**Exploratory Data Analysis**

**Objectif :** Comprendre au maximum les données dont on dispose pour définir une stratégie de modélisation.

**Checklist de base**

***Analyse de la forme :***

* Identification de la target : 'SARS-Cov-2 exam result'
* Nbres de lignes et colonnes : (5644, 111)
* Types de variables :: float64 : 70 ; Object : 37 ; int64 : 4
* Analyse des valeurs manquantes : Beaucoup de NaN ( moitié des variables > 90% de NaN)

***Analyse du fond :***

* Visualisation de la target (Histo/Box) :
* Comprehension des différentes viriables(Internet)
* Visualisation des relations features-target (Histo/Box)
* Identification des outliers

**Pre-processing**

***Objectif :*** transformer la data pour la mettre dans un format propice au Machine Learning

* Création du Train / Test set
* Elimination des NaN
* Encodage
* Suppression des outliers néfastes au modèle
* Feature Selection
* Feature MLengineering
* Feature Scaling

**Modeling**

**Objectif :** Developper un modèle de ML qui répond à l’objectif final.

* Définir une fonction d’évaluation
* Entrainement de différents modèles
* Optimisation avec GridSearchCV
* (Optionnel) Analyse des erreurs et retour au Preprocessing /EDA
* Learning Curve et prise de décision