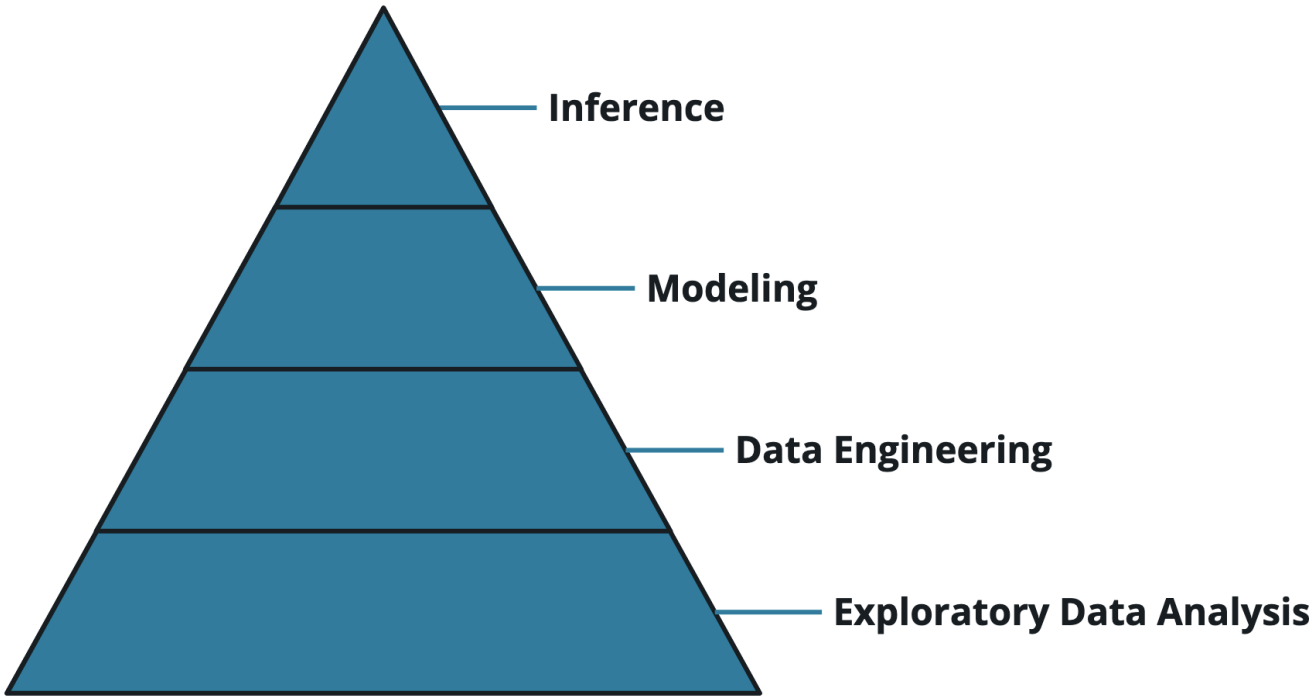
Exploratory Data Analysis (EDA) is the set of tools and processes that allow you to understand data better. This includes creating summary statistics, identifying outliers, and visualizing feature distributions.

With machine learning as a whole, EDA is the foundational layer. Together with data engineering, modeling, and inference, it creates the ML pyramid. EDA is the foundational layer because, without it, the rest of the ML workflow will not be successful. EDA provides an understanding and insight into the data that can be directly fed into data engineering.



ML Pyramid

QUIZ QUESTION

Exploratory data analysis is at what level in the machine learning pyramid?

At the base level, without a good understanding of your data and what it "looks" like, you will have a difficult time building a successful model.

The top level, it's useful but not as useful as creating a super accurate model.

Mid level, is not the most important, but also not the least important.

SUBMIT

Additional Resources

- If you would like a general overview and history of exploratory data analysis, check out its Wikipedia page.
- Here is an engineering handbook that has some additional information on EDA. Produced by the NIST.