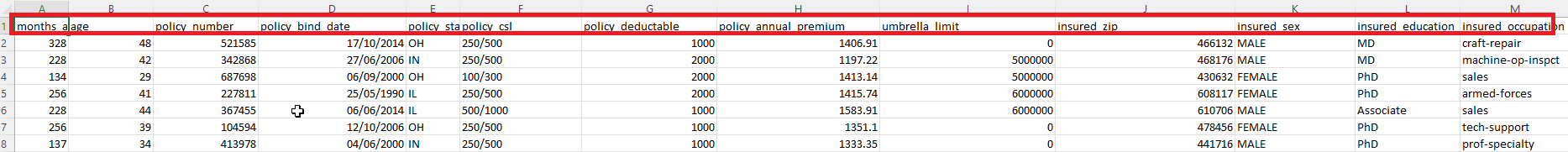
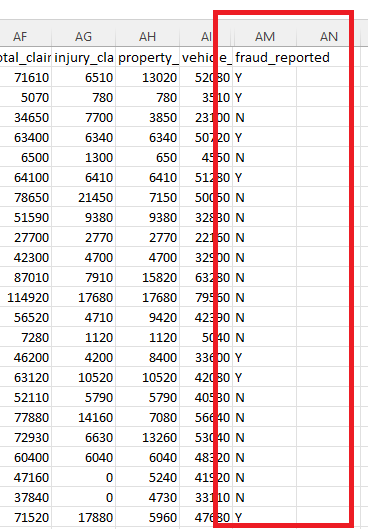
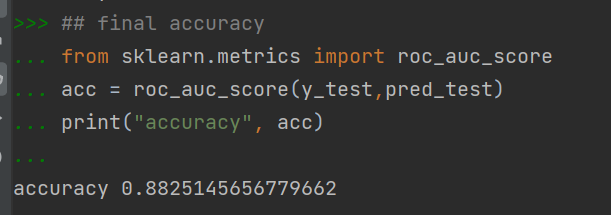
Dataset review:

<https://github.com/mwitiderrick/insurancedata>  
looking at the dataset, I like what I am seeing. It does have information about the applicants.   
  
  
  
The dataset also has labels to tell about the claim if it false or real claim.   
  
  
  
From experience, this sort of dataset is more suitable for tree based modelling, i.e random forest, XBoost etc.   
  
The only drawback is that it has only 1000 samples ! but it can be a start-up point. We may think to augment more samples from this dataset. That is a nice review regarding exploitability data analysis (EDA) on this dataset <https://www.kaggle.com/buntyshah/insurance-fraud-claims-detection>  
  
  
  
Results

Accuracy about 90 %



Explanation

Feature importance

