Importing data into Amazon Sagemaker

video available here: https://youtu.be/PCKP5pf-78M

callum@paxata.com
In Amazon Sagemaker create a Notebook instance
Open the Notebook – Select a coding environment
Get the Paxata information for the REST call you will make

- URL: https://dataprep.paxata.com
- Dataset ID: 7ce56be7c0d948868d38530018b80cb9/3
- Authorization Token: Token
Using Paxata docs get all of the information for the REST call

Documentation - REST API : Data Source Exports API – /datasource/exports

Endpoint for exporting items from the library.

HTTP GET
Not supported.

HTTP POST

URL format:
```
dataSourceId/exports/datasourceId
```

This request exports the given data set (specified by the ID in dataSourceId) using the specified data source (specified by ID in dataSourceId). If no version number is specified in the version argument, the request exports the latest version.

The table below lists the predefined arguments. There are additional variable arguments that are specific to the data source, and some are required. To get the data source variable arguments, use the REST API /datasource/fields endpoint. See the topic Data Source Fields API. Pass any data source variable arguments with the prefix "r.". For example, if the data source argument is jdbc.user with the value "user", the URL must include the text r.jdbc.user=user.

The Paxata server can export data to the local file system without data source configuration by setting the dataSource argument to the value local.

Important
For remote export requests, the REST API request can return before the export is not fully complete. Returning a result from the request does not imply that the export succeeded. Also, prepare your code for an export-related REST API request to timeout. Your export could be continuing between the server and the remote system. For exports to remote systems, check the status of export requests using the Export Audit Log REST API endpoint /library/exports. See Library Export API – library/exports. If the status property for one of the exports has the status property set to ".", the export is still running.
Paxata and Sagemaker

Import “requests” module to call the Paxata REST API and then make the call

Check the status is “200” (OK) (otherwise check the r.text variable)

Optional – Write the file out to the file system
A new JSON file has been created
View the JSON file in Jupyter