

データサイエンティスト

/// スキルチェックリスト ver4.00 ///

データサイエンティスト スキルチェックリスト

「データサイエンティスト スキルチェックリスト」は、データサイエンティストに必要とされるスキルをチェックリスト化したものです。チェックリスト内の各項目のスキルレベルは、★の数によって示し、見習いレベル（Assistant Data Scientist）～ 棟梁レベル（Full Data Scientist）までを判定します。

スキルレベル		判定基準
① Senior Data Scientist（業界を代表するレベル）	★★★★	－
② Full Data Scientist（棟梁レベル）	★★★	★★★の全項目のうち、50%を満たしている。
③ Associate Data Scientist（独り立ちレベル）	★★	★★の全項目のうち、60%を満たしている。
④ Assistant Data Scientist（見習いレベル）	★	★の全項目のうち、70%を満たしている。

※「必須スキル」に○がついている項目は、判定基準を満たしていても、この項目が達成されていないとそのレベルとは認められない項目として設定しています。

※ 独り立ちレベル以上のレベルは、下位のレベルを満たしていることが前提となります。

＊引用・改変時の注意事項

- ・チェックリストを引用される場合「一般社団法人データサイエンティスト協会 スキルチェックリストより引用」と引用元を明示願います。その他については著作権法に従っての引用を願います。
- ・チェックリストを改変された場合「一般社団法人データサイエンティスト協会 スキルチェックリストを改変」と改変した旨、明示願います。

スキルカテゴリー一覧							
データサイエンスカ	1	基礎数学	27	データエンジニアリングカ	1	環境構築	30
	2	データの理解・検証	25		2	データ収集	19
	3	意味合いの抽出、洞察	3		3	データ構造	11
	4	予測*	20		4	データ蓄積	17
	5	推定・検定	7		5	データ加工	15
	6	グルーピング	12		6	データ共有	15
	7	性質・関係性の把握	22			プログラミング	23
	8	サンプリング	5		7	ITセキュリティ	19
	9	データ加工	14		8	AIシステム運用	10
	10	データ可視化	38		項目数		159
	11	時系列分析	8	ビジネスカ	1	行動規範	16
	12	学習	49		2	契約・権利保護	8
	13	自然言語処理	16		3	論理的思考	18
	14	画像・映像認識	12		4	着想・デザイン	10
	15	音声認識	6		5	課題の定義	11
	16	パターン発見	4		6	アプローチ設計	16
	17	シミュレーション・データ同化	5		7	データ理解	6
	18	最適化	9		8	分析評価	4
	項目数		282			事業への実装	7
					9	PJマネジメント	26
			10		組織マネジメント	9	
			項目数		131		
項目数合計							572

＊データサイエンスカ 予測カテゴリーについての補足
予測スキルのうち、機械学習に関連するスキルは「学習」スキルに含めています

<<< データサイエンティスト スキルチェックリストについて >>>

【各シートの説明】

■ スキルレベル定義

- ・本リストでは、データサイエンティストに求められるスキルを「ビジネスカ」「データサイエンスカ」「データエンジニアリングカ」の3分野に分けて定義しているほか、それぞれの分野のスキルレベルとして、①業界を代表するレベル、②棟梁レベル、③独り立ちレベル、④見習いレベルの4段階を定義しています。「スキルレベル定義」では、各レベルの人材に求められるスキルのレベルを説明しています。

■ チェックリスト説明

- ・スキルチェックリストの見方のほか、各分野のサブカテゴリの一覧を掲載しています。また、スキルレベルを満たす要件について、リストを作成しています。

■ ビジネスカ／データサイエンスカ／データエンジニアリングカ

- ・ビジネスカ、データサイエンスカ、データエンジニアリングカ別にサブカテゴリを設定し、スキルレベルを満たす要件について説明しています。

■ 参考資料（データ可視化） sheet

- ・スキルチェックリスト：データサイエンスカの「データ可視化」サブカテゴリに関するイメージ共有のために参考資料(2021年11月19日時点)を例示しています。

データサイエンティストのスキルレベル 2021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、活用領域の開拓 - 組織・会社・産業を横断したデータカ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たデータ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・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む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 標準レベル	- 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 - 自身に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、実行 - 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証結果を取りまとめ、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（標準レベル）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 - 適切なクロスファンクショナルな非階層クロスファンクショナル分析 - 適切なクロスファンクショナルな非階層クロスファンクショナル分析 - 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成分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（標準レベル）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増進 - 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の基本的な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、必要なデータを入力し、分析、取りまとめることができる - データや事象のガバナンスの判断力 - 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 - 深層学習のメカニズムに関する理解など - 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 - 軸なし - 不適切な表現の理解	- 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図やデータモデル定義書の理解 - 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、コンピュータウイルス、改ざんの脅威を理解など
DS以前の方	- ビジネスは勤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ていると思っている	- 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ます 「平均年取」をそのまま割合みたりする - 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えない - 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

スキルチェックリスト 2021年版 <ビジネスカ>

▼ 他分野寄りのスキル

NO	SubNo	スキルカテゴリ	スキルレベル	サブカテゴリ	チェック項目	DE	DS	必須スキル
1	1	行動規範	★	ビジネスマインド	ビジネスにおける「論理とデータの重要性」を認識し、分析的でデータドリブンな考え方にに基づき行動できる			○
2	2	行動規範	★	ビジネスマインド	「目的やゴールの設定がないままデータを分析しても、意味合いが出ない」ことを理解している			○
3	3	行動規範	★	ビジネスマインド	課題や仮説を言語化することの重要性を理解している			○
4	4	行動規範	★	ビジネスマインド	現場に向かいヒアリングするなど、一次情報に接することの重要性を理解している			○
5	5	行動規範	★★	ビジネスマインド	社会における変化や技術の進化など、外的要因による分析プロジェクトへの影響をある程度見通し、柔軟に行動できる			
6	6	行動規範	★★	ビジネスマインド	ビジネスではスピード感がより重要であることを認識し、時間と情報が限られた状況下でも、言わば「ザッカリ感」を持って素早く意思決定を行うことができる			○
7	7	行動規範	★★	ビジネスマインド	作業ありきではなく、本質的な問題（イシュー）ありきで行動できる			○
8	8	行動規範	★★	ビジネスマインド	分析で価値ある結果を出すためには、しばしば仮説検証の繰り返しが必要であることを理解し、粘り強くタスクを完遂できる			○
9	9	行動規範	★★★	ビジネスマインド	プロフェッショナルとして、作業量ではなく生み出す価値視点で常に判断・行動でき、真に価値あるアウトプットを生み出すことにコミットできる			○
10	10	行動規範	★	データ・AI倫理	データを取り扱う人間として相応しい倫理を身に着けている（データのねつ造、改ざん、盗用を行わないなど）			○
11	11	行動規範	★	データ・AI倫理	データ、AI、機械学習の意図的な悪用（フェイクニュース、Botの悪用など）があり得ることを勘案し、技術に関する適切な知識と倫理を身につけている			○
12	12	行動規範	★★	データ・AI倫理	AI・機械学習がもたらす現在の倫理課題を説明できる（ディープフェイクによるプライバシーの侵害、バイアスによる人種差別、学習済みモデルのリバースエンジニアリングによる知的財産権の侵害など）	*		○
13	13	行動規範	★★★	データ・AI倫理	会社や組織全体におけるデータの取り扱いに関する倫理を維持・向上させるために、必要な制度や仕組みを策定し、その運営を主導することができる			
14	14	行動規範	★	コンプライアンス	直近の個人情報に関する法令（個人情報保護法、EU一般データ保護規則：GDPRなど）や、匿名加工情報の概要を理解し、守るべきポイントを説明できる			○
15	15	行動規範	★★	コンプライアンス	担当するビジネスや業界に係る直近の法令・ガイドラインを理解しており、データの保持期間や運用ルールに活かすことができる	*		
16	16	行動規範	★★	コンプライアンス	個人情報の扱いに関する法令、その他のプライバシーの問題、依頼元との契約約款に基づき、仮名化/匿名化すべきデータを選別できる（名寄せにより個人を特定できるもの、依頼元がデータ処理の結果をどのように保持し利用するのかなどの考慮）	*		
17	1	契約・権利保護	★	契約	二者間で交わされる一般的な契約の概念を理解している（請負契約と準委任契約の役務や成果物の違いなど）			
18	2	契約・権利保護	★★	契約	生成されたデータや学習済みモデルに関する権利保護に必要な法令を考慮し対処できる（契約法、特許法、著作権法、不正競争防止法など）			
19	3	契約・権利保護	★★	契約	性能保証を求められた際に、分析で作るモデルでは一般的に性能保証できないことを伝える、もしくは、事前に評価方法を定義するなど契約に盛り込むことができる	*		
20	4	契約・権利保護	★★★	契約	AI・モデル開発における生成されたデータや学習済みモデルに対し、成果物に対する責任の所在、権利の帰属を適切な専門家と協力しつつ、契約内容に盛り込むことができる（著作、利用許諾、営業機密、情報開示、利用範囲など）	*		
21	5	契約・権利保護	★★	契約	分析基盤実装などの開発・運用や、開発済み分析モデルの運用について、品質、可用性、責任などの観点で契約（SLA: Service Level Agreement など）にまとめることができる			
22	6	契約・権利保護	★★	権利保護	AI・モデル開発において、既存ライブラリを活用した場合の知財リスクの確認や、適切なガイドラインを参照・確認できる（経済産業省「AI・データの利用に関する契約ガイドライン」など）	*	*	○
23	7	契約・権利保護	★★★	権利保護	AI・モデルの活用・責任範囲に関し事業・現場に即したガイドラインを定義できる	*	*	
24	8	契約・権利保護	★★★	権利保護	手法・アルゴリズム構築時に、他者の権利を侵害をしない知財リスク管理や、他者からの権利侵害に備えた特許出願やデータ保全を含む適切な対応ができる			
25	1	論理的思考	★	MECE	データや事象の重複に気づくことができる			○
26	2	論理的思考	★★	MECE	初見の領域に対して、抜け漏れや重複をなくすことができる			
27	3	論理的思考	★★★	MECE	未知の領域であっても、類似する事象の推測などを活用し、抜け漏れや重複をなくすことができる			
28	4	論理的思考	★	構造化能力	与えられた分析課題に対し、初動として様々な情報を収集し、大まかな構造を把握することの重要性を理解している			○
29	5	論理的思考	★★	構造化能力	様々なデータや事象を、階層やグルーピングによって、適切に構造化できる			○
30	6	論理的思考	★	言語化能力	対象となる事象が通常見受けられる場合において、分析結果の意味合いを正しく言語化できる			○
31	7	論理的思考	★★	言語化能力	対象となる事象が通常見受けられない場合においても、分析結果の意味合いを既知の表現を組み合わせ、言語化できる			
32	8	論理的思考	★★★	言語化能力	データ表現に適した言葉がない場合でも、共通認識が形成できるような言葉を新たに作り出すことができる			
33	9	論理的思考	★	ストーリーライン	一般的な論文構成について理解している（序論⇒アプローチ⇒検討結果⇒考察や、序論⇒本論⇒結論 など）			○
34	10	論理的思考	★★	ストーリーライン	因果関係に基づいて、ストーリーラインを作ることができる（観察⇒気づき⇒打ち手、So What?, Why So?など）			○
35	11	論理的思考	★★★	ストーリーライン	相手や内容に応じて、自在にストーリーラインを組み上げることができる			
36	12	論理的思考	★	ドキュメンテーション	データの出自や情報の引用元に対する信頼性を適切に判断し、レポートに記載できる			○
37	13	論理的思考	★	ドキュメンテーション	1つの図表～数枚程度のドキュメントを論理立ててまとめることができる（課題背景、アプローチ、検討結果、意味合い、ネクストステップ）			○
38	14	論理的思考	★★	ドキュメンテーション	10～20枚程度のミニパッケージ（テキスト&図表）、もしくは5ページ程度の図表込みのビジネスレポートを論理立てて作成できる			○
39	15	論理的思考	★★★	ドキュメンテーション	30～50枚程度のフルパッケージ（テキスト&図表）、もしくは10ページ以上のビジネスレポートを論理立てて作成できる		*	
40	16	論理的思考	★	説明能力	報告に対する論拠不足や論理破綻を指摘された際に、相手の主張をすみやかに理解できる			○
41	17	論理的思考	★★	説明能力	論理的なプレゼンテーションができる			○
42	18	論理的思考	★★★	説明能力	プレゼンテーションの相手からの質問や反論に対して、説得力のある形で回答できる	*		○
43	1	着想・デザイン	★★	着想	ユーザーの視点に基づき、経験や体験を捉え、課題の発見や解決策を考えることで、データおよびテクノロジーを活用したビジネスモデルの着想ができる			
44	2	着想・デザイン	★★★	着想	企業の枠を超えて他社や大学、地方自治体、研究機関、起業家などが持つ技術やアイデア、サービスなどを組み合わせることで、事業にもたらす新たな価値を着想できる			

NO	SubNo	スキルカテゴリ	スキルレベル	サブカテゴリ	チェック項目	DE	DS	必須スキル
45	3	着想・デザイン	★★	デザイン	プライバシー・バイ・デザインやプライバシー影響評価（PIA:Privacy Impact Assessment）などを意識し、ユーザー向けサービスのデザインができる			
46	4	着想・デザイン	★★★	デザイン	データやAIの利活用において関係者や利用者の共感を得ることができるようなビジョンやコンセプトを設計し提言できる			
47	5	着想・デザイン	★	AI活用検討	弱いAI・強いAI、特化型AI・汎用型AIの違いを説明できる			
48	6	着想・デザイン	★★	AI活用検討	広く知られた学習済みモデルやライブラリについて、課題に適用できるか検討できる（RPAの業務適用、学習済みモデルの適用など）	*		
49	7	着想・デザイン	★★	AI活用検討	直近の世界的なAI倫理やプライバシー保護情勢を考慮し、担当事業における適切なデータ・AI活用を検討できる			
50	8	着想・デザイン	★★★	AI活用検討	直近の世界的なAI倫理やプライバシー保護情勢を考慮し、適切なサービス規定、データ処理工程、システム環境を考案できる	*	*	
51	9	着想・デザイン	★★★	AI活用検討	利害の異なる組織や事業を横断した課題に対して、AI・データの技術を用いた解決の枠組みを設計できる	*	*	
52	10	着想・デザイン	★★★	開示・非開示の決定	他社による模倣を防ぎ、競争力を保つ観点で、自社と市場の双方にメリットある開示方法を選択できる（自社にクローズした利用、製品に組み込んだ販売、APIとしての提供など）			
53	1	課題の定義	★	KPI	担当する分析プロジェクトにおいて、当該事業の収益モデルと主要な変数（KPI）を理解している			
54	2	課題の定義	★★	KPI	自らが関連する事業領域であれば、複数の課題レイヤーにまたがっていても、KPIを整理・構造化できる			
55	3	課題の定義	★★★	KPI	初見の事業領域であっても、KPIを構造化し、重要なKPIを見極められる			
56	4	課題の定義	★	スコーピング	担当する事業領域について、市場規模、主要なプレーヤー、支配的なビジネスモデル、課題と機会について説明できる			
57	5	課題の定義	★	スコーピング	主に担当する事業領域であれば、取り扱う課題領域に対して基本的な課題の枠組みが理解できる（調達活動の5フォースでの整理、CRM課題のRFMでの整理など）			○
58	6	課題の定義	★	スコーピング	既知の事業領域の分析プロジェクトにおいて、分析のスコープが理解できる			○
59	7	課題の定義	★★	スコーピング	事業モデルやバリューチェーンなどの特徴や事業領域の主たる課題を自力で構造的に理解でき、問題の大きさを整理できる			
60	8	課題の定義	★★★	スコーピング	事業領域の主要課題を他領域の課題との連関も含めて構造的に理解でき、問題の大きさを定義できる			
61	9	課題の定義	★★★	スコーピング	論理的な整理にとらわれず、批判的・複合的な視点で課題を識別できる			
62	10	課題の定義	★★★	スコーピング	仮説や可視化された問題がなくとも、解くべき課題を構造的に整理でき、見極めるべき論点を特定できる			○
63	11	課題の定義	★★★	価値の見積り	プロジェクト開始時点で、入手可能なデータ、分析手法、インフラ、ツールの生み出すビジネス価値を見積ることができる	*	*	
64	1	アプローチ設計	★	データ入手	仮説や既知の問題と与えられた中で、必要なデータにあたりをつけ、アクセスを確保できる	*	*	○
65	2	アプローチ設計	★★	データ入手	自身が担当するプロジェクトやサービスを越えて、必要なデータのあたりをつけ、アクセスを確保できる	*	*	
66	3	アプローチ設計	★★★	データ入手	組織全体及び関連する社外のデータを見渡して、必要なデータのあたりをつけ、アクセスを確保できる	*	*	
67	4	アプローチ設計	★★	AI-ready	検討目的に応じてこれまでデータ化されたことのないものを適切にデータ化するためのデータ取得とデータ構造の設計を行うことができる	*	*	
68	5	アプローチ設計	★★★	AI-ready	ドメイン特有のデータ構造と意味合いを理解した上で、データ発生からデータ利活用までの流れを俯瞰し、適切にノイズ、データの汚れの発生を抑制することができる	*	*	
69	6	アプローチ設計	★★★	AI-ready	業界、業務で求められるビジネス目的やモデル精度に応じて、必要なデータの品質、取得経路や範囲、契約条件などを勘案し、データの入手経路の確保と加工・運用設計ができる	*	*	
70	7	アプローチ設計	★★	アプローチ設計	分析の目的を検証すべき項目に分解し、アウトプットとなる比較結果やモデル作成の結果のイメージを描くことができる		*	○
71	8	アプローチ設計	★★	アプローチ設計	分析の目的に対して、プロジェクトの目標と評価方法を具体化し、定量的な成功基準を設定するとともに、成功基準の判定時期・判定者を決定できる			
72	9	アプローチ設計	★★	アプローチ設計	仮説検証思考で、論点ごとに検証すべき項目を識別できる			○
73	10	アプローチ設計	★★	アプローチ設計	最終的な結論に関わる部分や、ストーリーラインの骨格に大きな影響を持つ部分から着手するなど、検証すべき項目の優先度を判断できる		*	
74	11	アプローチ設計	★★★	アプローチ設計	データサイエンスを駆使して解くべき課題が否かを判断できる	*	*	○
75	12	アプローチ設計	★	分析アプローチ設計	スコープ、検討範囲・内容が明確に設定されていれば、必要な分析プロセスが理解できる（データ、分析手法、可視化の方法など）		*	
76	13	アプローチ設計	★★	分析アプローチ設計	解くべき課題がフレーミングされていれば、必要なデータ取得のあり方、粒度、サンプリングのあり方などを設計できる		*	
77	14	アプローチ設計	★★	分析アプローチ設計	解くべき課題がフレーミングされていれば、必要な分析手法、可視化などを適切に選択できる		*	○
78	15	アプローチ設計	★★★	分析アプローチ設計	複数の事業や課題にまたがっていても、必要なデータ、分析手法、可視化などを適切に選択し作業手順に落としこめる	*	*	
79	16	アプローチ設計	★★★	分析アプローチ設計	複数のアプローチの組み合わせでしか解けない課題であっても、その解決までの道筋を設計できる	*	*	
80	1	データ理解	★	データ理解	ビジネス観点で仮説を持ってデータをみることの重要性和、仮に仮説と異なる結果となった場合にも、それが重大な知見である可能性を理解している			○
81	2	データ理解	★★	データ理解	統計手法を用いる際の閾値の設定に対して、ビジネス観点で納得感のある調整ができる（年齢の刻み、商品単価、購入周期を考慮した量的変数のカテゴリーなど）		*	
82	3	データ理解	★★★	データ理解	分析プロセス全体を通して、ビジネス観点での妥当性をチェックし、データから得られた示唆が価値ある知見であるかを都度判断できる			○
83	4	データ理解	★	意味合いの抽出、洞察	分析結果を元に、起きている事象の背景や意味合い（真実）を見抜くことができる		*	
84	5	データ理解	★★	意味合いの抽出、洞察	分析結果を元に、特異点、相違性、傾向性、関連性を見出した上で、ビジネス上の意味を捉えるためにドメイン知識を持つ人に適切な質問を投げかけられる		*	
85	6	データ理解	★★	意味合いの抽出、洞察	分析結果を元に、意味合いの明確化に向けた分析の深掘り、分析対象データの見直しについて方向性を設計できる		*	
86	1	分析評価	★★	評価	担当する分析プロジェクトの分析結果を見て検討目的と合っているか評価できる		*	○
87	2	分析評価	★★	評価	分析結果が当初の目的を満たしていない場合に、問題を正しく理解し、目的達成に向けて必要な分析手順を追加・変更できる		*	○
88	3	分析評価	★★	業務へのフィードバック	分析的検討に基づき、担当業務に対する必要なアクション、改善案を整理して結論を導くことができる			
89	4	分析評価	★★★	業務へのフィードバック	分析的検討に基づき、経営レベルで必要なアクション、改革案を整理して結論を導くことができる			
90	1	事業への実装	★★	実装	現場に実装する際、実行可能性を考慮し適切に対応できる（AI活用に関する基礎理解促進、業務マニュアルの改訂・浸透や、現場のトレーニングなど）	*		○
91	2	事業への実装	★★	実装	担当する案件が予算内で解決するように取り組みをデザインし、現場に実装できる	*		
92	3	事業への実装	★★	実装	異なるスキル分野の専門家や事業者と適切なコミュニケーションをとりながら事業・現場への実装を進めることができる			
93	4	事業への実装	★★★	実装	費用対効果、実行可能性、業務負荷を考慮し事業に実装ができる	*		

NO	SubNo	スキルカテゴリ	スキルレベル	サブカテゴリ	チェック項目	DE	DS	必須スキル
94	5	事業への実装	★	評価・改善の仕組み	結果、改善の度合いをモニタリングする重要性を理解している			○
95	6	事業への実装	★★	評価・改善の仕組み	事業・現場へ実装するにあたりモニタリングの仕組みを適切に組み込むことができる	*	*	
96	7	事業への実装	★★★	評価・改善の仕組み	既存のPDCAサイクルに対し、次の改善的な取り組みにつなげることができる			
97	1	PJマネジメント	★	プロジェクト発足	プロジェクトにおけるステークホルダーや役割分担、プロジェクト管理・進行に関するツール・方法論が理解できる	*		
98	2	PJマネジメント	★★	プロジェクト発足	アジャイル開発体制のポイントを理解した上で、アジャイルな開発チームを迅速に立ち上げ、推進できる	*		
99	3	PJマネジメント	★★	プロジェクト発足	類似事例の実績やProof of Concept（PoC）を適宜利用して、プロジェクト計画に関わるステークホルダーとの合意を形成できる			
100	4	PJマネジメント	★★	プロジェクト発足	PoCのみで終わらないように、PoCプロジェクト立ち上げ時点で実務実装を想定した計画を策定できる			
101	5	PJマネジメント	★★	プロジェクト計画	ビジネス要件を整理し、分析・データ活用プロジェクトを企画・提案できる	*	*	○
102	6	PJマネジメント	★★	プロジェクト計画	プロジェクト立ち上げに向けて、目的・ゴールに沿ったアプローチ、成果物、納期、リソース配分を整理し、WBSに落とし込むことができる	*		
103	7	PJマネジメント	★★★	プロジェクト計画	依頼元やステークホルダーのビジネスをデータ面から理解し、分析・データ活用プロジェクトを立ち上げ、プロジェクトにかかるコストと依頼元の利益を説明できる	*	*	
104	8	PJマネジメント	★★	プロジェクト計画	分析プロジェクトのデータ、分析結果の中から、どれを顧客、外部に開示すべきか、あらかじめ判断できる			
105	9	PJマネジメント	★★★	運用	運用しているサービスやシステムに対して、稼働状況や情報漏洩の監視などを実施し、適切にサービスレベルを維持・管理できる	*	*	
106	10	PJマネジメント	★★★	横展開	特定のビジネス課題に向けた新しいソリューションを個別の現場の特性を考慮し横展開できる	*	*	
107	11	PJマネジメント	★★★	方針転換	プロジェクトの進捗や達成状況が芳しくない場合や、想定外の事象が起こった場合に、リカバリープランを見極めた上で、時には大幅な方針転換や終了の判断ができる	*	*	○
108	12	PJマネジメント	★★★	完了	プロジェクトやサービスが終了した場合に、契約内容に応じたデータ削除や運用管理体制の引継ぎ/終了等、適切にプロジェクト完了処理ができる	*		
109	13	PJマネジメント	★	リソースマネジメント	指示に従ってスケジュールを守り、チームリーダーに頼まれた自分の仕事を完遂できる			○
110	14	PJマネジメント	★★	リソースマネジメント	自身とチームメンバーのスキルを把握し、適切なプロジェクト管理ができる			
111	15	PJマネジメント	★★	リソースマネジメント	プロジェクトに設定された予算やツール、システム環境を適切に活用し、プロジェクトを進行できる	*	*	○
112	16	PJマネジメント	★★	リソースマネジメント	5名前後のチームをスケジュール通りに進行させ、ステークホルダーに対して、期待値に見合うアウトプットを安定的に生み出せる			
113	17	PJマネジメント	★★	リソースマネジメント	プロジェクトメンバーの特性を見極め、適切な業務範囲を設計し、曖昧な指示で終わらせず、明確な指示出しができる			
114	18	PJマネジメント	★★★	リソースマネジメント	プロジェクトに求められるスキル要件と各メンバーのスキル・成長目標・性格をふまえ、現実的にトレードオフ解消とシナジーを狙ったリソースマネジメントができる	*	*	
115	19	PJマネジメント	★★★	リソースマネジメント	プロジェクトメンバーの技量を把握した上で、プロジェクト完遂に必要なツール選定、予算策定、スコープ設定、またはアウトソーシング体制を検討・構築できる	*	*	
116	20	PJマネジメント	★★★	リソースマネジメント	複数のチームから編成されるプロジェクトにおいて、スケジュール通りに進行させ、複合的なステークホルダーに対し、期待値を超えたアウトプットを安定的に生み出せる	*	*	
117	21	PJマネジメント	★	リスクマネジメント	担当するタスクの遅延や障害などを発見した場合、迅速かつ適切に報告ができる			○
118	22	PJマネジメント	★★	リスクマネジメント	プロジェクトでの遅延や障害などの発生を検知し、リカバリーするための提案・設計ができる	*	*	
119	23	PJマネジメント	★★★	リスクマネジメント	期待される成果が達成できないケースを早期に見極め、プロジェクトの終了条件をステークホルダーと整理・合意できる			
120	24	PJマネジメント	★★★	リスクマネジメント	プロジェクトの推進に深く影響するようなリスクを早期に察知し、適切にギャップを埋め、つなぎ直し、それに合わせチームの再編成も随時行い、障害などの発生を大半を事前に抑制できる			
121	25	PJマネジメント	★★★	リスクマネジメント	プロジェクトに何らかの遅延・障害などが発生した場合、適切なリカバリー手順の判断、リカバリー体制構築、プロジェクトオーナーに対する迅速な対応ができる			○
122	26	PJマネジメント	★★★	リスクマネジメント	マルウェア、DDoS攻撃などの深刻なセキュリティ攻撃を受けた場合に対応する最新の技術を把握し、対応する専門組織（CSIRT）の構成を責任者にすみやかに提案できる	*		
123	1	組織マネジメント	★★	育成/ナレッジ共有	自身とチームメンバーのスキルを大まかに把握し、担当するプロジェクトを通して、チームメンバーへのスキル成長のためのアドバイスや目標管理ができる	*	*	
124	2	組織マネジメント	★★	育成/ナレッジ共有	チームメンバーのスキルに応じ、研修参加や情報収集への適切なアドバイスやチーム内でのナレッジ共有を推進できる	*	*	
125	3	組織マネジメント	★★	育成/ナレッジ共有	チーム全員がデータを取り扱う人間として相応しい倫理を持てるよう、適切にチームを管理できる			
126	4	組織マネジメント	★★★	育成/ナレッジ共有	チームの各メンバーが不安や恥ずかしさを感じることなくチャレンジすることができ、積極的に失敗を共有し改善に向けて問題提起と話し合いを行う、開かれた雰囲気醸成できる			
127	5	組織マネジメント	★★★	育成/ナレッジ共有	チームの各メンバーに対し、データサイエンティストとしてのスキル目標の設定、到達させるための適切なアドバイスができる	*	*	○
128	6	組織マネジメント	★★★	育成/ナレッジ共有	データサイエンティストに求められるスキルについて、育成制度の設計やナレッジ共有の仕組み構築と運営ができる	*	*	
129	7	組織マネジメント	★★★	育成/ナレッジ共有	チームに必要な情報やデータサイエンスの新しい技術・手法に関する情報収集方法や学習方法を主導し、自ら情報を取捨選択し、チームにフィードバックできる	*	*	
130	8	組織マネジメント	★★	組織マネジメント	所属する組織全体におけるデータサイエンスチームの役割を認識し、担当するプロジェクトにおいて、組織内や他部門・他社間でのタスク設定や調整ができる			
131	9	組織マネジメント	★★★	組織マネジメント	データサイエンスチームを自社・他社の様々な組織と関連付け、対象組織内での役割の規定、目標設定を行うことができる			○

スキルチェックリスト 2021年版 <データサイエンスカ>

▼ 他分野寄りのスキル

NO	SubNo	スキルカテゴリ	スキルレベル	サブカテゴリ	チェック項目	BZ	DE	必須スキル
1	1	基礎数学	★	統計数理基礎	順列や組合せの式 nPr , nCr を理解し、適切に使い分けることができる			○
2	2	基礎数学	★	統計数理基礎	確率に関する基本的な概念の意味を説明できる（確率、条件付き確率、期待値、独立など）			○
3	3	基礎数学	★	統計数理基礎	平均、中央値、最頻値の算出方法の違いを説明できる			○
4	4	基礎数学	★	統計数理基礎	与えられたデータにおける分散、標準偏差、四分位、パーセンタイルを理解し、目的に応じて適切に使い分けることができる			○
5	5	基礎数学	★	統計数理基礎	母（集団）平均と標本平均、不偏分散と標本分散がそれぞれ異なることを説明できる			○
6	6	基礎数学	★	統計数理基礎	標準正規分布の平均と分散の値を知っている			○
7	7	基礎数学	★	統計数理基礎	相関関係と因果関係の違いを説明できる			○
8	8	基礎数学	★	統計数理基礎	名義尺度、順序尺度、間隔尺度、比例尺度の違いを説明できる			○
9	9	基礎数学	★	統計数理基礎	ピアソンの相関係数の分母と分子を説明できる			○
10	10	基礎数学	★	統計数理基礎	5つ以上の代表的な確率分布を説明できる			○
11	11	基礎数学	★	統計数理基礎	二項分布は試行回数が増えていくどのような分布に近似されるかを知っている			○
12	12	基礎数学	★	統計数理基礎	変数が量的、質的どちらの場合でも関係の強さを算出できる			○
13	13	基礎数学	★	統計数理基礎	指数関数とlog関数の関係を理解し、片対数グラフ、両対数グラフ、対数化されていないグラフを適切に使いわけることができる			○
14	14	基礎数学	★	統計数理基礎	ベイズの定理を説明できる			○
15	15	基礎数学	★★	統計数理基礎	ベイズ推論が学習や予測、モデル評価などをすべて確率分布上の計算問題として扱っていることにより、これらの要素を確率分布として扱わない手法と比べ、どのようなメリットを生み出しているか理解している			
16	16	基礎数学	★★	統計数理基礎	自己情報量やエントロピーの意味について説明できる			
17	17	基礎数学	★★	統計数理基礎	カルバック・ライブラー情報量（KL divergence）、フィッシャー情報量の意味や利用方法について説明できる			
18	18	基礎数学	★★	統計数理基礎	尤度と最尤推定についての説明ができる（尤度関数、ネイマンの分解定理、十分統計量）			
19	19	基礎数学	★	線形代数基礎	ベクトルの内積に関する計算方法を理解し、線形式をベクトルの内積で表現できる			○
20	20	基礎数学	★	線形代数基礎	行列同士、および行列とベクトルの計算方法を正しく理解し、複数の線形式を行列の積で表現できる			○
21	21	基礎数学	★	線形代数基礎	逆行列の定義、および逆行列を求めることにより行列表記された連立方程式が解けることを理解している			
22	22	基礎数学	★	線形代数基礎	固有ベクトルおよび固有値の意味を理解している			
23	23	基礎数学	★	微分・積分基礎	微分により計算する導関数が傾きを求めるための式であることを理解している			○
24	24	基礎数学	★	微分・積分基礎	2変数以上の関数における偏微分の計算方法を理解している			
25	25	基礎数学	★	微分・積分基礎	積分と面積の関係を理解し、確率密度関数を定積分することで確率が得られることを説明できる			○
26	26	基礎数学	★	集合論基礎	和集合、積集合、差集合、対称差集合、補集合についてベン図を用いて説明できる			○
27	27	基礎数学	★	集合論基礎	論理演算と集合演算の対応を理解している（ANDが積集合に対応するなど）			
28	1	データの理解・検証	★	統計情報への正しい理解	単なるローデータとしての実数だけを見て判断出来ない事象が大多数であり、母集団に占める割合などの比率的な指標でなければ数字の比較に意味がないことがわかっている			○
29	2	データの理解・検証	★	統計情報への正しい理解	ニュース記事などで統計情報に接したときに、数字やグラフの持つメッセージを理解できる	*		○
30	3	データの理解・検証	★★	統計情報への正しい理解	自身の判断に必要な統計情報を積極的に収集するとともに、表現に惑わされず数字を正当に評価できる（原点が0ではないグラフ、不要な3D化、不要な2軸化、目盛りの未記載など）	*		
31	4	データの理解・検証	★★★	統計情報への正しい理解	数字やデータの検証のために、何と比較するべきかすみやかに把握し、収集・利用できる（業務データや過去に接触した統計情報の想起・活用を含む）	*		
32	5	データの理解・検証	★	データ確認	単独のグラフに対して、集計ミスや記載ミスなどがないかチェックできる	*		○
33	6	データの理解・検証	★	データ確認	データ項目やデータの量・質について、指示のもと正しく検証し、結果を説明できる	*		
34	7	データの理解・検証	★★	データ確認	複数のグラフや集計表で構成されているレポートに対して、全体として集計ミスや不整合が起きていないかチェックできる	*		○
35	8	データの理解・検証	★★	データ確認	データ項目やデータの量・質の検証方法を計画・実行し、その結果をもとにその後の分析プロセスを立案・修正できる	*		
36	9	データの理解・検証	★★★	データ確認	多数のグラフ、集計表、外部の統計情報、高度なデータ解析手法を用いた解析結果などを含むレポートに対して、不整合が起きていないか、妥当性の高い論理構造であるかチェックできる	*		
37	10	データの理解・検証	★★★	データ確認	分析に必要なデータを想定し、現在取得可能なデータの量・質で分析に耐えうるか、分析目的が達成可能であるかを判断できる	*		○
38	11	データの理解・検証	★	俯瞰・メタ思考	データが生み出された背景を考え、鵜呑みにはしないことの重要性を理解している	*		
39	12	データの理解・検証	★★	俯瞰・メタ思考	データを俯瞰して、変化をすみやかに察知するとともに、変化が誤差の範囲かどうか判断できる	*		○
40	13	データの理解・検証	★★★	俯瞰・メタ思考	複数のデータを多角的かつ大局的に俯瞰して、大きな動きや本質的な事実を見抜くことができる	*		○
41	14	データの理解・検証	★	データ理解	どのような知見を得たいのか、目的に即して集計し、データから事実を把握できる	*		○
42	15	データの理解・検証	★	データ理解	データから事実を正しく浮き彫りにするために、集計の切り口や比較対象の設定が重要であることを理解している	*		○
43	16	データの理解・検証	★	データ理解	普段業務で扱っているデータの発生トリガー・タイミング・頻度などを説明でき、また基本統計量や分布の形状を把握している	*		
44	17	データの理解・検証	★	データ理解	時系列データとは何か、その基礎的な扱いについて説明できる（時系列グラフによる周期性やトレンドの確認、移動平均の計算など）			
45	18	データの理解・検証	★★	データ理解	生データを眺めて、どのような切り口で集計・比較すればデータの理解や事実の把握につながるか検討できる	*		○
46	19	データの理解・検証	★★	データ理解	扱ったことのない新たなデータに内容の不明な項目があっても、生データの閲覧や集計を通して何の項目かあたりをつけられる	*		
47	20	データの理解・検証	★★	データ理解	扱っているデータの関連業務の知識と分析目的を踏まえて、どんな説明変数が効きそうか、あたりをつけて洗い出し、構造的に整理できる（変数のグループ化や階層化など）	*		
48	21	データの理解・検証	★★★	データ理解	データの変化から起きている事象の背景を構造的に推察し、仮説を立て、検証方法を企画実行できる	*		
49	22	データの理解・検証	★★★	データ理解	データを入手する前に、存在するであろうデータとその分布を想定して基礎俯瞰の方向性やその結果の想定ができ、それを前提とした解析方法の検討・ラフ設計をすることができる	*		
50	23	データの理解・検証	★★★	データ理解	扱ったことのない新たなデータであっても、ER図やテーブル定義、生データなどを見ることによってデータの発生源や欠損値の意味などのあたりをつけられる	*	*	
51	24	データの理解・検証	★★	データ粒度	分析目的とデータの量・質を踏まえて、想定されるメッセージと統計的観点から適切な集計単位やサンプリング率を決定できる	*		
52	25	データの理解・検証	★★★	データ粒度	分析目的とデータの量・質に加えて、想定しているメッセージ、深掘りの方向性・可能性、処理負荷、データ処理フローなども総合的に踏まえた最適な集計単位やサンプリング率を決定できる	*	*	
53	1	意味合いの抽出、洞察	★	洞察	分析、図表から直接的な意味合いを抽出できる（バラツキ、有意性、分布傾向、特異性、関連性、変曲点、関連度の高低など）	*		○
54	2	意味合いの抽出、洞察	★	洞察	想定に影響されず、分析結果の数値を客観的に解釈できる	*		
55	3	意味合いの抽出、洞察	★★	洞察	各種の解析手法（主成分分析、クラスター分析、決定木分析など）の結果を解釈し、意味合いを適切に表現・説明できる	*		○

NO	SubNo	スキルカテゴリ	スキルレベル	サブカテゴリ	チェック項目	BZ	DE	必須スキル
56	1	予測	★	回帰・分類	単回帰分析において最小二乗法、回帰係数、標準誤差、決定係数を理解し、モデルを構築できる			○
57	2	予測	★	回帰・分類	重回帰分析において偏回帰係数と標準偏回帰係数、重回帰係数について説明できる			
58	3	予測	★★	回帰・分類	重回帰分析や判別分析を実行する際に変数選択手法の特徴を理解し、適用できる			
59	4	予測	★★	回帰・分類	重回帰分析において多重共線性の評価ができ、適切に変数を統合・加工・削除して予測モデルが構築できる			○
60	5	予測	★	回帰・分類	線形回帰分析は量的な変数を予測し、ロジスティック回帰分析は二値の質的な変数を予測する手法であることを説明できる			
61	6	予測	★★	回帰・分類	過学習を防止するためにL1正則化（Lasso）、L2正則化（Ridge）、それらを組み合わせたElastic netを適用できる			
62	7	予測	★★	回帰・分類	一般化線形モデル（GLM）および一般化線形混合モデル（GLMM）を用いて分布形状に応じたモデルを構築できる			
63	8	予測	★★★	回帰・分類	対象データの分布をみて、分布形状に適合した計算式的非線形回帰モデルを構築できる			
64	9	予測	★★	回帰・分類	ロジスティック回帰分析において回帰係数とオッズ比の関係について説明できる			
65	10	予測	★★	回帰・分類	ロジスティック回帰分析を行う際に、最小二乗法ではなく最尤法を使う際の利点（回帰誤差が近似的に正規分布しなくても適用できるなど）を説明し、適用できる			
66	11	予測	★★	回帰・分類	Cox回帰（比例ハザードモデル）を用いて対象イベントの発生確率を予測する生存分析モデルを構築できる			
67	12	予測	★★	回帰・分類	対象の個体差やグループ毎の差が認められるデータに対し、階層ベイズモデルの構築ができる			
68	13	予測	★★	回帰・分類	予測アルゴリズムに応じ、必要な変数加工処理を設計し、実施できる（標準化やダミー変数化など、採用したアルゴリズムに適した変数加工）			
69	14	予測	★★★	回帰・分類	予測モデル構築において頑健性（Robustness）を維持するための具体的な方法を設計、実施できる			○
70	15	予測	★	評価	ROC曲線、AUC(Area under the curve)、を用いてモデルの精度を評価できる			
71	16	予測	★	評価	混同行列（正誤分布のクロス表）、Accuracy、Precision、Recall、F値、macro平均、micro平均、重み付き平均といった評価尺度を理解し、精度を評価できる			
72	17	予測	★	評価	RMSE（Root Mean Square Error）、MAE（Mean Absolute Error）、MAPE（Mean Absolute Percentage Error）、決定係数といった評価尺度を理解し、精度を評価できる			○
73	18	予測	★★	評価	不均衡データ（Imbalanced data）に対する分類モデルの評価尺度を、Precision、Recall、F値、PR（Precision Recall）曲線、マンユース相関係数などを用いてビジネス課題に合わせて適切に設定できる	*		
74	19	予測	★★★	評価	目的（予測・真のモデル推定など）に応じて、適切な損失関数とモデル選択基準（AIC：赤池情報量基準、BIC：ベイズ情報量基準、MDL：最小記述長など）を選択し、モデル評価ができる			
75	20	予測	★★★	評価	MSE、AUC、F値などは評価データ全体に対する平均的評価であることを理解し、必要に応じて予測値・誤差の可視化、データの部分集合に対する評価などを実施できる			
76	1	推定・検定	★	推定・検定	点推定と区間推定の違いを説明できる			
77	2	推定・検定	★	推定・検定	統計的仮説検定において帰無仮説と対立仮説の違いを説明できる			
78	3	推定・検定	★	推定・検定	第1種の過誤、第2種の過誤、p値、有意水準の意味を説明できる			
79	4	推定・検定	★	推定・検定	片側検定と両側検定の違いを説明できる			
80	5	推定・検定	★	推定・検定	検定する対象となるデータの対応の有無を考慮した上で適切な検定手法（t検定、z検定など）を選択し、適用できる			
81	6	推定・検定	★★	推定・検定	検定力やサンプルサイズ、分布など対象のデータを考慮したうえで適切な検定手法を選択し、結果を評価できる（パラメトリックな多群の検定、クラスカル・ウォリス検定、カイニ乗検定など）			
82	7	推定・検定	★★	推定・検定	p値だけでは仮説やモデルの正しさを評価できないことを理解し、p値以外のアプローチ（信頼区間、信用区間、ベイズファクターなど）と併せて透明性の高いデータ分析や結果の報告ができる	*		○
83	1	グルーピング	★	グルーピング	教師なし学習のグループ化（クラスター分析）と教師あり学習の分類（判別）モデルの違いを説明できる			
84	2	グルーピング	★	グルーピング	階層クラスター分析と非階層クラスター分析の違いを説明できる			
85	3	グルーピング	★	グルーピング	階層クラスター分析において、デンドログラムの見方を理解し、適切に解釈できる	*		
86	4	グルーピング	★★	グルーピング	非階層クラスター分析において、分析対象となるデータの特性や分析目的に応じ、適切なクラスター数を決定できる			○
87	5	グルーピング	★★	グルーピング	階層クラスター分析における代表的なクラスター間距離（Ward法、群平均法、最長一致法など）の概念を理解し、目的に合致した最適な手法で分析できる			
88	6	グルーピング	★★	グルーピング	クラスター分析を実行する際、各種距離関数（ユークリッド距離、マンハッタン距離、マハラノビス距離など）を理解し、分析目的に合致した最適な距離計算で分析できる	*		
89	7	グルーピング	★★	グルーピング	k-means法で得られる分析結果は局所最適解であるため初期値問題があることを理解し、適切な初期値を選択できる			
90	8	グルーピング	★★	グルーピング	k-meansやその派生方法、カーネルk-means、スペクトラルクラスタリングなどの違いを理解し、試行の中で最適な手法を選択・実行できる			
91	9	グルーピング	★★★	グルーピング	距離の公理を満たさない場合（1-cos類似度など）においてもクラスター分析を適切に実行できる			
92	10	グルーピング	★★★	グルーピング	自己組織化マップ（SOM）、Affinity Propagation、混合分布モデル、ディリクレ過程混合モデルなどを理解し、試行の中で最適な手法を選択できる			
93	11	グルーピング	★★	異常検知	手元のデータの多寡やデータの再現性をふまえ、教師あり異常検知と教師なし異常検知の違いを適切に使い分けことができる	*		
94	12	グルーピング	★★	異常検知	Local Outlier Factor（LOF）、オートエンコーダ、One-class SVM（Support Vector Machine）、マハラノビス距離などを用いた異常検知の手法を理解し、試行の中で最適な手法を選択・実行できる			
95	1	性質・関係性の把握	★	性質・関係性の把握	適切なデータ区間設定でヒストグラムを作成し、データのバラつき方を把握できる	*		○
96	2	性質・関係性の把握	★	性質・関係性の把握	適切な軸設定でクロス集計表を作成し、属性間のデータの偏りを把握できる	*		○
97	3	性質・関係性の把握	★	性質・関係性の把握	量的変数の散布図を描き、2変数の関係性を把握できる			○
98	4	性質・関係性の把握	★★	性質・関係性の把握	多重（質問間）クロス集計表などを駆使して、データから適切なインサイトを得ることができる（データの偏りの発見や独立性の検定など）	*		
99	5	性質・関係性の把握	★★	性質・関係性の把握	主成分分析と因子分析の違いや使い分けを説明し、実施できる			
100	6	性質・関係性の把握	★★	性質・関係性の把握	正準相関分析を説明し、実施できる			
101	7	性質・関係性の把握	★★	性質・関係性の把握	因子分析における、因子負荷量の意味を理解し、因子軸の回転（プロマックス回転、バリマックス回転など）について実施できる			
102	8	性質・関係性の把握	★★	性質・関係性の把握	コレスポンデンス（対応）分析と数量化III類を説明し、実施できる			
103	9	性質・関係性の把握	★★	性質・関係性の把握	コンジョイント分析を自ら設計し、効用値と寄与率からモデルを評価できる			
104	10	性質・関係性の把握	★★	性質・関係性の把握	適切な類似度を設定した上で、多次元尺度構成法を用いてポジショニングマップを描くことができる	*		
105	11	性質・関係性の把握	★★	性質・関係性の把握	空間的自己相関の手法を用いて空間的な類似性を数値化できる			
106	12	性質・関係性の把握	★★	性質・関係性の把握	行列分解（非負値行列因子分解[Non-negative Matrix Factorization:NMF]、特異値分解）を、目的に応じてパラメータを最適化して分析できる			

NO	SubNo	スキルカテゴリ	スキルレベル	サブカテゴリ	チェック項目	BZ	DE	必須スキル
107	13	性質・関係性の把握	★★★	性質・関係性の把握	テンソル分解（非負値テンソル因子分解[Non-negative Tensor Factorization:NTF]、CP分解[Canonical Polyadic Decomposition:CPD]、Tucker分解など）を、対象データの特性や目的に応じて適用する事で、より複合的な要因の性質や関係性を分析できる			
108	14	性質・関係性の把握	★★	グラフィカルモデル	グラフィカルモデルを用いて、確率変数間の関係性をグラフ表現できる			
109	15	性質・関係性の把握	★★	グラフィカルモデル	重回帰分析とパス解析の違いを説明できる			
110	16	性質・関係性の把握	★★	グラフィカルモデル	ベイジアンネットワーク分析結果から目的事象の事後確率を算出できる			
111	17	性質・関係性の把握	★★	グラフィカルモデル	共分散構造分析（構造方程式モデリング：SEM）を行い、観測変数・潜在変数の因果関係を説明できる			
112	18	性質・関係性の把握	★	因果推論	ある変数が他の変数に与える影響（因果効果）を推定したい場合、その双方に影響を与える共変量（交絡因子）の考慮が重要であると理解している（喫煙の有無と疾病発症の双方に年齢が影響している場合など）			
113	19	性質・関係性の把握	★	因果推論	分析の対象を定める段階で選択バイアスが生じる可能性があることを理解している（途中離脱者の除外時、欠損データの除外時など）			○
114	20	性質・関係性の把握	★★	因果推論	統計的因果推論における因果効果（平均処置効果、Average Treatment Effect; ATEなど）について説明できる			
115	21	性質・関係性の把握	★★	因果推論	ある変数の影響（因果効果）を推定したいがランダム化比較試験の実施が難しい場合、傾向スコアによる手法（傾向スコアマッチング、IPW、Doubly Robustなど）を用いることで観測されている共変量の影響を最小限に抑えることができる			
116	22	性質・関係性の把握	★★	因果推論	統計的因果推論における、強く無視できる割り当て条件（Conditional Exchangeability）を説明できる			
117	1	サンプリング	★	サンプリング	標本誤差およびサンプリングバイアス、およびそれぞれの違いについて説明できる			○
118	2	サンプリング	★	サンプリング	実験計画法の基本的な3原則（局所管理化、反復、無作為化）について説明できる			
119	3	サンプリング	★★	サンプリング	調査対象の母集団の規模・特性や調査コストに応じて、多段階抽出法や層化抽出法など適切な標本抽出方法を計画できる			
120	4	サンプリング	★★	サンプリング	属性数と水準数が決まれば適切な直交表を選択し実験計画ができる			
121	5	サンプリング	★★	サンプリング	調査に求められる信頼水準・誤差率から必要となるサンプル数を試算できる			
122	1	データ加工	★	データクレンジング	外れ値・異常値・欠損値とは何かを理解し、指示のもと適切に検出と除去・変換などの対応ができる	*		○
123	2	データ加工	★★	データクレンジング	各変数の分布・欠損率などをふまえて、外れ値・異常値・欠損値の対応方針を決定できる			○
124	3	データ加工	★	データ加工	標準化とは何かを理解し、適切に標準化が行える			
125	4	データ加工	★	データ加工	名義尺度の変数をダミー変数に変換できる			
126	5	データ加工	★★	データ加工	加工データに不具合がないか確認方法を設計し、検証できる（合計値の照合や、分布の確認など）	*		○
127	6	データ加工	★★★	データ加工	加工データの統計的な俯瞰によって不具合の早期発見ができるとともに、統計的観点で次ステップの解析に耐えるデータであるか評価できる			○
128	7	データ加工	★	特徴量エンジニアリング	数値データの特徴量化（二値化／離散化、対数変換、スケーリング／正規化、交互作用特徴量の作成など）を行うことができる		*	
129	8	データ加工	★★	特徴量エンジニアリング	カテゴリデータの特徴量化ができる（ダミー変数化、特徴量ハッスリング[カテゴリ数が多い場合の値の割り当て]、ヒックカウンティング、バックオフ、最小カウントスケッチ[レアなカテゴリのまとめ処理]など）		*	
130	9	データ加工	★★	特徴量エンジニアリング	k-means、k近傍法、t-SNEなどを用いた、類似度または非類似度に着目した特徴量を作成できる			
131	10	データ加工	★★	特徴量エンジニアリング	周期的な意味合いをもつ特徴量を適切な周期の三角関数を用いて変換できる			
132	11	データ加工	★★	特徴量エンジニアリング	時系列データに対し周期性やトレンドなどを考慮して、階差、移動平均、移動標準偏差、移動最大、移動最小、ラグなどの特徴量を作成できる	*		
133	12	データ加工	★★	特徴量エンジニアリング	複数のデータセットから、結合や集計などを組み合わせ、有効な特徴量を作成できる（ユーザごとの過去1ヵ月間の購買金額合計など）	*		○
134	13	データ加工	★★★	特徴量エンジニアリング	ドメイン知識に基づく洞察から有効な特徴量を効率的に作成できる（類似商品との価格差、借入額と返済額の比など）	*		
135	14	データ加工	★★★	特徴量エンジニアリング	有用と思われるデータを新たに調達し、既知のデータと組み合わせることで有効な特徴量を作成できる（業界特有なカレンダー、地理空間的な類似度を考慮した集計値など）	*		
136	1	データ可視化	★	方向性定義	データの性質を理解するために、データを可視化し眺めて考えることの重要性を理解している	*		○
137	2	データ可視化	★	方向性定義	可視化における目的の広がりについて概略を説明できる（単に現場の作業支援する場合から、ビッグデータ中の要素間の関連性をダイナミックに表示する場合など）	*		
138	3	データ可視化	★★	方向性定義	特異点を明確にする、データ解析部門以外の方にデータの意味を正しく伝える、現場の作業を支援するといった可視化の役割・方向性を判別できる	*		
139	4	データ可視化	★★★	方向性定義	データ量が膨大で構造が捉えにくい場合や、アウトプットが想像しにくい場合であっても、可視化の役割・方向性を判断できる（ビッグデータ中の要素間の関連性をダイナミックに表現する、細部に入りきらずに間に対して答えを出すなど）			
140	5	データ可視化	★	軸出し	散布図などの軸出しにおいて、目的やデータに応じて縦軸・横軸の候補を適切に洗い出せる	*		○
141	6	データ可視化	★	軸出し	積み上げ縦棒グラフでの属性の選択など、目的やデータに応じて適切な層化（比較軸）の候補を出せる	*		○
142	7	データ可視化	★★	軸出し	抽出したい意味にふさわしい軸・層化の粒度、順番を考慮して軸のきざみや層化方法を選択できる	*		○
143	8	データ可視化	★★	軸出し	膨大な属性を持つテーブルから目的に有用な属性を選択できる	*		
144	9	データ可視化	★★★	軸出し	非構造データから分析の軸にならう候補を抽出し、付加すべき属性候補を適切に出せる	*		
145	10	データ可視化	★	データ加工	サンプリングやアンサンブル平均によって適量にデータ量を減らすことができる	*		
146	11	データ可視化	★	データ加工	読み取りたい特徴を効果的に可視化するために、統計量を使ってデータを加工できる			
147	12	データ可視化	★★	データ加工	データの持つ分散量の観点で、高次元のデータを主成分分析（PCA）などにより1～3次元のデータに変換できる			○
148	13	データ可視化	★★★	データ加工	高次元の非線形な（高次の曲線、渦状の分布などの）データであっても、適切に1～3次元のデータに圧縮して、特徴（データの総分散量および各データの位置関係）を損なわずに可視化できる			
149	14	データ可視化	★★★	データ加工	ネットワーク構造、グラフ構造などの表現において、ノードとエッジが増えすぎて特徴抽出が困難であっても、データの絞り込みや抽象度を上げることで適切に可視化できる		*	
150	15	データ可視化	★★★	データ加工	データ量が膨大（ペタバイト以上）なために、処理しきれず描画できない規模のデータに対しても、適度なデータや情報の抽出（間引き）、クラスター分析などにより可視化しうる状態にデータを加工できる		*	
151	16	データ可視化	★★★	データ加工	大規模なデータへのリアルタイムな可視化が求められる場合であっても、特異点の抽出や次元圧縮を通じてデータを圧縮し、リアルタイム表示できる		*	
152	17	データ可視化	★★★	データ加工	大規模なデータへのリアルタイムな可視化が求められる場合であっても、データの分割転送、復元を通じて可視化できる		*	
153	18	データ可視化	★	表現・実装技法	データ解析部門以外の方に、データの意味を伝えるサインとしての可視化ができる	*		
154	19	データ可視化	★	表現・実装技法	適切な情報濃度を判断できる（データインク比など）			
155	20	データ可視化	★	表現・実装技法	不必要な誇張をしないための軸表現の基礎を理解できている（コラムチャートのY軸の基準点は「0」からを原則とし軸を切らないなど）			○
156	21	データ可視化	★	表現・実装技法	強調表現がもたらす効果と、明らかに不適切な強調表現を理解している（計量データに対しては位置やサイズ表現が色表現よりも効果的など）			○

NO	SubNo	スキルカテゴリ	スキルレベル	サブカテゴリ	チェック項目	BZ	DE	必須スキル
157	22	データ可視化	★	表現・実装技法	1～3次元の比較において目的（比較、構成、分布、変化など）に応じ、BIツール、スプレッドシートなどを用いて図表化できる		*	○
158	23	データ可視化	★	表現・実装技法	端的に図表の変化をアニメーションで可視化できる（人口動態のヒストグラムが経年変化する様子を表現するなど）			
159	24	データ可視化	★	表現・実装技法	1～3次元の図表を拡張した多変量の比較を適切に可視化できる（平行座標、散布図行列、テーブルレンズ、ヒートマップなど）			
160	25	データ可視化	★★	表現・実装技法	ネットワーク構造、グラフ構造、階層構造などの統計的な関係性の可視化ができる		*	
161	26	データ可視化	★★	表現・実装技法	地理空間情報などを地図に重ね合わせた可視化ができる（コロプレスマップ、ボロノイ図など）			
162	27	データ可視化	★★	表現・実装技法	挙動・軌跡の可視化ができる（店舗内でのユーザの回遊やEye tracking、ウェアラブル端末の3軸ジャイロセンサーや心拍数など）			
163	28	データ可視化	★★	表現・実装技法	適切な情報（意味）を押さえた上で、デザイン性を高めるための要件提示ができる	*		
164	29	データ可視化	★★	表現・実装技法	ドメイン知識とビジネス背景をふまえ、適切な情報項目を参照し、利用者の使い勝手を考慮した可視化したダッシュボードあるいは可視化ツールを設計できる	*	*	
165	30	データ可視化	★★★	表現・実装技法	人体、標高を持つ地図、球面など3次元空間上にデータを重ね合わせた可視化ができる		*	
166	31	データ可視化	★★★	表現・実装技法	地図上で同時に動く数百以上のポイントにおける時間変化を動的に表現できる（多地点での風の動き、飛行物の軌跡など）		*	
167	32	データ可視化	★★★	表現・実装技法	複雑で総合的な表現が求められる場合に、ユーザインターフェースの設計に加え、内部のアルゴリズム設計やシステム環境への負荷調整なども踏まえた可視化ができる		*	
168	33	データ可視化	★	意味抽出	外れ値を見出すための適切な表現手法を選択できる			
169	34	データ可視化	★	意味抽出	データの可視化における基本的な視点を挙げることができる（特異点、相違性、傾向性、関連性を見出すなど）			○
170	35	データ可視化	★★	意味抽出	統計値（代表値の指標、バラツキの指標、有意性の指標、関係性）を正しく読み、回帰式や移動平均線に意味付けできる			○
171	36	データ可視化	★★★	意味抽出	分類・グルーピングの分析において、分布傾向から原因を追究し、活用方針を計画・主導できる（分類に応じたDM発送による反応率の向上など）	*		
172	37	データ可視化	★★★	意味抽出	予測の分析において、関連性、特異点、変曲点から原因を追究し、活用方針を計画・主導できる（予測結果に基づく発注管理など）	*		
173	38	データ可視化	★★★	意味抽出	関係性の分析において関連が高い/低い原因を追究し、活用方針を計画・主導できる（レコメンドなど）	*		
174	1	時系列分析	★	時系列分析	時系列分析を行う際にもつべき視点を理解している（長期トレンド、季節成分、周期性、ノイズ、定常性など）			
175	2	時系列分析	★★	時系列分析	時系列データの時点間での関係を、自己相関やコログラムを利用して評価ができる			
176	3	時系列分析	★★	時系列分析	時系列データに対し、ライブラリを使用して、分析結果の比較を行い、適切なモデルを選択できる（自己回帰モデル[AR]、移動平均モデル[MA]、ARIMA、SARIMA、VAR、GARCH、Prophet、指数平滑法など）			
177	4	時系列分析	★★	時系列分析	グレンジャー因果などの手法を用い、時系列データにおける変数間の因果関係を把握できる			
178	5	時系列分析	★★	時系列分析	高速フーリエ変換（Fast Fourier Transform: FFT）などを用いてピリオドグラムの計算量を抑制できる			
179	6	時系列分析	★★	時系列分析	時系列データにおいて、前埋め/線形補完/移動平均などの手法を用い欠測値の補完を行うことができる			
180	7	時系列分析	★★★	時系列分析	状態空間モデルにおいて、カルマンフィルタを用いた欠測値の補完や予測モデルの構築ができる			
181	8	時系列分析	★★★	時系列分析	非線形・非ガウス型状態空間モデルにおいて、モンテカルロ・フィルタを用いて、複雑な時系列システムの予測モデルを構築できる			
182	1	学習	★	機械学習	機械学習にあたる解析手法の名称を3つ以上知っており、手法の概要を説明できる			
183	2	学習	★	機械学習	機械学習のモデルを使用したことがあり、どのような問題を解決できるか理解している（回帰・分類、クラスター分析の用途など）			
184	3	学習	★	機械学習	「教師あり学習」「教師なし学習」の違いを理解している			○
185	4	学習	★	機械学習	過学習とは何か、それがもたらす問題について説明できる			○
186	5	学習	★	機械学習	次元の呪いとは何か、その問題について説明できる			
187	6	学習	★	機械学習	教師あり学習におけるアンダーフィットの必要性を説明できる	*		
188	7	学習	★	機械学習	観測されたデータにバイアスが含まれる場合や、学習した予測モデルが少数派のデータをノイズと認識してしまった場合などに、モデルの出力が差別的な振る舞いをしてしまうリスクを理解している			○
189	8	学習	★	機械学習	機械学習における大域的(global)な説明（モデル単位の変数数の寄与度など）と局所的(local)な説明（予測するレコード単位の変数数の寄与度など）の違いを理解している			
190	9	学習	★	機械学習	ホールドアウト法、交差検証（クロスバリデーション）法の仕組みを理解し、学習データ、パラメータチューニング用の検証データ、テストデータを作成できる			
191	10	学習	★	機械学習	時系列データの場合は、時間軸で学習データとテストデータに分割する理由を理解している			
192	11	学習	★	機械学習	機械学習モデルは、データ構成の変化(データドリフト)により学習完了後から精度が劣化していくため、運用時は精度をモニタリングする必要があることを理解している			
193	12	学習	★★	機械学習	学習データ、検証データ、テストデータの分割時にデータの性質や目的に応じて、地理空間情報や様々な属性性を考慮すべきか判断できる	*		
194	13	学習	★★	機械学習	予測モデルにおけるリークとは何か、それがもたらす問題について説明できる			
195	14	学習	★★★	機械学習	リークの発生の有無を適切に判断し、対応できる（マスタデータが更新されており、本来予測時に使用できない未来の情報を含んでしまうパターン、本来は未知である評価データも含めて標準化するパターンなど）	*		
196	15	学習	★★★	機械学習	データと分析要件から、データ分布やモデル精度のモニタリング設計・実施と、劣化が見込まれるモデルに対するリモデルの設計ができる		*	
197	16	学習	★★	機械学習	次元の呪いの影響を受けやすいアルゴリズムを識別し対処するアプローチを知っている（特徴量選択、次元圧縮、L1/L2正則化など）			
198	17	学習	★★	機械学習	アンダーフィットにおいて、判断基準の明確化やアンダーフィットの精度管理を適切に実施することにより、高品質な教師付きデータを効率的に作成できる	*		
199	18	学習	★★	機械学習	教師あり学習と強化学習の違いを、前提、定義、応用先といった観点で説明できる			
200	19	学習	★★	機械学習	決定木分析においてCHAID、C5.0、CARTなどのデータ分割のアルゴリズムの特徴を理解し、適切な方式を選定できる			
201	20	学習	★	機械学習	ニューラルネットワークの基本的な考え方を理解し、出力される「ダイアグラム」の入力層、隠れ層、出力層の概要と、活性化関数の重要性を理解している			○
202	21	学習	★	機械学習	ライブラリを使ってサポートベクターマシンによる分析を実行・評価できる			
203	22	学習	★	機械学習	決定木をベースとしたアンサンブル学習（Random Forest、勾配ブースティング[Gradient Boosting Decision Tree : GBDT]、その派生形であるXGBoost、LightGBMなど）による分析を、ライブラリを使って実行でき、その結果を正しく解釈できる			
204	23	学習	★★	機械学習	データに「変更を加えず予測モデルを故意に誤らせる敵対的サンプル（Adversarial examples）」について、代表的な攻撃方法を理解している			

NO	SubNo	スキルカテゴリ	スキルレベル	サブカテゴリ	チェック項目	BZ	DE	必須スキル
205	24	学習	★★★	機械学習	最新の研究に基づき敵対的サンプル（Adversarial examples）に対する現実的な対策を検討できる			
206	25	学習	★★	機械学習	生成モデルと識別モデルの違いやそれぞれのメリットを理解しており、目的に応じて適切に使い分けられる			
207	26	学習	★★	機械学習	不均衡データ（Imbalanced data）がモデルに与える影響を理解し、サンプリングや評価/損失関数のチューニングなどを適切に対処できる			
208	27	学習	★★	機械学習	バイアスとバリエーションの関係について、未学習や過学習の観点から説明できる			
209	28	学習	★★	機械学習	機械学習・AIの公平性の評価指標について説明できる（デモグラフィックパリティ、均等オッズなど）	*		
210	29	学習	★★	機械学習	LIME、SHAPなどを用いて、ブラックボックス性の高いモデルの局所的な説明（レコード単位の予測根拠の提示）ができる			
211	30	学習	★★★	機械学習	分析要件に応じ、モデリング手法（線形モデル、決定木、サポートベクターマシン、ニューラルネットワーク、アンサンブル学習など）の選択とパラメータ設定、結果の評価、チューニングを適切に設計・実施・指示できる			○
212	31	学習	★★★	機械学習	複数の機械学習モデルを組み合わせるアンサンブル学習において、各モデルの予測スコアを用いた平均（Averaging）、多数決（Voting）、Stackingなどの手法を理解し、目的に応じて設計・実装できる		*	
213	32	学習	★★★	機械学習	半教師あり学習やアクティブラーニングにより効率的なアノテーションができる	*		
214	33	学習	★★	機械学習	課題やデータに応じて、汎化性能向上のためのデータ拡張（Data Augmentation）を設計・実装できる（画像データの回転・反転、テキストデータの単語入れ替え・置換・削除など）			
215	34	学習	★★★	機械学習	機械学習等の最新の論文を理解し、必要とあれば自分で実装し評価できる		*	
216	35	学習	★	深層学習	深層学習（ディープラーニング）モデルの活用による主なメリットを理解している（特徴量抽出が可能になるなど）			○
217	36	学習	★★	深層学習	バッチ勾配降下法（バッチ学習）、確率的勾配降下法（オンライン学習）、ミニバッチ勾配降下法（ミニバッチ学習）の違いを説明できる			
218	37	学習	★★	深層学習	ドロップアウト、L1/L2正則化などによる過学習の抑制や、バッチ正規化による学習の効率化について理解している			
219	38	学習	★★	深層学習	文字や単語といった質的（離散的）な特徴量をニューラルネットワークで学習する場合、特徴量の各値にベクトルを対応させるEmbeddingが有効であることを理解している			
220	39	学習	★★	深層学習	畳み込みニューラルネットワーク(CNN)、再帰型ニューラルネットワーク（RNN）、長期短期記憶（LSTM）、Transformerなどの深層学習の主要方式の特徴を理解し、目的に応じて適切に選定できる			
221	40	学習	★★	深層学習	ニューラルネットワークにおいてAttention機構の特徴を理解し、各ベクトルの重みを算出できる			
222	41	学習	★★	深層学習	転移学習とファインチューニングの基本的な枠組みを理解し、事前学習済みのニューラルネットワークを用いた効率的な学習方法を設計・実装できる		*	
223	42	学習	★★	深層学習	誤差逆伝播法（Backpropagation）における勾配消失、勾配爆発の問題を理解し、適切な活性化関数・重みの初期化方法の選択、Gradient Clippingなどの緩和策を実施できる			
224	43	学習	★★	深層学習	Residual Connection(Skip Connection)が誤差逆伝播法（Backpropagation）における勾配消失問題の緩和策となっていることを理解している			
225	44	学習	★★	深層学習	深層学習における最適化法/Optimizer（SGD、Momentum、RMSProp、Adamなど）や学習率の動的な調整法（Cyclical Learning Rateなど）の重要性について理解している			
226	45	学習	★★★	深層学習	深層学習の実装において、予想精度を向上するため、層の種類（全結合、畳み込み、プールなど）、層数、ニューロン数、活性化関数、学習率、学習回数などをチューニングできる			
227	46	学習	★★★	深層学習	深層学習モデルの推論時の計算コストを削減する方法（枝刈り、蒸留、量子化など）を設計・実装できる			
228	47	学習	★★	深層学習	主要な深層生成モデル（変分自己符号化器[Variational AutoEncoder：VAE]、敵対的生成ネットワーク[Generative Adversarial Network：GAN]）やその派生形を理解し、目的に応じて学習・評価できる			
229	48	学習	★★	強化学習	状態と行動、報酬による強化学習の基本概念を理解し、モンテカルロ法やQ学習などを実行できる			
230	49	学習	★★★	強化学習	主要な強化学習モデルを理解し、目的に応じて学習・評価できる(Deep Q-network [DQN]、Asynchronous Advantage Actor-Critic[A3C]、Proximal Policy Optimization [PPO]やその派生形など)			
231	1	自然言語処理	★	自然言語処理	テキストデータに対する代表的なクリーニング処理（小文字化、数値置換、半角変換、記号除去、ステミングなど）を目的に応じて適切に実施できる		*	
232	2	自然言語処理	★	自然言語処理	形態素解析や係り受け解析のライブラリを適切に使い、基本的な文書構造解析を行うことができる			
233	3	自然言語処理	★	自然言語処理	自然言語処理を用いて解けるタスクを理解し、各タスクの入出力を説明できる（GLUEタスクや固有表現抽出、機械翻訳など）			
234	4	自然言語処理	★★	自然言語処理	Unigram language modelやBPE（Byte Pair Encoding）によるサブワード分割手法を理解しており、必要に応じて適切に実施できる			
235	5	自然言語処理	★★	自然言語処理	使用メモリの削減やモデルの精度向上のため、名寄せやユーザ定義辞書の作成、ストップワードの除去・置換を実施できる			
236	6	自然言語処理	★★	自然言語処理	Word2Vec（Skip-gram、CBoW）などの分散表現（ベクトル表現）モデルを理解し、使いこなせる			
237	7	自然言語処理	★★	自然言語処理	TF-IDFやcos類似度などの基本的なアルゴリズムを使い、文書ベクトルの作成や文書間の類似度を計算できる			
238	8	自然言語処理	★★★	自然言語処理	形態素解析・構文解析・固有表現抽出のアルゴリズムを理解し、使いこなせる			
239	9	自然言語処理	★★★	自然言語処理	N-gram言語モデルの構築方法と代表的なスムージングアルゴリズムを理解し、使いこなせる			
240	10	自然言語処理	★★★	自然言語処理	索引型の全文検索の仕組み（転置インデックス、スコアリング、関連性フィードバック）を理解し、使いこなせる		*	
241	11	自然言語処理	★★★	自然言語処理	Trie、Suffix Arrayなどの代表的な高速文字列検索アルゴリズムを理解し、使いこなせる		*	
242	12	自然言語処理	★★★	自然言語処理	主要なトピックモデル（確率的潜在意味解析[pLSA]、潜在的ディリクレ配分法[LDA]）の仕組みを理解し、使いこなせる			
243	13	自然言語処理	★★★	自然言語処理	データの特性に合わせ、適切な言語処理アルゴリズムを選択し、誤り分析、辞書作成などを行い、成果を最大化できる			
244	14	自然言語処理	★★★	自然言語処理	再帰型ニューラルネットワーク（RNN）、長期短期記憶（LSTM）、Transformerなどを用いたニューラル言語モデルを理解し、使いこなせる			
245	15	自然言語処理	★★★	自然言語処理	隠れマルコフモデル（Hidden Markov Model：HMM）、Conditional Random Fields：CRF）などを用いた系列ラベリング手法を理解し、使いこなせる			
246	16	自然言語処理	★★★	自然言語処理	GPT（Generative Pre-Trained transformer）、BERT（Bidirectional Encoder Representations from Transformer）などのAttentionベースのニューラル言語モデルを理解し、使いこなせる			
247	1	画像・映像認識	★	画像認識	画像のデジタル表現の仕組みと代表的な画像フォーマットを知っている			
248	2	画像・映像認識	★	画像認識	画像に対して、目的に応じた適切な色変換や簡単なフィルタ処理などを行うことができる		*	
249	3	画像・映像認識	★	画像認識	画像データに対する代表的なクリーニング処理（リサイズ、バディング、正規化など）を目的に応じて適切に実施できる		*	

NO	SubNo	スキルカテゴリ	スキルレベル	サブカテゴリ	チェック項目	BZ	DE	必須スキル
250	4	画像・映像認識	★★	画像認識	画像認識を用いて解けるタスクを理解し、入出力とともに説明できる（識別、物体検出、セグメンテーションなど）			
251	5	画像・映像認識	★★	画像認識	画像の処理や解析において、効果的なパターン検出や画像特徴抽出などを既存手法から選ぶことができる		*	
252	6	画像・映像認識	★★	画像認識	画像の処理や解析において、既存のAPI化したクラウドサービスなどを目的に即して、選定・活用できる			
253	7	画像・映像認識	★★	画像認識	ResNet、EfficientNet、MobileNetなどの代表的な画像認識モデルを理解し使いこなせる			
254	8	画像・映像認識	★★	画像認識	ライブラリ、学習済みモデルなどを用いて識別・物体検出・セマンティックセグメンテーションなどの画像処理を実装・評価できる			
255	9	画像・映像認識	★★★	画像認識	画像・動画処理を行う環境（モバイル上での処理など実行に制約を伴う環境）と処理パフォーマンスを考慮し、適切な手法の選択とモデルの構築ができる		*	
256	10	画像・映像認識	★	映像認識	動画のデジタル表現の仕組みと代表的な動画フォーマットを理解しており、動画から画像を抽出する既存方法を使うことができる			
257	11	画像・映像認識	★★	映像認識	物体が動いている方向・速度（オプティカルフロー）を動画から推定し、トラッキングできる			
258	12	画像・映像認識	★★★	映像認識	動画の解析手法の現況について理解し、専門家とともに実装を検討できる（Kernel Temporal Segmentation[KTS]による動画の映像区間切り分け、動画内の重要度推定によるハイライト作成など）			
259	1	音声認識	★	音声認識	wavやmp3などの代表的な音声フォーマットの特徴や用途、基本的な変換処理について説明できる（サンプリングレート、符号化、量子化など）			
260	2	音声認識	★	音声認識	目的に応じ音声認識関連のAPIを選択し、適用できる（Speech to Text など）			
261	3	音声認識	★★	音声認識	音声データから、分析目的にあった波形データの抽出やノイズの除去ができる		*	
262	4	音声認識	★★	音声認識	音声認識や本人認証、感情分析などの代表的な音声処理分野について理解し、用いられる分析手法を説明できる			
263	5	音声認識	★★★	音声認識	短時間フーリエ分析、メルフィルタバンク処理、ケプストラム分析、LPC（線形予測分析）などの代表的な音声信号分析手法を理解し、使いこなすことができる			
264	6	音声認識	★★★	音声認識	音声認識、音声合成、認証・感情分析などの目的に合わせて、パラメータ調整や手法変更、言語モデル・音響モデルなどを差し替え、モデル構築・精度評価ができる			
265	1	パターン発見	★	パターン発見	条件Xと事象Yの関係性を信頼度、支持度、リフト値を用いて評価できる			
266	2	パターン発見	★★	パターン発見	アプリオリアルアルゴリズムのアソシエーション分析において、パラメータをチューニングして出力件数を調整できる			
267	3	パターン発見	★★	パターン発見	分析要件から適切なレコメンダアルゴリズム（協調フィルタリング、コンテンツベースフィルタリング、Factrization Machineなど）を選定できる			
268	4	パターン発見	★★★	パターン発見	行動履歴に加え、アイテム属性、ユーザ属性、時間変化などの情報を用いて、目的に応じた推薦アルゴリズムを実装・評価できる			
269	1	シミュレーション・データ同化	★★	シミュレーション・データ同化	離散型・連続型シミュレーションについて、説明できる（モンテカルロ、ヒストリカル、Agentベースなど）			
270	2	シミュレーション・データ同化	★★	シミュレーション・データ同化	マルコフ連鎖の特徴を理解し、MCMC（マルコフ連鎖モンテカルロ法）シミュレーションをライブラリを用いて実装できる			
271	3	シミュレーション・データ同化	★★★	シミュレーション・データ同化	データ同化の概念を理解し、実行できる（データを用いてシミュレーション内の不確実性を減少させる計算技法など）			
272	4	シミュレーション・データ同化	★★★	シミュレーション・データ同化	シミュレーションにおける問題を理解し、対処を考えることができる（初期条件・境界条件・パラメータの不確実性、データ分布の不均一性、実験計画の最適性など）			
273	5	シミュレーション・データ同化	★★★	シミュレーション・データ同化	MCMC（マルコフ連鎖モンテカルロ法）における各種アルゴリズム（メトロポリス・ヘイスティングス法、ギブスサンプラー、ハミルトニアン・モンテカルロ法など）について理解し、活用できる			
274	1	最適化	★★	最適化	複数のA/Bテスト、スプリットテストの統計的結果を踏まえ、デザイン等の最適化を実施できる		*	
275	2	最適化	★★	最適化	一定の制約下で最適解の識別と報酬の最大化がともに求められ、かつ報酬分布が時間経過で変化するような問題に対して、多腕バンディットアルゴリズムを適用・実装できる			
276	3	最適化	★★	最適化	線形計画法について説明できる			
277	4	最適化	★★	最適化	凸関数および、凸計画問題の条件や特徴を説明できる			
278	5	最適化	★★	最適化	連続最適化問題（制約なし）において、使用可能なアルゴリズムを説明できる（ニュートン法、最急降下法など）			
279	6	最適化	★★	最適化	連続最適化問題（制約あり）において、使用可能なアルゴリズムを説明できる（ラグランジュ未定乗数法、内点法、逐次2次計画法など）			
280	7	最適化	★★	最適化	組み合わせ最適化問題において、代表的な解法の概念を説明できる（厳密解法（分枝限定法、動的計画法、切除平面法）、近似解法（局所探索、貪欲法など）、メタヒューリスティック解法（遺伝的アルゴリズム、タブーサーチなど））			
281	8	最適化	★★★	最適化	ビジネス課題にあわせて、変数、目的関数、制約を定式化し、線形・非線形を問わず、最適化モデリングができる		*	
282	9	最適化	★★★	最適化	代表的な最適化問題に関して、モデリングを行い、ソルバーを使い、最適化できる（ナップザック問題、ネットワークフロー問題、巡回回路問題など）			

スキルチェックリスト 2021年版 <データエンジニアリングカ>

▼ 他分野寄りのスキル

NO	SubNo	スキルカテゴリ	スキルレベル	サブカテゴリ	チェック項目	BZ	DS	必須スキル
1	1	環境構築	★	システム企画	オープンデータを収集して活用する分析システムの要件を整理できる			
2	2	環境構築	★★	システム企画	数千万レコードのデータを保持するシステムにおいて、データの重要性や分析要件に則したシステム構築、初期データ投入方法、システム運用の要件定義が行える			○
3	3	環境構築	★★	システム企画	ソフトウェア開発プロジェクトの管理方法、設計、テスト方法を理解した上で、データ管理・分析システムの要件定義を行うことができる			
4	4	環境構築	★★★	システム企画	組織の情報システム戦略、全体システム化計画の方針に従い、全体最適の観点を持って、データ分析システムの企画・要件定義を推進できる	*		
5	5	環境構築	★★★	システム企画	データ活用戦略を正しく理解し、各種業務プロセスについての専門知識とシステムに関する知識を有し、双方を活用して、適切な要件定義が行える	*		○
6	6	環境構築	★★★	システム企画	数十億レコードの内部データ（CRM、Webログ、ユーザー購買データなど）、外部データ（購入しているデータ、オープンデータなど）を理解し、複数のデータソースを統合する要件を整理できる	*		
7	7	環境構築	★★★	システム企画	数十億レコードのデータを持つ業務要件やリソース負荷に応じて、データフローや管理機構の統合、またバッチ実行スケジュールの最適化について要件定義が行える			
8	8	環境構築	★	システム設計	サーバー1～10台規模のシステム構築、システム運用を設計書を元に行える			
9	9	環境構築	★	システム設計	数十万レコードを持つデータベースのバックアップ・アーカイブ作成など定常運用ができる			
10	10	環境構築	★★	システム設計	分析システムにおけるシステム運用（監視、フェイルオーバー、バックアップ、リカバリ処理、サービス開始・停止処理など）の運用設計や手順書作成が行える			
11	11	環境構築	★★	システム設計	数千万レコードのデータを保持するシステムのキャパシティ要件（データ容量）と必要処理性能（スループット）を定義できる			
12	12	環境構築	★★	システム設計	HTTPを活用したオープンAPIと分析システムのサーバー環境及びデータベースの連携設計ができる			
13	13	環境構築	★★	システム設計	クラウドのメリットとデメリットを比較して、必要に応じてオンプレミスのデータ分析システムをクラウド環境に移行できる			
14	14	環境構築	★★	システム設計	既存システムの各機能が参照・更新するデータの構造や関連付けを把握し、分析機能の追加や改修時にデータモデルの見直しが必要か判断できる	*		○
15	15	環境構築	★★	システム設計	分析実行頻度の少ないシステムにおいて、分析非稼働時にはサービス削除し、必要時のみサービスを有効にすることでクラウドコストを最小化するシステム運用を設計できる			
16	16	環境構築	★★★	システム設計	数十億レコードのデータに対してリバランシングなども含めてシステム拡張方法や最適化の要件を整理できる			
17	17	環境構築	★★★	システム設計	扱うデータのデータ規模や機密性、分析要件を理解した上で、オンプレミスで構築するか、クラウド上で構築するかの要件を整理できる			○
18	18	環境構築	★	アーキテクチャ設計	ノーコード・ローコードツールを組み合わせ、要件に応じたアプリやツールを設計できる			
19	19	環境構築	★	アーキテクチャ設計	コンテナ技術の概要を理解しており、既存のDockerイメージを活用して効率的に分析環境を構築できる			
20	20	環境構築	★	アーキテクチャ設計	分析環境を提供するクラウド上のマネージドサービス(Amazon SageMaker、Azure Machine Learning、Google Cloud AI Platform、IBM Watson Studioなど)を利用し、機械学習モデルの開発機能を利用することができる			
21	21	環境構築	★★	アーキテクチャ設計	分析者向けのRDBMS、NoSQL、ETL、Visualizationなど単一コンポーネントのユーザー利用機能の設計が行える			○
22	22	環境構築	★★	アーキテクチャ設計	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			
23	23	環境構築	★★	アーキテクチャ設計	Dockerコンテナ技術を用いてデータの蓄積環境・分析環境を構築し、再利用できる形でイメージを管理できる			
24	24	環境構築	★★	アーキテクチャ設計	開発した予測モデルをREST APIから呼び出せるように実装し、マイクロサービスのコンポーネントの一つとしてシステムに組み込むことができる			
25	25	環境構築	★★	アーキテクチャ設計	量子コンピュータにおけるロジックを検討する際に計算結果のゆらぎについて理解した上で、適切な処理フローやテスト手順を検討できる			
26	26	環境構築	★★★	アーキテクチャ設計	数十億レコード規模のデータに対し、HadoopやSparkなどを組み合わせた適切なシステム構成を設計できる			
27	27	環境構築	★★★	アーキテクチャ設計	KVS、カラム指向、ドキュメント指向などデータ構造の異なる複数のシステムからデータ取得と分析環境への連携を設計できる			
28	28	環境構築	★★★	アーキテクチャ設計	費用対効果が高いPaaS型サービスと設計自由度の高いIaaS環境のメリット、デメリットを理解し、PaaS/IaaSを組み合わせた効率的な分析システムを設計できる			
29	29	環境構築	★★★	アーキテクチャ設計	Amazon Web Services（AWS）、Microsoft Azure、Google Cloud Platform（GCP）、IBM Cloudなど複数社のクラウドサービスを組み合わせるマルチクラウド構成を設計し、各社が提供する固有の分析サービスを連携して活用することができる			
30	30	環境構築	★★★	アーキテクチャ設計	量子コンピュータの量子計算の特徴や制約を理解し、量子コンピュータの計算性能を生かすユースケースを検討できる			
31	1	データ収集	★	クライアント技術	対象プラットフォームが提供する機能（SDKやAPIなど）の概要を説明できる			
32	2	データ収集	★	クライアント技術	Webクローラー・スクレイピングツールを用いてWebサイト上の静的コンテンツを分析用データとして収集できる			
33	3	データ収集	★★	クライアント技術	対象プラットフォーム（IoTデバイス、スマートフォン、クラウドサービスなど）で取得可能なデータ項目から目的に応じてデータ収集システムを設計できる			
34	4	データ収集	★★	クライアント技術	対象プラットフォームにおけるバッテリー消費や通信速度などを含めたシステム要件を作成できる			
35	5	データ収集	★★	クライアント技術	エッジコンピューティング構成においてエッジデバイス側（端末側）で収集したログを、適切なタイミングでサーバー側に転送・集約するデータ収集の仕組みを設計できる	*		
36	6	データ収集	★★★	クライアント技術	収集するデータが増えることを想定し、拡張性を考慮したエッジデバイスのロガー機能（or データログ）を設計できる	*		
37	7	データ収集	★	通信技術	対象プラットフォームに用意された通信サービス（HTTP、FTPなど）を用い、データを収集先に格納するための機能を実装できる			
38	8	データ収集	★★	通信技術	データ収集対象の要件に応じて、MQTT（Message Queue Telemetry Transport）によるパブリッシュ/サブスクライブ型の通信を検討・実装できる			
39	9	データ収集	★★	通信技術	既存のデータ収集システムの通信を5GやLPWA(Low Power Wide Area)など新しい通信方式を変更することによって発生する影響を評価できる			
40	10	データ収集	★★	通信技術	ネットワークプロトコルや暗号化などの通信技術を用い、通信のボトルネックと可用性（継続的に通信が成立していること）を考慮した上で、必要な通信機能を実装できる			
41	11	データ収集	★★★	通信技術	データ通信において、機能・性能の問題に対し根本原因を特定できるだけでなく、必要に応じて新規技術の適用を検討できる			
42	12	データ収集	★	データ抽出	データベースから何らかのデータ抽出方法を活用し、小規模なExcelのデータセットを作成できる			○
43	13	データ収集	★	データ収集	既存のサービスやアプリケーションに対して、分析をするためのログ出力の仕様を整理することができる	*	*	

NO	SubNo	スキルカテゴリ	スキルレベル	サブカテゴリ	チェック項目	BZ	DS	必須スキル
44	14	データ収集	★★	データ収集	Fluentdなどのログ収集ツールを用いて形式の異なる複数サーバのログを収集し分析可能な形式に変換して蓄積することができる			
45	15	データ収集	★★★	データ収集	GDPRや個人情報保護法、企業のプライバシーポリシーを理解し、Webログやユーザ情報の適切なデータ取得設計やID連携設計ができる	*		
46	16	データ収集	★★★	データ収集	入手可能なデータに加え、分析結果の品質・効果を向上させる新たなデータ入手方法を提案できる（IoTでの新設センサーの種類、配置場所、データ入手間隔など）	*	*	○
47	17	データ収集	★★	データ統合	データフォーマットの異なる同一内容のデータ（会員情報、商品マスタなど）を統合する処理を設計できる			○
48	18	データ収集	★★★	データ統合	組織統合などにより異なるコード定義やフォーマットが混在する複数システムのデータを統合し利活用するシステムを設計できる	*		
49	19	データ収集	★★★	データ統合	システム分析・業務分析をもとに、必要なデータフロー管理やジョブ管理などのツールを選定・評価できる			○
50	1	データ構造	★	基礎知識	扱うデータが、構造化データ（顧客データ、商品データ、在庫データなど）か非構造化データ（雑多なテキスト、音声、画像、動画など）なのかを判断できる			○
51	2	データ構造	★	基礎知識	ER図を読んでテーブル間のリレーションシップを理解できる			○
52	3	データ構造	★★	要件定義	業務で使用するデータのライフサイクル（いつ、どんなデータが発生し、いつまで保持されているのかなど）を把握して、論理モデルを作成できる	*		○
53	4	データ構造	★★	要件定義	データ保持ルール（データアクセス、性能、保持期間、セキュリティなど）に基づき、データベース・DWHの運用ルールを定義できる			
54	5	データ構造	★	テーブル定義	正規化手法（第一正規化～第三正規化）を用いてテーブルを正規化できる			
55	6	データ構造	★★	テーブル定義	ビジネスプロセスを理解・整理して、データフロー図、論理データモデル、ER図、テーブル定義書を作成できる	*		○
56	7	データ構造	★★	テーブル定義	業務特性や基幹システムの特徴をもとに、検索で頻繁に使用するデータのキー（顧客IDなど）を想定し、インデックスを作成・設定できる	*		
57	8	データ構造	★★	テーブル定義	データ集計を高速化またはSQLを単純化するため、スタースキーマ、スノーフレークスキーマなどを用いたデータモデルを設計できる			
58	9	データ構造	★★	テーブル設計	稼働中の複数のシステム間で発生するデータ項目の差異を、変換テーブルを活用して、埋めることができる			
59	10	データ構造	★★	テーブル設計	DWHに入れる元データ（基幹DBのデータなど）のキーに変更があった場合に、サロゲートキーやナビゲーションブリッジテーブルを用いて対応できる			
60	11	データ構造	★★	テーブル設計	非正規化テーブルや一時テーブルなどを作成し、アプリケーションの処理速度を高速化できる			
61	1	データ蓄積	★	DWH	DWHアプライアンス（Oracle Exadata、IBM Integrated Analytics System、Teradataなど）に接続し、複数テーブルを結合したデータを抽出できる			
62	2	データ蓄積	★★	DWH	DWHアプライアンス（Oracle Exadata、IBM Integrated Analytics System、Teradataなど）の機能と特徴を理解し、適切な管理対象データを選定できる			
63	3	データ蓄積	★	分散技術	HadoopやSparkの分散技術の基本的な仕組みと構成を理解している			
64	4	データ蓄積	★	分散技術	NoSQLデータストア（HBase、Cassandra、Mongo DB、CouchDB、Redis、Amazon DynamoDB、Cloudant、Azure Cosmos DBなど）にAPIを介してアクセスし、新規データを登録できる			
65	5	データ蓄積	★★	分散技術	HadoopやSparkのメリット・デメリットを理解し、HadoopやSparkにて管理すべきデータを選定できる			
66	6	データ蓄積	★★	分散技術	KVSの特性（集計・ソートが苦手、データの一貫性保証など）を理解し、KVSがデータストア要件を満たすかを判断できる			
67	7	データ蓄積	★★	分散技術	HadoopやSparkなどのデータ基盤上で、パーティション分割されたデータに対して更新・管理ができる			
68	8	データ蓄積	★★★	分散技術	HadoopやSparkの分散アーキテクチャを理解し、大容量データ処理の設計やパフォーマンスチューニングができる			
69	9	データ蓄積	★	クラウド	クラウド上のストレージサービス（Amazon S3、Google Cloud Storage、IBM Cloud Object Storageなど）に接続しデータを格納できる			○
70	10	データ蓄積	★★	クラウド	クラウド上のDWHサービス（Amazon Redshift、Google BigQuery、IBM Db2 Warehouseなど）にデータをロードし公開できる			
71	11	データ蓄積	★★★	クラウド	クラウド上のデータストアサービスが対象システムの機能要件・非機能要件に合致するか評価を行い、採用可否を判断できる			
72	12	データ蓄積	★★★	リアルタイム処理	リアルタイムに入力されるストリームデータから指定条件のイベントを即時に抽出する複合イベント処理（CEP）を実現するサーバー環境・構成を設計できる			
73	13	データ蓄積	★★★	リアルタイム処理	Apache Kafka、Apache Stormなどのストリーム処理プラットフォームを活用して、入力データをリアルタイムに処理して連携できる			
74	14	データ蓄積	★★★	リアルタイム処理	適切なライブラリや実行環境を選択して、ストリーム処理やイベント駆動アプリケーションを実装できる			
75	15	データ蓄積	★★★	キャッシュ技術	基盤設計において、どこのシステム要素にCache Serviceやmemcachedなどのキャッシュ機能を採用すると処理が高速化されるか判断できる			
76	16	データ蓄積	★★★	データ蓄積技術	データストア選定の際にリレーショナルDBだけでなく、要件に応じてNoSQL、グラフDB・時系列DBなどの適切なデータ蓄積機能を選定できる			
77	17	データ蓄積	★★★	検索技術	Apache Solr、Elasticsearchなどの全文検索エンジンを用いて、テキストキーワードやログの検索用DBを構築することができる			
78	1	データ加工	★	フィルタリング処理	数十万レコードのデータに対して、条件を指定してフィルタリングできる（特定値に合致する・もしくは合致しないデータの抽出、特定範囲のデータの抽出、部分文字列の抽出など）			○
79	2	データ加工	★	フィルタリング処理	正規表現を活用して条件に合致するデータを抽出できる（メールアドレスの書式を満たしているか判定をするなど）			
80	3	データ加工	★	ソート処理	数十万レコードのデータに対して、目的の並び替えになるように複数キーのソート条件を設定ができる			○
81	4	データ加工	★	結合処理	数十万レコードのデータに対して、単一条件による内部結合、外部結合、自己結合ができ、UNION処理ができる			○
82	5	データ加工	★	前処理	数十万レコードのデータに対して、NULL値や想定外・範囲外のデータを持つレコードを取り除く、または既定値に変換できる			○
83	6	データ加工	★★	前処理	フラットファイルやバイナリファイルに対するデータロードの前処理（クレンジング操作、禁則処理やバイナリ処理）ができる			
84	7	データ加工	★★	前処理	線形補間など、複数のレコードを考慮したクレンジング処理ができる			
85	8	データ加工	★	マッピング処理	数十万レコードのデータに対して、規定されたリストと照合して変換する、都道府県名からジオコードに変換するなど、ある値を規定の別の値で表現できる			○
86	9	データ加工	★★	マッピング処理	データ定義や実際の観測データの状況をもとに、一般的な名寄せ処理を設計・実装できる			
87	10	データ加工	★	サンプリング処理	数十万レコードのデータに対して、ランダムまたは一定間隔にデータを抽出できる			○

NO	SubNo	スキルカテゴリ	スキルレベル	サブカテゴリ	チェック項目	BZ	DS	必須スキル
88	11	データ加工	★	集計処理	数十万レコードのデータを集計して、合計や最大値、最小値、レコード数を算出できる			○
89	12	データ加工	★	変換・演算処理	数十万レコードのデータに対する四則演算ができ、数値データを日時データに変換するなど別のデータ型に変換できる			○
90	13	データ加工	★	変換・演算処理	変化元データと変換先データの文字コードが異なる場合でも変換処理ができる			
91	14	データ加工	★★	変換・演算処理	数千万レコードのカラムナ型データベースに対して効率的な処理を行うために、Parquetなどのカラムナファイル形式でデータを作成することができる			
92	15	データ加工	★★	変換・演算処理	手書き文字や画像などの非構造化データを、OCRや画像認識ライブラリ、CNN画像判定モデルなどを用いて分析可能な構造化データに変換できる		*	
93	1	データ共有	★	データ出力	加工・分析処理結果をCSV、XML、JSON、Excelなどの指定フォーマット形式に変換してエクスポートできる			○
94	2	データ共有	★	データ出力	加工・分析処理結果を、接続先DBのテーブル仕様に合わせてレコード挿入できる			○
95	3	データ共有	★	データ展開	データ取得用のWeb API（REST）やWebサービス（SOAP）などを用いて、必要なデータを取得できる			
96	4	データ共有	★★	データ展開	利用者の要件に合致したレポート（図、表）を、PDFやPostScriptなどの印刷用フォーマットで出力する変換機能を設計できる			
97	5	データ共有	★★	データ展開	BIツール用のサーバーを構築し、単一データソースのレポート用データマートを設計・構築できる			
98	6	データ共有	★★	データ展開	データストア上のデータをメールやメッセージ（Webサービスなど）を用いてプッシュ配信するシステムのサーバー・ネットワーク・ソフトウェアの構成を設計できる			
99	7	データ共有	★★	データ展開	Webアプリケーションの実装において、WebSocketを用いてクライアント側にリアルタイムにデータ提供できる機能を設計できる			
100	8	データ共有	★★	データ展開	RSS、RDFや業界標準フォーマットなど要件に合致したデータ形式・配信形式で、情報提供するシステムのインターフェースを設計できる			
101	9	データ共有	★★	データ展開	Web API（REST）やWebサービス（SOAP）などを用いて、必要なデータを提供するシステムの公開インターフェースを設計できる			
102	10	データ共有	★	データ連携	FTPサーバー、ファイル共有サーバーなどから必要なデータファイルをダウンロードして、Excelなどの表計算ソフトに取り込み活用できる			○
103	11	データ共有	★	データ連携	BIツールのレポート編集機能を用いて新規レポートを公開できる	*	*	
104	12	データ共有	★	データ連携	BIツールの自由検索機能を活用し、必要なデータを抽出して、グラフを作成できる	*	*	
105	13	データ共有	★★	データ連携	連携対象システムの仕様に合わせて、ETLツールを用いたデータ変換、ファイル転送処理を実装できる			
106	14	データ共有	★★★	データ連携	他データや他システム連携を容易にする目的で、適切なオープンデータ系の標準フォーマット（IMI：Infrastructure for Multilayer Interoperability[情報共有基盤]に記載のデータモデルや、GTFS：General Transit Feed Specificationなど）を用いて、データモデルを設定できる			
107	15	データ共有	★★★	データ連携	ESBやAPI Gatewayなどのデータ連携サービスを活用してシステム間のデータ連携（データ配信・交換）を行うインターフェースを設計できる			
108	1	プログラミング	★	基礎プログラミング	小規模な構造化データ（CSV、RDBなど）を扱うデータ処理（抽出・加工・分析など）を、設計書に基づき、プログラム実装できる			○
109	2	プログラミング	★	基礎プログラミング	プログラム言語や環境によって、変数のデータ型ごとに確保するメモリサイズや自動型変換の仕様が異なることを理解し、プログラムの設計・実装ができる			
110	3	プログラミング	★	基礎プログラミング	データ処理プログラミングのため分岐や繰り返しを含んだフローチャートを作成できる			○
111	4	プログラミング	★	基礎プログラミング	オブジェクト指向言語の基本概念を理解し、スーパークラス（親クラス）を継承して、スーパークラスのポリバティやメソッドを適切に活用できる			
112	5	プログラミング	★	基礎プログラミング	ホワイトボックステストとブラックボックステストの違いを理解し、テストケースの作成とテストを実施できる			○
113	6	プログラミング	★	拡張プログラミング	JSON、XMLなど標準的なフォーマットのデータを受け渡すために、APIを使用したプログラムを設計・実装できる			
114	7	プログラミング	★	拡張プログラミング	外部ライブラリが提供する関数の引数や戻り値の型や仕様を調べて、適切に呼び出すことができる			○
115	8	プログラミング	★★	拡張プログラミング	異なるタイプの複数の処理を効率よく行うために、スクリプトを用いたプログラムを設計・実装できる（パイプライン処理のluigiなど）			
116	9	プログラミング	★★	拡張プログラミング	TensorFlow、PyTorchなどのフレームワークを用いてニューラルネットワークを設計、実装できる		*	
117	10	プログラミング	★★	拡張プログラミング	SNSから抽出した非構造化データを、適切な言語やライブラリを選んでデータ処理のプログラム実装ができる		*	
118	11	プログラミング	★★★	拡張プログラミング	CUDAやOpenCLを活用してGPU上で稼働する並列プログラミングを実装できる		*	
119	12	プログラミング	★★★	拡張プログラミング	IoTデバイスに関連する各分野におけるデータ取得標準（HEMSなど）を理解しており、今後の技術動向や規制についてのリスクを提示できる			
120	13	プログラミング	★★★	拡張プログラミング	単一サーバーの物理メモリを超える複数のデータソースを組み合わせたデータ処理において、分散処理アーキテクチャやデータのインメモリ処理の特性を意識してプログラム設計ができる			
121	14	プログラミング	★	アルゴリズム	入れ子の繰り返し処理（二重ループ）など計算負荷の高いロジックを特定したアルゴリズムの改善策を検討できる			
122	15	プログラミング	★★	アルゴリズム	システム化対象業務のビジネスロジックを理解して、外部ライブラリやAPIを有効に活用し、プログラムを実装することができる	*		
123	16	プログラミング	★★★	アルゴリズム	論文や書籍に記載された計算式や機械学習ロジックを理解してプログラム実装できる		*	
124	17	プログラミング	★	分析プログラム	Jupyter Notebook（Pythonなど）やRStudio（R）などの対話型の開発環境を用いて、データの分析やレポートの作成ができる		*	○
125	18	プログラミング	★★	分析プログラム	分析プログラムのロジックと処理手順を理解した上で正しい分析結果を出力しているか検証ができる		*	○
126	19	プログラミング	★★	分析プログラム	比較的シンプルな分析課題において、クラウド、SaaS分析環境を用いてデータインポートから加工、集計、アウトプットまでの一連の実装ができる			
127	20	プログラミング	★	SQL	SQLの構文を一通り知っていて、記述・実行できる（DML・DDLの理解、各種JOINの使い分け、集計関数とGROUP BY、CASE文を使用した縦横変換、副問合せやEXISTSの活用など）			○
128	21	プログラミング	★★	SQL	RDBにおける分析関数の構文と挙動を理解し、分析関数を用いて複雑な副問合せや自己結合を解消できる			
129	22	プログラミング	★★	SQL	N：Nの結合や完全外部結合の危険性（計算量の増大、結果の不完全性）、暗黙の型変換の危険性（インデックス不使用、小数点以下の切り捨てなど）を考慮したSQLを記述できる			
130	23	プログラミング	★★★	SQL	記述したSQLの実行計画の確認と判断ができ、SQLの修正やインデックス作成により、処理時間を大幅に改善するようなパフォーマンスチューニングができる			
131	1	ITセキュリティ	★	基礎知識	セキュリティの3要素（機密性、完全性、可用性）について具体的な事例を用いて説明できる			○
132	2	ITセキュリティ	★★	プライバシー	ハッシュ化、マスキング、k-匿名化、差分プライバシーなどのプライバシー保護の仕組みを理解し適用できる		*	○
133	3	ITセキュリティ	★	攻撃と防御手法	マルウェアなどによる深刻なリスクの種類（消失・漏洩・サービスの停止など）を理解している	*	*	○
134	4	ITセキュリティ	★	攻撃と防御手法	OS、ネットワーク、アプリケーション、データなどの各レイヤーに対して、ユーザーごとのアクセスレベルを設定する必要性を理解している			

NO	SubNo	スキルカテゴリ	スキルレベル	サブカテゴリ	チェック項目	BZ	DS	必須スキル
135	5	ITセキュリティ	★★	攻撃と防御手法	DoS攻撃、不正アクセス、マルウェア感染や内部不正などのセキュリティインシデントが発覚した場合に既存のルールに基づき対応できる			
136	6	ITセキュリティ	★★	攻撃と防御手法	OS、ネットワーク、アプリケーション、データに対するユーザーごとのアクセスレベルを設計できる			
137	7	ITセキュリティ	★★	攻撃と防御手法	SQLインジェクションやバッファオーバーフロー攻撃の概要を理解し、防止する対策を判断できる			
138	8	ITセキュリティ	★★★	攻撃と防御手法	なりすまし、改ざん、盗聴などのセキュリティ侵害を防御するための対策とセキュリティポリシーを設計し実践できる			
139	9	ITセキュリティ	★★★	攻撃と防御手法	侵入検知システム（IDS）やファイアウォール、エンドポイント対策（EPP/EDR）などを用いて、外部からの不正アクセスを検知、防御、内部侵入後の対策をおこなう環境を設計できる			
140	10	ITセキュリティ	★	暗号化技術	暗号化されていないデータは、不正取得された際に容易に不正利用される恐れがあることを理解し、データの機密度合いに応じてソフトウェアを使用した暗号化と復号ができる			○
141	11	ITセキュリティ	★	暗号化技術	なりすましや改ざんされた文書でないことを証明するために、電子署名が用いられることを理解している			○
142	12	ITセキュリティ	★	暗号化技術	公開鍵暗号化方式において、受信者の公開鍵で暗号化されたデータを復号化するためには受信者の秘密鍵が必要であることを知っている			○
143	13	ITセキュリティ	★	暗号化技術	ハッシュ関数を用いて、データの改ざんを検出できる			
144	14	ITセキュリティ	★★	暗号化技術	SSHやSSL/TLSなどのセキュアプロトコルの概要と必要性を説明できる			
145	15	ITセキュリティ	★	認証	OAuth認証が求められるデータ提供サービスに対して、認証処理で取得したトークンを付与してデータ取得用のREST APIを呼び出すことができる			
146	16	ITセキュリティ	★★	認証	Kerberos認証とRadius認証の違いを理解し、それぞれの認証の特徴やユースケースを説明できる			
147	17	ITセキュリティ	★★	認証	SAMLやOAuthを用いて一度のログインで複数のWebアプリケーションのログイン認証を連携するシングルサインオンの仕組みを設計できる			
148	18	ITセキュリティ	★★★	ブロックチェーン	ブロックチェーン技術を用いてストレージに蓄積されたデータの安全性と品質を保証するシステムを設計できる			
149	19	ITセキュリティ	★★	ゼロトラスト	ゼロトラストの概念を理解し、クラウド利用やリモートワークに対応した情報セキュリティの担保と、データ活用の利便性を両立させることができる			
150	1	AIシステム運用	★	ソース管理	GitやSubversionなどのバージョン管理ソフトウェアを活用して、開発した分析プログラムのソースをリポジトリに登録しチームメンバーと共有できる			
151	2	AIシステム運用	★	AutoML	AutoMLを用いて予測対象を判定するために最適な入力データの組み合わせと予測モデルを抽出できる	*	*	
152	3	AIシステム運用	★	MLOps	MLOpsの概要を理解し、AIモデル性能の維持管理作業の基本的な流れを説明できる			
153	4	AIシステム運用	★	MLOps	AIシステムのモニタリング項目を理解し、AIモデルの劣化状況や予測対象データの不偏、AIシステムの異常を検知できる			
154	5	AIシステム運用	★★	MLOps	AIシステムの予測モデル評価において、求められている予測性能を満たしているかを評価し、業務システムへの適用を判断できる			
155	6	AIシステム運用	★★	MLOps	AIモデルの管理方針（更新頻度、バージョン管理、リリース手順など）を決定し、適切にAIモデルを運用できる			
156	7	AIシステム運用	★★★	MLOps	AIシステム運用において、モデル作成からモデルの評価、デプロイ、学習データの更新、モデル再学習のパイプラインを設計し、AIモデルのライフサイクルを自動化できる			
157	8	AIシステム運用	★★★	MLOps	AIシステムに求められる要件（予測性能、リアルタイム性、コストなど）を考慮し、適切な学習環境および分析環境（サーバ/エッジなど）を設計できる			
158	9	AIシステム運用	★	AIOps	ITシステムの運用におけるAIOpsの概要とメリットを説明できる			
159	10	AIシステム運用	★★	AIOps	ITシステムのアプリケーションログと過去の障害実績より、統計的手法や機械学習を用いて大量のログから障害の予兆を検出し障害を未然に防止するシステムを設計することができる			

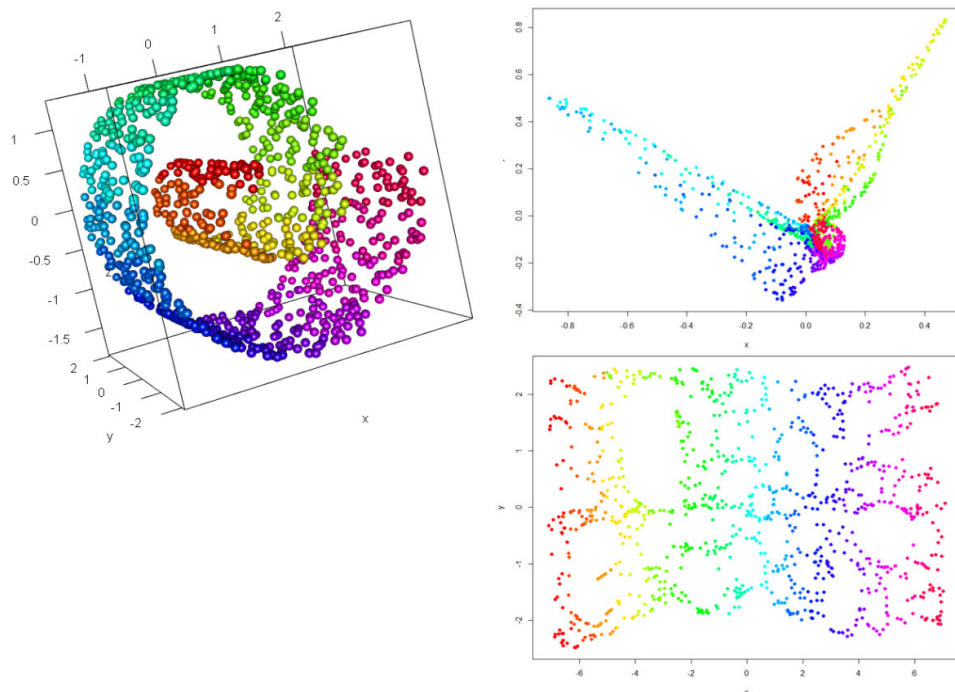
【付別紙】

スキルチェックリスト：データサイエンス力

「データ可視化」に関するイメージ共有のための参考資料（例示）

NO

2 データの持つ分散量の観点で、高次元のデータを主成分分析（PCA）などにより1～3次元のデータに変換できる



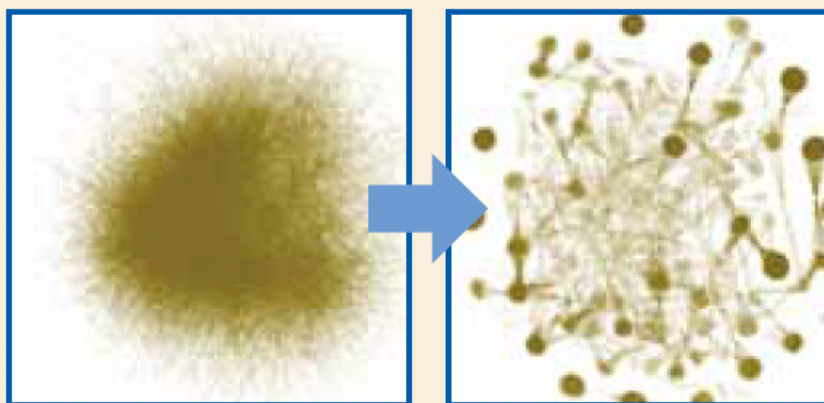
元：株式会社ALBERT WEBサイト

<https://blog.albert2005.co.jp/2014/12/11/>

4 ネットワーク構造、グラフ構造などの表現において、ノードとエッジが増えすぎて特徴抽出が困難であっても、データの絞り込みや抽象度を上げることで適切に可視化できる

図2 企業間取引データをクラスター化して把握しやすくした例

	研磨前	研磨後
頂点数	3,282	3,282
枝数	35,168	73,132
クリーク数	32,953	343



元：（出版物）大学共同利用機関法人 情報・システム研究機構 国立情報学研究所 発行

http://www.nii.ac.jp/userdata/results/pr_data/NII_Today/66/all.pdf

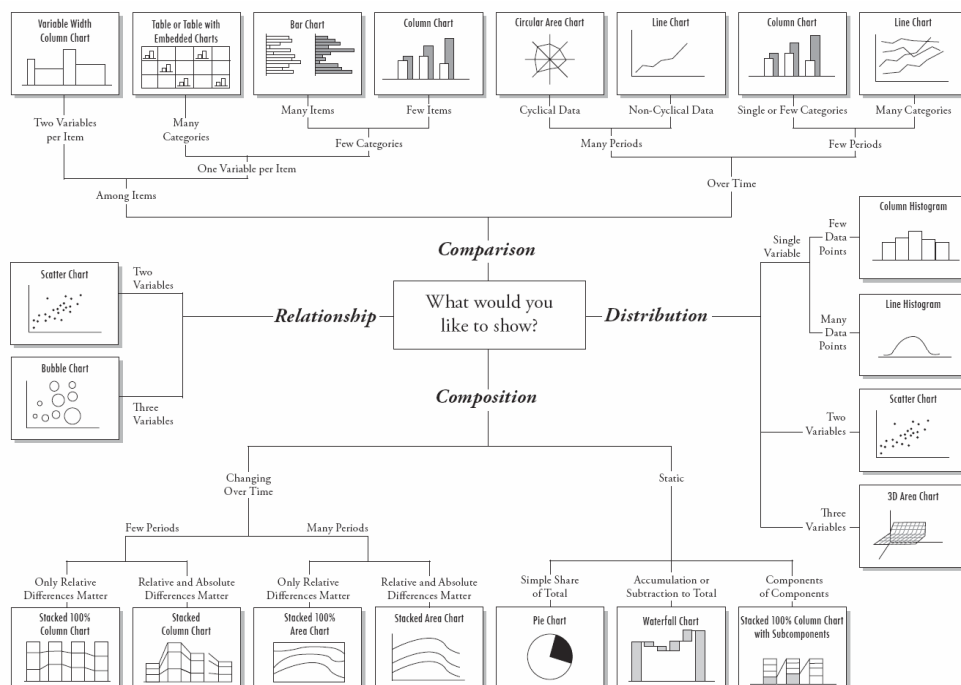
強調表現がもたらす効果と、明らかに不適切な強調表現を理解している（計量データに対しては位置やサイズ表現が色表現よりも効果的など）

表現例	量を表現する	順序を表現する	分類を表現する	関係を表現する
位置 	✓	✓	✓	✓
長さ 	✓	✓		
明度/彩度・色 		✓		
			✓	

元：「solutions-2」WEBサイト

<http://www.solutions2.be/Books.aspx>（2015年11月時点）をもとに独自に修正加筆

2 図表化できる

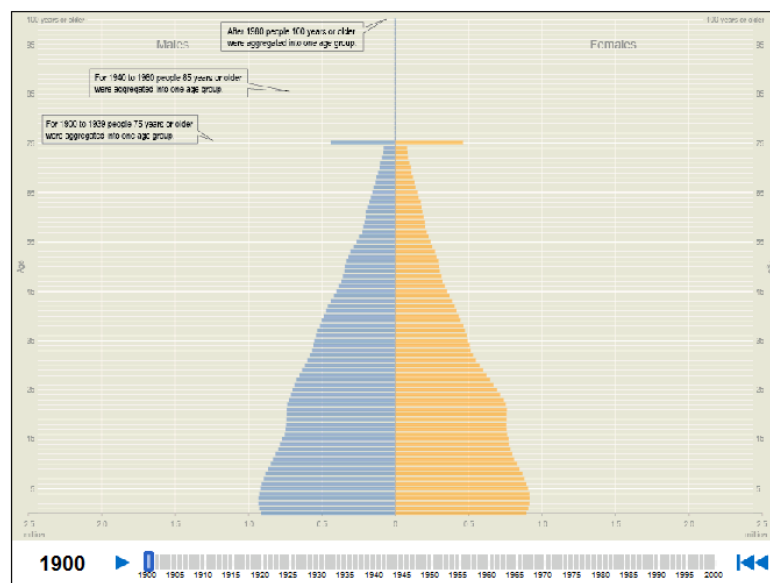


元：Amit Agarwal 「Digital Inspiration」WEBサイト

<http://www.labnol.org/software/find-right-chart-type-for-your-data/6523/>

A Century of Population Change in the Age and Sex Composition of the Nation

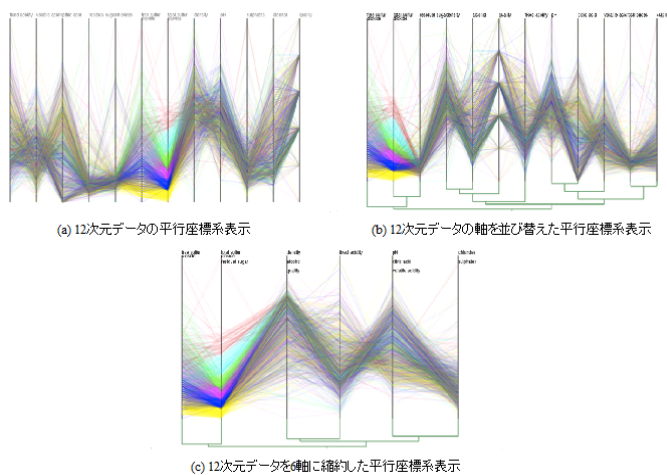
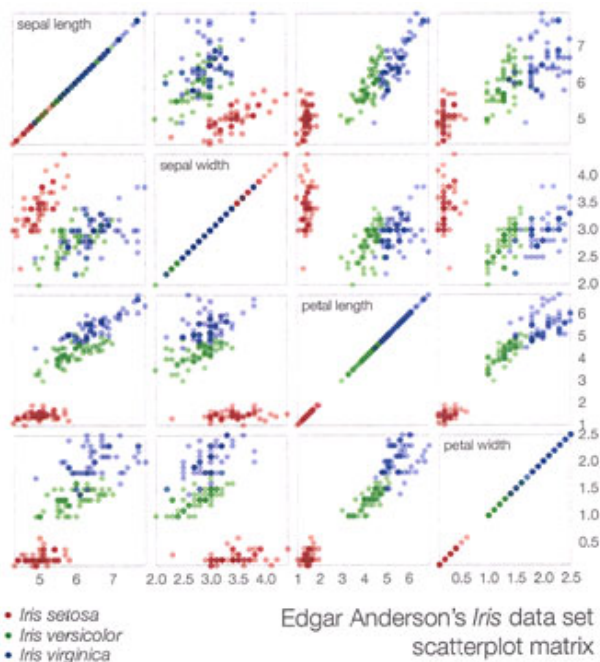
September 12, 2013



元：アメリカ国税調査局 WEBサイト

<http://www.census.gov/dataviz/visualizations/055/>

1 1～3次元の図表を拡張した多変量の比較を適切に可視化できる（平行座標、散布図行列、テーブルレンズ、ヒートマップなど）

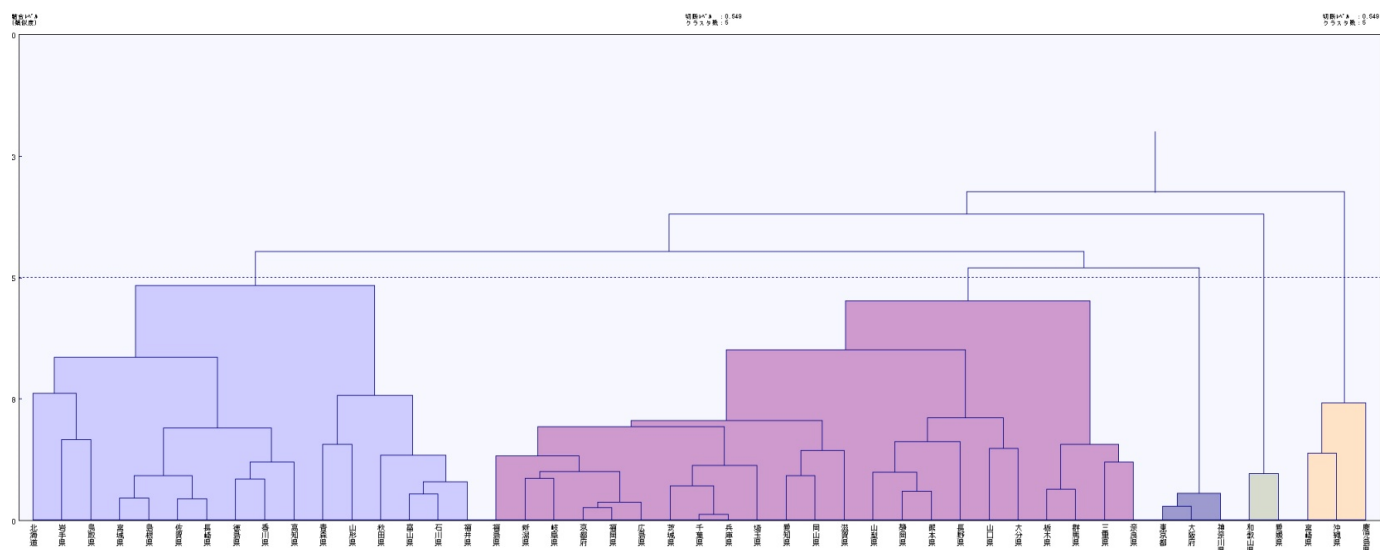


元：[左図]「Data-Driven Documents」WEBサイト <http://mbostock.github.io/d3/talk/20111116/iris-splom.html>

[右図]「グラフスペクトル解析を用いた軸縮約可能平行座標系」WEBサイト

<http://web-ext.u-aizu.ac.jp/~shigeo/research/cpc/index-j.html>

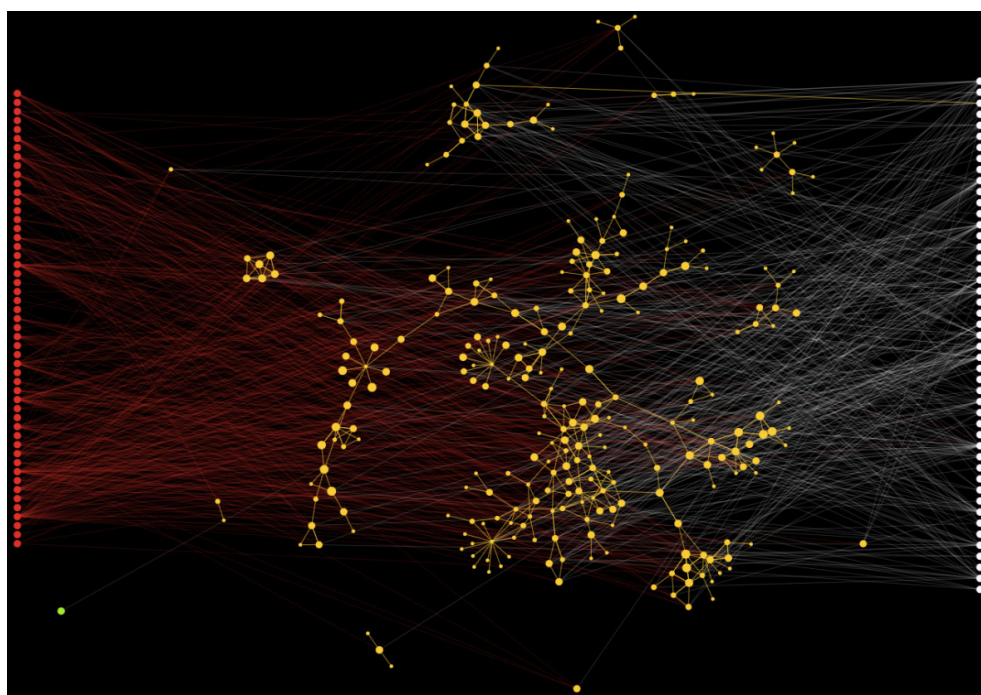
5 ネットワーク構造、グラフ構造、階層構造などの統計的な関係性の可視化ができる



元：株式会社ALBERT WEBサイト

https://www.albert2005.co.jp/knowledge/images/tech_mining_img111.jpg

5 ネットワーク構造、グラフ構造、階層構造などの統計的な関係性の可視化ができる



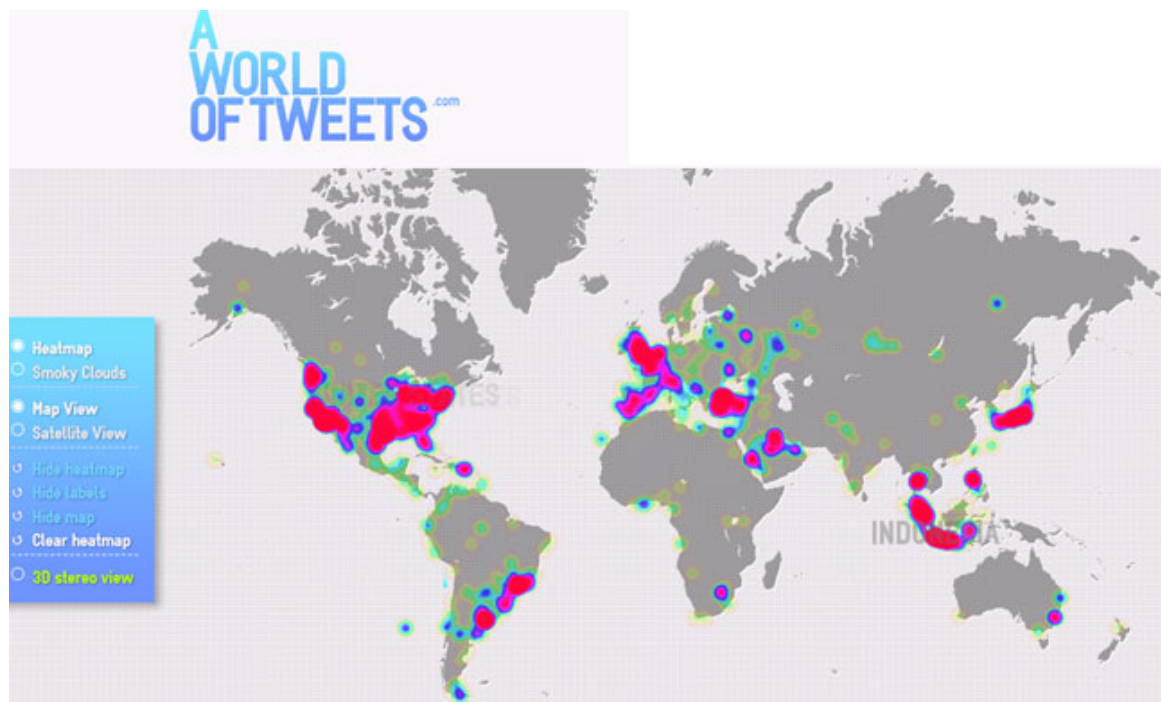
赤・白点：Win

黄点：Cheese

元：「Cytoscape.js」WEBサイト (Wine & cheese)

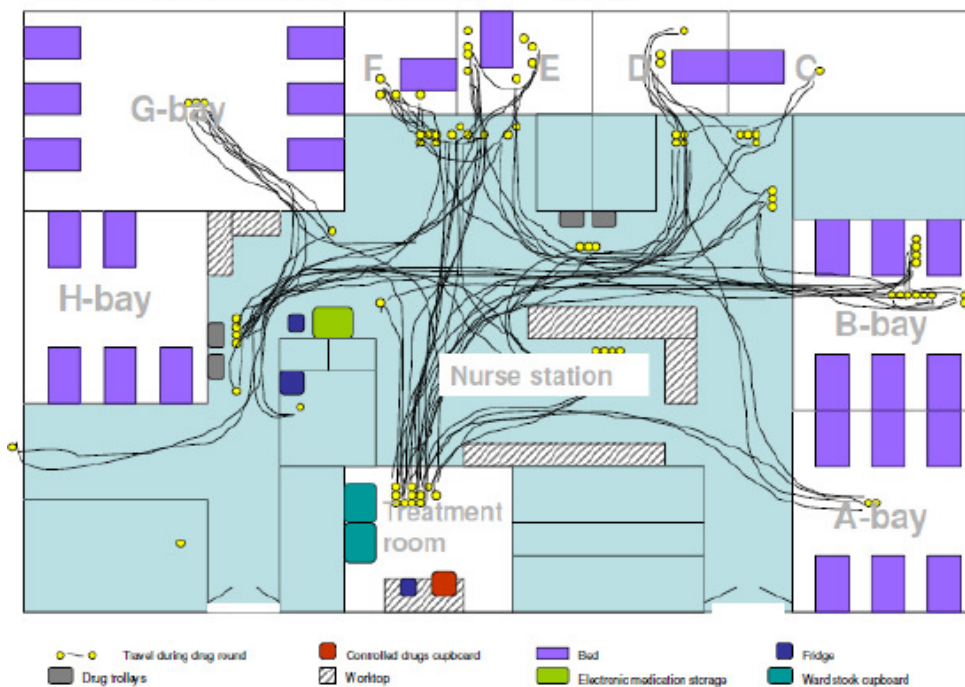
<http://cytoscape.github.io/cytoscape.js/demos/cde4db55e581d10405f5/>

5 地理空間情報などを地図に重ね合わせた可視化ができる（コロプレスマップ、ボロノイ図など）



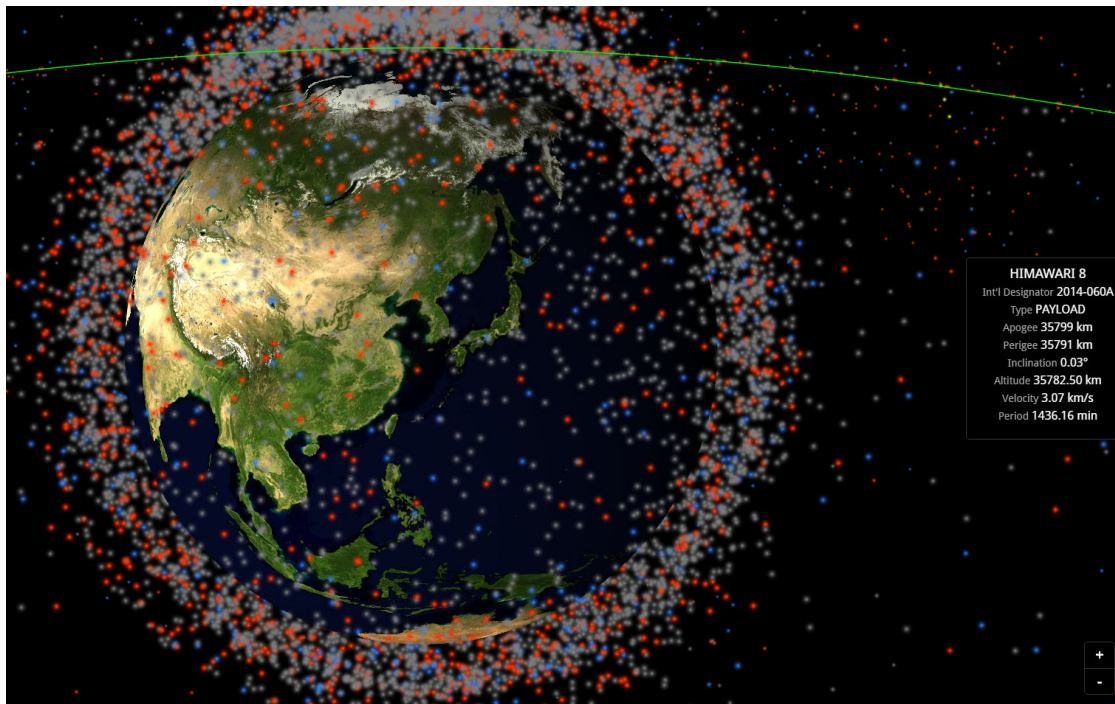
元：「A WORLD OF TWEETS」WEBサイト（2015年11月時点）
<http://aworldoftweets.frogdesign.com/>

7 や心拍数など



元：「Imperial College London」WEBサイト（2015年11月時点）
<https://www1.imperial.ac.uk/cpssq/ms/m/>

) 人体、標高を持つ地図、球面など3次元空間上にデータを重ね合わせた可視化ができる



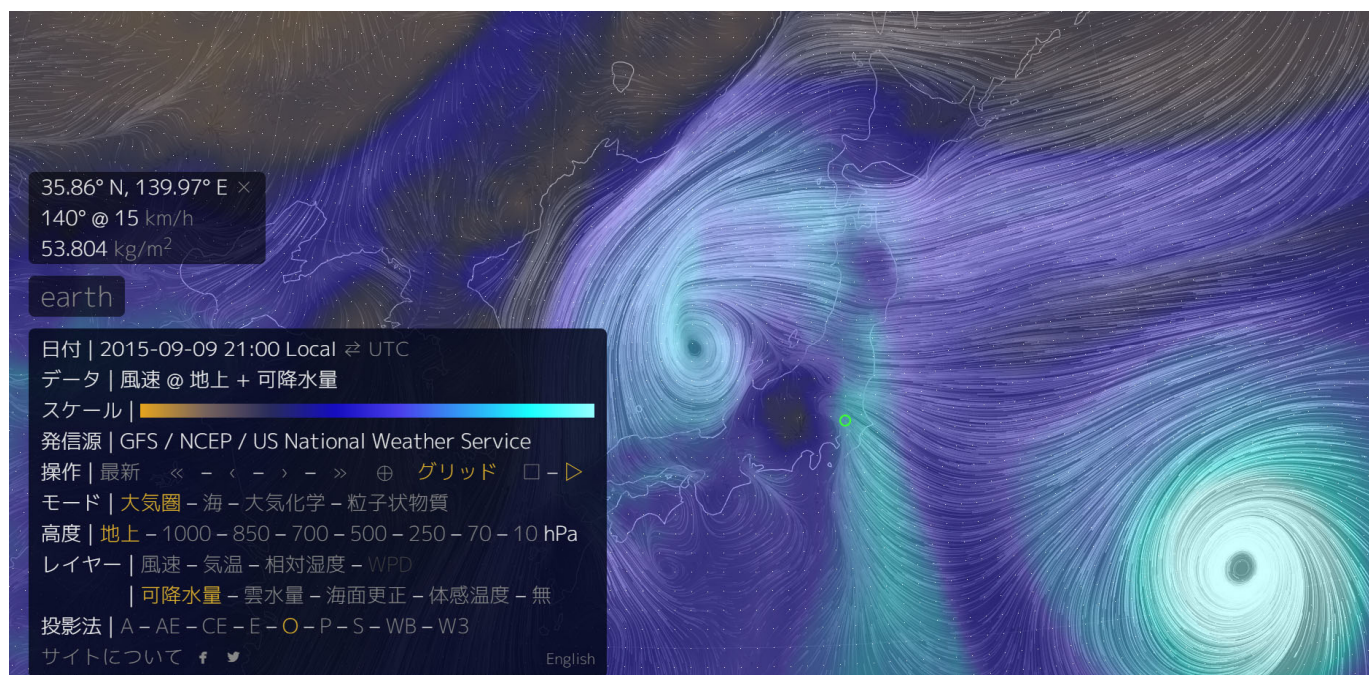
元 : James Yoder 「Stuff in Space」 WEBサイト
<http://stuffin.space/>

跡など)



元： Kaspersky Lab 「CYBERTHREAT REAL-TIME MAP」 WEBサイト
<https://cybermap.kaspersky.com/>

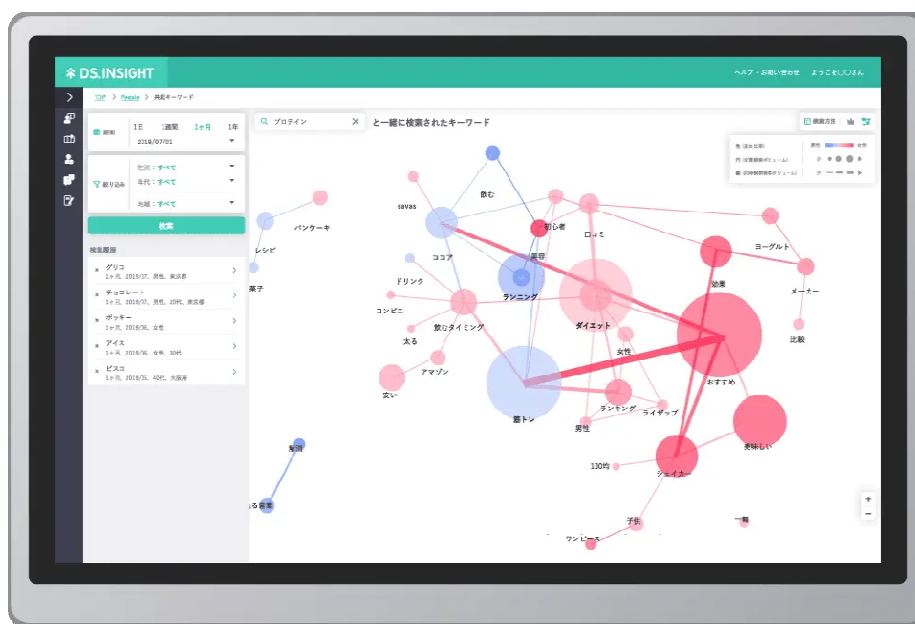
跡など)



元：Cameron Beccario「earth」WEBサイト

http://earth.humanschool.net/jp/#2015/09/09/1200Z/wind/surface/level/overlay-total_precipitable_water/orthographic--224_70_38_05_3000/grid-on

テム環境への負荷調整なども踏まえた可視化ができる



元：ヤフー株式会社 Yahoo Japan Corporation「DS.INSIGHT」WEBサイト

<https://ds.yahoo.co.jp/insight/trial-account.html>