

AIT 580 – Assignment 8 – R Data Analysis

Due Sunday October 07, 2018

Please submit on Blackboard

Data Analysis using R: Employee Attrition Dataset and Acme Dataset

Location: Course Content - Datasets - EmployeeAttrition.csv and Acme.csv

Tool: You need to use RStudio for this assignment inside Virtual Box VM

1. Open RStudio by clicking on blue circle R icon on the left side launch bar. This will open RStudio main screen.
2. On the top menu, click on File > Open File and locate the Assignment-5-R-Data-Analytics-Source-File.R provided with the assignment.
3. Also make sure to add the complete path of EmployeeAttrition.csv file inside source file. For example, if you store EmployeeAttrition.csv in Documents folder then your complete path should be `~/Documents/EmployeeAttrition.csv`
4. Press Source button (green arrow icon) inside RStudio window on top. This will compile and run your code.
5. If you want to compile and run single line/statement in your code, press Run button (green arrow icon) inside RStudio window on top.
6. If everything goes well, you will see output in blue or black color. If your code has any error, it will be written in red color.
7. Write your remaining code in the same source file to answer all the questions below.

Questions:

Please answer the following questions

NOTES: Submit one **R source code file** of your code and **output document (word or pdf)** containing all the outputs. Make sure to **print** all outputs to questions in your R source code file so that when class TA runs your code in RStudio, it will show all the output for each question.

Please change your source code filename with your FullName-GNumber and submit to blackboard.

Part 1

1. Use EmployeeAttrition.csv and write R code to find the following:
 - a. Find the number of rows and columns in the dataset **(5 points)**
 - b. Find the maximum Age in the dataset **(5 points)**
 - c. Find the minimum DailyRate in the dataset **(5 points)**
 - d. Find the average/mean MonthlyIncome in the dataset **(5 points)**
 - e. How many employees rated WorkLifeBalance as 1 **(5 points)**

- f. What percent of total employees have TotalWorkingYears less than equal to 5? Also calculate the percentage for TotalWorkingYears greater than 5 **(5 points)**
- g. Print EmployeeNumber, Department and MaritalStatus for those employees whose Attrition is Yes and RelationshipSatisfaction is 1 and YearsSinceLastPromotion is greater than 3 **(10 points)**
- h. Find the mean, median, mode, standard deviation and frequency distribution of EnvironmentSatisfaction for males and females separately. (Hint: For frequency distribution use table() function) **(10 points)**

Part 2

Acme corporation is accused of gender bias in setting starting salaries for newly hired workers. The accompanying synthetic dataset lists recent salary data of new hires, showing college degree (BS, MS, PhD), gender (M, F), years of previous experience and starting salary (thousands of dollars).

1. Use Acme.csv and write R code to find the following:
 - a. Identify data types for each attribute in the dataset **(5 points)**
 - b. Produce a summary statistics for each attribute in the dataset **(5 points)**
 - c. Produce visualizations for each attribute (**Hint:** use hist() function) **(5 points)**
2. Display the relationship between
 - a. Years of Experience and Starting Salary for all employees **(5 points)**
 - b. Years of Experience and Starting Salary for each gender **(5 points)**
 - c. Years of Experience and Starting Salary for each degree **(5 points)**

(Hint: use Scatter Plots)
3. Find the correlation between Starting Salary and Years of Experience? **(5 points)**
 - a. Is the correlation different for each gender? **(5 points)**
 - b. Is the correlation different for each degree? **(5 points)**
4. What can you conclude about Acme with respect to gender bias after your overall analysis? **(5 points)**