Data set of a bank

10000 customers – 10000 ROWS

The bank has been seeing unusual churn rates (churn is when people leave the company) and they've seen customers leaving at unusually high rates and they want to understand what the problem is

This bank has millions of customers it operates this fictional bank operates in Europe in three countries France Spain and Germany and they have lots and lots of customers so what they did is they took this sample of 10000 customers and they measured six months ago everything they knew about them they had customer id their surname credit score their geography their gender their age their Tenure (how long they've been with the bank) the balance of the customers at that point in time the number of products they had at that point in time.

So number approaches things like how many products do they have to have a savings account have a credit card.

do they have a loan?

Did the customer have a credit card or not. Is the customer an active member? - measured differently by different organizations. Whether or not the customer logged into their online banking in the past month. whether they did a transaction in the past two months or some other measurement like that and estimated salary (so he bank doesn't know the salary of the customers but based on the other things they know they could estimate a salary for that customer and they also gave you this information.)

six months ago they measured all of these things

So for these 10000 are randomly selected customers are we going to just watch them so we just going to wait six months and six months on the track we're going to check who of those customers left and who those customers state - tells you whether or not the person left the bank within those six months.

your goal is to create a demographic segmentation model to tell the bank which of their customers are at highest risk of leaving.