

■app_Wide2Long (テストトップアプリ)

ワイドフォーマットで整理されたデータ表を、ロングフォーマットに変換する自作アプリ。(Interactive Visualizer で扱いやすいデータ構造にするための前処理を行う)

※Google Drive からダウンロードしようとする、「ウイルスに感染しています」「ウイルススキャンが出来ません」という警告文が出るが、これは.exe ファイル形式が Google Drive にサポートされていないために表示されるもの。

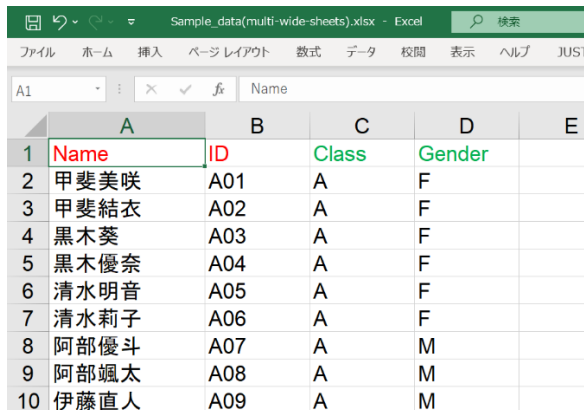
■Interactive_Visualizer (Colab ノートブック)

Python の Altair ライブラリによるインタラクティブなグラフ描画を、ボタン選択とスライダー調整のみ(コード修正不要)で操作可能にした.ipynb ファイル。要は、手軽に本格的な探索的データ分析(EDA)を実行できるようコードをまとめた Python 実行環境。特に Line Graph 系では『個人』『集団』『科目』それぞれに特化した時系列分析モードと層別分析モードを実装しており、それぞれの視点に対して適切なインサイトが得られやすくなるよう設計している。

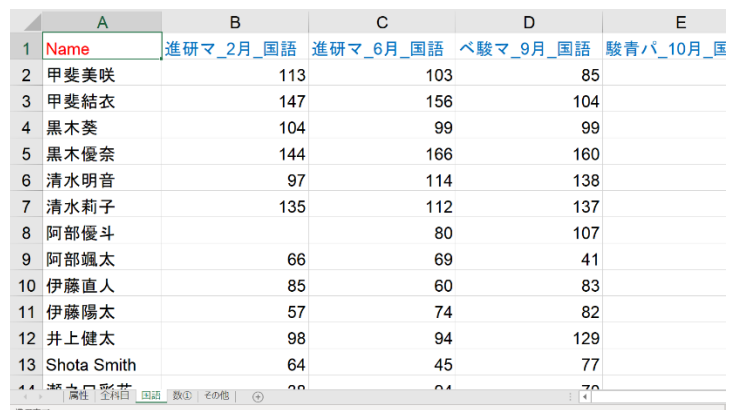
データ読み込みはローカル環境から直接ドロップインする形式にすることで、Google Drive と連携する手間すらなく、すぐに使えることも魅力の一つ。

ステップ1： ワイドフォーマットデータ整理

以下の要領でデータを蓄積、整理する。



	A	B	C	D	E
1	Name	ID	Class	Gender	
2	甲斐美咲	A01	A	F	
3	甲斐結衣	A02	A	F	
4	黒木葵	A03	A	F	
5	黒木優奈	A04	A	F	
6	清水明音	A05	A	F	
7	清水莉子	A06	A	F	
8	阿部優斗	A07	A	M	
9	阿部颯太	A08	A	M	
10	伊藤直人	A09	A	M	



	A	B	C	D	E
1	Name	進研マ_2月_国語	進研マ_6月_国語	ベ駿マ_9月_国語	駿青バ_10月_国
2	甲斐美咲	113	103	85	
3	甲斐結衣	147	156	104	
4	黒木葵	104	99	99	
5	黒木優奈	144	166	160	
6	清水明音	97	114	138	
7	清水莉子	135	112	137	
8	阿部優斗		80	107	
9	阿部颯太	66	69	41	
10	伊藤直人	85	60	83	
11	伊藤陽太	57	74	82	
12	井上健太	98	94	129	
13	Shota Smith	64	45	77	

①ファイル名は空白（スペース）を含まない文字列。拡張子はxlsx / csv に対応。

例；○「data_file.xlsx」 × 「data file.xlsx」

xlsx の場合、複数シートにまとめても構わない（複数科目を整理するにはその方が向いている）。

②以下は**必須**の列

「**Name**」 列名固定。テスト受験者の名前。

「**ID**」 列名固定。重複のない英数字。順番がわかりやすい規則にしておくのがよい。

「**試験スコア**」 列名は下記のフォーマットに準ずること。

- ・'試験名' '時期' '科目名' の3要素をアンダースコア '_' で繋いで記載。

例；「進研マーク_6月_全科目」、「進研・記_R71112_数学」、「小テスト_0423_英単語」

- ・半角シャープ '#' は、コードエラーに繋がる可能性があるため、使用を避けた方がよい。
- ・入力データは、テストのスコア。満点が異なるスコアを入力して構わない。
- ・欠損値である場合は、セルを空白にしておく。

※複数シートにデータをまとめる場合、受験者とスコアデータを対応づけるため、「Name」か「ID」いずれか一方は各シートに必ず含めておくこと。

③以下は**任意**の列（あった方が分析の幅は広がる）

「**ラベルデータ**」

- ・列名は、必須の列のルールに当てはまらないようにすれば自由。
- ・コース、性別、処置-対照群など、2群以上にグルーピングできる属性データ。
- ・数値ではなく、文字列で入力するのが無難（1/2ではなく、1組/2組 or A/B など）

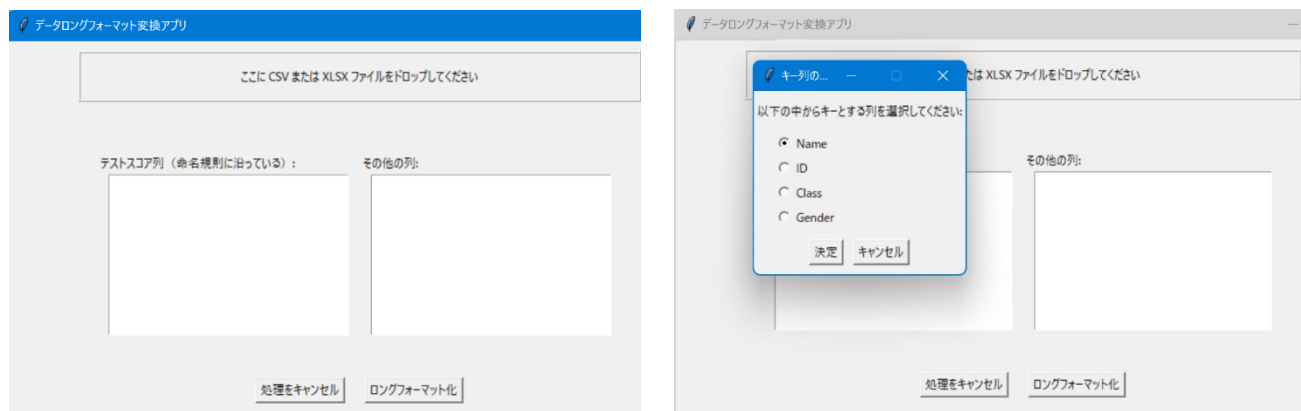
「**共変量データ**」

- ・試験スコア以外の数値データ。アンケートスコア、提出物の状況(%)、睡眠時間など。（小数点可）
- ・列名は「試験スコア」と同様のフォーマットに準ずる。
- ・時刻データには非対応なので、時間の長さ等に変換する。

ステップ 2 : app_Wide2Long によるデータフォーマットの変換

①「app_Wide2Long.exe」ファイルを実行すると、デスクトップアプリが起動する。

画面上部エリアに、ステップ 1 で整理したデータファイルをドラッグ&ドロップで入力。

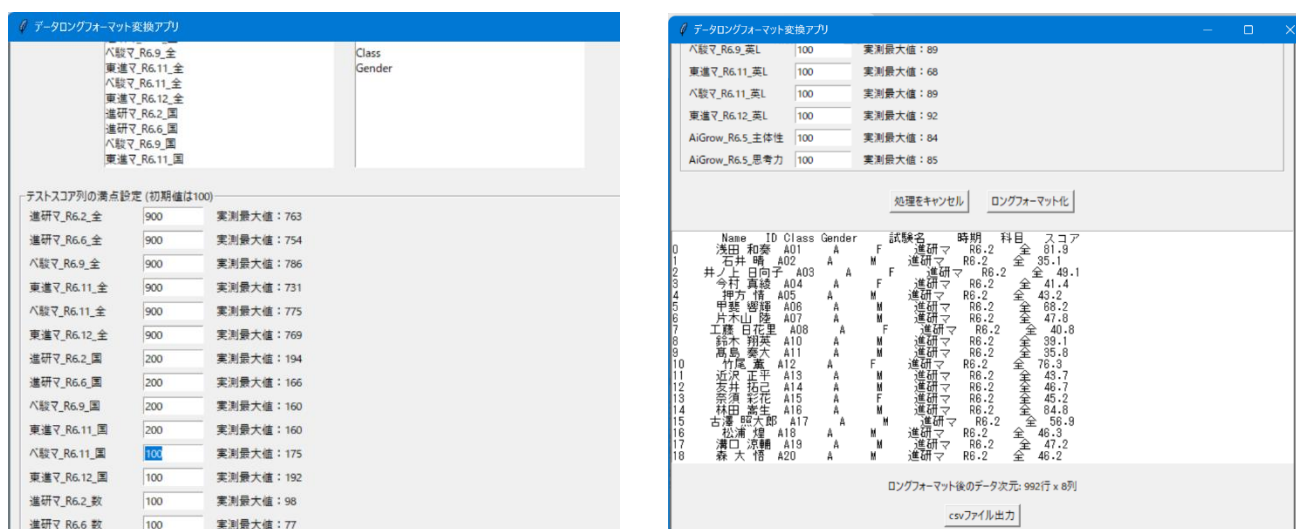


②複数シートに分けた xlsx ファイルを入れると、「主キー」を選択するように求められるので、「Name」か「ID」を選択（ステップ 1 -②で選んだ方）。

※マウスホイールで画面をスクロール出来ない場合は、右側のスライドバーで調整する。

③欠損値のカウント、「テストスコアとして見なされた列」「それ以外として見なされた列」が表示されるので、それぞれ問題ないか確認。問題がある場合はステップ 1 からやり直す。

④「テストスコアの満点設定」をそれぞれ調整する（初期値は 100 に設定されているが、後の分析で比較可能にするため全て%スケールに変換する）。全ての値の調整が済んだら、画面下部の「ロングフォーマット化」ボタンを押す。



④ロングフォーマットのプレビューが表示されたら、画面最下部の「csv ファイル出力」を押すことで、csv ファイルが出力される（ファイル名は何でもよい）。

ここで作成したファイルを、ステップ 3 で用いる。

ステップ 3 : Interactive_Visualizer でグラフを描画する

「Interactive_Visualizer.ipynb」を Google Drive 上においてファイルを開く。基本的には、データを読み込んで操作してみれば使い方が分かるようになっているので、詳細な説明は省く。

主な用途と細かい仕様上の注意点について、以下記載。

◆共通事項

- ・グラフ内の要素にカーソルを当てると、各種統計値が表示される。
- ・Boxplot 以外、層別分析時に凡例（図の右側に示される表）の項目をクリックすると、該当する個人やグループがハイライトされる。また、Shift キーを押しながらクリックすることで複数選択できる。
- ・データ点が描画されているグラフは、凡例だけでなくデータ点をクリックすることでも属性別ハイライトができる。

●Interactive Line Graph (for Group Analysis)

用途：複数試験スコア×複数メンバーの時系列分析

複数試験スコア×属性別平均の時系列分析

●Interactive Line Graph (for Row-/Column-wise Analysis)

用途：複数試験スコア（科目別）×個人の時系列分析

複数試験スコア（科目別）×属性別平均の時系列分析

●Interactive Boxplot & Confidence Interval

用途（箱ひげ図モード）：試験スコアの分布把握（サンプルサイズ小）

用途（群間比較モード）：単一試験スコアの信頼区間表示（簡易的な統計検定）

- ・Jitter 幅スライダー …データ点のばらつき幅を調整するために加えたが、実際にはばらつきの乱数を変えるだけで、幅の調整はできない（Altair に実装されていない）。ただし、値が 0 のときのみ、全データ点が直線状に並ぶ。
- ・群間比較モード時、信頼区間の有意水準 α を設定できる。

●Interactive Histogram & Density Line Graph

用途：試験スコアの分布把握（サンプルサイズ大）

●Interactive Bar Graph

用途：個人の単一試験における順位把握

●Interactive Scatter Plot

用途：試験 2 種類の相関分析

- ・x 軸、y 軸には「試験スコア」「共変量データ」を設定できる。