

Which of the following approach should be used to ask Data Analysis question?

- a) Find only one solution for particular problem
- b) Find out the question which is to be answered
- c) Find out answer from dataset without asking question
- d) None of the mentioned

View Answer

Answer: b

Explanation: Data analysis has multiple facets and approaches.

Which of the following is key characteristic of hacker?

- a) Afraid to say they don't know the answer
- b) Willing to find answers on their own
- c) Not willing to find answers on their own
- d) All of the mentioned

View Answer

Answer: b

Explanation: Hacker is an expert at programming and solving problems with a computer.

In a statistical research, correlations are published as below. One or more of these correlations is/are invalid. Which one is it?

- a) -0.395
- b) 1.827
- c) 0.738
- d) -1.03
- e) 0

Choose the correct interpretation of these findings.

Answer: Both b & d

In scientific study on wild fires in the forest, correlation with three different (independent) variables were reported as below

X1 : + 0.345

X2 : + 0.689

X3 : - 0.701

Which variable(s) shows strong correlation with wild fires in forest?

X3

What is the difference between confidence interval and standard hypothesis testing?

Confidence interval is not a test but usually more informative than hypothesis testing.

What makes confidence intervals more informative than the results of standard hypothesis testing?

replication

Hold out validation

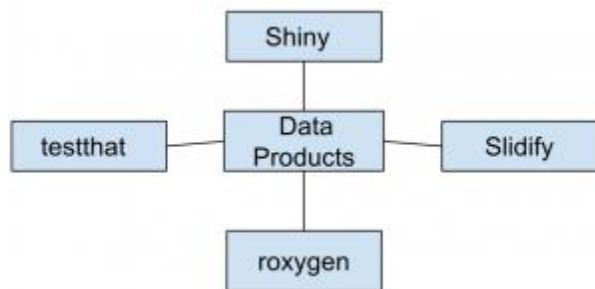
Cross validation

Answer: (all of the above)

Which is used to query and get the data from external sources?

SQL

Data Products shown in the below figure is built using which programming language ?



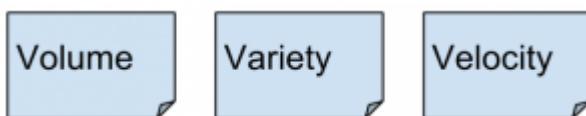
- a) S
- b) Python
- c) R
- d) Java

[View Answer](#)

Answer:c

Explanation:Products mentioned in the figure are web application frameworks written in R.

Which of the following term is appropriate to the below figure ?



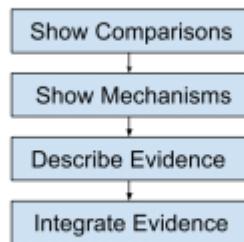
- a) Large Data
- b) Big Data
- c) Dark Data
- d) None of the mentioned

[View Answer](#)

Answer:b

Explanation:Big data is a broad term for data sets so large or complex that traditional data processing applications are inadequate

Which of the following principle is incorrectly represented in the below figure?



a) Show Comparisons

b) Integrate Evidence

c) Describe Evidence

d) None of the Mentioned

View Answer

Answer:d

Which graph type would you use to visualize the shape of a distribution?

What are/is the concerns when using Open Data?

What is used to import structured web data?

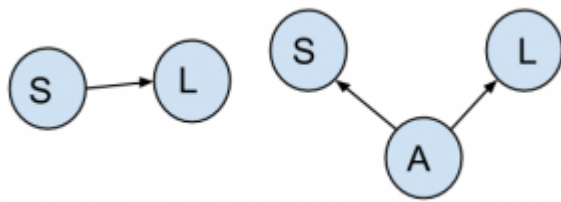
APIs

Which programming platform is popular among Data Scientists?

R & Python

The Data Science Venn Diagram is composed of Hacking Skills, Math or Statistics Skill and --
Domain Expertise

Which of the following design term is perfectly applicable to the below figure ?



- a) Correlation
- b) Confounding
- c) Causation
- d) None of the mentioned

View Answer

Answer:b

Explanation:Confounding can be dealt with either at the study design stage, or at the analysis stage.

Which below sentences applies to the Data Science Projects well?

Data Science projects consists of multiple roles from business person to data gathering, machine learning and visual analyst.

Choose the correct statement about various types of Business Analytics.

Prescriptive Analytics is the highest form of analytics, which takes inputs from other analytics models.

Which data type(s) has True zero.

Ratio Only

Which sampling method is best suited when there are specific sub-groups to investigate?

Stratified sampling

Which test can be performance on two or more samples of quantitative data ?

One-way ANOVA

In statistical Hypothesis testing, what is p-value?

The p-value or probability value is the probability for a given statistical model that, when the null hypothesis is true

What kind of machine learning algorithm used for house price estimation?

Supervised Machine Learning

What is percentage work is machine learning modelling and evaluation in entire data science project?

20% approx..

What process allows you to determine that if a particular event occurs another event will likely happen next, in a time-ordered data?

Sequence mining

The most common statistical method, and the method used in market-pricing analysis, of the options shown is _____

Regression

The problem of having a large number of dimensions in which variables are associated with each other is _____

overfitting

What is critical thinking?

About finding critical question

Which of the following is the probability calculus of beliefs, given that beliefs follow certain rules ?

- a) Bayesian probability
- b) Frequency probability
- c) Frequency inference
- d) Bayesian inference

View Answer

Answer:a

Explanation:Data scientists tend to fall within shades of gray of these and various other schools of inference.

What is reasoning?

Your belief, experience, evidence etc., that support conclusion.

Which Industry is Data Science most applicable?

Most of the Industries, including above all.

Which of the following can be considered as random variable ?

- a) The outcome from the roll of a die
- b) The outcome of flip of a coin
- c) The outcome of exam
- d) All of the Mentioned

View Answer

Answer:d

Explanation:The probability distribution of a discrete random variable is a list of probabilities associated with each of its possible values.

What is Tableau popular for?

Visualization

Which of the following below is/are an example(s) of empirical approach?

- a) Following an exercise routine.
- b) Learning a receipe
- c) Trying on new clothes
- d) Following a map to reach distination

Trying on new clothes

What one is the best example of structured data?

Monthly sales data

Which of the following is also referred to as random variable ?

- a) stochast
- b) aleatory
- c) eliette
- d) All of the Mentioned

View Answer

Answer:b

Explanation:Random variable is also known as stochastic variable.

What is a popular application of predictive analytics ?

To speculate the stock prices in the financial markets.

Choose the correct statement.

Raw data is original source of data

Which of the following is performed by Data Scientist?

Option (1) Define the question

Option (2) Create reproducible code

Option (3) Challenge results

Option (4) All of the above.

Answer : All of the above

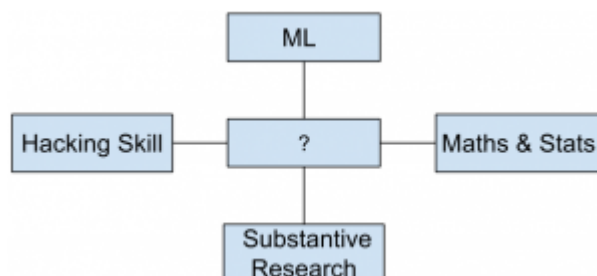
Which of the following technique is also referred to as Bagging ?

Bootstrap aggregating

Which of the following command allows you to update the repository in Git?

push

Which of the following would be more appropriate to be replaced with question mark in the following figure?



a) Data Analysis

b) Data Science

c) Descriptive Analytics

d) None of the mentioned

View Answer

Answer:b

Point out the wrong statement:

- a) Merging concerns combining datasets on the same observations to produce a result with more variables
- b) Data visualization is the organization of information according to preset specifications
- c) Subsetting can be used to select and exclude variables and observations
- d) All of the Mentioned

View Answer

Answer:b

Explanation:Data formatting is the organization of information according to preset specifications.
