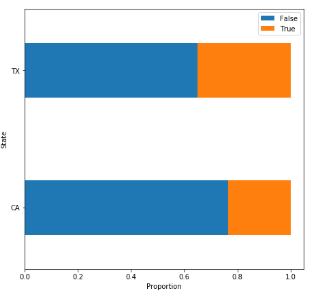
**Data Report**

The Data contains the details of people who made their First payment Detail, there are two sets of data train set to train the data and the test set to Predict the outcome of the loan. The prediction is done based on the variables in the data set.

I have chosen Jupyter notebook to do the analysis because it excels in a form of programming called “literate programming”.

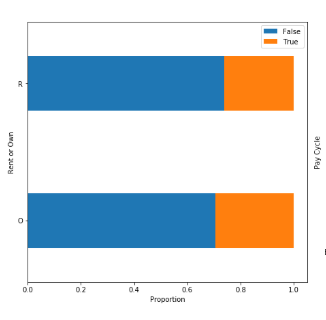
**Descriptive Analytics:**

Exploring state wise First payment Default:

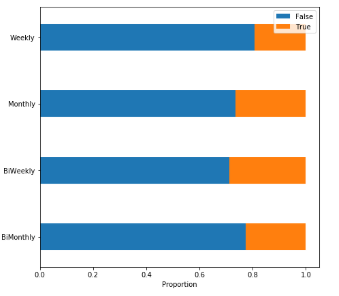


From the graph it is evident that in Texas there are less people with First Payment Default as False than in California.

Exploring if the Lenders have their own house or rented house:



For Rent or Own Graph there is almost same proportion of true and false, hence it does not matter wether the person is living in rent or owning.



**Solution:**

* For Pay Cycle we can see that if person has pay cycle monthly Bi-Weekly then there is least chance that his First Payment Default is False.
* In the dataset there are columns like setID which are of no use to us because they donot help in prediction.
* There is date of funding and loan due date, which are important, but it will take a considerable amount of time to get information from the dates, hence to apply my model, I have ignored those columns.]
* The Algorithm I have chosen to predict is Logistic regression, I have applied it and got an Accuracy of 79% which is because the data is less.
* From the algorithm if we enter the values of a customer with all the details the algorithm will be able to predict the outcome of the loan.
* I have also cross checked by using another Algorithm which is “Support Vector Machine”.
* Finally I got the Output in a file named Output file.