

社会人のためのデータサイエンス演習

第5週:ビジネスでデータサイエンスを実現するために

第5回:組織におけるデータサイエンスの実現

講師名:杉山 聡

第5週の内容紹介

第1回

● 各週のおさらい

第2回

● データ分析に基づく問題解決ケーススタディ①

第3回

● データ分析に基づく問題解決ケーススタディ②

第4回

● 様々な組織で活躍するデータサイエンティスト

第5回

● 組織におけるデータサイエンスの実現

第6回

講座のまとめ①

第7回

講座のまとめ②

ータサイエンス活用における課題

データを活用して 成果をあげたい



課題

- プロジェクトを遂行できる人材が 集まらない
- ・ 必要データが揃わない・ データが分析できる状態にない

- 経営層の理解が得られない
- 企業方針・風土とのミスマッチ

分析実行の体制を整える

データサイエンス活用に重要な3つのこと

人材育成·調達

データサイエンス活用のプロジェクトを 遂行できるスキルを持つ人材

分析環境の整備

必要データを取得・加工・活用できる システム基盤の構築

組織風土の改革

現場、経営層ともに、組織をあげたデータサイエンス活用に対する理解と共通意識の浸透

人材育成・調達

人材育成·調達

必要な人材のスキルやスキルレベルの定義 を行い、プロジェクトを組む

内部での育成

- トレーニングによる育成他部門からの登用
- 組織全体の底上げ

新規採用

- ・ 経験者の採用(独自採用、エージェント等)
- データサイエンスを学んだ学生の採用

外部連携

- 分析コンサルティングサービスの活用
- 分析ツール / システムサービスの導入

分析環境の整備

分析環境の整備

● 必要なデータの取得・分析が行える環境整備をする



● データサイエンスを組み込んだ エンタープライズアーキテクチャ(EA)

データ基盤の構築 (データソース、DWH, アナリティクス環境)

データオーナーシップの明確化とデータ共有

組織風土の改革

経営層

● データ活用の組織方針、組織体制づくり

経営意思決定に おけるデータサ イエンスの活用

人材の育成、採用 プロジェクトアサイン

実業務上の 課題把握と解決

事業マネジメント層

事業推進におけるデータ活用プロジェクトの発足

運用スタッフ

データサイエンススキル の習得、向上

組織全体のデータ活用に対する理解浸透・意識改革も必要

【事例】データ分析基盤を構築。組織全体で「デジタルの民主化」を実現

株式会社LIXIL(リクシル)

概要

自社のデータソースを統合し、全社で使える分析基盤を構築。特定の人材に依存せず、データに基づく意思決定を従業員全員が行うための基盤構築と、スキル定着のための教育の仕組みを構築。

組織内における活用事例



数か月先の売上を予測し、営業活動を効率化

営業部門では、過去実績の可視化に加え、数か月先の未来の 指標を可視化し、重点訪問先のターゲティングをするなどし、 営業活動の効率化を実現した。

今後の展開

組織内のデジタルの民主化加速のため、以下の取り組みを実施。

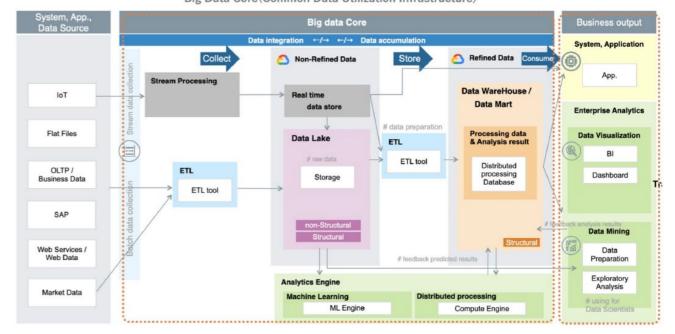
- ・マネジメント層向けデータリテラシー向上トレーニングの実施
- ・全従業員が手軽に使えるAI活用ツールを独自開発

LDP (LIXIL Data Platform) の構造

①Big Data Core(共通データ活用基盤)

2 Tool

Big Data Core(Common Data Utilization Infrastructure)



次回のテーマ

次回は

「講座のまとめ①」

お疲れ様でした!



社会人のためのデータサイエンス演習

第5週:ビジネスでデータサイエンスを実現するために

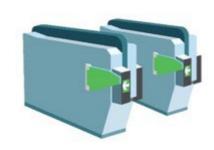
第6回:講座のまとめ①

講師名:菅 由紀子

社会におけるデータサイエンスの役割



オンラインショッピング SNS 動画配信サイト など



交通ICカードなどの 移動履歴データ

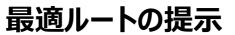


車載カーナビの 移動履歴データ

データサイエンスを活用し、社会課題の解決へ

快適な買い物の実現 生活者ニーズの把握







混雑·渋滞緩和

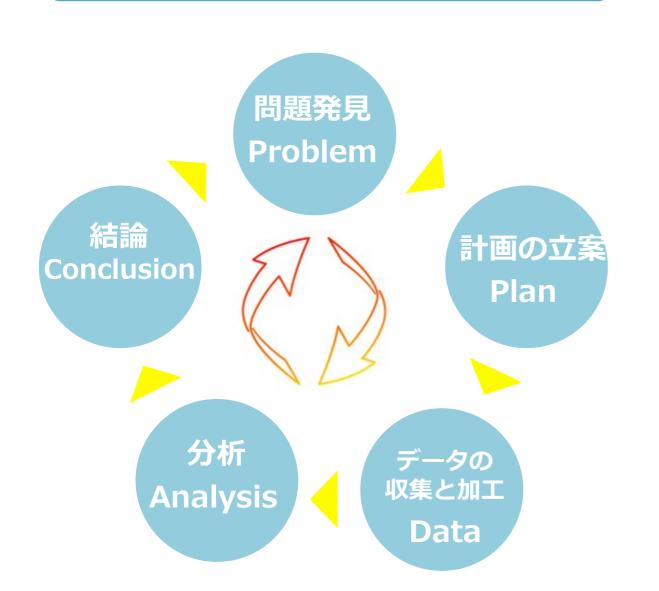


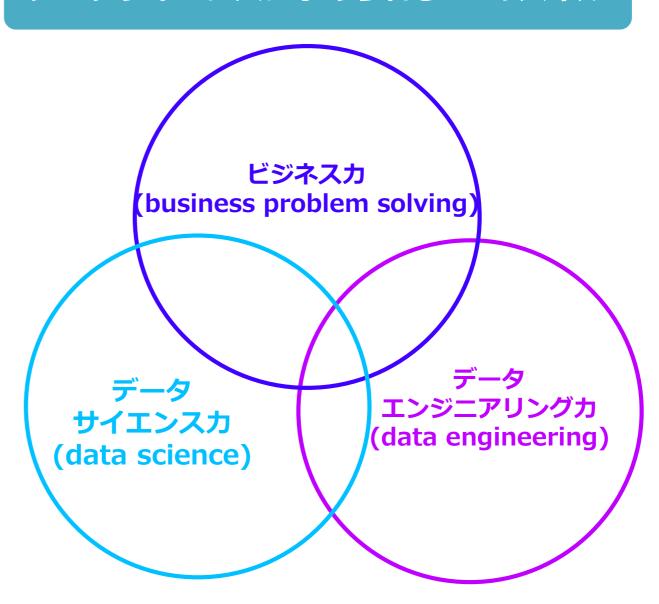
など

データサイエンスによる問題解決

データサイエンスの問題解決プロセス

データサイエンスに求められる3つのスキル





変数の状況把握 [データの尺度]

分析の第一歩は「変数間の関係性の把握」

4つのデータ尺度

質的

名義尺度

・分類を目的とした尺度。順序に意味がない 例:性別、都道府県、血液型など

順序尺度

・順序に意味がある尺度。間隔は定かではない 例:満足度評価、順位

量的

間隔尺度

・順序尺度で、かつ"間隔"に意味がある尺度。0に意味がない 例:温度、暦年(和暦、西暦)

比例尺度

・間隔尺度で、かつ"原点"に意味がある尺度 距離・大きさ、体重、お金

変数の状況把握 [可視化]

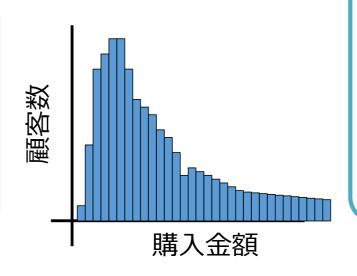
クロス集計

	男	女		
サイトA	18	3		
サイトB	4	16		

質的データの分布の可視化に用いられる。データを比較したい項目の要素ごとに分け、表を作成する。

列、行・・・各項目の要素

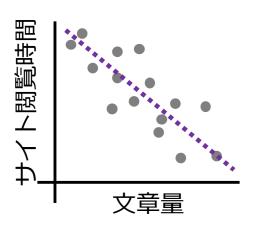
ヒストグラム



量的データの分布の可視化に用いられる。データを階級ごとに分け、度数分布表を作成し、グラフ作成を行う。

横軸・・・データの階級 縦軸・・・階級ごとのデータ

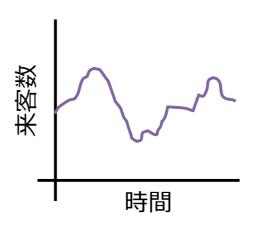
散布図



量的データの可視化に用いられる。 縦軸と横軸の項目に対して各データ が当てはまるところに打点して作成 する。

横軸、縦軸・・・各項目

時系列グラフ



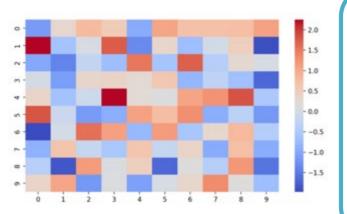
時系列データの可視化に用いられる。 データを時間ごとに分け、グラフ作 成を行う。

横軸・・・時間

縦軸・・・時間ごとのデータ

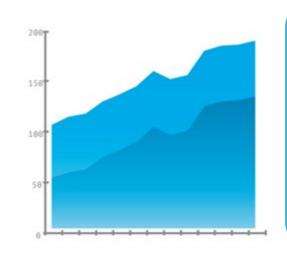
変数の状況把握 [可視化]

ヒートマップ



行列型のデータを、色の濃淡によって分布度合いを把握する方法。 色調で直感的にデータの把握を行うことができ、地図などと組み合わせて地理的な特性の把握に使われることもある。

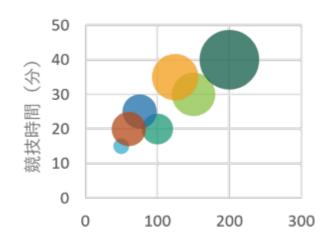
エリアチャート



時系列データを折れ線グラフで表し、数値の大きさを面積で可視化 する。

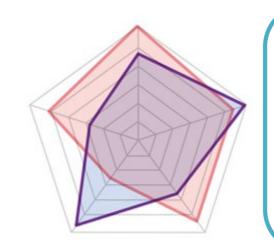
色づけによって視覚的にデータの 大小関係を把握しやすくする。

バブルチャート



散布図を応用した可視化の方法。 データの分布に対してそれぞれ量 的データの値を円の大きさで表し たもの。

レーダーチャート

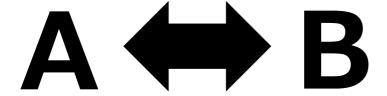


複数のデータの比較を正多角形 上で放射状に表現したグラフ。 系列間のバランスや特徴を把握 する上で有効な可視化手法。

相関と因果の違い

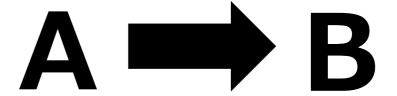
- 相関は、ある変数が変化すると、他方の変数も同時に変化する関係
- 因果は、ある変数が、他方の変化を引き起こす関係 (一方通行の関係、原因と結果)

相関関係



背が高い人ほど体重も重い

因果関係



台風で航空便が欠航

相関と因果は異なることに注意

統計的因果推論

統計的因果推論とは、ある事柄(原因)が、生じる効果(結果)に どの程度影響を及ぼしているかを統計データを通して定量的に評価する方法論。

ある事柄(原因)

どの程度影響を及ぼしているか

ある事象

例:ある商品の売上分析結果

商品







インターネット広告経由の購入者が多い



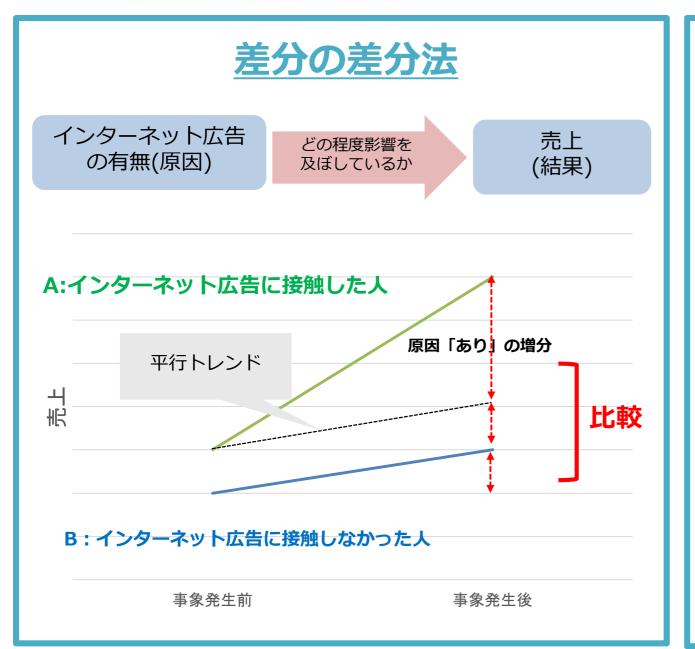
インターネット広告は商品の購入に影響を及ぼすのか?

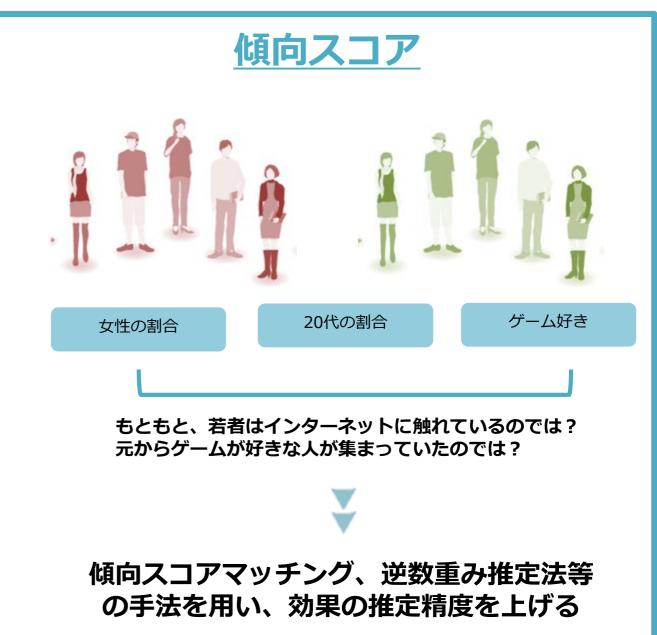
購入者がたまたまインターネット使用者が多かっただけかも?(若者に人気、など)

なるべく同じ条件になる ように補正しつつ 評価・推定したい

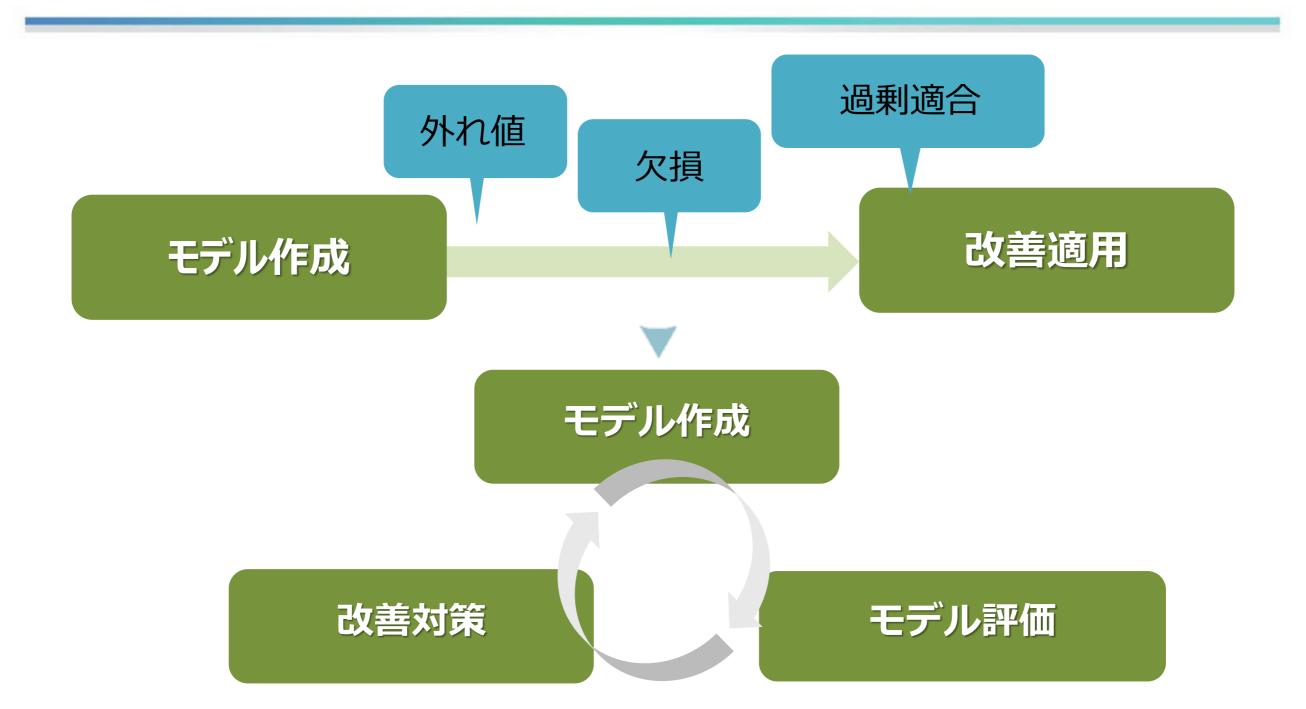
インターネット広告に触れなかった場合の売上は?

統計的因果推論の手法





モデル評価のプロセス



仮説検定とは

- 仮説に対して、正しいか否かを統計学的に検証する手法
- 統計的な判断として、差に意味があるかを明らかにする

有意差とは

立てた仮説と結果の差について、統計的に意味があるものを「有意差」という。統計調査などによって得られた2つの値の差が、統計的に信頼できるものか、偶然のものかどうかを判定する方法を有意差検定という。

ミュージカル鑑賞の1年間の平均支出金額

男性・女性それぞれ100人に対してのアンケート



男性 1500円



女性 3000円



男性 1500円



女性 1550円

男女の結果に差があるといえる

男女の結果差はある・・・?

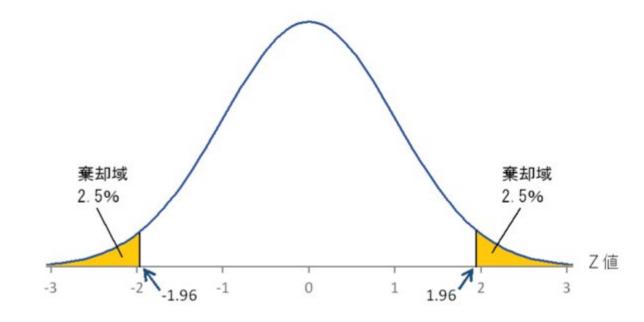
仮説検定における有意確率

P値(有意確率)とは

帰無仮説において、その結果以上または以下がでる確率を指す。母集団から 無作為抽出したサンプルの標本平均や標本標準偏差から、それらの母平均が 等しいと言えるかを判定する。

P値(有意確率)は一般的に、0.05(5%) を境に、有意水準とされる。

(0.05を下回ると、「偶然ではない」 「有意である」と判断されることが多い)



総務省「なるほど統計学園」

https://www.stat.go.jp/naruhodo/11_tokusei/kentei.html

次回のテーマ

次回は

「講座のまとめ②」

お疲れ様でした!



社会人のためのデータサイエンス演習

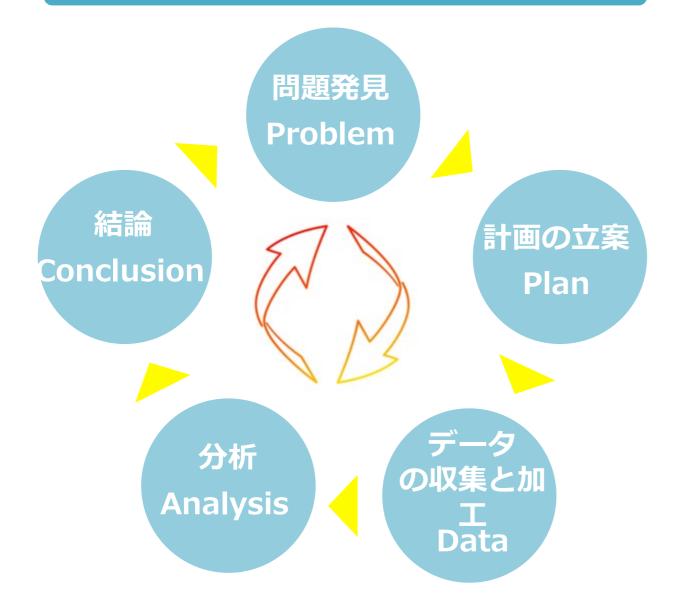
第5週:ビジネスでデータサイエンスを実現するために

第7回:講座のまとめ2

講師名:菅 由紀子

組織におけるデータサイエンス実現のポイント

データサイエンスの問題解決プロセス



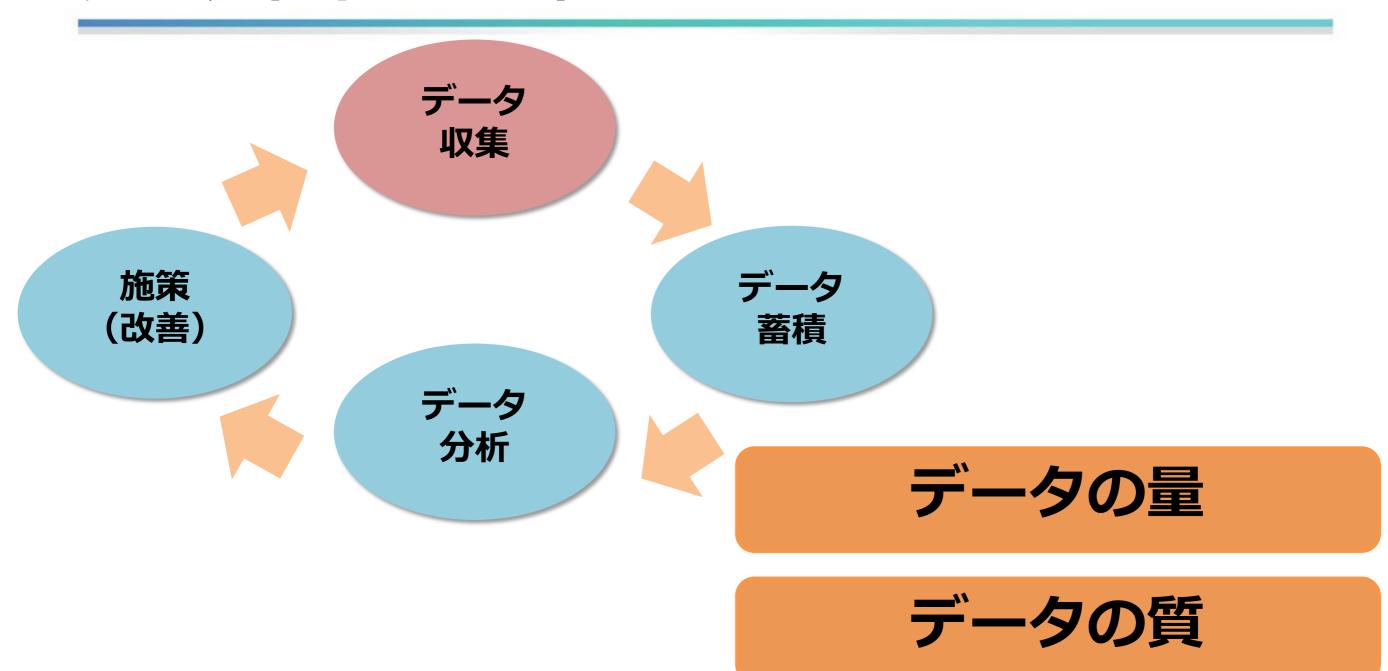
実施体制の確保

人材育成·調達

分析環境の整備

組織風土の改革

データ収集の重要性



さまざまなデータ

データの一例

内部データ

ウェブサイトアクセスデータ

サイト内閲覧データ、購買データ、その他登録情報、 問い合わせ履歴

会員情報

顧客アンケート

ポイントカード履歴

外部データ

調査会社等により取得されたデータ

マーケティングリサーチを目的に行われた アンケート調査結果

外部サービスによる販売データ

POSデータ、市場動向調査など

公的統計

政府統計の総合窓口 "e-Stat"



統計で見る日本

e-Statは、日本の統計が閲覧できる政府統計ポータルサイトです



お問い合わせ | ヘルプ | English

ログイン

新規登録

統計データを探す 統計データの活用 統計データの高度利用 統計関連情報 リンク集

●統計データを探す (政府統計の調査結果を探します)

その他の絞込

in すべて

政府統計一覧の中から探します

会 分野

17の統計分野から探します

1 組織

統計を作成した府省等から探します

キーワード検索: (例:国勢調査

検索

●統計データを活用する

ゴゴグラフ

主要指標をグラフで表示 (統計ダッシュボード)

時系列表

主要指標を時系列表で表示 (統計ダッシュボード)



地図上に統計データを表示 (統計GIS)



地域

都道府県、市区町村の 主要データを表示

利用ガイド

●統計データの高度利用

ミクロデータの利用

公的統計のミクロデータの利用案内

開発者向け

API、LODで統計データを取得

●統計関連情報

統計分類・調査計画等

政府統計の総合窓口 "e-Stat"

分野から探す

統計データを17の分野から探します。 ※主な調査には「基幹統計」を表示しています。



データの数、種類が豊富

整備済のデータ

政府統計の総合窓口 "e-Stat"

データセット一覧

< 戻る

政府統計名	<u>作物統計調査</u>
提供統計名	作物統計調査
提供分類1	作況調査(果樹)
提供分類2	<u>速報</u>
提供分類3	<u>令和3年産日本なし、ぶどうの結果樹面積、収穫量及び出荷量</u>
提供周期	<u>年次</u>

<u>ل</u>	挿入		描画 ペ	ージ レイア	ウト 数式	データ	校閲	表示(〉操作ア	シスト							
۱,۰			MS Pゴシ	ノツク	v 11	- A^	A :	= = [87	· 8	b 折	り返して	全体を表示する	5 v	標準		~
スト	□ ~		ВІ		· 4 ·	Α ν	bc,	= = =	= ←=	∓ ≡	→ +	ルを結合	して中央揃え		<u>~</u> ~ %	9	€0 .00 0,€ 00.
						-	А					77 C (IC IC)	D C I Min	- L	EB - 70	, ,	.00 →.0
	*	×		*													
С	D	E	F	G	Н	I	J	K	L	М	N	0	Р	Q	R	S	Т
統	計表】	4															
1	令和3	年産	₹日本なしの	結果樹面積、	10 a 当たり	収量、収穫	量及び出荷	量			+						
全		围		10 a 当たり	100 410 7	NEC NIX	<u> </u>	対 前 年	産 比								
都	· 道 府 J		結果樹面積	収量	収 穫 量	出 荷 量	結果樹 面 積	10 a 当たり 収 量	収穫量	出荷量	ik						
^		-	ha 10, 300	kg	184, 700	172, 700	%		%		%						
全		国		1, 790			96		108		09						
うち		成	136	1, 120	1,520	1, 360			63		64						
	5101 0	形	162 110	1, 030 1, 100	1, 670 1, 210	1, 520 1, 000		1	75 79		75 78						
		1 1	789	1, 510	11, 900	11, 000	96		92		92						
		成	894	2, 140	19, 100	17, 700	97		141		44					1	
		k	730	2, 180	15, 900	14, 900	99	142	141		41						
	群!	馬	202	2,000	4, 040	3, 900	99	108	107	1	06						
		E	322	2, 010	6, 470	6, 050	98		104		04						
		能	1, 290	1, 590	20, 500	20, 200			113		12					-	
	東神奈月	気	80 211	2, 030 1, 840	1, 620 3, 880	1, 590 3, 800		270000	104		05 11					-	
		易易	376	1, 600	6, 020	5, 540			87		87				-	-	
		LI LI	160	1, 500	2, 400	2, 330	99		81		81						
		11	133	1, 600	2, 130	2, 020	100		116		18						
	福	#	65	1, 330	865	809	98	115	113	1	14						
	長	野	648	1, 850	12,000	11,500	94	93	88	3	89						
		4	117	1, 540	1, 800	1, 650	100		96		96						
		EII	304	1, 580	4, 800	4, 510		500000	101		03						
		買	46	1, 120	515	470	100		101		01					-	
		郵車	64 62	1, 660 1, 470	1,060	1, 000 851	93 100		123 103		27 03				_	-	
		反	599	1, 470	11, 100	10, 200	92		108		06						
			136	1, 230	1, 670	1, 500	100		74		74						
		1	162	1, 670	2, 710	2, 460	98		124		24						
		1	200	2, 180	4, 360	4, 080			107	1	07						
		II	32	1, 240	397	362			115		15						
		岡	339	2, 210	7, 490	7, 140			124		24						
		買	172	2, 270	3, 900	3, 660			122		24					-	
	熊	*	450	1, 760	7, 920	7, 290	98	116	113	1	13						

表番号	統計表	調査年月	公開(更新)日	表示・ダウンロード
令和3年産日本	なし、ぶどうの結果樹面積、収穫量及び出荷量			
	統計表(速報)	2021年	2022-02-15	≵ EXCEL

求められる「データ倫理」



データサイエンティスト協会

「データサイエンティスト スキルチェックリスト」
「2021年版 タスクリスト ver.3」にデータ倫理に関する項目が追加

〈ビジネスカ〉

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 必要スキル		必要スキル
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			0
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			0
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			0
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)			0
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			0
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)	*		0
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			0
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関係する直近の法令・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	*		