# Introduction

For this assignment, you will build an address lookup tool for Spanish addresses. You will create a database for this tool in MongoDB, and you will build an application so users can look up addresses in your database.

Store your code repository or repositories on GitHub and share it with us. Optionally, you can upload your Docker image to Docker Hub.

We expect you to spend 4 to 6 hours on this assignment. You can leave your answers where applicable for each part of the assignment below. Good luck!

# Description

## Part 1

First, you will set up the data pipelines including the ETL to create the database for this tool. Make sure you do this in a way where the data can be regularly or even continuously updated (additionally, you are allowed to make use of a workflow management system such as Apache Airflow). You may store the raw data in SQL if you prefer, but the transformed data has to be stored in MongoDB (additionally, you are allowed to make use of a search engine such as Elasticsearch). Make sure your data is securely accessible by us. Save the field mapping / document structure in JSON format and share it with us.

Use the following sources to build your database:

* [https://results.openaddresses.io/sources/es/25829](about:blank)
* [https://results.openaddresses.io/sources/es/25830](about:blank)
* [https://results.openaddresses.io/sources/es/25831](about:blank)
* [https://results.openaddresses.io/sources/es/32628](about:blank)
* [https://results.openaddresses.io/sources/es/nc/statewide](about:blank)

Pay attention to the following:

* Use the sources to create one or more collections in a MongoDB database
* Create your data pipeline / ETL in a way that enables regular or continuous updates
* Write your data pipeline / ETL (mainly) in Python
* Share a JSON file with the field mapping / document structure

## Part 2

Second, you will build a small web application (RESTful API only) that accesses the data. This API will provide the address lookup functionality. Write your API in Python. You can choose any framework you prefer, but this API will be used by many users performing many requests, so make sure it is optimized for high performance. It should accept a string containing address information (this address line can be encapsulated in a JSON object, if you prefer). The API should return a JSON object with the address information from your database. Make sure the API is able to handle input strings with address lines in different format or order, possibly containing spelling errors or abbreviations. Dockerize your application and share it with us.

Pay attention to the following:

* Create an application with Python consisting of a high-performance RESTful API with at least one endpoint for address lookups, with string or JSON input and JSON output
* Run your application and share screenshots of:
  + at least one successful request
  + at least one unsuccessful request (also explain why this request was unsuccessful)
* Share your Dockerized application with us and make sure it's securely accessible

## Part 3

Store your code repository or repositories on GitHub and share it with us. Optionally, you can upload your Docker image to Docker Hub.

Please answer the following questions:

1. *How much time did it cost?*

Part 1: 5 hours

Part 2: 2 hours

1. *How difficult was this assignment to achieve?*

Part 1

1. very easy - 2.easy - 3. moderate - 4. **difficult** - 5. very difficult

Part 2

1. very easy - 2.easy - 3. moderate - 4. **difficult** - 5. very difficult

1. *What did you think of the assignment?*

The assignment is a complete set of practices for data engineering role that needs having experience with setup different environments and connect them by the data pipeline.