**Exploratory Data Analysis Overview**

In this video, we will review Exploratory Data Analysis. Exploratory Data Analysis is the first step of any data science project. In the first lab, you will perform some Exploratory Data Analysis using a database. In the second lab, you will see if the data can be used to automatically determine if the Falcon 9’s first stage will land. Some attributes can be used to determine if the first stage can be reused. We can then use these features with machine learning to automatically predict if the first stage can land successfully, for example. You can observe that the success rate since 2013 has improved. We can incorporate this as a feature via launch Number. We see that different launch sites have different success rates. As a result, they can be used to help determine if the first stage will land successfully. CCAFS LC-40 has a success rate of 60%, while KSC LC-39A and VAFB SLC 4E have a success rate of around 77%. Combining attributes also gives us more information. If we overlay the result of the landing outcomes as a color we see that CCAFS LC-40, has a success rate of 60%, but if the mass is above 10,000 kg the success rate is 100%. Therefore, we will combine multiple features. In the lab, you will determine what attributes are correlated with successful landings. The categorical variables will be converted using one hot encoding, preparing the data for a machine learning model that will predict if the first stage will successfully land.