**Assignment – ETL Using Tableau Prep**

**Data Files:** *employee\_info.xlsx*

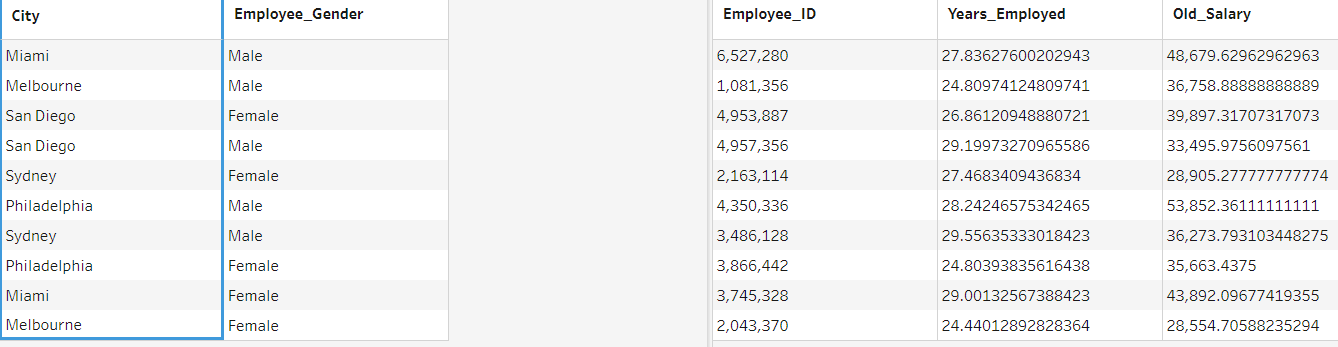
*employee\_payroll.csv*

1. **Connecting to Data**
   1. Add the **Addresses** spreadsheet as input from the **employee\_info.xlsx** workbook.
   2. Change the type of the **Postal\_Code** column to **String**.
   3. Rename the Step as **Employee Addresses**.
   4. Add the data set **employee\_payroll.csv** to the workflow.
   5. Do not include the columns for **Marital\_Status** or **Dependents**.
   6. Rename the Step as **Employee Payroll**.
2. **Examining, Cleaning, and Transforming Data**
   1. Add a Clean step to **Employee Payroll**.
   2. Create a new column named **Bonus** that represents 1.5% of the **Salary** column.
   3. Include only active employees in the output table. In other words, only include those that have a missing value for **Employee\_Term\_Date**.
   4. Create a new column named **New\_Salary**, which is the current salary plus an additional   
      2% raise.
   5. Rename the **Salary** field to **Old\_Salary**.
   6. Create a new column named **Years\_Employed** that calculates the number of years that each employee worked at Orion Star, based on the hire date and today's date.

Hint: The TODAY() function can be used to provide the current date. You also need to transform the **Employee\_Hire\_Date** to the Date type using the function of DATE().

* 1. Change the values for the **Employee\_Gender** column (i.e., change “F” to “Female” and “M” to “Male”).
  2. Add a Clean step to **Employee Address**.
  3. Examining the **City** column, we found that **Miami-Dade** is a county name, not a city name. Change the value **Miami-Dade** to **Miami**.
  4. In the data, you may notice that the **State** column values have inconsistent spellings (e.g., “PA” and “Pa”). Make all values in the **State** column to Uppercase.
  5. In the **Country** column, spell out the code “AU” and “US” to “Australia” and “United States”, respectively. This will help create the data role.
  6. Assign data role to the columns of City, State, and Country to their corresponding roles.

1. **Joining Data from Two Tables**
   1. Use the Join step to join **Employee Payroll** and **Employee Addresses** on the **Employee\_ID** column. Rename the step as **Payroll Location Join Query**.
   2. Only keep useful columns by excluding the following data columns: **Employee\_ID-1**, **Street\_ID**, **Employee\_Hire\_Date**, **Employee\_Term\_Date**.
2. **Aggregating Data**
   1. Add a Step of **Aggregate** after the Payroll Location Join Query. Rename the aggregation step as **Employee Information by City**.
   2. Show the total number of employees, the average years employed, the average salary (use **Old\_Salary** column) of male and female at each city. Your result will look like:



1. **Output**
   1. Add an Output step after the joining query.
   2. Rename it as **Output\_Employee\_Info** and save it to a Tableau Data Extract file.

**After you complete the assignment, save the Tableau Prep flow. Turn this file to the Canvas Folder.**