Tools for Data Science

▼ Week 2

Languages of Data Science:

- You should select a language to learn depending on your needs, the problems you are trying to solve, and who you are solving them for
- The popular languages are Python, R, SQL, Scala, Java, C++, and Julia
- JavaScript, PHP, Go, Ruby, and Visual Basic all have their own unique use cases as well
- The problems you need to solve can be related to your company, role, and age of the existing application

Introduction to Python:

- Python uses clear and readable syntax
- Python has a huge global community and a wealth of documentation
- For data science you can use Python's scientific computing libraries like Pandas, NumPy, SciPy, and Matplotlib
- Python can also be used for Natural Language Processing (NLP) using the Natural Language Toolkit (NLTK)
- Python community has a well-documented history of paving the way for diversity and inclusion efforts in the tech industry as a whole

Introduction to R:

- The Open Source Initiative (OSI) champions open source, while the Free Software Foundation (FSF) defines free software
- Python is open source, and R is free software
- R language's array-oriented syntax makes it easier to translate from math to code for learners with no or minimal programming background
- R has become the world's largest repository of statistical knowledge

Introduction to SQL:

- SQL is different from other software development languages because it is a non-procedural language
- SQL's scope is limited to querying and managing data
- SQL was designed for managing data in relational databases
- SQL behaves like an interpreter between you and the database
- If you learn SQL and use it with one database, you can apply your SQL knowledge with many other databases easily

Other Data Science Languages:

- Data science tools built with Java include Weka, Java-ML, Apache MLlib, and Deeplearning4
- For data science, popular program built with Scala is Apache Spark that includes Shark, MLlib, GraphX, and Spark Streaming
- For data science, TensorFlow, MongoDB and Caffe were built with C++
- Programs built for Data Science with JavaScript include TensorFlow.js and R-js
- One great application of Julia for Data Science is JuliaDB

Summary:

- You should select a language to learn depending on your needs, the problems you are trying to solve, and whom you are solving them for.
- The popular languages are Python, R, SQL, Scala, Java, C++, and Julia.
- For data science, you can use Python's scientific computing libraries like Pandas, NumPy, SciPy, and Matplotlib.
- Python can also be used for Natural Language Processing (NLP) using the Natural Language Toolkit (NLTK).
- Python is open source, and R is free software.
- R language's array-oriented syntax makes it easier to translate from math to code for learners with no or minimal programming background.
- SQL is different from other software development languages because it is a non-procedural language.
- SQL was designed for managing data in relational databases.
- If you learn SQL and use it with one database, you can apply your SQL knowledge with many other databases easily.
- Data science tools built with Java include Weka, Java-ML, Apache MLlib, and Deeplearning4.
- For data science, popular program built with Scala is Apache Spark which includes Shark, MLlib, GraphX, and Spark Streaming.
- Programs built for Data Science with JavaScript include TensorFlow.js and R-js.
- One great application of Julia for Data Science is JuliaDB.