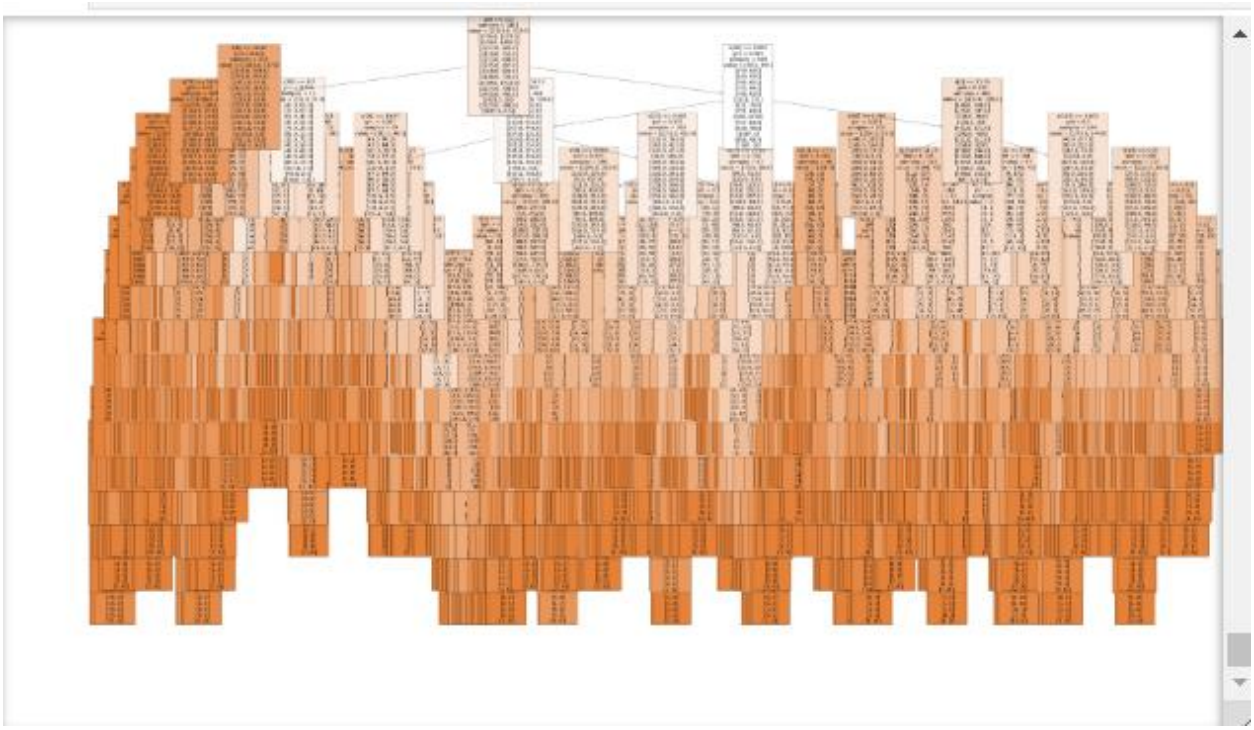
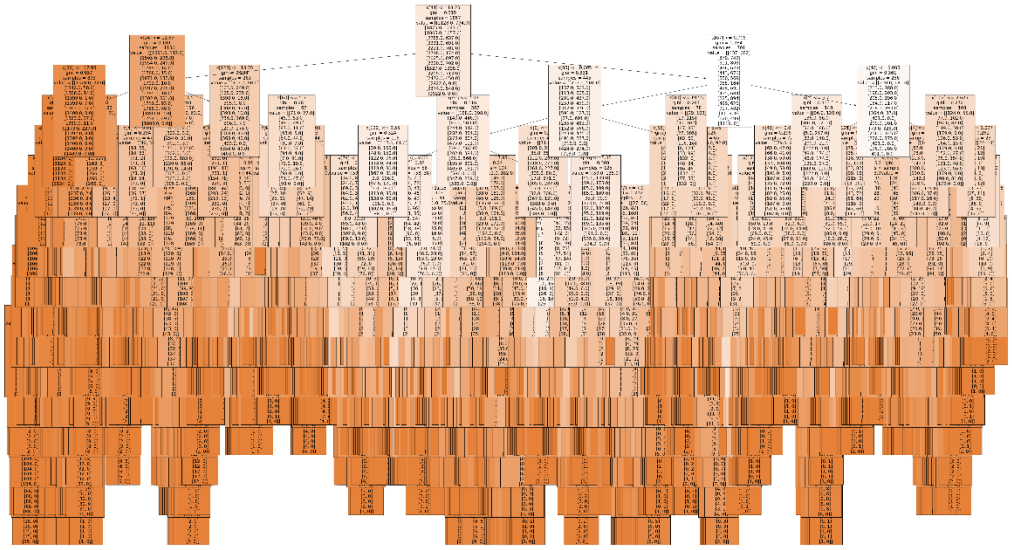
For the years 2010-2020, the random forest model accuracy was at 59%. Used 15 as the max_depth and 100 as the n_estimators. The decision tree is pretty complicated. The three most important features were Dusseldorf, Munchneb and Basel.

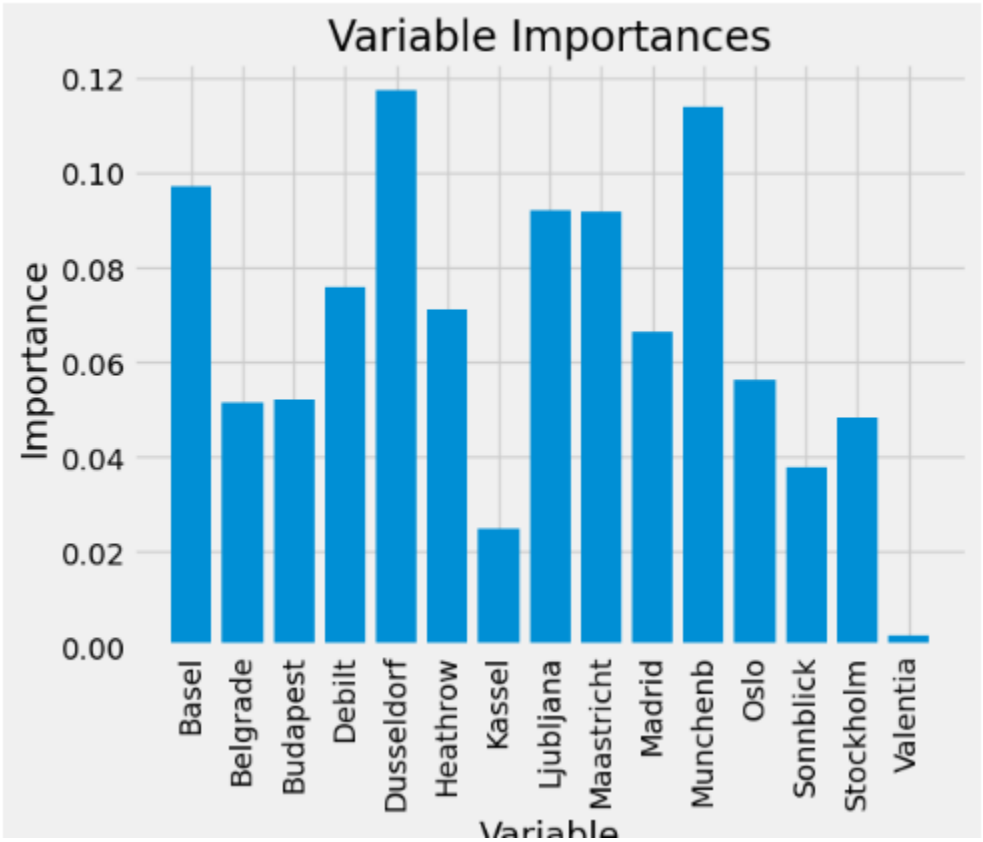
```
#04.2 performing predictions on the test dataset
y_pred = clf.predict(X_test)

# using metrics module for accuracy calculation
print('Model Accuracy: ', metrics.accuracy_score(y_test, y_pred))
```

```
Model Accuracy:  0.5909712722298222
```



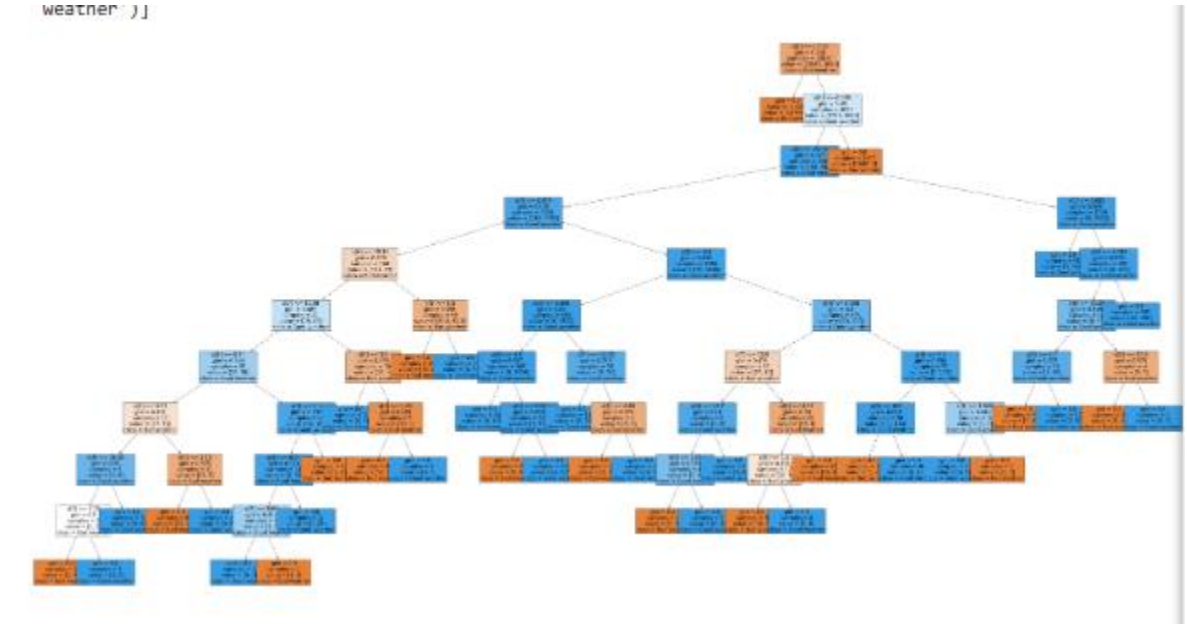


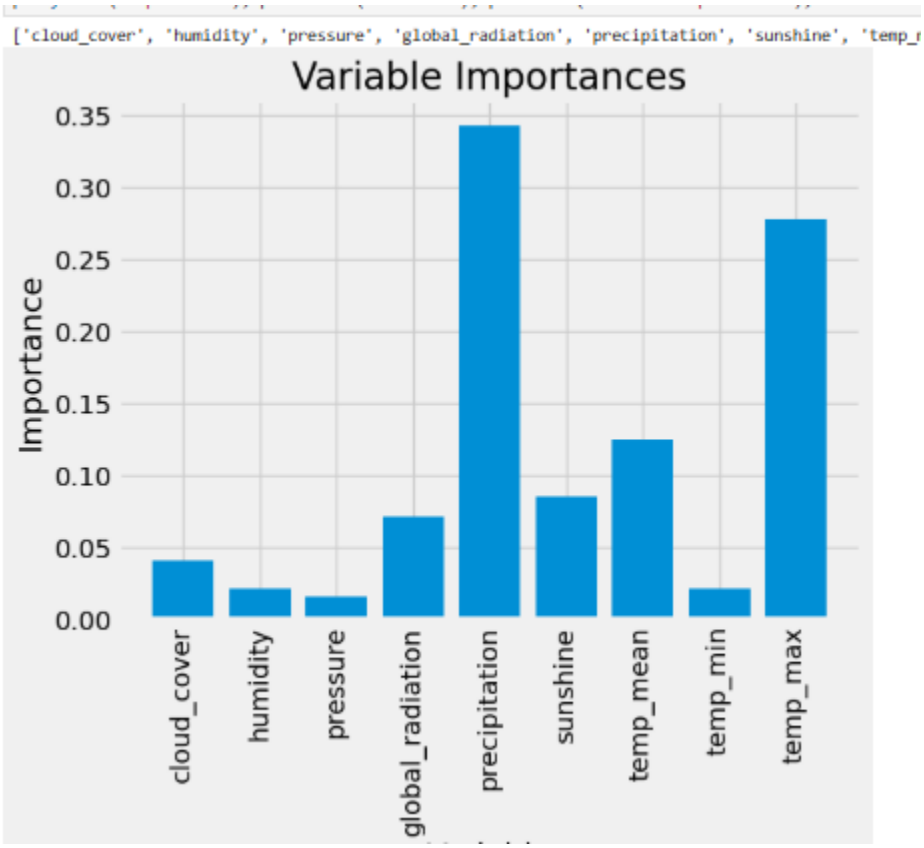
```
['Basel', 'Belgrade', 'Budapest', 'Debilt', 'Dusseldorf', 'Heathrow', 'Kassel', 'Ljubljana', 'Maastricht', '
adrid', 'Munchenb', 'Oslo', 'Sonnblick', 'Stockholm', 'Valentia']
```



Dusseldorf

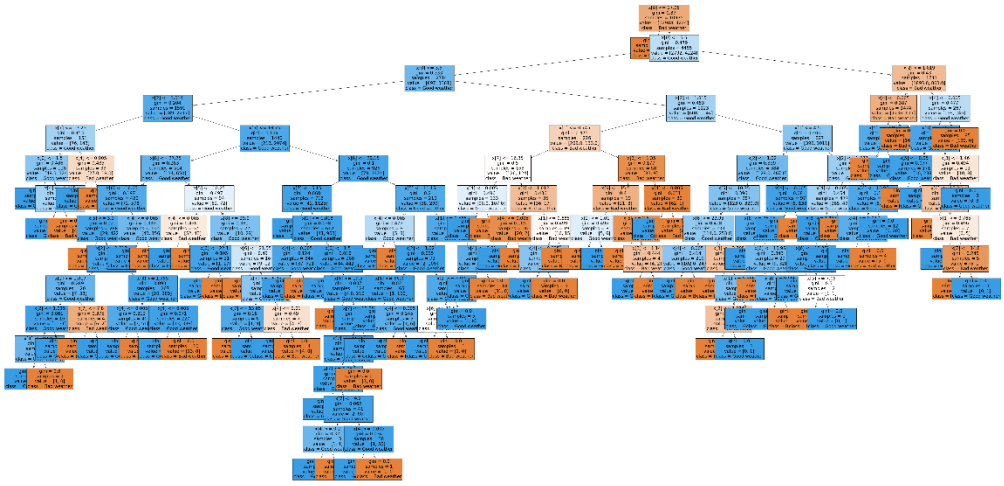
For Dusseldorf used n_estimators=100 and max_depth=15. The model was 100 percent accurate, sorting everything correctly into good and bad weather. The most important variables were precipitation, max temp, and mean temp.

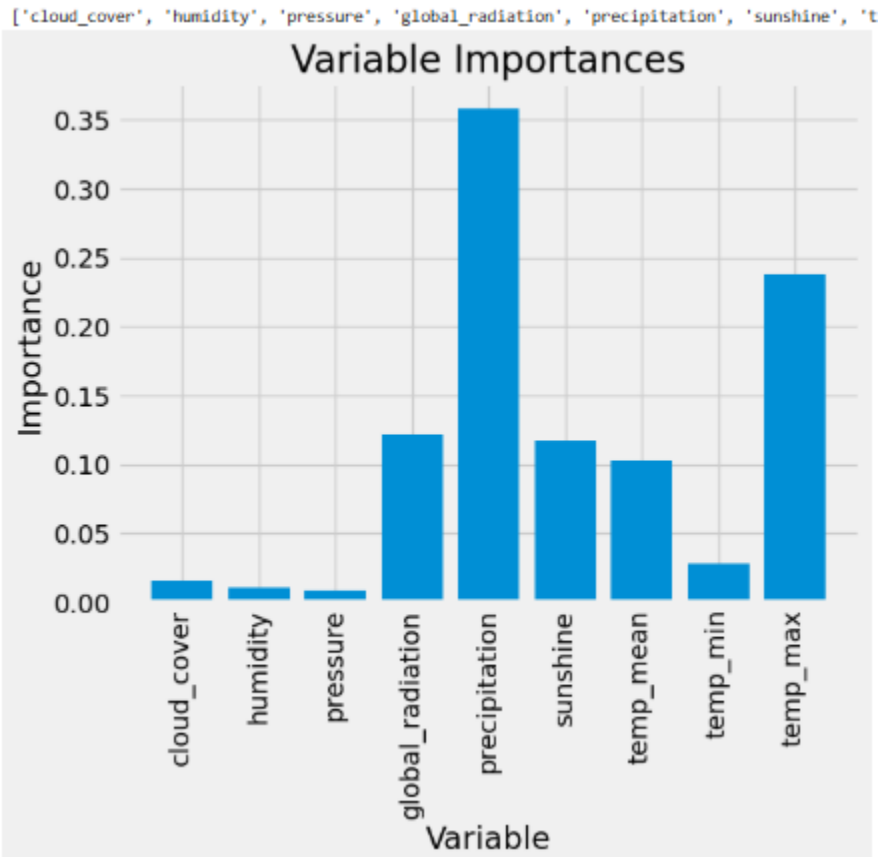




Basel

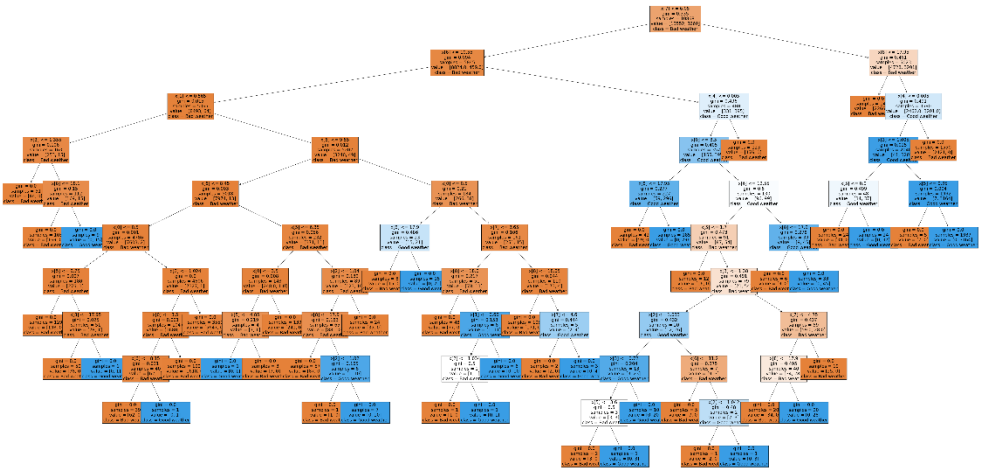
The Basel model was also 100 percent accurate with the precipitation, max temp and radiation as the most important factors.

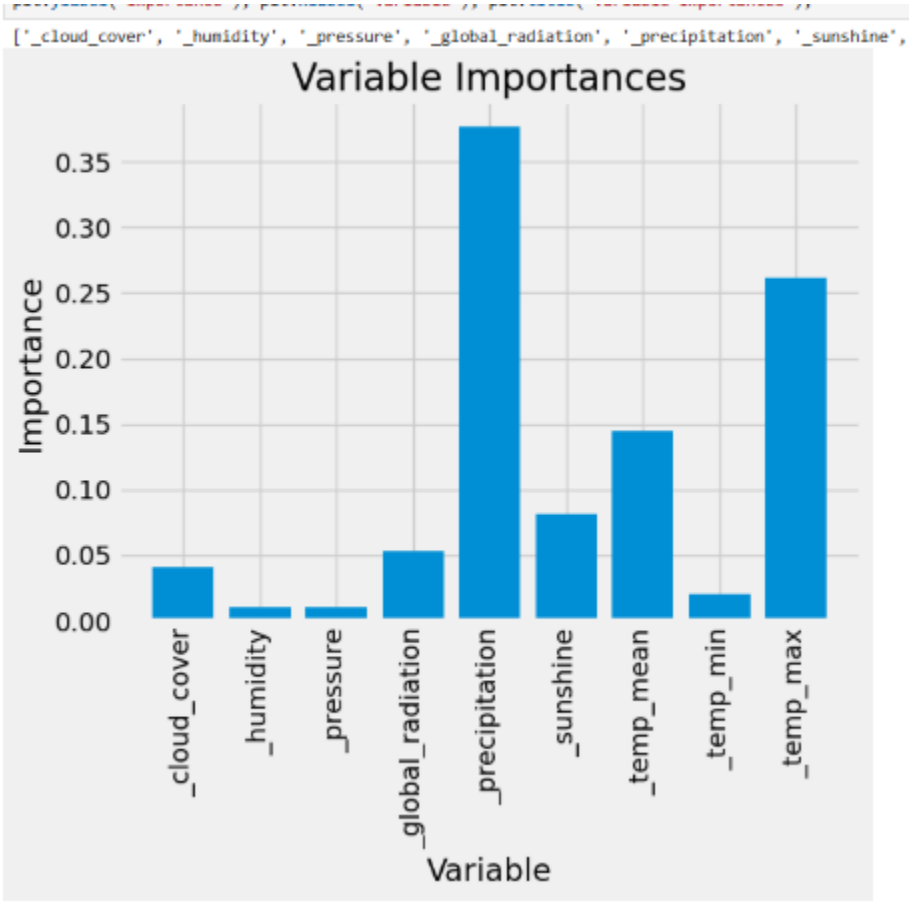




Muncheb

Also had 100 percent accuracy with the three most important factors of precipitation, temp max and temp mean





Conclusion

Precipitation and max temp were important factors at each of three individual weather stations in determining if a day would have pleasant or unpleasant weather conditions. ClimateWins should invest in equipment that accurately records these values. The apparatus that records temperature will also be able to collect mean and minimum temperatures as well, giving it additional value. I think it is important to note that we carefully assess these variables before using them to predict weather in other areas. The three cities we looked at all receive more than 30 inches of rain on average each year. Madrid, meanwhile, has an average of less than 20 inches. Weather and geography are so closely related that careful research needs to be done before assuming that the factors with the most impact in one city will have the same importance in another city.