Isabelle Bricaud

IB349

MSc Artificial Intelligence

COMP8481: Solving Problems with Data and Text

Practical Assessment

Number of Words : 1482 (without the names of the figures & appendix)

Introduction

The REF (Research Excellence Framework) is the UK's system for assessing the quality of research in UK higher education institutions. It takes place every 5 to 7 years and play a key role in determining how much research funding each university receives from the Government.

Universities are evaluated based on three criteria: Outputs, Impact, and Environment, which together form an overall quality score. The evaluations are done using stars assigned by criteria: 4 stars being the best and unclassified the worst. When analyzing the ratings given to the University of Kent for the Computing Department, it is identified that its lowest score concerns the Environment criterion (15%). The goal is to compare Kent's School of Computing with other selected universities, divided into two groups: those whose research environment scores are better than Kent's (b_env) and those whose research environment scores are worse than Kent's (w_env).

In further analyzing the ratings assigned across all criteria for the two groups of universities, it is important to note that the universities with better scores have a majority of 4* and 3* ratings (Figure 1). However, the University of Kent is rated 2* for both the Outputs and Environment criteria.

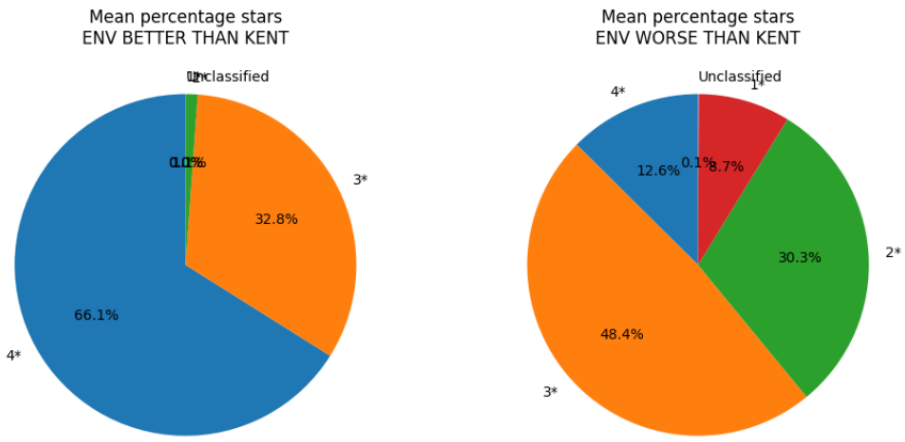
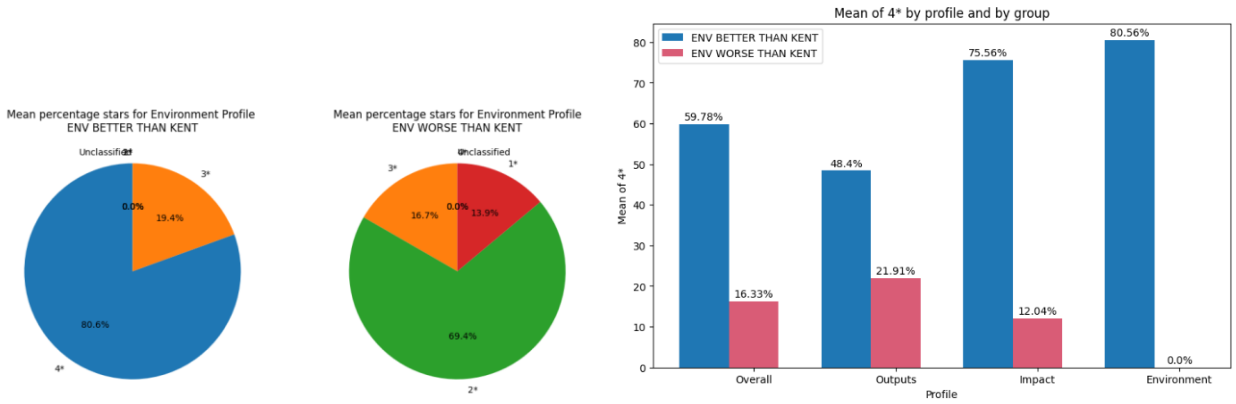


Figure 1 : Mean percentage stars for the two groups

Since the Environment criterion is the least represented for the University of Kent, it may be interesting to examine the distribution of ratings for this criterion. (Figure 2)



Figures 2 & 3 : Mean percentage 4 stars for Environment Profile and by group

Universities that perform well on the environment score have 80% of 4* ratings, while those with a lower score have a majority of 2* ratings and no 4* ratings (Figure 3).

To improve this score, it is important for the University of Kent to identify the associated factors and implement changes based on these observations.

Analysis of the unit's environment statements

Through the environment statements of the universities in both groups (b_env and w_env), it is possible to identify the most common words in both groups. By taking the 10 most common words in these statements, many appear in both groups: 'research', 'staff', 'support', 'university', and 'include'. This suggests that the lexical fields used in the statements are well-oriented. In terms of lexical redundancy, each unique word is repeated an average of just over 7 times in the texts for both groups, indicating the importance of repeating key terms.

Key topics can be identified in the statements. Regarding the b_env universities, they focus more on *the academic support for doctoral degrees, the university environment, and the allocation of resources*. As for the w_env universities, they emphasize *the international dimension and the impacts and collaborations of projects*. The difference in the topics addressed suggests that an area for improvement in Kent's statement could be to place more emphasis on subjects such as funding and academic support.

The texts in the b_env group have a neutral or positive tone (33%), while those in the w_env group do not adopt a positive tone, suggesting that their discourse is less impactful. Regarding the tone to adopt, it is important to reflect a positive sentiment and enthusiasm towards the research environment, though this does not seem to be overly impactful. Furthermore, an analysis of similarity with certain words reveals that the texts in the b_env

corpus use a more structured language, focused on concrete actions and tied to the institution, whereas those in the w_env corpus use words in a more generic and highly descriptive manner without anchoring them in a clear context.

Analysis of the institution's environment statements

In the same way as with the unit's environment statements, it is possible to identify the most common words in the b_env and w_env groups. Among the most common words in both groups are those found also in the unit's environment statements: ‘research’, ‘support’, ‘university’, ‘staff’, and ‘include’. The lexical fields of both groups appear to be the same. The word repetition-to-unique-word ratio is approximately 6, which remains high, further emphasizing the importance of lexical repetition.

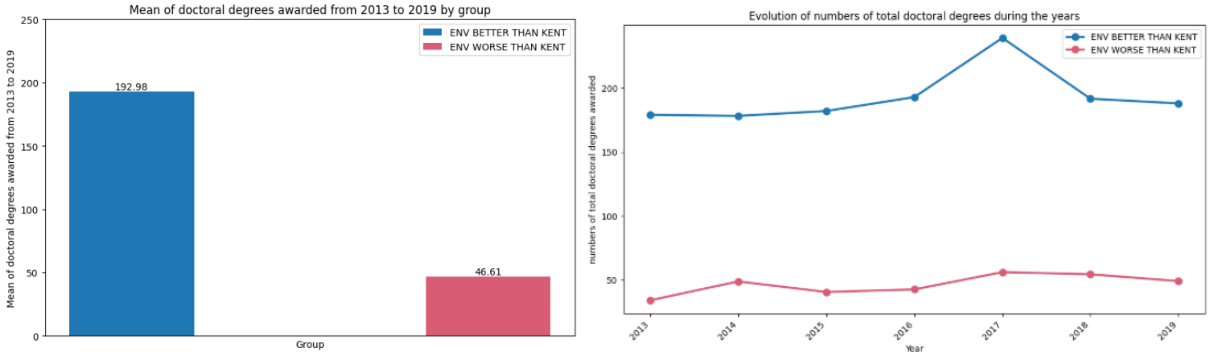
Regarding the themes identified in these statements, the b_env group highlights a *research environment with a focus on staff, such as academic support, investments in infrastructure, and interdisciplinarity*. In contrast, the w_env group emphasizes *career development support, the development of collaborative projects*, but with less emphasis on concrete results and *the development of applied research*. To align more closely with the b_env scores, the themes of the statements should be directed towards allocating budgets to invest in infrastructure, focusing research on interdisciplinarity, and reinforcing the availability of staff to support research. Discussing career development does not seem to have an impact on the REF rating.

The tone used in the texts about the institution's environment seems to be a key factor for having a positive impact on the rating. The majority of the texts in the b_env group adopt an enthusiastic tone, so it is important to express enthusiasm about the development of research within the institution. The similarity analysis characterizes the b_env texts as emphasizing strong involvement of stakeholders and infrastructure to support research, whereas w_env focuses on individual achievements. This suggests that research should be supported both financially and logistically, with available staff, and that rewards should not be individual but oriented towards collaboration. With these recommendations, the overall impact will be increased.

Analysis of REF2021 Results

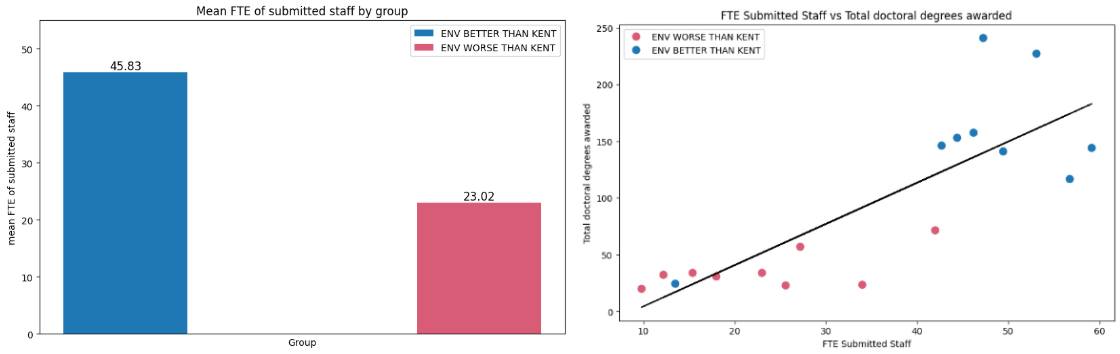
After reviewing the previous statements, academic staff and funding were identified as key points for Kent to improve its environmental score for research. To confirm this, additional observations and correlations between different factors can be examined.

Firstly, the results may be influenced by the number of doctoral degrees. From 2013 to 2019, universities in the b_env group had approximately 193 doctoral degrees per year, compared to around 47 for the w_env group (Figure 5). Therefore, increasing the number of doctoral degrees could be considered.



Figures 4 & 5 : Mean of doctoral degrees and evolution each year from 2013 to 2019

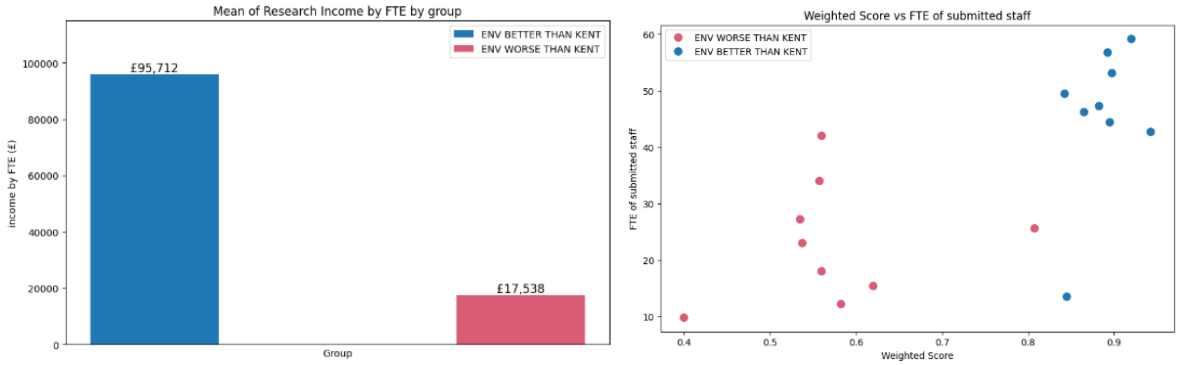
By calculating the average dedicated staff (FTE of submitted staff) per group, the b_env group's average is approximately double (45.83) that of the w_env group (23.02) (Figure 6). The number of doctoral and FTE Submitted Staff is strongly correlated (0.81). This difference in averages is likely explained by the lower number of doctoral, which requires fewer FTEs.



Figures 6 & 7 : Mean of FTE Staff and correlation with doctoral degrees

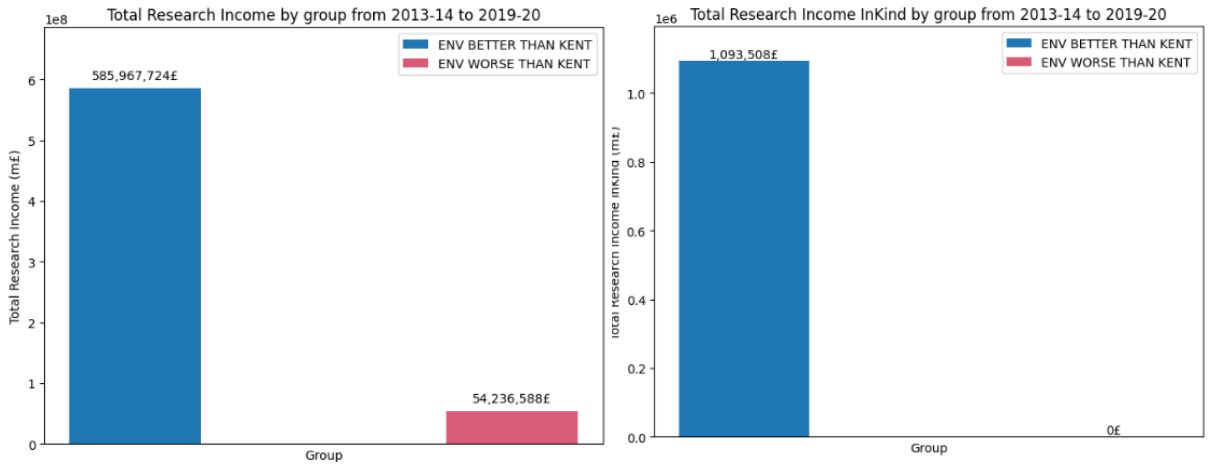
To increase the number of FTE Submitted Staff, in addition to increasing the number of doctoral, it would be necessary to raise their salaries to attract them and ensure their enthusiasm for conducting research or managing research services. There is an average salary difference of £78,174 between the two groups (Figure 8). Furthermore, increasing the

number of FTE Submitted Staff would also raise the weighted score, as these two factors are moderately correlated (0.71).



Figures 8 & 9 : Mean Income by FTE and Relationship between FTE and weighted score

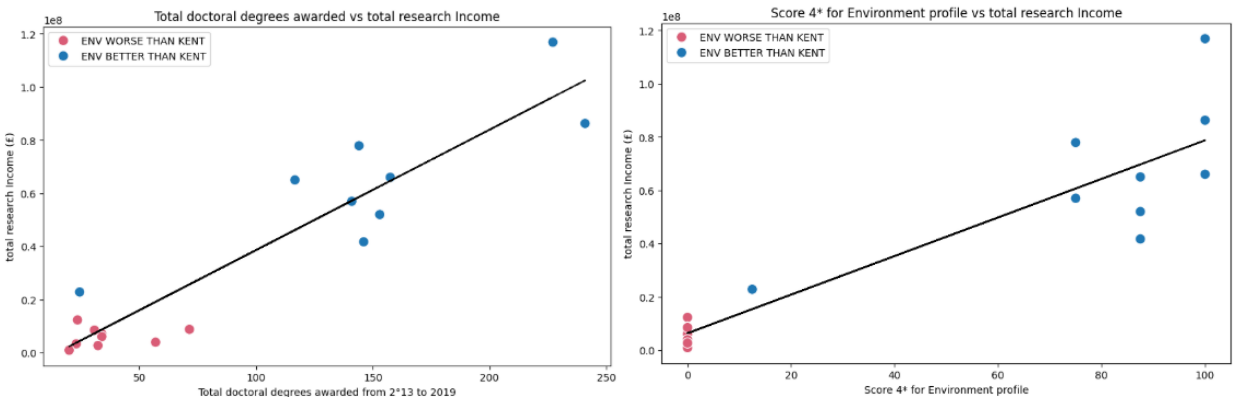
Looking at the total research income for both groups, the difference is striking: universities in the b_env group have 10 times more income than those in the w_env group (£585,967,724 vs £54,236,588, see Figure 10). Additionally, some universities in the b_env group also benefit from in-kind research income.



Figures 10 & 11 : Total type of Research Income by group

The number of doctoral and total research income being very high in the b_env groups suggests that these two factors are linked. Indeed, there is a correlation of 0.93. However, is it the number of doctoral or the income that has an impact on the final score? A correlation

of 0.92 exists between the environment score and income. Therefore, it could be concluded that the total research income plays a major role in achieving the 4* rating for the environment.



Figures 12 & 13 : Correlation between doctoral degrees/score 4 for Environment and research income*

From the previous results, funding seems to be the major factor influencing the environment criterion rating. It is therefore interesting to study where these revenues come from. The main categories identified as sources of research income are: 'EU Funding', 'UK Public Funding', 'Non-EU Funding', 'UK Charities', and 'UK Industry and Other'. However, after observing their distribution within the two groups, it appears to be approximately the same (Figure 14).

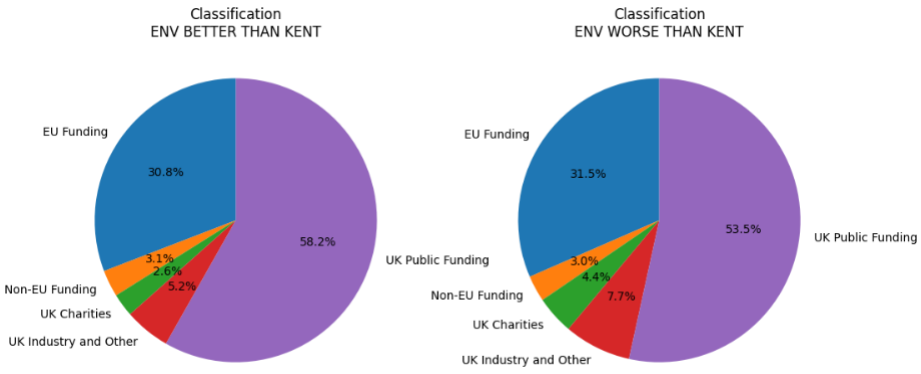


Figure 14 : Distribution of research income by sources

Conclusion

The analysis of universities in both groups (b_env and w_env) has identified key factors for improving the environment criterion rating in the context of REF2021. Universities in the b_env group stand out due to high research income, a large number of doctoral degrees, and a higher number of FTE Submitted Staff. Each of these aspects is strongly correlated with each other and with the environment score.

The statements on the unit's and institution's environment highlight the importance of adopting an enthusiastic tone and focusing on themes such as academic support, the global impacts (rather than individual) of research, and funding and investments.

In conclusion, to improve the environmental score, it is necessary to increase research income, increase the number of doctoral degrees, expand staff, and focus on the right themes while maintaining a positive tone in the statements.

The University of Kent could aim for a better environmental score in the next REF by aligning itself with the universities in the b_env group and implementing these recommendations.

APPENDIX

Note :

The code used to carry out these analyses can be found in the following notebook: *Practical_coursework_PartB_ib349.ipynb*