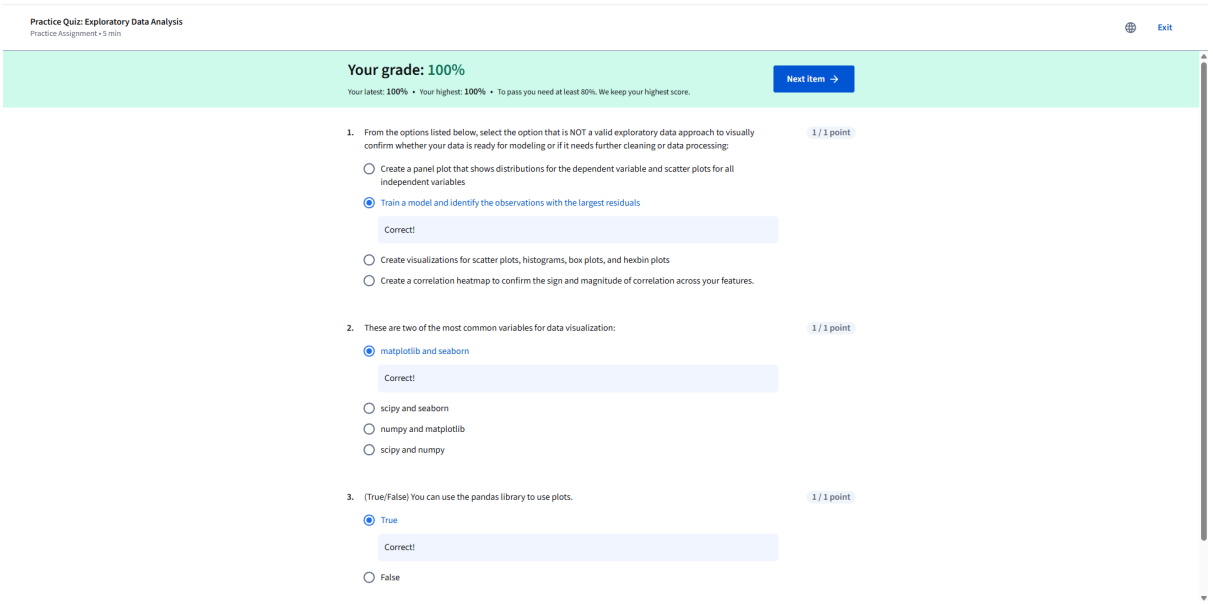
Practice Quiz: Exploratory Data Analysis



Practice Quiz: Exploratory Data Analysis
Practice Assignment • 5 min

Exit

Your grade: 100%

Your latest: 100% • Your highest: 100% • To pass you need at least 80%. We keep your highest score.

Next item →

1. From the options listed below, select the option that is NOT a valid exploratory data approach to visually confirm whether your data is ready for modeling or if it needs further cleaning or data processing: 1 / 1 point

- Create a panel plot that shows distributions for the dependent variable and scatter plots for all independent variables
- Train a model and identify the observations with the largest residuals

Correct!

- Create visualizations for scatter plots, histograms, box plots, and hexbin plots
- Create a correlation heatmap to confirm the sign and magnitude of correlation across your features.

2. These are two of the most common variables for data visualization: 1 / 1 point

- matplotlib and seaborn

Correct!

- scipy and seaborn
- numpy and matplotlib
- scipy and numpy

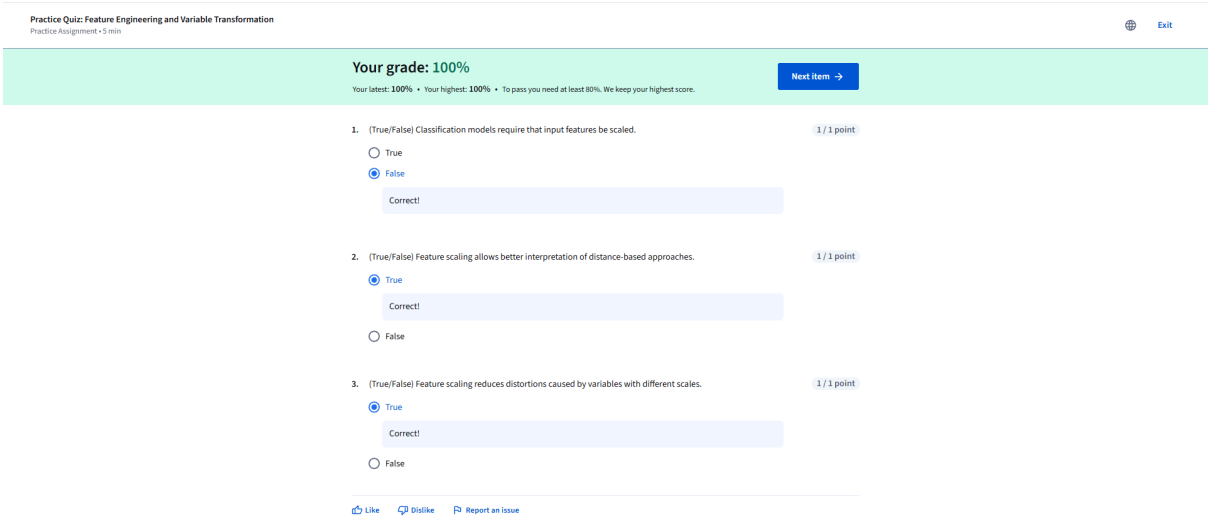
3. (True/False) You can use the pandas library to use plots. 1 / 1 point

- True

Correct!

- False

Practice Quiz: Feature Engineering and Variable Transformation



Graded Quiz: Module 3 - Exploratory Data Analysis and Feature Engineering

Your grade: 100%

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

Next item →

1. Which scaling approach converts features to standard normal variables? 1 / 1 point

- ○ Robust scaling
- ○ MinMax scaling
- ◉ Standard scaling
- ○ Nearest neighbor scaling

Correct
Correct. Standard scaling converts variables to standard normal variables.

2. Which variable transformation should you use for ordinal data? 1 / 1 point

- ○ Standard scaling
- ◉ Ordinal encoding
- ○ One-hot encoding
- ○ Min-max scaling

Correct
Correct. Use ordinal encoding if there is some order to the categorical features.

3. What are polynomial features? 1 / 1 point

- ○ They are logistic regression coefficients.
- ○ They are lower order relationships in the data.
- ○ They are represented by linear relationships in the data.
- ◉ They are higher order relationships in the data.

Correct
Correct. Polynomial features are estimated by higher order polynomials in a linear model, like squared, cubed, etc.

4. What does Boxcox transformation do? 1 / 1 point

- ◉ It transforms the data distribution into more symmetrical bell curve
- ○ It transforms categorical variables into numerical variables.
- ○ It makes the data more right skewed.
- ○ It makes the data more left skewed

Correct
Correct. Boxcox is one of the ways we can transform our skewed dataset to be more normally distributed.

5. Select three important reasons why EDA is useful. 1 / 1 point

- ○ To analyze data sets, to determine the main characteristics of data sets, and to use sampling to examine data
- ○ To examine correlations, to sample from dataframes, and to train models on random samples of data
- ○ To utilize summary statistics, to create visualizations, and to identify outliers
- ◉ To determine if the data makes sense, to determine whether further data cleaning is needed, and to help identify patterns and trends in the data

Correct
Correct. EDA helps us analyze data to summarize its main characteristics.

6. What assumption does the linear regression model make about data? 1 / 1 point

- ◉ This model assumes a linear relationship between predictor variables and outcome variables.
- ○ This model assumes an addition of each one of the model parameters multiplied by a coefficient.
- ○ This model assumes a transformation of each parameter to a linear relationship.
- ○ This model assumes that raw data in data sets is on the same scale.

Correct
Correct. The linear regression model assumes a linear relationship between predictor and outcome variables.

7. What is skewed data? 1 / 1 point

- ○ Data that has a normal distribution.
- ○ Raw data that may not have a linear relationship.
- ◉ Data that is distorted away from normal distribution; may be positively or negatively skewed.
- ○ Raw data that has undergone log transformation.

Correct
Correct. Often raw data, both the features and the outcome variable, can be negatively or positively skewed.

8. Select the two primary types of categorical feature encoding. 1 / 1 point

- ○ Log and polynomial transformation
- ○ Encoding and scaling
- ○ Frequency encoding and label encoding
- ◉ One-hot encoding and ordinal encoding

Correct
Correct. Encoding that transforms non-numeric values to numeric values is often applied to categorical features.

9. Which scaling approach puts values between zero and one? 1 / 1 point

- ○ Standard scaling
- ◉ Min-max scaling
- ○ Nearest neighbor scaling
- ○ Robust scaling

Correct
Correct. Min-max scaling converts variables to continuous variables in the (0, 1) interval by mapping minimum values to 0 and maximum values to 1.

10. Which variable transformation should you use for nominal data with multiple different values within the feature? 1 / 1 point

- ○ Ordinal encoding
- ○ Standard scaling
- ◉ One-hot encoding
- ○ Min-max scaling

Correct
Correct. Use one-hot encoding if there are multiple different values within a feature.