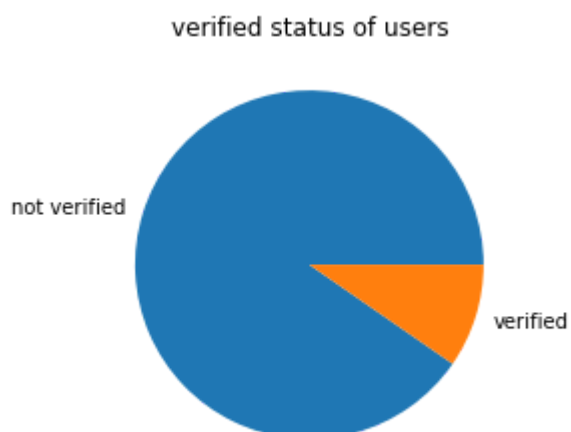


We will analyse the corpus of tweets we got when we looked for tweets with the hashtag #boxing, which was the trending tweet in hyderabad when this project was made.

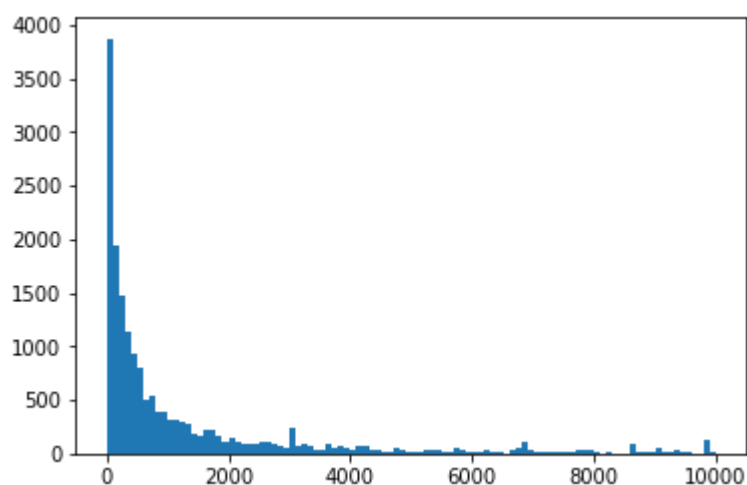
First we will see the verification status of the users.

We see that around 90% of these users are not verified, while around 10% of users are verified. Now, as we know that there are alot more unverified users compaired to verified users, this tells us that this particular trend is not being tweeted impartially by either category. If, for instance the ratio were 50-50, then we could say that this hashtag is being tweeted more by verified users than unverified users, suggesting possibilities like the hashtag being related to an advertisement campaign.

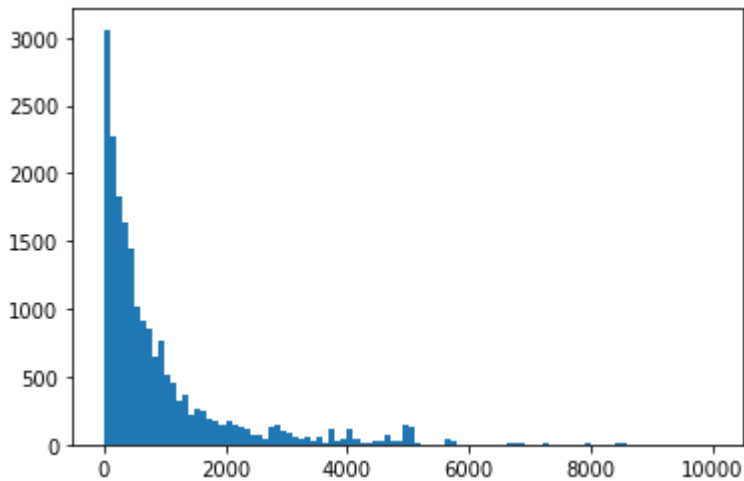


Next, we look at the number of followers the users that tweeted the tweets have. majority of the users have less than 2000 followers, with 75% having less than 2,750 followers. Again, considering that most accounts aren't "famous", this shows that this tweet isn't impartially tweeted by "famous" people. This tells us that the topic at hand is something everyone is talking about and is not for instance a movement by a certain group of people or something related to an advertisement campaign.

We note that the user with the highest number of followers that made a tweet on this has 1.5 million followers.



We also look at the number of friends our users have. We note that most of them have less than 1500 friends, with 75% having less that 1116 friends.



```
data['tweet_lang'].describe()
labels = []
for i in range(0,51):
    labels.append(None)
labels[0] = 'english'
labels[1] = 'spanish'
data['tweet_lang'].value_counts().plot(kind='pie', labels=labels)
plt.show()
```

We now look at the language in which the tweet was tweeted. The distribution we see is something you would expect from a topic that is being talked about worldwide, and not something that would be specific to a certain geographic location like a country or city. If, it had been a topic that was only being talked about in India, you probably wouldn't see a lot of spanish, portuguese, or japanese tweets, which were the top 3 languages after english.

