

PYTHON 3 DATA ANALYSIS EXAM

2026年10月版

Python 3 エンジニア認定 データ分析試験 想定問題集

全240問【解説つき】

6

問収録

無料サンプル版
製品版は全240問

SKILL BASE CAMP

本書は非公式の学習教材です。試験を実施する団体の公認・監修を受けたものではありません。
公的機関の公表資料および一般に公開された情報のみに基づき、著者が独自に作成しています。

本書について

本書は、Python 3 エンジニア認定データ分析試験の想定問題集（全240問）の無料サンプル版です。製品版の構成・難易度・解説の書き方を確認していただくため、6問を抜粋しています。

対象	Python 3 エンジニア認定データ分析試験の受験者
出題形式	選択式（CBT 方式。本試験は40問・60分）。本書は四肢択一で作成
収録数	6問（各章の問題のあとに解答・解説）
準拠	「Python 3 エンジニア認定データ分析試験 受験案内」（オデッセイ コミュニケーションズ、2026年10月25日確認）と、2026年1月5日以降の出題の改定のお知らせ（実施団体、2025年12月1日公表）

本試験の形式と時間配分

試験時間	60分
設問数	40問
1問あたり	約1分30秒
合格基準	7割正解（正答率70%）

1問あたり約1分30秒で、コードや表を読んで答える試験です。本書は本試験40問の6回分に当たる240問を収めています。解くときも、1問1分30秒を目安に時間を測ってみてください。

科目と配点

実施団体は、主教材の章・節ごとの出題数（40問中）を公表しています。本書は同じ比率で240問を配分しています。

データ分析エンジニアの役割 5% (2/40) 12問

Pythonと環境	12.5% (5/40)	30問
数学の基礎	15% (6/40)	36問
ライブラリによる分析実践	67.5% (27/40)	162問

ライブラリによる分析実践の内訳は、NumPy 36問・pandas 42問・Matplotlib 36問・scikit-learn 48問です。配点と科目別の合格最低基準は公表されていません。

解き方の目安

コードを読ませる設問は、実行して確かめる前に、各行が何を返すかを頭の中で追ってください。pandas や NumPy の設問では、配列の形 (shape)、軸 (axis)、戻り値の型を先に押さえると迷いにくくなります。数学の計算問題や、混同行列から指標を求める問題は、式を書き出してから計算します。

収録内容

サンプル問題	6問	6
--------	----	---

本書のご利用にあたって

本書をお読みになる前に、以下の事項をご確認ください。ご購入・ご利用をもって、これらに同意いただいたものとみなします。

- 1. 本書は非公式の学習教材です。** Python 3 エンジニア認定データ分析試験を実施する機関（一般社団法人Pythonエンジニア育成推進協会。試験の運営・申込はオデッセイ コミュニケーションズ）の公認・監修・推奨を受けたものではなく、同機関および関係する事業者と 本書の著者との間に、資本関係・提携関係・その他一切の関係はありません。
- 2. 本書の問題・解説は、すべて著者が独自に作成したものです。** 試験問題を再現・引用したものではありません。また、公式テキスト、講座資料、他社の教材から文章を転載した箇所也没有ありません。作成にあたっては、各ライブラリの公式ドキュメント、官公庁・大学等が公表している資料など、一般に公開されている情報のみを参照しています。
- 3. 本書は試験に出題される内容を予測・保証するものではありません。** 収録した問題は、公開されている出題範囲をもとに著者が独自に構成した想定問題であり、実際の出題との一致を意図したものでも、保証するものでもありません。本書を学習しても合格を保証するものではありません。
- 4. 内容は制作時点の情報に基づきます。** Python や各ライブラリは更新され、試験の出題範囲も改定されます。本書は Python 3.13・NumPy 2.2・pandas 2.2・Matplotlib 3.10・scikit-learn 1.6 の公式ドキュメントに基づいており、本書の記述が最新の状況と異なる場合がありますので、実際の受験および実務にあたっては、必ず実施団体の受験案内と 各ライブラリの公式ドキュメントをご確認ください。
- 5. 内容の正確性について。** 著者は内容の正確性に努めていますが、誤りがないことを保証するものではありません。本書の利用によって生じたいかなる損害についても、著者は責任を負いかねます。

6. **複製・再配布の禁止。** 本書は購入者ご本人が学習に利用する目的で提供されています。全部または一部を問わず、複製、転載、譲渡、貸与、公衆送信（インターネット上での共有・アップロードを含む）、翻案することを禁じます。
7. **商標について。** 本文中に記載される試験名、団体名、製品名、サービス名は、各社の商標または登録商標です。本文中では ™ および ® を省略しています。これらの表示は、当該商品・サービスを説明する 目的で用いているものであり、権利者による本書の承認を示すものではありません。

サンプル問題

製品版の構成・難易度・解説の書き方を確認していただくため、全収録問題から抜粋しています。コードや表を読ませる問題と、計算問題の例も収めています。

問 1

データ分析エンジニアの役割

機械学習の定義

経済産業省「AI・データの利用に関する契約ガイドライン」(AI編)による機械学習の説明として、最も適切なものはどれか。

- A 仕様を詳細に定め、既知の法則をもとに実装を段階的に詳細化していく、演繹的な開発手法の一つに位置づけられている
- B データの中から一定の規則を見つけ、その規則に基づいて未知のデータを推測・予測する学習手法の一つとされている
- C ディープラーニングとは別系統の技術とされ、ニューラルネットを多層にして推論の精度を高める手法は機械学習に含めていない
- D コンピュータサイエンスの研究者の間で定義が確立しており、機械学習に分類される手法は決定木とSVMの2つに整理されている

問 2

Pythonと環境

for 文と range()

次の表のコードを実行したときに出力されるものとして、最も適切なものはどれか。表の「字下げ」は行頭の空白の数を表す。

実行するコード

行	字下げ	コード
1	0	for n in range(2, 12, 3):
2	4	if n % 4 == 0:
3	8	break
4	4	print(n, end=' ')
5	0	else:
6	4	print('end')

- A 2 5 8 11 end
- B 2 5 end
- C 2 5
- D 2 5 8

問 3

数学の基礎

分割表と条件付き確率

次の表は、ある工程でつくった部品200個について、良品か不良品かと、出荷前の検査の判定をまとめたものである。検査で不合格と判定された部品の中から1個を無作為に選ぶとき、それが不良品である確率として、最も適切なものはどれか。

部品200個の区分と検査の判定（個）

区分	不合格	合格	計
不良品	18	2	20
良品	27	153	180
計	45	155	200

- A 0.09
- B 0.225
- C 0.4
- D 0.9

問 4

ライブラリによる分析実践

重複行の削除

次の DataFrame df に `df.drop_duplicates("店舗", keep="last")` を実行したとき、結果に残る行の行ラベルとして、最も適切なものはどれか。

DataFrame df

行ラベル	店舗	売上
0	東	12
1	西	8
2	東	15
3	南	9
4	西	11
5	東	7

A 3・4・5

B 0・1・3

C 3

D 0・1・2

問 5

ライブラリによる分析実践

欠損値を含むデータの描画

次の DataFrame df について、`df.plot(kind='line')` と `df.plot(kind='bar')` でグラフを描いた。欠損値 (NaN) の扱いの組合せとして、最も適切なものはどれか。

DataFrame df の中身

行ラベル	売上	来店数
1月	120	30
2月	NaN	28
3月	150	35
4月	140	NaN

- A 折れ線は NaN を0とみなして線を引き、棒グラフは NaN の値の棒だけを描かずに位置を詰める
- B 折れ線は NaN の位置で途切れ、棒グラフは NaN を0で埋めて高さ0の棒として扱う
- C 折れ線は NaN をはさむ前後の点を直線でつなぎ、棒グラフは欠けた値を0とみなして描く
- D 折れ線は欠けた月の値を0として線を落とし、棒グラフは NaN を含む2月と4月の行がまるごと表示されない

問 6

ライブラリによる分析実践

二値分類の評価

ある工場で、製品の画像から不良品（陽性）を見つける二値分類モデルを作り、テスト用の200件で予測した。結果の内訳は次の表のとおりである。このモデルの precision（適合率）と recall（再現率）を、この順に並べたものとして正しいものはどれか。

テストデータ200件の予測結果の内訳

実際のクラスと予測の組	件数
実際が不良品で、不良品と予測した	30
実際が良品で、不良品と予測した	10
実際が不良品で、良品と予測した	20
実際が良品で、良品と予測した	140

- A 0.75 と 0.60
- B 0.60 と 0.75
- C 0.75 と 0.75
- D 0.60 と 0.60

解答・解説

問 1 正解 B データ分析エンジニアの役割 機械学習の定義

ガイドラインは、機械学習を「データの中から一定の規則を発見し、その規則に基づいて未知のデータに対する推測・予測等を実現する学習手法の一つ」と定義している。開発の過程に着目すると、観察されたデータに基づいて帰納的にソフトウェアを開発する点が共通するとしている。したがって B が正しい。

他の選択肢 A は誤り。仕様を詳細に定めて実装を段階的に詳細化する演繹的な手法は、従来型のソフトウェア開発の説明である。C は誤り。ディープラーニングは、ニューラルネットを多層にして実行することで精度の高い推論を目指す機械学習の一手法とされている。D は誤り。機械学習という用語は多義的に使われ、確立した定義は必ずしもないとされ、手法の例として SVM、決定木、ニューラルネットワーク、クラスタリング等が挙げられている。

出典 AI・データの利用に関する契約ガイドライン 1.1版 AI編 第2 1(3) (機械学習)、2(1) (従来型のソフトウェア開発との違い)

問 2 正解 C Pythonと環境 for 文と range()

`range(2, 12, 3)` は 2 から始めて 3 ずつ増やし、終端の 12 を含まないので、`n` は $2 \cdot 5 \cdot 8 \cdot 11$ の順に変わる。2 と 5 は 4 で割り切れないので出力される。`n` が 8 のとき $8 \% 4$ が 0 になり、`break` で `for` ループを抜ける。ループの `else` 節は `break` されずにループが終わったときだけ実行されるので、`'end'` は出力されない。したがって C が正しい。

他の選択肢 A は誤り。`break` を見落とし、ループが最後まで回ったとみなした場合の出力である。B は誤り。`else` 節は `break` でループを抜けたときには実行されない。D は誤り。`n` が 8 のときは `print` より前の `break` でループを抜けるので、8 は出力されない。

出典 Python 3.13 チュートリアル 4.3 「`range()` 関数」・4.4 「`break` 文と `continue` 文」・4.5 「ループの `else` 節」

問 3 正解 C 数学の基礎 分割表と条件付き確率

「不合格である」という条件のもとで「不良品である」確率は条件付き確率である。どの部品も同じ確からしきで選ばれるので、場合の数の数え上げで求められる。不合格の45個のうち不良品は18個なので、 $18 \div 45 = 0.4$ となる。式で書くと、 $P(\text{不合格かつ不良品}) = P(\text{不合格}) \times P_{\text{不合格}}(\text{不良品})$ の関係から、 $0.09 \div 0.225 = 0.4$ と同じ値になる。したがって **C** が正しい。

他の選択肢 A は誤り。200個全体の中で「不合格かつ不良品」である確率 ($18 \div 200 = 0.09$) であり、条件を付けていない。B は誤り。200個全体の中で不合格と判定される確率 ($45 \div 200 = 0.225$) である。D は誤り。条件と結果を逆にした、不良品の中で不合格と判定される確率 ($18 \div 20 = 0.9$) である。

出典 高等学校学習指導要領（平成30年告示）解説 数学編 数学A「(2) 場合の数と確率」（条件付き確率、 $P(A \cap B) = P(A)P_{A|B}(B) = P(B)P_{B|A}(A)$ ）

問 4 正解 A ライブラリによる分析実践 重複行の削除

`drop_duplicates()` は、引数に渡した列の値が重なる行を取り除く。`keep="last"` を指定すると、重なりのある組のうち最後に現れた行だけを残す。店舗ごとに最後の行を見ると、東は 5、西は 4、南は 3（1行だけ）なので、残るのは行ラベル 3・4・5 である。

店舗ごとの行と残る行

店舗	現れる行	残る行
東	0・2・5	5
西	1・4	4
南	3	3

したがって A が正しい。

他の選択肢 B は誤り。既定の `keep="first"`（最初に現れた行を残す）で実行した場合の結果である。C は誤り。`keep=False`（重なりのある行をすべて取り除く）で実行した場合の結果である。D は誤り。`keep="last"` で取り除かれる側の行である。 `duplicated("店舗", keep="last")` が `True` を返すのはこの3行になる。

[出典](#) pandas 2.2 User Guide 「Indexing and selecting data」 「Duplicate data」

問 5 正解 B ライブラリによる分析実践 欠損値を含むデータの描画

pandas はグラフの種類ごとに欠損値の扱いを決めている。**折れ線は NaN の位置で途切れ、棒グラフは NaN を0で埋める**。2月の売上の線は途切れ、4月の来店数の棒は高さ0になる。したがって **B** が正しい。

他の選択肢 A は誤り。0で埋めるのは積み上げの折れ線・棒・面・円グラフで、通常の折れ線は途切れる。棒グラフも NaN の棒を省かず0で埋める。C は誤り。折れ線は NaN の前後を結ばずに途切れさせる。D は誤り。折れ線は NaN を0とせずに途切れさせ、棒グラフは NaN を0で埋めて行を除かない。値を除くのは散布図・六角ビンや、列ごとに除くヒストグラム・箱ひげ図・KDE である。

出典 pandas User Guide Chart visualization (2.2) 「Plotting with missing data」

問 6 正解 A ライブラリによる分析実践 二値分類の評価

陽性を不良品とすると、tp（真陽性）は30、fp（偽陽性）は10、fn（偽陰性）は20である。**precision は $tp \div (tp + fp)$ 、recall は $tp \div (tp + fn)$ で求める**。

precision と recall の計算

指標	計算	値
precision	$30 \div (30 + 10) = 30 \div 40$	0.75
recall	$30 \div (30 + 20) = 30 \div 50$	0.60

precision は「陽性と予測したもののうち本当に陽性だった割合」、recall は「本当の陽性のうち見つけられた割合」を表す。したがって **A** が正しい。

他の選択肢 B は誤り。precision と recall の式を取り違えた値である。C は誤り。recall の分母にも fp を使い、 $30 \div 40$ としてしまった場合の値である。D は誤り。precision の分母にも fn を使い、 $30 \div 50$ としてしまった場合の値である。

出典 scikit-learn 1.6.1 ユーザーガイド「Metrics and scoring」3.4.4.9 Precision, recall and F-measures・3.4.4.9.1 Binary classification

本章の出典

- デジタルスキル標準 DSS-P 分冊版 データサイエンティスト編 ver.2.0（独立行政法人情報処理推進機構、2026年4月）
- AI・データの利用に関する契約ガイドライン 1.1版（経済産業省、令和元年12月）
- 令和6年版 情報通信白書 第1部第3章第1節（総務省、2024年7月）
- The Python Tutorial 1. Whetting Your Appetite（Python Software Foundation、Python 3.13）
- Python チュートリアル 6. モジュール／12. 仮想環境とパッケージ（Python Software Foundation、3.13 日本語版）
- venv — 仮想環境の作成（Python Software Foundation、3.13 日本語版）
- Python のセットアップと利用 4. Windows で Python を使う（Python Software Foundation、3.13 日本語版）
- scikit-learn Getting Started／2.5 Decomposing signals in components／3.1 Cross-validation（scikit-learn 開発チーム、1.6）
- pandas Package overview（pandas 開発チーム、2.2）
- NumPy What is NumPy?（NumPy 開発チーム、2.2）
- Matplotlib Quick start guide（Matplotlib 開発チーム、3.10）
- Anaconda Distribution and Miniconda（Anaconda, Inc.）
- conda Managing environments（conda 公式ドキュメント）
- Python 3.13 チュートリアル 第3章「形式ばらない Python の紹介」（Python Software Foundation、日本語版 3.13.17）
- Python 3.13 チュートリアル 第4章「その他の制御フローツール」（同）
- Python 3.13 チュートリアル 第5章「データ構造」（同）
- Python 3.13 チュートリアル 第7章「入力と出力」（同）
- Python 3.13 チュートリアル 第8章「エラーと例外」（同）
- Python 3.13 チュートリアル 第9章「クラス」（同）
- Python 3.13 チュートリアル 第10章「Brief tour of the standard library」（同）
- JupyterLab 4.3 ドキュメント「Get Started」「Installation」（Project Jupyter）
- JupyterLab 4.3 ドキュメント「The JupyterLab Interface」「Notebooks」「Commands」（Project Jupyter）

-
- 文部科学省「高等学校学習指導要領（平成30年告示）解説 数学編 理数編」
 - 九州大学「微分積分・同演習B 多変数関数と偏導関数」（講義資料）
 - 総務省統計局「なるほど統計学園」（中心的な傾向を捉える、データの散らばりを捉える、データの散らばりを見る、ヒストグラム、箱ひげ図、散布図、複数の変数の関係性を見る、推測統計）
 - 総務省統計局「統計学習用補助教材」統計基本用語集（案）
 - 北海道大学 情報理論 配布資料 #2「エントロピー」（平成17年）
 - pandas 2.2 Package overview（pandas 開発チーム）
 - pandas 2.2 User Guide「Intro to data structures」（pandas 開発チーム）
 - pandas 2.2 User Guide「Essential basic functionality」（pandas 開発チーム）
 - pandas 2.2 User Guide「IO tools (text, CSV, HDF5, …）」（pandas 開発チーム）
 - pandas 2.2 User Guide「Indexing and selecting data」（pandas 開発チーム）
 - pandas 2.2 User Guide「Reshaping and pivot tables」（pandas 開発チーム）
 - Matplotlib Introduction to Figures（Matplotlib 開発チーム、3.10）
 - Matplotlib Introduction to Axes (or Subplots)（Matplotlib 開発チーム、3.10）
 - Matplotlib Introduction to Artists（Matplotlib 開発チーム、3.10）
 - Matplotlib Plot types（Matplotlib 開発チーム、3.10）
 - Matplotlib Customizing Matplotlib with style sheets and rcParams（Matplotlib 開発チーム、3.10）
 - pandas User Guide Chart visualization（pandas 開発チーム、2.2）
 - scikit-learn 1.6.1 ユーザーガイド「Decomposing signals in components (matrix factorization problems)」（scikit-learn 開発チーム）
 - scikit-learn 1.6.1 ユーザーガイド「Metrics and scoring: quantifying the quality of predictions」（scikit-learn 開発チーム）
 - scikit-learn 1.6.1 ユーザーガイド「Cross-validation: evaluating estimator performance」（scikit-learn 開発チーム）
 - scikit-learn 1.6.1 ユーザーガイド「Tuning the hyper-parameters of an estimator」（scikit-learn 開発チーム）
 - scikit-learn 1.6.1 ユーザーガイド「Clustering」（scikit-learn 開発チーム）
 - scikit-learn 1.6.1「Getting Started」（scikit-learn 開発チーム）

いずれも公的機関等が公表している資料です。制度は改正されるため、受験にあたっては必ず最新版をご確認ください。

Python 3 エンジニア認定データ分析試験 想定問題集

2026年10月版（第1版）

発行：Skill Base Camp <https://skillbasecamp.com>

本書は非公式の学習教材であり、試験を実施する団体の公認・監修を受けたものではありません。

内容は公開情報に基づき著者が独自に作成しています。制作時点の情報に基づくため、制度の最新の状況は一次情報をご確認ください。

文中に記載される試験名・団体名・製品名は、各社の商標または登録商標です。本文中では ™ および ® を省略しています。