**Overview**

Your team represents the advanced analytics practice that services multiple business lines at a bank XGD.

You have the sample data that showcases the demographics of the existing customers,

relationships of these customers with the bank and relationship scoring, by the bank, of these customers.

You are given a sample of 9,231 customers who have at least one savings/checking account

relationship with the bank. There are total 32 variables in this sample

------------------------------------------------------------------------------------

Stage 1 Challenge

1.The Problem Statement

--------------------------------

We have divided the sample into a training (6,924) and a test (2,307) dataset.

Using the variables at hand predict the application score and the behavioral score of the customer.

2.Evaluation Criterion

--------------------------------

Use the 'training'(file - Stage 1 Train) sample to train the model , using this model compute the predicted application score and behavioral

score for the test dataset(file - Stage 1 Test).

The RMSE for the "test" sample predictions will be used in creating the leaderboard .

The accuracy of the prediction will be judged by using RMSE (root mean square error) metric.

The submission should contain only 3 columns viz.; Identifier, P\_Application\_Score, P\_Behavioural\_Score.

Where P\_Application\_Score is the predicted application score and P\_Behavioural\_Score is the

predicted behavioral score

The final result should be uploaded in a csv format **(submission format provided**)

3.Leaderboard Creation

--------------------------------

For each model prediction, percentile rank will be computed (with lowest RMSE having highest rank),

an average of the two (for both models) percentile rank will give the standing of the team on the leaderboard.

Shortlisted teams will be required to submit their final codes for

sanity checking and ensuring proper methodology was deployed in solving the challenge.