**DIY Project – 2: Employee Attrition and Satisfaction Analysis**

**Time: 15 hours**

The HR Leads of a software product-based company wants to analyze the root cause of attrition of employees and perform satisfaction analysis based on Employee data collected from their portal. Assume that you have been hired as a Data Warehousing and Mining engineer within the team to build the model of warehouse and mine and analyze the employee data to perform root cause analysis of the situation and identify possibilities to utilize the work-force to its full potential.

Download the sample dataset “HR Employee Attrition.csv” from your iON Login into your virtual hands-on environment to perform the necessary task.

The dataset “HR Employee Attrition.csv” contains 36 columns and 1470 records.

Along with other columns certain fields/columns have numeric values related to the rating/meaning of certain criteria which are listed below:

**Education**

|  |  |
| --- | --- |
| **Qualification ID** | **Qualification Level** |
| 1 | Below College |
| 2 | College |
| 3 | Bachelor |
| 4 | Master |
| 5 | Doctor |

**EnvironmentSatisfaction**

|  |  |
| --- | --- |
| **Work Environment Satisfaction ID** | **Work Environment Satisfaction Level** |
| 1 | Low |
| 2 | Medium |
| 3 | High |
| 4 | Very High |

**JobInvolvement**

|  |  |
| --- | --- |
| **Job Involvement Level ID** | **Job Involvement Level** |
| 1 | Low |
| 2 | Medium |
| 3 | High |
| 4 | Very High |

**JobSatisfaction**

|  |  |
| --- | --- |
| **Job Satisfaction ID** | **Job Satisfaction Level** |
| 1 | Low |
| 2 | Medium |
| 3 | High |
| 4 | Very High |

**PerformanceRating**

|  |  |
| --- | --- |
| **Performance Rating** | **Performance Level** |
| 1 | Low |
| 2 | Good |
| 3 | Excellent |
| 4 | Outstanding |

**RelationshipSatisfaction**

|  |  |
| --- | --- |
| **Relationship Satisfaction ID** | **Relationship Satisfaction Level** |
| 1 | Low |
| 2 | Medium |
| 3 | High |
| 4 | Very High |

**WorkLifeBalance**

|  |  |
| --- | --- |
| **Work-Life Balance Level** | **Work-Life Balance Measure** |
| 1 | Bad |
| 2 | Good |
| 3 | Better |
| 4 | Best |

**Assigned Tasks:**

You need to perform the following activities:

1. Load the csv file and perform necessary EDA and Data Preprocessing activities.
2. Add necessary column(s) to the dataset to implement the time-invariant, non-volatile features of data warehouse.
3. Develop models using Classification and Prediction techniques to identify the best fitted cause of Employee dissatisfaction.
4. Generate management reports to provide the following and generate graphs/charts accordingly.

i) Sum of monthly income, average job satisfaction as per Job Role and Education Field.

ii) Gender wise, job role wise average JobSatisfaction and Environment satisfaction.

iii) EducationField wise, JobRole wise average Hourly Rate, average monthlyincome, average JobSatisfaction.

iv) Department wise, gender wise average job satisfaction.

v) Find average breakdown of average DistanceFromHome related to gender, department and JobRole.

vi) Find average monthly income by education and attrition.

**N.B:**

1. The entire work should be stored in a single folder within the Virtual Environment. Name the folder as DIY2\_YourName.

2. To explain certain conditions (if required) create a word document and explain with reference to QNo. Screenshots of executable code and its outputs are mandatory.

3. Create .zip/7z file of the folder created which should contain the entire solution. Upload the .zip/7z file to submit. To create a zip file, the following command may be used:

zip -r DIY2\_YourName.zip DIY2\_YourName