Problem Statement

<u>Credit Card Lead Prediction</u>

Happy Customer Bank is a mid-sized private bank that deals in all kinds of banking products, like Savings accounts, Current accounts, investment products, credit products, among other offerings.

The bank also cross-sells products to its existing customers and to do so they use different kinds of communication like tele-calling, e-mails, recommendations on net banking, mobile banking, etc.

In this case, the Happy Customer Bank wants to cross sell its credit cards to its existing customers. The bank has identified a set of customers that are eligible for taking these credit cards.

Now, the bank is looking for your help in identifying customers that could show higher intent towards a recommended credit card, given:

- Customer details (gender, age, region etc.)
- Details of his/her relationship with the bank (Channel_Code,Vintage, 'Avg_Asset_Value etc.)

## ★ Brief on the Approach, used to solve the problem:

1. Imported the data (from [AV problem](#)) into data frames and performed EDA.
2. Performed the data preprocessing the converting to numeric.
3. Applied ANN classification that worked well.

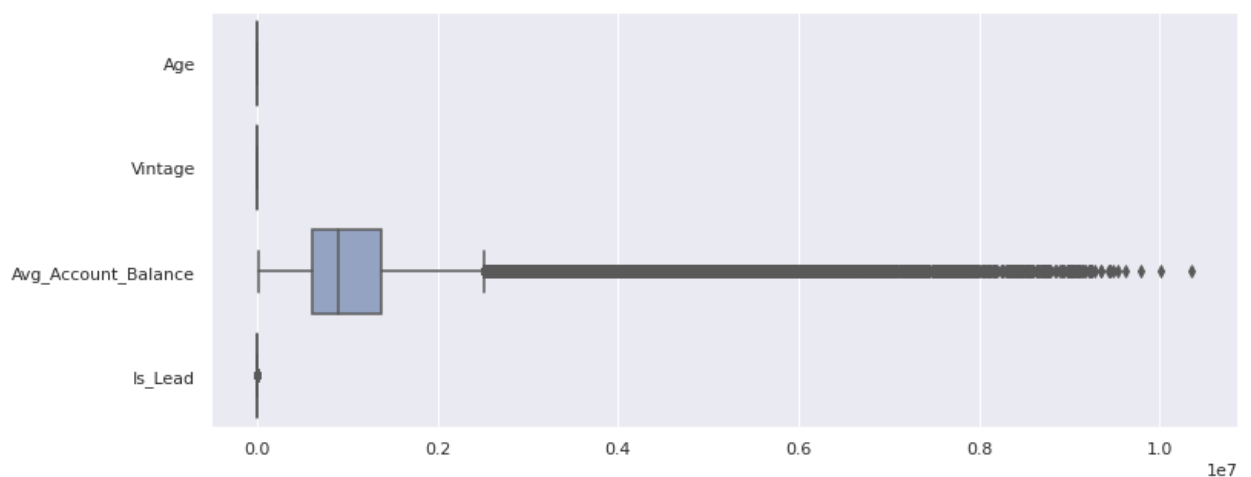How I have approached all the points are covered in next steps.

The below EDA represents the count of people from various occupations.
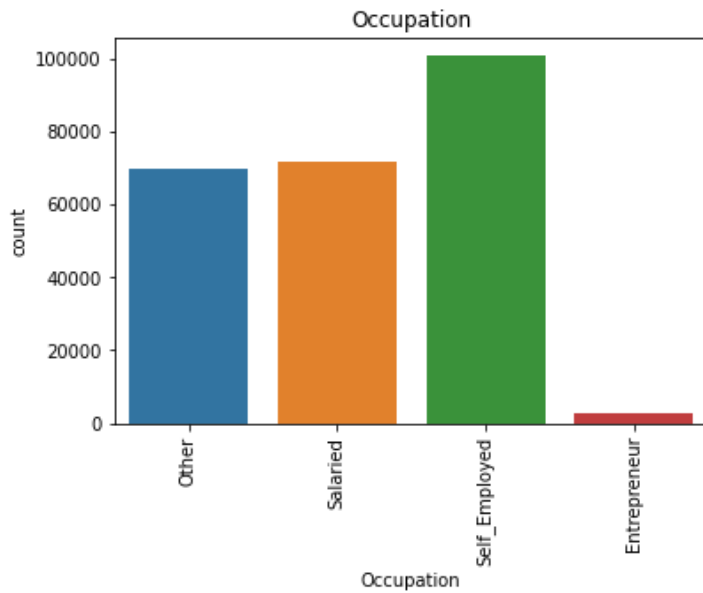
The below EDA represents, How many people are there with respective ages. I can see age 26 to 31 have more people with the bank.
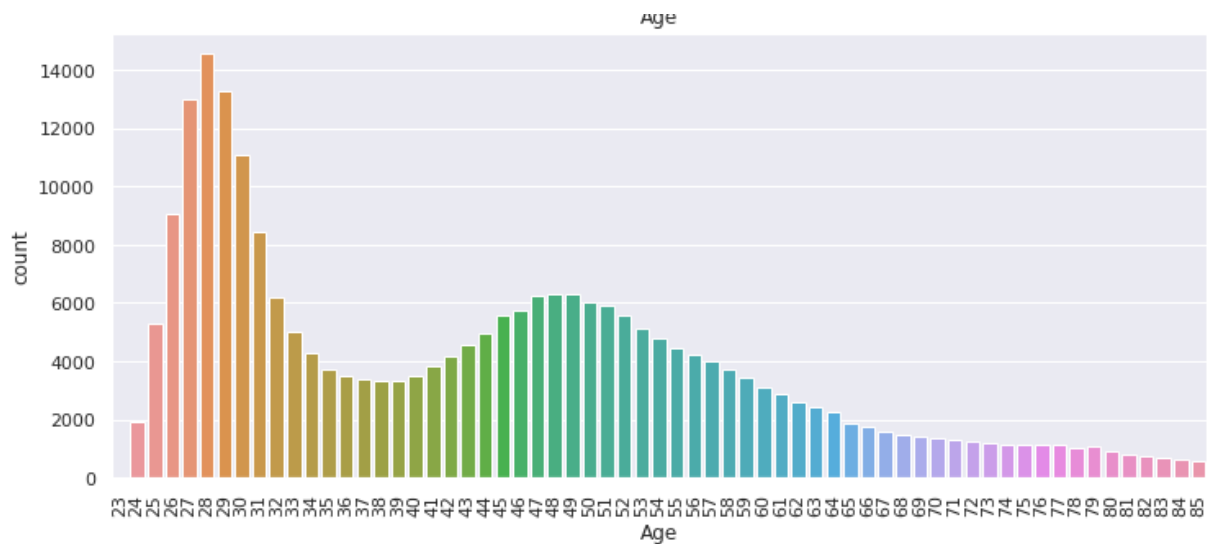
The below EDA represents the outliers that I have removed in the data cleaning and preprocessing steps.

★ Data-preprocessing / Feature engineering:

For the train and test data I followed the below data preprocessing steps:

1. Converted Credit_Product,Gender, Is_Active to binary.
2. Encoded Occupation, Channel_code to numeric.
3. Dropped ID from train as it has no correlation with our target, checked with correlation matrix.
4. Encode labels in column 'Region_Code'.
5. Encode labels in column 'Avg_Account_Balance'.
6. For ANN, converted the features to Scalar and Standardized.

★ How did I discover the approaches?

Started with simple logistic regression and then calculated the AUC-ROC and I got 0.69. That was low so again I went back to the data and did some label encoding and then tried with RFE for better feature selection and was getting the AUC-ROC the same. So then I thought of XGboost and Random forest to handle the problem and while exploring I got to know about Catboost.

★ Final model:How did I reach it?

I have applied Catboost with RFE and got AUC of 0.75. That was a good improvement but not that great. Then I thought of applying some deep learning concepts to make predictions better, I tried with ANN and yes I got the AUC of 0.87 which is great.

In ANN I improved the model by:

- By Changing the the batchsize
- By changing the unit size
- By changing the epochs
- By changing the hidden layers
- By applying correct activation functions