

Data Wrangler

Data Wrangler is a cloud-based solution for exploratory data analysis and data transformations on AWS

Helps with extract, transfer, and load (ETL) pipelines.

S3 Upload

S3 is your object store. Data Wrangler uses S3 to import data into its application in order for it to process it. You'll need to have a bucket, which one should be created for you already.

Upload the `data.csv` file into your bucket. Remember where you put it, you'll need it for the next step.

Data Wrangler: Pipeline Creation

To process your `data.csv` file, you'll select the `Amazon S3` option to import data.

First, you'll create a transformation using one-hot encoding for the `b` feature. Output style will be columns. Then create an analysis of your modified dataset, using the Table Summary option.

Data Wrangler: Pipeline Export

To export your dataset, you export your flow with the **Save to S3** option. With the new notebook that is created for you, execute steps 1 and 2. Step 3 is not required for our use case as we will build our own model separately.

When completed, a new bucket will be created with your Data Wrangler exports.

Additional Resources

- For more information about Data Wrangler features, their [main page](#) has a good overview.
- If you want to learn more about Data Wrangler, check out [AWS's official documentation](#).