# Decision Tree Fraud Check algorithm assignment

Problem statement: Use decision trees to prepare a model on fraud data

Treating those who have taxable\_income <= 30000 as "Risky" and others are "Good"

Data Description:

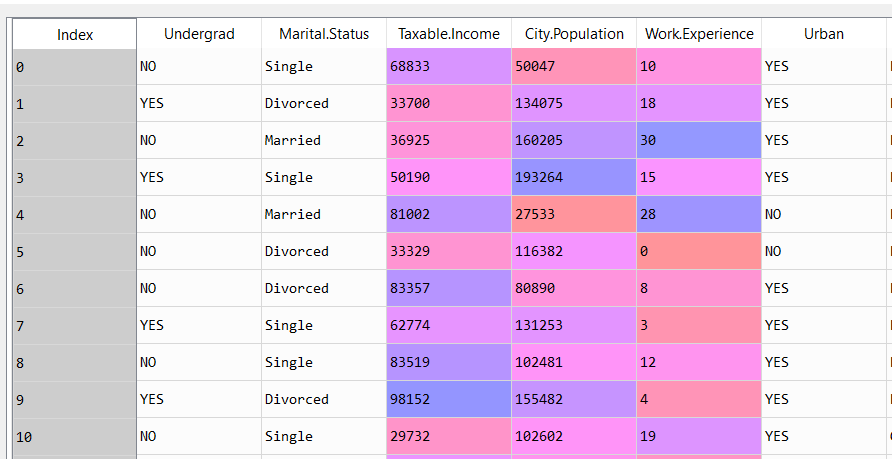
Undergrad : person is under graduated or not

Marital.Status : marital status of a person

Taxable.Income : Taxable income is the amount of how much tax an individual owes to the government

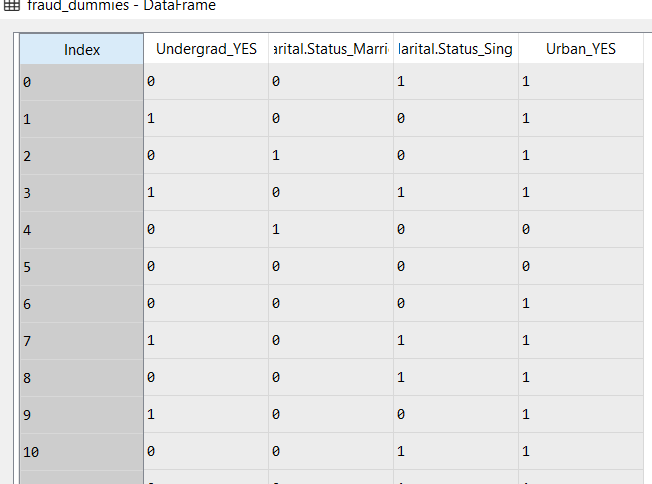
Work Experience: Work experience of an individual person

The data looks like the following:

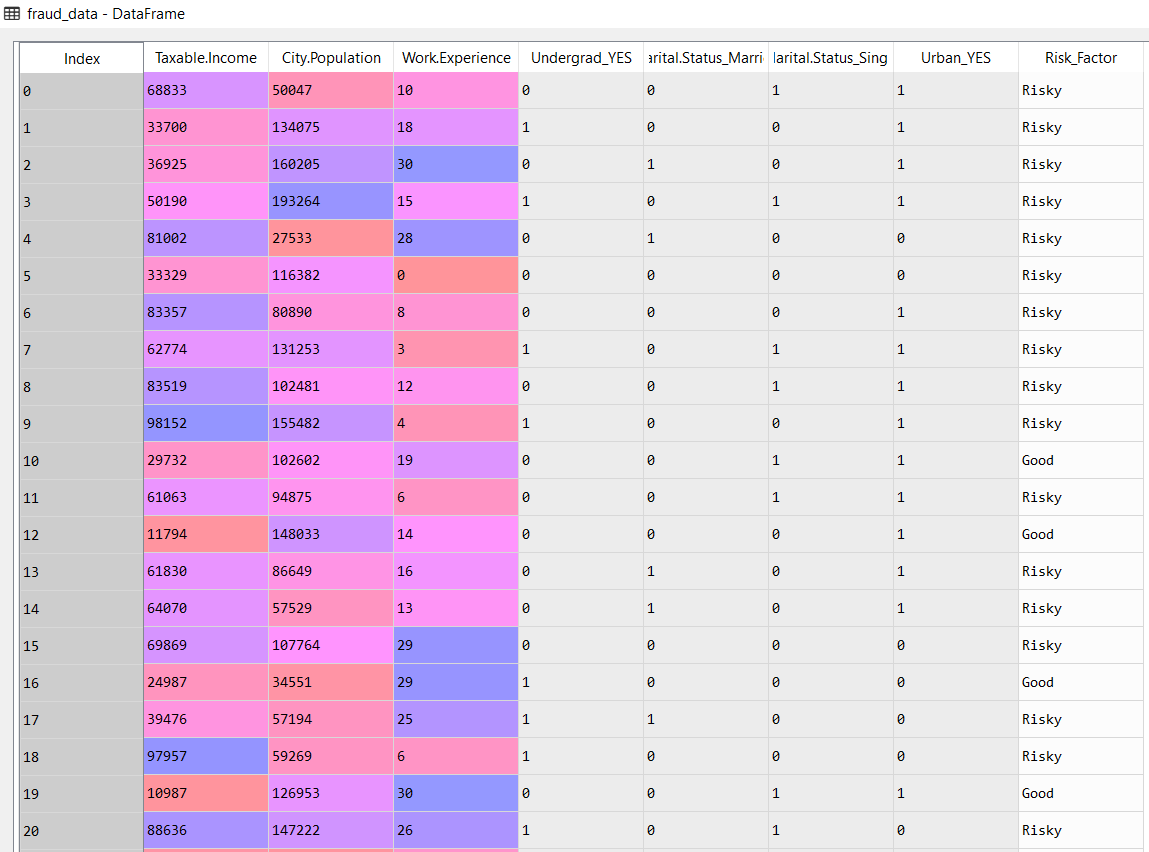


In the above data we have the variables Undergrad, Marital Status and Urban as the categorical variables and the rest Taxable income, City population and work experience as continuous variables

Of the above categorical data, we create dummy variables as follows:

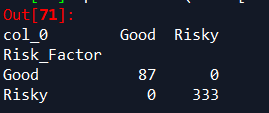


With the above converted categorical variables, the continuous variables will be concatenated like the following:



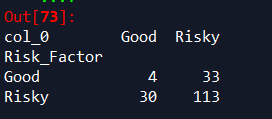
We split the data into training and test data and we pass the train data into the decision tree algorithm

First we pass the train data and obtain the following results:



From the above results we have zero True Negative and False Negative. Hence we get an accuracy of 100%

We pass the test data and obtain the following results:



The accuracy of the test data is 0.65