

PES University, Bengaluru
UE17CS322- Data Analytics
Session: Aug-Dec 2019
Week 2 - Assignment-1

Date of Submission: 29/8/2019

Max. Marks: 15

***Note:** It's important to work on your approach than the final answer. Since solutions will vary from individual to individual, right justification for technique used will be vital.*

TOPIC : EXPLORATORY DATA ANALYSIS (STOCKTAKING OF THE DATA) USING R

About the Dataset:

This is the [HR Analytics Case Study dataset](#) from Kaggle.

Some information about the columns:

Education 1 'Below College' 2 'College' 3 'Bachelor' 4 'Master' 5 'Doctor'

EnvironmentSatisfaction 1 'Low' 2 'Medium' 3 'High' 4 'Very High'

JobInvolvement 1 'Low' 2 'Medium' 3 'High' 4 'Very High'

JobSatisfaction 1 'Low' 2 'Medium' 3 'High' 4 'Very High'

PerformanceRating 1 'Low' 2 'Good' 3 'Excellent' 4 'Outstanding'

RelationshipSatisfaction 1 'Low' 2 'Medium' 3 'High' 4 'Very High'

WorkLifeBalance 1 'Bad' 2 'Good' 3 'Better' 4 'Best'

For the purpose of this assignment : **general_data.csv**, **employee_survey_data.csv** and **manager_survey_data.csv** files are sufficient, you needn't use the other files.

***Ignore the rows with NA**

(Hint : use the merge() function after removing the NA tuples to perform natural join on all the three files)

Question-1

(6 marks)

a) Starting from age 20, with an interval of ten, group the ages (20-29, 30-39 and so on) and plot a histogram. What is the nature of the graph?

Pick the age interval with the maximum average monthly income. Use this interval to answer the following:

(b) Find the employee(s) (employeeID, MonthlyIncome, TotalWorkingYears) based on the 5 highest monthly incomes.

(c) Plot a bar graph of the average income of each job role.

Question-2**(1 mark)**

How many females are there in each group (single, married and divorced) for each level of environment satisfaction. Come up with a suitable visualization to depict the same.

Question-3**(2 marks)**

Which department has the least attrition in terms of percentage of employees in that department?

Question-4**(4 marks)**

- a) Create a box plot separately for the distribution of monthly incomes of the various education fields. Compare all the fields and print the summary statistics (mean, median, mode, quartiles, IQR).
- b) Comment on the gender wage gap.
- c) Find the 95th percentile monthly income for the Medical Education Field.

Question-5**(2 marks)**

Explain skewness and kurtosis in terms of the tails of the distribution. You may take an example for better explanation.