As part of the literature review, it is considered to understand the scoring method used by the REF (or RAE) expert panels in order to have an idea about the possible factors that affect the final overall scores. One particular report was found documenting the approaches used by the main five Funding Councils in analysing the impact case studies submitted to REF, namely The Nature, Scale and Beneficiaries of Research Impact. Impact case studies are part of the submissions to REF made by institutions, outlining how past researches benefit "the economy, society, culture, public policy and services, health, environment and quality of life". According to the report, the approach to assessing impact was a combination of text mining and qualitative analysis. In particular, three approaches were used during text mining - topic modelling, keyword searches and information extraction. In topic modelling, a cluster of words that frequently occurred in documents of similar contexts was defined as the "topic". Latent Dirichlet Allocation (LDA) algorithm was used during this process. Keyword searching made it easier to look at specific content in documents. The case studies were also matched against third party information through information extraction.

Given that the time constraint of the current team project, it might be difficult to replicate the text mining analysis for the 6679 impact case studies. It is also noticed that the published case studies may have certain words or phrases removed from the original submissions for publication purposes, which might be the keywords that lead to higher scores. Therefore, it is considered to use a similar approach to analyse the texts involved in the tabulated REF submission data - 100-word additional information and publication title.

The actual report will be uploaded as a reference.