# Analytics mindset

## ETL

### Case 3 – Advanced ETL text extraction and unique identifiers – Alteryx

In older computer systems, multiple values were often stored in a single cell to save space. This practice is sometimes still followed today. For example, an employee identification number may tell you the employee number, plant number and business function. That is, 143-01-Acc could mean employee 143 from plant 01, who works in accounting.

If you already performed Case 2, this case is the same as Case 2, except the data is "messier." For this case, you need to use the Excel file titled **Analytics_mindset_case_studies_Case3_Alteryx.xlsx,** which contains 597 rows of employee data. In the tab labeled Case 3 data, you will find three columns: EmployeeCode, FirstName and LastName. The EmployeeCode is the combination of four different fields: Location, EmpID, PlantID and PayPeriod. Each of these fields is defined as follows (note that these definitions are not the same as in Case 2):

- Location: The location code shows the location where the employee works. The company operates in eight different countries. Employees in different countries do not have to use the standard three-digit codes; thus, the codes for each country are as follows: Argentina (ARG), Australia (AUS), Canada (Canada), England (ENG), Germany (GER), Japan (Japan), Mexico (MEX) and the United States of America (US). The country codes are the first digits in the EmployeeCode, reading from left to right.
- EmpID: The company assigns a random employee identification number from 1 to 597. Reading the EmployeeCode from left to right, the EmpID is made up of the first set of numbers immediately after the Location code and *preceding* the dash.
- PlantID: The company has various plants throughout the different countries. Each country numbers its plants starting at one and adds one more number for each additional plant. The PlantID is contained in the EmployeeCode, reading from left to right, immediately *after* the dash.
- PayPeriod: Employees are paid either weekly or monthly. The system records this as a W for weekly and as either an M or Mo for monthly. M and Mo mean the same thing; however, sometimes the employee just records them differently. The PayPeriod is the last letter or letters, reading from left to right.

You have been asked by your manager to extract data using the employee code and also to create a new unique identifier that will provide the plant number by location.



SCORE No. 06753-191US_21

**Required**

- Complete the ETL overview case, which covers the fundamental considerations in the ETL process.
- Use Alteryx to:
  - Create separate columns for each of the four fields described above.
  - Add a concatenated field to create the unique identifier combining Location and PlantID so the output looks like the following: USA-12.
- Produce an output file in Excel. Submit your Alteryx workflow as a packaged workflow (.yxzp file type [Options > Export Workflow >]) and include your name in the file name.



SCORE No. 06753-191US_21