

コンペ

<users ユーザー情報>

id:ユーザーID generation_id:世代ID prefecture_id:都道府県ID created:アプリ登録日時

```
users.shape= (2935754,4)
```

generation_id missing rate= 86% 2526262/2935754

prefecture_id missing rate=67% 1963598/2935754

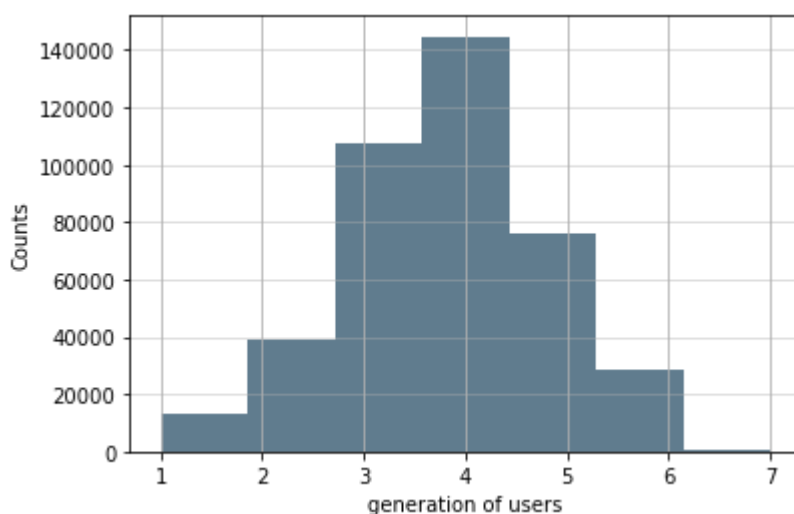
個人情報の欠落度が高いという情報は、ユーザーのアプリケーションに対する信頼度が高くないことをある程度推察できると思います。// I think the information that the lack of personal information is high can infer to a certain extent that the user's trust in the application is not high.//

```
#generation.describe() count=409492.000000
```

```
#prefecture.count()=972156
```

- **generation of users**

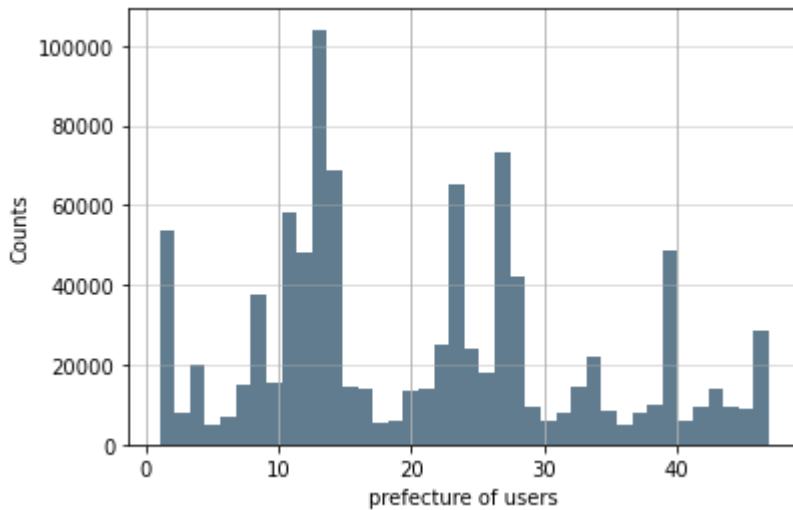
4.0	144705
3.0	107773
5.0	75788
2.0	39080
6.0	28426
1.0	13152
7.0	568



40年代のusersは多いので、日本社会の老齢化についての関係？

- **prefecture of users**

13.0	104020
27.0	73469
14.0	68753
23.0	65069
11.0	58259
12.0	48067
1.0	44744
40.0	44251
28.0	42074
22.0	25174
8.0	22270
34.0	21930
4.0	20145
26.0	18004
47.0	15998
10.0	15265
9.0	15215
7.0	15022
33.0	14399
15.0	14394
43.0	14176
21.0	13803
20.0	13657
24.0	12585
46.0	12533
25.0	11254
38.0	9810
42.0	9636
44.0	9616
29.0	9553
45.0	8963
2.0	8725
35.0	8417
37.0	8029
3.0	7925
17.0	7557
6.0	7161
16.0	6347
41.0	6044
19.0	5976
30.0	5951
18.0	5381
5.0	5165
36.0	5015
39.0	4449
32.0	4010
31.0	3896



13-東京都 27-大坂府 14-神奈川 23-愛知県 11-埼玉県

31-鳥取県 32-島根県 39-高知県 36-徳島県 5-秋田県

最も多い5つの地区と最も少ない5つの地区

Is it related to the size of the population in these areas?

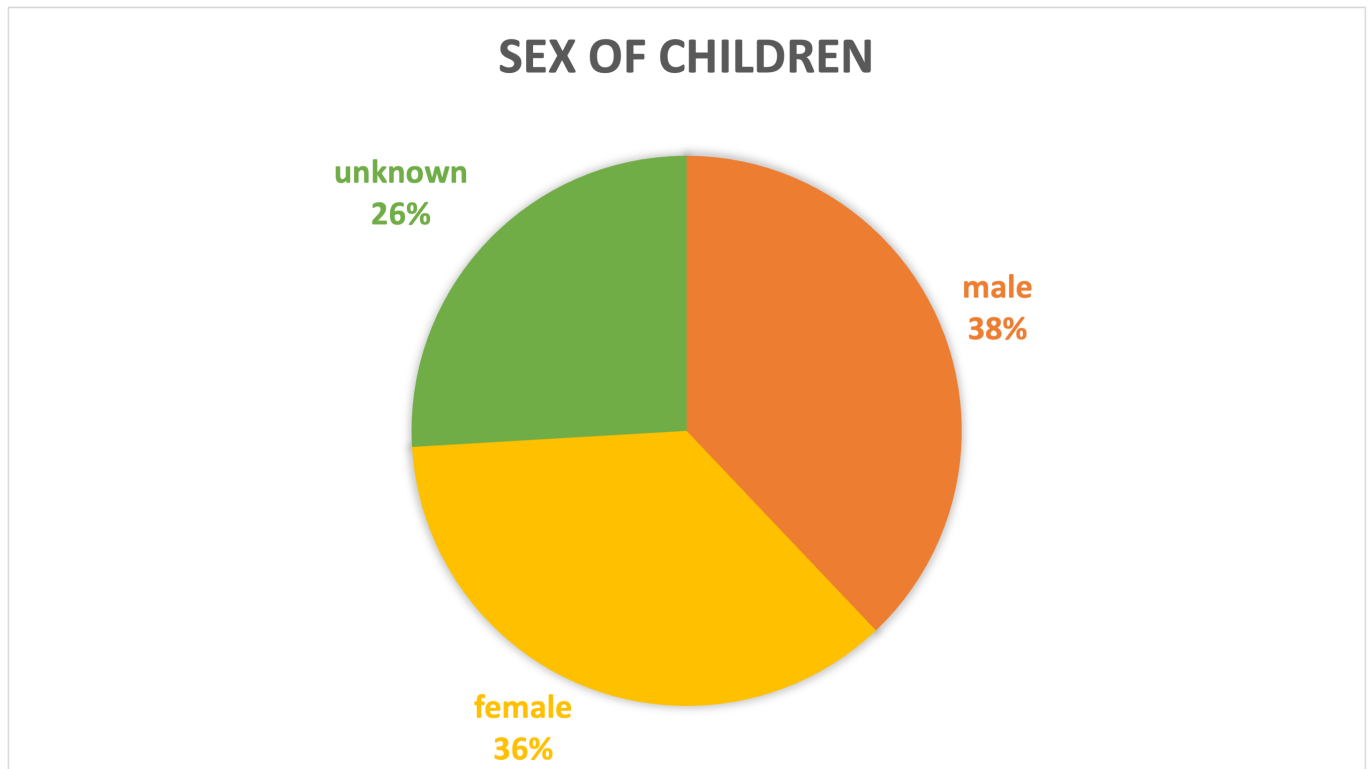
<children 子供情報>

id:子供ID user_id:紐づくユーザーID birthday:子供の誕生日 sex:子供の性別

- 性別

```
1    713041
2    679819
0    487602
Name: sex, dtype: int64
```

0-unknown487602 1-male713041 2-female679819



(male>female)

$679819/487602=1.3942088014405192$

male/female=765/728w=1.0508241758241759 (2021.4.1)

male/female=774/738=1.048780487804878 (2020.4.1)

male/female=785/747=1.0508701472556894 (2019.4.1)

(***Data from 総務省統計局 <https://www.stat.go.jp/data/jinsui/index.html>)

- 年齢

birthday transfer to age

`birthday.min(), birthday.max()`

`('1614-03-01', '2026-05-01')`

#insert new column age

`age.shape(1880462,)`

`unborn=children[(children['age'] < 0.0)] #unborn child`

26745/1880462 #Ratio of unborn children=0.014222568709178915

`ageover10=children[(children['age'] > 10.0)] #Children over 10 years old`

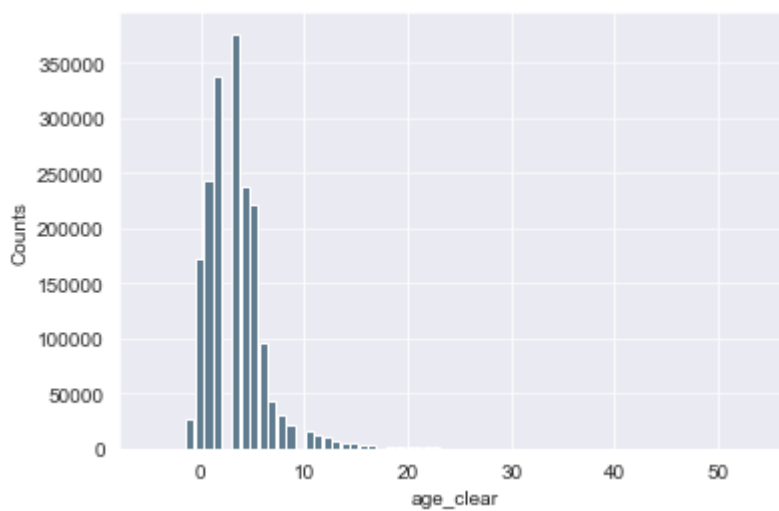
56839/1880462 Ratio of children over 10 years of age=0.03022608273924174

interval_statistic of age

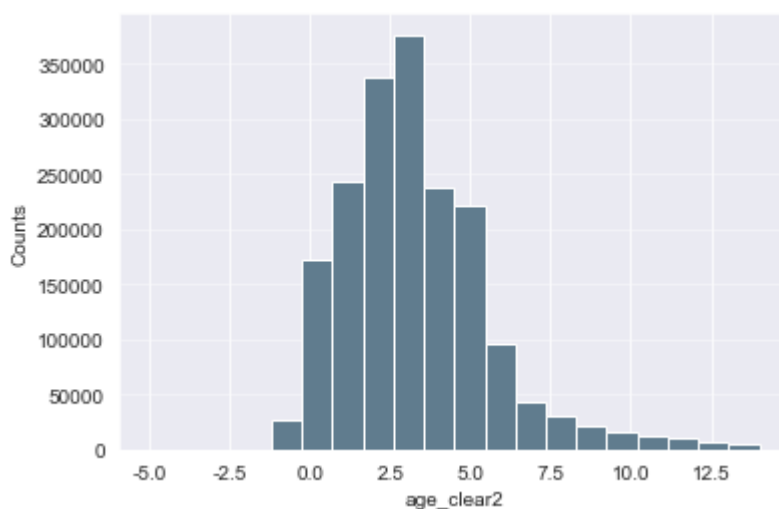
-5.000~0.000	199126	10.589%
0.000~5.000	1588528	84.475%
5.000~10.000	429185	22.823%
10.000~15.000	57393	3.052%
15.000~20.000	15039	0.800%
20.000~25.000	4916	0.261%
25.000~30.000	1488	0.079%
30.000~35.000	610	0.032%
35.000~40.000	219	0.012%
40.000~45.000	84	0.004%
45.000~50.000	25	0.001%
50.000~55.000	13	0.001%
55.000~60.000	6	0.000%
60.000~65.000	3	0.000%
65.000~70.000	5	0.000%
70.000~75.000	5	0.000%
75.000~80.000	2	0.000%
80.000~85.000	1	0.000%
85.000~90.000	0	0.000%
90.000~95.000	0	0.000%
95.000~100.000	1	0.000%
100.000~105.000	2	0.000%
105.000~110.000	2	0.000%
110.000~115.000	2	0.000%
115.000~120.000	0	0.000%
120.000~125.000	4	0.000%
125.000~130.000	0	0.000%
130.000~135.000	0	0.000%
135.000~140.000	0	0.000%
140.000~145.000	0	0.000%
145.000~150.000	0	0.000%
150.000~155.000	0	0.000%
155.000~160.000	0	0.000%
160.000~165.000	0	0.000%
165.000~170.000	0	0.000%
170.000~175.000	0	0.000%
175.000~180.000	0	0.000%
180.000~185.000	0	0.000%
185.000~190.000	0	0.000%
190.000~195.000	1	0.000%
195.000~200.000	1	0.000%
200.000~205.000	0	0.000%
205.000~210.000	0	0.000%
210.000~215.000	0	0.000%
215.000~220.000	0	0.000%
220.000~225.000	0	0.000%
225.000~230.000	0	0.000%
230.000~235.000	0	0.000%
235.000~240.000	0	0.000%
240.000~245.000	0	0.000%
245.000~250.000	0	0.000%
250.000~255.000	0	0.000%
255.000~260.000	1	0.000%

260.000~265.000	1	0.000%
265.000~270.000	0	0.000%
270.000~275.000	0	0.000%
275.000~280.000	1	0.000%
280.000~285.000	1	0.000%
285.000~290.000	0	0.000%
290.000~295.000	0	0.000%
295.000~300.000	0	0.000%
300.000~305.000	0	0.000%
305.000~310.000	0	0.000%
310.000~315.000	0	0.000%
315.000~320.000	0	0.000%
320.000~325.000	0	0.000%
325.000~330.000	1	0.000%

age<55



age<15



1508406 -5.0 count1. →As invalid data

5<age<-1 count0. →As invalid data

pregnancy period: -1<age<0 count 26745

0<age<1 count 172380

Effective age range : from -1 to 15

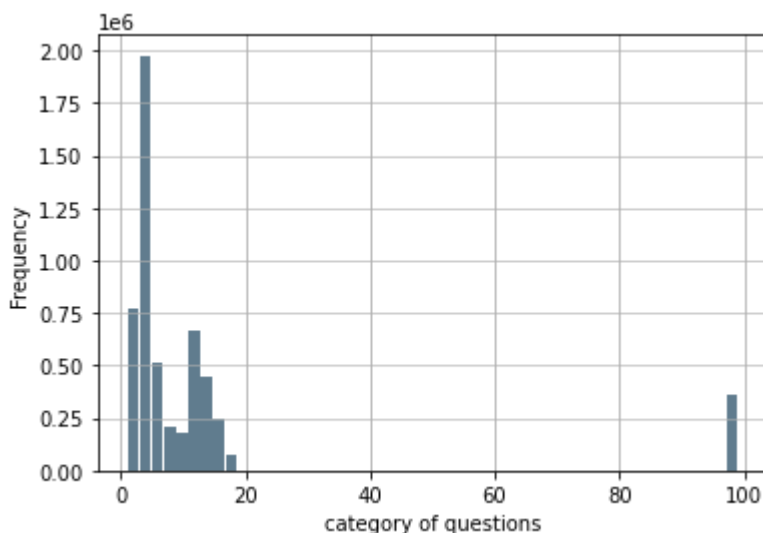
*こどもの数（15歳未満人口） from 総務省統計局

<questions 質問>

id:質問ID user_id:質問したユーザーID category_id:質問のカテゴリID content:質問本文 created:質問日時

category id

[(3, 1874721), (1, 768923), (11, 529749), (99, 362606), (5, 328195), (14, 311963), (7, 209871), (6, 184784), (9, 182860), (15, 140269), (12, 135019), (13, 133031), (16, 109876), (4, 101349), (18, 78370)]



Three level:

level 1:

The most concerned question is 3, which is 子育て・グッズ Parenting / Goods （圧倒的な数の質問！）

then 1:妊娠・出産Pregnancy / childbirth #then 11: 雑談・つぶやきChat / Tweet

level 2:

99:その他の疑問, 5:ココロ・悩み, 14:家族・旦那 Has a close number of questions

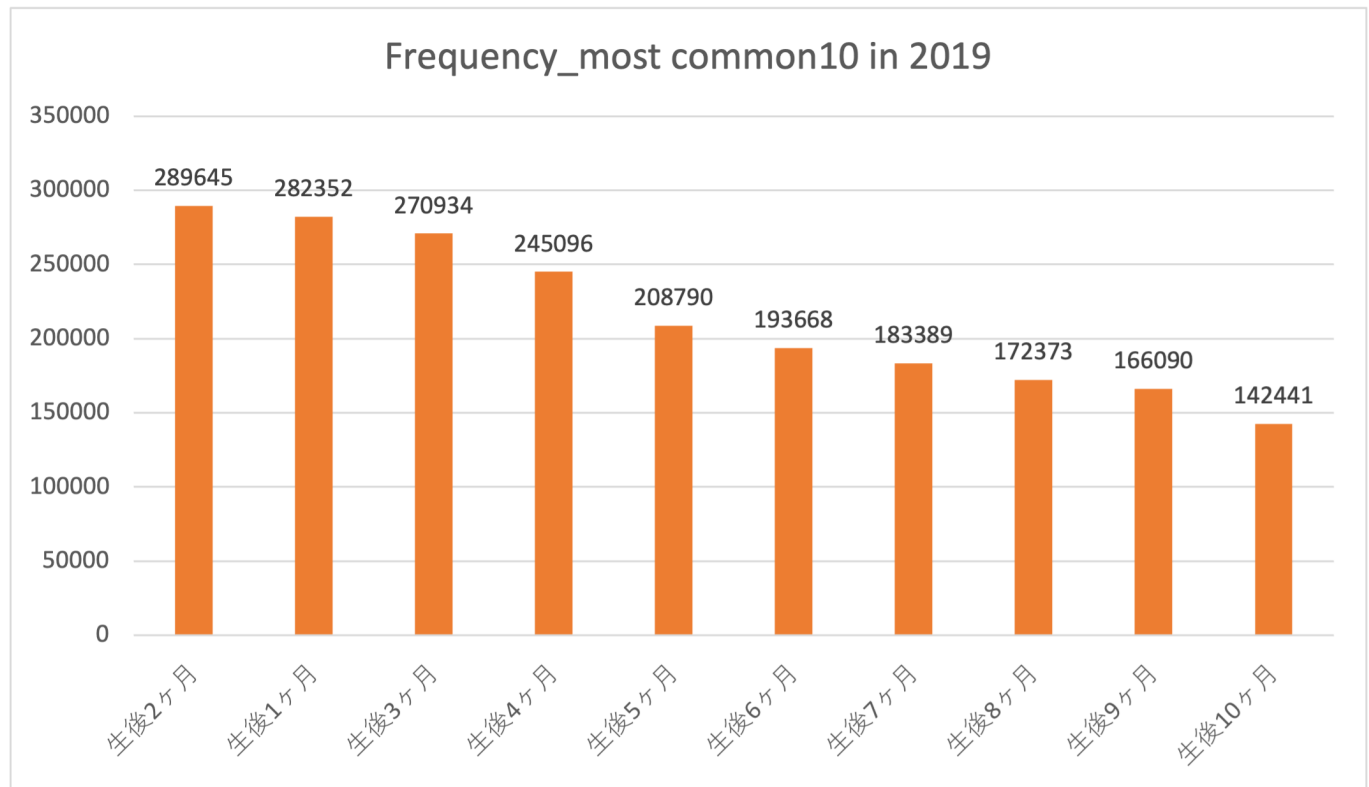
level 3:

Remaining category: 7家事・料理,6妊活Pregnancy,9お金・保険,15お出かけgo out?,12お仕事,13ファッション・コスメ,16産婦人科・小児科,4サプリSupplement・健康,18住まい

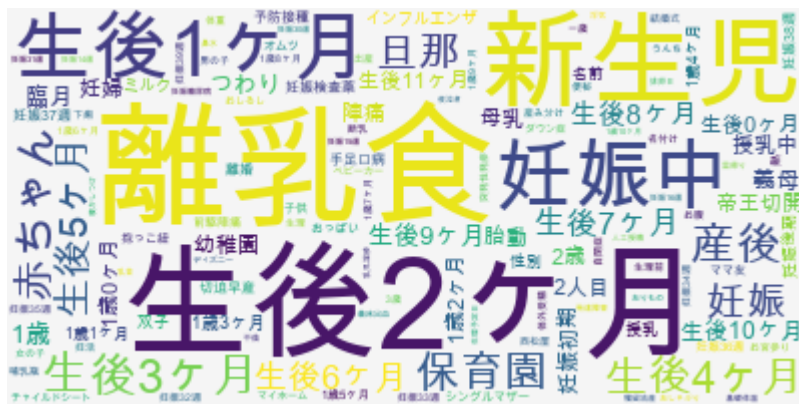
<search 検索>

word_2019

[('生後2ヶ月', 289645), ('生後1ヶ月', 282352), ('生後3ヶ月', 270934), ('生後4ヶ月', 245096), ('生後5ヶ月', 208790), ('生後6ヶ月', 193668), ('生後7ヶ月', 183389), ('生後8ヶ月', 172373), ('生後9ヶ月', 166090), ('生後10ヶ月', 142441)]



WordCloud of 2019



```
fdist1.hapaxes()[10] #Return a list of all samples that occur once 2019
```

```
[ '6ヶ月 発熱 鼻水',  
  '1歳3ヶ月 出す',
```



```
'義理 弟 お年玉',
'1歳半 嫉妬',
'結膜炎 7ヶ月',
'凍結胚盤胞6日目 性別',
'野崎クリニック 産後2週間検診',
'創価学会とは何を信仰しているの',
'横浜ベイクォーター 年始',
'bcg 消えない']
```

exp. of length of words(2019):

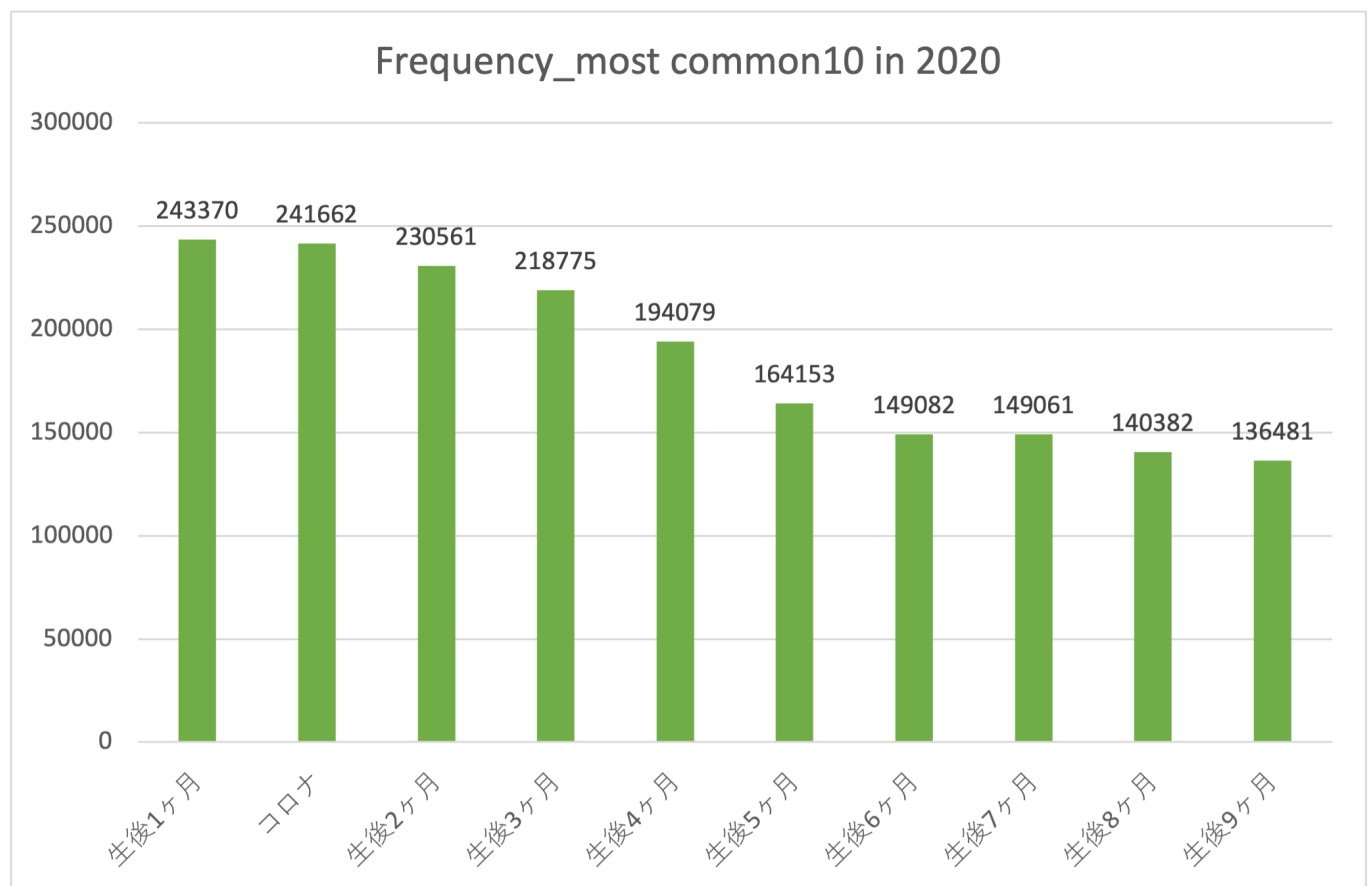
```
V=set(word2019)
print([w for w in V if len(w)==8 and fdist1[w]>20000])

['産後 ダイエット', '離乳食 いつから', 'チャイルドシート', '離乳食 食べない']
```

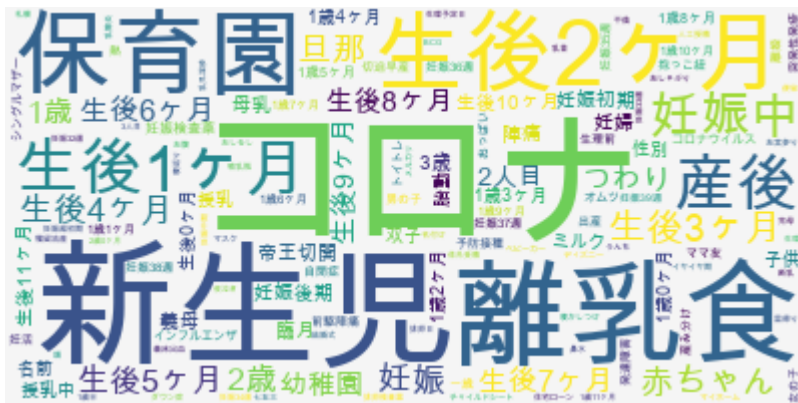
2019年searchのwordcloudの中で離乳食がとても目立ちます。

word_2020

```
[('生後1ヶ月', 243370), ('コロナ', 241662), ('生後2ヶ月', 230561), ('生後3ヶ月',
218775), ('生後4ヶ月', 194079), ('生後5ヶ月', 164153), ('生後6ヶ月', 149082), ('生後7
ヶ月', 149061), ('生後8ヶ月', 140382), ('生後9ヶ月', 136481)]
```



WordCloud of 2020



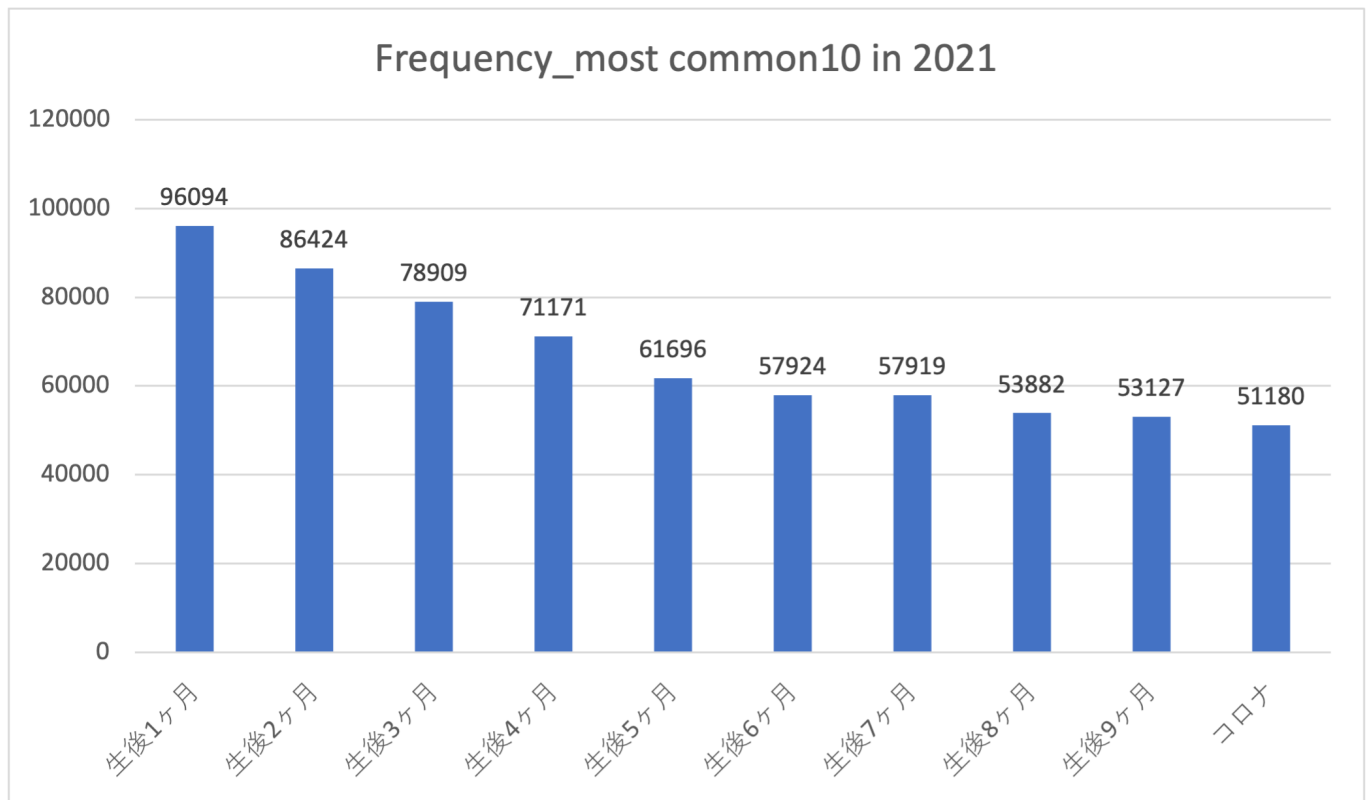
['10連休',
'12週体重',
'1歳 マウスピース',
'ミルク 水筒 熱湯消毒',
'足立クリニック 小倉',
'足立クリニック 小倉 里帰り',
'暴力 エスカレート',
'2歳の我が子を亡くしもうすぐで半年。 未だにまだ毎日居るかのように、見渡してしまったり
タオル一枚見ても思い出す毎日。。',
'寝かしつ 遅い',
'発熱 妊娠超初期']

```
V=set(word2020)
print([w for w in V if len(w)==3 and fdist2[w]>200000])

['ヨリナ']
```

word2021

```
[('生後1ヶ月', 96094), ('生後2ヶ月', 86424), ('生後3ヶ月', 78909), ('生後4ヶ月', 71171), ('生後5ヶ月', 61696), ('生後6ヶ月', 57924), ('生後7ヶ月', 57919), ('生後8ヶ月', 53882), ('生後9ヶ月', 53127), ('コロナ', 51180)]
```



WordCloud of 2021



```
fdist3.hapaxes()[1:10] #Return a list of all samples that occur once 2021
```

```
[ '裁縫作業台',
  '元日 夜ご飯',
  '年末 つわり',
  '子どもの前で泣いてしまった',
  '中絶 火葬届',
  '女 翠',
  'エコー 女の子 19w',
  'マンマ 後追い',
  '離婚 話し合いできない',
  '認定日 面談']
```

exp. of length=87 & freq>1 of words(2021):

(*fdist2021.tabulate() After checking this item, the comment with a length of 91 is the least frequently seen, with only 3 occurrences.)

```
V=set(word2021)
print([w for w in V if len(w)==91 and fdist3[w]>=0])
```

['大和屋さんから出ているすくすくチェアプラスかすくすくチェアスリムプラスのどちらにするかで悩んでいます！ どちらかを購入された方は決め手は何でしたか？宜しければ教えていただきたいです！',
'昨日質問させてもらいました。 今日、仲良し中に子供が泣いたら一旦やめて欲しいことを伝えましたが、無理だと言われました。泣いたところで死ぬわけじゃないんだから、別に泣かせとけばいいと。',
'2人目は何歳差で、ご出産されましたか？☺ いいねで回答お願いします♪ 何歳差で産んでも、大変なこと、楽だったりすることを 聞くのですが、、、みなさんの体験談お聞きしたいです☺']