



This assignment requires you to perform the Extract, Transform, Load process on an Excel workbook. This assignment is to be done using the Excel workbook "ETL-Exercise-1A.xlsx".

You will be taking two sets of data and combining it into a single source that can be analyzed. Right now, the two sets of data are similar, but have too many differences which prevent records from being directly compared across sets. You will need to create and implement rules that resolve the differences in the data.

So you will be:

- Extracting the data from the original worksheet.
- Transforming the data using an Excel formula.
- Loading the data into a new worksheet that contains a single set of combined data.

The xls file is a collection of orders, organized by line item. Multiple rows can be associated with a single order because you can order multiple items in a single order.

There are four worksheets in this workbook:

- Source 1
The first data source (29 records)
- Source 2
The second data source (30 records)
- Full Set
An empty sheet that will contain a consolidated set of all 59 records
- Lookups
Where we'll store the tables used to look up values. You'll use this throughout the assignment.

Inconsistencies

To understand the inconsistencies between the data, open the workbook and look at the Source 1 and Source 2 worksheets. You'll notice that the data doesn't quite match up. For example, order is represented in Source 1 as a five-digit number (i.e., 10001) but in source 2 as an "A" followed by a five-digit number (i.e., A10001). Left as is, an analysis (such as a Pivot Table) would see this as two different orders. The data must be reconciled so that the format is the same. You should not make any changes to the Source 1 or Source 2 worksheets to complete this assignment!

Guidelines

You must submit the final version of your "ETL-Exercise-1A.xlsx" worksheet as "999-StudentName-ETL-Exercise-1A.xlsx".

Your worksheet should be emailed, as an attachment.

The email must be sent by the end of day the assignment is due.



Part 1: ETL With The Orderid Field

The business rule states that we leave OrderID in Source 1 as it is and remove the "A" from Source 2.

Part 2: ETL With The Customer State Field

Source 1 has the values in "Customer State" field as two letter codes.

Source 2 has the values in "Customer State" field full form.

The business rule require that:

- CustomerState field has full form
- StateCode field as two letter code

Part 3: Finish The Worksheet

Perform the ETL process on the rest of these fields in the Full Set worksheet:

- Customer Full Name*
- Customer City
- Customer City
- State Code
- Customer Status*
- Order Date
- Product ID
- Product
- Unit Price
- Quantity
- Discount
- Full Price
- Extended Price
- Total Discount*

Here are a summary of the remaining inconsistencies:

Source 1 Field	Source 2 Field	In the Full Set tab
Customer Full Name as one field	Customer First Name and Customer Last Name as separate fields	Customer Full Name should appear as FirstName LastName with a single space in-between for all customers
Customer Status as "Silver," "Gold," and "Platinum." Platinum is best.	Customer Status as 1, 2, and 3. 3 is the best.	Customer Status should appear as Silver, Gold, or Platinum for all customers
Total Discount included	Total Discount not computed	Total Discount should be computed for all customers (use a formula to subtract the extended price from the full price)

Part 4: Credit Line Field

Add the column for CreditLine to the "Full Set" worksheet. A minimum credit line of \$2,000 has been established, so that even if the customer has a credit line of \$0 it is changed to \$2,000.