# Your grade: 100%

Your latest: 100% • Your highest: 100% • To pass you need at least 70%. We keep your highest score.

Next item →

**1.** What is a CSV file? 1 / 1 point

- ○ CSV files are a standard way to store data across platforms.
- ◉ CSV files are rows of data or values separated by commas.
- ○ CSV makes data readily available for analytics, dashboards, and reports.
- ○ CSV is a method of JavaScript Object Notation.

**Correct**
Correct. CSV, or Comma Separated Value, files are rows of data or values separated by commas.

**2.** What are residuals? 1 / 1 point

- ○ Residuals are a method for handling identified outliers.
- ○ Residuals are data removed from the dataframe.
- ○ Residuals are a method to standardize data.
- ◉ Residuals are the difference between the actual values and the values predicted by a given model.

**Correct**
Correct. Residuals are model prediction errors.

**3.** If removal of rows or columns of data is not an option, why must we ensure that information is assigned for missing data? 1 / 1 point

- ○ Information must be assigned to prevent outliers.
- ○ Missing data may bias the dataset.
- ○ Assigning information for missing data improves the accuracy of the dataset.
- ◉ Most models will not accept blank values in our data.

**Correct**
Correct. Information must be given for every feature and label in a dataset.

**4.** What are the two main data problems companies face when getting started with artificial intelligence/machine learning? 1 / 1 point

- ○ Lack of training and expertise
- ◉ Lack of relevant data and bad data
- ○ Data sampling and categorization
- ○ Outliers and duplicated data

**Correct**
Correct. Companies need to collect and organize their data to make it ready before leveraging it for machine learning.

**5.** What does SQL stand for and what does it represent? 1 / 1 point

- ○ SQL stands for Sequential Query Language, and it represents a set of relational databases with fixed schemas.
- ○ SQL stands for Structured Query Language, and it represents databases that are not relational, they vary in structure.
- ◉ SQL stands for Structured Query Language, and it represents a set of relational databases with fixed schemas.
- ○ SQL stands for Sequential Query Language, and it represents a set of sequential databases with fixed schemas.

**Correct**
Correct. SQL is the set of highly structured relational databases with fixed schema.

**6.** What does NoSQL stand for and what does it represent? 1 / 1 point

- ◉ NoSQL stands for Not-only SQL, and it represents a set of databases that are not relational, therefore, they vary in structure.

- NoSQL stands for Non-Structured Query Language, and it represents a set of relational databases with fixed schemas.
- NoSQL stands for Not-only SQL, and it represents a set of databases that are relational, therefore, they have fixed structure.
- NoSQL stands for Non-Structured Query Language, and it represents a set of non-relational databases with varied schemas.

**Correct**
Review the Feature Engineering and Variable Transformation-Background video.

**7.** What is a JSON file? 1 / 1 point

- JSON stands for JavaString Object Notation, and they have very similar structure to Python Dictionaries.
- (•) JSON stands for JavaScript Object Notation, and it is a standard way to store the data across platforms.
- JSON stands for JavaScript Object Notation, and it is a non-standard way to store the data across platforms.
- JSON stands for JavaString Object Notation, and it is a standard way to store the data across platforms.

**Correct**
Correct. JSON stands for JavaScript Object Notation, and those files are going to be a standard way to store data across platforms.

**8.** What is meant by the Messy Data? 1 / 1 point

- Duplicated or unnecessary data.
- Inconsistent text and typos.
- Missing data.
- (•) All of the above.

**Correct**
Correct. Duplicated or unnecessary data, inconsistent text and typos, and missing data are all examples of the messy data.

**9.** What is an outlier? 1 / 1 point

- (•) Outlier is an observation in dataset that is distant from most other observations.
- Outlier is a data point that does not belong in our dataset.
- Outlier is a data point that is very close to the mean value of all observations.
- Outlier is a data point that has the highest or lowest value in the dataset.

**Correct**
Correct. An outlier is an observation in data that is distinct from most other observations.

**10.** How do we identify outliers in our dataset? 1 / 1 point

- We can only identify outliers visually through building plots.
- (•) We can identify outliers both visually and with statistical calculations.
- We can only identify outliers by using some statistical calculations.
- We can identify outliers only by calculating the minimum and maximum values in the dataset.

**Correct**
Correct. We can use plots, such as histograms, density, and box plots, as well as making some statistical calculations, such as calculating the interquartile ranges.