

TMI4013 Data Mining

Assignment 2 (15%)

This is a group assignment.

Dataset

Use the Petrol/Shampoo dataset given in the link in eLeap. The dataset consists of real data collected from interviews. Please refer to the questions (pdf) for their dimension descriptions.

***Please do not post/publish/share the data/question without the lecturers' permission on any website/social media/forum. These data are meant for assignment of this course ONLY.**

Assuming that you are the marketing manager of a company, your task is to improve the sale of shampoo and whitener of a brand. You collected data from a survey to find out some important insights.

Tasks

The tasks for this assignment are:

1. Rename attributes so they can be interpreted easily.
2. You can also focus on a very specific investigation, e.g Why rural folks like to buy Nissan?
3. Create various appropriate visualizations using the data from your dataset to answer the following questions: [8%]
 - a. What are the THREE (3) factors that drive a customer to buy shampoo?
 - b. How frequently do long hair users buy shampoo and do they buy whitener at the same frequency?
 - c. What is the brand of whitener that is normally bought together with shampoo? Are they the same brand?
 - d. Is it reasonable to bundle shampoo and whitener together, for sale at a lower price, to increase the sales of both?
 - e. Which working group of respondents care about having beneficial ingredients in the shampoo?
 - f. Does having children affect the choice of shampoo bought?
 - g. Does the highest level of education attained affect the choice of shampoo bought?
 - h. Does ethnicity affect the choice of shampoo bought?
4. Design a dashboard to show the visualizations. Make sure that you have a dashboard title. [2%]
5. Discuss analytically what insights you can gain from your data through your dashboard. [5%]

Note

Maximum 6 pages.

1. Page 1 – 4 should contain the visualizations and the answer to each question.
2. Page 5 should be a full-page screenshot of your dashboard. Make sure that all your visualizations are clear and legible.
3. Page 6 should be an analytical discussion of what insights you can gain from your data through your dashboard.

Software

You may use any tool of your choice to do this assignment. The fancier, the better.

Report Format

- 1.5 spacing
- Times New Roman
- Font size 12
- Justified alignment
- Margin 1" for top, bottom, right and left

Submission

1. Assignment report (6 pages) in PDF.
2. Softcopy of your charts and dashboard (format either in Microsoft Excel, Tableau, PowerBI or any software you used).

The assignment report and softcopy of your charts can be uploaded to eLeap through the link provided. Put the assignment report and softcopy of your charts in one folder and submit a zipped folder containing the items. Name the zipped file as <course_code-group_num>.zip/rar. For example, TMI4013-Group1.zip.

Due date: 20 November 2023, 4pm (Late penalty -1% per day)

Marking Rubric

Criteria	Marks	7 – 8	5 – 6	3 – 4	1 – 2	0
Visualizations (8%)	8	Clear, insightful and advanced visualizations with title, axis label, legend, colour (whichever applicable).	Clear visualizations with title, axis label, legend, colour (whichever applicable).	Visualizations missing one of either title, axis labels, legends, colour (whichever applicable).	Poorly constructed visualizations.	None.
Criteria	Marks	2.0	1.5	1.0	0.5	0
Dashboard (2%)	2	Design of dashboard has very high clarity and aesthetics. Visualizations on dashboard are neatly arranged and easily interpreted.	Design of dashboard has high clarity and aesthetics. Visualizations on dashboard are neatly arranged and easily interpreted.	Design of dashboard has average clarity and aesthetics. Visualizations on dashboard are neatly arranged and easily interpreted.	Design of dashboard has low clarity and aesthetics. Visualizations on dashboard can be interpreted.	None.
Criteria	Marks	5	4	3	2	1
Discussion (5%)	5	Comprehensive and analytical discussion on the insights gain from the data through the dashboard.	Clear and analytical discussion on the insights gain from the data through the dashboard.	Clear and straightforward interpretation of visualizations on the dashboard.	Minimal interpretation of visualizations on the dashboard.	Poor interpretation of visualizations on the dashboard.