**Machine Learning Model**

**Data Source**: ml\_data.csv

**Value:**  ‘high\_q’

Description: The ‘high\_q’ value was set based on the top 69% quality of life index ratings. If a city was in the category, ‘high\_q’ was scored with a 1.

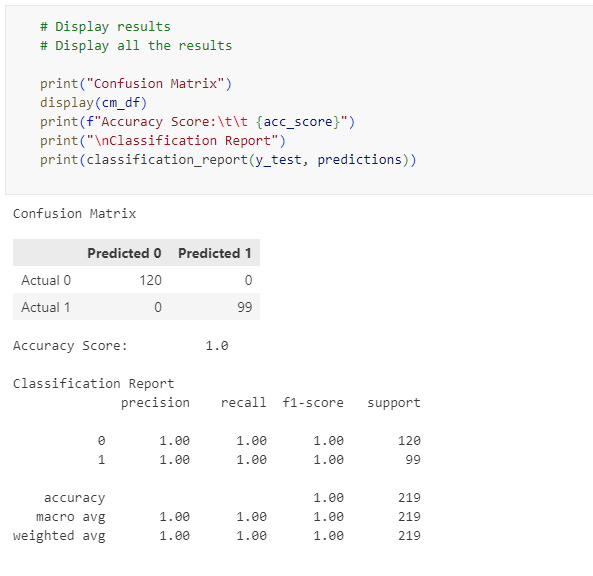
**Features:**’max\_temp’, ‘humidity’, ‘cloudiness’, ‘wind\_speed’, ‘2017’, ‘2018’, ‘2019’, ‘2020’, ‘2021’, ‘purchase\_power’, ‘safety’, ‘health\_care’, ‘cost\_of\_living’, ‘prop\_to\_income\_ratio’, ‘traffic\_comm\_time’, ‘pollution’, ‘climate’

Description: The features listed as years each represent the amount of total government expenditures, by state, on total outdoor activities including travel and tourism.

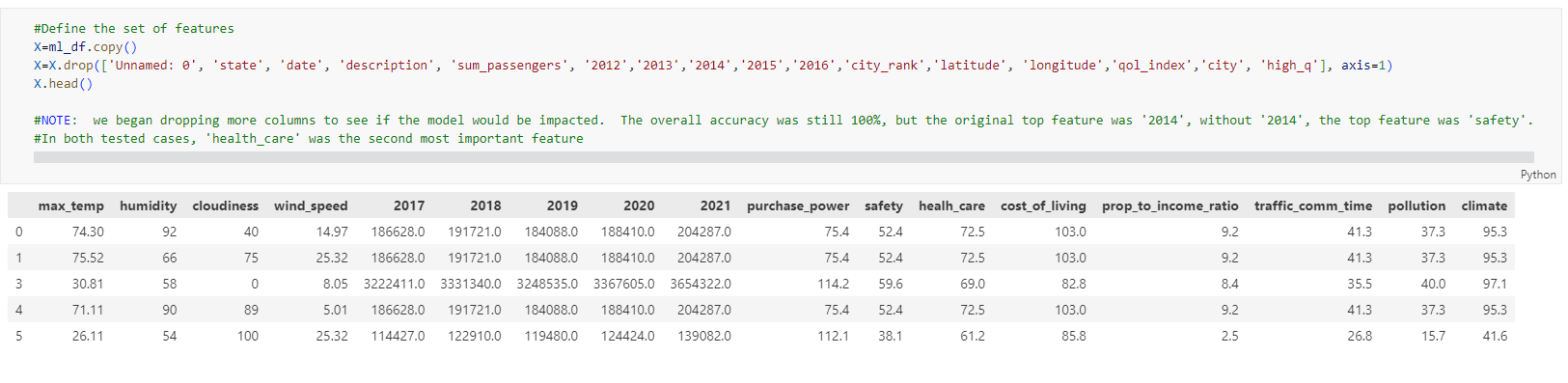
**Random Forest Model**



**Outcome:**



The Random Forest model produced 100% accuracy on every attempt. After the first attempt of running the model, several features were added to the ‘drop’ list.

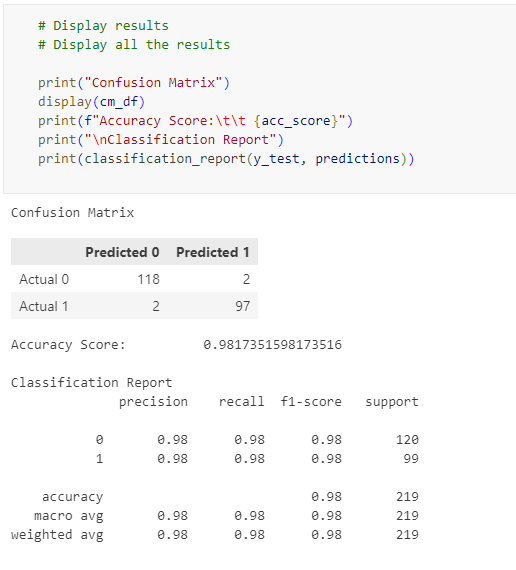


Regardless of the features kept for the model, or the number of n\_estimators, Random Forest produced 100% accuracy for this data set.

**Logistic Regression Model**



**Outcome:**



The Logistic Regression Model produced a 98% accuracy score keeping the original set of features.

**Results**

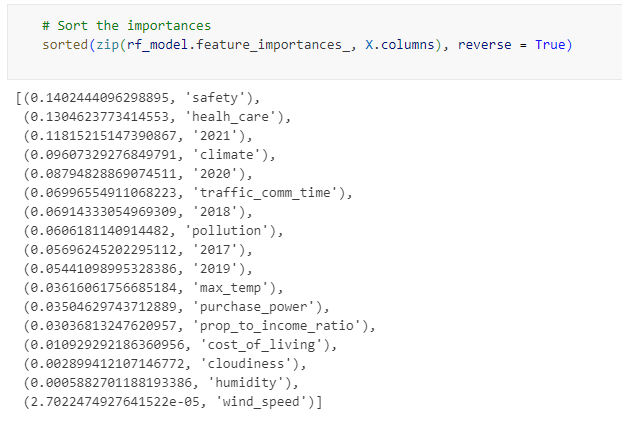
**Issues:**

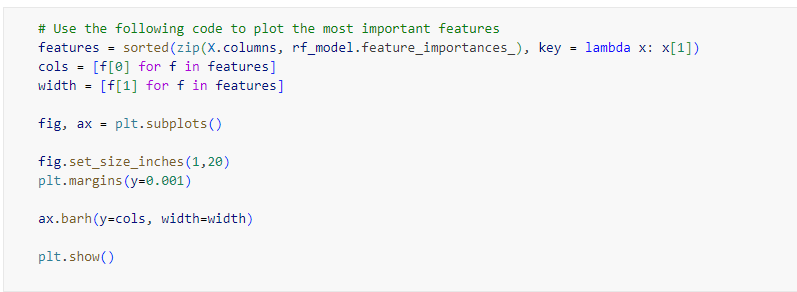
The data file used for the model was merged from four different sets of data, each with differing values and amounts of data. The issue was how the Quality of Life Index was transferred. The original file with this data represented only 60 cities in the US. When the data was merged with the 2000 city file, the Quality of Life Index was merged on “state” to ensure that most of the 2000 cities had that data point. In the merged file, any city in the same state would have the same Quality of Line Index score.

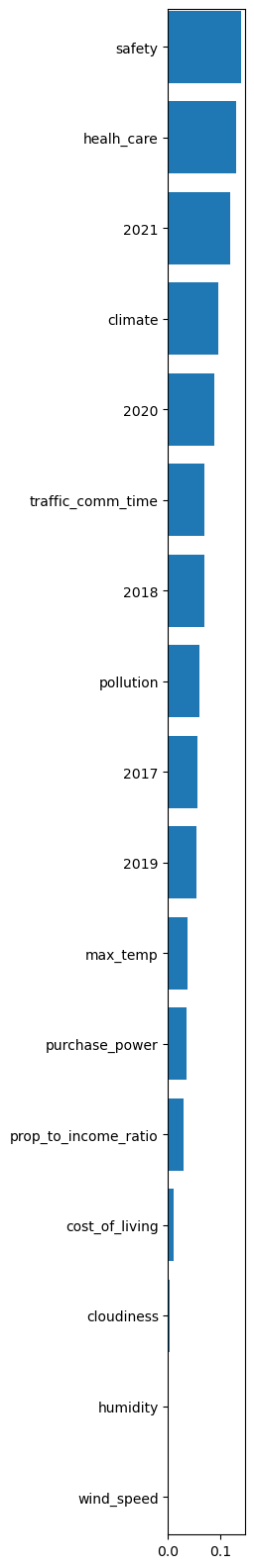
**Takeaways:**

The strength of the Random Forest Model is not the 100% accuracy, it is the ability to measure `feature\_importances\_`

**Feature Importances**







This represents the 2nd set of measured Feature Importances. In the first set, the features of “2012” through “2016” were kept. The feature ‘2014’, representing the government expenditures of outdoor activities, by state, was the top feature in the Random Forest Model, with ‘health\_care’ as 2nd. When limiting the features to the past five years of data for government expenditures by date, ‘safety’ became the number one feature.

As investors consider locations for property, safety, health care, climate, and government expenditures on outdoor activities would be factors that require serious consideration.