1.0 COURSEWORK TITLE

Employee Dataset

THE COURSEWORK OVERVIEW

For the assignment, you are asked to explore the application of data analytics techniques to the dataset which is provided. You must study data problems related to the dataset, giving special consideration to the unique properties of the problem domain, and testing one or more techniques on it.

Your analysis should be deep and in details, also it must go further than what has already been covered in this course. You must adopt the data Exploration, Manipulation, Transformation and Visualization concepts to guide you through the solution process. It is very important to explain and justify the techniques that have been chosen.

You also may need to pre-process your data to get it into an appropriate format. The assignment should involve a number of techniques by categorize it into different criteria and a detailed exploration with the commands using in each criteria. Outline the findings, analyze them and justify correctly with an appropriate graph. Also, a supporting document is needed to reflect the graph and code using R programming concepts.

2.0 OBJECTIVES OF THIS COURSEWORK

This assignment will help you to explore and analyse a set of data and reconstruct it into meaningful representations for decision making.

3.0 TYPE

Individual Assignment

4.0 COURSEWORK DESCRIPTION

This dataset contains the data of staffs within an organization that could determine some hidden issue in human resources management. Human resource department manager assigned you to perform analysis with the given dataset to identify hidden problem in the organization and provide meaningful insight for decision making.

Techniques

The techniques used to explore the dataset using various data exploration, manipulation, transformation, and visualization techniques which covered in the course. And as an additional feature must explore the further concepts which can improve the retrieval effects.

The dataset provided for this assignment is related to the employees' job information and attribution. It contains 18 columns and 49654 rows. The dataset includes the personal detail of the staff, job department, position, location, working status, and reason of termination.

5.0 GENERAL REQUIREMENTS

- The R program should compile and be executed without errors.
- Validation should be done for each entry from the users to avoid logical errors.
- No duplication is allowed in dataset.
- You should;
    - Include good programming practice such as comments, variable naming conventions and indentation.
    - Carried out additional research from Internet to comprehend the knowledge and information on the given dataset when examine the data.
- The analysis should be meaningful and effective in providing the information for the decision making.
- Any additional features implemented must improve the retrieval effects.
- In a situation where a student:
    - ***Failed to demo the assignment, overall marks awarded for the assignment will be adjusted to 50% of the overall existing marks if it is more than 50%.***
    - ***Found to be involved plagiarism, the offence and will be dealt in accordance to APU and Staffordshire University regulations on plagiarism.***

6.0 DELIVERABLES:

- The complete **RScript** (source code) and **report** must be submitted to **APU Learning Management System (Moodle)**.
- RScript (Program Code):
  - Name the file under your name and TP number.
  - Start the first two lines in your program by typing your name and TP number. For example:

    # NAME

    #TP123456
  - For each question example, give an id and explain what you want to discover. For example:

    # Question 1: Why staff would leave the company.

    # Analysis 1-1: Find the relationship between job position with attrition…

    # Analysis 1-2: Find the relationship between job age and ….

    # Analysis 1-3: Find the relationship between …
  - For each extra feature example, give an id and provide the explanation.

    # Extra feature 1

    # comments about the extra feature

8.2 *DOCUMENTS: COURSEWORK REPORT*

✓ As part of the assessment, you must submit the project report in printed and softcopy form, which should have the following format:

A) Cover Page:
All reports must be prepared with a *front cover*. A protective transparent plastic sheet can be placed in front of the report to protect the front cover. The front cover should be presented with the following details:

- Module
- Coursework Title
- Intake
- Student name and id
- Date Assigned (the date the report was handed out).
- Date Completed (the date the report is due to be handed in).

B) Contents:

- Introduction and assumptions (if any)
- Data import / Cleaning / pre-processing / transformation
- Each question must start in a separate page and contains:
  - ✓ Analysis Techniques - data exploration / manipulation / visualization
  - ✓ Screenshot of source code with the explanation.
  - ✓ Screenshot of output/plot with the explanation.
  - ✓ Outline the findings based on the results obtained.
- The extra feature explanation must be in a separate page and contains:
  - ✓ Screenshot of source code with the explanation.
  - ✓ Screenshot of output/plot with the explanation.
  - ✓ Explain how adding this extra feature can improve the results.

C) Conclusion

D) References

- The font size used in the report must be 12pt and the font is Times New Roman. Full source code is not allowed to be included in the report. The report must be typed and clearly printed.
- You may source algorithms and information from the Internet or books. Proper referencing of the resources should be evident in the document.
- All references must be made using the **APA (American Psychological Association)** referencing style as shown below:

  The theory was first propounded in 1970 (Larsen, A.E. 1971), but since then has been refuted; M.K. Larsen (1983) is among those most energetic in their opposition…….

```
/**
 * Following source code obtained from (Danang, S.N. 2002)
 */
int noshape=2;
noshape=GetShape();
```

- List of references at the end of your document or source code must be specified in the following format:

Larsen, A.E. 1971, A Guide to the Aquatic Science Literature, McGraw-Hill, London.

Larsen, M.K. 1983, British Medical Journal [Online], Available from http://libinfor.ume.maine.edu/acquatic.htm (Accessed 19 November 1995)
*Danang, S.N., 2002, Finding Similar Images [Online], The Code Project, *Available from http://www.codeproject.com/bitmap/cbir.asp, [Accessed 14th *September 2006]*

- Further information on other types of citation is available in *Petrie, A., 2003, UWE Library Services Study Skills: How to reference [online], England, University of Western England, Available from http://www.uweac.uk/library/resources/general/info_study_skills/harvard2.htm, [Accessed 4th September 2003].*

7.0 ASSIGNMENT ASSESSMENT CRITERIA

The assignment assessment consists of three components: Coding (50%), Documentation (30%) and Presentation (20%). Details of the division for each component are as follows:

<table>
<tr><th colspan="2">Coding (50%)</th><th colspan="2">Documentation (30%)</th></tr>
<tr><th>Criteria</th><th>Marks Allocated</th><th>Criteria</th><th>Marks Allocated</th></tr>
<tr><td>Data Exploration</td><td>10%</td><td rowspan="4">Structure of the report and references
Content:
• Description and justification of the R concepts incorporate.
• Program out screenshots, graphs
Project description, limitation, and conclusion</td><td rowspan="4">30%</td></tr>
<tr><td>Data Manipulation</td><td>10%</td></tr>
<tr><td>Data Transformation</td><td>10%</td></tr>
<tr><td>Data Visualization</td><td>20%</td></tr>
<tr><th colspan="4">Presentation (20%)</th></tr>
<tr><th colspan="3">Criteria</th><th>Marks Allocated</th></tr>
<tr><td colspan="3">• Ability to answer questions addressed by the lecturer pertaining to the work done and presented</td><td>20%</td></tr>
</table>

8.0 DEVELOPMENT TOOLS

The program written for this assignment should be written in R Studio

9.0 ACADEMIC INTEGRITY

- You are expected to maintain the utmost level of academic integrity during the duration of the course.
- Plagiarism is a serious offence and will be dealt with according to APU and Staffordshire University regulations on plagiarism.