# Challenge

## Problem Statement

**Part 1 – Joining data sets**

You have been furnished with sales data (CUSTOMER\_SALES.xlsx) and customer accounts data (CUSTOMER\_CAMPAIGN.xlsx), which are pulled from two different data sources/databases. Unfortunately, the sales data does not have a unique identifier to join directly to the customer account data. These different sets of data are to be joined on a regular basis, and the joining process must be automated. It is your goal to produce a single CUSTOMER\_ CAMPAIGN \_SALES dataset.

Note, there are no duplicates in either datasets; Therefore, all 1000 customers in the CUSTOMER\_ CAMPAIGN.xlsx must match directly to the 1000 customers in CUSTOMER\_SALES.xlsx. Finally, your strategy must be robust to join any other two datasets in the same format so automation is key.

**Part 2 – Data analyses**

After joining the data sets, you will notice that there is a column called ‘Sales’ which contains either a 0 (customer did not make a sale) and 1 (customer did make a sale), based on a specific marketing campaign that was run. Considering this, can you tell us something interesting about the data?!

Report

Please prepare a report in any format you wish within 7 days of receiving this challenge to be presented at your next interview via screen sharing.