WAF Challenge 2020

The WAF challenge for 2020 will be drawing data from a popular kaggle competition, "Trending videos on Youtube"(https://www.kaggle.com/datasnaek/youtube-new). The kaggle competition contains data collected from Youtube videos across different geographic zones. For this challenge, we would like you to only focus on the specific dataset of **US audiences.** The data dictionary and context can be found in the link above.

We designed **three prompts** for this challenge, of which you're **only required to do one**.You may choose to do two or all three if you'd like, but the number of prompts completed will not impact our assessment.

Prompt 1: Imagine that you are working on the recommendation engine team of Youtube, please design a recommendation engine (i.e., clustering engine) for Youtube videos (Note: we are not looking for extremely fancy methods involving video processing or recognition. We want to gauge your ability to perform <u>unsupervised learning tasks with both structured and unstructured data</u>). Explain your rationale in designing the engine and make sure to include some examples of working recommendations.

Prompt 2: Imagine that you are working on the digital ranking team of Youtube, please create a predictive model to assess which Youtube videos should be on the front page (Note: We purposefully left the specific metric of this open-ended and up to your discretion. For example, you can predict the number of views, number of likes, or likelihood of being a mega-hit or receiving more than 600 million views). Like most <u>classic ML tasks</u>, while the quality of the model is important, your thought process, including why you chose the metric that you chose, is just as essential for our assessment.

Prompt 3: Imagine that you are a product analyst at Youtube and you work extremely close with the product management team in designing new features, please explore the dataset to find interesting inferences and come up with a product recommendation for Youtube. This prompt is less technical relative to the other prompt, so we would like you to focus on showcasing your abilities in <u>data story-telling and product intuition</u>. For the assessment, we would be focusing on your data exploration, creativity, and your ability to turn data into real-world insights.

When putting up the final report, we would advise you to explain your work as concisely as possible. The deliverable that we are expecting is a **3-page summary slide deck**, along with the **code/excel document you used for the challenge**. Please do not surpass the page limit because only the first 3 pages will get reviewed. The deadline for the submission will be **January 31st at midnight.** Please submit the challenge to whartonanalyticsfellow@wharton.upenn.edu. Thank you and we look forward to seeing your work!