Course Project

Flight Price Prediction

This data set contains 21 variables that are described in the data dictionary document and include several categorical variables, some binary variables, and numerical data. You can approach this data from a variety of perspectives using the techniques you learned in class to answer questions. Use the tool of your choice but make sure you know how to use it correctly. There are 10000 observations, more than enough to provide valid and reliable statistical analysis.

An agency has hired you to analyze this dataset to develop models that predict flight prices. You must conduct **two ANOVA analyses and three Regression analyses** to answer five questions you believe will help the agency provide the best guidance to your agency. As with any project, you will start with EDA to get a sense of your data. For categorical variables, using a pivot table to get counts and proportions is an excellent way to get a better understanding of those variables. Example questions could focus on how to flight pries vary by airline or class type, what is the impact of flight duration and distance on pricing, how do time factors (day of the week, time of day, advance booking days) affect flight prices?

Deliverables

Perform EDA and include information about the data in the report that helps the reader get an understanding of the data set (useful graphics should be included). Develop five research questions. Perform the analysis and write a detailed description of the results and what they mean (how you would use them). The report should be a minimum of 10 pages (excluding the cover page) of content, double-spaced, with 1-inch margins, in Times Roman 12-point font, free from grammatical errors, and appropriately using APA style for citations and reference list. The paper will be submitted for grading via software that checks for plagiarism so do not submit a zip or compressed file. Plagiarism is a violation of the Student Code of Conduct and will be handled per university policy.

Create a slide presentation that would support a very brief presentation (8-10 minutes) of your analysis. Each group can either submit a recording of the presentation using Powerpoint's Record slide Show feature (or another recording system), or must include a transcription of what would be said on each slide.

EDA analysis will be due on January 28, 2024
Hypotheses are due on February 18, 2024
Final Report is due on March 7, 2024
Final Presentation is due on March 7, 2024
Peer Review is due on March 7, 2024