# Assignment - Data Transform Activity (20%)

Due: Module 4

## Purpose:

The purpose of this assignment is to transform data into an appropriate format for analysis and then extract that data for further problem solving.

Assignment Objective:  
For the business problem you selected, you will extract data for further analysis using a series of transformations, including: select, mutate, filter, and arrange.

## Instructions:

This activity relates to the human resources analysis data.

**Click on the link below to download your data:**

* [**Human Resources analysis - Data**](https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset/data)

In this assignment, you will perform simple transformations on your data (you might need to reload the data in HDInsight and Hive).  Do the following steps:

1. a) Create a new table for your analysis called “employee\_sales”.
2. Load the table “employee” into this table.

c) Select these columns: Attrition, Department, JobSatisfaction & MonthlyIncome.

2) Round the data found in the “MonthlyIncome” column to the nearest $1000. (HINT: the SQL function to round a number is ROUND(obs, -3))

3) Filter the data to only look at those items in the “Sales Department”.

4) Order the data by “JobSatisfaction” from highest to lowest. (HINT: Use the DESC query)

**1)**

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generatedTable

Description automatically generated

2)

Graphical user interface, text, application, email

Description automatically generated

Table

Description automatically generated

3)

Graphical user interface, text, application, email

Description automatically generatedGraphical user interface, text, application, email

Description automatically generated

Table

Description automatically generated

4)

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Table

Description automatically generated