データサイエンティストのスキルレベル 2017年版

	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 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行お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 ・解決に必要な結果を総合した上で、説得力ある形で共有し、関連する組織、人を動かし、知見の横展開、組織を超えるつなぎ込みができる ・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、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	・予測、グルーピング、機械学習、大量データの可視化、言語処理、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、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 -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 -適切な学習データとテストデータの作成 ・基礎的な分析活動については、自律的に実施できる -多重共線性を考慮した重回帰分析 -パラメトリックな2群の検定の活用（t検定） -適切な初期値設定を行った非階層クラスタ-分析 -主成分分析や因子分析 -機械学習における過学習の理解 -形態素解析などを用いた基本的文書構造解析 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、論理データモデル作成 -HadoopやSparkでの管理対象データ選定 -SDKやAPI、ライブラリなどの適切な活用 -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 など ・深層学習（ディープラーニング）の学習を高速化するために、GPU（GPGPU）環境を設計・実装できる ・データ匿名化方法の理解と加工処理の設計ができる
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リとモレの判断力 -分析前の目的、ゴール設定 -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、ベイズの定理など） ・データ分析の基礎知識を有している -予測（回帰係数、標準誤差‥） -検定（帰無仮説、対立仮説‥） -グルーピング（教師あり学習、教師なし学習‥） など ・適切な指示のもとに、データ加工を実施できる -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） ・データ可視化の基礎知識を有している （ヒストグラム、散布図、積み上げ棒グラフなど）	・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 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 -SQLで簡単なSELECT文を記述・実行できる -設計書に基き、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 -同種のデータを統合するシステムの設計 -インポート、レコード挿入、エクスポート ・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	・基本統計量の意味を正しく理解していない ・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	・レポートされてくる数値サマリに目は通すが、特に記憶には残らない ・アクセス解析システムを使っていない ・ExcelやAccessは数字しか入れない

ver 2.00（as of 10/23/2017)