

ANALYTICS / DATA SCIENCE 201 (ADY201M) COURSE 1 AND 2 PRACTICE TEST (SAMPLE) STUDY GUIDE



SAMPLE STUDY GUIDE WITH LIMITED QUESTIONS

VISIT US ONLINE FOR
THE FULL VERSION STUDY GUIDE

<https://ady201mcourse1and2.examzify.com>



Copyright © 2026 by Examzify - A Kaluba Technologies Inc. product.

ALL RIGHTS RESERVED.

No part of this book may be reproduced or transferred in any form or by any means, graphic, electronic, or mechanical, including photocopying, recording, web distribution, taping, or by any information storage retrieval system, without the written permission of the author.

Notice: Examzify makes every reasonable effort to obtain accurate, complete, and timely information about this product from reliable sources.

SAMPLE

Table of Contents

Copyright	1
Table of Contents	2
Introduction	3
How to Use This Guide	4
Questions	5
Answers	8
Explanations	10
Next Steps	16

SAMPLE

Introduction

Preparing for a certification exam can feel overwhelming, but with the right tools, it becomes an opportunity to build confidence, sharpen your skills, and move one step closer to your goals. At Examzify, we believe that effective exam preparation isn't just about memorization, it's about understanding the material, identifying knowledge gaps, and building the test-taking strategies that lead to success.

This guide was designed to help you do exactly that.

Whether you're preparing for a licensing exam, professional certification, or entry-level qualification, this book offers structured practice to reinforce key concepts. You'll find a wide range of multiple-choice questions, each followed by clear explanations to help you understand not just the right answer, but why it's correct.

The content in this guide is based on real-world exam objectives and aligned with the types of questions and topics commonly found on official tests. It's ideal for learners who want to:

- Practice answering questions under realistic conditions,
- Improve accuracy and speed,
- Review explanations to strengthen weak areas, and
- Approach the exam with greater confidence.

We recommend using this book not as a stand-alone study tool, but alongside other resources like flashcards, textbooks, or hands-on training. For best results, we recommend working through each question, reflecting on the explanation provided, and revisiting the topics that challenge you most.

Remember: successful test preparation isn't about getting every question right the first time, it's about learning from your mistakes and improving over time. Stay focused, trust the process, and know that every page you turn brings you closer to success.

Let's begin.

How to Use This Guide

This guide is designed to help you study more effectively and approach your exam with confidence. Whether you're reviewing for the first time or doing a final refresh, here's how to get the most out of your Examzify study guide:

1. Start with a Diagnostic Review

Skim through the questions to get a sense of what you know and what you need to focus on. Your goal is to identify knowledge gaps early.

2. Study in Short, Focused Sessions

Break your study time into manageable blocks (e.g. 30 – 45 minutes). Review a handful of questions, reflect on the explanations.

3. Learn from the Explanations

After answering a question, always read the explanation, even if you got it right. It reinforces key points, corrects misunderstandings, and teaches subtle distinctions between similar answers.

4. Track Your Progress

Use bookmarks or notes (if reading digitally) to mark difficult questions. Revisit these regularly and track improvements over time.

5. Simulate the Real Exam

Once you're comfortable, try taking a full set of questions without pausing. Set a timer and simulate test-day conditions to build confidence and time management skills.

6. Repeat and Review

Don't just study once, repetition builds retention. Re-attempt questions after a few days and revisit explanations to reinforce learning. Pair this guide with other Examzify tools like flashcards, and digital practice tests to strengthen your preparation across formats.

There's no single right way to study, but consistent, thoughtful effort always wins. Use this guide flexibly, adapt the tips above to fit your pace and learning style. You've got this!

Questions

SAMPLE

- 1. Which of these are the common elements associated with Big Data?**
 - A. Velocity, Volume, Variety, Veracity, Value
 - B. Speed, Size, Sameness, Security, Satisfaction
 - C. Frequency, Footage, Features, Factuality, Fun
 - D. Efficiency, Effectiveness, Emotion, Engagement, Economics
- 2. Which approach is used to process data in parallel for efficient analysis?**
 - A. Batch Processing
 - B. Stream Processing
 - C. Distributed Computing
 - D. Centralized Processing
- 3. What is a key consideration when setting goals for data mining?**
 - A. The frequency of data collection
 - B. The number of visualization techniques
 - C. The level of accuracy expected from the results
 - D. The number of attributes needed
- 4. In the context of data analysis, what is data cleaning?**
 - A. The process of categorizing data
 - B. The method of eliminating irrelevant data
 - C. The technique used for fixing or removing incorrect data
 - D. The practice of archiving old data
- 5. The term "data repositories" exclusively refers to RDBMs and NoSQL databases that are used to collect, organize, and isolate data for analytics.**
 - A. True
 - B. False
- 6. What should be your initial step to harness the power of data science in a small manufacturing business?**
 - A. Begin hiring a team of data scientists.
 - B. Invest in expensive data analytics software.
 - C. Increase your advertising budget to attract more customers.
 - D. Document your existing data collection and archiving practices.

- 7. What is the main purpose of data modeling in the data science methodology?**
- A. To develop models for descriptive or predictive purposes
 - B. To select an appropriate analytical approach
 - C. To refine and adjust the problem statement
 - D. To collect and compile raw data
- 8. Which aspect is important when ensuring a data science model's findings are actionable?**
- A. The number of data points collected.
 - B. Stakeholder familiarity with digital tools.
 - C. The clarity of problem definition.
 - D. The analytical method chosen by data scientists.
- 9. Which deep learning model would you choose as the foundational approach for generating new instances of data resembling your original dataset's patterns?**
- A. Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs)
 - B. Neural Networks
 - C. Linear Regression
 - D. Decision Trees
- 10. Why is there a growing demand for data scientists and analytics professionals in various industries?**
- A. Because data scientists are primarily involved in managing customer analytics initiatives
 - B. Because data scientists are mainly responsible for convincing C-suite executives about the benefits of data and analytics
 - C. Because of the digital revolution and the need to analyze big data for effective decision-making
 - D. Because data scientists are primarily focused on conducting biological experiments

Answers

SAMPLE

1. A
2. C
3. C
4. C
5. B
6. D
7. A
8. C
9. A
10. C

SAMPLE

Explanations

SAMPLE

1. Which of these are the common elements associated with Big Data?

- A. Velocity, Volume, Variety, Veracity, Value
- B. Speed, Size, Sameness, Security, Satisfaction
- C. Frequency, Footage, Features, Factuality, Fun
- D. Efficiency, Effectiveness, Emotion, Engagement, Economics

The common elements associated with Big Data are encapsulated in the concepts of Velocity, Volume, Variety, Veracity, and Value. Velocity refers to the speed at which data is generated and processed. This aspect emphasizes the importance of real-time or near real-time data processing capabilities that modern technologies provide, which is essential for timely decision-making in many applications. Volume highlights the sheer amount of data that organizations must manage. With the proliferation of digital technologies and the Internet, data is being generated at an unprecedented scale, and organizations must have strategies to handle this massive volume. Variety signifies the diverse types of data that exist, which can include structured, semi-structured, and unstructured data. This variety necessitates the use of advanced analytical techniques to extract meaningful insights from different data sources. Veracity points to the quality and reliability of the data. In the context of Big Data, ensuring high levels of data integrity and accuracy is crucial for making trustworthy decisions. Finally, Value reflects the significance of deriving actionable insights from the data. The ultimate goal of Big Data initiatives is not just to analyze information but to gain meaningful value that can drive business decisions and innovation. In contrast, the other answer choices present elements that lack direct correlation with the defining characteristics of Big

2. Which approach is used to process data in parallel for efficient analysis?

- A. Batch Processing
- B. Stream Processing
- C. Distributed Computing
- D. Centralized Processing

Distributed Computing is the correct approach for processing data in parallel to enable efficient analysis. This method involves splitting a large dataset across multiple machines or nodes within a network, allowing them to process different parts of the data simultaneously. By leveraging parallel processing, distributed computing can significantly reduce the time required to analyze large datasets, making it particularly effective for big data applications. This approach supports scalability, as more computing resources can be added as needed to handle increasing volumes of data. It also enhances fault tolerance since if one node fails, others can continue processing, ensuring that the overall system remains operational. In contrast, batch processing involves processing a large volume of data in predefined batches, often resulting in longer wait times before analysis can be completed. Stream processing is focused on real-time data processing and is more suited for continuous data flows rather than parallel analysis of large datasets. Centralized processing relies on a single location or system to handle all data processing tasks, which can create bottlenecks and inefficiencies if the data is extensive. Thus, distributed computing stands out for its ability to efficiently manage and analyze large datasets through parallel processing.

3. What is a key consideration when setting goals for data mining?

- A. The frequency of data collection
- B. The number of visualization techniques
- C. The level of accuracy expected from the results**
- D. The number of attributes needed

Setting goals for data mining is a critical step in ensuring that the analysis is both meaningful and effective. One of the most important considerations in this process is the level of accuracy expected from the results. Accuracy directly influences the reliability of insights drawn from the data and determines how well the mined data meets the objectives set for the project. High accuracy in results ensures that decision-makers can trust the findings and take informed actions based on those insights. When establishing goals, it's vital to align the expected accuracy with the overall purpose of the data mining endeavor. For example, if a project aims to predict customer behavior, the required accuracy might be high to ensure that decisions based on those predictions lead to successful outcomes. Conversely, if the analysis is exploratory in nature, a lower level of accuracy might suffice, but still, a clear understanding of the expectation remains essential. Focusing on the level of accuracy helps guide the selection of methods, tools, and metrics used throughout the data mining process, leading to more effective and actionable results.

4. In the context of data analysis, what is data cleaning?

- A. The process of categorizing data
- B. The method of eliminating irrelevant data
- C. The technique used for fixing or removing incorrect data**
- D. The practice of archiving old data

Data cleaning refers to the technique used for fixing or removing incorrect, corrupted, or irrelevant data from a dataset. This process is crucial because the accuracy and quality of data significantly impact the integrity of analysis and modeling results. Poor data quality can lead to erroneous conclusions and decisions, making data cleaning a vital step in any data analysis workflow. Through data cleaning, analysts ensure that the data is consistent, complete, and accurate. This might involve correcting typos, filling in missing values, standardizing formats, or removing duplicate entries. By addressing these issues, data cleaning helps maintain the reliability of datasets and supports producing trustworthy insights. The other choices relate to different aspects of data management but do not accurately capture the essence of data cleaning. For instance, categorizing data pertains to organizing data based on specific attributes rather than correcting it. Eliminating irrelevant data addresses a different concern—filtering out data that does not contribute to the analysis but does not encompass the broader scope of correcting data quality issues. Lastly, archiving old data refers to storing data for future access and does not relate to the correction or enhancement of active datasets.

5. The term "data repositories" exclusively refers to RDBMs and NoSQL databases that are used to collect, organize, and isolate data for analytics.

A. True

B. False

The statement that "data repositories" exclusively refers to RDBMs and NoSQL databases is inaccurate because the term encompasses a broader range of storage solutions. While RDBMS (Relational Database Management Systems) and NoSQL databases are indeed important types of data repositories, they are not the only ones. Data repositories can also include data lakes, data warehouses, cloud storage solutions, flat files, and other formats where data is stored and managed for analytical purposes. Each of these options serves different use cases and can work individually or in conjunction with one another. For example, data lakes are often used for storing unstructured data at scale, which allows for flexible analytics and big data processing. Furthermore, other platforms such as document stores, graph databases, and even simple spreadsheets can be viewed as data repositories under the right context. By recognizing that the term "data repositories" includes a variety of technologies and methods for storing data, it becomes clear that the initial statement is misleading, validating the assertion that the statement is false.

6. What should be your initial step to harness the power of data science in a small manufacturing business?

A. Begin hiring a team of data scientists.

B. Invest in expensive data analytics software.

C. Increase your advertising budget to attract more customers.

D. Document your existing data collection and archiving practices.

The initial step to harnessing the power of data science in a small manufacturing business is to document your existing data collection and archiving practices. This is crucial because understanding what data is currently available and how it is structured lays the foundation for any future data analysis efforts. By cataloging the existing data, a business can identify gaps, ensure data quality, and establish a clear framework for organizing data so that it can be effectively used later for analysis. This practice also helps in determining what additional data might be needed for meaningful insights and drives better decision-making processes. Without a comprehensive understanding of the existing data, any subsequent efforts to analyze or leverage data science could be misguided or ineffective. The other options generally represent more advanced steps that would occur after establishing a solid foundation of data practices. For instance, hiring a team of data scientists or investing in expensive software would be premature without having a clear understanding of the available data. Similarly, increasing an advertising budget without first analyzing the existing data would not provide the strategic insight needed.

7. What is the main purpose of data modeling in the data science methodology?

- A. To develop models for descriptive or predictive purposes**
- B. To select an appropriate analytical approach
- C. To refine and adjust the problem statement
- D. To collect and compile raw data

The main purpose of data modeling in the data science methodology is to develop models for descriptive or predictive purposes. Data modeling is a critical phase where the relationships among various data points are established and mathematical representations of the data are created. This allows data scientists to examine trends, make inferences, and predict future outcomes based on existing data. When models are created, they serve both descriptive and predictive functions. Descriptive models summarize past behaviors and relationships within the data, while predictive models forecast future events or trends based on historical data. This modeling process ultimately leads to insights that inform decision-making and strategic initiatives. While selecting an appropriate analytical approach, refining the problem statement, and collecting raw data are all important components of the data science workflow, they don't encapsulate the essence of data modeling's purpose. The focus of data modeling is directly tied to constructing a formal representation of data that supports analysis, which facilitates the generation of insights crucial to data science projects.

8. Which aspect is important when ensuring a data science model's findings are actionable?

- A. The number of data points collected.
- B. Stakeholder familiarity with digital tools.
- C. The clarity of problem definition.**
- D. The analytical method chosen by data scientists.

The clarity of problem definition is crucial when ensuring a data science model's findings are actionable because it sets the foundation for the entire analytical process. A well-defined problem helps in identifying the right questions to answer, determining the relevant data to collect, and selecting appropriate methodologies for analysis. When the problem is clearly articulated, it enables data scientists to align their work with the actual needs of the stakeholders, ensuring that the insights derived from the model can effectively inform decision-making and drive meaningful actions. If the problem definition is vague or poorly understood, it can lead to irrelevant findings, wasted resources, and a disconnect between the data insights and the stakeholders' requirements. Thus, emphasizing clarity in defining the problem helps in translating complex data findings into actionable recommendations, ultimately enhancing the impact of the data science efforts on business goals or operational strategies.

9. Which deep learning model would you choose as the foundational approach for generating new instances of data resembling your original dataset's patterns?

- A. Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs)**
- B. Neural Networks
- C. Linear Regression
- D. Decision Trees

Choosing Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) as the foundational approach for generating new instances of data is appropriate due to their design purpose and functionality. Both GANs and VAEs are specifically tailored for generating new samples that mimic the distribution and patterns of a given dataset. GANs operate through a dual-model architecture where the generator creates synthetic data, while the discriminator evaluates its authenticity, thereby improving the generator's output over iterations. This adversarial process allows GANs to produce highly realistic data that can reflect complex patterns inherent in the training set. On the other hand, VAEs use a probabilistic approach in combination with neural networks to encode data into a latent space and then decode it back to the original space. This process not only allows for data generation but also for exploring variations in the dataset while maintaining statistical coherence. In contrast, the other options are not designed for generating new instances of data in the way GANs and VAEs are. Neural Networks can be utilized for a variety of tasks, including classification and regression, but they do not inherently focus on data generation. Linear Regression is specifically aimed at modeling the relationship between dependent and independent variables and is not equipped for data generation. Decision Trees

10. Why is there a growing demand for data scientists and analytics professionals in various industries?

- A. Because data scientists are primarily involved in managing customer analytics initiatives
- B. Because data scientists are mainly responsible for convincing C-suite executives about the benefits of data and analytics
- C. Because of the digital revolution and the need to analyze big data for effective decision-making**
- D. Because data scientists are primarily focused on conducting biological experiments

The growing demand for data scientists and analytics professionals is significantly driven by the digital revolution and the increasing necessity to analyze vast amounts of data for informed decision-making. As businesses and organizations generate more data than ever before, they require skilled professionals who can extract meaningful insights from this data. These insights are crucial for making strategic decisions, optimizing operations, and enhancing customer experiences. The digital transformation across various industries has led to the creation of complex data ecosystems, which include not only structured data from traditional databases but also unstructured data from sources like social media, online interactions, and IoT devices. Data scientists are equipped with the analytical skills and tools needed to navigate this complexity and derive actionable intelligence from diverse data sets. Furthermore, organizations recognize that data-driven decision-making can significantly improve performance, increase competitive advantage, and foster innovation. This recognition spurs the need for data scientists who can utilize statistical methods, machine learning, and data visualization techniques to translate data into insights that drive business strategy. Other options do not encompass the broad scope of the demand for data scientists. While customer analytics and convincing executives may be part of a data scientist's role, they represent a narrower focus compared to the overarching influence of the digital revolution and the critical need for data analysis across all sectors

Next Steps

Congratulations on reaching the final section of this guide. You've taken a meaningful step toward passing your certification exam and advancing your career.

As you continue preparing, remember that consistent practice, review, and self-reflection are key to success. Make time to revisit difficult topics, simulate exam conditions, and track your progress along the way.

If you need help, have suggestions, or want to share feedback, we'd love to hear from you. Reach out to our team at hello@examzify.com.

Or visit your dedicated course page for more study tools and resources:

<https://ady201mcourse1and2.examzify.com>

We wish you the very best on your exam journey. You've got this!

SAMPLE