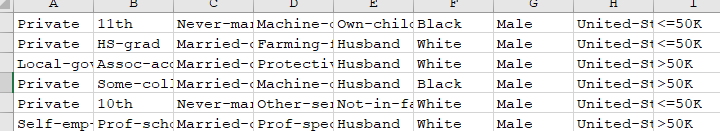
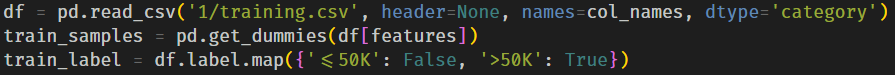
**Part 1:**

Removal of the unknown values and continuous attributes was done in excel. The first five entries of testing.csv is shown here:

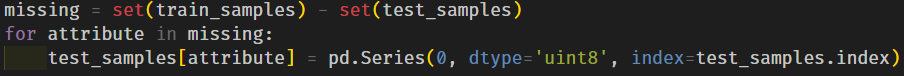
  
  
**Part 1 a:**

The decision tree was built in Python using Pandas and Scikit-Learn, and this guide was followed: https://www.ritchieng.com/machine-learning-evaluate-classification-model/  
  
There were two tricky parts:

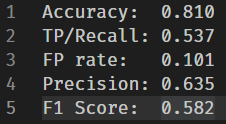
* Scikit-Learn does not accept categorical attributes, so pandas’ get\_dummies() function was used:



* The test labels were missing an attribute value that existed on the training labels. The missing values were found with the set difference, and then inserted with 0 values. A for loop was used to keep the code more general.



The results were:



**Part 1 b:**