本日は自由席です

青森県庁市町村課｜市町村共通課題解決分科会・第2回

以下のシートを開いてください（前回と同じリンクです）

https://00m.in/fAGww

【整える】
データクレンジング

青森県庁市町村課｜市町村共通課題解決分科会・第2回

本日のゴール｜

データ整形の重要度を正しく認識する

前回のおさらい

【おさらい】フォームによるデータ収集と利活用

なぜフォームなのか

- 「収集」と「集計」が同時に完了するから
- 回答者が送信した瞬間にデータベース化が完了
- 転記ミス・手作業ゼロ/ 予算ゼロ・手続きなしで今日から使える

こんな業務はもうやめましょう！

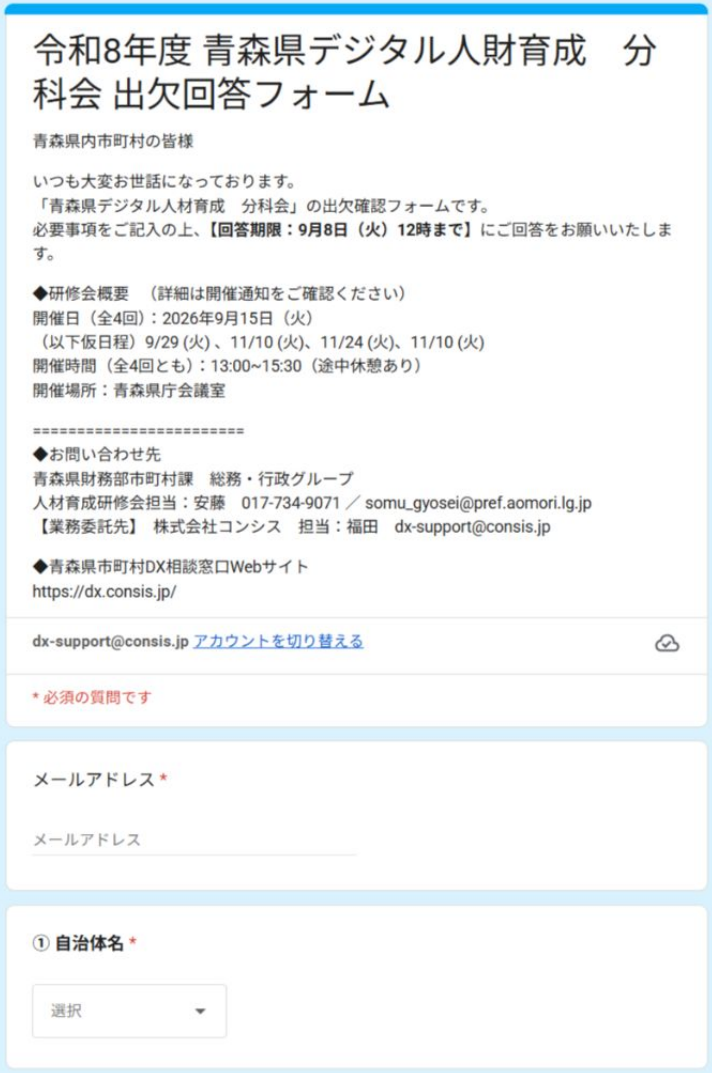
- 各課にWordやExcelの様式をメールで配って回収
- 届いたファイルを1冊ずつ開いて、手作業で1つのファイルに転記・集計
- 誤字脱字や入力漏れを1件ずつ目視でチェック、電話で確認する

【おさらい】
Googleフォームの使い方

集計用スプレッドシートと自動連携し、リアルタイムでのデータ可視化・分析が可能

AI・スプレッドシートとの強い親和性

回答が即座にデータベース化され、生成 AI（Gemini等）やBIツールと直接連動できるため、「データ収集から AI分析までの一連の体験」を学ぶのに適しています。


令和8年度 青森県デジタル人財育成　分科会 出欠回答フォーム

青森県内市町村の皆様

いつも大変お世話になっております。
「青森県デジタル人材育成　分科会」の出欠確認フォームです。
必要事項をご記入の上、**【回答期限：9月8日（火）12時まで】**にご回答をお願いいたします。

◆研修会概要　（詳細は開催通知をご確認ください）
開催日（全4回）：2026年9月15日（火）
（以下仮日程）9/29 (火) 、11/10 (火)、11/24 (火)、11/10 (火)
開催時間（全4回とも）：13:00~15:30（途中休憩あり）
開催場所：青森県庁会議室

==========
◆お問い合わせ先
青森県財務部市町村課　総務・行政グループ
人材育成研修会担当：安藤　017-734-9071 ／ somu_gyosei@pref.aomori.lg.jp
【業務委託先】 株式会社コンシス　担当：福田　dx-support@consis.jp

◆青森県市町村DX相談窓口Webサイト
https://dx.consis.jp/

dx-support@consis.jp アカウントを切り替える

* 必須の質問です

メールアドレス *

メールアドレス

① 自治体名 *

選択

【おさらい】フォームの2大活用シーン

【庁内】手作業転記の「ゼロ化」

- 各課からの予算要望のとりまとめ
- 庁内研修・説明会の受講申込み
- 各種定期報告・アンケートの集約

【庁外】市民の声を集める

- イベント参加者の満足度・属性調査
- 公園施設や道路の破損通報フォーム
- 地域の課題抽出・パブリックコメント

【おさらい】Googleフォームの使い方

庁内向け作例





【おさらい】Googleフォームの使い方

庁外向け作例

窓口満足度アンケート

dx-support@consis.jp アカウントを切り替える

共有なし

* 必須の質問です

ご年代をお教えください

- 20代以下
- 30代
- 40代
- 50代
- 60代
- 70代以上

本日、どちらの窓口にお越しになりましたか？ *

- 町民課
- 税務会計課
- 福祉課
- その他:

待ち時間について、該当するものを1つずつお選びください。 *

	1	2	3	4	5	
不満	○	○	○	○	○	満足

説明の分かりやすさ について、該当するものを1つずつお選びください。 *

	1	2	3	4	5	
不満	○	○	○	○	○	満足

補助金申請について

農業関係の補助金申請フォームです

dx-support@consis.jp アカウントを切り替える

共有なし

* 必須の質問です

補助金の種類を選択してください *

- 農業機械や施設について
- 農作物関係について
- 関連団体について
- その他:

氏名(申請される方) *

回答を入力

連絡先(電話番号ハイフンなし) *

回答を入力

ご連絡可能な時間帯 *

- 午前中
- 13時～15時
- 15時～17時
- その他:

送信　フォームをクリア

【おさらい】グループワーク＆ハンズオン

作成されたフォーム

業務効率化（77%）：通勤届、検査依頼、PC権限申請、時間外申請など

市民の声集約（23%）：敬老会アンケート、窓口満足度など

ご意見・ご感想

「DXには関わりがないと思っていたが、思いのほか簡単に活用できると感じた」

「スプレッドシートが思っていたよりも色々なことができそう」

「初めて使ったが使いやすかった。業務の効率化につながると思う」

【おさらい】グループワーク＆ハンズオン

主なハードル：

• 時間・リソース:
日常業務が忙しい、テスト運用に不安(6名)

• スキル不安: Googleフォームの細かい操作方法(6名)
→できる範囲でOKです。
「知る・ふれる・試してみる」をできたのが重要！

• セキュリティ: 個人情報や部外秘データの取扱(5名)
→ゼロトラスト（許可した人だけがアクセスできる）を徹底！

懸念と課題：

• 「セキュリティ的（LGWAN環境など）に庁内で使うことが難しそう。」

• 「庁内の規則等の問題で、実際にデータ収集に使用することはできない。」

→LoGoフォームなど使用可能なツールで試してみましょう！

アイスブレイク

ポータルシート【グループワーク用フォーム】を開いてください

【お題】きょうのお昼ごはん、何食べましたか？

（※昼食をスキップした方は「朝ごはん」をご記入ください）
入力して送信 ➔ 「別の回答を送信」で何回でも連打投稿OKです！

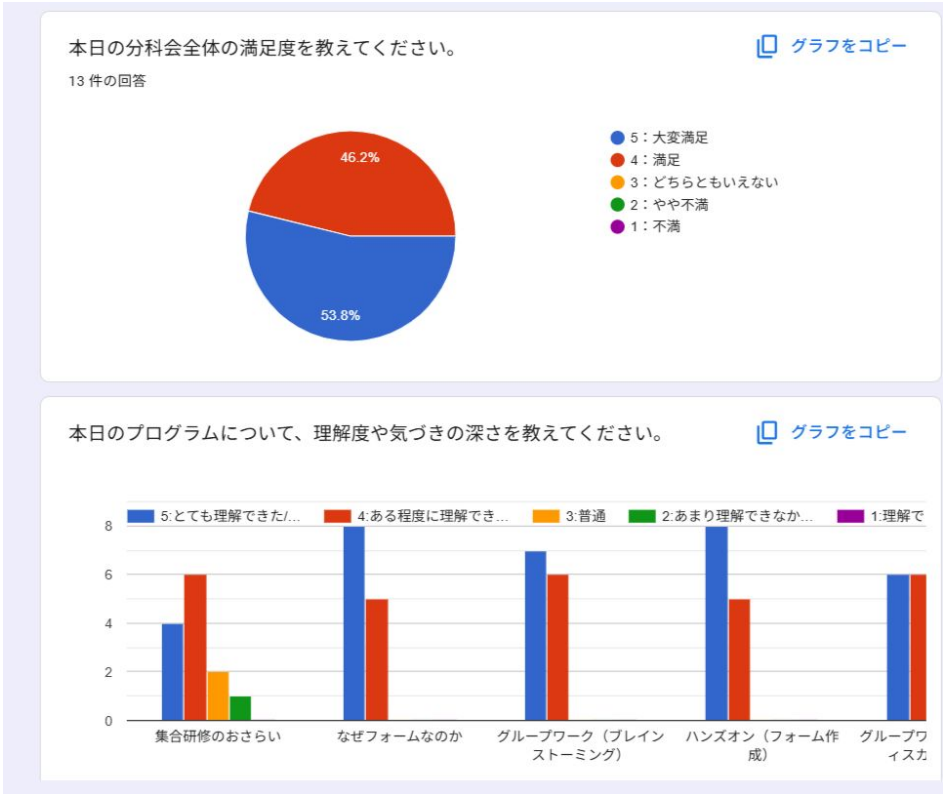
【集める】→【整える】

BI（Business Intelligence）ツール

収集したデータを可視化・視覚化

例）googleフォームの回答要約機能

※基本的な集計データを確認するには必要十分



BI（Business Intelligence）ツール

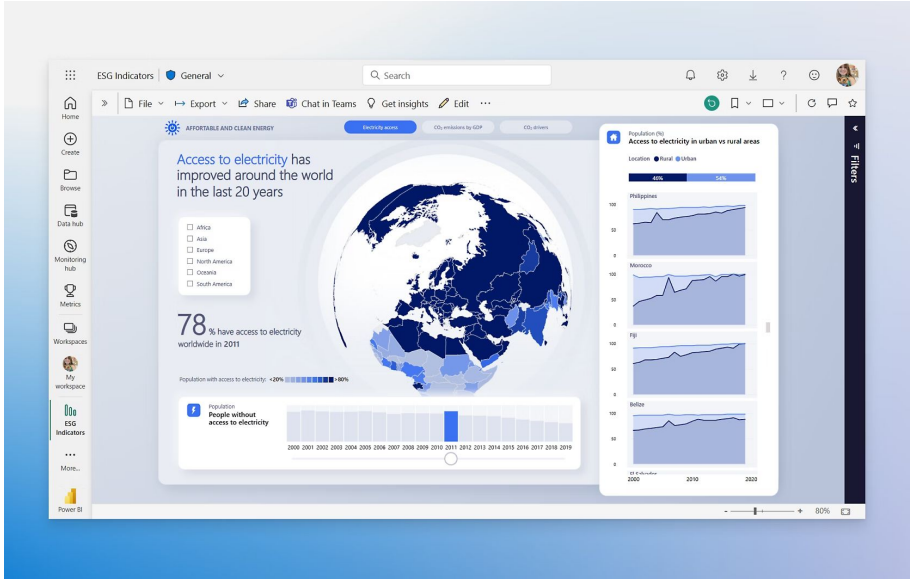
収集したデータを可視化・視覚化

クロス分析や複数データの掛け合わせなど、
精緻な分析が可能

主なBIツール

- Microsoft Power BI
- Googleデータポータル（旧Looker Studio）

など



今回使うBIツール

- （比較的）簡単に作成可能
- 無料で利用できる
- 共有・連携機能

cf) MS Power BI

- 関数の知識がやや必要
- 無料で利用可、共有・連携は有償版のみ



Googleデータポータルの実演

架空イベントのアンケート

https://docs.google.com/forms/d/e/1FAIpQLSfGYx8Smr77ukn0kArPvOaJkFeabJD2r26XWSpeqPulpGaETA/viewform

管理画面

https://docs.google.com/forms/d/14uyHoYTUF-ggxpUZT9LxM9lGdGgUP27yx63jlGGQPyY/edit



Googleデータポータルの実演

Googleフォーム標準集計画面

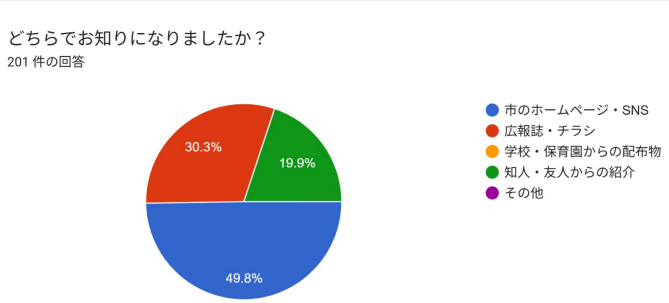
メリット：

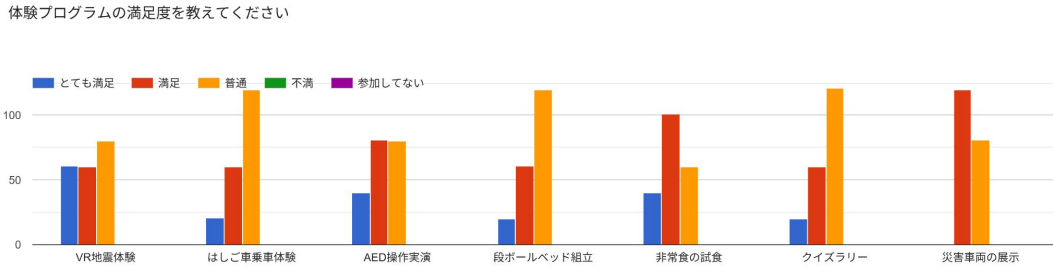
設定不要で自動的にグラフ化される

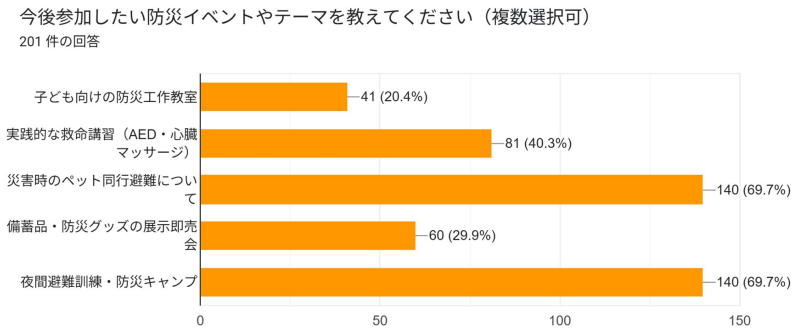
デメリット：

「年代別×満足度」などの**クロス分析**ができない

フィルターによる絞り込みが不可



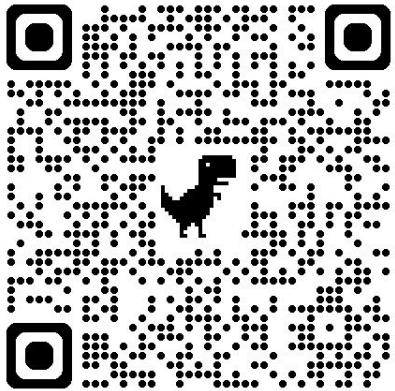




Googleデータポータルの実演

データポータルを開く

https://datastudio.google.com/reporting/d59f0b77-edd8-4551-937a-806281c0e7f9





◎◎市ファミリー防災フェスタ2026参加者アンケート

本日のイベントにお越しいただきありがとうございます。こちらは今回のイベントに関する満足度調査です。今後の参考にさせていただきますのでご協力をお願いいたします。
主催：◎◎市　主担当：◎◎市地域防災課、◎◎地区消防事務組合

fukuda@consis.jp アカウントを切り替える

共有なし

* 必須の質問です

回答者様について

年代 *

- 10代以下
- 20代
- 30代
- 40代
- 50代
- 60代
- 70代以上

性別 *

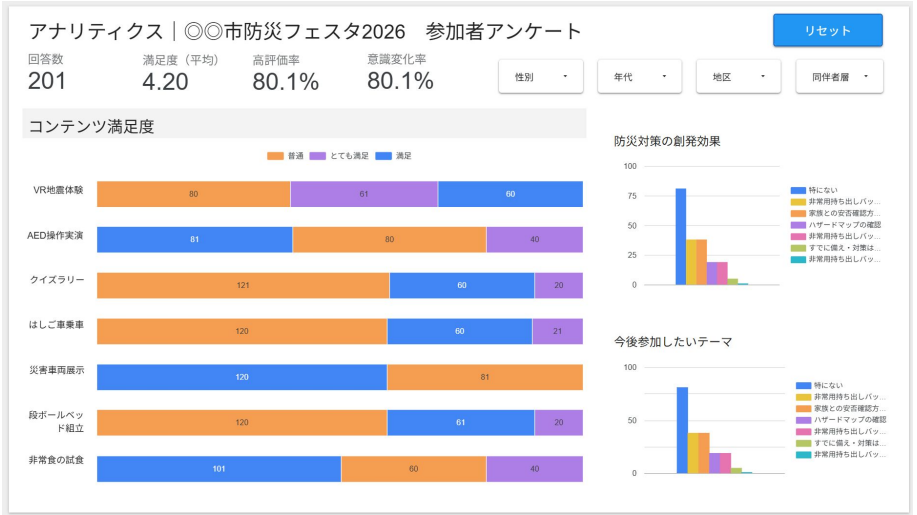
- 男性
- 女性
- 回答しない

Googleデータポータルの実演

BIの強み

- フィルター機能
- リアルタイム計測

高度な分析が可能に

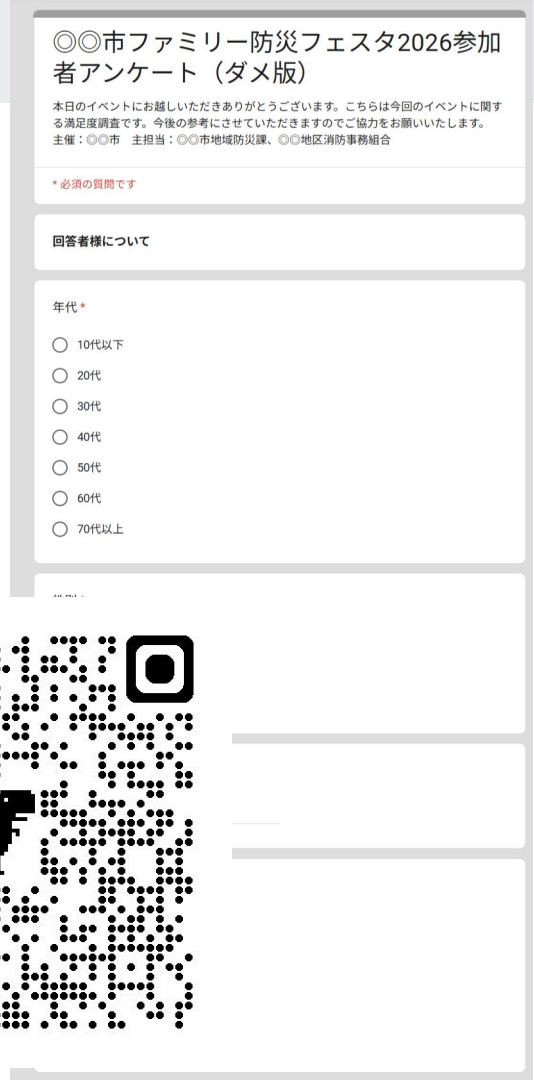


もうひとつのアンケート

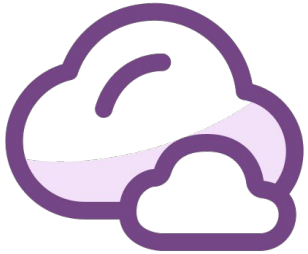
https://docs.google.com/forms/d/e/1FAIpQLSfruUNLMi_kirWDOjG0WXyrGWAR9nhjeQKGL9e0u18alqIujg/viewform?usp=header

先のフォームと、どこが違う？

実際に回答して、見つけてみてください。



Do not edit
How to change the design



このフォーム、どこがダメですか？

Presenting with animations, GIFs or speaker notes? Enable our Chrome extension

slido

このフォーム、どこがダメですか？

- 「お住まい」が**自由記述**
- 「体験プログラムの満足度」が**複数選択可＆無回答可**
- 「始めたい対策」が**自由記述**
- 「参加したいテーマ」が**自由記述**

→データの正確性・完全性が確保できない！



◎◎市ファミリー防災フェスタ2026参加者アンケート（ダメ版）

本日のイベントにお越しいただきありがとうございます。こちらは今回のイベントに関する満足度調査です。今後の参考にさせていただきますのでご協力をお願いいたします。
主催：◎◎市　主担当：◎◎市地域防災課、◎◎地区消防事務組合

* 必須の質問です

回答者様について

年代 *
- 10代以下
- 20代
- 30代
- 40代
- 50代
- 60代
- 70代以上

性別 *
- 男性
- 女性
- 回答しない

お住まい *

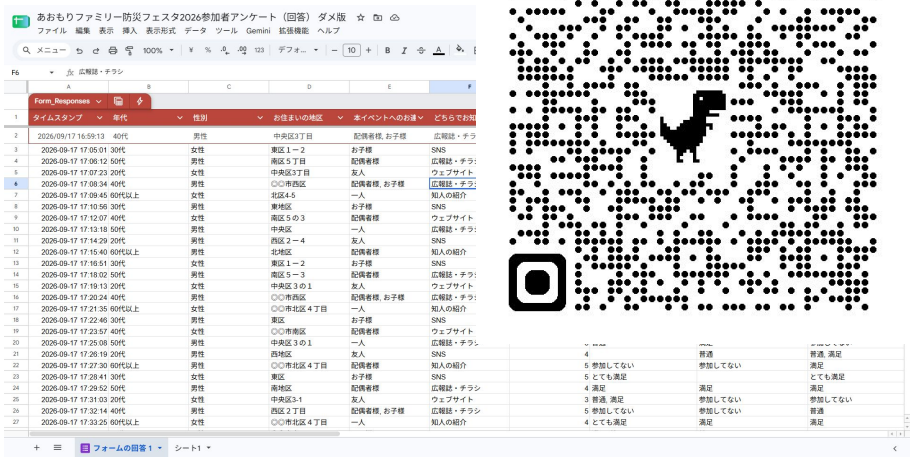
回答を入力

本イベントへのお連れ様 *
- いない（回答者様のみ参加）
- 配偶者様
- お子様
- 親御様
- お孫様

見比べる｜収集されたデータ

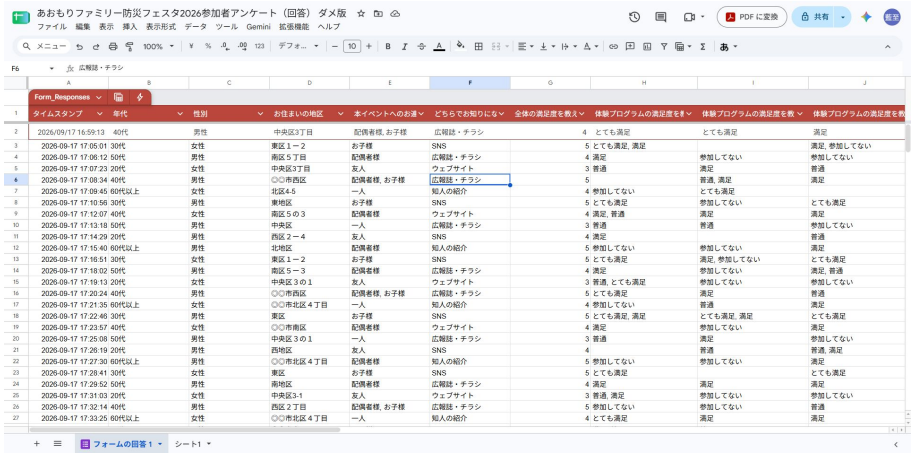


https://docs.google.com/spreadsheets/d/1tGDK6xJbkSJ18CmrY0ZMEXxvec3O5iMnUBJYssXXWjE/edit?usp=sharing



https://docs.google.com/spreadsheets/d/1cVG_OnAU6emp2gCEx645HyrPVopbCeOonC8B7ktwzO4/edit?usp=sharing

見比べる｜収集されたデータ



お住まい：自由記述
→**表記ゆれ**

PG満足度：複数選択可＆無回答可
→**データなし＆無効回答が混入**

始めたい対策：自由記述
参加したいテーマ：自由記述
→**回答が個別化して集約不可**

＝汚れたデータ（分析の信頼性が低い）

汚れたデータを BIに通した場合

データポータルを開く

https://datastudio.google.com/reporting/bd227631-fe78-4fa0-b2be-aae811b37df7



Garbage In, Garbage Out

- ゴミを入れたらゴミが出てくる
- 意思決定の精度は「データの品質」で決まる

→ データクレンジングが必要！

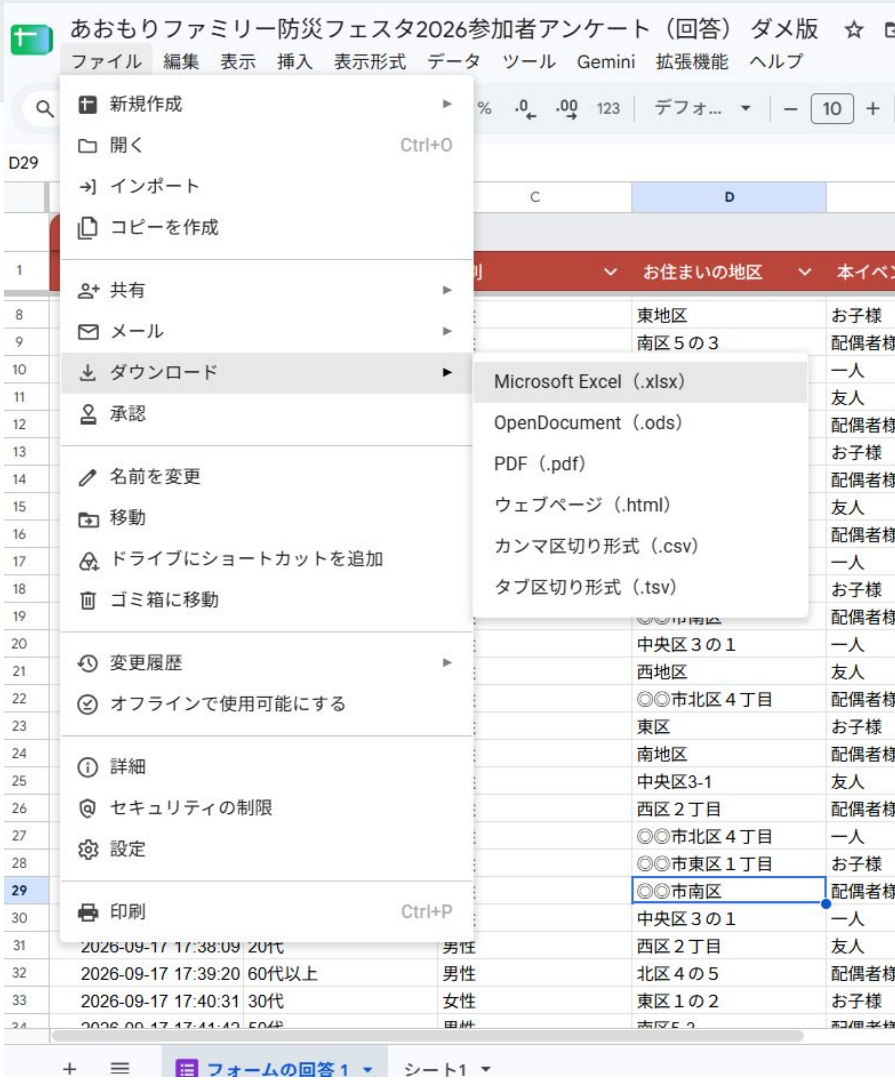
データ品質｜6つの基本

① 正確性	データが事実に則しているか	❌ 顧客の引越し前の古い住所がそのまま残っている	⭕ 実際の住所とシステム上の住所が完全に一致している
② 完全性	必要なデータが欠けずに揃っているか	❌ 必須入力であるはずの「メールアドレス」が空欄（NULL）になっている	⭕ 分析に必要な項目がすべて埋まっている
③ 一貫性（整合性）	複数のシステム間でも矛盾がないか	❌ 社内システムAとシステムBで、同じ顧客の電話番号が違っている	⭕ どのシステムを見ても同じデータが同期されている
④ 適時性（最新性）	必要な最新データが利用できるか	❌ 昨日の売上データを見たいのに、反映されるのが3日後になる	⭕ 意思決定をしたいタイミングで、常に最新の鮮度のデータがある
⑤ 有効性（妥当性）	決めたルールや形式に合格しているか	❌ 年齢の欄に「-5歳」や「1000歳」といった値が入っている	⭕ 定められた桁数、日付フォーマット、選択肢（マスタ）の範囲内である
⑥ 一意性	データに重複がないか	❌ 同じ顧客が「会員ID：001」と「002」で二重登録されている	⭕ 1つの実体（人やモノ）に対して、データも1件だけ存在する

データクレンジングの実践

汚れたデータをローカル環境で洗浄（整理）してみましょう

.xlsx形式でDLしてMicrosoft Excelで開きます



データクレンジングの実践

手法1：置換（Ctrl＋H） 誤った表記を正しい表記に置き換える

検索する文字列	**置換後の文字列**	
◎◎市	（空白）	「◎◎市」を除外
中央	中央地区	表記の統一
中央区	中央地区	**※注意：** 中央地区 が 中央地区地区 にならないよう置換順序に注意
せいぶ	西部地区	ひらがな表記の修正
東区	東部地区	表記の統一
（全角スペース）	*(空欄)*	末尾や間に挟まった見えないゴミスペースの削除

データクレンジングの実践

手法1：置換（Ctrl＋H） 誤った表記を正しい表記に置き換える

検索する文字列	置換後の文字列
非常用持ち出し袋	非常用持ち出しバッグの準備
非常用バッグと家具固定	非常用持ち出しバッグの準備 家具の転倒防止対策
家具固定	家具の転倒防止対策
水と食料	水・食料の備蓄（3日分以上）
安否確認方法を決める	家族との安否確認方法の取り決め
特にない	*（空欄または「なし」）*

データクレンジングの実践

手法1：置換（Ctrl＋H） 誤った表記を正しい表記に置き換える

ちょっと無理 …

- 作業量が多い
- 完全な自由記述には置換ルールを 1件ずつ考えて設定
- 単純な置き換えだけでは正しきれないデータが多い

→ 大量かつ複雑なデータに対しては現実的ではない手法

データクレンジングの実践

手法2：Power Queryを使って整形

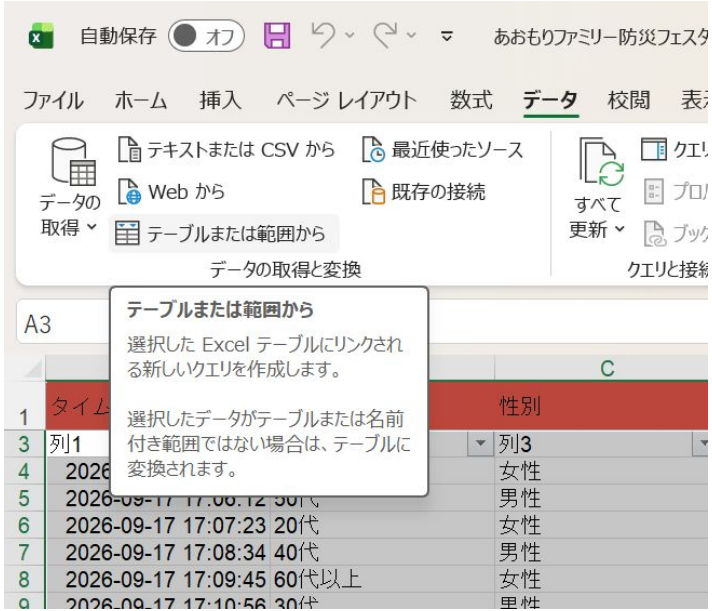
Power Query：

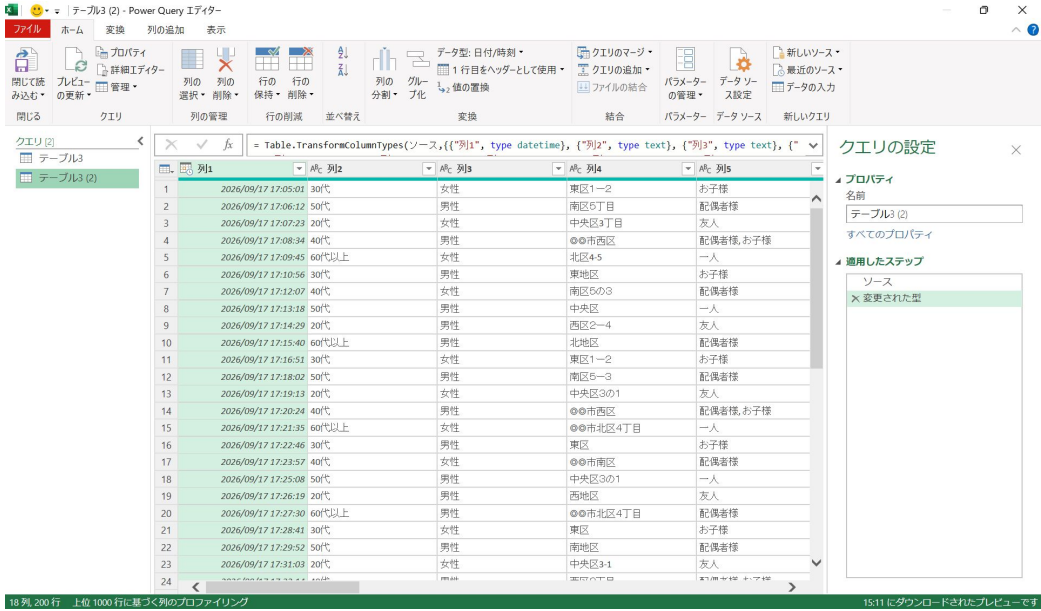
Microsoft Excel / Power BIに組み込まれているデータ処理ツール。
接続した外部データの変換・組み合わせ・読み込みができる

特徴：

- 関数の知識がなくてもある程度の処理が可能（ノーコード・ローコード）
- 元データに変更を加えず処理ができる
- 大量・複数のデータを統合処理可能、元データの変更点も反映

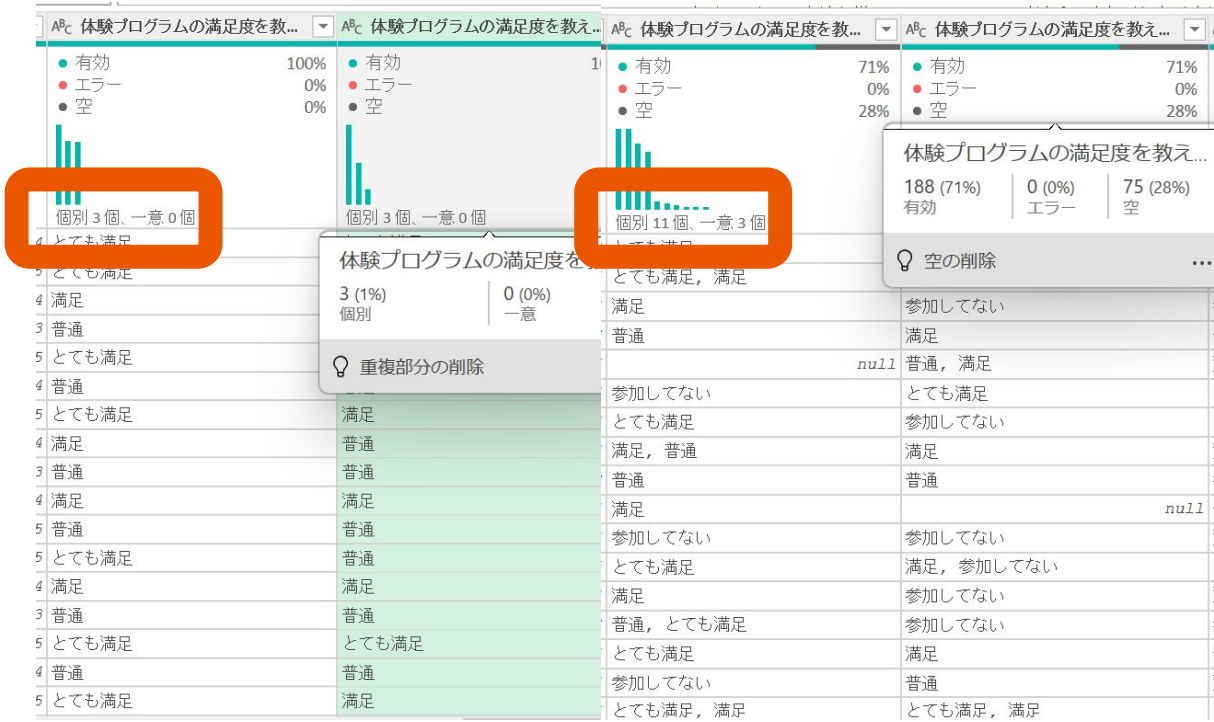
データクレンジングの実践｜ Power Query





データクレンジングの実践｜Power Query

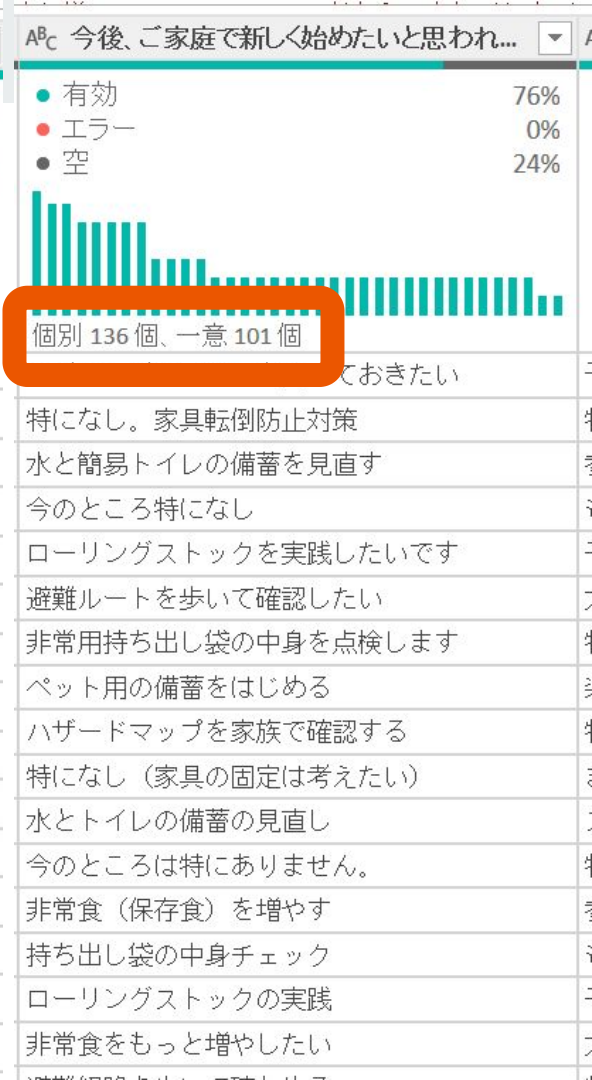
データの汚れ具合がわかる


体験プログラムの満足度を教...
有効 100%
エラー 0%
空 0%
個別 3 個、一意 0 個
体験プログラムの満足度を教え...
有効
エラー
空
個別 3 個、一意 0 個
3 (1%)
個別
0 (0%)
一意
重複部分の削除
体験プログラムの満足度を教...
有効 71%
エラー 0%
空 28%
個別 11 個、一意 3 個
体験プログラムの満足度を教え...
有効 71%
エラー 0%
空 28%
体験プログラムの満足度を教え...
188 (71%)
有効
0 (0%)
エラー
75 (28%)
空
空の削除
とても満足
満足
普通
とても満足
満足
普通
満足
普通
満足
普通
とても満足
満足
普通
とても満足
満足
普通
とても満足

データクレンジングの実践｜Power Query

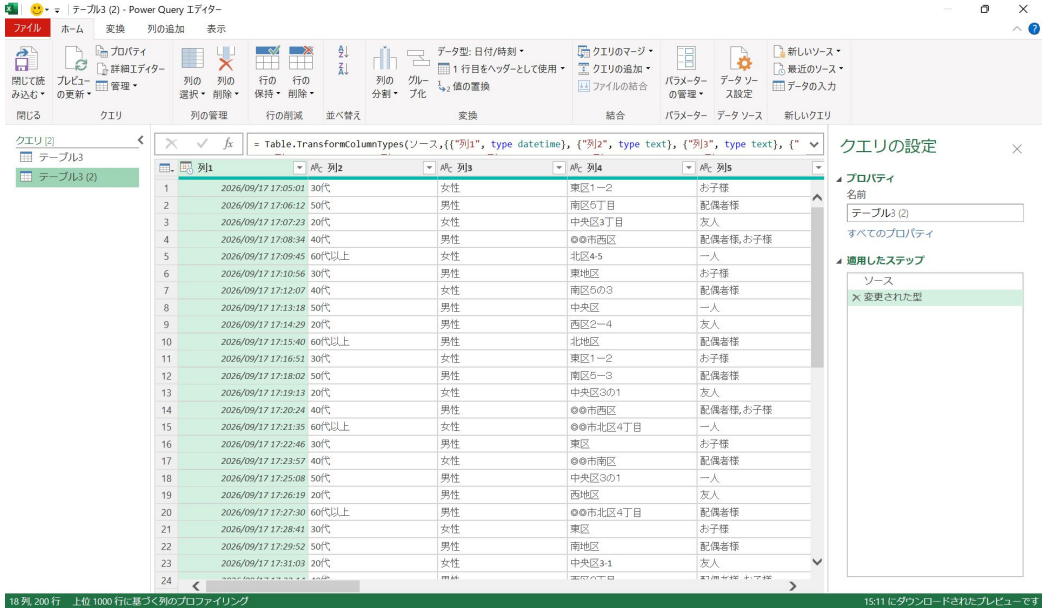
データの汚れ具合がわかる





データクレンジングの実践｜Power Query

実演・条件列

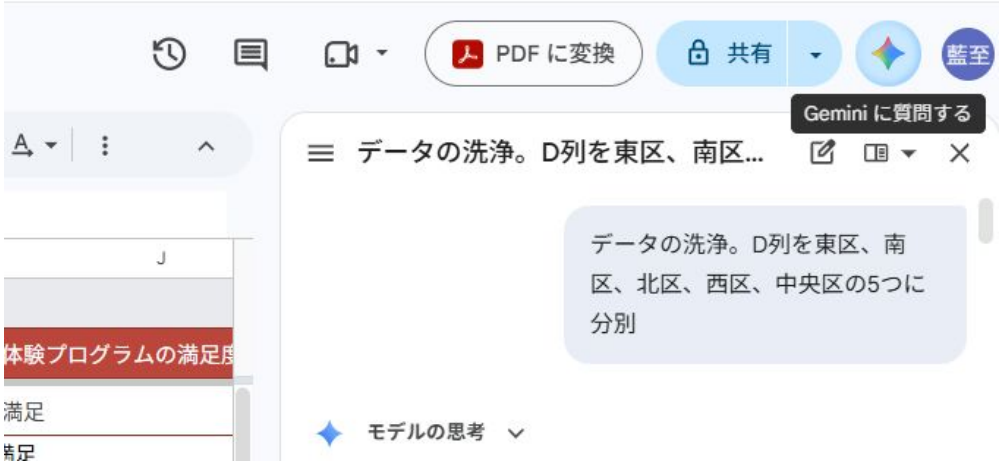


データクレンジングの実践

手法3：AIにやってもらう

GoogleスプレッドシートでGeminiによる処理

自然言語による指示（ノーコード）

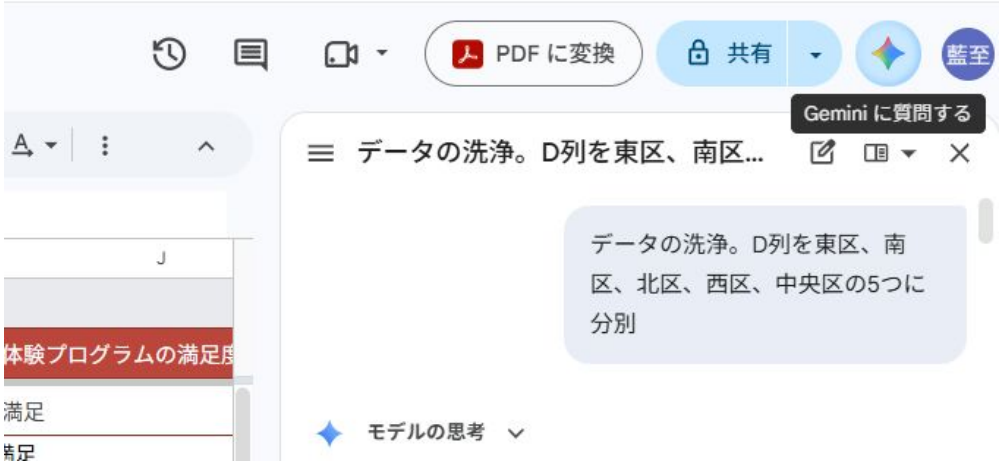


データクレンジングの実践

手法3：AIにやってもらう

GoogleスプレッドシートでGeminiによる処理

自然言語による指示（ノーコード）



データクレンジングの実践

手法3：AIにやってもらう

Gemini in Google

「H～N列をとても満足、満足、普通、不満、参加してないの5つに分別。nullはすべて「普通」に置換。複数回答の場合は評価が最も高いものに置換。」



データクレンジングの実践

手法3：AIにやってもらう

Gemini in Google

「Q列を回答の内容でいくつかに分類した上でマージして」



データクレンジングの実践

手法3：AIにやってもらう

Gemini in Google

「Q列を回答の内容でいくつかに分類した上でマージして」

作業完了のご報告

「フォームの回答 1」シートのQ列（`Q2:Q202`）に記録されていた備えや対策に関する回答データについて、表記揺れの名寄せとカテゴリ集約を実施しました。実施内容

1. **表記揺れの名寄せとカテゴリ集約**
   - 多様な表現（「買っておきたい」「購入」「モバブ」や「見直す」「増やす」「点検」など）を標準化された以下の主要9分類、および「特になし」に統一しました。
     - **モバイルバッテリーの準備**
     - **家具の転倒防止対策**
     - **水・簡易トイレの備蓄見直し**
     - **ローリングストックの実践**
     - **避難経路の確認**
     - **非常用持ち出し袋の点検**
     - **ペット用の備蓄**
     - **ハザードマップの確認**
     - **非常食の備蓄**
     - **特になし**
2. **複数回答・補足記述の集約**
   - 「特になし。家具転倒防止対策」などの複数該当記述については、具体的な対策内容（家具の転倒防止対策）を主カテゴリとして集約しました。
3. **データ検証**
   - 全201件（`Q2:Q202`）が正常に更新され、想定通りのカテゴリ名に統一されていることを確認しました。

Gemini に質問する

2

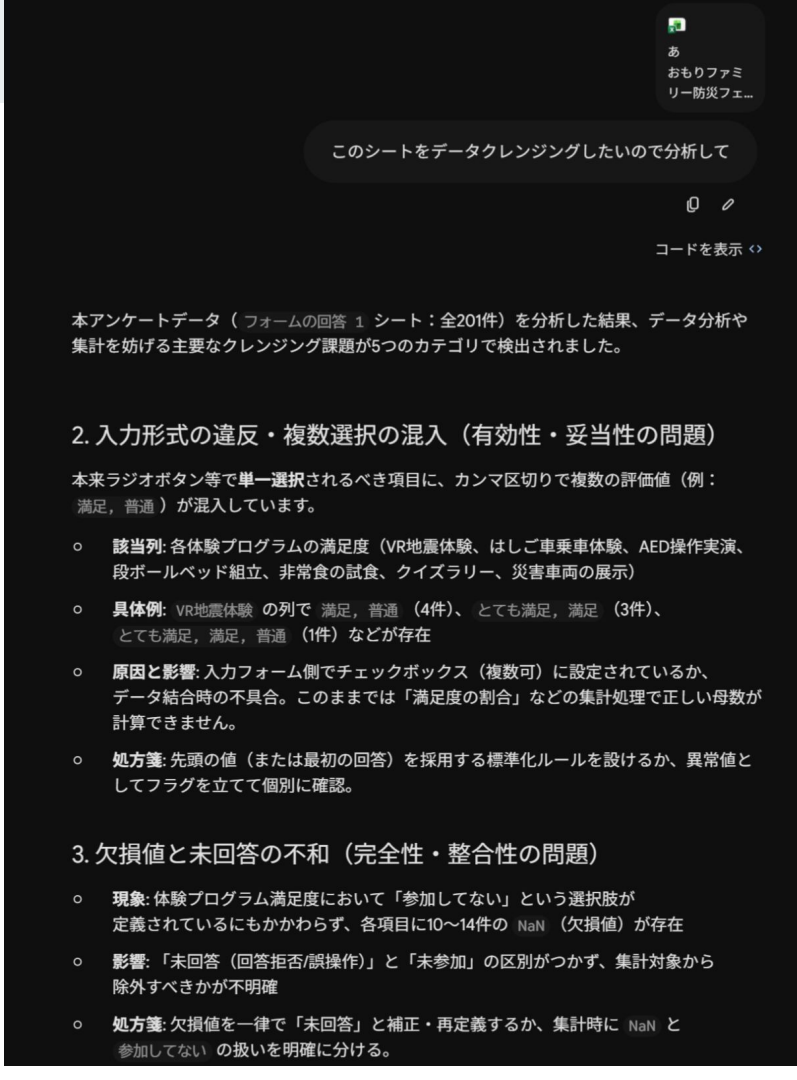
Gemini in Workspace は、不正確な情報を表示することがあります。 詳細

データクレンジングの実践

手法3：AIにやってもらう

元データをアップして丸ごと解析

「このシートをクレンジングしたいので分析して」



データクレンジングの実践

手法3：AIにやってもらう

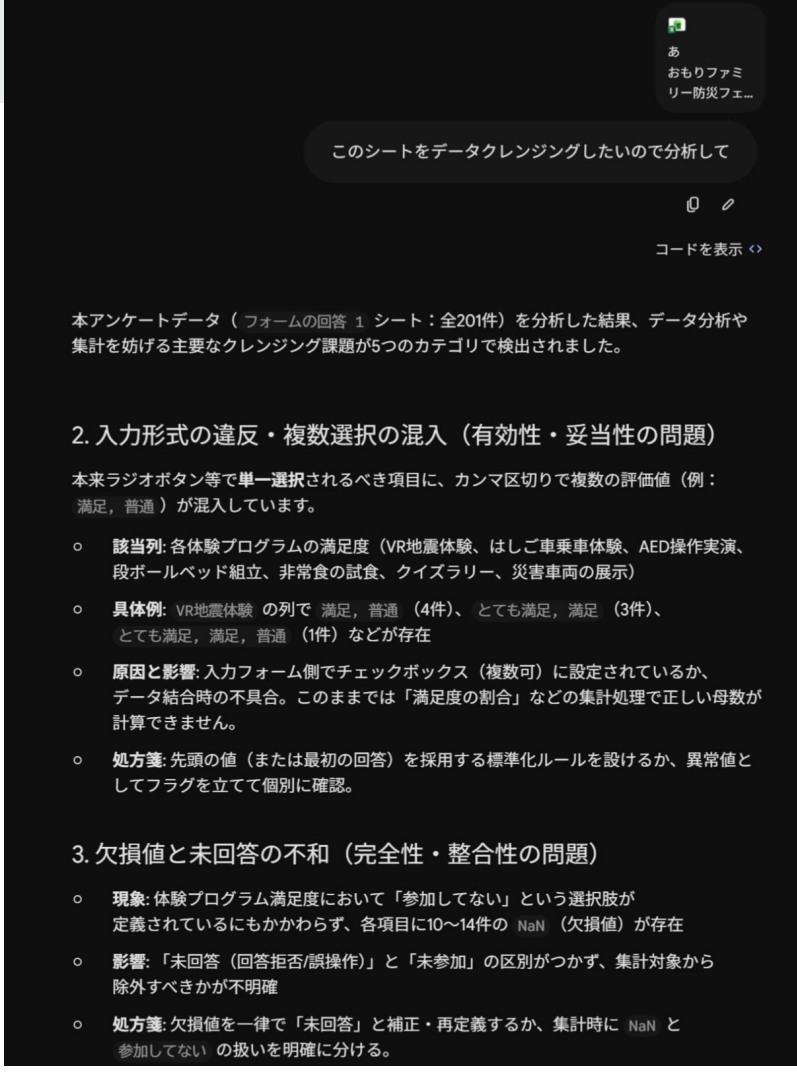
「このシートをクレンジングしたいので分析して」

入力形式の違反・複数選択の混入（有効性・妥当性の問題）

欠損値と未回答の不和（完全性・整合性の問題）

表記揺れ・フォーマットの不統一（一貫性の問題）

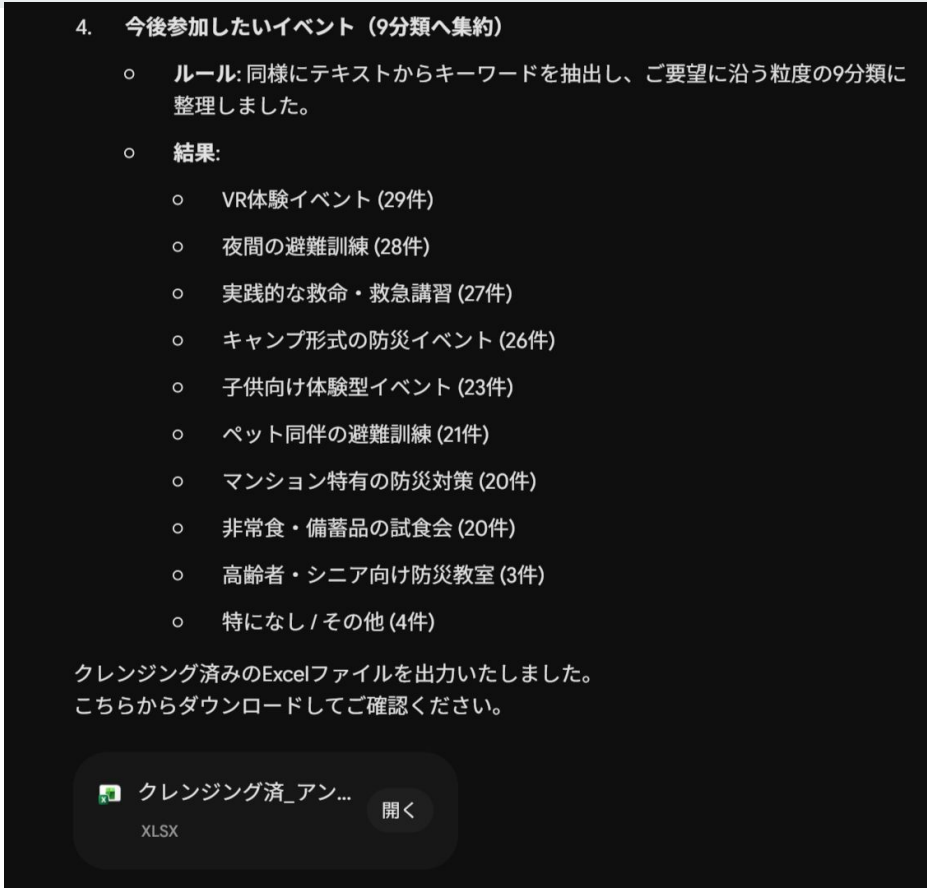
タイムスタンプの構造的異常（データの人工性）



データクレンジングの実践

手法3：AIにやってもらう

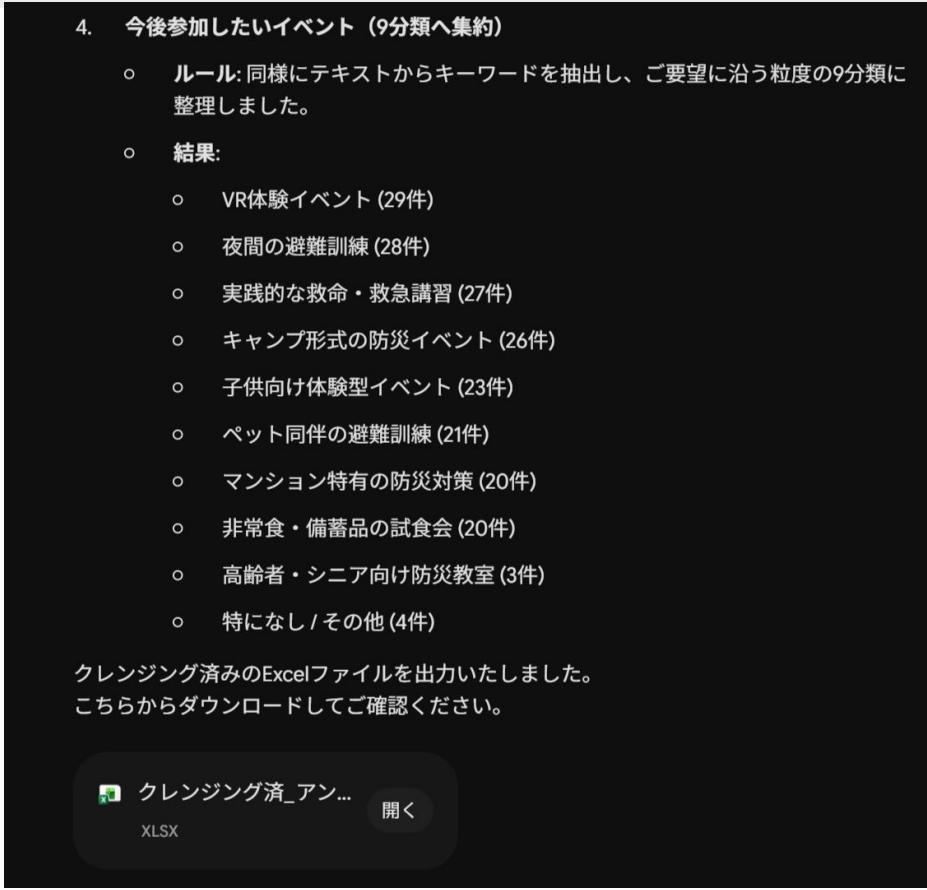
「お住まいの地区」は5分類、プログラムの満足度は5分類、複数回答は評価の良いものに置換、nullは「普通」に置換。今後始めたい備え / 今後参加したいイベントは妥当な6〜9分類にまとめて



データクレンジングの実践

手法3：AIにやってもらう

「お住まいの地区」は5分類、プログラムの満足度は5分類、複数回答は評価の良いものに置換、nullは「普通」に置換。今後始めたい備え / 今後参加したいイベントは妥当な6〜9分類にまとめて



データクレンジング｜ AI活用の注意点

データの秘匿性とセキュリティ：個人/機密情報の入力による情報漏洩や規約違反

→ 事前スクリーニング

ハルシネーション：
存在しないデータへの書き換え、残すべきデータを消去・改ざん

データの保持と品質の劣化：データが壊れるリスク、整合性・正確性の棄損

→ 人によるチェック、バックアップ

ヒューマン・イン・ザ・ループ（HITL）

- AI処理プロセスの中に人間が介在する
- 「評価」「修正」「最終承認」は人の役目

データ品質｜6つの基本

① 正確性	データが事実に則しているか	❌ 顧客の引越し前の古い住所がそのまま残っている	⭕ 実際の住所とシステム上の住所が完全に一致している
② 完全性	必要なデータが欠けずに揃っているか	❌ 必須入力であるはずの「メールアドレス」が空欄（NULL）になっている	⭕ 分析に必要な項目がすべて埋まっている
③ 一貫性（整合性）	複数のシステム間でも矛盾がないか	❌ 社内システムAとシステムBで、同じ顧客の電話番号が違っている	⭕ どのシステムを見ても同じデータが同期されている
④ 適時性（最新性）	必要な最新データが利用できるか	❌ 昨日の売上データを見たいのに、反映されるのが3日後になる	⭕ 意思決定をしたいタイミングで、常に最新の鮮度のデータがある
⑤ 有効性（妥当性）	決めたルールや形式に合格しているか	❌ 年齢の欄に「-5歳」や「1000歳」といった値が入っている	⭕ 定められた桁数、日付フォーマット、選択肢（マスタ）の範囲内である
⑥ 一意性	データに重複がないか	❌ 同じ顧客が「会員ID：001」と「002」で二重登録されている	⭕ 1つの実体（人やモノ）に対して、データも1件だけ存在する

データクレンジングでも限界がある

Do not edit
How to change the design



データクレンジングで整形できる要素はどれ？

Presenting with animations, GIFs or speaker notes? Enable our Chrome extension

slido

Do not edit
How to change the design



データクレンジングで整形しきれない要素はどれ？

Presenting with animations, GIFs or speaker notes? Enable our Chrome extension

slido

構造化データ

6つの品質指標（正確性・完全性・一貫性・適時性・有効性・一意性）を満たし、
列と行で整理され、BIツールやAIでそのまま分析できるデータ

↔　非構造化データ
例）メールやWord・PDFなどの文書、写真、画像、動画、音声ファイルなど

世の中に存在するデータのうち、
構造化されているデータはわずか10～20%

「あとから構造化」は地獄

- 膨大な手間・時間・人的コスト
- 完璧な構造化には限界がある

構造化は「収集」の段階で

最良のデータクレンジングは、
最初からクレンジング不要な形（構造化済み）で集めること

→ **フォームの設計が重要**
（ラジオボタン化・回答必須設定など。記述形式は極力除外）

データの「粒度」を意識する

粒度が小さい（詳細）：生年月日、詳細な住所、自由記述コメント

粒度が大きい（粗い）：年代（30代など）、地区（中央地区など）、選択肢

収集する側：**「できるだけ詳細なデータがほしい」**

回答する側：記述式や細かい入力は負担（離脱リスク） / 汚データの発生リスク

基本はすべて選択式にして粒度を粗く抑える。
記述式はあらかじめ選択肢化できない内容、重要で詳細な情報が必要な設問に限定。

皆様の作例から


部署名 *
選択
端末管理番号（TOOOOO） *
回答を入力
この質問は必須です
権限付与希望日 *
日付
2026/10/02
時刻
10 : 00
権限付与終了日 *
日付
2026/10/03
時刻
1 : 0
時間の値が無効です

皆様の作例から


申請金額
回答を入力
支払目的・内容
回答を入力
金融機関名
回答を入力
支店名
回答を入力
口座番号
回答を入力
送信
フォームをクリア

皆様の作例から

2．ランニング歴

- ○ 1年未満
- ○ 1〜3年
- ○ 3〜5年
- ○ 5年以上

3．月間の平均走行距離（km）

回答を入力

4．普段のランニングペース（スロージョグ時）

- ○ キロ7分台より遅い
- ○ キロ6分台
- ○ キロ5分台
- ○ キロ4分台より速い

4．普段のランニングペース（スピード練習時）

- ○ キロ7分台より遅い
- ○ キロ6分台
- ○ キロ5分台
- ○ キロ4分台より速い


ランニング同好会アンケート（仮）

dx-support@consis.jp アカウントを切り替える

共有なし

目標・ニーズ

1．現在の目標・目的

- ☐ 健康維持
- ☐ 5km走完走
- ☐ 10km走完走
- ☐ ハーフマラソン完走
- ☐ フルマラソン完走
- ☐ サブ4などのタイム更新（サブ4とはフルマラソン4時間以下での完走という意味です）
- ☐ 他種目への挑戦（トレイルラン・トライアスロンなど）

2．同好会で行いたい練習の形式

- ☐ ジョグ（おしゃべりラン程度）
- ☐ LSD（ロング・スロー・ディスタンス 10km以上レベル）
- ☐ スピード練習（インターバル走など）
- ☐ タイムトライアル

3．同好会練習で参加可能な時間

- ☐ 30分以下
- ☐ 30分〜1時間
- ☐ 1時間〜2時間
- ☐ 2時間以上

4．走りたいエリア・ロケーション

- ☐ 町外
- ☐ 青森市などの町外
- ☐ 大きめの公園
- ☐ 街中
- ☐ 郊外

5．同好会で参加したい大会の形式

- ☐ 5km走
- ☐ 10km走
- ☐ ハーフマラソン
- ☐ フルマラソン
- ☐ リレーマラソン
- ☐ 県外の大会（遠征）

6．大会参加費の上限目安（1大会あたり出せる最大金額）

- ○ 3,000円以下
- ○ 5,000円以下
- ○ 10,000円以下
- ○ 内容次第でいくらでも

7．【！重要！】飲み会の頻度（年間）

- ○ 総会のみ参加したい
- ○ 総会＋1〜2回程度
- ○ 総会＋大会前に決起集会（3〜4回）
- ○ 総会＋お誘いがあったら是非

戻る　次へ　フォームをクリア

本日のゴール｜

データ整形の重要度を正しく認識する

おつかれさまでした。

アンケートにご協力お願いします。