Hiring Challenge

>>> The Dataset <<<

groover_dataset_challenge.csv is a file containing a sample of 10,000 feedbacks that were given on Groover.

A submission is a track sent by one band/artist to an influencer (a music industry professional). The feedback is written by the influencer in response to the artist sending his/her track.

The dataset contains the following columns:

- **band_id:** int representing the band
- **influencer_id:** int representing the influencer
- **feedback:** Free text written by the influencer to the band/artist
- **score:** float between 0 and 1 corresponding to the influencer's appreciation

>>> The Challenge <<<

The goal is to evaluate your ability to treat a given dataset containing a Natural Language Processing task. We want to integrate a NLP component to a greater machine learning model, your task is to provide the best way to start this project!

Exercise 1: General Statistics

Describe the dataset using metrics and visualizations. This part's goal is to make sure you understand the data you will be using and to show us how you are able to explain it.

Exercise 2: Preprocessing Natural Language

Find a way to clean the feedback column. Describe the processing steps you choose to apply and why are you deciding to do so.

Exercise 3: Fitting Into Analysis

Transform the preprocessed text into a format that would fit the input of a maching learning model and explain your choice. You may use any pre-existing model or transformation technique you can find, but don't forget to explain your choice, its advantages and drawbacks!

>>> Rules <<<

- Jupyter Notebooks are our preferred output format, whether as *HTML or .ipynb.*

- Even if you do not have a full answer to an Exercise, *please explain your ideas and reasoning*, leads that you would follow with more time or resources. Bear in mind that this is an open challenge.

- *You are free to use any tool you can find*, and indicate it when you do. You can include any source you find relevant (publication, blog post, tutorial...)

- If you have any question, don't hesitate to contact me: **santiago@groover.co**

Thank you and good luck!