# Assignment - Data Aggregation Activity (10%)

Due: Module 5

## Purpose:

The purpose of this assignment is to aggregate data in a format for analysis and extract those data for further problem solving.

Assignment Objective:  
For the business problem you selected, perform the required steps to aggregate data for visualization using a series of summarizations, including: Count and Average.

## Instructions:

**This assignment uses the data related to the “Human Resources Analysis” business problem:**

**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 aggregate the data for extracting insights (you might need to reload the data in HDInsight and Hive).  Do the following steps:

1) Create 2 tables. The first table shows how many salespeople have attrition. Store this information in a column called “attrition”. The second table counts how many salespeople have no attrition. Store this information in a column called “Non Attrition”.

2) Create 3 statistics tables (average, min, max) showing the monthly income for: all salespeople, those with attrition and those without attrition. For each table, label the columns “average\_monthly\_income”, “min\_monthly\_income”, and “max\_monthly\_income”.

3) Create 2 tables containing the monthly income and count of salespeople making that monthly income for: salespeople with attrition and salespeople without attrition. Label the columns as: “monthlyincome” and “count”

## This assignment relates to the following Course Learning Requirements:

### CLR 2: Illustrate how current tools are used to address these problems, including the strength and weaknesses of each tool

### CLR 4: Extract and load data to support problem solving and decision-making

## Grading for This Assignment

Please take a screen shot of each step, showing the result. Also provide the query you used.

|  |  |
| --- | --- |
| **Steps required** | **Grade Earned** |
| 1) Create two tables, one showing the number of salespeople with attrition and the other showing the number of salespeople with non-attrition | 50% |
| 2) Create three statistics tables showing the monthly income for all salespeople, those with attrition and those without attrition. | 80% |
| 3) Create two tables (one for salespeople with attrition and one for salespeople without attrition) showing the monthly income and the count of salespeople making that amount of income | 100% |

1)

Graphical user interface, text, application, email

Description automatically generatedGraphical user interface, text, application, email

Description automatically generated

2)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 generated

3)

Graphical user interface, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Table

Description automatically generated