データサイエンティストのスキルレベル 2021年版

赤字は2019年版に比較した更新部分

	ビジネス（business problem solving）力	データサイエンス（data science）力	データエンジニアリング（data engineering）力
①Senior Data Scientist **業界を代表するレベル**	・業界を代表するデータプロフェッショナルとして、組織全体や市場全体レベルでのインパクトを生み出すことができる -対象とする事業全体、産業領域における課題の切り分け、論点の明確化・構造化 -新たなデータ分析、解析、利活用領域の開拓 -組織・会社・産業を横断したデータコンソーシアムの構築、推進 -事業や産業全体に対するデータ分析を核としたバリューチェーン創出 -技術や環境の変化に応じデータ×AI関連のガイドラインや社会のあり方について提言・働きかけ など	・業界を代表するデータプロフェッショナルとして、データサイエンス における既存手法の限界を打ち破り、新たに課題解決可能な領域を 切り拓いている -既存手法では対応困難な課題に対する新規の分析アプローチの開発・実践・横展開 -高難度の分析プロジェクトのアプローチ設計、推進、完遂能力 など	・業界を代表するアーキテクトとして、データサイエンス領域で取り組もうとしている分析アプローチを、挑戦的な課題であっても安定的に実現できる -複数のデータソースを統合した例外的規模のデータシステム、もしくはデータプロダクトの構築、全体最適化 -技術的限界を熟知し、これまでにない代案の提示・実行 など
②Full Data Scientist **棟梁レベル**	・生み出す価値にコミットするプロフェッショナルとして、データサイエンティストとは何かを体現したビジネス判断、課題解決ができる ・初見の事業領域に向かい合う場合や、スコープが複数の事業にまたがる場合であっても本質的な課題を見出し、構造化・深掘りができる ・入り組んだステークホルダー構造の中で、Win-Winの形で価値を設計・創造し、そこからの発展を見据えた仕込みと推進するハブとしての役割を担うことができる ・プロフェッショナルからなる複数のチームによるプロジェクトの役割、目標を定義、リスクをマネージしつつ推進し、全体としてのアウトプットにコミットできると共に、メンバーを育成、さらには持続的な育成システムを作り出すことができる ・自身が関わる事業や組織において、社会の変化に応じたAI-Ready化を推進できる	・予測、グルーピング、機械学習、深層学習、大量データの可視化、 自然言語処理、画像・映像認識、音声認識、最適化問題などの 応用的なデータサイエンス関連のスキルを活かし、データ分析 プロジェクトの技能的主軸を担うことができる ・初見の事業領域に向かい合う場合や、スコープが複数の事業にまたがる 場合であっても、適切な分析・解析アプローチの設計、実行、深掘り ができる ・複数もしくは高度な分析プロジェクトを持つチームにおいて、 Associate Data Scientist（独り立ちレベル）以下のメンバー の技能を育成することができる	・数十億レコード程度の分析環境の要件定義・設計、データ収集/蓄積/ 加工/共有プロセスやITセキュリティに関するデータエンジニアリング 関連のスキルを活かし、データ分析プロジェクトを中核的に推進する ことができる ・全体を統括するアーキテクトとして、サービス上のそれぞれの機能が どのデータに関連があるか総合的に把握し、新たな技術を理解しつつ 設計や開発に活かすことができる ・複数もしくは高度な分析プロジェクトを持つチームにおいて、 Associate Data Scientist（独り立ちレベル）以下のメンバーの技能を 育成することができる
③Associate Data Scientist **独り立ちレベル**	・大半のケースで自立したプロフェッショナルとして、ビジネス判断、課題解決ができる -ビジネス要件の整理、プロジェクトの企画・提案・遂行 -分析のアプローチ設計および結果への適切な対応と評価 -AI・機械学習がもたらす倫理課題への対応 -データや分析結果の開示範囲、知財リスクの確認などの適切な対応 ・既知の領域、テーマであれば、新規課題であっても解くべき問題の見極めや構造化、深掘りができる ・データ、分析結果に対する表面的な意味合いを超えた洞察力を持ち、担当プロジェクトの検討結果を取りまとめ、現場への説明、実装を自律的かつ論理的に行うことができる ・５名前後のプロフェッショナルによるチームでのプロジェクトを推進しアウトプットにコミットできる -タスクの粘り強い完遂 -イシュードリブンでスピード感のある判断 -プロジェクトマネジメントと個別メンバーの育成 -異なるスキル分野の専門家、事業者との協働 など	・単一プロジェクトにおけるデータ分析をFull Data Scientist （棟梁レベル）に相談しつつ推進できる ・Assistant Data Scientist（見習いレベル）の日々の活動に適切な指示 ができる ・既知の領域、テーマであれば、新規課題であっても適切な分析・解析 アプローチの設計、実行、深掘りができる ・基礎的なデータ加工については、自律的に実施できる -外れ値・異常値・欠損値の対応 -適切な学習データ、検証データ、テストデータの作成 -特徴量エンジニアリングによる効果的なデータの作成 など ・基礎的な分析活動については、自律的に実施できる -p値の限界の理解と現実的な対応 -多重共線性を考慮した重回帰分析 -傾向スコアなどによる因果効果の推定 -系列データの特性を理解した時系列分析 -適切なクラスター数による非階層クラスター分析 -ライブラリなどを活用した機械学習や深層学習 など ・非構造化データに対する基礎的な分析を自律的に実施できる -画像のパターン抽出や音声のノイズ除去 -形態素解析などを用いた基本的文書構造解析やベクトル表現 など	・単一プロジェクトにおけるデータ処理・環境構築をFull Data Scientist（棟梁レベル）に相談しつつ推進できる ・Assistant Data Scientist（見習いレベル）の日々の活動に適切な指示ができる ・数千万レコード程度のデータ処理・環境構築については自律的に実施できる -データの重要性や分析要件に則したシステム要件定義 -適切なデータフロー図、論理データモデル、ER図の作成 -HadoopやSparkでの管理対象データ選定 -SDKやAPI、ライブラリ、コンテナ技術などの適切な活用 -SQLの構文理解と実行 -分析プログラムのロジック理解と分析結果検証 -データ匿名化方法の理解と加工処理の設計 など ・分析要件に合わせたインフラ環境（GPU/CPU、クラウド/オンプレミスなど）を設計・実装できる ・AIシステム運用を推進できる -AIモデルの作成、デプロイ、学習データの更新、モデル再学習といったライフサイクル管理の実践
④Assistant Data Scientist **見習いレベル**	・ビジネスにおける論理とデータの重要性を理解したデータプロフェッショナルとして行動規範と判断が身についている -データやAIを取り扱う倫理と法令の理解 -引き受けたことは逃げずにやり切るコミットメント -迅速な報告や、報告に対する指摘のすみやかな理解 など ・データドリブンな分析的アプローチの基本が身についており、仮説や既知の問題が与えられた中で、必要なデータを入手し、分析、取りまとめることができる -データや事象のダブりとモレの判断力 -分析前の目的、ゴール設定 -データの出自や情報引用元の信頼性の判断 -目的に即したデータ入手 -分析結果の意味合いの正しい言語化 -モニタリングの重要性理解 など ・担当する検討領域についての基本的な課題の枠組みを理解できる -担当する業界の主要な変数（KPI） -基本的なビジネスフレームワーク など	・統計数理や線形代数、微分積分、集合理論の基礎知識を有している （代表値、分散、標準偏差、正規分布、条件付き確率、母集団、相関、ベイズの定理、対数関数・指数関数、ベクトルや行列の計算方法、関数の傾きと微分の関係、論理演算と集合演算の関係など） ・データ分析の基礎知識を有している -分析用データの整備 -予測、グルーピングなどのモデリング -標本抽出、点推定と区間推定 -モデルの評価 ・機械学習の基本的な概念を理解している -教師あり学習と教師なし学習の違い -機械学習における過学習の理解 -深層学習のメリットに関する理解など ・適切な指示のもとに、データ加工と基礎的な分析を実施できる -基本統計量や分布の確認、および前処理（外れ値・異常値・欠損値の除去・変換や標準化など） -クロス集計による偏りの把握や、単回帰分析の実行 など ・データ可視化の基礎知識を有している -軸だし -不適切な表現の理解 -意味合いの導出	・データやデータベースに関する基礎知識を有している -構造化/非構造化データの判別、論理モデル作成 -ER図やテーブル定義書の理解 -SDKやAPIの概要理解 -クラウドストレージにファイルを格納できる など ・数十万件程度のデータ加工技術を有している -ソート、結合、集計、フィルタリングができる -設計書に基づき、プログラム実装できる ・適切な指示のもとに、以下を実施できる -データベースから条件を満たすデータを抽出できる -インポート、レコード挿入、エクスポート - システムや予測モデルのテスト実施 ・セキュリティの基礎知識を有している - 機密性、可用性、完全性の3要素 - 暗号化、認証、認可の理解 - マルウェア、コンピュータウィルス、改ざんの脅威を理解 など
DS以前の方	・ビジネスは勘と経験だけで回すものだと思っている ・課題を解決する際に、そもそも定量化する意識が無い ・データに付帯する権利や個人情報についての意識がない ・とりあえずAIを使えば大抵の課題解決ができると思っている	・基本統計量の意味を正しく理解していない ・線形代数や微分・積分の基本が理解できていない ・指数を指数で割り算したりする ・「平均年収」をそのまま鵜呑みにしたりする ・グラフ・チャートの使い方が不適切	・スプレッドシートで関数が使用できない ・繰り返しや分岐による処理フローが理解できない ・自分が使っているツールの仕組みに興味がない ・テキストや画像はデータでないと思っている ・データは自動的に整備されていると思っている