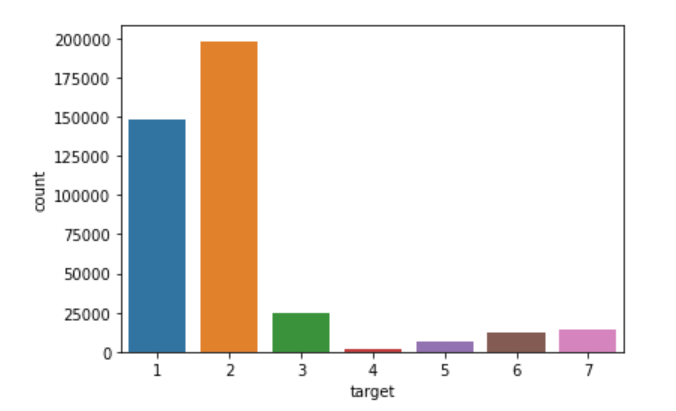
The idea behind this classification task is to identify different types of forest cover types in undisturbed forests.

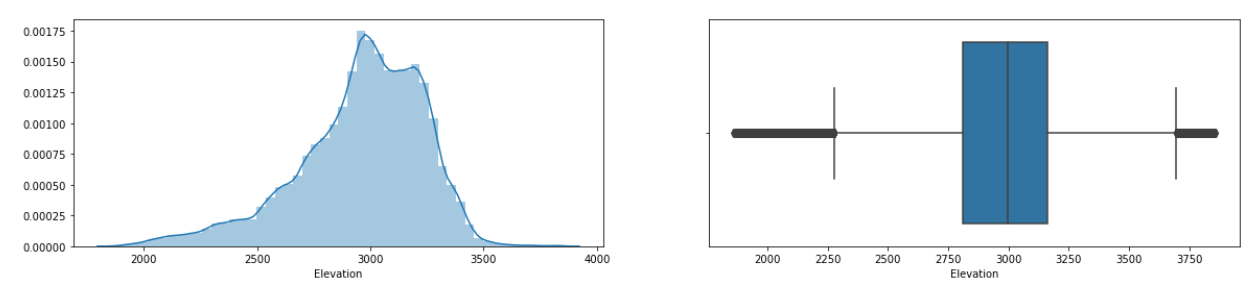
The data is complete or doesn’t need any type of data interpolation since there is not data missing.

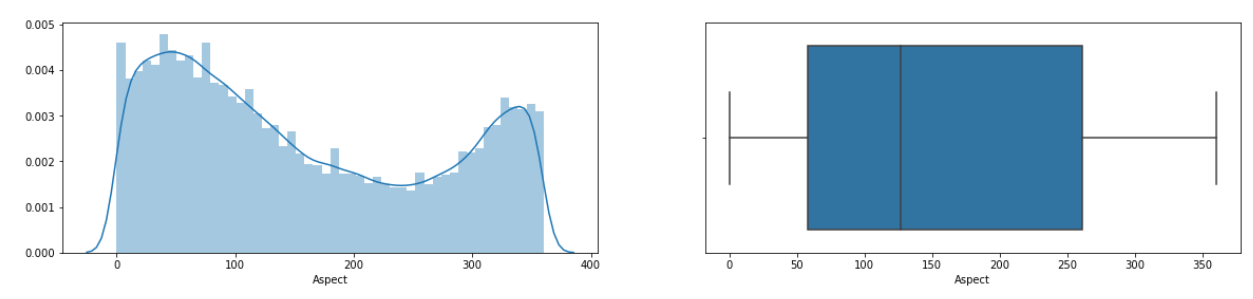
* EDA

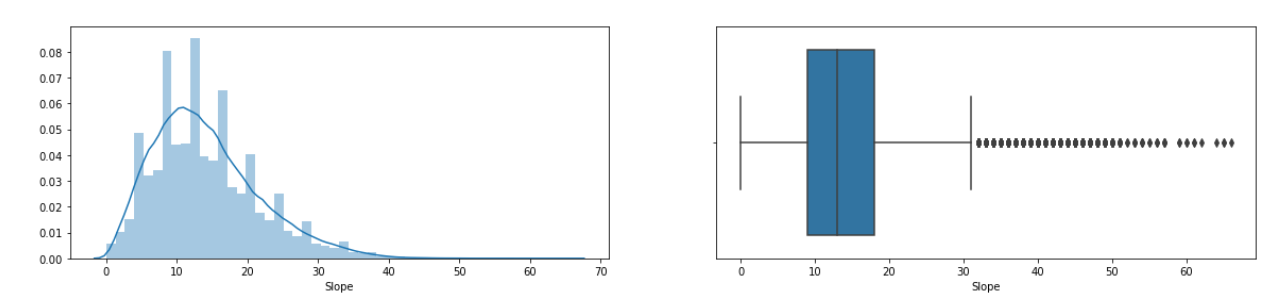
Conducting EDA to understand our data, lets join our target labeling or classification into the training data to try to find patterns or relationships between the variables. From the image we can see that the cover type 1 and 2 are the most predominant features over the dataset, while labelings from 3 to 7 are much lower.

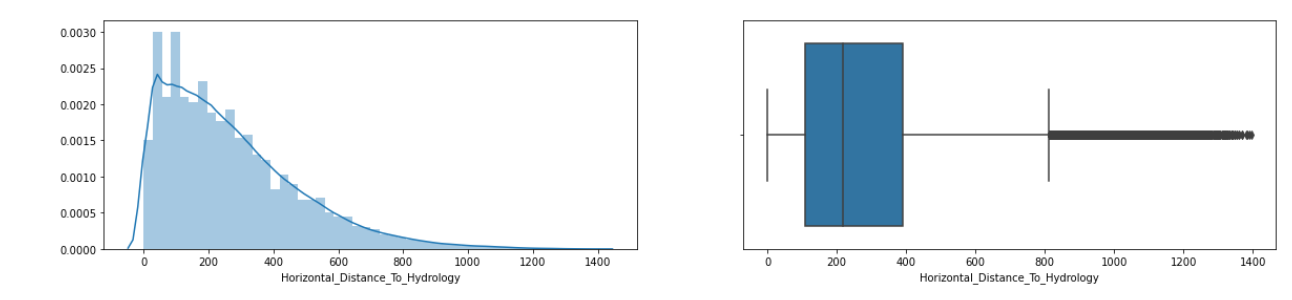


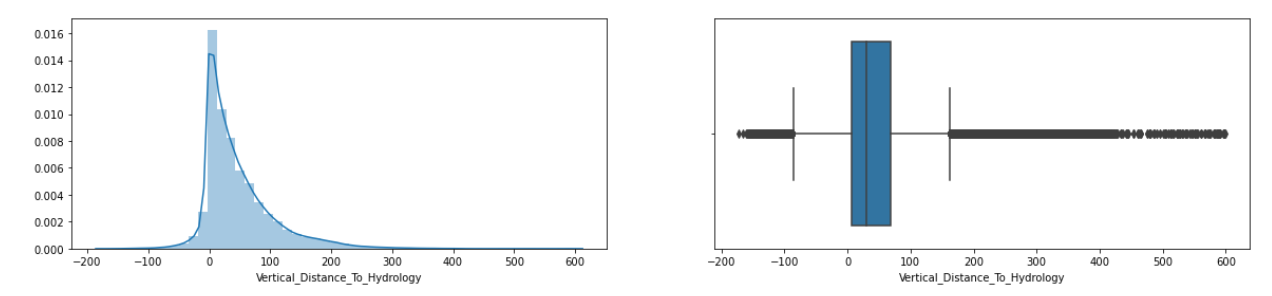
* Plot

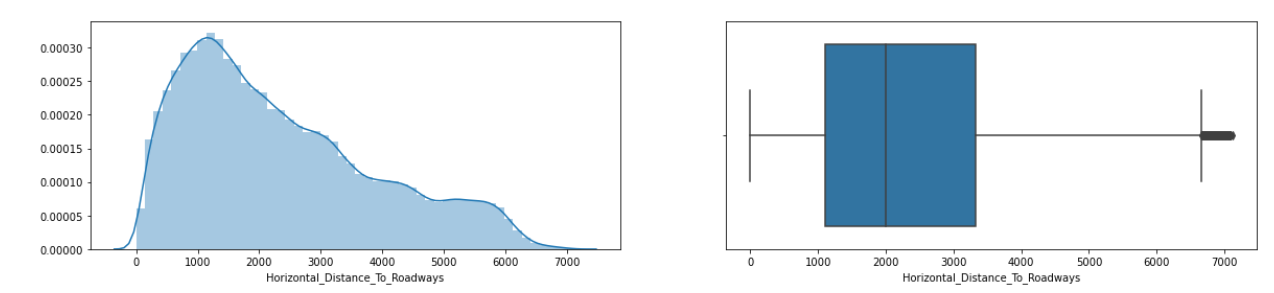


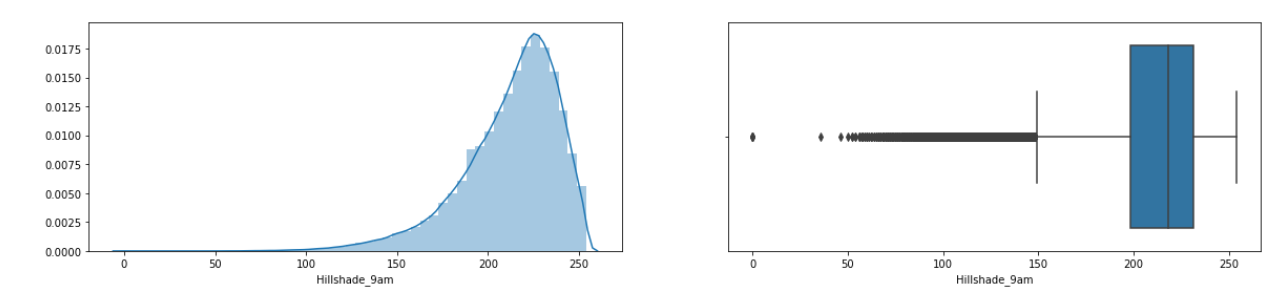


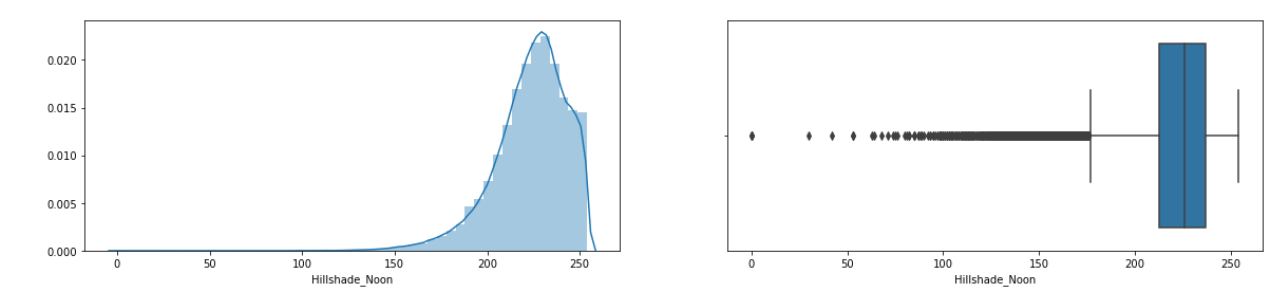


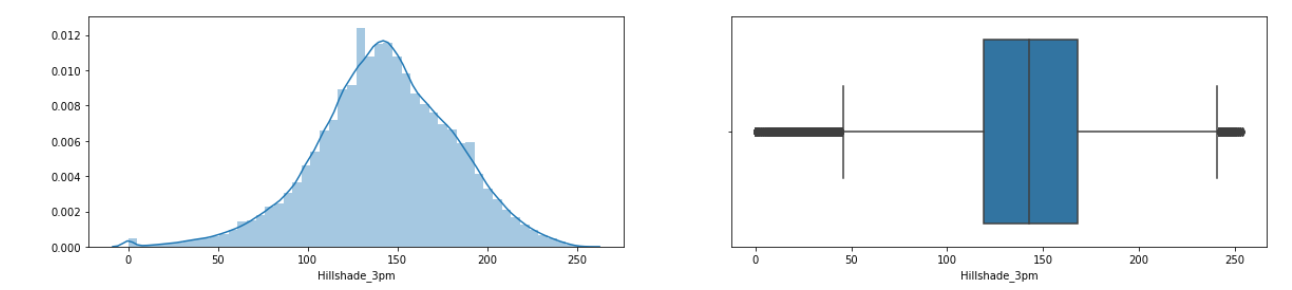


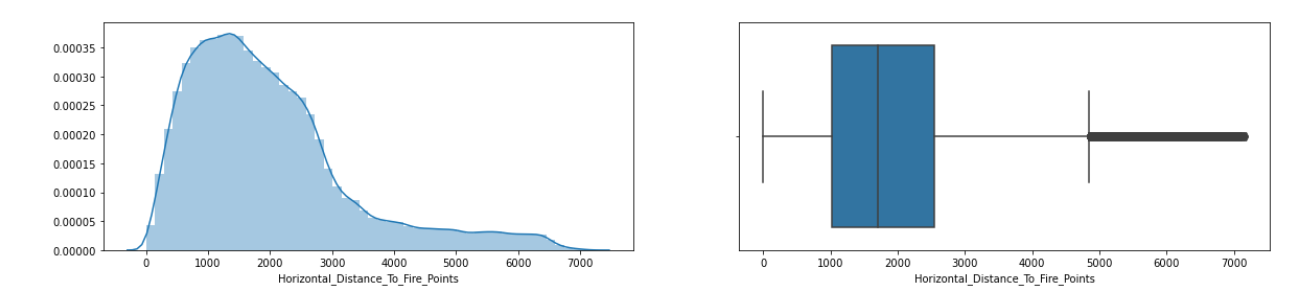


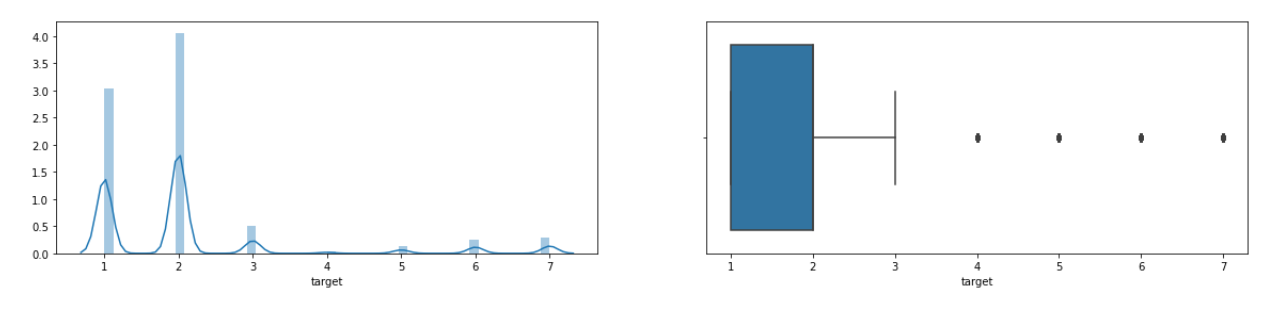




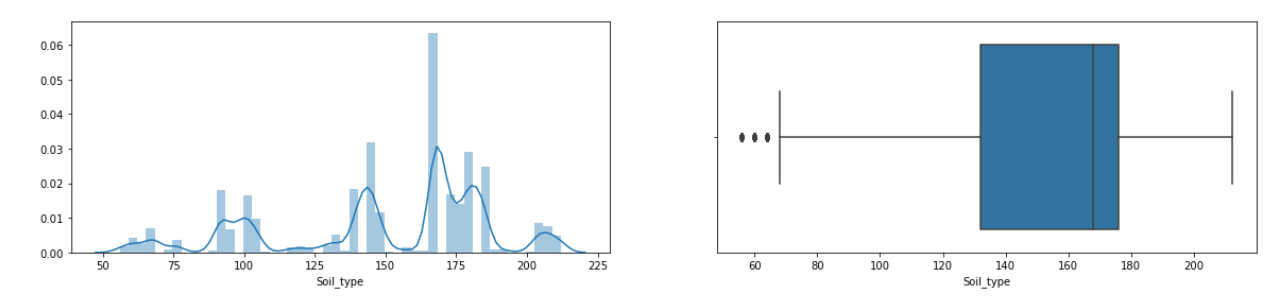




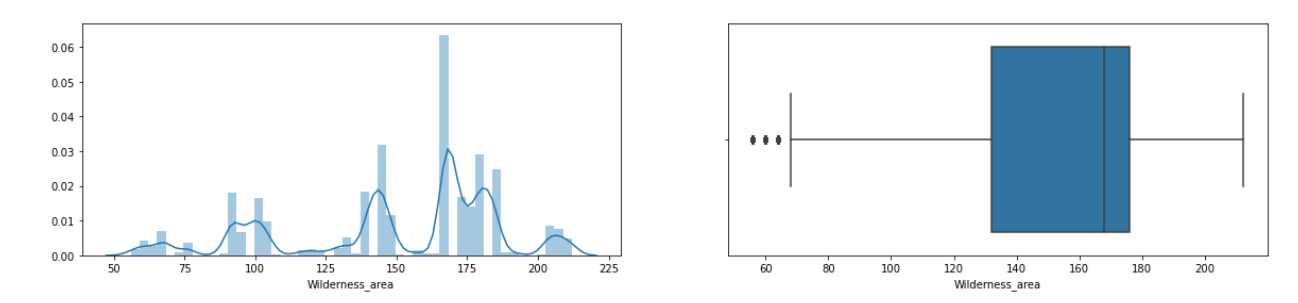




* Convert soil type 1-40 from binary type, to number type and merged them into one single column called soil\_type.



* Convert Wilderness areas 1-4 from binary type, to number type and merged them into one single column called wildernas\_Area.



Using the KNN model the accuracy improved to 95.4554 %

Changing number of neighbors to n=3, the accuracy improved to 96.0553%