

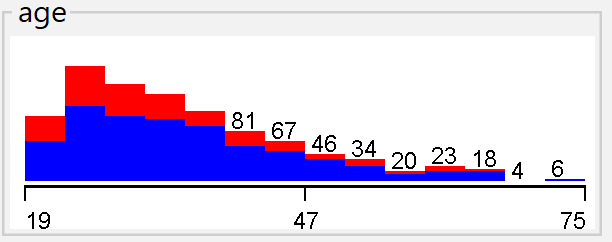
**Normalize Data**

The normaly distributed data attribures nclude “personal status, property management “

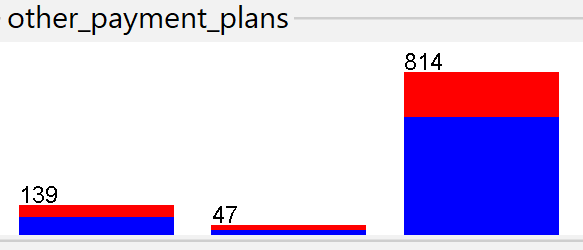


**Right Skewed Data**

The age attribute is right skewed



**Left Skewed data**

****

**8. Why is it worth taking a look at the data before attempting a**

**classification task?**

To check the any missing value, check data skewed ness in data because it might give wrong classifier and give skewed and bias classifier