

Extremely Important: This is an example of a group report, which is to be used to discuss in your groups how you might tackle such an assignment. It is **NOT** the actual group report assignment - that assignment will be released next week.

1 Introduction

You have been hired to explore data related to housing prices across the country. The price of 1000 houses will be given to you, alongside the post-code, number of bedrooms, number of bathrooms, number of floors, and distance from the nearest a) supermarket, b) train station, and c) pub.

2 Assignment

The people who have hired you are interested in finding houses on the market for comparatively little money, so they can buy them and then sell them at a higher price. To do this, they need you to create a statistical model which will estimate a house's price based on the variables given above. They will then be able to estimate prices for houses on the market, and buy houses which are particularly cheap compared to the estimated price.

You are to use the data given to create such a statistical model. You are also to comment on the performance of the model, noting any potential limitations or weaknesses. During the writing of your report, you should consider

- Are there outliers/extreme values in the data, and if so, what should be done with them?
- What statistical values/diagrams could best support any points you wish to make?
- Is your model performing well?
- Is your model plausibly meeting the assumptions needed for such a model?

(**Note:** the above list is not intended to include all considerations you should bear in mind).

3 Report

The report you write should be divided into four sections.

1. A one page executive summary, allowing a **non statistician** to understand the results you have presented. Any recommendations for how the people have hired you should consider house prices should be given here, in bullet point form.

2. Findings (max 3 pages, including figures and tables).

This section should contain descriptions of your main findings and recommendations for the future, as you would present them to the people tasked with comparing estimated house prices with actual house prices. These officials will not be statisticians, so keep statistical jargon to a minimum, and use figures or tables to support your findings and recommendations.

This section's aim is to provide an overview of what you found, to persuade them your recommendations are worth following. You should also summarise any limitations of your approach, and of the data used. Honest and constructive criticism will be valuable if the people who hire you want to build upon your work later.

3. Statistical methodology (max 6 pages, including figures and tables).

This is a section written for other data scientists who may wish to check your work, or build on it themselves. You should describe here the methods you used in statistical language. This description should include any changes you made to the data, any decisions you made on variables to exclude or transform, how you assessed the performance of your model, and how you performed residual diagnostics on your model.

A major goal of this section is to provide other data scientists with enough information that, should they wish to, they could completely reproduce your results.

4. Appendix (max 3 pages, including figures and tables).

Here you should include any annotated R code and any additional figures or results supporting statements in Sections (1) and (2) but not included there. Do **not** put R code in Sections (1)-(2)

Note that the above page counts are strict - you will lose marks if you go over the page count in any section. The page counts are also deliberately quite low. This is to get you to think very carefully about what information does and does not truly **need** to be included in your report. We are imagining you are handing this report to people who will be extremely busy indeed!

You will also need to produce a presentation of your results, as detailed in the presentation assignment.

4 Mark Scheme

The report will be marked out of 20. The marking criteria for the report is given in Table 2, and is adapted from the university's marking criteria.

Mark	Criteria
18-20	The report is exemplary, providing clear evidence of a complete grasp of both the statistical methods employed, and their interpretation in the given context. The report is exceptionally well-designed, concisely offering relevant information and strong commentary at all points. The English used is faultless.
15-17	The report is excellent, providing strong evidence of a complete grasp of the statistical methods employed, along with a thorough understanding of how to interpret them in the given context. The report is very well-designed, frequently offering relevant information and strong commentary. The English used is very strong.
13-14	The report is good, providing evidence of a strong grasp of the statistical methods employed, along with some understanding of how to interpret them in the given context. The report is well-designed, offering relevant information and satisfactory commentary. The English used is good.
11-12	The report is acceptable, and provides evidence of a reasonable grasp of the statistical methods employed, although with limited evidence of an ability to interpret them in context. The report is acceptable in design, though some information is not relevant, or is inappropriately placed. There are some flaws in the English used.
8-10	The report contains insufficient evidence of a reasonable grasp of the statistical methods employed, although there is some evidence present. The context of the data is under-served or presented inaccurately. The report is flawed in design, with limited examples of relevant information. The English used is flawed, and sometimes poor.
6-7	The report is unacceptable, containing little evidence of a reasonable grasp of the statistical methods employed. The context of the data is under-served or presented inaccurately. The report is badly flawed in design, difficult to read and with limited examples of relevant information and numerous errors. The English used is often poor.
3-5	The report is unacceptable, containing little evidence of a reasonable grasp of the statistical methods employed. The context of the data is ignored or presented with major errors. The report is extremely badly flawed in design, extremely difficult to read and with almost no examples of relevant information and numerous serious errors. The English used is very poor.
0-2	The report is completely unacceptable, containing no evidence of any grasp of the statistical methods employed. The context of the data is ignored or presented without any accuracy. The report is exceptionally badly flawed in design, almost impossible to read and with no examples of relevant information, and with numerous serious errors. The English used is essentially unreadable.

Table 1: Marking criteria