

データサイエンティスト 体験講座

本日の予定

1. 事前説明
2. データサイエンティストとは
3. 実技演習
(データサイエンティスト体験講座)

本日のインターンについて

4時間で体験できること…

- ①データサイエンティストが なんとなく わかる！
- ②分析案件の流れが おおむね わかる！
- ③分析手法を ある程度 体感できる！

諸連絡

休憩について

休憩は**適宜(計3回)**とります。
具合の悪い方はZoomのチャット機能でお知らせください。

資料の取り扱いについて

配布した資料は、お持ち帰りいただけます。
無断複製・転載・配布行為はご遠慮ください。

諸連絡【Web版】

Zoomのマイクについて

発言時、または、グループワーク時以外は「OFF」に設定をお願いします。

Zoomのカメラについて

受講者の皆様の反応を見させていただきながら進めたいため、可能な限り常時「ON」に設定をお願いします。

下記に該当する場合は「OFF」のまま受講いただいても問題ございません。

- ・ 顔を出すことに抵抗がある
- ・ 回線が重い
- ・ PCにカメラが付いていない、または、故障している

諸連絡【Web版】

配布した資料を閲覧できない場合

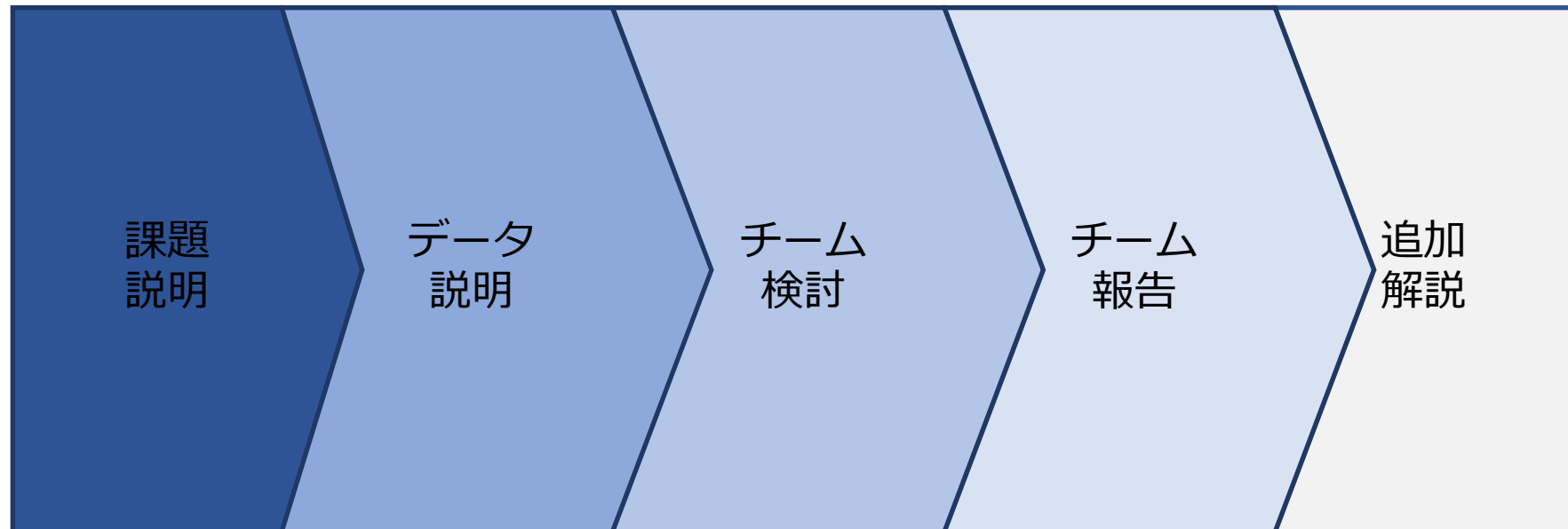
原因の調査や復旧を試みる必要はございません。
スタッフが適宜画面共有をしますので、
スタッフの画面を見ながらご参加ください。

演習中の見学について

グループワーク時の社員の見学をご了承ください。

演習の進め方について

これから提示する各課題については、下記フローを1セットとし進めて行きます。



<報告時に意識して欲しいこと>

課題の検討結果を報告する際は、分析を進める上でどのように予想や検討結果が変化していったのかも報告してください。手元にあるデータでどのような事が考えられるかを予想し **意見交換** して下さい。

本日のストーリー説明

今から皆様は…

ある**ゲーム会社**のデータサイエンス部門に
配属された**データサイエンティスト**です！

売り上げが**伸び悩んでいる**ゲームについて
データ分析によって**原因を特定してほしい**と依頼されています。

2018									2019											
1月	～	6月	7月	8月	9月	10月	11月	12月	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月
ア プ リ リ リ ス	～																			
		参 画																		

ゲームの説明

ゲームのディテールは皆様のご想像に**お任せ**します！

モンスターがいて、クエストがあって、
強化したり、コレクションしたり、課金したり、
そのようなイメージです。

ただし、下記**クエスト解禁ルール**だけは設定します。

- ・ 1日、1クエストまで解禁される
- ・ クリア後、翌日に次レベルのクエスト解禁がある
- ・ クリアしていない場合、解禁はない

1.基礎集計

いくつかのデータを渡され、
下記について調べてほしいと言われました。

- どのような**特徴**があると感じるか
- 分析を**上手くいかせなくする要因**がないか

まずは、次スライドのデータを見てみましょう。

1.基礎集計



01_1_お客様データ

ゲームをプレイしている**お客様データ**。
量が膨大になってしまうので、
偏りなく抜粋して提供しています。

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F	G
1	user_id	地域	年齢	性別	機種	登録日	累計課金額
2	3527	広島	49	女	iphone	2018/6/1	8096
3	3528	宮城	26	女	Android	2018/6/1	0
4	3529	静岡	26	男	iphone	2018/6/1	86848

1.基礎集計【Web版】

【練習問題】

「01_1_お客様データ」の中身を見て
下記 を埋めてください。

(目安：15分 **00:00**まで)

※グループのメンバー確認や自己紹介含む



Google Sheets

1.基礎集計 ▾

というシートに記入してください

	A	B	C	D	E	F	G	H
1	No	事実				解釈		記入者
2	1	男性：女性 =	<input type="text"/>	:	<input type="text"/>	<input type="text"/>	性ユーザーが多い	<input type="text"/>
3	2	年齢 最大値：	<input type="text"/>					<input type="text"/>
4		最小値：	<input type="text"/>					

1.基礎集計【Web版】

練習問題の要領でデータの中身を見て
傾向や特徴を確認してみましょう。

(目安：20分 **00:00**まで)

※この時間にやっていただきたいこと

- ・ お客様データを見て、データから何ができるかをチーム内で議論
- ・ 発表のためにスプレッドシートにチームの意見を記載



Google Sheets

1.基礎集計 ▾

というシートに記入してください

こちらは現時点では
入力・考慮不要です

	A	B	C	D	E	F	G	H
1	No	事実	解釈	記入者	発表		発表者：	
2	1				▼			
3	2				▼			
4	3				▼			

1.基礎集計

追加提供された下記データについても見てみましょう。



01_2_DAU
(daily_active_user)



01_3_日次課金率



01_4_課金単価

1.基礎集計



01_2_DAU
(daily_active_user)

日毎の**ログインユーザー数**。
※登録されているユーザー数ではなく、
ログインしたユーザー数です。

The screenshot shows an Excel spreadsheet titled "01_2_DAU(daily_active_user).xlsx". The spreadsheet has a header row with "日付" (Date) in column A and "DAU" (Daily Active Users) in column B. The data rows show dates from 2018/6/1 to 2018/6/4 and corresponding DAU values: 1062, 1073, 1080, and 1070.

	A	B	C	D
1	日付	DAU		
2	2018/6/1	1062		
3	2018/6/2	1073		
4	2018/6/3	1080		
	2018/6/4	1070		

1.基礎集計



01_3_日次課金率

日毎の**課金率データ**。
ログインユーザー数における、
課金ユーザー数の割合を示しています。

The screenshot shows an Excel spreadsheet with the following data:

	A	B
1	日付	課金率 (課金人数/ログイン人数)
2	2018/6/1	0.351095736
3	2018/6/2	0.353674321
4	2018/6/3	0.344590724

1.基礎集計



01_4_課金単価

日毎の**課金額**について、
平均と**中央値**を示しています。

01_4_課金単価.xlsx - Excel			
ファイル ホーム 挿入 ページレイアウト 数式 データ 校閲 表示 開発			
A1	日付		
	A	B	C
1	日付	1回の課金の課金額(平均)	1回の課金の課金額(中央値)
2	2018/6/1	1391.710269	864
3	2018/6/2	1460.111287	972
4	2018/6/3	1345.767612	972
	2018/6/4	1372.686700	1026

1.基礎集計

生のデータでは傾向が分かり辛かったと思います。

一般的な基礎集計の手法として
「集計」と**「可視化」**があります。

集計と可視化した下記データも見てみましょう。



お客様データ基礎集計

1.基礎集計



01_1_お客様データ



- 年齢の最大、最小は？
- 年齢の平均値は？
- 年齢の中央値は？

お客様データ基礎集計.xlsx - Excel

ファイル ホーム 挿入 ページレイアウト 数式 データ 校閲 表示 開発

A1 : X ✓ fx 集計結果\項目名

	A	B	C	D	E	F	G	H	I
1	集計結果\項目名	補足	user_id	地域	年齢	性別	機種	登録日	累計課金額
2	データ数	全ての項目	1227	1227	1227	1227	1227	1227	1227
3	重複しないデータ数	文字・日付項目のみ	1227	47		2	2	61	
4	最頻値	文字・日付項目のみ		東京		男	iphone	2018/07/09	
5	最頻値の度数	文字・日付項目のみ		134		752	798	30	
6	データ内最古の日付	日付項目のみ						2018/06/01	
7	データ内最新の日付	日付項目のみ						2018/07/31	
8	平均値	数値項目のみ			34.64873676				14879.43276
9	標準偏差	数値項目のみ			20.72100867				30805.63542
10	最小値	数値項目のみ			5				0
11	第一四分位数	数値項目のみ			21				0
12	第二四分位数(=中央値)	数値項目のみ			30				0
13	第三四分位数	数値項目のみ			44				16560
14	最大値	数値項目のみ			118				388608

集計結果 ヒント 参考資料

準備完了 100%

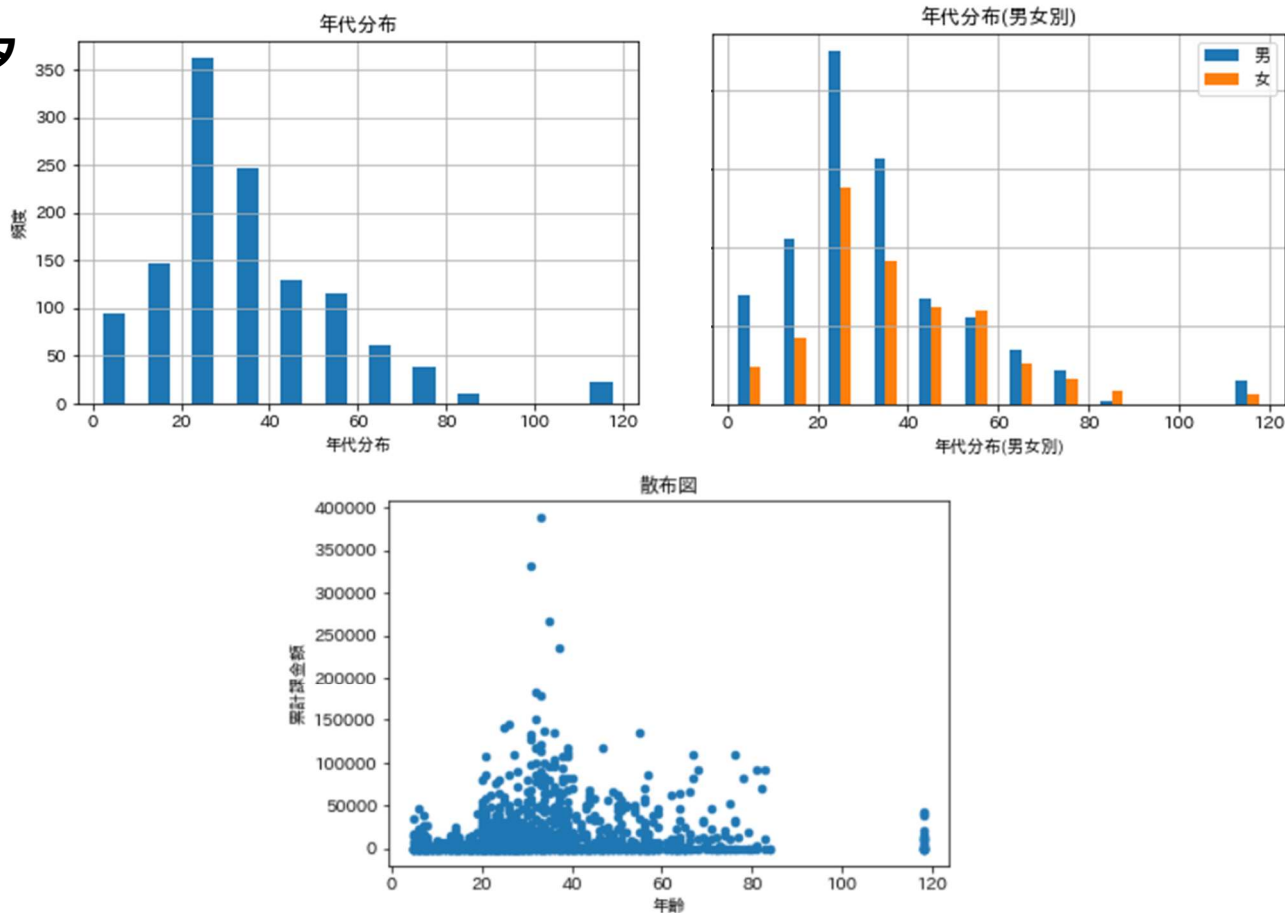
1.基礎集計



- データの分布はどうか？
- 視覚的な比較でわかることはあるか？

01_1_お客様データ

グラフ化



1.基礎集計【Web版】

追加提供データ、図表化・可視化データを確認しましょう。
報告資料をまとめましょう。

(目安：30分 **00:00**まで)

時刻になりましたらチーム毎に
下記の通り発表していただきます。

- ・1チーム3分以内
- ・特に共有したいポイント2～3個

発表に向けてこちらの
検討もお願いします

	A	B	C	D	E	F	G	H
1	No	事実	解釈	記入者	発表		発表者：	
2	1				▼			
3	2				▼			
4	3				▼			

本日のストーリー説明

ゲームリリース後、広告や話題性的には問題ないのですが、**アクティブユーザーは増えなくなっている**ことに気付きました。
原因を特定していきましょう。

2018										2019											
1月	～	6月	7月	8月	9月	10月	11月	12月		1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月
アプリリリース ～ 参画																					
					基礎集計 原因特定																

2.クロス集計

クロス集計とは…

1つの表に、縦列(=表側)、横列(=表頭)に異なるカテゴリーの項目を持たせ、それぞれの項目が交わる箇所に数量や比率を記載した表のこと。

⇒手早く、グループ化して数を数えたい、存在を確認したい等で非常に有効です。

		表頭		
		卒業	退学	
表側	資格あり	500	64	= 11.3%
	資格なし	875	153	= 14.9%

単位：人

2.クロス集計



縦: **日付**、横: **登録日からの経過日数**

※データ部分はログイン率(%)

=日付のログイン人数/～日前に登録した人数×100

02_1_継続率
(登録日別ユーザー×日付)

【データ ※ の読み方紹介】

※登録日からの経過日数4～6日の人の
2018/6/6のログイン率

2018/6/6を基準(=1)に
4～6日前に該当する日付を調べる

2018/6/1～2018/6/3に
ゲームに登録した人の内、79%が
2018/6/6にログインしたことを示す。

	A	B	C	D
1	登録日からの経過日数別ログイン率			
2	日付	1～3日	4～6日	7～9日
3	2018/6/1	99	⑥	82
4	2018/6/2	100	⑤	78
5	2018/6/3	100	④	81
6	2018/6/4	100	③	80
7	2018/6/5	100	②	80
8	2018/6/6	100	①	79
9	2018/6/7	100		80
10	2018/6/8	100		77
11	2018/6/9	100		83

2.クロス集計



縦: **日付**、横: **登録日からの経過日数**

※データ部分はログイン率(%)

=日付のログイン人数/～日前に登録した人数×100

02_1_継続率
(登録日別ユーザー×日付)

【補足】

登録日からの日数が経過する毎に
同じカテゴリの人()は
右図のように遷移する。

つまり、2018/6/1～2018/6/3に
ゲームに登録した人のログイン率は
下記のように推移していると言える。

2018/6/3 : 100%

2018/6/6 : 79%

2018/6/9 : 66%

	A	B	C	D
1	登録日からの経過日数別ログイン率			
2	日付	1～3日	4～6日	7～9日
3	2018/6/1	③ 99	⑥ 82	⑨ 64
4	2018/6/2	② 100	⑤ 78	⑧ 66
5	2018/6/3	① 100	④ 81	⑦ 60
6	2018/6/4	100	③ 80	⑥ 64
7	2018/6/5	100	② 80	⑤ 58
8	2018/6/6	100	① 79	④ 64
9	2018/6/7	100	80	③ 63
10	2018/6/8	100	77	② 64
11	2018/6/9	100	82	① 66

2.クロス集計【Web版】

クロス集計結果を確認し、下記を検討しましょう。

（目安：15分 00:00まで）

- ①：ログイン率の傾向が変わった日付を特定してください。
- ②：傾向が変わった原因について仮説を考えてください。
- ③：②で考えた原因について手持ちのデータから裏付けられることはありますか。
裏付けられるデータがない場合、どのようなデータがあれば裏付けられますか。

時刻になりましたらチーム毎に
下記の通り発表していただきます。

- ・ 1チーム3分以内
- ・ 特に共有したいポイント2～3個



2.クロス集計 ▾

というシートに記入してください

本日のストーリー説明

原因を依頼主に報告してから数ヶ月。

原因に対して、様々な対策検討、コンテンツの増強が行われた。
(ガチャのデザイン変更、お助け機能の追加 等)

年明け以降も安定稼働を続けている中、
既存顧客満足度向上施策が検討されるのであった。

2018									2019											
1月	～	6月	7月	8月	9月	10月	11月	12月	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月
アプリリリース ～ 参画																				
					基礎集計 原因特定		対策 検討・対応		既存顧客 満足度向上施策											

4.ユーザー行動のグループ化

サービス初期から、
高難度クエスト、コンプリート要素、
フレンド機能、ランキング、対人戦 等
の要素が追加されました。

その為、**ユーザー毎**に楽しみ方が
多様化しているのではないかと思います。

ゲーム内の行動からどのような楽しみ方をしているか
グループに分けましょう。

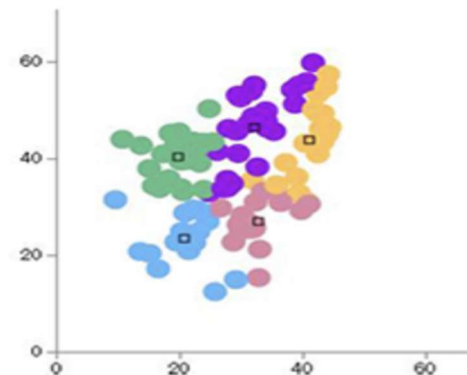
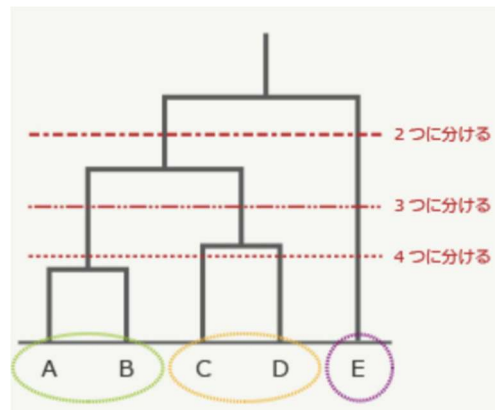


4. ユーザー行動のグループ化

【グループ化を行う統計的な分析手法：クラスター分析とは】

データの集合を共通の特徴を持つような**部分集合に切り分ける**手法。

切り分け方は様々あり、
大きく分けて**階層的**な手法と**非階層的**な2つの手法が存在する。



この分析は探索的な分析手法であり、
結果は、**人間の解釈が必要**なものとなる。

その為、**客観的な証拠にはならず**、
あくまでデータの知見、直感的な理解を深めるための手法である。

4.ユーザー行動のグループ化



ユーザー毎の下記**行動傾向集計**結果。

- ・クエスト系行動
- ・フレンド系行動
- ・コンプリート系行動

04_ユーザー行動傾向集計

	A	B	C	D	E
1		クエスト系行動			
2	ユーザーID	クエスト実施回数(ソロ)	クエスト実施回数(マルチ)	総与ダメージ量	総被ダメージ量
3	1	1943	1907	481840	32569
4	2	1556	1694	480322	43749
5	3	1523	1572	405623	44976

	A	F	G
1		フレンド系行動	
2	ユーザーID	フレンド数	メッセージ送受信数
3	1	211	1611
4	2	268	1645
5	3	295	1944

	A	H	I	J
1		コンプリート系行動		
2	ユーザーID	有料ガチャ回数	無料ガチャ回数	図鑑数
3	1	423	4258	1542
4	2	404	4740	1890
5	3	470	3948	1633

4.ユーザー行動のグループ化



04_ユーザー行動傾向
集計_クラスタ結果付き

ユーザー毎の行動傾向集計結果に対して
クラスター分析を行い、
ユーザーを**5つのグループに分けた**結果。
0~4のクラスターIDが割り振られた。

04_ユーザー行動傾向集計_結果付き.xlsx - Excel

	A	H	I	J	K
1		コンプリート系行動			
2	ユーザーID	有料ガチャ回数	無料ガチャ回数	図鑑数	クラスターID
3	1	423	4258	1542	1
4	2	404	4740	1890	1
5	3	470	3948	1633	1

準備完了 160%

4.ユーザー行動のグループ化



クラスター分析でグループ分けした
グループ毎に各項目を集計した結果。

04_ユーザー行動傾向集計_ クラスター毎集計結果

	A	B	C	D	E	F	G	H
1			クエスト系行動				フレンド系行動	
2	クラスターID	件数	クエスト実施回数(ソロ)	クエスト実施回数(マルチ)	総与ダメージ量	総被ダメージ量	フレンド数	メッセージ送受信数
3	0	30	1256.53	728.80	221441.83	21713.33	76.40	942.27
4	1	10	1729.90	1667.20	407458.60	38850.50	243.40	1759.60
5	2	10	151.00	177.00	35255.30	3526.30	21.00	171.30
6	3	20	374.70	387.90	33689.55	3559.40	36.90	387.15
7	4	30	723.43	1249.27	229544.97	21987.40	156.93	2460.47

	A	B	I	J	K	L	M	N
1			コンプリート系行動					
2	クラスターID	件数	有料ガチャ回数	無料ガチャ回数	図鑑数	カテゴリ	クラスターの特徴	このクラスターに有効な施策
3	0	30	244.70	1420.57	990.13			
4	1	10	449.10	3921.40	1735.30			
5	2	10	71.80	757.70	420.70			
6	3	20	351.15	2472.80	1388.05			
7	4	30	149.73	1458.57	1005.97			

4.ユーザー行動のグループ化 【Web版】

データを確認し、下記を検討しましょう。

(目安：15分 **00:00**まで)

- ③結果を確認し、各クラスタの特徴を考えてみましょう。
- ④各ユーザー層向けの施策を検討してみましょう。



Google Sheets

4-2.ユーザー行動のグループ化 ▼

というシートに記入してください