Capstone Project II

Transforming Variables

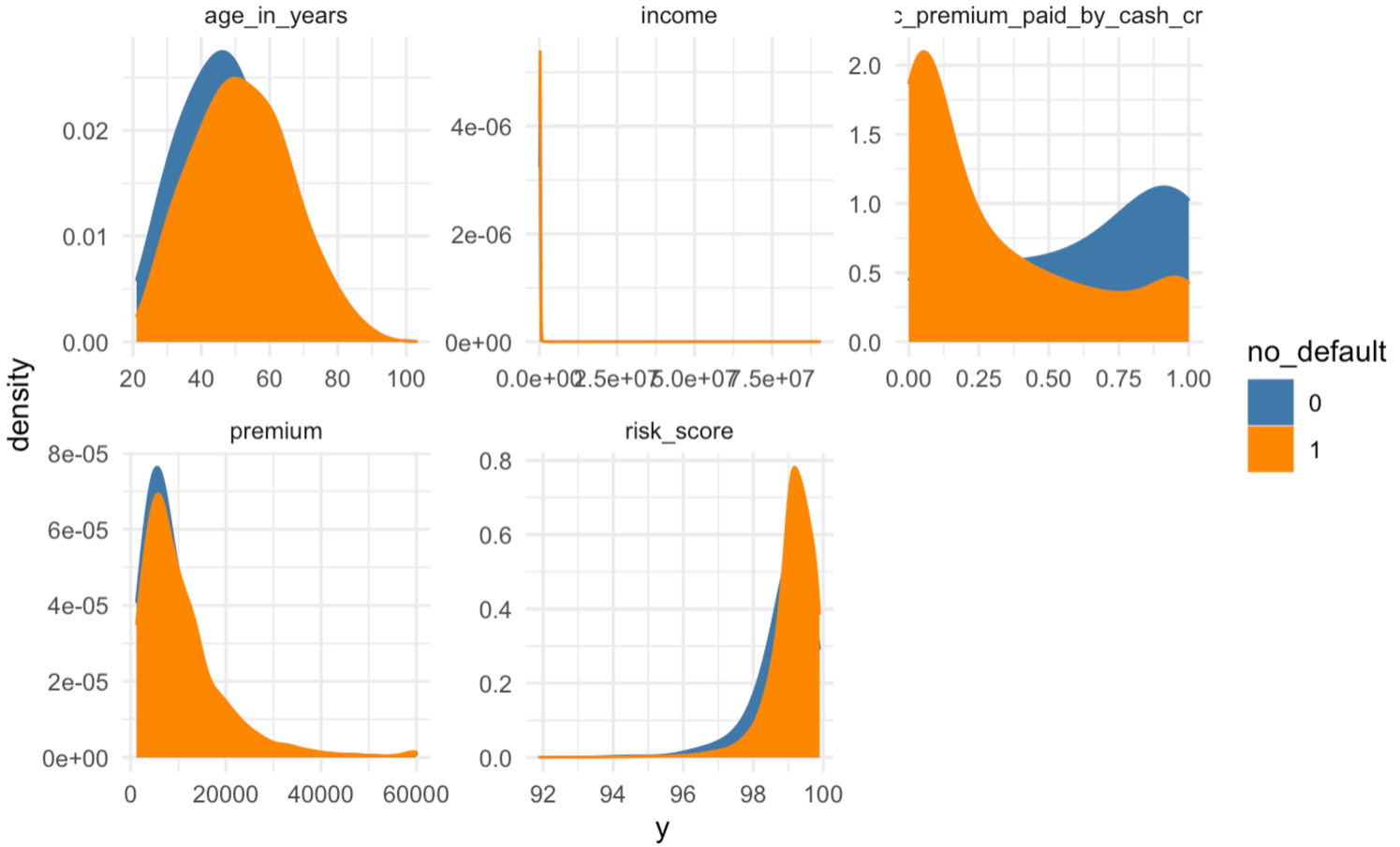
There were no missing values in the data

Age in days was converted to age in years

A variable called the late payment was created to measure whether a person missed a payment or not

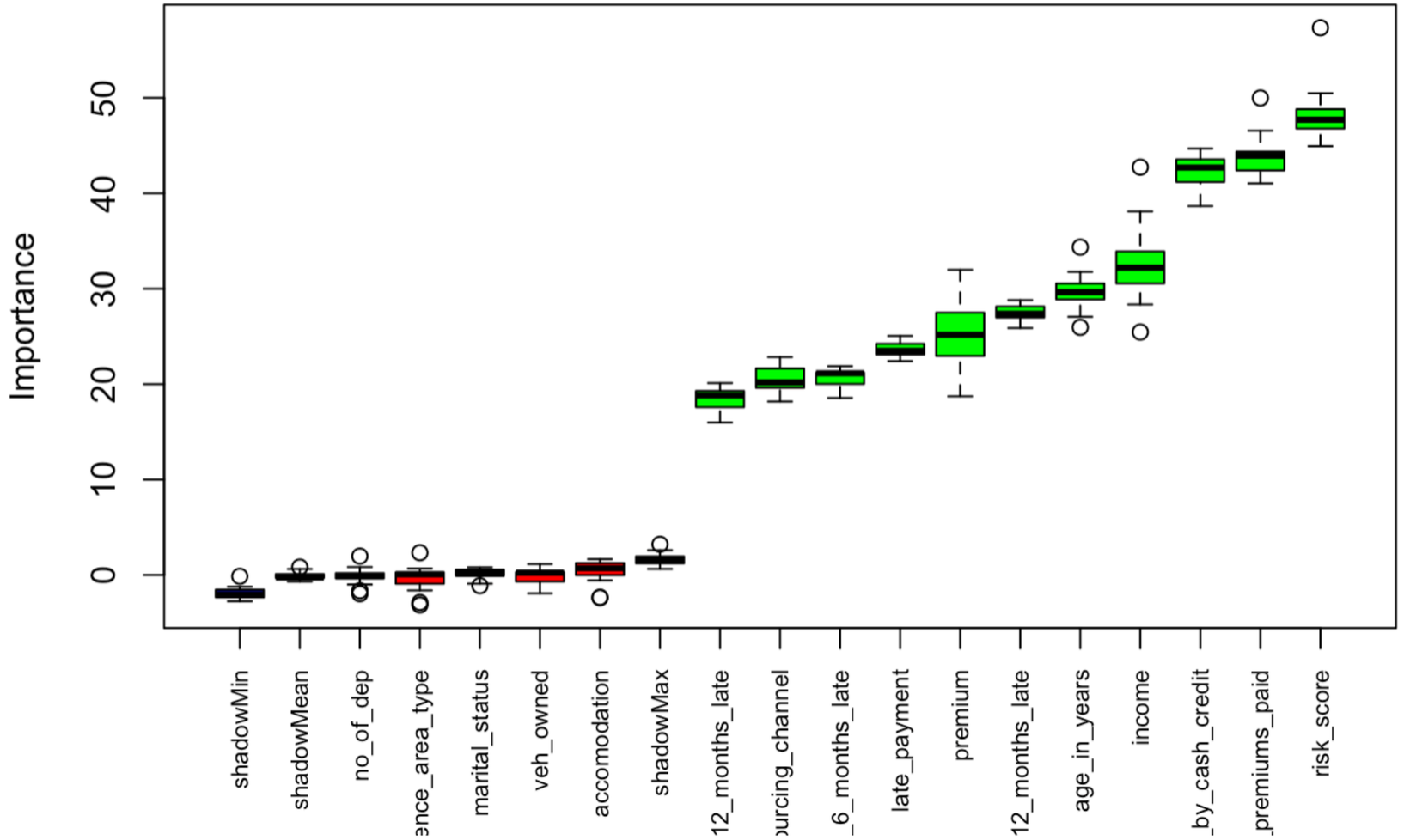
The premium has large values which are outliers (some of the clients make more than most), a decision was taken not to delete the outliers but to model using a method resistant to the effects of outliers

IMPORTANT VARIABLES



Individuals with a higher percentage of their premium paid by cash or credit are more likely to default

Important Variables



The most important variable in predicting whether someone will leave to stay is their risk score followed by the number of premiums paid

Variables like the number of dependents, residence area type

Marital status, vehicle owned and the accommodation are not meaningful interdicting the important variables

Analytical Approach

A random forest classification model or a bagging model can be built to predict the class.