There are two parts: Written Assignment and Coding Assignment

1. Written Assignment

You have received this email from a customer. Use the available resources online to draft a response to this email.

"Hi There!
We're interested in integrating our current scraping solution to work with SPM. (Zyte Smart Proxy Manager).
At the moment, we're using python + headless browser (selenium). What should we do to integrate SPM into this stack?
Also, we're kinda confused in how we'll rotate the proxies to get successful responses. Can you clarify how we do this?
In regards to IP regions, how do we select which region/country we want to perform requests?
Regards,
Customer"

Deliverable: An email draft as a text file or a document. You can share it via email or provide a link to the file in github.

2. Coding Assignment

Scrapy is an open-source framework for extracting data from websites. In order to do this assignment, you would have to install scrapy from https://scrapy.org/ and you would need python available on your machine.

Zyte has an offering called Scrapy Cloud, to deploy your spider in the cloud. You will be signing up for a free account in Zyte Dashboard https://app.zyte.com/ and use the free unit of Scrapy Cloud. No Credit card information is required. No other subscription is necessary for this task.

2.1: Basic Spider and Deploy in Scrapy Cloud

The task is to build a spider to extract the following fields from all the 1000 books available in the website: books.toscrape.com:

Fields to extract: book title, book price, book image URL, book details page URL

Constraints:

1. Your spider should visit all the category pages and all the information should be extracted from those.
2. Your spider should not visit the individual book pages. This way, you'll save a considerable amount of bandwidth.
3. Use of plugins and custom pipelines (optional - bonus points)
4. Stop the spider after first 750 items are scraped (optional - bonus points)

Deliverables:
Please provide the following two files in a github or bitbucket repo, and share the link:

a) Code of the spider
b) CSV or XML or JSON file of the four fields for the 1000 books
c) Link to the project deployed in Scrapy Cloud in Zyte Dashboard

2.2: Spider with Javascript

Scrape and extract quote, author and tags from http://quotes.toscrape.com/js/ You would need to use your choice of Headless browser or Splash to be able to execute the javascript. No need to deploy this spider on Scrapy Cloud.

Deliverables:
Please provide the github or bitbucket link to the code of the spider and extracted data: quote, author, and tags (in any format)