# Questions

1. In the dataset the column credit limit what does it indicates whether it is monthly or average credit limit?

Answer: Monthly credit limit.

1. What does it mean when the data have an Avg\_Utilizatio\_Ratio of 0?
2. Is it 0 because they can no longer use the credit card since they are already churned or
3. is it 0 because they havent used it for a few months?

Answer: It is the ratio of (credit card spent + money withdrawal)/(Total available limit for credit card spends + Total money withdrawal limit)

1. How would you treat ‘Unknown’ label?

Answer: replace them with nulls first, since they don't provide any information. Thereafter imputed them with values based on K nearest neighbor to remove all nulls. However, all three have minimal effect on the dependent variable, so they can be left alone as well

1. Could you please elaborate this column? What do "Total\_Trans\_Amt" and "Total\_Revolving\_Bal " mean?
2. What is the difference?

Answer: Total\_Trans\_amt is the sum of transactions one has done in the last 12 months. This basically tells us the total usage of a credit card by the user.

Revolving balance is the unpaid amount that carries off on your next credit card's cycle. Total\_revolving\_bal would be its sum. This tells us the amount that they don't pay on time.

1. However, I'm not sure what the attributes "Total\_Relationship\_Count", "Contacts\_count", "Total\_Amt\_Chng\_Q4\_Q1", "Total\_Ct\_Chng\_Q4\_Q1" represent for the costumers.
2. Furthermore, how was "Avg\_Utilization\_Ratio" calculated? Do you think it should be included in the training of a machine learning model?

Answer:

Total\_Relationship\_Count: total number of products held by the customers (cards, accounts, etc.)

Contacts\_count\_12\_mont: I guess it holds the number of times the bank contacted the customer and/or viceversa. There doesn't seem to have any relationship with other fields (e.g. contact clients that left the bank, customers with revolving balance, etc). I guess it has to do with ad campaigns.

Total\_Amt\_Chng\_Q4\_Q1: represents how much the customer increased their expenditure when comparing the 4th quarter agains the 1st.

Total\_Ct\_Chng\_Q4\_Q1: similar to the previous but in number of transactions.

1. Is this dataset authenticated or somebody mad it for visualization purpose?

Answer: I don’t know the authenticity of the dataset. But personally I feel the most of the dataset available on Kaggle are good for learning purpose.