Exercises

Select only ONE of the following two exercises to work on
Feel free to use any tool you are comfortable working with.
As a result of the test, we are expecting:

- A worked solution in code (optional)
- A document or a 10 min presentation with
    - Research resources that you used to approach the problem
    - What is the solution that you are proposing?
    - What are your conclusions on your solution?
    - Is there a better way to resolve the exercise if you were to have more time to work on it?

Exercise 1: Language classification

Characters used in a language have a direct correlation with the language itself.
Design a solution that will allow a user to provide a document and identify (classify) the language it was written on.
The languages that this tool should identify are: Spanish, English, Italian, German, French, Portuguese, and Danish.
You can use the following documents as a sample or test data for your solution:

https://es.wikipedia.org/wiki/Divina_comedia
https://it.wikipedia.org/wiki/Divina_Commedia
https://en.wikipedia.org/wiki/Divine_Comedy
https://de.wikipedia.org/wiki/Göttliche_Komödie
https://fr.wikipedia.org/wiki/Divine_Comédie
https://pt.wikipedia.org/wiki/Divina_Comédia
https://nl.wikipedia.org/wiki/De_goddelijke_komedie

Exercise 2: Categorizing music

Music has a deep history of recordings; also, there are ways we can split music into different categories.
Trying to classify the entire music data is a big challenge, so let's focus just on Jazz music.
Here you will find the best 100 jazz albums: http://www.scaruffi.com/jazz/best100.html, using https://www.discogs.com API you should be able to determine the style of jazz of each one.
So your goal will be to classify and group the albums in 25 grouping categories.