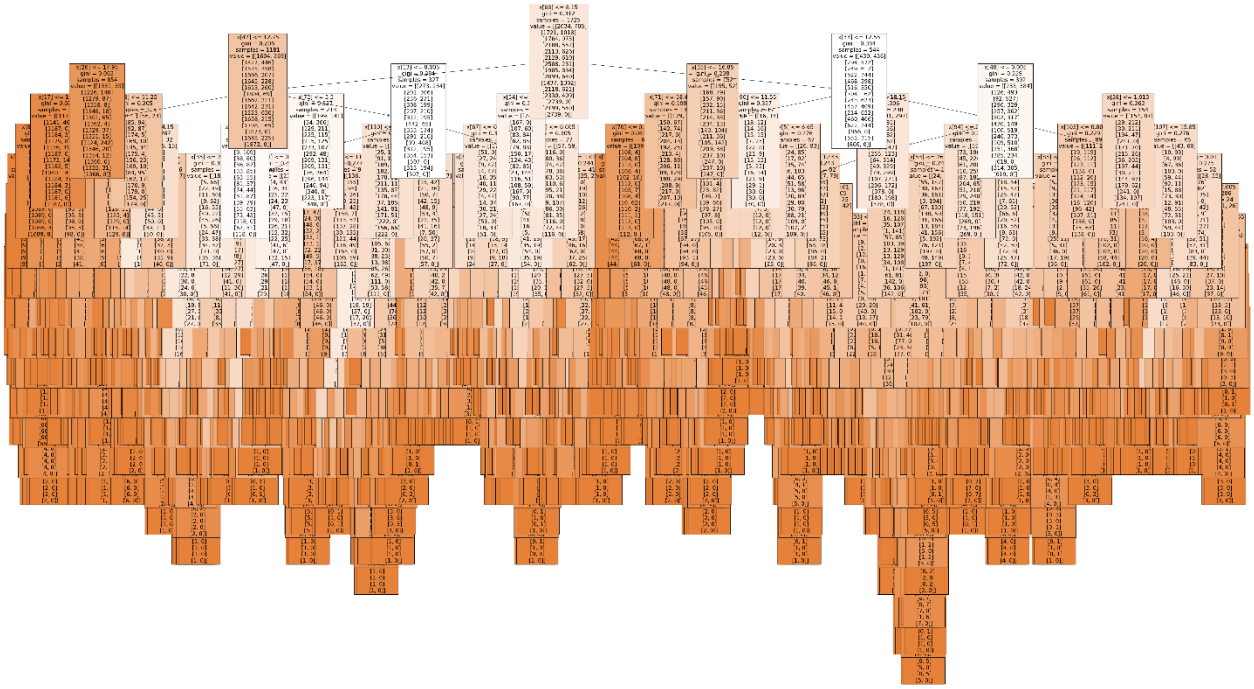
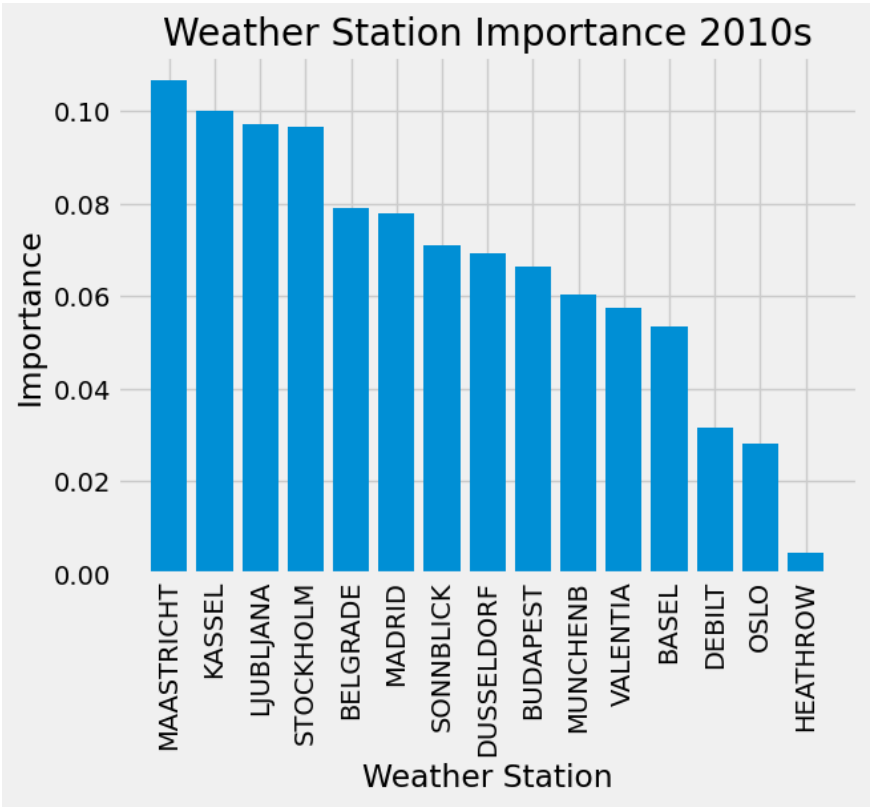
Random Forest and Feature Importance

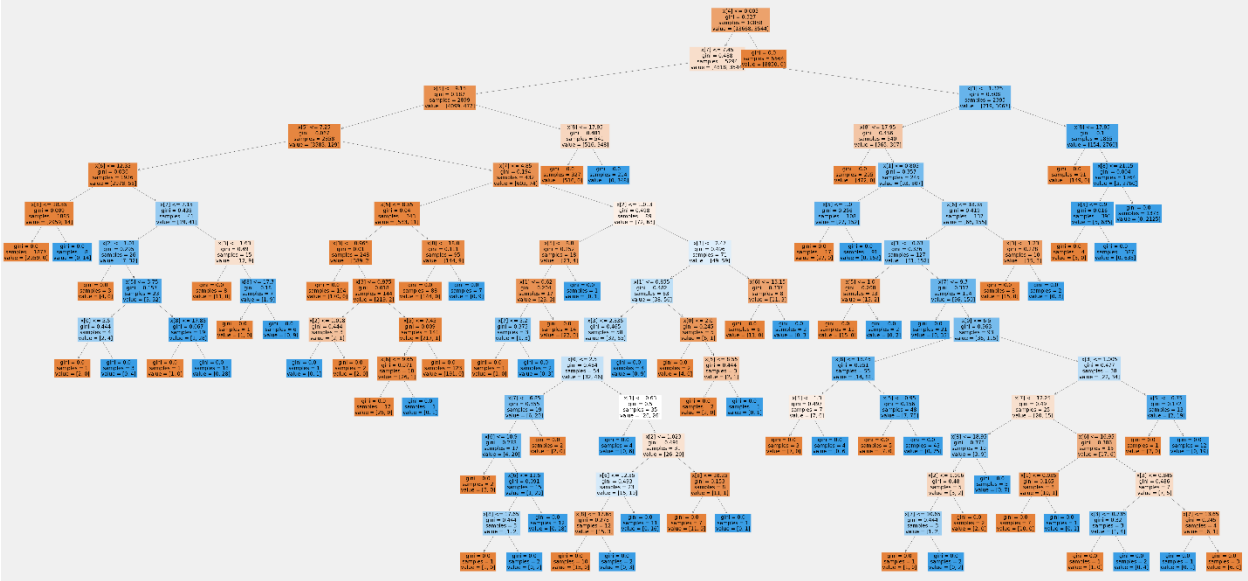


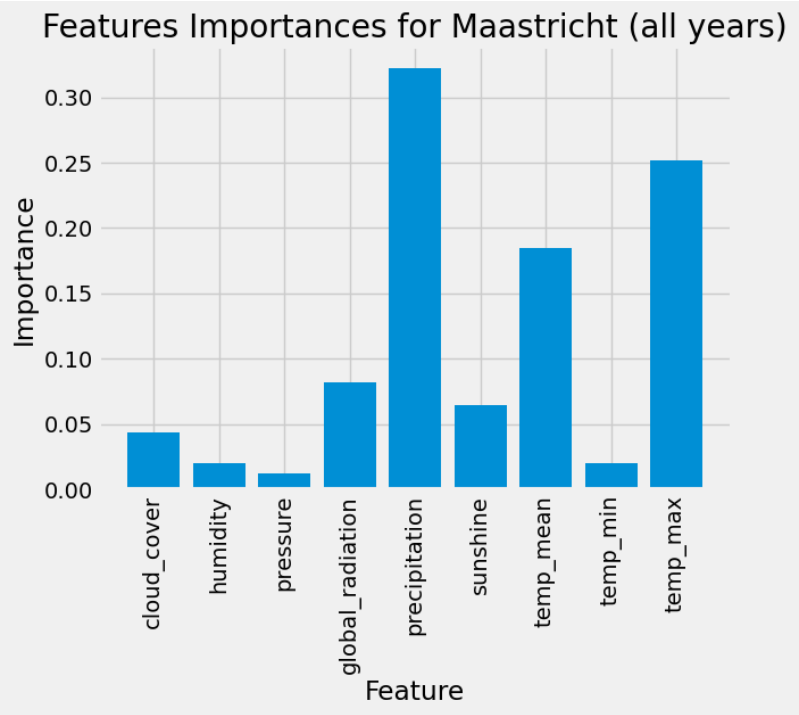


Initial model accuracy: 0.597

Top 3 stations: Maastricht, Kassel, Ljubljana

Maastricht Random Forest

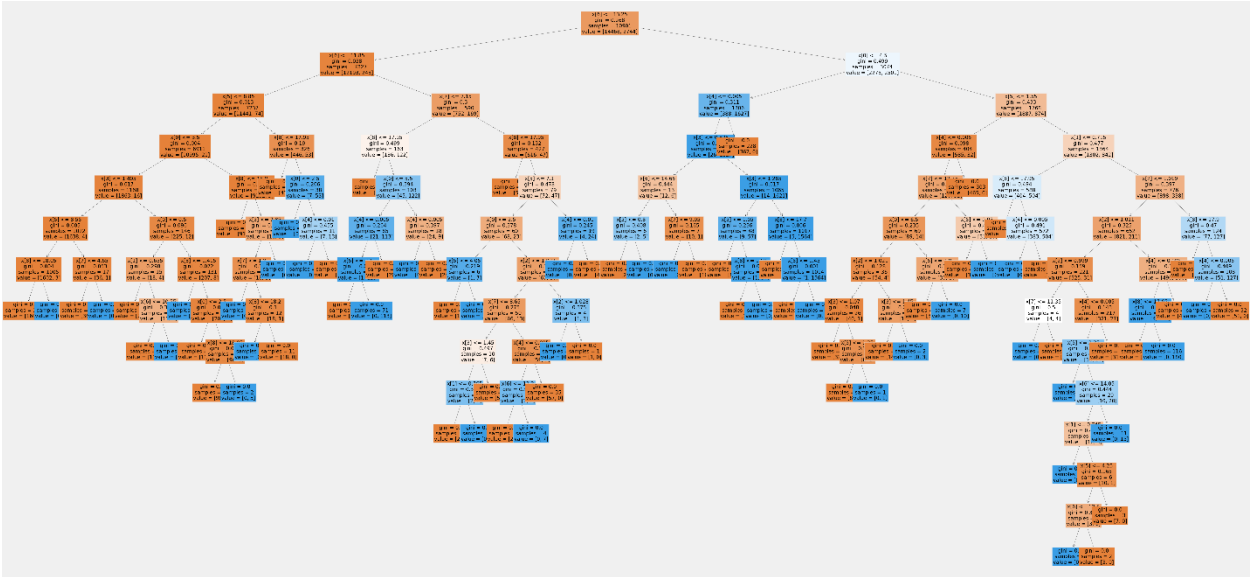


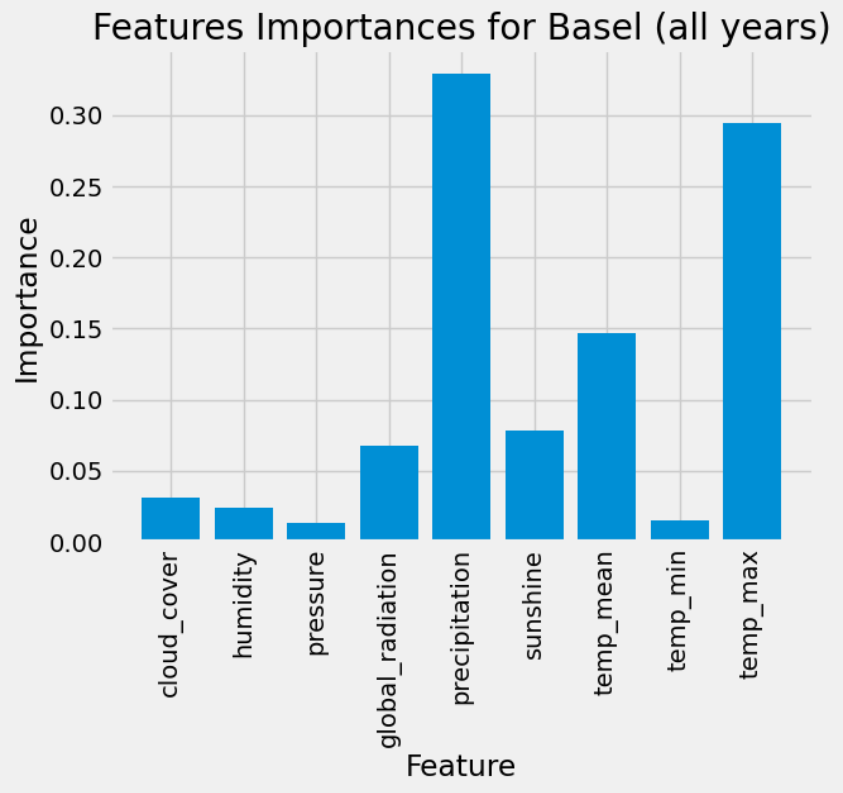


Model accuracy: 1.0

Top 3 features: precipitation, max temperature, mean temperature

Kassel Random Forest

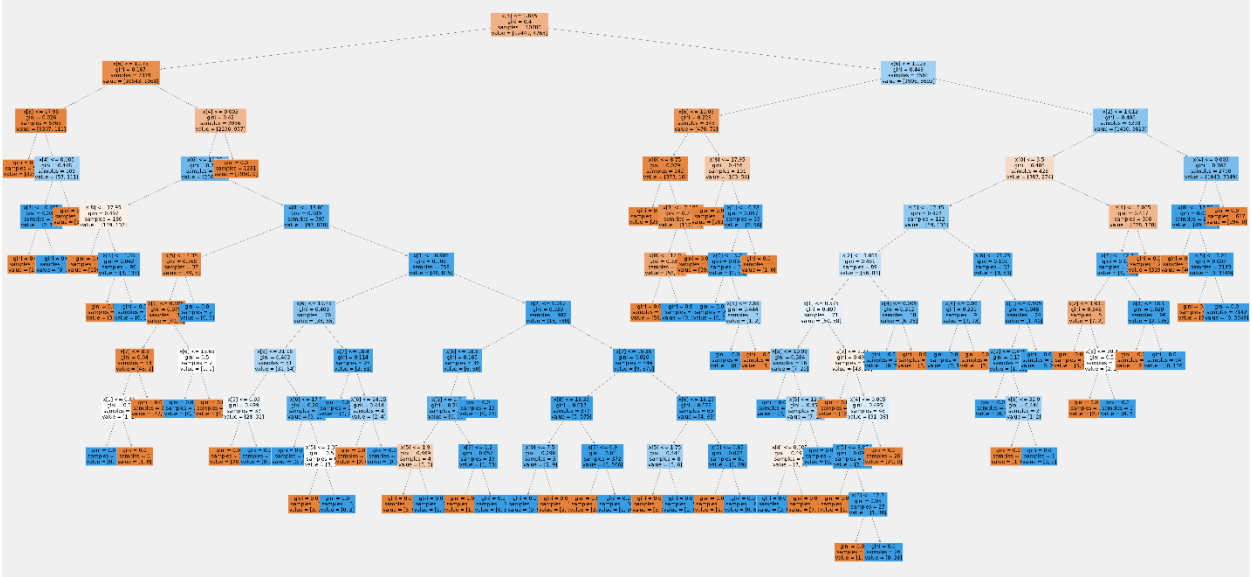


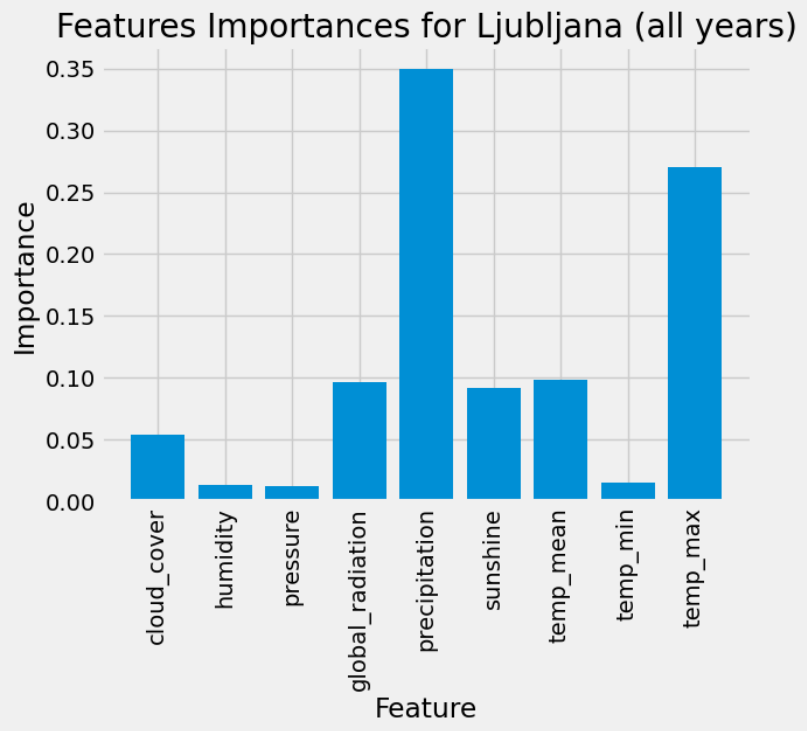


Model accuracy: 1.0

Top 3 features: precipitation, max temperature, mean temperature

Ljubljana Random Forest





Model accuracy: 1.0

Top 3 features: precipitation, max temperature, mean temperature

Conclusion and recommendations for ClimateWins: For each of the top 3 stations (Maastricht, Kassel, Ljubljana) the top 3 most important features were the same: **precipitation**, **max temperature**, and **mean temperature**. In other words, they were the

most useful when it came to discerning whether the day's weather would be pleasant or not in a given area. Therefore, these three variables will also be paramount in our ability to predict weather patterns and overall climate change in the coming years. This is especially important given the rising global mean temperatures over time. If we want to stay ahead of the curve and have a good idea of what's going to come, I suggest investing in our temperature detection tools and radar technology for detecting and tracking storms. The better idea we have of the temperatures to come, and any potential storms in the area, the better prepared we will be in the short term. In the long term, continuing to track and analyze precipitation and temperature trends will be useful for us in planning for the future.

Now, one note of caution: these three models produced a training accuracy of 100%. While this initially sounds great, it instead can be an indication that the model is overfitting; rather, the model's complexity allows it to memorize the data it was given, rather than generalize patterns within the data. What this means is that the same model may not perform as well on unseen data, which would be problematic for us, since we're relying on it to help us make important business decisions. Despite the overfitting problem, it is encouraging to see the same three variables emerge as the top three contributing features for each model. **Before making investment decisions** like I recommended above, it would be prudent to test these models with additional data from unseen periods of time and validate the findings on other models to ensure that precipitation, max temperature and mean temperature are playing an important role across other stations.