Report 1: Sussex Budget Production

This feedback is automatically generated from the mark scheme. It contains a grade for each section, a set of things that you did and/or did not do that were desired, and some written feedback and recommendations. Use both the recommendations and the things that you didn't do to reflect on how you can improve your skills in preparation for your next report. Please also read the general feedback on Canvas as it contains general remarks that might also be helpful.

# Quality of report writing (report): 14 out of 20

* You included all the appropriate headings for a scientific lab report.
* If you have a figure or a table in the report, make sure you are using it in the report text for an explanation or exposition and referencing it (e.g., we see in Figure 1 that ...).
* If you have a figure or a table in the report, make sure you give it a label (e.g., Figure 1) and a caption which sufficiently explains what is in the image.
* As well as the above, the examiner has left the following feedback about the written and structural aspects of your report: *"I would recommend making your headings larger, and not using code snippets in the report"*

# Quality of report content (report): 18 out of 30

* Make sure you give a little background on IMDb to help set the scene of the report, such as what it the acronym means and what its typical function is.
* When you are dealing with a dataset, you should let the reader know what is in the data, such as the features (or a summary of them) and some examples of rows.
* If you remove data from the dataset, make sure you write the number of rows you remove.
* Make sure you read the information about the assignment properly, as you did not provide the summary statistics of the IMDb scores for the top 5 genres. But you did provide the top 10 movies by IMDb score, the top 5 genres by frequency, the distribution of the IMDb scores for the top 5 genres.
* You provided sufficient justification for the choice of a director or an actor, but did not clearly give a choice of who to approach for the new movie.
* As well as the above, the examiner has left the following feedback about the content of your report: *"I wouldn't recommend imputing data using a summary statistic, since this could massively bias the other summary statistics, it is best just to remove the rows (since there is enough data). Otherwise all the content was there, you just need to pick a director and an actor"*

# Hypothesis Test (report): 12 out of 15

* You did not include a clear testable hypothesis statement (IF ... THEN ...).
* You did not include clear justification for using the hypothesis test (e.g., each review is independent).
* You did not include clear assumptions on the data and reality of the statistical model (e.g., there are enough reviews to assume normality).
* You did not include information on the data being analysed (e.g., number of samples).
* You included clear written null and alternative hypotheses.
* You included which statistical hypothesis test you undertook.
* You included what type of tail you used for the hypothesis test.
* You included the significance level you considered to conclude the hypothesis test.
* You included the test statistic obtained from the hypothesis test.
* You included the p-value obtained from the hypothesis test.
* You did not include the confidence interval obtained from the hypothesis test.
* You included a clear conclusive statement saying whether you rejected (or failed to reject) the null hypothesis.
* As well as the above, the examiner has left the following feedback about your hypothesis test: *"The hypothesis test seems to be done correctly, just lacking in some details. I would recommend writing your results rather than putting in a screenshot of the output of code"*

# Quality of exploratory data analysis techniques (notebook): 19 out of 25

* You did use df.info().
* You wrangled the data to obtain the romance and horror genres.
* You did not consider removing movies that had a low number of reviews.
* You did not determine if there were movies that were in both the horror and romance genre.
* You investigated the NaNs in the dataset.
* You did investigate the duplicates in the dataset.
* The figures and tables in the report were generated from the code.

# Quality of the code (notebook): 8 out of 10

* You displayed the versions of the packages you used.
* You wrote and used the function "read\_function()" that I asked for.
* It is good practise to only import the packages and functions that you intend to use.
* It is good practise to omit warnings that come from packages when you send your code to be read.
* Your code ran without any errors when I ran it on my machine.
* As well as the above, the examiner has left the following feedback about your code: *"I suspect you can optimise the population bit of code that takes 10 minutes, if you want to talk about it then do come see me at some point. Very well written code, just make sure the print statements are actually necessary"*

# Score: 71 out of 100

*Note that this score does not reflect late submissions.*