# Project proposal: Coronavirus tweets NLP - Text Classification

Question/Need:

* What is the question behind your analysis or model and what practical impact will your work have?

***Throughout covid pandemic time, what are the hot topics that people care about?***

***What is people’s attitude towards covid (or certain topic, such as vaccination) as the time goes by?***

* Who is your client and how will that client benefits from exploring this question or building this model/system?

***Policy makers will be more interested to know the sensitivity of certain topic on world trend. If they can understand what the hottest topic on tweet is, they can see the policy impact and it helps them to better make future plans.***

Data Description:

* What dataset(s) do you plan to use, and how will you obtain the data? Please include a link! (The link can be to the dataset you’re downloading, the site you’re scraping, etc.)

***I plan to use tweet that is covid relative from Mar 2020. I download the dataset from Kaggle: https://www.kaggle.com/datasets/datatattle/covid-19-nlp-text-classification?sort=votes***

* What is an individual sample/unit of analysis in this project? In other words, what does one row or observation of the data represent?

***One row represents one tweet that is covid related.***

* What characteristics/features do you expect to work with? In other words, what are your columns of interest?

***Tweet date and tweet content.***

* If modeling, what will you predict as your target?

***What topic does one tweet talk about?***

***What is the sensitivity level trend along the time on certain topics (e.g. vaccination)?***

Tools:

* How do you intend to meet the tools requirement of the project?

***Use python NLP library such as such as NLTK, spaCy, gensim, scikit-learn.***

* Are you planning in advance to need or use additional tools beyond those required?

***No.***

MVP Goal:

* What would a [minimum viable product (MVP)](https://app.thisismetis.com/courses/155/assignments/1479) look like for this project?

***Data cleaning have been done and some initial key visuals will be presented. Ideally, I should be able to identify several topic patterns.***