

Glossary

For your reference, here are all the new terms we introduced in this lesson:

- **Exploratory Data Analysis (EDA)** - The method of analyzing data to understand its main ideas.
- **SageMaker** - AWS's suite of tools for machine learning
- **Feature Engineering** - Changing or creating features from a dataset into a form that is used by a machine learning model.
- **DataFrame** - A highly modifiable row and column based table of data containing mixed value types.
- **Histogram** - A diagram that represents the frequency distribution of set of data.
- **Correlation** - A relationship between two or more things.
- **Extract, transform, and load (ETL)8** - A multistep process often done with data to acquire and prepare it for higher level tasks,
- **Labeled data** - Data that has target values representing individual data points
- **Unlabeled data** - Data that does not contain a target value, meaning there is no way to create predictions from its feature values.
- **JSON Lines (JSONL)** - a file that contains json objects on each line
- **Sentiment** - point of view from an individual related to a review or body of text. Positive means happy sentiment, negative means unhappy sentiment.