

# ユーザー調査の可視化

国勢調査、Twitterのツイートを用いてブログターゲットユーザー＝フルキャリア女性を理解する

# 目次

- テーマの選定理由
- 分析プロジェクトサマリー
- 分析アプローチ全体像
- 統計データへのアプローチ
- ツイートデータへのアプローチ
- 改善と今後に向けて

# 1. テーマ選定理由

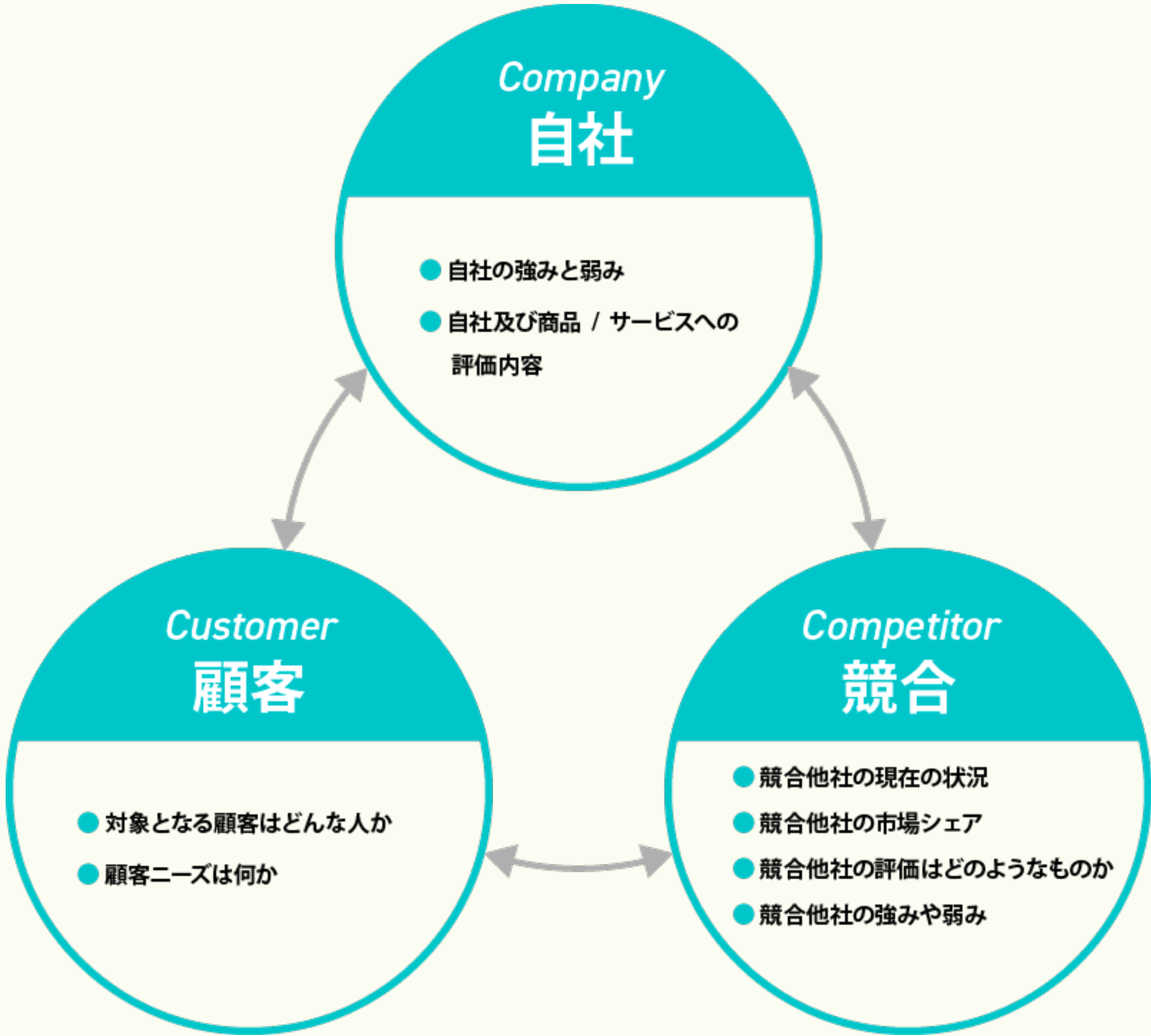
## 理由① フルキャリア女性へ向けたブログを開始したいから

フルキャリア:ライフもキャリアも前向き



**フルキャリア女性を応援したい&情報発信をしていきたい**

理由② ターゲットを知るの大切だと思っているから



# 3C分析の顧客分析をテーマとして設定

## 2.分析プロジェクトサマリー

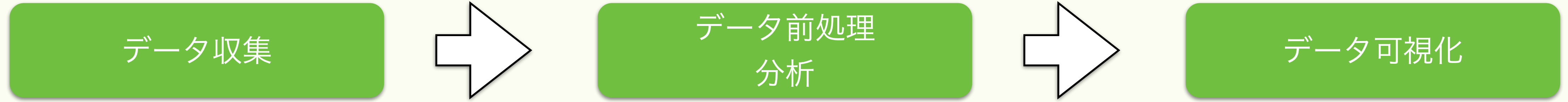
# 分析プロジェクトサマリー

ビジネス課題	<ul style="list-style-type: none"><li>● フルキャリアになりたい女性向けのブログ開始予定だが、ターゲット像が掴めていない</li></ul>
分析ゴール	<ul style="list-style-type: none"><li>● フルキャリアの女性の興味・関心を把握</li></ul>
KPIとその目標	<ul style="list-style-type: none"><li>● 月間5000PV</li></ul>
アウトプット	<ul style="list-style-type: none"><li>● フルキャリアになりたい女性の思考性や現状を可視化するダッシュボード</li></ul>
使用するデータ	<ul style="list-style-type: none"><li>● 国で実施の調査、Twitterデータ</li></ul>
分析手法の仮説	<ul style="list-style-type: none"><li>● 可視化のためのデータ前処理</li><li>● フリーテキスト情報は、自然言語処理をしWordCloudで可視化</li></ul>
評価方法	<ul style="list-style-type: none"><li>● ABテスト(分析結果を用いている記事と用いていない記事)</li></ul>

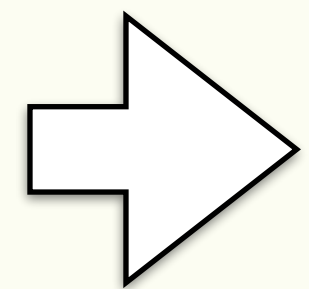
# 3.分析アプローチ全体像



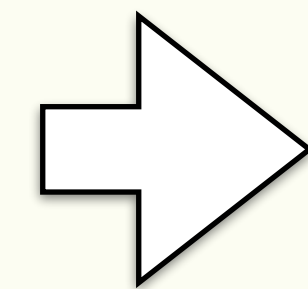
# 分析アプローチ全体像



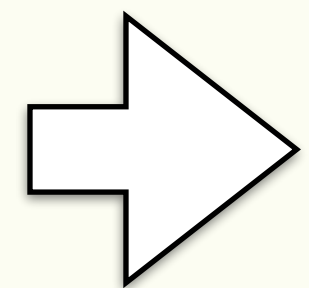
統計データ



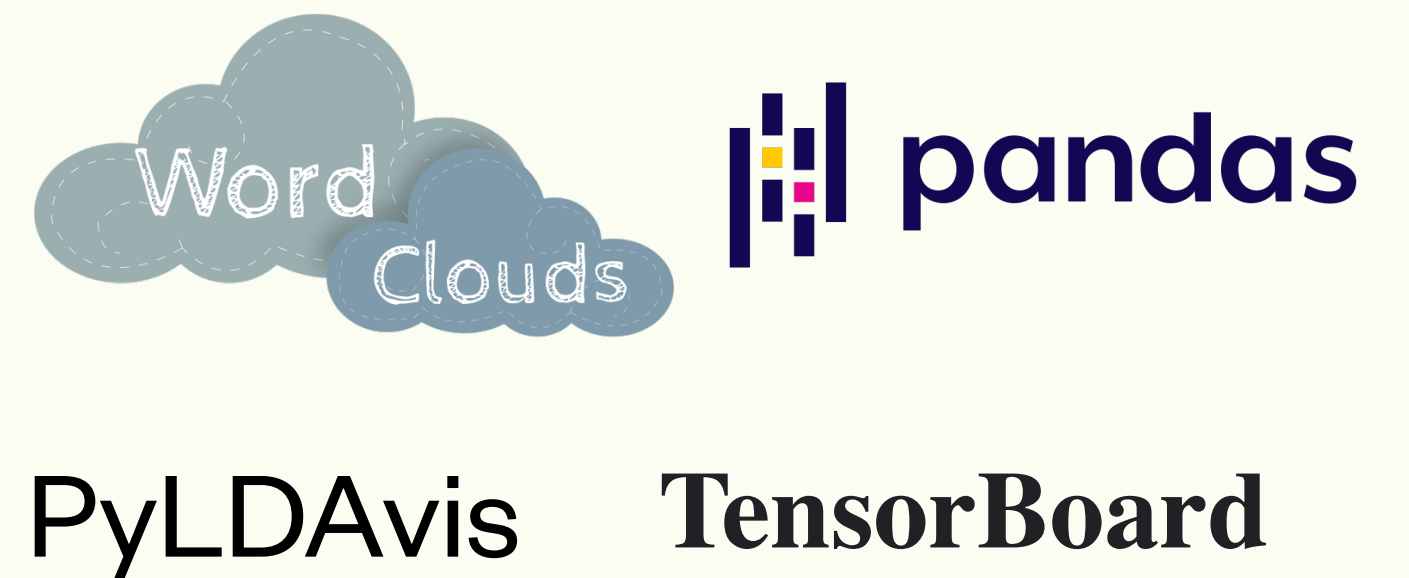
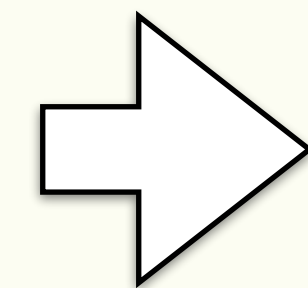
- 分析対象データのみ抽出
- 欠損が多いデータを整理



ツイートデータ



- ストップワード設定
- 形態素解析
- トピックモデル



## 4.統計データへのアプローチ

# データ収集

## 1.e-StatサイトやGoogleで検索

## 使えるようなデータリストアップ

## 該当データをエクセルでダウンロード

The screenshot shows the e-Stat website interface. On the left, there are search filters for 'データ種別' (Data Type) with 'データベース' (282) and 'ファイル' (692) options. Below are dropdown menus for filtering by '統計分野', '組織', '統計の種類', '政府統計名', '提供周期', '調査年', '調査月', '50音', '統計表フォーマット', and '集計地域区分'. The main content area is titled '政府統計一覧' (List of Government Statistics) and shows a table of statistics with columns for '政府統計コード', '政府統計名', and '概要'. The table lists various statistics such as '民間企業退職金実態調査', '一般職国家公務員在職状況統計表', and '職種別民間給与実態調査'. A pagination bar at the top right indicates '1/14ページ'.

取得データ	リンク
男女,ふだんの就業状態,ライフステージ,頻度,学習・自己啓発・訓練の種類別行動者率(15歳以上)ー全国	<a href="https://www.e-stat.go.jp/stat-search/files?page=1&amp;layout=datalist&amp;cycle=0&amp;toukei=00200533&amp;tstat=000001158160&amp;tclass1=000001158164&amp;tclass2=000001158165&amp;tclass3=000001158167&amp;tclass4=000001158168&amp;tclass5val=0&amp;stat_infid=000032222788">https://www.e-stat.go.jp/stat-search/files?page=1&amp;layout=datalist&amp;cycle=0&amp;toukei=00200533&amp;tstat=000001158160&amp;tclass1=000001158164&amp;tclass2=000001158165&amp;tclass3=000001158167&amp;tclass4=000001158168&amp;tclass5val=0&amp;stat_infid=000032222788</a>
男女,ふだんの就業状態,ライフステージ,目的,学習・自己啓発・訓練の種類別行動者率(15歳以上)ー全国	<a href="https://www.e-stat.go.jp/stat-search/files?page=1&amp;layout=datalist&amp;cycle=0&amp;toukei=00200533&amp;tstat=000001158160&amp;tclass1=000001158164&amp;tclass2=000001158165&amp;tclass3=000001158167&amp;tclass4=000001158168&amp;tclass5val=0&amp;stat_infid=000032222790">https://www.e-stat.go.jp/stat-search/files?page=1&amp;layout=datalist&amp;cycle=0&amp;toukei=00200533&amp;tstat=000001158160&amp;tclass1=000001158164&amp;tclass2=000001158165&amp;tclass3=000001158167&amp;tclass4=000001158168&amp;tclass5val=0&amp;stat_infid=000032222790</a>
付表99 女性が職業をもつことに対する意識	<a href="https://www.mhlw.go.jp/bunya/koyoukintou/josei-jitsujo/21.h">https://www.mhlw.go.jp/bunya/koyoukintou/josei-jitsujo/21.h</a>
曜日,年齢,男女,ふだんの就業状態,ライフステージ,行動の種類別行動者率(15歳以上)ー全国	<a href="https://www.e-stat.go.jp/stat-search/database?page=1&amp;que">https://www.e-stat.go.jp/stat-search/database?page=1&amp;que</a>

# データ前処理

令和3年社会生活基本調査 生活行動—全国(調査票A) 学									
第3-1表 男女,配偶関係,ふだんの就業状態,ふだんの健康状									
					行動者数	行動者数	行動者数	行動者数	行
					学習・自己啓 発・訓練の種類	学習・自己啓 発・訓練の種類	学習・自己啓 発・訓練の種類	学習・自己啓 発・訓練の種類	学 発
					22_商業実務・ ビジネス関係	3_介護関係	4_家政・家事(料 理・裁縫・家庭 経営など)	5_人文・社会・ 自然科学(歴史・ 経済・数学・生 物など)	6_
					(千人)	(千人)	(千人)	(千人)	(千
男女	配偶関係	ふだんの	ふだんの	年齢					
0_総数	0_総数	0_総数	4_あまり良	R6_(再掲)24	242	92	294	220	
0_総数	0_総数	0_総数	5_良くない	00_総数	81	45	188	99	
0_総数	0_総数	0_総数	5_良くない	01_15~19歳	2	-	-	4	
0_総数	0_総数	0_総数	5_良くない	02_20~24歳	13	5	7	3	
0_総数	0_総数	0_総数	5_良くない	03_25~29歳	11	3	11	1	
0_総数	0_総数	0_総数	5_良くない	04_30~34歳	7	1	26	19	
0_総数	0_総数	0_総数	5_良くない	05_35~39歳	3	2	11	5	
0_総数	0_総数	0_総数	5_良くない	06_40~44歳	6	3	14	5	
0_総数	0_総数	0_総数	5_良くない	07_45~49歳	4	7	16	11	
0_総数	0_総数	0_総数	5_良くない	08_50~54歳	11	6	17	9	
0_総数	0_総数	0_総数	5_良くない	09_55~59歳	5	3	8	7	
0_総数	0_総数	0_総数	5_良くない	10_60~64歳	6	4	9	6	
0_総数	0_総数	0_総数	5_良くない	11_65~69歳	3	0	10	6	
0_総数	0_総数	0_総数	5_良くない	12_70~74歳	3	1	16	4	
0_総数	0_総数	0_総数	5_良くない	13_75~79歳	4	5	20	7	
0_総数	0_総数	0_総数	5_良くない	14_80~84歳	-	0	10	6	
0_総数	0_総数	0_総数	5_良くない	15_85歳以上	2	5	13	6	
0_総数	0_総数	0_総数	5_良くない	R1_(再掲)20	79	45	188	95	
0_総数	0_総数	0_総数	5_良くない	R2_(再掲)34	48	36	144	72	

1.分析対象のデータを抽出

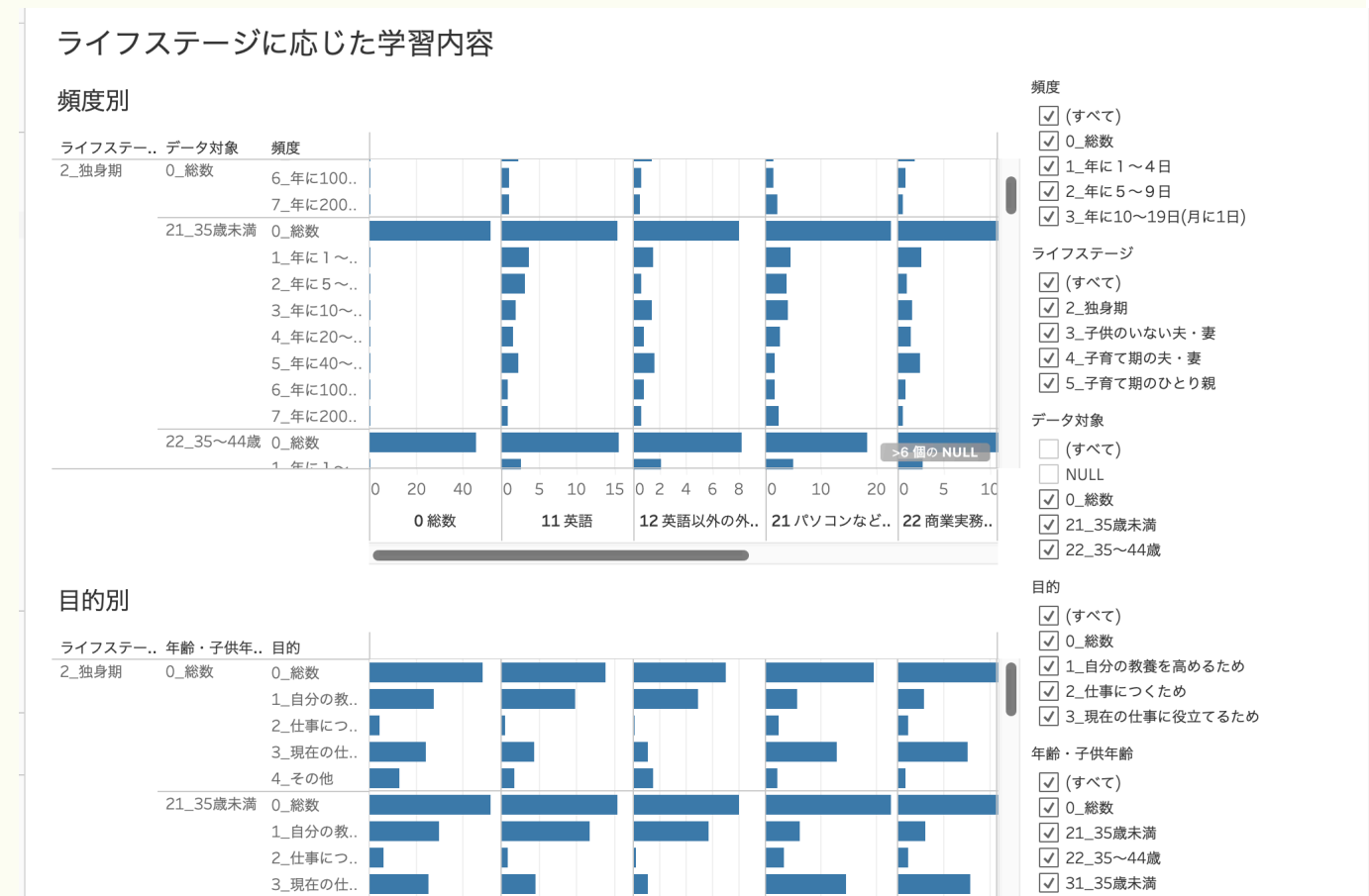
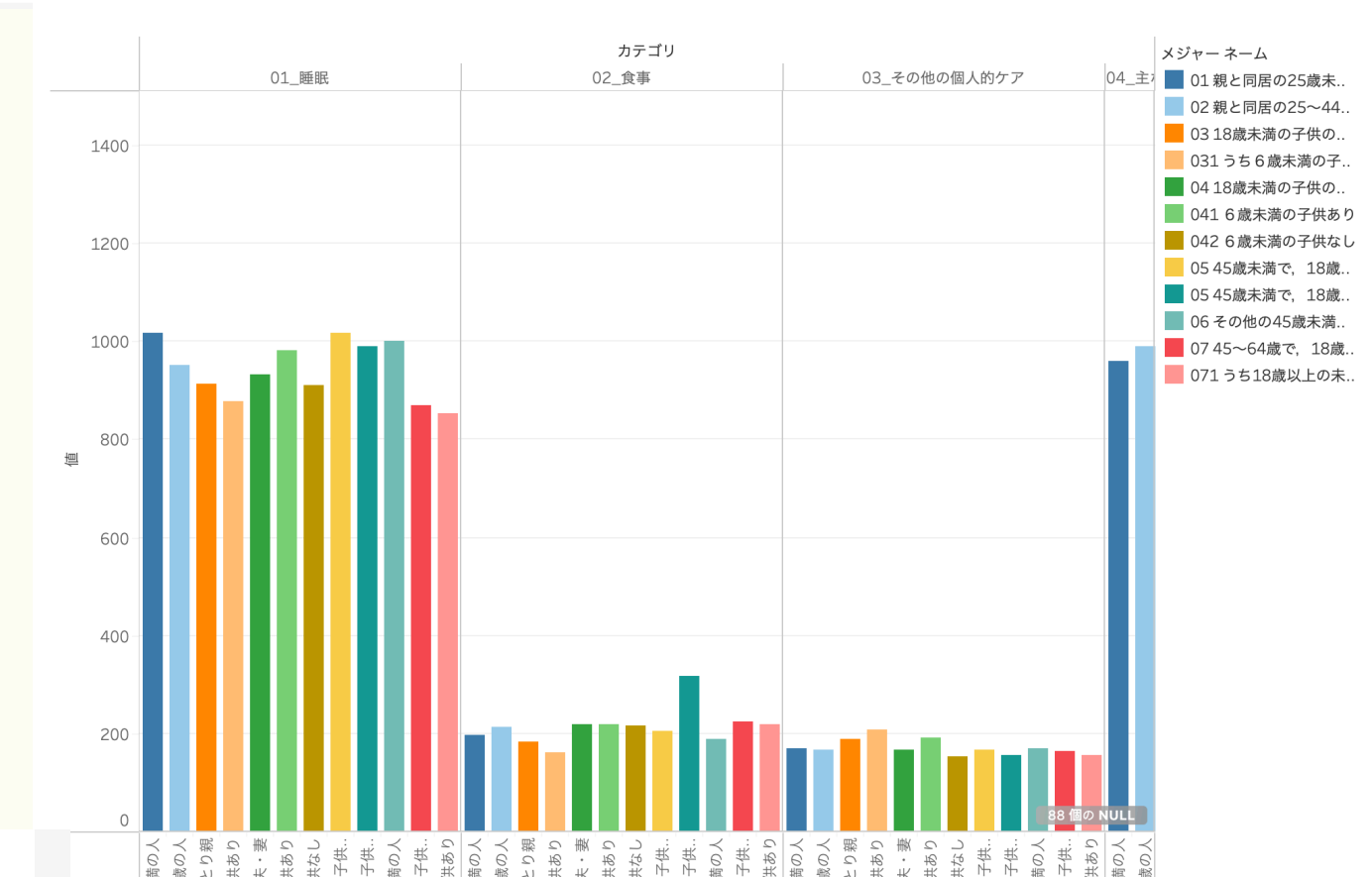
欠損が多いデータを置き換え

# データの取り込み

## 1. ExcelファイルをTableau Publicに取り込み

## データごとに適切な表示方法を選択

## ストーリーボードを作成



# アウトプット:女性が職業を持つことへの意識

1.70代以上は女性は仕事しないと考えている

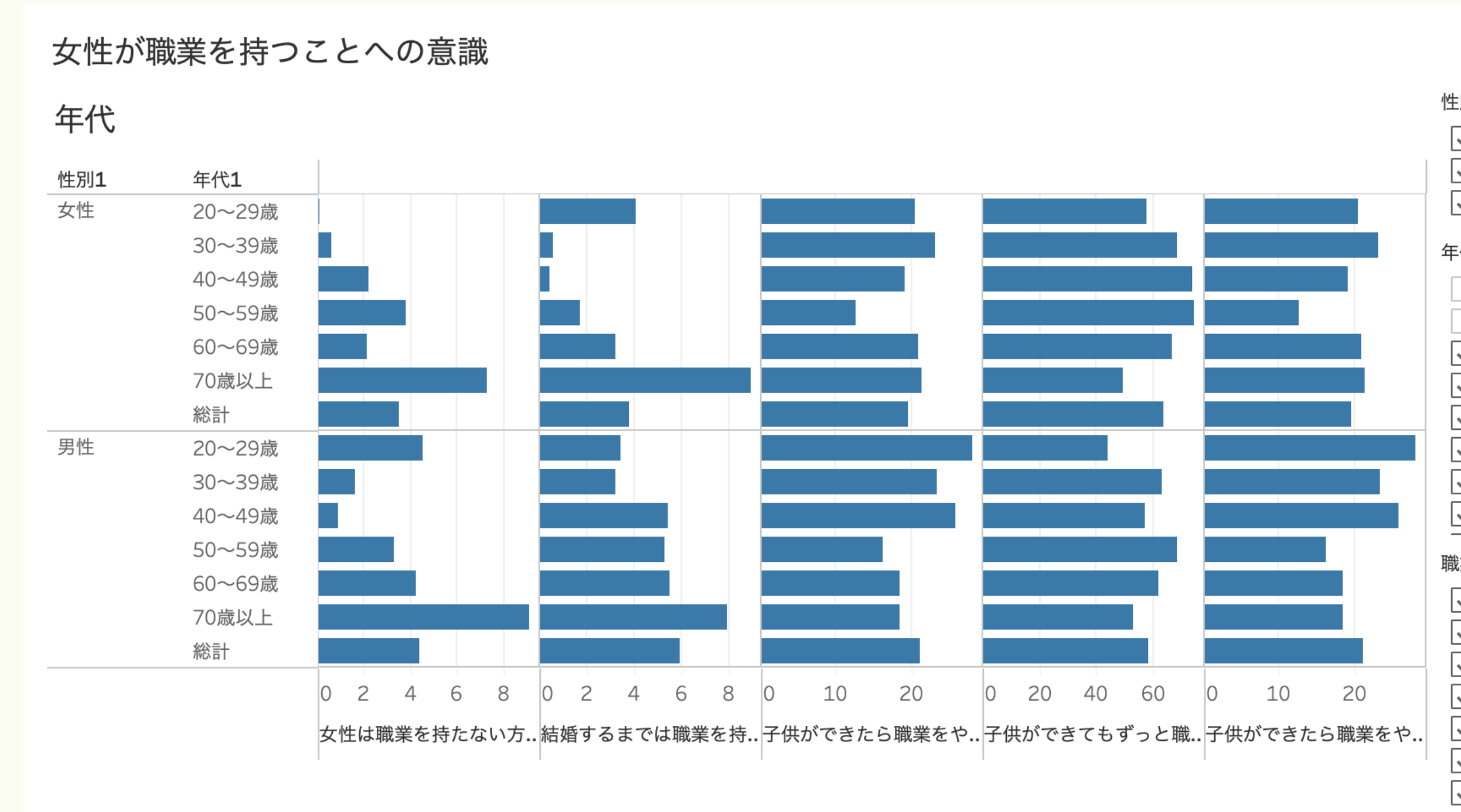
人が一定数いる

30代は男女ともに子供ができてても職業を続

けた方がいいと考えている傾向がある

20代男性は、子供ができたなら一度仕事を辞

めて欲しいと考えている傾向にある





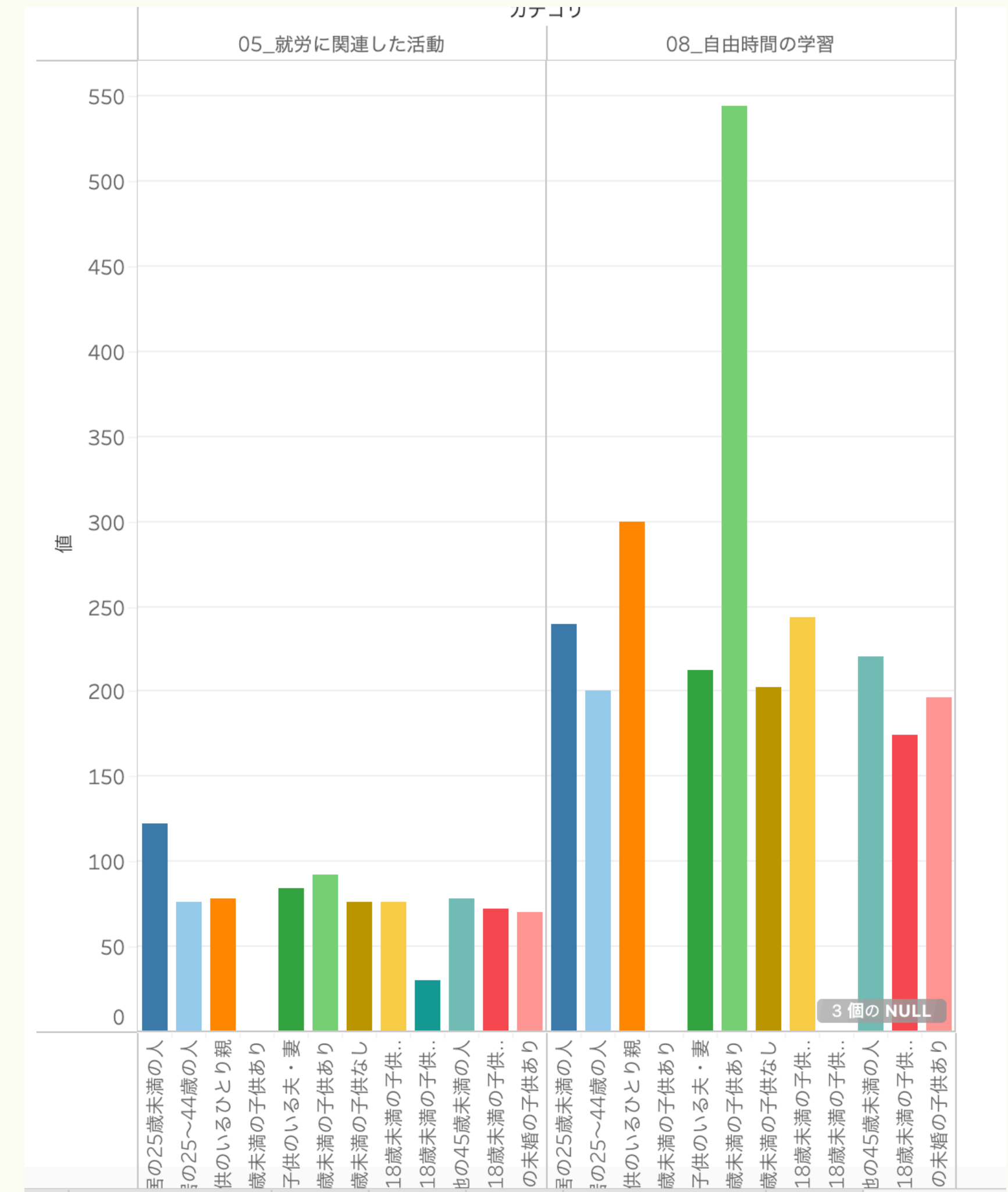
# アウトプット：ライフステージ別の1週間の1人ごとの平均時間

1.6才以下の子供を持つ親は、自由な学習時

間が多い

6才以下の子供を持つ親は就労に関連した活

動が多い





# アウトプット:ライフステージごとの学習内容と頻度

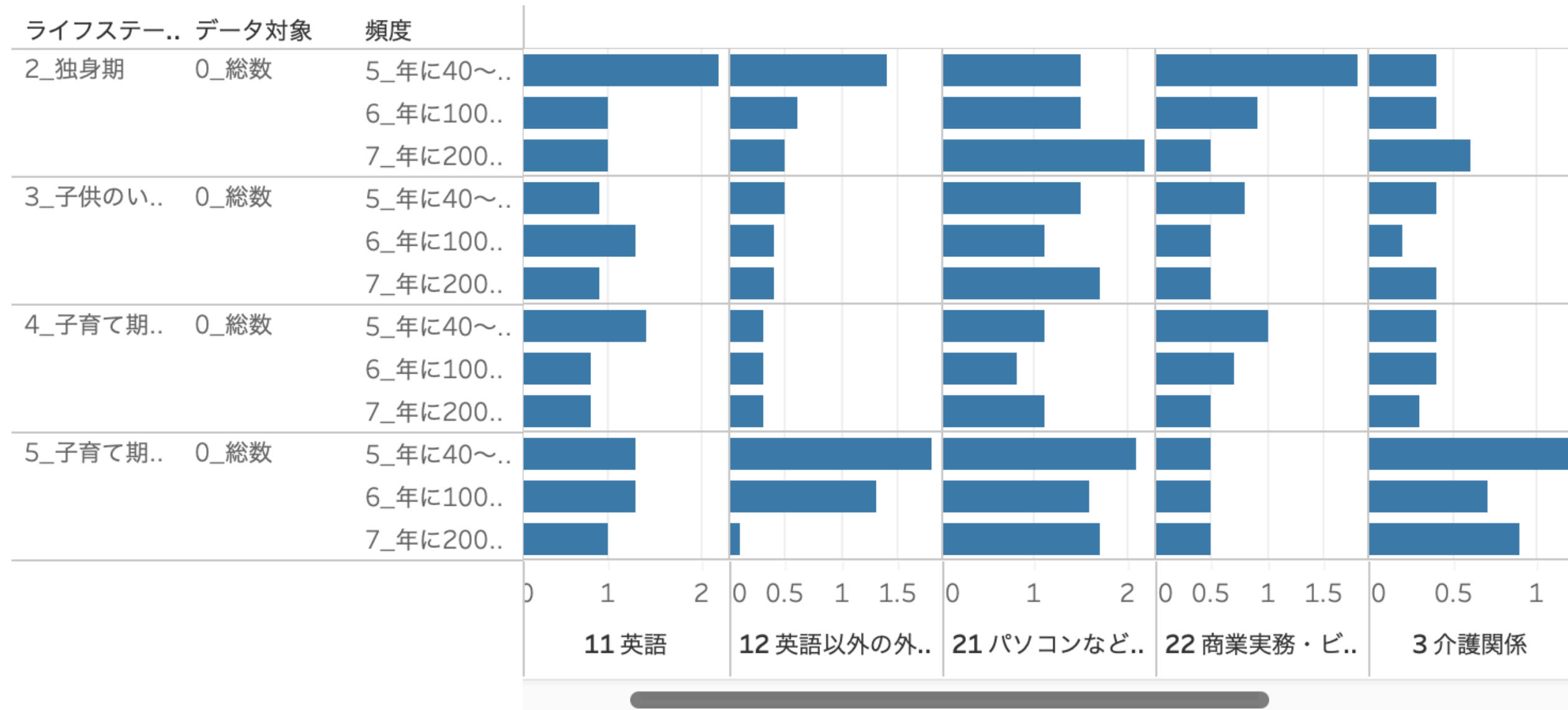
## 1.独身期はビジネスに関することも学習する

傾向がある

どのライフステージもパソコン関連の学習を

している

頻度別

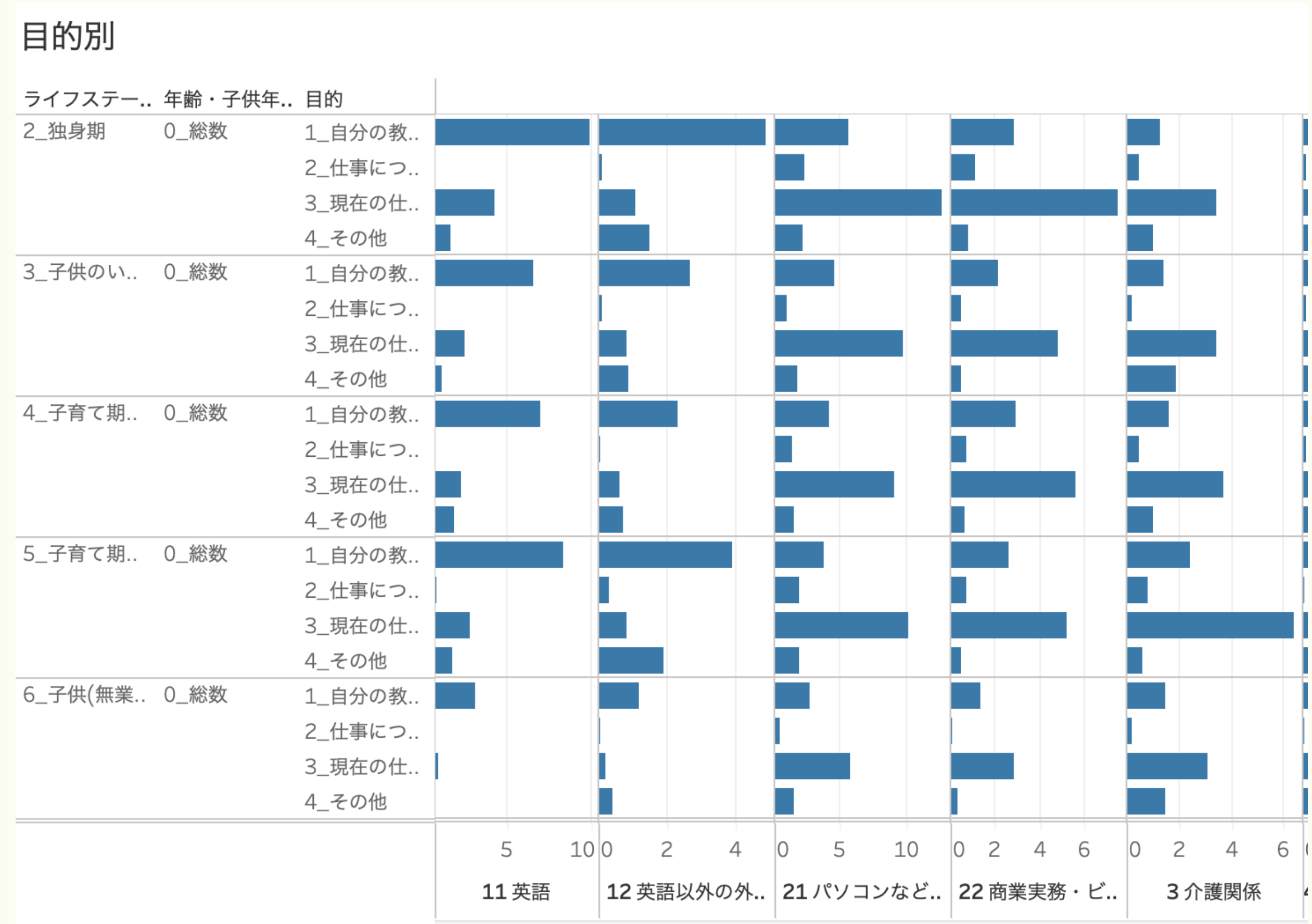


# アウトプット:ライフステージごとの学習内容と目的

1.英語は教養のために学習する傾向がある

パソコンやビジネスに関する内容は、現在の

仕事のために学習している



## 国勢調査のデータ分析からの考察

- 女性の社会進出は推奨されている傾向
- 独身期 仕事中心→結婚 夫中心矢印→子育て 子供中心  
に生活が変わっていている

## 国勢調査考察

### 仕事

- 女性の社会進出は推奨
- 30代は共働き家庭が多そう

### 夫婦

- 女性と男性の仕事時間は3時間差
- 家事や育児は女性が3時間以上多い

### 子供あり

- 6才以下の子供の場合は子供と時間が中心
- 6才以上の子供の場合は仕事中心な生活に変わっていく

**仕事と家庭の両立・家事分担・キャリアアップが悩みのタネ？**

# 5. ツイートデータへのアプローチ

# データ収集

## 1. Octoparse というスクレイピングツールを使用

フリーキャリア女性がつぶやきそうな単語をリストアップ

上記の単語を含むツイートを取得

テンプレートギャラリー

\* Keyword

共働き 家事  
共働き 疲れた  
仕事 家庭 両立  
仕事 家事 疲れた  
仕事 育児  
仕事 勉強  
キャリアアップ 勉強

\* Start time (yyyy-mm-dd)

2022-04-01

\* End time (yyyy-mm-dd)

2023-03-31

タスク名

全てのキーワード

グループ

デフォルト ▼

# データ前処理

## 1. 広告ツイートを対象外にする

いいね数やリツイート数などに値がないものは0

に置換

いいね数やリツイート数などの値を10Kの表記に

なっているものを、数字に置換

Category	Keyword	Tweet_W	Author_N	Author_V	Tweet_Ti	Tweet_Co	Tweet_In	Tweet_Vi	Tweet_Al	Tweet_Nr	Tweet_Nr	Tweet_Nr	Tweet_Nr
Post&am	共働き	https://tv	名前が浮	https://tv	2023-04-	これはき	https://pl	https://tv	False		71	280	21.7K
Post&am	共働き	https://tv	関口美奈	https://tv	2023-04-	妻「私もフルタイ	https://tv	https://tv	False	23	96	862	138.8K
Post&am	共働き	https://tv	けだま	https://tv	2023-04-	結婚したら妻に共	https://tv	https://tv	False	235	6,435	50.9K	7.4M
Post&am	共働き	https://tv	ソリティ	https://tv	2023-04-	一人目のときは夫	https://tv	https://tv	False	1	37	590	82.6K
Post&am	共働き	https://tv	あやめ@	https://tv	2023-04-	実は3月は超繁忙	https://tv	https://tv	False	7	3,478	21.7K	2.4M
Post&am	共働き	https://tv	ぽこみ	https://tv	2023-04-	既婚者男性たちに	https://tv	https://tv	False	27	165	1,189	153.2K
Post&am	仕事 家庭	https://tv	白石貴子	https://tv	2023-04-	重要！女性が輝く	https://tv	https://tv	False	38	95	328	4,319
Post&am	仕事 家庭	https://tv	ぽこみ	https://tv	2023-04-	結婚したり子供が	https://tv	https://tv	False	25	387	3,055	248.9K
Post&am	仕事 家庭	https://tv	選挙へ行	https://tv	2023-04-	これ分かるな。い	https://tv	https://tv	False	1	9	86	7,458
Post&am	仕事 家庭	https://tv	あかね@	https://tv	2023-04-	ママがこれやると	https://tv	https://tv	False	17	34	195	3,999
Post&am	仕事 家庭	https://tv	ふじみ	https://tv	2023-04-	これか	https://pl	https://tv	False	1	9	46	816

## データ分析:形態素解析

- **janomeのanalyzerモジュールを使用**

**ユーザーの趣向性把握しやすいかつブログ記事のキーワードとしても使用可能な名**

**詞と形容詞のみを取り出す**

**取得時に使ったキーワード、記号などをストップキーワードに設定**



アウトプット:出現単語をWordCloud表示

●2回以上出現する名詞と形容詞のみ

家族/家庭に関することが多め

楽しい=大変の気持ち



## データ分析:トピックモデル

- **複数のカテゴリに存在すると考え、トピックモデルで使用**

**今後もモデルを更新する可能性があるため、pLSAではなくLDAを使用**

# アウトプット:主成分分析とPyLDAvisで可視化

- 綺麗に分けられた4トピックに分割
- 分割数はより具体的な傾向を見た  
いので、大きい値で設定

## 4つの分割

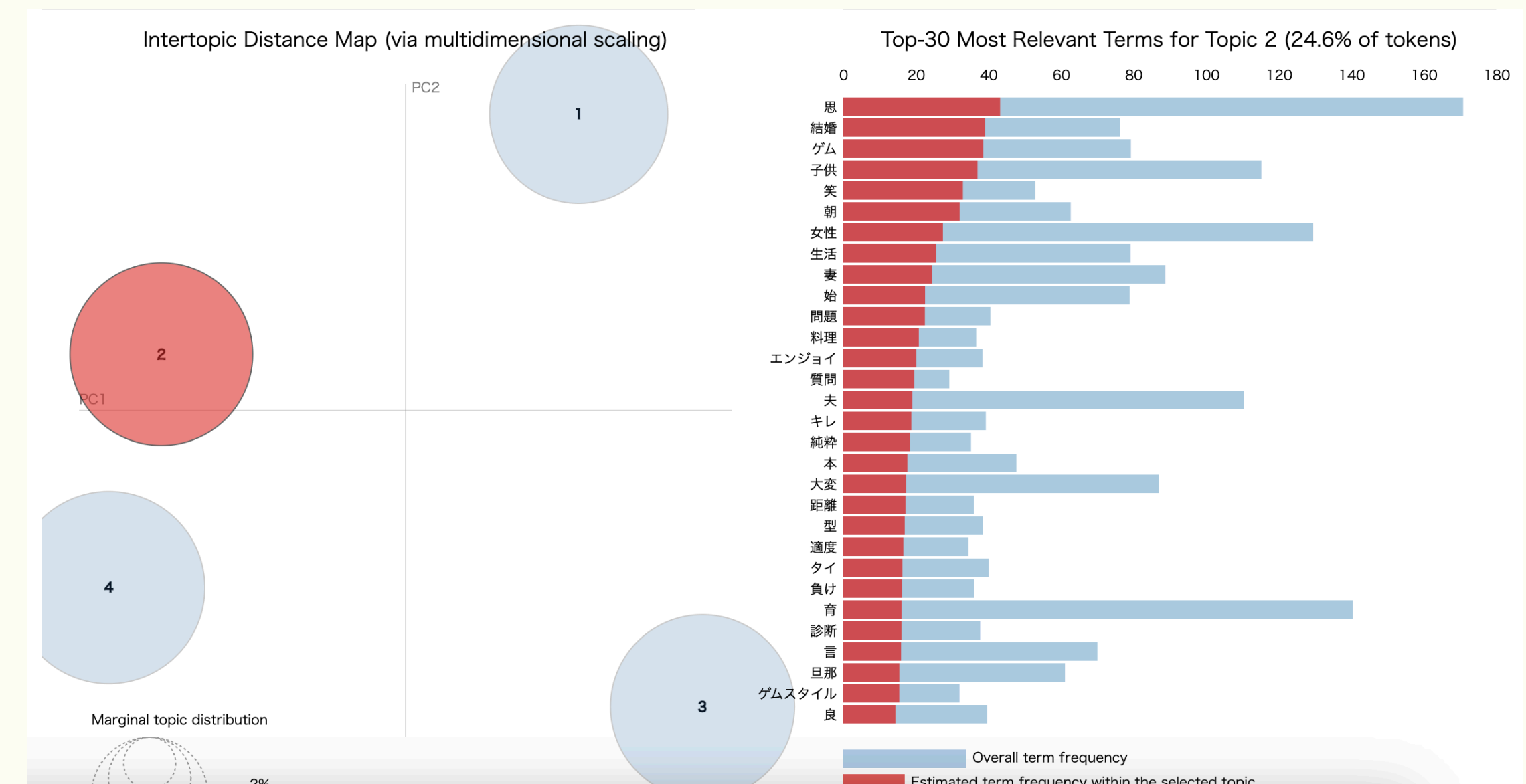
キャリアアップについて

結婚生活

子育て

夫との関係性

	topic1	topic2	topic3	topic4
0	思	思	育	育
1	女性	結婚	子供	夫
2	男性	ゲーム	妻	思
3	英語	子供	女性	ゲーム
4	旦那	笑	夫	大変
5	言	朝	大変	夫婦
6	キャリアアップ	女性	思	子供
7	始	生活	活	生活
8	座	妻	言	男性
9	希望	始	無理	分担
10	介護	問題	転職	時短
11	育	料理	育休	女性
12	子ども	エンジョイ	男性	家族
13	稼	質問	始	タイ
14	娘	夫	ママ	キャリアアップ



# 国勢調査→ツイッターデータの考察

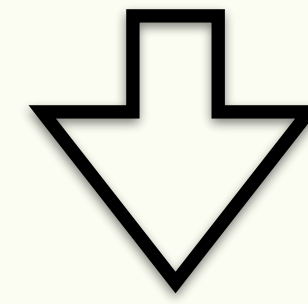
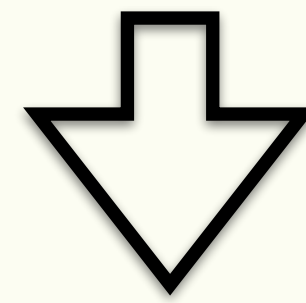
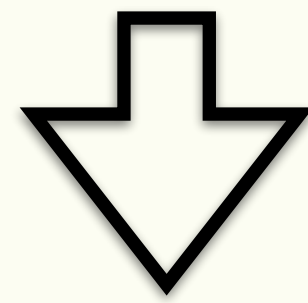
国勢調査

悩みのタネ

仕事と家庭の両立

家事分担

キャリアアップ



ツイッター

子育ての負担大

旦那の助け不足  
時短したい

稼ぎたい  
転職に興味  
フルタイム勤務

# 6.改善と今後に向けて

## 問題と改善ポイント

### 問題

- 国勢調査はクロス集計の結果になってしまい、1ユーザーごとの情報ではない
- ツイートデータ量が少ない

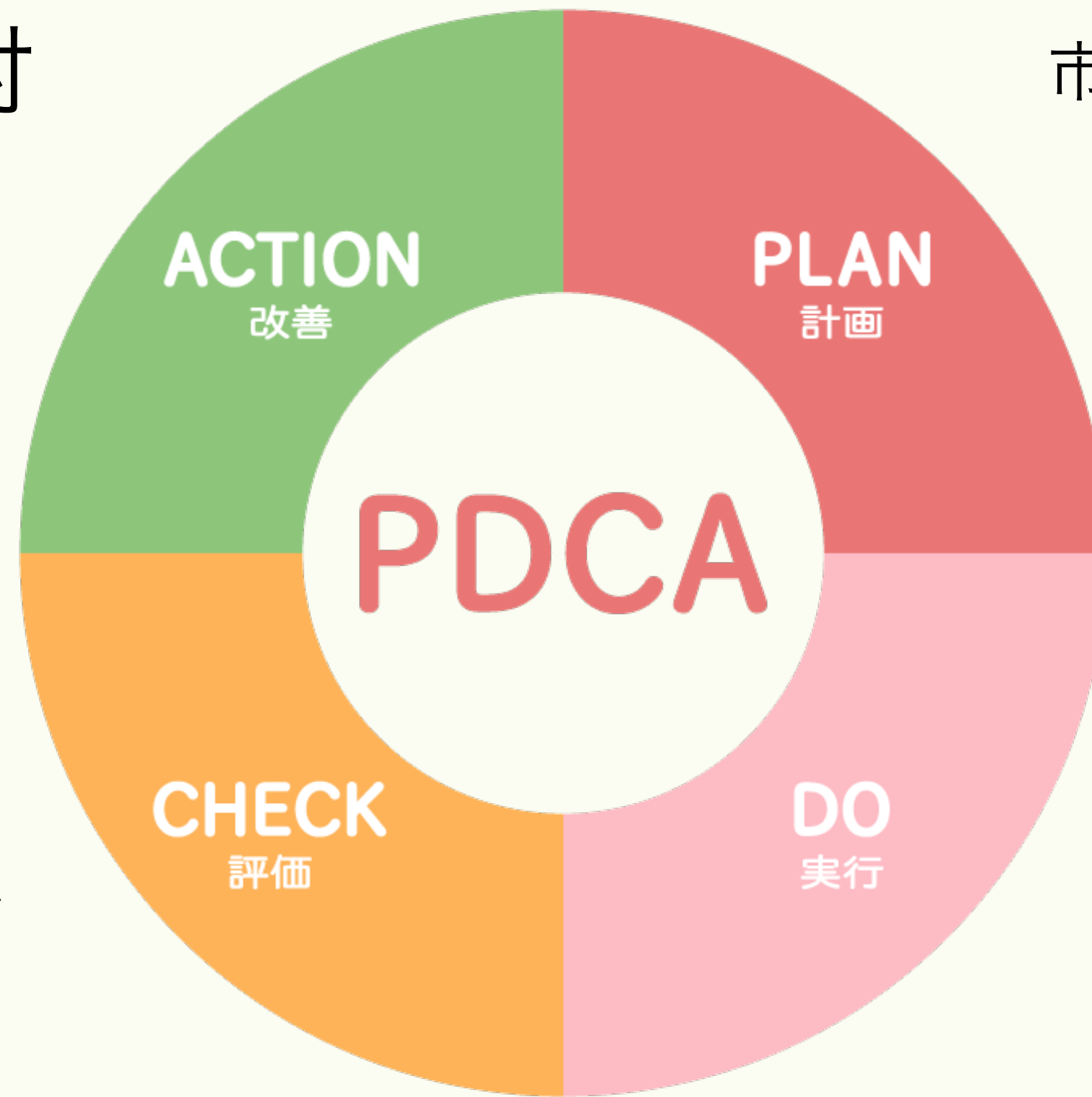
### 改善ポイント

- 自身でのアンケート実施し、データ習得
- Twitterデータを定期的に習得できる仕組みを作る

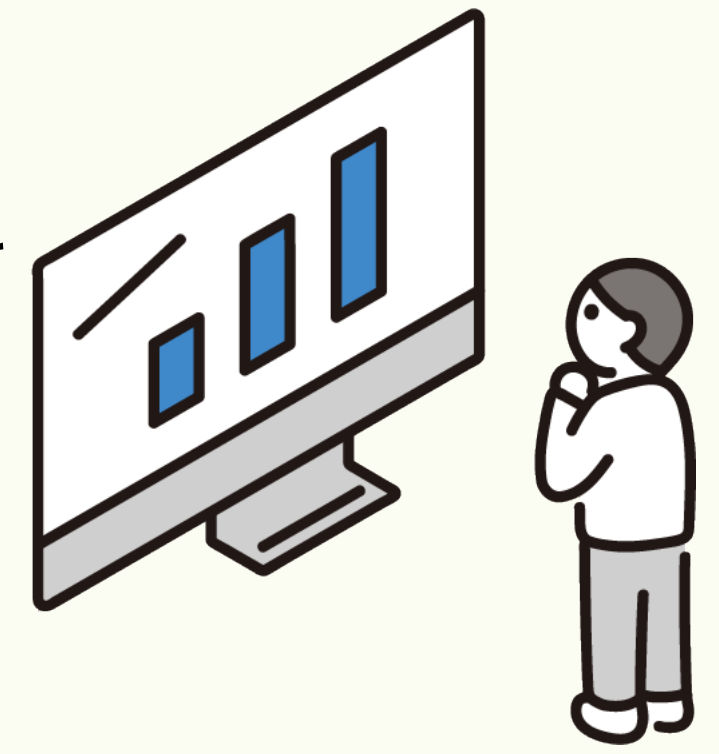
今後



施策検討



市場データも踏まえた  
計画



KPI確認

CHECK  
評価

DO  
実行

ブログ執筆



Google Analytics

<https://socc.../9759>

データ分析を取り入れながらブログ発信

# Appendix



# TF-IDF

各ツイートをTF-IDFを求めてツイートの重要キーワードを確認

```
wordcloud2= WordCloud(  
    font_path='/Library/Fonts/Arial Unicode.ttf',  
    width=800,  
    height=600,  
    background_color="white"  
).generate_from_frequencies(vecs_dic[1])  
wordcloud2.to_image()
```

	ac	ach	ad	adhd	adko	adl	af	affcs	agn	ai	...	カ マ	ノ ハ オ	ノ	バ タ	ヒ ン	フ イ ン	フ ウ	フ
前半 稼女性 財布 希望 男性 メッ 年 取 財布 オッサン 年取 イケメン	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.
妻フルタイム 働 分担 夫 沢山 妻	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.
結婚 妻 男性 分担 間最近 たて 答え 最低 定時 子供 間黙 ゃ 認識	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.
夫婦 合 里帰り 出産 最短 保育園 目 指 時短 勤務 メイン むし 当たり前 後悔	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.
繁忙期 週間 保育園 送り 夫 出社 娘 時期 生活 娘 めめ 食事 風呂 手 間 去 肌 ケア 歯磨き	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.

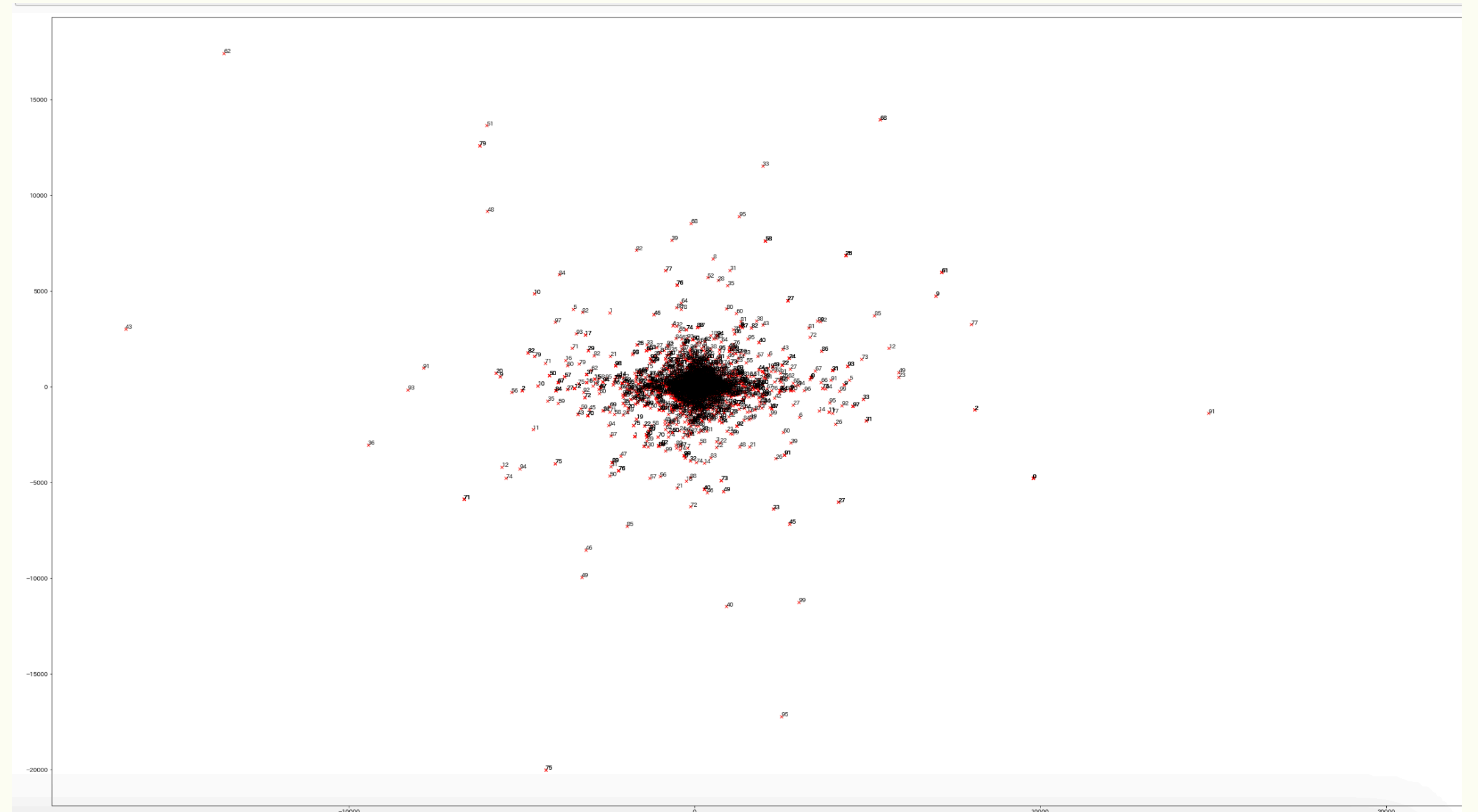
沢山 分担  
フルタイム

## Word2Vec

- それぞれのツイートの名詞・形容詞をベクトルの足し算をし特徴を把握

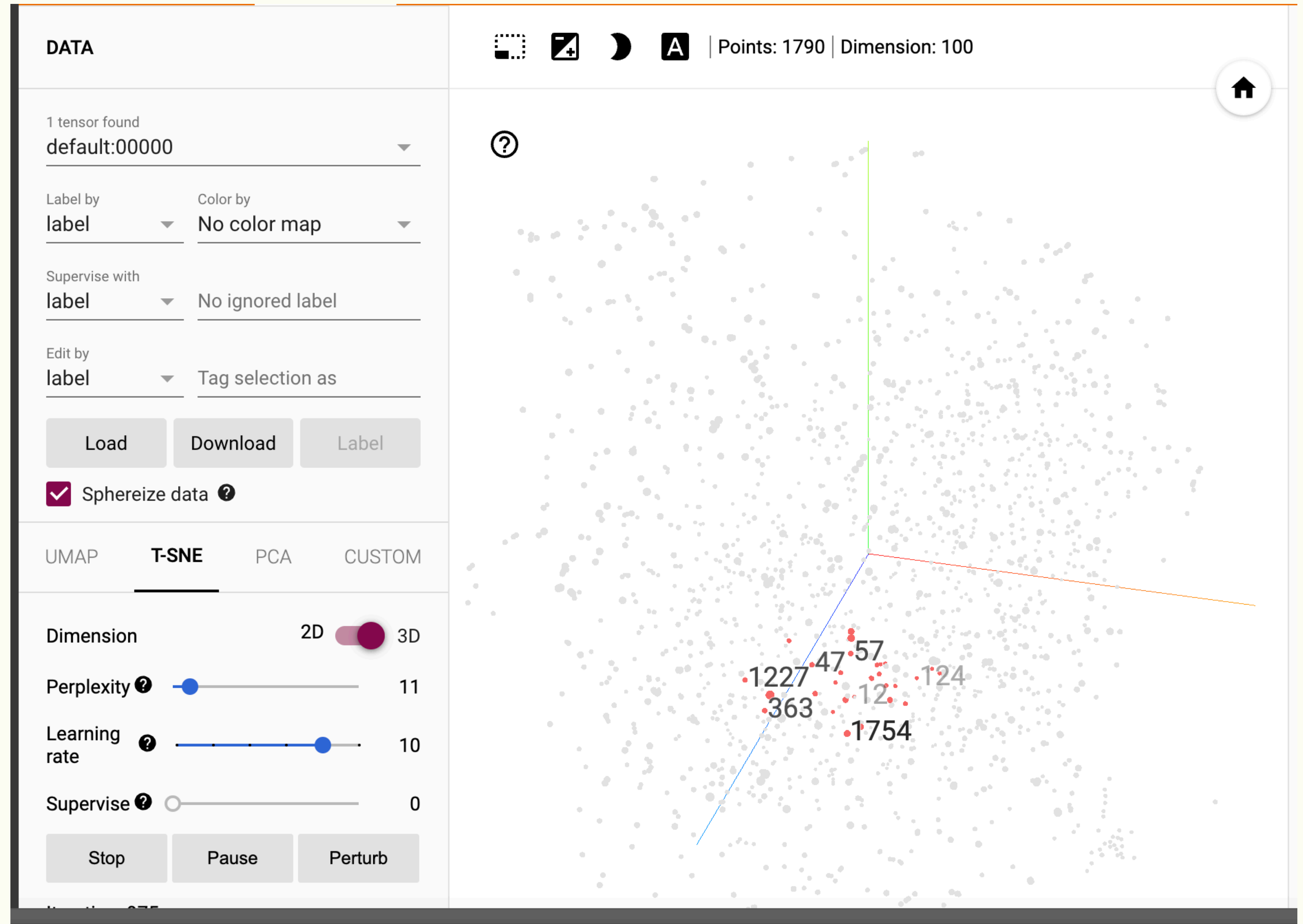
可視化のためt-SNEで2次元に圧縮

それぞれのワードベクトルの近さから特徴を捉えたかったため主成分分析ではなく、t-SNEを選択



- それぞれのツイートのフィルターをかけず、ツイートごとのベクトルを求める

## Tensor Boardで可視化



# K-meansを使ったクラスタリング

- 3つに分類

分類結果を3Dにプロットし、主成分でプロット  
主成分が同じようなツイートが多いことがわかる

