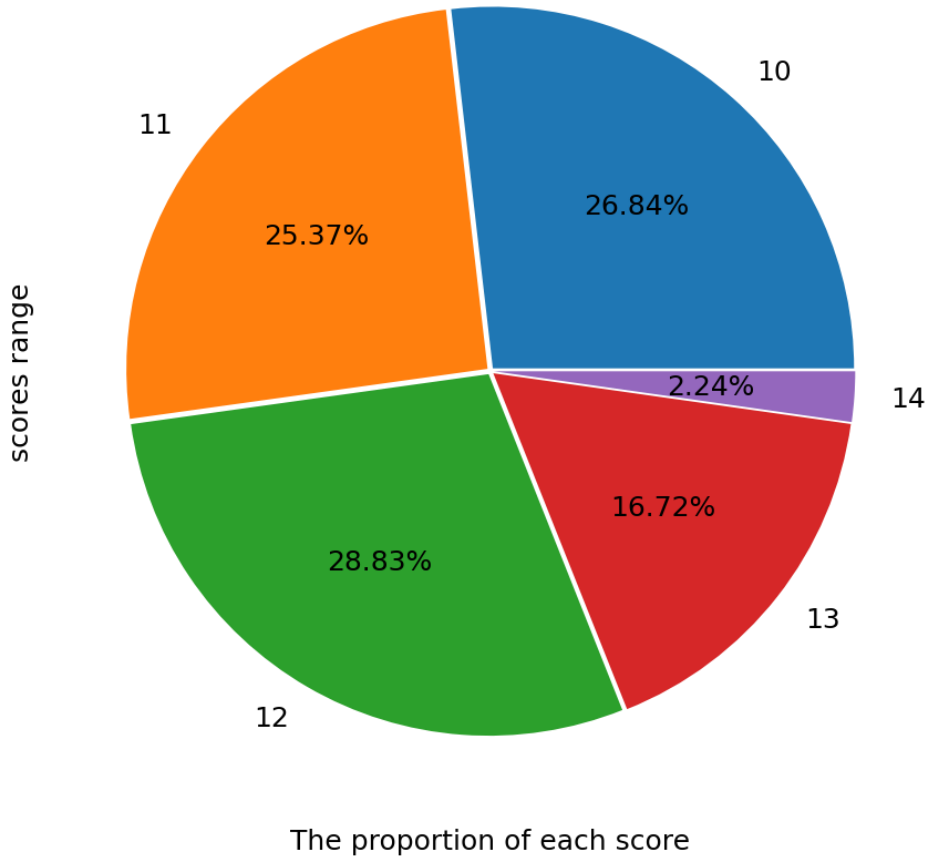
Act report

Before acting we import our clean data in the form of two dataframes tweets_info which contains the previously merged tables, and tweets_predic representing image predictions, then we sort those dataframes by their tweet_id to match each other in the order.

The first insight is calculating the average of text length in tweets having True value in the three columns p1, p2 and p3 of tweets_predic, those columns represents the first, second and third prediction of the algorithm recognizing dogs, so we filter by the product of the three columns elements in order for the result to be only True or 1 if and only if the three values are all 1, after calculating the mean we find its value 111 characters.

The second one is about finding the proportion of tweets that scored 10/10 or more after eliminating non-repeated scores which means scores counted only 1, for that we filter tweets_info by rating_numerator values +10, then we eliminate unneeded elements hence we keep the range of scores between 10 and 14, finally we draw a pie chart showing the proportion of each score counts in the form of percentage, bellow is an illustrating to findings:



The last insight is finding the name and the race p1 of the dog of the dog from top 10 favorited tweets and retweeted and at the same time it has the highest probability p1_conf, thus we begin by creating two variables top10_fovorite and top10_retweeted containing the top 10 for each of the two columns, after we leave only 8 dogs that are in top 10 of both favorited and retweeted, so visually we can see that “French_bulldog” has the highest probability p1_conf with 0.905334, his name after extracting is “Jamesy”.

The visualization is representing the most common names of dogs, we eliminate “a” and “the” we graph bars showing the count of the name of the dog, we find that those values are near to each other, and here is the plot: