**Tableau for Data Science and Data Visualization**

**The Data Science Workflow**

Collection

Exploration (Tableau)

Munging

Modeling

Validation

Reporting (Tableau)

**What is Tableau**

Tableau is a business intelligence and data visualization tool allowing users to make sense of their data through interactive charts, graphs, and diagrams.

Dimensions and Measures

Converting measures to dimensions

**Two Types of DBs**

Relational DB (SQL)

Organize data into one or more tables

Each table has columns and rows

A unique key identifies each row

Non-Relational DB (noSQL / not just SQL)

Organize data in anything but a traditional table

Key-value stores

Documents (JSON, XML, etc)

Graphs

Flexible tables

NRDBMS help users create and maintain a non-relational DB (mongoDB, dynamoDB, apache cassandra, firebase, etc…)

Any non-relational DB falls under this category, so there’s no set language standard. Most NRDBMS will implement their own language for performing CRUD and administrative operations on the DB.

**DB Queries**

Requests made to the DBMS for specific information

As DB structure becomes more complicated, it becomes more difficult to get the specific pieces of information; queries are critical at this juncture

**Tables and Keys**

Three primary features of a DB table -> key (primary [surrogate and natural], foreign keys, composite keys), columns (attributes), and rows (data)

**SQL**

Three primary features of a DB table -> key (primary [surrogate and natural], foreign keys, composite keys), columns (attributes), and rows (data)

REMEMBER – SQL implementations vary between systems. Not all RDBMS’ follow the SQL standard the same way. For the most part, conceptually every vendor is the same but the implementation details will vary.

SQL is a hybrid of four different languages:

DQL – data query language (querying the DB and getting information from it)

DDL – data definition language (defining DB schemas)

DCL – data control language (controlling user access and permission to data)

DML – data manipulation (CRUD operations)

A query is a set of instructions given to the RDBMS, written in SQL, that tell the RDBMS what information you want it to retrieve for you.