

3.5.2 Preparing and cleaning data for Analysis Quiz

Question 1

What term is used for the process of conditioning data into a form that can be analyzed?



transforming

extracting

loading

processing

Question 2

What task is performed as part of the loading process in data preparation?



transferring the data into a database

providing values for missing data

collecting and recording the data

locating and removing duplicate data

Question 3

Which two tasks are part of the transforming data process? (Choose two.)

collecting data required to perform the analysis



reviewing the data and addressing erroneous or missing values



eliminating the data that is not relevant to the analysis question to be answered

creating visual representations of the data

presenting the knowledge gained from the data

Question 4

In a data analysis project, which process gathers all grades for students enrolled in a college course?

action



extract

load

transform

Question 5

What is a characteristic of open data?

data that lacks intellectual property restrictions



data that lacks predefined organization

data that does not need to be stored

data that does not generate new knowledge

Open data is not protected by intellectual property restrictions and can be used and redistributed without legal, technical, or social restrictions.

Question 6

An analyst is reviewing gathered data from various sources and modifying incompatible metrics into a usable form. In which ETL step would this task be completed in?

action

extract

load



transform

Question 7

What are Web scraping tools used for in big data analytics?

to automatically extract data from HTML pages

to provide standardized interfaces for data collection

to store real-time data in its original format



to create historical record of everything and anything that happens within a system, including events such as transactions, errors, or intrusions

Web scraping tools are used to automatically extract data from HTML pages. Typically, web scraping is an automated process which uses a bot or web crawler to find and obtain data. Specific data is gathered and copied from the web to a database or spreadsheet.

Question 8

Which universal file format is organized in rows and has the columns separated by a delimiter character?

HTML



CSV

JSON

XML

Question 9

Which two methods are used to collect unstructured data for analysis? (Choose two.)

querying a relational database



using a web-scraping tool

copying a spreadsheet



obtaining data through an API

downloading a CSV file

Question 10

What are two important considerations when selecting data for an analysis project? (Choose two.)



the data is relevant to the original business question

the data must be free and unlicensed



the data is current and updated frequently

the data can be downloaded from the internet

the data is not being streamed live