



✓ **Congratulations! You passed!**

TO PASS 80% or higher

Keep Learning

GRADE
100%

Extracting Data

LATEST SUBMISSION GRADE

100%

1. Which of the following best describes a dataset?

1 / 1 point

- ☐ An industry standard data storage format for data science projects.
- ☒ A collection of data directly used to accomplish business goals.
- ☐ A table containing the results of a database query.
- ☐ A set of repositories where relevant data is scattered throughout.

✓ **Correct**

Often multiple data sources will contain the necessary data, which may also be in many formats. Through extraction of that data, a dataset can be built that can be directly used to accomplish the business goals stated in the project specifications.

2. Which items best represent structured data?

1 / 1 point

- ☐ Spreadsheets and videos
- ☐ Images and database tables
- ☒ Spreadsheets and database tables
- ☐ Videos and images

✓ **Correct**

These are both structured data because it is easy to search and retrieve specific data from them.

3. What is the term for the part of data used to store the variables used by a model to make estimations and whose quantity represents the model's dimensions?

1 / 1 point

- ☐ Row
- ☐ Data example
- ☐ Value
- ☒ Feature

✓ **Correct**

A feature, also called an attribute or column, defines the variables that exist in the dataset, and the total number of different features make up the model's dimensions.

4. What are the three primary factors of big data?

1 / 1 point

- ☐ Volume, variety, and verbosity
- ☒ Volume, variety, and velocity
- ☐ Volume, veracity, and velocity
- ☐ Volume, variety, and veracity

✓ **Correct**

These are considered the three primary factors, although veracity is often included as a fourth factor.

5. What type of data source is most likely to contain both structured and unstructured data?

1 / 1 point

- ☒ Data lake
- ☐ Data warehouse
- ☐ Operational data store
- ☐ Data mart

✓ **Correct**

A data lake can store data that is kept for possible future needs. It can store both structured and unstructured data from many sources and it normally will keep the data in its original forms.

6. What is the most common type of architecture used to access APIs for public data sources?

1 / 1 point

- ☐ XML
- ☐ HTML
- ☐ JSON
- ☒ RESTful

✓ **Correct**

A RESTful API provides a web-based interface that uses fundamental HTTP operations to create, read, update, and delete values that have been exposed by the interface. The standard architectural method to do this is called representational state transfer (REST).

7. Which of the following best describes the ETL process?

1 / 1 point

- ☒ The process of combining data from multiple sources, converting the data into a consistent and useful format, and loading the result into a destination for later analysis.
- ☐ The process of combining data from multiple sources, doing upfront data validation, and loading the original for comparison.
- ☐ The process of querying data from multiple databases, transforming all values to ANSI standard format, and loading the result into a destination for later analysis.
- ☐ The process of querying data from a database, removing any obvious errors, and storing the result for later analysis.

✓ **Correct**

Extraction is pulling data from various sources, often with differing structures. Transformation is converting data into a proper and uniform format. Loading is inserting the resulting data into a destination for later analysis and modeling.

8. What specific file type would store this record: 2021-01-27 03:00:02,92,"Sold",0.14

1 / 1 point

- ☒ CSV
- ☐ JSON
- ☐ DataFrame
- ☐ Text

✓ **Correct**

A comma-separated values file is a type of text file that stores each value separated by a delimiter, which in this case is a comma. The first row in a CSV file contains the column labels that define the variables for each row. CSV is commonly used as a means of exporting or importing records from dissimilar systems.

9. What is an SQL query that retrieves all the rows from Table A (left) and just the matching rows from Table B (right) called?

1 / 1 point

- ☐ Right outer join
- ☐ Inner join
- ☒ Left outer join
- ☐ Full outer join

✓ **Correct**

A left outer join returns all the rows from Table A and the rows from Table B that have a matching key.

10. Which of the following databases are NOT considered to be NoSQL?

1 / 1 point

- ☐ Wide-column
- ☐ Document
- ☐ Key-value
- ☒ Relational

✓ **Correct**

Relational databases normally support SQL queries, and therefore do not qualify as NoSQL.