Lesson Review

Using AWS Sagemaker, we began to understand the importance of EDA and use their tools for exploration. In this lesson, we learned how to:

- Using Sagemaker Studio, we were able to perform some basic exploratory data analysis and understand how S3 plays a role in data availability.
- We created a similar analysis with a Sagemaker service called Data Wrangler. This allowed us to quickly build a pipeline of analysis without any code.
- Using the Pandas framework, we created a new dataset and performed some basic EDA using the frameworks builtin methods.
- For creating new datasets, we used Ground Truth to build an annotation project and label data using a private workforce.

With the skills you gained here and the tools available through AWS Sagemaker, you can truly understand the data and feel confident to proceed to the next step in data cleaning, feature engineering, and model training!