PART 2: Checks for quality assurance (Includes Assessment)

Sometimes it is known in advance which types of data integrity issues to expect, but other times it is during the Exploratory Data Analysis (EDA) process that these issues are identified. After extracting data it is important to include checks for quality assurance even on the first pass through the AI workflow. Here you will combine the data into a single structure and provide a couple checks for quality assurance.

QUESTION 3:

Implement checks for quality assurance

1. Remove any repeat customers based on *customer_id*
2. Remove stream data that do not have an associated *stream_id*
3. Check for missing values

An answer key has been provided in the form of an online Jupyter Notebook for your to review upon completion of this exercise.

QUESTION 4:

Combine the data into a single data structure

For this example, the two most convenient structures for this task are Pandas dataframes and NumPy arrays. At a minimum ensure that your structure accommodates the following.

1. A column for *customer_id*
2. A column for *country*
3. A column for *age* that is created from *DOB*
4. A column *customer_name* that is created from *first_name* and *last_name*
5. A column to indicate churn called *is_subscriber*
6. A column that indicates *subscriber_type* that comes from *invoice_item*
7. A column to indicate the total *num_streams*

Resources

- [Python's datetime library](#)
- [NumPy's datetime data type](#)

An answer key has been provided in the form of an online Jupyter Notebook for your to review upon completion of this exercise.