Capstone Project 2: Broad Project Ideas

1. **Predicting a Technology's Position in the Hype Cycle**
    a. Case. The Gartner Hype Cycle displays various technology's maturity on a graph to predict, among other things, if a certain technology has reached a 'bubble'. You could buy this information from Gartner at a hefty price, or share hype cycles yourself for the benefit of investors and entrepreneurs.
    b. Method. Categorising news article headlines frequency is an academically proven way to determine a product or technology's hype. I will be using the https://newsapi.org/ API to access news articles from sources such as ABC News, Al Jazeera, Associated Press, BBC News, Business Insider, CNN, Buzzfeed and many more. An obvious technology to analyse with this approach is Bitcoin. The hype of Bitcoin is the most recent tech that followed a typical hype cycle bubble trend.
    c. Note. The classification of articles into different categories should be straight forward, however, the completed model will be productionised and published in the form of a blog post (Streamlit?). On completion, any technology, as a keyword, could be analysed in the same approach.

2. **Building a corpus of scholarly articles on nutrition (or any scientific field of study) to identify contextual limitations**
    a. Case. The scientific field of nutrition is controversial due to the context in which most studies were conducted. Nutritionists can not agree on how the control groups itself influenced findings. A better understanding is required of how features in control groups would interact with features from other studies to gain better insight into the complex mechanisms.
    b. Method. Academic papers through an API such as https://libraries.mit.edu/scholarly/publishing/apis-for-scholarly-resources/ from MIT could contribute to a large corpus from which contextual keywords (word2vec) could visually identify and group relevant relational features. From here the effect of one feature on another could reveal meaningful insight.
    c. Note. This option will be more complex and there will not be enough time to productionise this model.