

# データアナリストコース

## シラバス

## SQL 言語

学習時間	総学習時間	20 時間 00 分
	e ラーニング講座	12 時間 00 分
	課題・実践	8 時間 00 分
単位	対象外	

### 前提知識・経験

情報技術やデータベースの初歩的な知識

### 到達目標

- ① 表内の行および列を、任意の順番に取り出すことができるようになります。
- ② 表へ新規に行を追加したり、表内の行を変更／削除したりできるようになります。
- ③ 表内の行をさまざまな形で加工して取り出すことができるようになります。
- ④ 複数の表を結合したり、複雑な問合せを行ったりできるようになります。

### 学習内容

- ① リレーショナル・データベース
- ② データの検索
- ③ 行の制限
- ④ データの操作
- ⑤ 単一行関数
- ⑥ グループ関数
- ⑦ データの結合
- ⑧ 副問合せ

### 構成

【e ラーニング講座】

SQL 言語

### 成績評価の基準

評価対象外

## データサイエンティストマインド

学習時間	総学習時間	7時間 34分
	eラーニング講座	5時間 34分
	課題・実践	2時間 00分
単位	2単位	

### 前提知識・経験

義務教育で習得した知識

### 到達目標

#### 【1.学習指針】

理論を確認することがいかに実務に活きるのか、今後の学習の指針とともに理解する。

#### 【2.代表的スキル要件】

##### ① Drew Conway (2010)

データサイエンティストに求められる代表的なスキルについて、Conwayの主張を通して理解する。

##### ② Dominik Haitz (2019)

データサイエンティストに求められる中心的スキルの変遷について、Haitzの主張を通して理解する。

##### ③ データ分析プロセス

スキル要件と対応付けながら、データ分析の大まかな流れを理解する。

#### 【3.私家版 DS 必須スキル】

##### ① 批判スキル (防御)

データや情報を正しく読み解く際に必要な用語の定義と主張の仮定について、実例を通して理解する。また、理論的なバックグラウンドやエンジニアリングに強みを持つ相手から強い主張にあった場合の心構えを身につける。

##### ② 提案スキル (攻撃)

一般的に問題解決が3つの方針に基づくことを理解する。そして、専門外のドメイン知識を周囲から引き出すための問い(Q)の類型化を把握できる。また、広い視点からの仮設構築がデータ分析の基盤となるだけでなく、価値創出のためには必ずしもデータに偏重す

る必要がないことを理解する。

③ 調整スキル（補助・強化）

データサイエンティストとして建設的に議論するためのヒントを把握する。また、人が必ずしも論理だけでは動かないことを認知バイアスとレトリックの文脈で理解する。

## 学習内容

---

### 【学習指針】

- ① 実務に生きる理論
- ② 導入
- ③ 汎用性
- ④ 再現性
- ⑤ 現実性
- ⑥ まとめ

### 【代表的スキル要件】

- ① Drew Conway (2010)
  - Drew Conway (2010) 3つのスキル
- ② Dominik Haitz (2019)
  - Business Mindset
  - Business Mindset と 3つのスキル
- ③ データ分析プロセス
  - CRISP-DM

### 【私家版 DS 必須スキル】

- ① 批判スキル（防御）
  - 定義する力
  - 仮定を見出す力
  - 理論屋・実装屋から身を守る
- ② 提案スキル（攻撃）
  - 問題解決の一般的アプローチ
  - 問題を換言する
  - 「Q」の運用
  - 仮説構築
- ③ 調整スキル（補助・強化）
  - 説得・依頼の技術

- 認知バイアスの把握
- 修辞法/弁論術の取扱い

## 構成

---

### 【eラーニング講座】

- ① 学習指針
- ② 代表的スキル要件
- ③ 私家版 DS 必須スキル

### 【提出課題】

全1回

## 成績評価の基準

---

eラーニングの受講、課題の提出をもって6段階にて評価する。

## クリティカルシンキング

学習時間	総学習時間	3時間04分
	eラーニング講座	1時間04分
	課題・実践	2時間00分
単位	2単位	

### 前提知識・経験

- ① 義務教育で習得した知識
- ② 「01 データサイエンティストマインド」を修了していること

### 到達目標

- ① Logical - Lateral - Critical  
ロジカルシンキング、ラテラルシンキングとクリティカルシンキングそれぞれの基本概念について理解する。
- ② Logical Thinking  
具体例を通してロジカルシンキングについて理解する。
- ③ Lateral Thinking  
具体例を通してラテラルシンキングについて理解する。
- ④ Critical Thinking  
具体例を通してクリティカルシンキングについて理解する。

### 学習内容

- ① Logical - Lateral - Critical
  - 概論
  - Thinking とスタートラインと焦点
  - Logical Thinking
  - コイン投げのギャンブル
  - 格上にとっての5番勝負と7番勝負
  - おまけ：測定の工夫
- ② Lateral Thinking
  - トップと最下位の同時改革
  - 同じ弁当を倍の値段で売る
  - 自動改札機の情報処理
- ③ Critical Thinking

- 星座と運動神経
- 縁起の良い日と宝くじ/ 簡単すぎる懸賞のクイズ
- 外資系コンサルの福利厚生？
- 過去最高の記録
- おまけ：平均の推定

## 構成

---

【eラーニング講座】

クリティカルシンキング

【提出課題】

全1回

## 成績評価の基準

---

eラーニングの受講、課題の提出をもって6段階にて評価する。

## データ理解

学習時間	総学習時間	5 時間 09 分
	e ラーニング講座	3 時間 09 分
	課題・実践	2 時間 00 分
単位	2 単位	

### 前提知識・経験

義務教育で習得した知識

### 到達目標

#### 【1.概論】

情報とデータの定義を確認し、データ取得過程の重要性とデータ自身の限界について理解する。

#### 【2.恣意性の発生源】

- ① データの切り出し  
データ取得の段階で生じ得る恣意性について選択バイアスを中心に理解し、シンプソンのパラドクスを自身の言葉で説明できる状態を目指す。
- ② データの測定  
データを測定する段階で生じ得る恣意性と、恣意性を助長する質問の禁じ手を理解する。
- ③ データの自意識  
データ自身の置かれた状況で恣意性が発生することを理解する。

#### 【3.尺度水準】

- ① 定義と具体例  
尺度の4水準について、定義と実例、許される数学的操作について理解する。
- ② 尺度から見えること  
典型的な誤用や非自明な事実を尺度の定義から確認することで各種尺度への理解を深める。

### 学習内容

#### 【1.概論】

- ① 情報とデータ
- ① 定義

- ② 取得過程の重要性とデータの恣意性
- ③ データ取得の実例
- ④ 定量データの限界

#### 【2.恣意性の発生源】

- ① データの切り出し
  - 選択バイアス
  - 因果モデル
- ② データの測定
  - 切断データ
  - 打ち切りデータ
  - 丸め込まれたデータ
  - 質問の禁じ手
- ③ データの自意識
  - 回答の動機
  - 回答バイアス
  - 質問と回答者・回答の分離
  - 代表的心理効果

#### 【3.尺度水準】

- ① 定義と具体例
- ② 尺度から見えること

### 構成

---

#### 【eラーニング講座】

- ① 概論
- ② 恣意性の発生源
- ③ 尺度水準

#### 【提出課題】

全1回

### 成績評価の基準

---

eラーニングの受講、課題の提出をもって6段階にて評価する。

## 環境構築

学習時間	総学習時間	1 時間 15 分
	e ラーニング講座	1 時間 15 分
単位	1 単位	

### 前提知識・経験

- ① 義務教育で習得した知識
- ② コマンドラインインタフェース (CLI) の利用
- ③ ターミナル (macOS)、あるいはコマンドプロンプト (Windows OS) の利用 ※推奨
- ④ Python 仮想環境の利用 ※推奨
- ⑤ パッケージ管理 ※推奨
- ⑥ デバッグの概念
- ⑦ 入力補完機能 (UNIX 系 OS のタブキーを用いた補完など) の利用 ※推奨
- ⑧ コードのデバッグ ※推奨
- ⑨ logger の利用 ※推奨

### 到達目標

- ① 仮想環境  
データの分析に必要な仮想環境を作成し、必要に応じたパッケージ管理ができる。
- ② Jupyter Notebook  
Jupyter Notebook 特有のコマンドおよび最低限必要となる機能について理解する。

### 学習内容

- ① 仮想環境
  - 環境の確認
  - 環境の新規作成
  - 環境のコピー
  - パッケージのインストール
  - パッケージの更新とファイルの削除
- ② Jupyter Notebook
  - Jupyter Notebook と仮想環境
  - ショートカットキーの基礎
  - Markdown の活用
  - 基本機能

- マジックコマンド
- デバッグとおすすめ拡張機能

## 構成

---

【e ラーニング講座】

環境構築

## 成績評価の基準

---

e ラーニングの受講をもって2段階にて評価する。

## データサイエンティスト体験

学習時間	総学習時間	0 時間 41 分
	e ラーニング講座	0 時間 41 分
単位	1 単位	

### 前提知識・経験

義務教育で習得した知識

### 到達目標

データサイエンティストの仕事としてイメージしやすい機械学習アルゴリズムの構築について、実データでのコードと合わせながら概観する。

### 学習内容

#### 推薦アルゴリズムの構築

- データ取得と概要の把握
- データ理解と前処理
- 比較ベースラインの準備
- 提案モデルの作成
- 提案モデルの検証
- Future Work の整理検討

### 構成

#### 【e ラーニング講座】

データサイエンティスト体験

### 成績評価の基準

e ラーニングの受講をもって2段階にて評価する。

## 実戦のための代表的ライブラリ

学習時間	総学習時間	16 時間 32 分
	e ラーニング講座	7 時間 01 分
	課題・実践	9 時間 31 分
単位	2 単位	

### 前提知識・経験

#### 【NumPy】

- ① Python の基本的なデータ構造と制御フローについての知識
- ② デシリアライズや reduce 関数の知識 ※推奨
- ③ データの縦持ち・横持ちの概念（RDB のテーブルで問題となる程度）についての知識
- ④ 諸関数および数学定数の定義についての知識
- ⑤ Python の基本的なデータ構造と制御フロー、諸統計量の定義や関係（性質）についての知識
- ⑥ 大学学部 1 年生程度の線形代数の知識

#### 【pandas】

- ① Python の基本的なデータ構造と制御フローについての知識
- ② NumPy の ndarray での配列操作についての知識
- ③ ソートアルゴリズムの知識 ※推奨
- ④ 諸統計量の定義や関係（性質）についての知識
- ⑤ 基本的な集計関数についての知識
- ⑥ クロス集計の知識
- ⑦ SQL における GROUP BY 句についての知識

### 到達目標

#### 【1.NumPy】

- ① NumPy のデータ構造  
NumPy 特有のデータ構造である ndarray の生成と演算にともなうブロードキャストについて理解する。
- ② NumPy を介した配列の読み書き  
多次元配列の書き出しと読み書きが自由に行えるようになることを目指す。また、pickle を用いる際のセキュリティ上の注意点について最低限おさえておくべき内容を理解する。

- ③ NumPy における要素の参照  
インデックスやスライス、関数を用いて、多次元配列から必要な要素や配列の一部を参照できる。
- ④ NumPy における配列操作  
多次元配列に対して、形状変化をとまなうさまざまな配列操作ができる。
- ⑤ 数学関数  
初等関数、数学定数、要約統計量について、NumPy で用意されている関数を介して自由に演算が行える。
- ⑥ 行列代数  
行列代数について、NumPy で用意されている関数を介して自由に演算が行える。

## 【2.pandas】

- ① pandas のデータ構造  
pandas 特有のデータ構造である Series と DataFrame について、生成から特徴まで ndarray との差異まで含めて理解する。
- ② pandas を介したファイルの読み書き  
ローカルや Web 上にあるさまざまなデータ (csv に限らず Excel や tsv、zip ファイル等も含む) を DataFrame として読み込める。また、必要に応じて DataFrame をローカルに csv として書き出すことができる。
- ③ pandas における要素の参照  
DataFrame および Series の特定要素を参照する方法を習得する。同時に実データに対して求める処理が実行されるコードが書ける。
- ④ pandas における配列操作  
pandas ですすでに入っている要素を置き換える方法を理解する。カラムやレコードの追加、結合および削除できる。
- ⑤ データ集計  
pandas を介したデータの集計について理解する。

## 学習内容

---

### 【NumPy】

- ① NumPy のデータ構造
  - ndarray オブジェクト
  - Broadcasting
- ② NumPy を介した配列の読み書き
  - 配列の書き出し
  - 配列の読み込み

- pickle を用いる際の一般的注意
- ③ NumPy における要素の参照
  - インデックスを指定した参照
  - スライスを用いた参照
  - np.where 関数を用いた参照
- ④ NumPy における配列操作
  - 形状の変換
  - 軸の順序交換
  - 要素の削除
  - 配列の結合
- ⑤ 数学関数
  - 初等関数と数学定数
  - 記述統計
  - 確率変数間の関数を捉える統計量
- ⑥ 行列代数
  - 行列積
  - 行列の特徴を捉える

### 【pandas】

- ① pandas のデータ構造
  - Series オブジェクト
  - DataFrame オブジェクト
  - Index オブジェクト
- ② pandas を介したファイルの読み書き
  - ファイルの読み込み
  - ファイルの書き出し
- ③ pandas における要素の参照
  - pandas における要素の参照
  - 「インデックス」を用いた参照
  - property(loc/iloc)を用いた参照
  - bool 型を用いた参照
  - query を用いた参照
- ④ pandas における配列操作
  - 要素の置換
  - カラム/レコードの追加・結合
  - カラム/レコードの削除

- おまけ：縦持ち・横持ち変換
- 文字列操作
- ⑤ データ集計
  - データ集計
  - ソート
  - 要約統計量
  - クロス集計
  - 関数の適用
  - GroupBy オブジェクト

## 構成

---

### 【e ラーニング講座】

- ① NumPy
- ② pandas

### 【提出課題】

専門課題 1 回 + 複合課題 1 回

## 成績評価の基準

---

e ラーニングの受講、課題の提出をもって 6 段階にて評価する。

## データの前処理

学習時間	総学習時間	8時間 31分
	eラーニング講座	3時間 41分
	課題・実践	4時間 50分
単位	2単位	

### 前提知識・経験

#### 【目的指向の前処理】

- ① 義務教育で習得した知識
- ② 「05 実戦のための代表的ライブラリ」を修了していること
- ③ 諸統計量の知識
- ④ 重複データを扱った経験 ※推奨
- ⑤ 欠損データを扱った経験 ※推奨
- ⑥ 外れ値を扱った経験 ※推奨

#### 【2.尺度に応じた変換】

- ① 義務教育で習得した知識
- ② 「05 実戦のための代表的ライブラリ」を修了していること
- ③ 諸統計量の知識
- ④ ハッシュ関数の概念についての知識

### 到達目標

#### 【1.目的指向の前処理】

- ① 重複の処理  
単純集計で対応可能な重複データの代表的処理法について理解する。
- ② 欠損値の処理  
単純集計で対応可能な欠損値の代表的処理法について学び、また、欠損のメカニズムについても概観し、不適切な対応についても理解する。
- ③ 外れ値の処理  
単純集計で対応可能な外れ値の代表的処理法について理解する。

#### 【2.尺度に応じた変換】

- ① 量的データの変換  
量的データに対して用いられる変換について理解する。

② 質的データの変換

質的データに対して用いられる変換について理解する。

## 学習内容

---

### 【1.目的指向の前処理】

① 重複の処理

- データの準備/重複の確認
- 重複の除去
- 重複の取り扱い検討

② 欠損値の処理

- データの準備
- 欠損の確認
- 欠損のメカニズム
- 代表的な欠損対応
- 欠損の除去
- 欠損の補完
- おまけ：すべての欠損, NaN/Null/None にあらず
- 欠損の取り扱い検討

③ 外れ値の処理

- データの準備
- 外れ値の確認 z-score/四分位範囲
- 外れ値の Clipping/おまけ

④ 【2.尺度に応じた変換】

⑤ 量的データの変換

- データの準備
- 線形な変換
- 非線形な変換
- 順位ベースでの変換

⑥ 質的データの変換

- データの準備
- label encoding
- one-hot encoding
- hash encoding
- frequency encoding
- target encoding

## 構成

---

### 【eラーニング講座】

- ① 目的指向の前処理
- ② 尺度に応じた変換

### 【提出課題】

専門課題 1 回 + 複合課題 1 回

## 成績評価の基準

---

eラーニングの受講、課題の提出をもって6段階にて評価する。

## データの可視化

学習時間	総学習時間	9 時間 19 分
	e ラーニング講座	4 時間 13 分
	課題・実践	5 時間 06 分
単位	2 単位	

### 前提知識・経験

#### 【1. Matplotlib, Seaborn, Plotly】

- ① 義務教育で習得した知
- ② 「05 実戦のための代表的ライブラリ」を修了していること
- ③ 要約統計量の知識

#### 【2. 可視化情報の展開】

- ① 義務教育で習得した知識
- ② 「05 実戦のための代表的ライブラリ」を修了していること

### 到達目標

#### 【1. Matplotlib, Seaborn, Plotly】

- ① 可視化の必要性  
データサイエンティスト業務においてなぜ可視化が必要となるのか自らの言葉で説明できる。
- ② 描画について基礎的な導入  
グラフ描画ライブラリである Matplotlib について、そのコーディングスタイルである object-oriented style を中心に描画の基礎を理解する。
- ③ 散布図  
データから散布図を描けるようになるとともに、付随する問題についても理解する。
- ④ ヒストグラム  
データからヒストグラムを描けるようになるとともに、付随する問題についても理解する。
- ⑤ 棒グラフ  
データから棒グラフを描けるようになるとともに、付随する問題についても理解する。
- ⑥ 折れ線グラフ  
データから折れ線グラフを描けるようになるとともに、付随する問題についても理解する。

- ⑦ 箱ひげ図  
データから箱ひげ図を描けるようになるとともに、付随する問題についても理解する。
- ⑧ その他,さまざまな可視化  
遭遇頻度の比較的高い可視化手法について概観する。

## 【2. 可視化情報の展開】

可視化の一形態であるアニメーションについて、2つの方法で作成できる。

### 学習内容

---

#### 【1. Matplotlib, Seaborn, Plotly】

- ① 可視化の必要性
  - Anscombe's Quartet
  - Datasaurus
- ② 描画について基礎的な導入
  - 描画の基礎
  - object-oriented style
  - pyplot style
- ③ 散布図
  - 散布図の基礎
  - 散布図の「 $+\alpha$ 」
- ④ ヒストグラム
  - ヒストグラムの基礎
  - ヒストグラムの「 $+\alpha$ 」
- ⑤ 棒グラフ
  - 棒グラフの基礎
  - 棒グラフの「 $+\alpha$ 」
- ⑥ 折れ線グラフ
  - 折れ線グラフの基礎
  - 折れ線グラフの「 $+\alpha$ 」
- ⑦ 箱ひげ図
  - 箱ひげ図の基礎
  - 箱ひげ図の「 $+\alpha$ 」
- ⑧ その他,さまざまな可視化
  - 簡易 EDA
  - 参考サイトーThe Python Graph Gallery

#### 【2. 可視化情報の展開】

#### アニメーション作成の基礎

- 概論
- ArtistAnimation による描画例
- FuncAnimation による描画例
- アニメーションの保存

#### 構成

---

##### 【e ラーニング講座】

- ① Matplotlib, Seaborn, Plotly
- ② 可視化情報の展開

##### 【提出課題】

専門課題 1 回 + 複合課題 1 回

#### 成績評価の基準

---

e ラーニングの受講、課題の提出をもって 6 段階にて評価する。

## データ分析実践

実践時間	6時間 30分
単位	8単位

### 前提知識・経験

データアナリストコースの知識及び以下の学習を修了していること。

- ① データサイエンティストマインド
- ② クリティカルシンキング
- ③ データ理解
- ④ 環境構築
- ⑤ データサイエンティスト体験
- ⑥ 実戦のための代表的ライブラリ
- ⑦ データの前処理
- ⑧ データの可視化

### 到達目標

データアナリストコースで学習した内容をアウトプットすることで、即戦力データアナリストとして必要な実践能力を身に着けること。状況や目標からデータ分析の必要性について思考し、データごとに適切な前処理を実施した上で適切な可視化を実行することができること。また個人・チームで実施した分析結果からビジネスの課題解決や新たな価値創造のヒントとなるようなインサイトを適切にとらえ、提示できること。

### 内容

計3回の実践演習に取り組み、成果発表をする。成果発表に加え、以下の提出物を提出するものとする。複数人でチームを組み取り組む場合は、チームで相談しながら進めること。

### 構成

以下3回の実践演習を実施する。

- ① データから価値ある情報や知見を獲得する
- ② 与えられた選択肢の中から、データをもとに決定する
- ③ データを解釈しやすいように整形・可視化する

### 成績評価の基準

成果物の提出・発表をもって6段階にて評価する。