**Name:** Taha Shahid

**Capstone Project 1:** Data Story for Default of Credit Card Clients dataset

Some of the questions that were investigated from the surface of the above-mentioned dataset are:

1. Is the proportion of defaults the same for men and women?

Ans: It was found out that Proportion of male defaulters is 24.16 % and Proportion of female defaulters is 20.77. Also, total number of males in data was 11888 and total number of females in data was 18112.

2. Does Education level matter for predicting default or not default?

Ans: It was found out that the highest numbers of defaulters were university students with 23.74%, second highest were High School student with 23.45%, graduate students had a defaulter proportion of 19.24% and all others has the lowest default proportion of 5.7%.

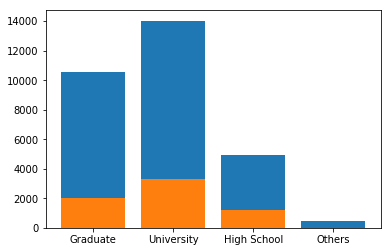
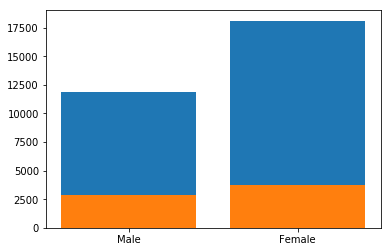
3. Does credit card limit have a connection with being a defaulter or not?

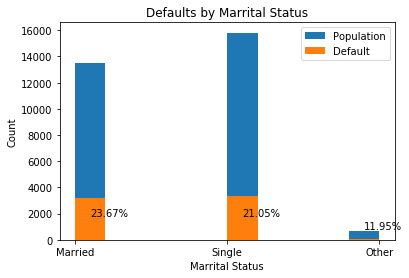
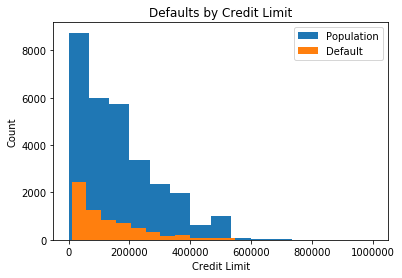
Ans: Proportion of defaulters from 0 limit balance to 200000: 25.34%. Proportion of defaulters from 200000 limit balance to 400000: 15.48%. Proportion of defaulters from 400000 limit balance to 600000: 12.28%. Proportion of defaulters from 600000 limit balance to 800000: 7.69%. So, it was found out that with higher credit limit balance there were fewer number of defaulters.

4. Does Marital Status have anything to do with being a defaulter or not?

Ans: Married defaulters were the highest at 23.67 %. Single defaulters were 21.05 %. And all others were 11.95 %.

So, from above four questions it was found out that generally men had a higher percentage of default as compared to women. Secondly, students in university and high school had some of the highest percentages when it came to their accounts being defaulted. Thirdly, credit card limit balance almost had a direct correlation with the account being defaulted, as the limit balance increased lower the number of defaulters. Lastly, married and single people had a pretty close default proportion but married people for some reason had a higher proportion of default. The above finding will be investigated further using exploratory data analysis. To get a visual of the above questions investigated below are the four graphs.



****