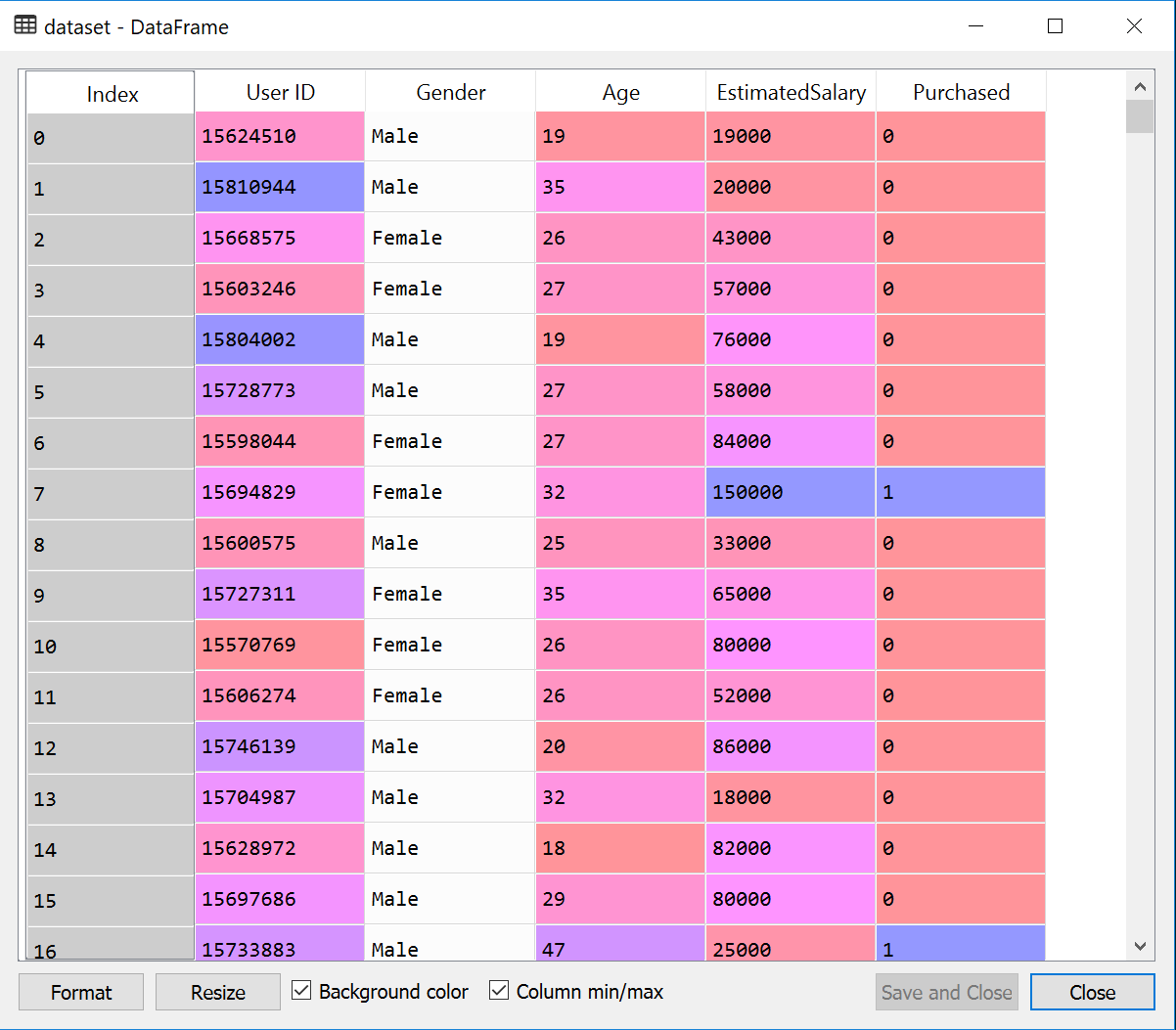
**Project Documentation – Turcu Gabriel**

The dataset:



The columns that we will need from our dataset are: Age, Estimated Salary and Purchased.

Our dataset represents a database of customers that have been approached for a commercial campaign to buy a SUV and whether or not they actually ended up buying the car. We want to see if we can predict for a new set of customers if they’re going to buy the SUV or not.

1. Loading the dataset.
2. Splitting the dataset into Training set and Test set with a test size of 25%.



1. Performing feature scaling on the Age and Estimated Salary.



1. Fitting the 4 models on the Training set.



For KNN, we set the metric to that value and p to 2, so our KNN algorithm would use the Euclidian distance.

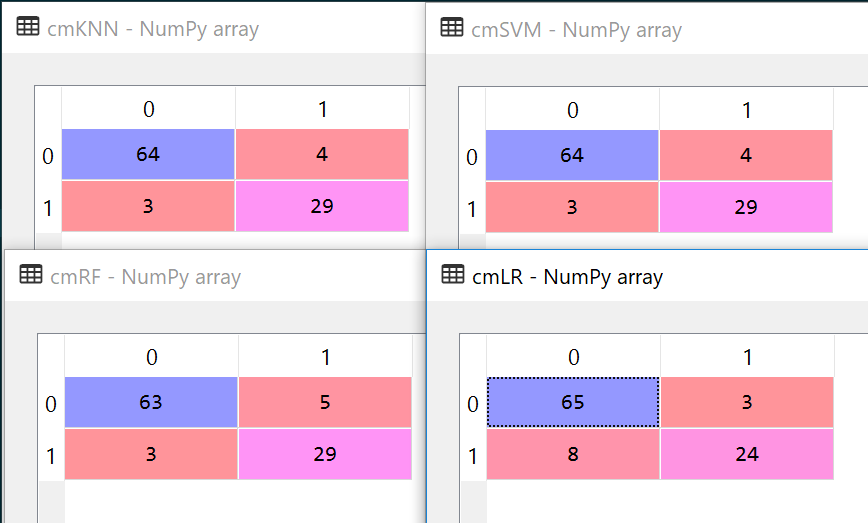
1. Predicting the Test set results



1. Making the Confusion Matrixes



The confusion matrixes for the 4 models look like this:



The numbers on the main diagonal represent the number of correct predictions while the numbers of the secondary diagonal show the number of incorrect predictions.

1. Visualizing the Training and Test set results

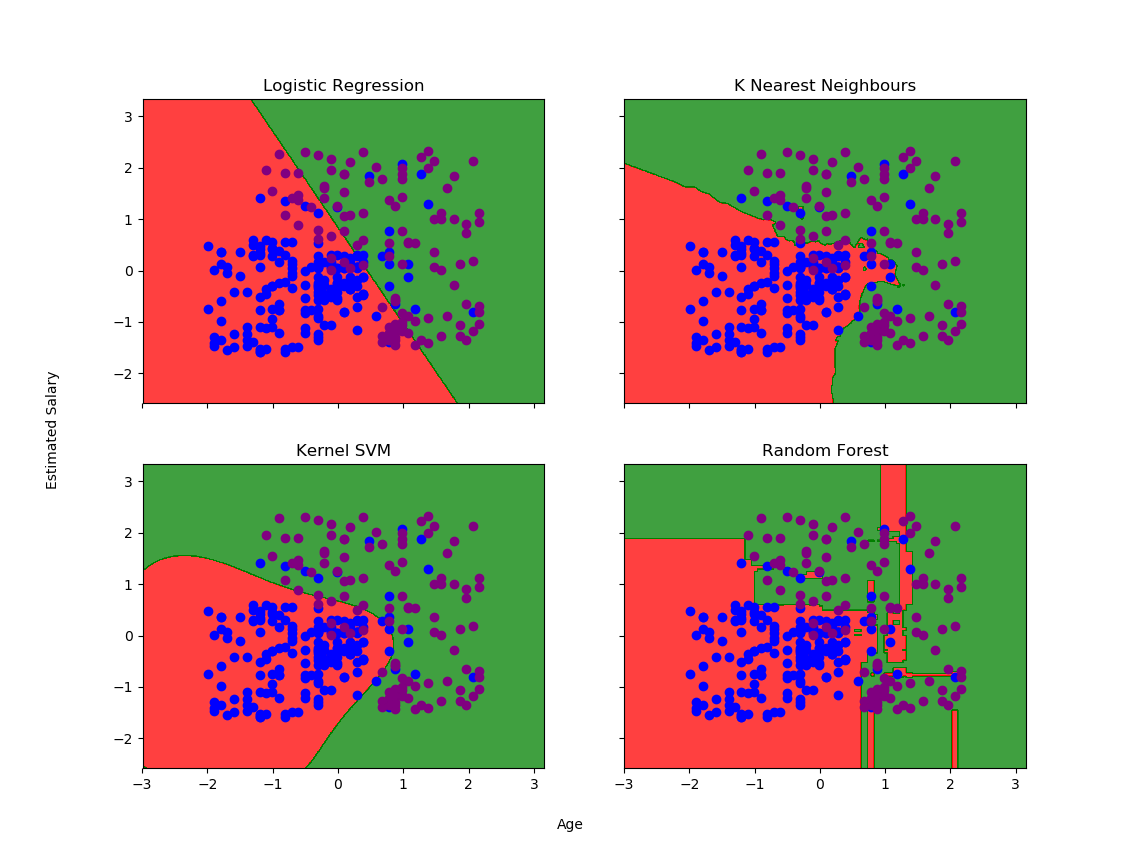


We create a figure with 4 plots so that we can visualize the 4 models at the same time.

Some observations and clarifications about the code:

* np.mesgrid is used to create a mesh grid of all the points that fall within our dataset and we use contour to draw them on the subplots.
* We use subplot to plot all 4 plots on the same figure, each contour having a different classifier.
* We then scatter the points from our dataset with the people that ended up buying the SUV and the people that didn’t.

For the Training set, the results for the 4 models look like this:



In the RED area, we have the customers that the model predicts are not going to buy the SUV, while in GREEN, we have the customers that the model predicts will buy the SUV. In BLUE, we have the customers that did not buy the SUV, while in PURPLE, we have the customers that actually bought the SUV.

This is how the models perform when ran on the test data:

