

Python 3 エンジニア認定デ ータ分析試験 学習用テキ スト

予備知識ゼロから合格ラインへ —— 読んで理解する1冊目

無料サンプル(第1章まで)

Pythonエンジニア育成推進協会「Python データ分析試験」出題範囲(主教材
『Pythonによるあたらしいデータ分析の教科書 第3版』)準拠
2026年10月版

ご利用にあたって

- このテキストは「IT資格道場」(www.shikaku-doujou.com)の購入者限定テキストのうち、学習用テキストの第1章までを収録した**無料サンプル**です。どなたでもご覧いただけます。
- 無料サンプルも著作物です。転載・改変しての再配布・二次販売はできません。
- AI モデルの学習データとしてのご利用(学習目的の入力・データセット化を含む)はできません。
- 製品版(学習用テキスト・頻出度別ノート・暗記用ノートの3点)は、ご購入者を識別できる透かし入りのPDFとして提供されます。
- 本PDFはスマートフォン・タブレットでの閲覧に合わせたA5版面・文字サイズで組版しています。

目次

この教材群の使い方(3点セット)

この試験が問うもの

第1章 データエンジニアの役割 [1.1]

役割とデータ分析の流れ [1.1]

ETLとデータ基盤とデータの種類 [1.1]

機械学習の種類とAIの関係 [1.1]

サンプルはここまでです

この教材群の使い方(3点セット)

種類	役割
① 学習用テキスト	これだけで問題が解けるようになる本編
② 頻出度別ノート	テキストを読んだら次はこれ。頻出順に絞った問題と解説で「解ける」に橋渡し。試験直前の最終チェックにも
③ 暗記用ノート	覚えるべき要点だけを一問一答に凝縮。空き時間の反復と用語の再確認用

使い方: ① 学習用を読む → ② 頻出度別で解き方を掴む → ③ 問題演習で腕試し → ④ 頻出度別に戻って再確認(試験直前もここ)。暗記用は空き時間や用語の再確認に使います。

この試験が問うもの

Python 3 エンジニア認定データ分析試験は、Python でデータを分析する基礎を問う試験で、出題範囲は主教材『Pythonによるあたらしいデータ分析の教科書 第3版』の内容です。章ごとの出題数が公式に決まっており、第4章のライブラリ(NumPy・pandas・Matplotlib・scikit-learn) が40問のうち27問を占め、残りは役割・Python の基礎・数学です。

項目	内容
運営	一般社団法人Pythonエンジニア育成推進協会
問題数	40問(すべて選択問題)
試験時間	60分
合格ライン	正答率70%(28問)
出題数の内訳	1章 2問・2章 5問・3章 6問・4章 27問
4章の内訳	NumPy 6問・pandas 7問・Matplotlib 6問・scikit-learn 8問
対象の版	Python 3.13・NumPy 2.2・pandas 2.2・Matplotlib 3.10・scikit-learn 1.6

このテキストの章は、公式の出題範囲の並びとは別に、学習しやすい順で並べています。第1章が1章、第2章が2章、第3章が3章に当たり、公式の4章は第4章から第7章の4つ(NumPy・pandas・Matplotlib・scikit-learn)に分けました。

例のコードは、手元の Python で一度実行して出力を確かめると記憶に残ります。

第1章 データエンジニアの役割 [1.1]

役割とデータ分析の流れ [1.1]

データ分析の現場には、役割の違う職種が三つあります。違いは「何を仕事の中心に置くか」で、次の表のとおりです：

役割	仕事の中心
データエンジニア	データを集めて整え、分析に使える状態で届ける基盤を作って運用する
データサイエンティスト	基盤のデータから統計や機械学習で知見やモデルを引き出す
データアナリスト	集計や可視化で、業務の判断を支える

複数の基幹システムに散らばった売上や在庫のデータを集め、蓄積する基盤を設計して運用するのは、データエンジニアです。システムの安全性を点検するセキュリティ監査担当者とは、仕事の中心が異なります。

分析の作業は、収集、前処理（加工）、可視化、モデル化の順に進みます。収集は元になるデータを集める工程です。前処理はデータクレンジングとも呼び、分析に使える形に整える工程です（欠損値を補う、「東京」と「東京都」のような表記ゆれをそろえる、など）。可視化は結果をグラフなどで見せる工程で、モデルの精度を測る評価は、モデル化の後に行います。

ETLとデータ基盤とデータの種類 [1.1]

データ基盤は、分析に使うデータを集めて蓄えておく仕組みの全体です。データを整理して置くデータウェアハウスと、元の形のまま置くデータレイクがあり

ます。基盤にデータを入れる処理は ETL と呼びます。三つの段階の頭文字を並べた名前で、段階は次のとおりです:

- Extract (抽出): 元のシステムからデータを取り出す
- Transform (変換): 形式や単位をそろえたり、集計したりして、分析しやすい形に加工する
- Load (書き出し): 加工後のデータを保存先に入れる

元のデータを加工するのは Transform の段階で、ETL はこの 3 つの段階の頭文字を並べた呼び名です。

データは形によっても分かります。行と列で表せて、列の意味が決まっているデータは構造化データです (CSV ファイル、データベースのテーブル、集計した表が当たります)。メールの本文のように、決まった形に収まらないデータは非構造化データです。問い合わせメールの本文は、行と列に収まらないので非構造化データです。

機械学習の種類とAIの関係 [1.1]

機械学習は、正解のラベルを使うかどうかで大きく分かります。

教師あり学習は、ラベル付きのデータで学習する方法です。予測する対象で、次の二つに分かれます:

- 分類: カテゴリを当てるタスクです。「迷惑メール」か「通常のメール」かの判定が当たる
- 回帰: 連続した数値を予測するタスクです。気温や曜日から来月の売上金額を当てる場合が当たる

売上金額のような数値の予測は、回帰で行います。

教師なし学習は、ラベルを使わずに、データの似ている度合いから構造を見つける方法です。似た買い方をする顧客をグループに分けるのがクラスタリングです(特徴量の数を減らして扱いやすくするのが次元削減です)。どちらも正解を与えずに構造を探る方法で、迷惑メールの判定のような分類や、金額の予測のような回帰は、正解を与える教師あり学習の仕事です。強化学習は、行動の結果として得る報酬を手がかりに、よい行動を学習する方法です。

人工知能(AI)、機械学習、ディープラーニングは入れ子の関係になっていて、広い順に、人工知能、機械学習、ディープラーニングです。機械学習は、データから規則を学ぶ人工知能の手法です(ディープラーニングは、多層のニューラルネットワークを使う機械学習の手法にあたります)。三つは、同じシステムの中で組み合わせて使うことも多い技術です。

サンプルはここまでです

続き(第2章～巻末)は、Python 3 エンジニア認定データ分析試験 問題集のご購入で、頻出度別ノート・暗記用ノートと合わせて3点すべてダウンロードできます。

SAMPLE

Python 3 エンジニア認定データ分析試験 学習用テキスト(無料サンプル)

版

2026年10月版(Pythonエンジニア育成推進協会「Python データ分析試験」出題範囲(主教材『Pythonによるあたらしいデータ分析の教科書 第3版』) 準拠)

発行

IT資格道場(<https://www.shikaku-doujou.com>)

お問い合わせ

info@shikaku-doujou.com

Python は Python Software Foundation の登録商標です。記載されているその他の会社名・製品名・サービス名は、各社の商標または登録商標です。

本教材は公式の出題範囲(主教材の章立て)と NumPy・pandas・Matplotlib・scikit-learn の対象の版の仕様を基に作成したもので、一般社団法人Pythonエンジニア育成推進協会との提携・承認関係はありません。

本文・図表・設問はすべてオリジナルの著作物です。無断での転載・複製・再配布・二次販売はできません。お気づきの点は info@shikaku-doujou.com までご連絡ください。