# Exercise 6.3: Original Analysis Case Study[¶](#Exercise-6.3:-Original-Analysis-Case-St)

In this case study, as part of phase I, we will perform exploratory data analysis by graphing the features in the dataset.

The dataset is composed of 10,000 customer's record at a bank. The dataset has a total of 14 features 13 of which can be considered as independent variables and 1 as the dependent variable. The goal is to build a model that can predict whether a customer is likely to stay or exit the bank. The model will predict the dependent variable 'Exited' using the approrpiate set of independent variables 'CreditScore','Geography','Gender','Age','Tenure','Balance','NumberOfProducts','HasCrCard', and 'IsActiveMember'.

We will perform model selection and model validation exercises and use the model the make the desired prediction. The accuracy and percision of the model will be analyzed in the next phases of the study. Code below shows the content of the dataset as read from the file BankCustomers.xlsx

In [3]:

# Load Libraries

import pandas as pd

import matplotlib.pyplot as plt

import numpy as np

import xlrd

In [4]:

#Step 1: Load data into a dataframe

DataFile = "Data/BankCustomers.xlsx"

data = pd.read\_excel(DataFile)

print(data.head(5))

RowNumber CustomerId Surname CreditScore Geography Gender Age \

0 1 15634602 Hargrave 619 France Female 42

1 2 15647311 Hill 608 Spain Female 41

2 3 15619304 Onio 502 France Female 42

3 4 15701354 Boni 699 France Female 39

4 5 15737888 Mitchell 850 Spain Female 43

Tenure Balance NumOfProducts HasCrCard IsActiveMember \

0 2 0.00 1 1 1

1 1 83807.86 1 0 1

2 8 159660.80 3 1 0

3 1 0.00 2 0 0

4 2 125510.82 1 1 1

EstimatedSalary Exited

0 101348.88 1

1 112542.58 0

2 113931.57 1

3 93826.63 0

4 79084.10 0

## Analysis and steps for next phase[¶](#Analysis-and-steps-for-next-phase)

The graphs have shown us the number of customers who have stayed/left by various attributes such as gender, age, balance and other attributes. The parallel graph does not show much information. The graphs generally show that there are moe stays than exists.

We will create a classification graph that will show exits/stays based on classification of credit score, age, salary and balance in the next phase.We will then use appropriate techniques in selection, evaluation and prediction