**PROJECT ANALYSIS**

**Finding the data**

**Main Data Source**

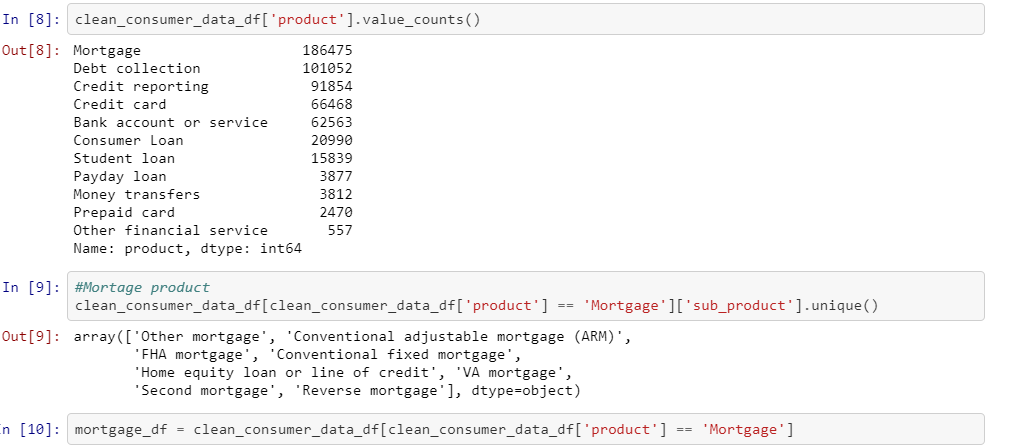
* <https://www.kaggle.com/cfpb/us-consumer-finance-complaints>
* <https://data.world/veteransaffairs/veteran-populations-by-state>
* <https://www.kaggle.com/doyouevendata/state-abbreviations>

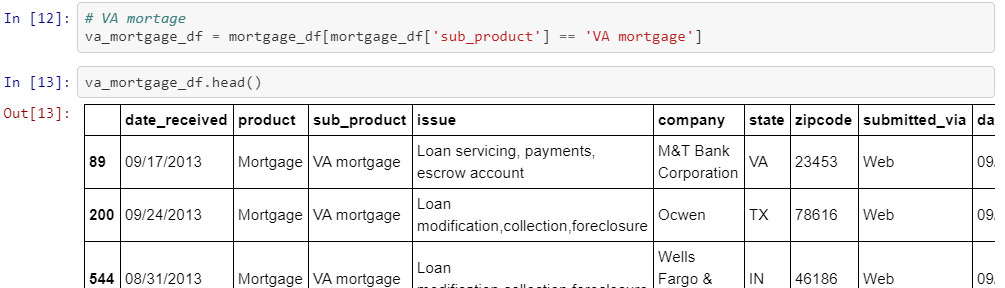
**Other Sources**

* [https://data.world/veteransaffairs/veteran-populations-by-state](https://data.world/veteransaffairs/veteran-populations-by-state" \t "_blank)
* <https://www.kaggle.com/doyouevendata/state-abbreviations>

**Data Clean-up and analysis**

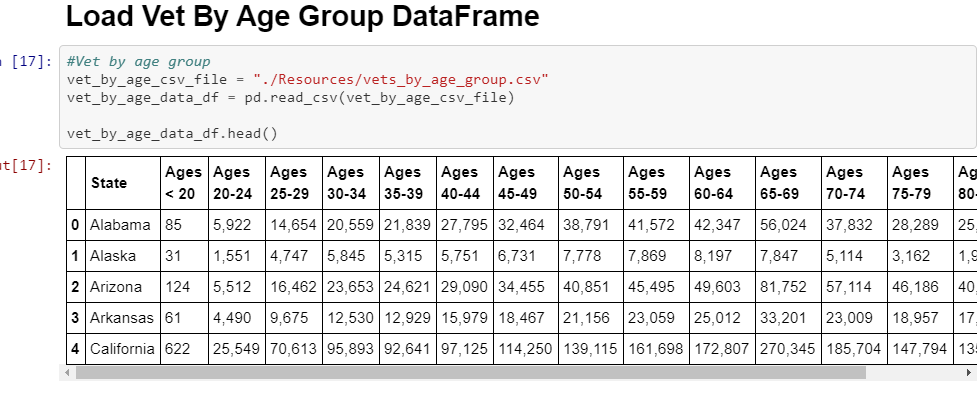
We put the consumer complaint dataset into a data frame using pandas library, we further trimmed down to the columns we needed and checking the number of items we had. Realizing that the highest number of products we had was mortgage then the sub product was the Veterans Mortgage. We dropped the null columns and remained with the columns that we needed. We narrowed down to the product mortgage and sub-product Veterans Mortgage and turned it into a data frame using pandas and renamed the some of the columns.





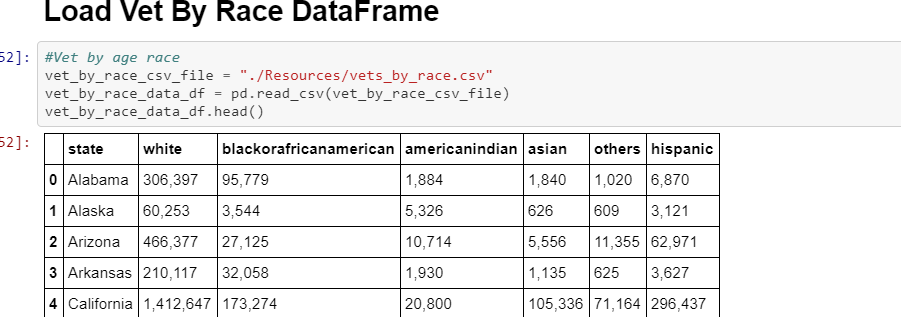
**Age-group data cleaning**

We obtained the veterans data and loaded the csv file using pandas library and converted it into a data frame , the data was pretty much clean so we did not have to do much cleaning. Below is a screenshot of the procedure we used;



**Race data cleaning**

We got the Veterans by race dataset since it was a csv we loaded it using pandas and converted it in to a data frame here is a screenshot of how we did it;



**State data cleaning**

Since the states in the consumer complaints data were in abbreviations we decided to obtain the data for all the states ; 