Congratulations! You passed!

Grade received 100%

Latest Submission Grade 100%

To pass 80% or higher

Go to next item Retake the assignment in **7h 58m**

1. Why perform EDA? **1 / 1 point**

- [ ] Necessary step before ML Modeling
- [ ] Use visual techniques to explore data
- [x] Find hidden insights

Correct
EDA or Exploratory Data Analysis is a necessary step in the lifecycle of a data science problem.

- [ ] Use statistical techniques to explore data

2. What is the purpose of a histogram? **1 / 1 point**

- [x] Summarize the distribution of a dataset

Correct
Histograms impart essential information about the distribution of a dataset, such as in the gaussian distribution, where it creates a bell curve.

- [ ] Visually inspect a dataset
- [ ] Part of an EDA process
- [ ] Find the median of a dataset

3. Why perform clustering on a dataset? **1 / 1 point**

- [x] Discover hidden patterns

Correct
Clustering is an unsupervised machine-learning technique to find hidden insights in a dataset

- [ ] To create a visualization of relationships in the data
- [ ] Create new labels
- [ ] A business wants to identify customer segments

4. What is Feature Engineering? **1 / 1 point**

- [ ] Determining the importance of features on model accuracy
- [ ] A feedback loop
- [ ] Creating features
- [x] Using domain knowledge as part of data science workflow

Correct
Feature engineering allows domain experts to find useful features from raw data to create a machine-learning model.

5. Why scale data? **1 / 1 point**

- [ ] Improves accuracy of machine-learning model
- [ ] Important pre-processing step in machine-learning
- [ ] Can help with outliers depending on the scaling technique
- [x] Most machine learning algorithms require data to be scaled

Correct
Machine learning requires data to be in a numerical format and, most of the time, in a scaled format.