Data Science MCQ PDF

Data science MCQs test quick recall of Python, statistics and machine learning ideas, such as the difference between precision and recall, what overfitting means and which plot suits a distribution. Use them for exam practice, then read the explanation for every wrong answer to fix the underlying concept.

Key takeaways

- MCQs test recall and recognition, so pair them with coding practice.
- Read each explanation, including for questions you answered correctly.
- Group mistakes by topic and revise the weakest topic first.
- Timed practice builds exam speed.
- Distractors usually reflect real misunderstandings, such as confusing correlation with causation.

What is data science multiple choice questions?

Data science MCQs are multiple-choice questions that check knowledge of programming, statistics, data preparation and machine learning concepts, usually with one correct option and distractors based on common mistakes.

How do I use MCQs well?

Attempt a set without notes, mark the ones you guessed, then read the answers. For each miss, write the idea in one line. Repeat the set after a few days. Practice with explanations teaches more than a score.

Statistics practice

Q1. Which measure is least affected by outliers? A) Mean B) Median C) Range D) Variance. **Answer: B.** The median depends on order, not on extreme values.

Q2. A correlation of -0.9 means: A) No relation B) Weak positive C) Strong negative linear relation D) Causation. **Answer: C.** The sign gives direction and the size gives strength; it does not prove cause.

Q3. A small p-value suggests: A) The null hypothesis is certainly false B) Data would be unlikely if the null were true C) The effect is large D) The sample is biased. **Answer: B.** It is about the data given the null, and it says nothing about effect size.

Python and pandas practice

Q4. Which call shows missing values per column? A) df.isna().sum() B) df.null() C) df.count_na() D) df.missing(). **Answer: A.**

Q5. In pandas, which joins keep every row from the left table? A) inner B) left C) outer only D) cross. **Answer: B.**

Q6. What does df.groupby("city")["sales"].mean() return? A) One row per row B) Average sales per city C) Total sales D) Number of cities. **Answer: B.**

Machine learning practice

Q7. Overfitting means the model: A) Performs well on training data but poorly on new data B) Performs poorly everywhere C) Is too simple D) Has no features. **Answer: A.**

Q8. For a rare disease screening test, which metric is most important to watch? A) Accuracy only B) Recall C) Mean error D) Variance. **Answer: B.** Missing real cases is costly, though precision should be checked too.

Q9. Which algorithm groups unlabeled data? A) Logistic regression B) K-means C) Linear regression D) Naive Bayes. **Answer: B.**

Q10. Why scale features for k-nearest neighbours? A) To add rows B) Because distance depends on feature size C) To remove labels D) To increase accuracy always. **Answer: B.**

Evaluation and leakage practice

Q11. Data leakage happens when: A) Test information influences training B) The data file is large C) The model is saved D) A chart is wrong. **Answer: A.** Fitting a scaler on all data before splitting is a common cause.

Q12. Cross-validation is used to: A) Remove outliers B) Estimate performance more reliably C) Create new labels D) Speed up training. **Answer: B.**

How do I review my results?

Count correct answers by topic. Anything below your target means returning to notes and coding examples. Create ten of your own questions to teach the idea back.

What question styles appear in written tests?

Written tests mix direct definition questions, scenario questions and code-reading questions. A scenario question may describe a model with high training accuracy and low test accuracy and ask for the likely cause. A code-reading question may show three lines of pandas and ask what the output is. To prepare, practise saying why each wrong option is wrong, since that exposes shallow understanding. Also practise reading the question twice, because words like "not", "least" and "except" change the answer entirely.

How can I make my own practice questions?

After each topic, write five questions with four options each, making the wrong options from real mistakes you or classmates have made. Swap sets with a friend and explain each answer aloud. This method turns passive reading into active recall and gives you a personal bank that targets your weak areas. Keep a running list of mistakes, review it weekly and retire questions you answer correctly three times in a row so that your practice time stays focused where it matters.

How do I manage time in a timed test?

Do a first pass answering only the questions you are sure about, mark doubtful ones and return to them. Do not spend more than a minute on a single item in the first pass. In the second pass eliminate clearly wrong options before choosing. Leave a few minutes at the end to check marked answers and any rule about negative marking in your exam.

Which resources help beyond MCQs?

Work through the official scikit-learn and pandas tutorials, solve small coding exercises and explain concepts aloud. Interviewers often follow a multiple-choice style fact with a "why" question, so understanding the reason behind each answer is more valuable than memorising options. Combine quizzes with a project so ideas connect to real data.

Frequently asked questions

Are MCQs enough to prepare for data science interviews?

No. MCQs help recall, but interviews also test coding, projects and explanation, so combine quizzes with hands-on practice.

How should I review wrong answers?

Write the correct idea in one line, find the matching topic in your notes and retry a similar question after a few days.

What topics do data science MCQs cover?

They cover Python, pandas, statistics, probability, machine learning algorithms, evaluation metrics, overfitting and data leakage.

How many questions should I practise daily?

Ten to twenty questions with careful review is usually more useful than a long set done quickly without reading explanations.

Do these questions match a specific exam?

No. They are general practice questions, so check your exam syllabus and past papers for its exact format and topics.

Sources and further reading

- Python documentation
- pandas documentation
- scikit-learn user guide

About this PDF

These PDFs are study material for learning and revision. They are not a certificate, a degree or a promise of a job, and technology changes, so check official documentation for the latest details. Spot an error? Write to info@uncodemy.com and the trainer team will review it.

More free resources and courses: uncodemy.com | info@uncodemy.com | +91 98183 66550