MSCI562 Coursework, Part 1
Lent Term, 2022-2023
Management Science Department
Lancaster University Management School

Disclaimer

This is the first part of the two-part coursework. The second part will be released at the end of the module.

Problem Description

Customer satisfaction is important in retaining customers, especially in the service industry. An airline in the UK is currently conducting a study regarding its customer satisfaction. The airline aims to understand what the main factors influencing the customer satisfaction are, so that they can improve their service level. The marketing manager will use this information to improve their service quality regarding tickets booking, boarding and onboard experience. The marketing manager also wants to have a predictive model that will determine whether a new customer will be satisfied given their answers and their characteristics.

Each student has a sample of data, which contains a target variable and 24 features. In this study, the features are classified into four groups: (a) personal details, (b) flight details, (c) pre-boarding, (d) onboard. The last two groups were collected from a survey during the flight.

Details of the dataset

Target variable

- Satisfaction: airline satisfaction level ("neutral or dissatisfied" and "satisfied")

Personal details

- Gender: (female; male)
- Typecustomer: (loyal customer; disloyal customer)
- Age: (age of each customer)
- Traveltype: (types of journeys - business travel; personal travel)
- Class: (ticket class - business; economy; eco plus)

Flight details

- Flightdistance: (the flight distance of this journey in miles)
- Departdelay: (minutes of delay when departing)
- Arrivedelay: (minutes of delay when arriving)

Pre-boarding (5 – is totally satisfied, 1 – totally dissatisfied)

- Easeonlinebooking: (satisfaction level of online booking; 0:NA; 1 – 5)
- Timeconvenient: (satisfaction level of departure/ arrival time convenient; 0:NA; 1 – 5)
- Gatelocation: (satisfaction of gate location; 0:NA; 1 – 5)
- Onlineboarding: (satisfaction level of online boarding; 0:NA; 1 – 5)

Onboard (5 – is totally satisfied, 1 – totally dissatisfied)

- Wifi: (satisfaction level of the inflight wi-fi service; 0:NA; 1 – 5)

- Fooddrink: (satisfaction level of food and drink; 0:NA; 1 – 5)
- Seatcomfort: (satisfaction level of seat comfort; 0:NA; 1 – 5)
- Inflightentertainment: (satisfaction level of inflight entertainment; 0:NA; 1 – 5)
- Onboard: (satisfaction level of on-board service; 0:NA; 1 – 5)
- Legroom: (satisfaction level of leg room service; 0:NA; 1 – 5)
- Checkin: (satisfaction level of check-in service; 0:NA; 1 – 5)
- Inflight: (satisfaction level of inflight service; 0:NA; 1 – 5)
- Clean: (satisfaction level of cleanliness; 0:NA; 1 – 5)

Task 1. Exploratory data analysis and visualization

Using appropriate visualization methods and statistical measures covered in the first part of the course, develop general and specific insights from the data which are relevant to the classification problem at hand. Your report should discuss all the variables contained in the dataset, and for each variable your answer should address the questions:

- Which variables appear to be important for the task at hand, and why? Support your claims with appropriate visualization and outputs that show whether and how important each variable is.
- Are different variables related? Which variables convey information similar to that provided in other variable(s)?
- What are the most common characteristics among the passengers? Use appropriate dimensionality reduction techniques to answer this question.

You should also report key findings related to issues of data quality such as incorrect observations, outliers, unexpected findings. What else have you learned from the process of data analysis (data understanding phase) that you think is important and relevant to the final task? You can refer to the Cross-Industry Standard Process for Data Mining.

Guidelines

- You need to submit a report explaining your findings.
- Your report should include the following sections:
    - Introduction, explaining the problems you need to address and the structure of your report,
    - Conclusions, summarising the main findings and recommendations .
    - Executive Summary, explaining the main findings in one paragraph.
- You can include any graphs and outputs that you find relevant to the problem in the report. Make sure that they are properly referenced and interpreted in the text. Remember that it is up to you to interpret what a figure or table shows, not for the marker to infer.
- It is also important to explain how you came to one or the other conclusion. For example, you should explain **why** you think that there is an effect in a model. Simply stating “there is” is not acceptable.
- Format sections properly, use styles (see this to learn how to do that: https://modules.lancaster.ac.uk/mod/url/view.php?id=1987585).
- Do not include any appendices (they will be discarded during the marking). You do not need to attach R code either, except for the cases, when the code helps in explaining what you did and why.

- Attach outputs of models that you are going to interpret – otherwise, how are we going to find out whether your interpretations are correct or not?
- Each student will have a different dataset, so make sure that you use the one assigned to you, not to your colleague. Failing to use the correct dataset might result in the "fail" mark for the coursework.
- Your report must be between 1000 and 2000 words and should not exceed 15 pages.

Assessment criteria

Your work will be assessed on the following:

1. **Your ability to use correctly the tools that we covered in the module**. Your report needs to clearly show that you understand what the visualization and statistical measures you use mean, and why they are relevant to this specific problem. It is not enough to simply present relevant figures and measures; you also need to explain why you chose them.
2. **Your ability to draw correct conclusions from the visualisation and statistical measures you use**. Again including a figure or a number/statistic is not sufficient. You have to inform the reader what it means and why it is relevant.
3. **Your ability to express and justify your key findings succinctly** (rather than reporting every possible figure/table/statistic you created).
4. You will also be assessed on **report quality**, including: formatting, presenting figures and tables, having captions, numbering and proper references in the text.

You must use R in your report, all outputs should be generated using the R functions.