***Problem statement - Mine the Social media data from Twitter, LinkedIn and Online news channels to find out the key topics of discussion, influencers, sentiments change in the medical community regarding global fund raising initiatives, including level of crowd sourcing, amount raised, platform (fees levied by platform (if any) for spinal muscular atrophy (SMA) in pediatric patients***

***Disease Areas: spinal muscular atrophy (SMA)***

***Data to be used – Last 90 days***

***Time to complete the case – 3 working days***

***Tools – R or Python***

***Output expectation –***

1. ***Deck explaining the process and results***
2. ***Graphical representation of key topic of discussion, Sentiment analysis***
3. ***Online Influencers/Organizations/Groups***
4. ***Code files used to mine the data***
5. ***Output files including the data fetched***

The management decided to adopt an NLP and AI backed strategy, which would involve identification of barriers of fund raising, key concerns expressed by audience/patients, experience of fundraising initiatives and topics of discussion.

**Suggestions (Optional):**

You may design a NLP/AI backed strategy to solve for the above problem:

* Filter the noise (irrelevant) data
* Extract frequently used phrases
* Extract key actors (audience, organizations, etc.) and who are the influencers
* Grouping clusters of conversations with labels for clusters
* Sentiment analysis (analyze model performance)
* Generate insights from extracted metadata and provide summary/conclusion (feel free to use any visualization/outcomes)

You do not have to limit yourself to the above points but feel free to add your own ideas as you deem fit. You have to present your submission before a hiring panel and answer questions that they have in the course of the interview.