MA30091 2021/2022 - Coursework 3

Set: 10:00 Wednesday 6 April

Due: 18:00 Wednesday 4 May

Size of Groups: 3–4 students per group.

Estimated Time: This assignment should take each group member roughly 8–10 hours.

Submission: Submission is via the unit's page on Moodle only. Submissions up to five working days late (without a DoS-approved extension) will receive a maximum mark of 40% and after that will be 0%.

Background and Resources

Airbnb operates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities. The company operates world-wide. Homeowners can list their houses / apartments on the platform and set a price (as well as other policies which won't be considered in the analysis). After their stay, guests are invited to leave a review, including a score.

In this coursework you will consider the properties listed for Rio de Janeiro, Brazil, on the platform in April 2018. The analysis will focus on the three districts with the highest number of listed properties and the most common types of properties. Properties with less than five reviews have been removed from the data set, as have properties with a daily price of more than $1,000. A description of the available variables can be found in the file "Airbnb Data Description.docx" and the data are provided in the file "AirbnbRio.Rdata". If you prefer ".csv" you can alternatively load the data from "AirbnbRio.csv".

Instructions

This is a group coursework and you should not discuss any elements of the coursework with anyone other than your group members or the lecturers for MA30091.

In this coursework, your group has to produce a report (maximum 5 pages) which addresses the following issues:

1. Which factors / variables significantly influence the rental price?
2. Can we predict whether a listed property will be available for at least 1 night in the next 30 days based on the provided variables? Which other variables may be important to know?

3. Are there any factors/ variables that lead to the property receiving very high review scores?

All arguments should be quantitative and using the techniques from the course. Your analysis should include all aspects of a statistical analysis and the report should be divided into two parts as follows:

Part A: A section aimed at members of the general public, presenting and explaining your results. This section of the report should be no longer than one A4 page and should be readable without specialist statistical knowledge.

Part B: A technical section aimed at statisticians, explaining exactly what you did, why you did it and what the conclusions were, in a manner that would allow the analysis to be repeated. By this, it is meant that the statistical structure of what you have done should be clear, so that it could be repeated. This means that models and results should be presented mathematically and with graphs or tables as necessary. When writing this section, you can assume the reader has read the first section.

The report should include no computer commands and no raw output. The minimum acceptable font size is 11pt and your report should have at least 2cm margins.

In addition to the report, you should submit an appendix of commented R code which could be used to exactly reproduce your results. This will not be marked for credit, but it might be checked to see if it is unclear what you have done in your report. There is no page limit for the appendix, but note that the main report must stand alone, and should not rely on references to parts of the appendix. The report and appendix should be submitted online via Moodle by 18:00, 4th May 2022. The report should be in pdf format, and the appendix may be either an R script or an R markdown file

Questions can be posted on Moodle or the Lino board. Your questions should be formulated such that you don't give away your modelling approach. **Questions that are deemed to break this instruction will not be answered and removed from the Lino board.** Due to annual leave and university closure, questions posted between April 13 and April 23 won't be answered before Monday April 25. Questions posted after 12:00 pm on Tuesday May 3 will also not be answered.

Marking

The coursework will be marked out of 36 and count 36% towards your final grade for MA30091. There is no single correct analysis for this type of project, so you will not be marked on the basis of how close you get to some particular model answer. The marks are partially subdivided, with 25% of the marks allocated to the non-scientific summary (Part A of the report). The remaining 75% are allocated to the combination of statistical approaches and justifications, interpretation of results in context and presentation.

25 - 36 (First) A report that could be presented with little or no revision. Analysis is sound so that conclusions are well-supported statistically. Interpretation is reasonably mature. The

project should demonstrate a clear overview of the work, without getting lost in details, and be free of all but minor statistical errors.

22 - 24 (2.1) A report that could be presented after a round of revision, but without having to re-do much of the actual analysis. Some flaws in the analysis or presentation (or minor flaws in both), but basically sound. A good grasp of the statistics and context, so that interpretation is reasonable.

18 - 21 (2.2) Major re-working required before the report could be presented, but containing some sound statistics demonstrating understanding of statistical modelling and its application. Reasonable presentation and organisation.

15 - 17 (Third) Major flaws in analysis and presentation, but demonstrating some understanding of statistics, and a reason-able attempt to present the results.

0 - 14 (Fail) Flawed analysis demonstrating little or no understanding of statistics, and/or incomprehensible or overly bad organisation/presentation