**Categories of Data Science Tools**

Open source tools are available for various

data science tasks. In this video, we’ll have a look at the

different data science tasks. In subsequent videos we’ll walk through

the most commonly used open source tools for those tasks. The most important tools are covered throughout

this course.

**Data Management** is the process of persisting and retrieving data.

**Data Integration and Transformation**, often referred to as Extract, Transform, and Load, or “ETL,” is the process of retrieving data from remote data management systems. Transforming data and loading it into a local data management system is also part of Data Integration and Transformation.

**Data Visualization** is part of an initial data exploration process, as well as being part of a final deliverable.

**Model Building** is the process of creating a machine learning or deep learning model using an appropriate algorithm with a lot of data.

**Model deployment** makes such a machine learning or deep learning model available to third-party applications.

**Model monitoring and assessment** ensures continuous performance quality checks on the deployed models. These checks are for accuracy, fairness, and adversarial robustness.

**Code asset management** uses versioning and other collaborative features to facilitate teamwork.

**Data asset management** brings the same versioning and collaborative components to data. Data asset management also supports replication, backup, and access right management. Development environments, commonly known as

**Integrated Development Environments, or “IDEs”,** are tools that help the data scientist to

implement, execute, test, and deploy their work.

**Execution environments** are tools where data preprocessing, model training, and deployment take place.

Finally, there is **fully integrated, visual tooling available** that covers all the previous tooling components, either partially or completely. This concludes this video. In the next video we’ll start looking at open source tools for data science tasks.science tools