Congratulations! You passed!
TO PASS 80% or higher

Keep Learning

GRADE
100%

Data Preparation

LATEST SUBMISSION GRADE
100%

1. Which of the following is NOT a data quality issue? **1 / 1 point**

- ○ Duplicate data
- ○ Inconsistent data
- ◉ Scaled data
- ○ Missing values

Correct
That's correct! This is a data wrangling technique.

2. Imputing missing data means to **1 / 1 point**

- ○ drop samples with missing values.
- ◉ replace missing values with something reasonable.
- ○ merge samples with missing values.
- ○ replace missing values with outliers.

Correct
That's correct!

3. A data sample with values that are considerably different than the rest of the other data samples in the dataset is called an/a _____. **1 / 1 point**

- ○ Noise
- ○ Inconsistent data
- ○ Invalid data
- ◉ Outlier

Correct
That's correct!

4. Which one of the following examples illustrates the use of domain knowledge to address a data quality issue? **1 / 1 point**

- ○ None of these
- ○ Drop samples with missing values
- ○ Simply discard the samples that lie significantly outside the distribution of your data
- ◉ Merge duplicate records while retaining relevant data

Correct
That's correct! This requires some logic to resolve conflicting values.

5. Which of the following is NOT an example of feature selection? **1 / 1 point**

- ○ Adding an in-state feature based on an applicant's home state.
- ○ Removing a feature with a lot of missing values.
- ◉ Replacing a missing value with the variable mean.
- ○ Re-formatting an address field into separate street address, city, state, and zip code fields.

Correct
That's correct!

6. Which one of the following is the best feature set for your analysis? **1 / 1 point**

- Feature set with the smallest number of features
- Feature set that contains exclusively re-coded features
- Feature set with the largest number of features
- (selected) Feature set with the smallest set of features that best capture the characteristics of the data for the intended application

✓ **Correct**
That's correct!

7. The mean value and the standard deviation of a zero-normalized feature are **1 / 1 point**

- mean = 1 and standard deviation = 1
- mean = 0 and standard deviation = 0
- (selected) mean = 0 and standard deviation = 1
- mean = 1 and standard deviation = 0

✓ **Correct**
That's correct!

8. Which of the following is NOT true about PCA? **1 / 1 point**

- PC1 and PC2, the first and second principal components, respectively, are always orthogonal to each other.
- (selected) PCA is a dimensionality reduction technique that removes a feature that is very correlated with another feature.
- PCA stands for principal component analysis
- PC1, the first principal component , captures the largest amount of variance in the data along a single dimension.

✓ **Correct**
That's correct!