

テキストマイニングを用いた  
HOT PEPPER Beautyクチコミデータの分析

---

東海大学 濱屋亜依香

# 目次

---

1.研究背景

2.先行研究・研究目的

3.データ概要

4.HOT PEPPER Beautyとは

5.分析使用ソフト概要

6.基礎集計

6.1.基礎集計まとめ・分析手順

7.3.テキストマイニング

7.3.1. 評判抽出

7.3.2. 単語頻度解析

7.3.3. 特徴表現抽出

8.まとめ・提案

9.今後の課題

参考文献

Appendix

# 1.研究背景

## ❖ 美容院・美容師数推移

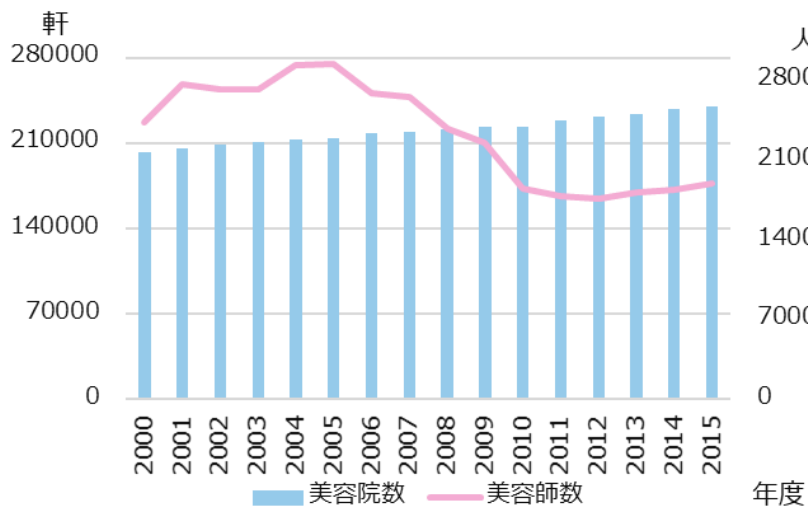


図1.美容院・美容師数推移<sup>[1][2]</sup>

## ❖ 美理容市場規模推移

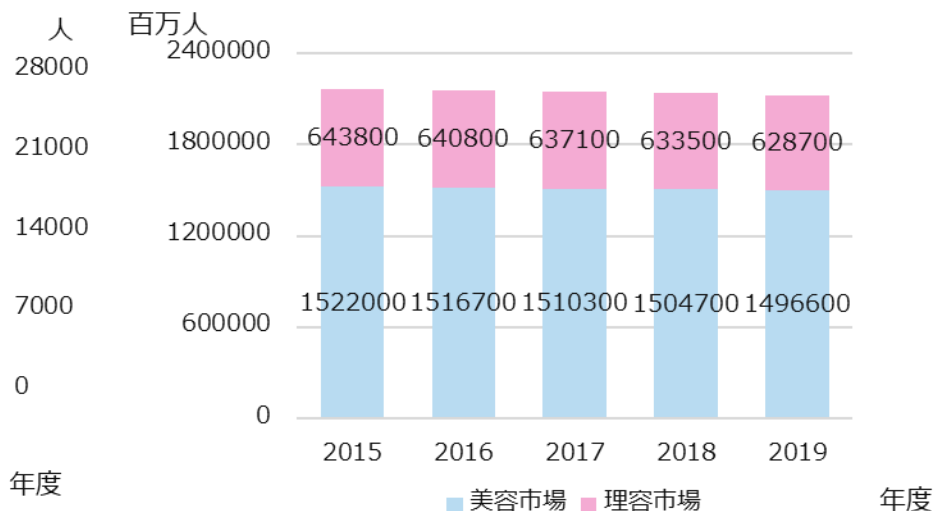


図2. 美理容市場規模推移<sup>[3]</sup>

- 美容院数は増加傾向にある
- 美容師数は2004～2005年にピークを迎え、2012年から増加

- 美容市場は理容市場の約2.5倍の規模
- 美理容師市場は年々減少

※2010年度宮城県は仙台市以外の市町村,福島県は相双保険福祉事務所の管轄内の市町村が含まれていない

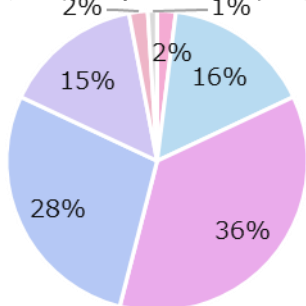
**美容院数・美容師数は増加しているが美理容市場の規模は減少している**

# 1.研究背景

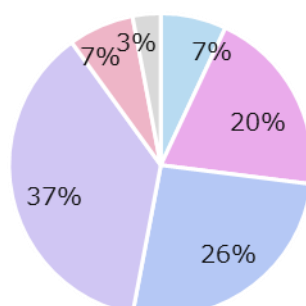
## ❖ 緊急事態宣言解除後の美容室利用に関するアンケート調査

Q.美容室へ行く頻度はどのように変化したいと思いますか？

< ビフォーコロナ >



< ウィズコロナ >



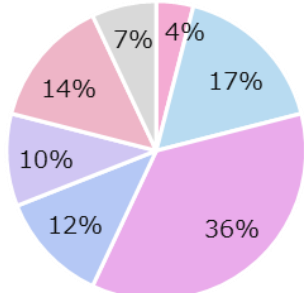
- 月に2回以上
- 1ヶ月に1回程度
- 2ヶ月に1回程度
- 3ヶ月に1回程度
- 4~6ヶ月に1回程度
- 7~12ヶ月に1回程度
- 数年に1回程度

年間6.1回/一人  
から  
年間4.1回/一人  
に減少

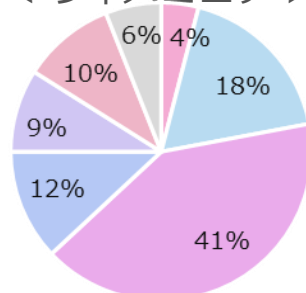
図3.4. 美容室へ行く頻度 (ビフォーコロナ・ウィズコロナ) [4]

Q.美容室での平均利用金額はどのように変化したいと思いますか？

< ビフォーコロナ >



< ウィズコロナ >



- 1000円以下
- 2000~3000円代
- 4000~5000円代
- 6000~7000円代
- 8000~9000円代
- 10000~11000円代
- 12000円以上

6370円/一人一回  
から  
5919円/一人一回  
に減少

図5.6. 美容室で平均利用額 (ビフォーコロナ・ウィズコロナ) [4]

**コロナ前に比べ美容室に行く頻度,利用金額ともに減少している**

# 1.研究背景

## ❖ 初回来店時の美容室選びで重要視すること(抜粋)

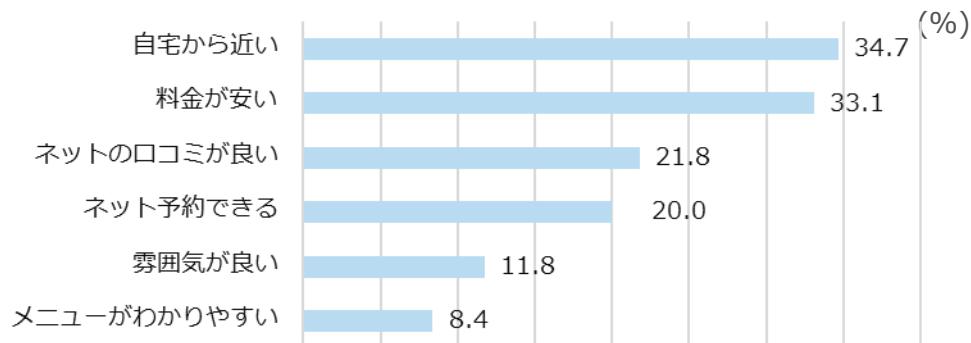


図7.初回来店時にサロン選びで重要視すること<sup>[5]</sup>  
(美容室利用者/MA/ 15~69歳女性/n=5584)

- 自宅からの距離, 料金が重要
- 5人に1人は口コミを見てから来店している

## ❖ 現在利用している美容室の継続理由(抜粋)

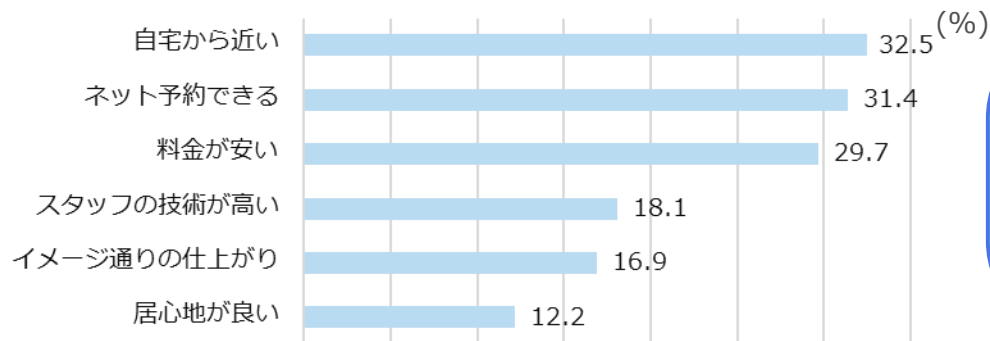


図8.現在利用しているサロンの継続理由<sup>[5]</sup>  
(美容室利用者/MA/ 15~69歳女性/n=5584)

- 自宅からの距離, 料金に加えネット予約が重要
- 技術や仕上がりによる満足度が低い?

どちらも自宅から近いこと以外は自店舗で改善・向上することができる

## 2. 先行研究・研究目的

### 研究背景・課題

- 美容室の新規・継続利用の理由には「自宅から近い」「料金」が主になっている
- コロナウイルスの影響により来店頻度や利用料金は減少していて美容市場規模はさらに減少してしまう

### 先行研究

正岡・二宮<sup>[6]</sup>

- 若年層(20代女性)の美容室の利用状況を調査(大阪市内)
- 新規・継続利用促進戦略を提案している

### 研究目的

本研究では口コミのデータを用いて地域や年代によってどのような特徴があるのかを探り、客足・利用料金を増やすために自店舗で改善・向上できる案を提示することを目的とする

# 3.データ概要

提供	株式会社リクルートテクノロジーズ
期間	2012年1月～2014年1月
概要	ホットペッパービューティーデータ

## 店舗データ (約8万件)

- 緯度経度
- 住所

## スタイリストデータ (約18万件)

- キャリア年数
- 性別

## セットメニューデータ (約7万件)

- メニュー名
- 料金
- 施術時間

## 口コミデータ (約73万件)

- 性別
- 世代
- レビュー内容
- ユーザー評価
- 総合評価

## メニューデータ (約52万件)

- メニュー名
- 料金
- 施術時間

※各々のデータを店舗IDによる紐づけを行う  
(73436件)

- ・ +や&を含むセットメニュー
- ・ 1100円以下, 39800円以上は除外

## 4.HOT PEPPER Beautyとは

### ❖ サロン利用者向け 概要<sup>[7]</sup>



◎国内最大級のヘア、リラク&ビューティーサロンの検索・予約サイト

- 株式会社リクルートライフスタイルさんが運営
- 行きたい日時での検索, 信頼できる口コミ, お得なクーポンなどがあり自分の行きたいサロンを探ることができる

### ❖ 掲載希望サロン向け 概要<sup>[7]</sup>



◎様々なサービスやプランでお店のサポートを行う

- 店舗に合わせた幅広い価格・内容のプランがあり費用に合わせた集客効果が見込める
- ネット予約の自動スケジューリングや, 予約内容をレジに接続できる「サロンボード」を活用し店舗の運用の効率化する



## 5.分析使用ソフト概要

### ❖ Visual Mining Studio (VMStudio)

データの前処理から、マイニング処理、他アプリケーションとの連携機能を備え、さらにその結果をグラフィカルな表示で表現することができるツール

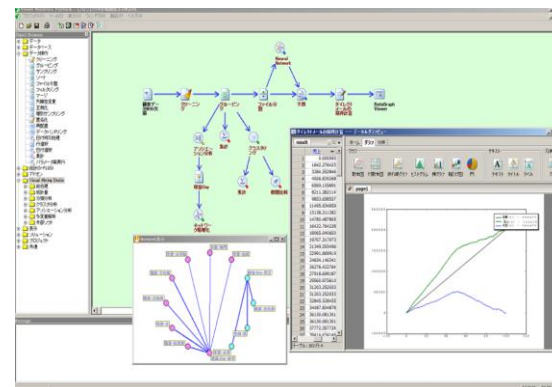


図9. VMStudio 例[8]

### ❖ Text Mining Studio (TMStudio)

頻度の分析から、テキストに付随する属性を活かした特徴分析、話題分析など豊富な分析機能を備え、さらにその結果をグラフィカルな表示で表現することができるツール

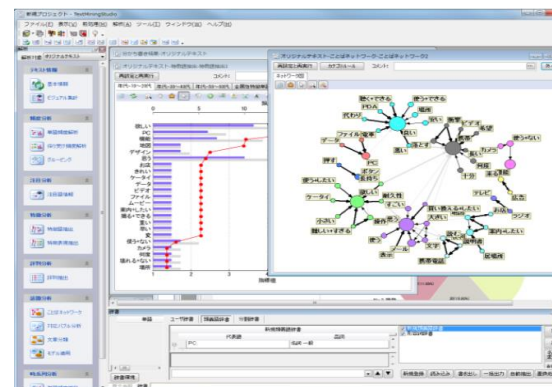


図10. TMStudio 例[8]

※どちらもNTTデータ数理システムさんの製品

# 6.基礎集計

## ❖ 口コミ 性別割合

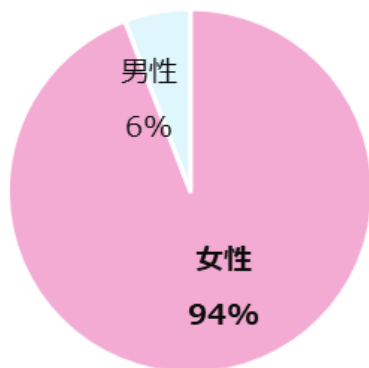


図11. 口コミ 性別割合

## ❖ 口コミ 年代割合

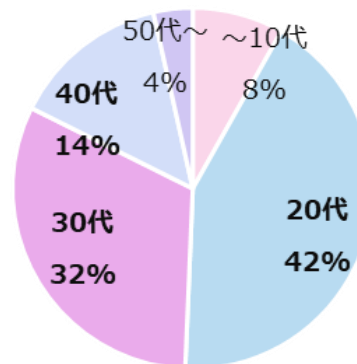


図12. 口コミ 年代割合

## ❖ 美容院地域割合

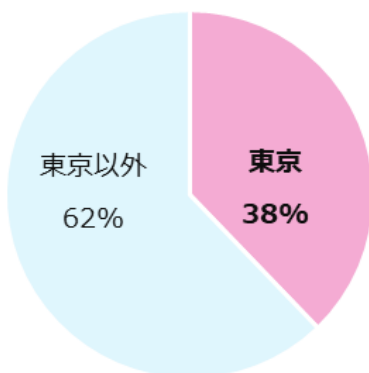


図13. 美容院 地域割合

性別を「女性」、  
美容店の住所が「東京」に  
データを絞り(26132件)  
年代ごとに比較する

※以後表記がないグラフでは東京かつ女性のみでのデータとする

## 6.基礎集計

### ❖ 年代別 使用料金割合

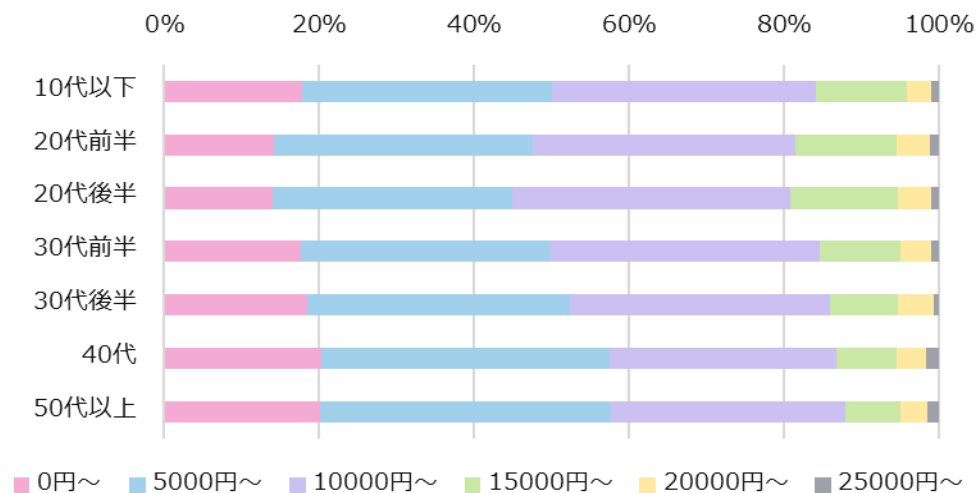


図14. 年代別 使用料金割合

- 20代後半が一番多くお金をかけている
- 30代以上の約半数は美容室にかけるお金が10000円未満となり減少している

**美容にかけるお金は20代後半になるまで増加しているが、  
それ以降は減少している**

# 6.基礎集計

## ❖ all「5」評価割合

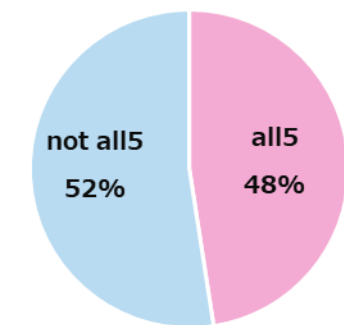


図15.all「5」評価割合

※ムード・サービス・テクニック・メニュー・総合の評価をすべて5の人を「all5」としそれ以外を「not all5」とする

## ❖ not all「5」各評価件数

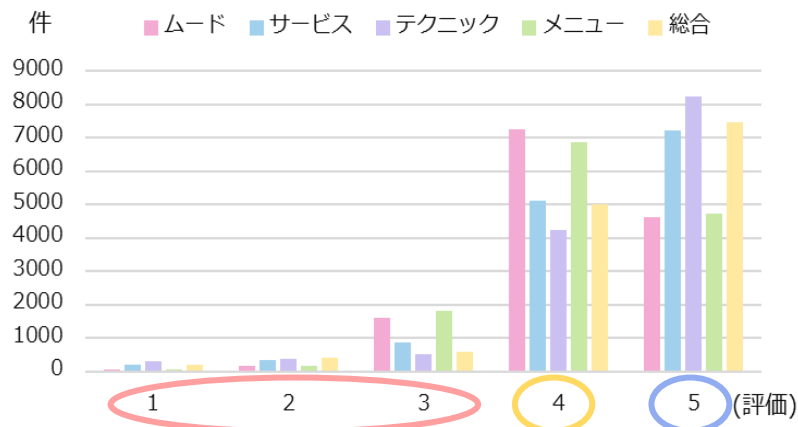


図16.not all5 各評価件数

all5と答えた人を除く

▪ 全て満足している人は約半数いる

▪ ムードとメニューの「5」評価の割合は下がるが満足している人が多い

評価が1・2・3は「不満」, 4は「やや不満」, 5は「満足」に分けて考える

※スライドp.16クラスター数に利用

# 6.基礎集計

※1,2,3⇒不満 4⇒やや不満 5⇒満足

## ❖ 評価 年齢ごと割合

### <ムード>

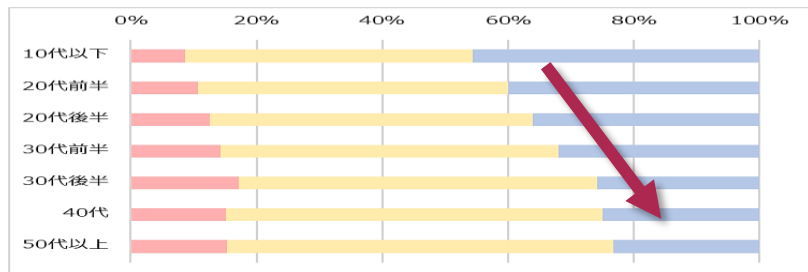


図17. 評価 年齢ごと割合 <ムード>

### <テクニック>

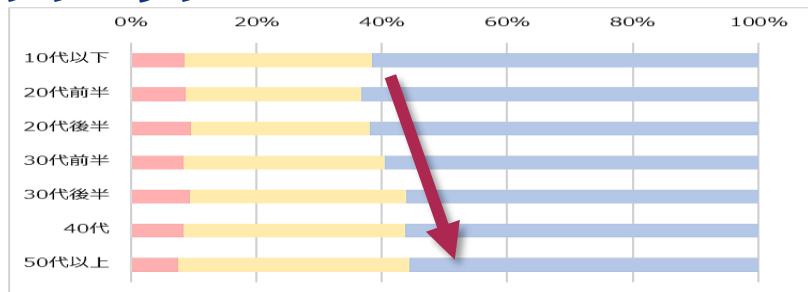


図19. 評価 年齢ごと割合 <テクニック>

### <サービス>

■ 不満 ■ やや不満 ■ 満足

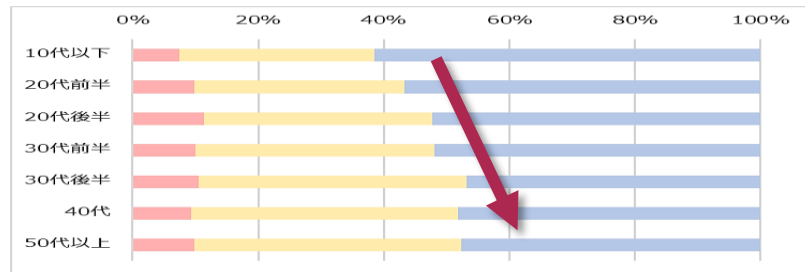


図18. 評価 年齢ごと割合 <サービス>

### <メニュー>

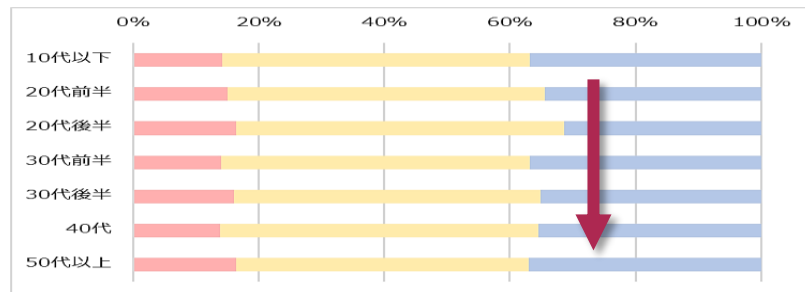


図20. 評価 年齢ごと割合 <メニュー>

### <総合>

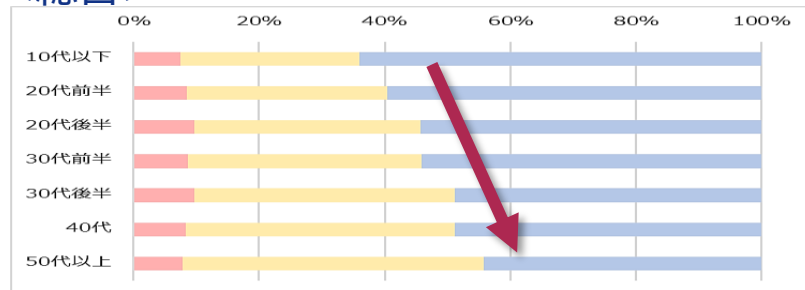


図21. 評価 年齢ごと割合 <ムード>

**年齢が上がると  
満足しなくなる割合が増加  
(メニューを除く)**

## 6.1.基礎集計まとめ・分析手順

### まとめ

- 美容にかけるお金は20代後半になるまで増加しているが、30代からは減少している
- 年齢が上がると満足しなくなる割合が増加する

### 分析手順

- ① 相関分析：各々の評価がどれくらい関係があるかを見る
- ② クラスタ分析：基礎集計をもとに3つのクラスターに分け、それぞれにどのような特徴があるかを見る
- ③ テキストマイニング：②の結果から大きな不満を持つ人がどの地域にいてどのような傾向にあるかを見る

# 7.1. 相関分析

## ❖ 相関分析

<全体>

表1. 相関分析 (全体)

Correlation.TitleName	ムードポイント	サービスポイント	テクニックポイント	メニューポイント	総合ポイント
ムードポイント	1.00	0.62	0.48	0.41	0.60
サービスポイント	0.62	1.00	0.63	0.47	0.75
テクニックポイント	0.48	0.63	1.00	0.51	0.82
メニューポイント	0.41	0.47	0.51	1.00	0.60
総合ポイント	0.60	0.75	0.82	0.60	1.00

どちらも相関アリ

<not all 5>

表2. 相関分析 (not all 5)

Correlation.TitleName	ムードポイント	サービスポイント	テクニックポイント	メニューポイント	総合ポイント
ムードポイント	1.00	0.48	0.31	0.06	0.46
サービスポイント	0.48	1.00	0.54	0.27	0.69
テクニックポイント	0.31	0.54	1.00	0.37	0.78
メニューポイント	0.06	0.27	0.37	1.00	0.46
総合ポイント	0.46	0.69	0.78	0.46	1.00

特にサービスとテクニックが総合に大きく関係している

## 7.2. クラスタ分析

### ❖ クラスタ分析

初期クラスターの設定方法：k-means法  
距離の計測方法：Euclid法  
クラスターの数：3

<not all 5>

表3. クラスタ分析 (not all 5)

id	size	ムードポイント	サービスポイント	テクニックポイント	メニューポイント	総合ポイント	residual
1	7828	4.45	4.80	4.87	4.40	4.92	7564.30
2	4926	3.96	4.05	4.20	4.01	4.00	5574.26
3	945	3.23	2.57	2.08	3.11	2.17	1889.49

#### ID 1

- ムードとメニューに不満
- 10代以下, 20代前半が多い

#### ID 2

- 全体的に「4」の評価でやや不満
- 30代後半, 40代, 50代以上が多い

#### ID 3

- 全体的に大きく不満を持ち, 特にサービスとテクニックに不満
- 20代後半が多い

一番不満があるID3, 割合の多い20代後半と30代前半を比較する



## 7.3.テキストマイニング

### ❖ 評判抽出

- ◎ 単語に対して好意的・非好意的表現それぞれで語られた回数をカウントし、それをもとに好評語・不評語のランキングを作成する

### ❖ 単語頻度解析

- ◎ 文章に現れる単語の出現回数をカウントし、表やグラフに表わす解析

ピンク : positive  
青 : negative とする

### ❖ 特徴表現抽出

- ◎ データに付随する属性ごと、特徴的に出現する係り受け表現を抽出する

属性毎に抽出された特徴語を表し、結果は**指標値**の大きい順に表示される  
「属性頻度」 : その属性内における単語の出現頻度  
「全体頻度」 : テキスト全体にわたってのその単語の出現頻度

<抽出設定>

抽出指標 : 補完類似度

頻度 : 2以上

品詞設定 : 名詞 - 形容詞・形容動詞

# 7.3.1.テキストマイニング 評判抽出

## ✦評判抽出

### <好評語ランキング>

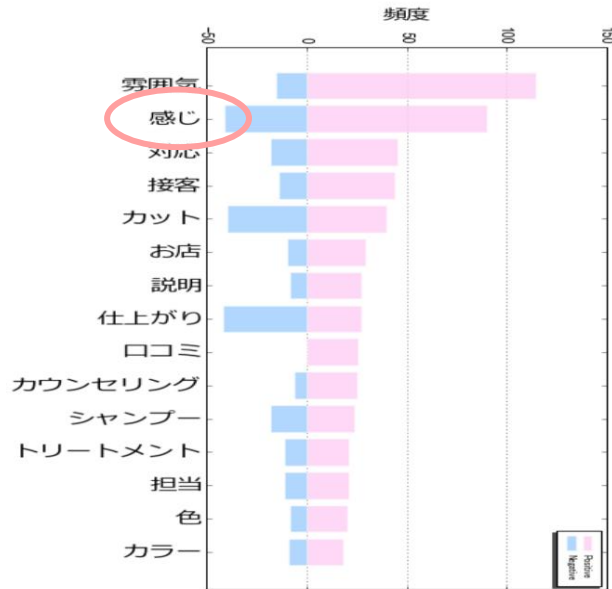


図22.好評語ランキング (id3)

### <不評語ランキング>

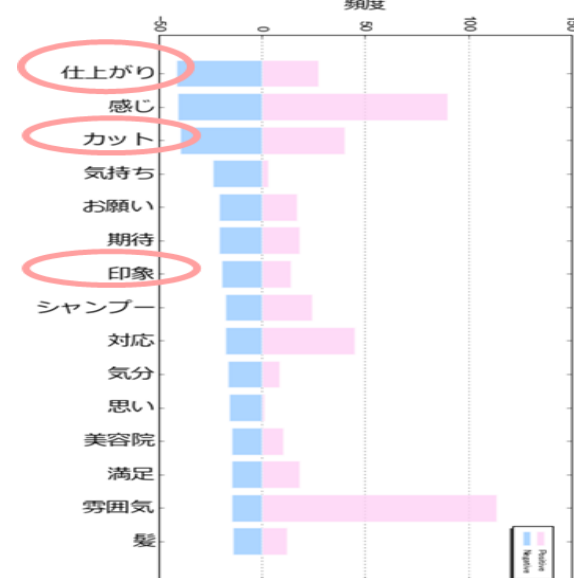


図23. 不評語ランキング (id3)

雰囲気やお店の「ムード」は好評であるが仕上がりや  
カットの「テクニック」は不評である⇒5段階評価と同じ傾向

# 7.3.2.テキストマイニング 単語頻度解析

## ❖ 単語頻度解析

<住所 出現回数 id3>

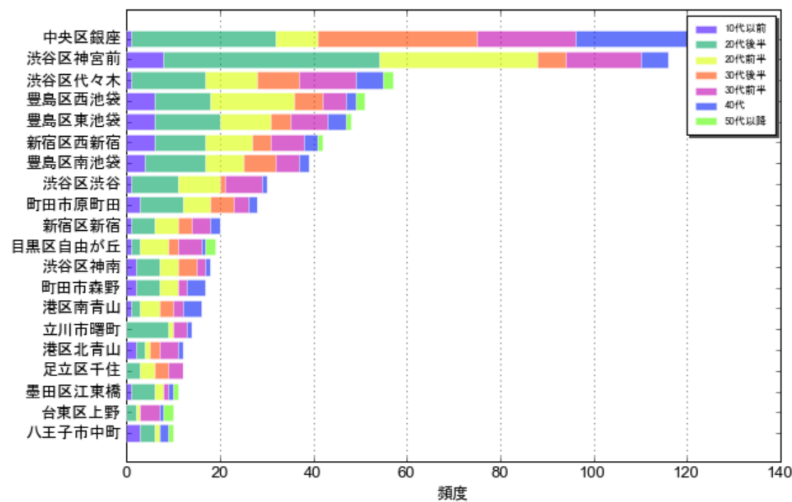


図24.単語頻度解析 (id3)

- 銀座，神宮前，池袋の値が特に多い

<住所 出現回数 all5>

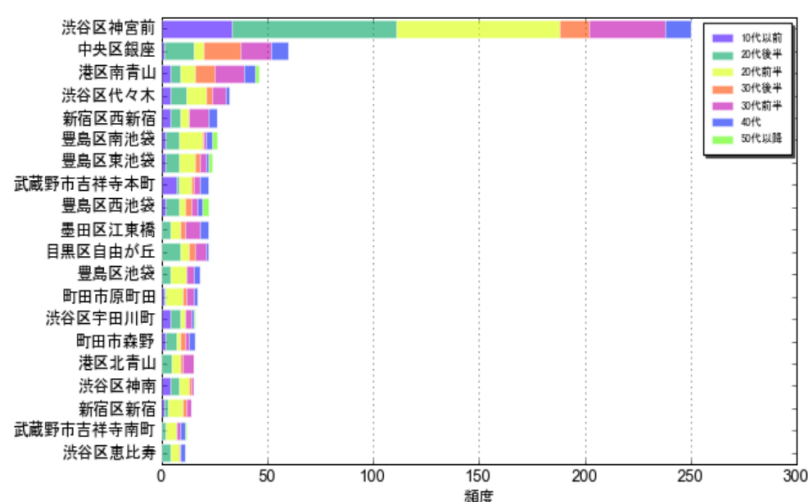


図25.単語頻度解析 (all5)

- 神宮前の値が特に多い
- 青山や吉祥寺がid3に比べ多い

神宮前の値はどちらも多いが，出現回数は違う傾向にある

## 7.3.3.テキストマイニング 特徴表現抽出

### ◆特徴表現抽出

<住所頻度 20代後半 >

表4.住所頻度 20代後半 比較

	id 3	all 5
1	神宮前	神宮前
2	東池袋	自由が丘
3	南池袋	西池袋, 青山

※属性頻度の多い順

<住所頻度 30代前半 >

表5.住所頻度 30代前半 比較

	id 3	all 5
1	代々木	西池袋
2	渋谷	上大崎
3	自由が丘	東・南池袋

※属性頻度の多い順

- 渋谷区は20代後半, 30代前半どちらも評価が低くなる
- 西池袋は20代後半, 30代前半どちらも評価が高くなる
- 東・南池袋と自由が丘の評価が逆になっている

20代後半から30代前半になると同じ地域でも評価が真逆になる

# 7.3.3.テキストマイニング 特徴表現抽出

## ◆特徴表現抽出

<レビュー評価 20代後半>

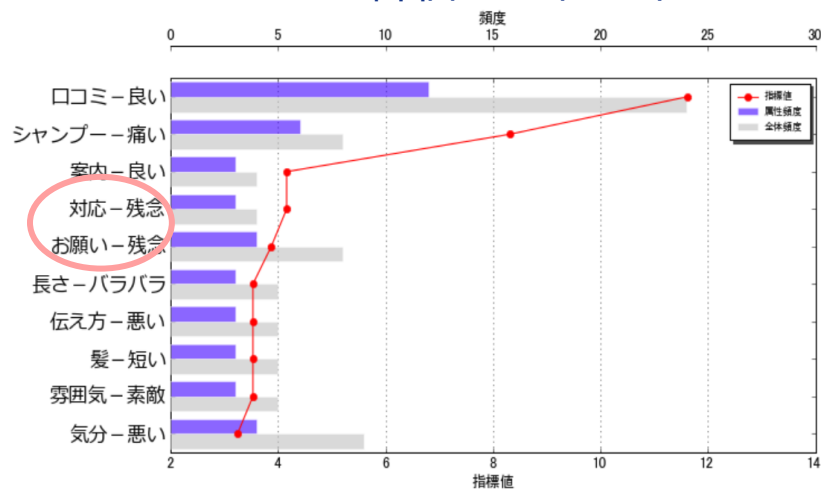


図26.特徴表現抽出 (id3)

- お店のクチコミは良い
- 店員のサービス, 技術に不満

<レビュー評価 30代前半>

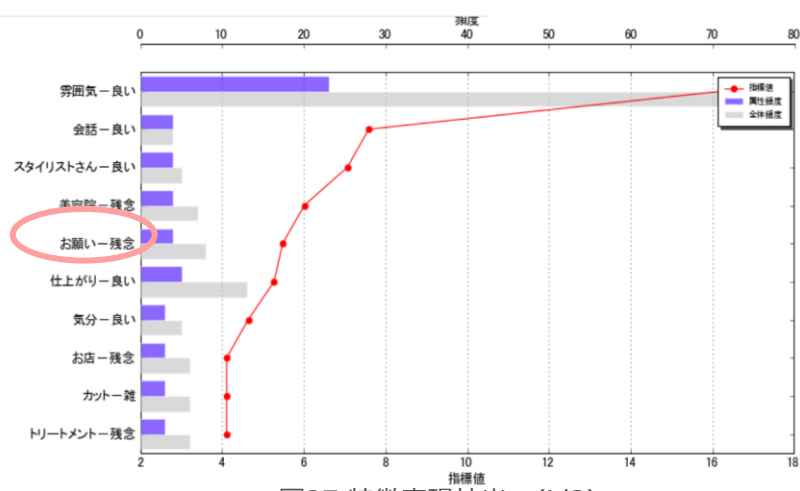


図27.特徴表現抽出 (id3)

- お店の雰囲気や仕上がりは良い
- 店員の融通, 技術に不満

どちらも店員の融通やテクニックに不満を持っている

## 8.まとめ・提案

### まとめ

- 20代後半・30代前半どちらも**サービス・テクニック**に不満を持っている
- 特に「**お願いー残念**」の項目から、頼むというアクションを起こしたにも関わらず期待通りにならなかったことが不満を持つ原因なのではないか
- 住所の比較により年齢が少し変わるだけで、**同じ場所でも評価が逆**になる場合がある

### 提案

学割や初回来店のみの価格やクーポンがあるが、東池袋・南池袋・自由が丘のように同じエリアでも評価が変わるようなところには「**年齢ごとに細かな設定**」を試みるのが良いと思う

## 9. 今後の課題

---

- テキストマイニングではクラスターid3に着目し, 年齢も絞ったがさらに幅広い年代と比較したい (appendix掲載内容に着手)
- データ数は少なくなるが男性にどのような特徴があるのかを調べる
- 提案の内容を料金設定などのさらに具体的な案を提示し, 年齢・地域が違う場合でも適応できるものを考えたい

# 参考文献

---

[1]理容師美容師試験研修センター「新規免許登録者件数」[http://www.sb.rbc.or.jp/2006/11/post\\_5.html](http://www.sb.rbc.or.jp/2006/11/post_5.html)  
(最終閲覧日2020年11月15日)

[2]厚生労働省衛生行政報告例 「美容所の施設数・従業理容」  
[https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou\\_iryuu/kenkou/seikatsu-eisei/seikatsu-eisei21/index.html](https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryuu/kenkou/seikatsu-eisei/seikatsu-eisei21/index.html) (最終閲覧日2020年11月15日)

[3]株式会社矢野経済研究所 「プレスリリース No.2409 理美容市場に関する調査」  
<https://www.yano.co.jp/> (最終閲覧日2020年11月15日)

[4]ヘアログ 「緊急事態宣言解除後の美容室利用に関するアンケート調査」  
<https://hairlog.jp/special/article/9664> (最終閲覧日2020年11月15日)

[5]株式会社リクルートライフスタイル ホットペッパービューティーアカデミー  
美容センサス2020上期 「15～69歳男女の美容サロン利用実態」 (最終閲覧日2020年11月15日)  
[https://hba.beauty.hotpepper.jp/wp/wp-content/uploads/2020/06/census\\_fullreport\\_202006.pdf](https://hba.beauty.hotpepper.jp/wp/wp-content/uploads/2020/06/census_fullreport_202006.pdf)

[6]正岡幹之, 二宮正司, 美容室のマーケティング戦略に関する一考察, 大阪経大論集・第66巻台4号・2015年11月  
p.185-p.212

[7] HOT PEPPER Beauty 「HOT PEPPER Beauty とは」  
<https://beauty.hotpepper.jp/> (最終閲覧日2020年12月10日)

[8]株式会社NTTデータ数理システム 「VMS TMS」  
<https://www.msi.co.jp/index.html> (最終閲覧日2020年12月10日)



# Appendix

# Appendix[3]

## ❖美美容市場規模

- 調査方法  
当社専門研究員による直接面談，電話・e-mailによるヒアリング，  
郵送アンケート調査，ならびに文献調査併用
- 調査対象  
理美容チェーン，シェアサロン運営企業，理美容商社・卸，理美容  
化粧品・機器メーカー等
- 調査期間  
2020年1月～3月
- 調査実施機関  
株式会社矢野経済研究所

### ＜理美容市場とは＞

本調査における理美容市場とは国内における理容所（理容室）と美容所（美容室）で提供する施術（理髪・パーマメント・カット・セット・カラー等その他サービス）及び店頭・専用サイトにおける物販サービスの市場をさす。また訪問理美容サービス等の出張，無店舗型サービスも含む

### ＜市場に含まれる商品・サービス＞

理容美容の各施術（理髪・パーマメント・カット・セット・カラー等その他サービス）及び店頭・無店舗での物品販売

# Appendix[4]

---

## ❖緊急事態宣言解除後の美容室利用に関するアンケート調査

- 調査方法  
インターネット調査
- 調査対象  
全国20代男女
- 調査期間  
2020年6月15日～6月16日
- 調査実施機関  
株式会社ノーマリズム ヘアログ実施

期間を以下のように定める

- ・ビフォーコロナ: 新型コロナウイルス感染拡大以前
- ・ウィズコロナ: 新型コロナウイルス感染拡大～新型コロナウイルス収束
- ・アフターコロナ: 新型コロナウイルス収束以降

# Appendix[5]

---

## ❖15～69歳男女の美容サロン利用実態

- 調査目的

女性・男性それぞれの過去1年間における美容サロンの利用実態を把握し美容に対する意識をとらえる

- 調査方法

インターネットによるアンケート調査 - 株式会社マーケティングアプリケーションズの「MApps Panel」を利用

- 調査対象

全国人口20万人以上の都市に居住する15～69歳の男女 各6,600サンプル

- 調査期間

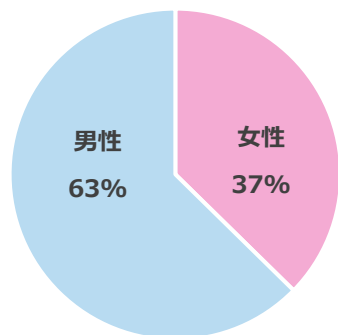
2020年2月14日（金）～2月25日（火）

- 調査実施機関

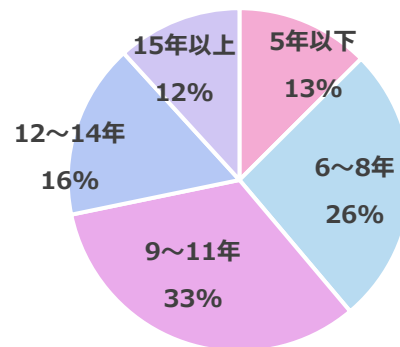
株式会社アンド・ディ

# Appendix 基礎集計

## ❖ スタイリスト 性別割合

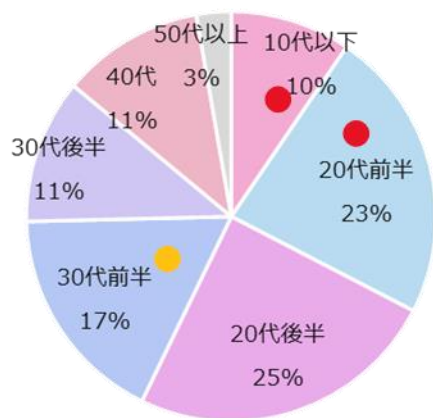


## ❖ スタイリスト キャリア割合

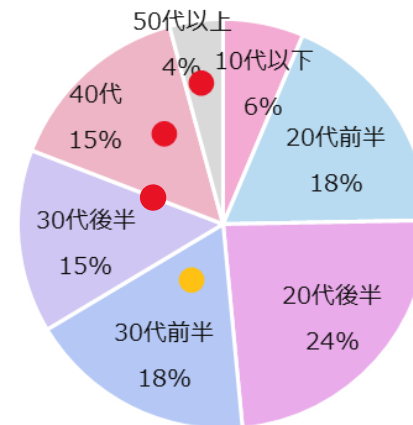


# Appendix 分析 クラスター 基礎集計

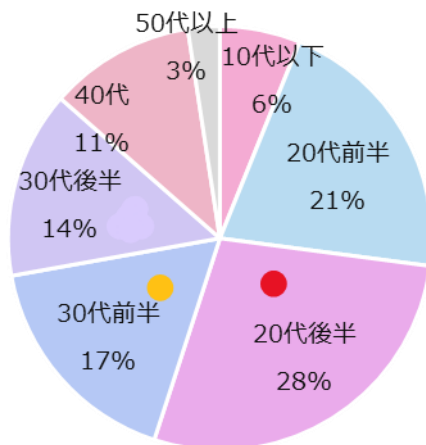
## ❖Id 1 年代別割合



## ❖Id 2 年代別割合

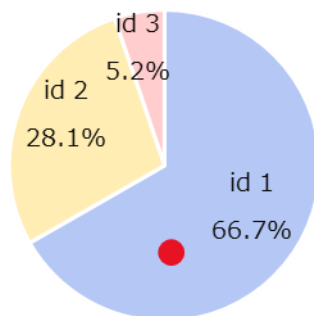


## ❖Id 3 年代別割合

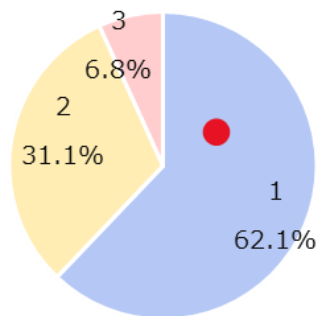


# Appendix 分析 クラスタ 基礎集計

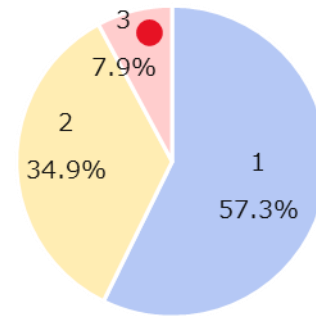
## 10代以下



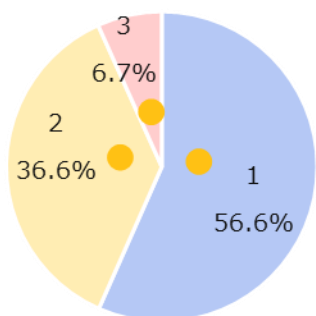
## 20代前半



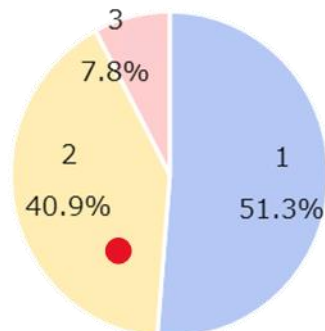
## 20代後半



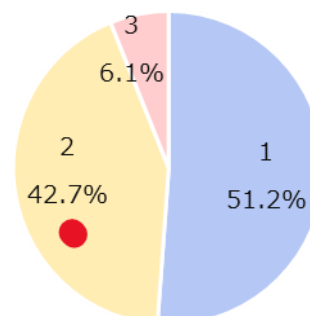
## 30代前半



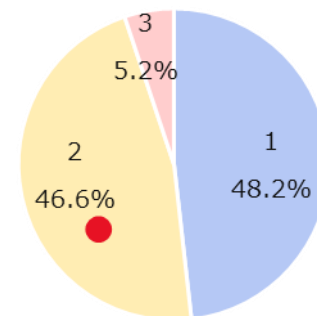
## 30代後半



## 40代



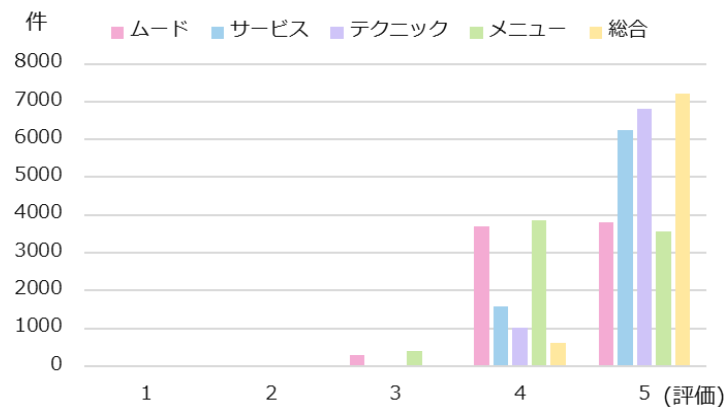
## 50代以上



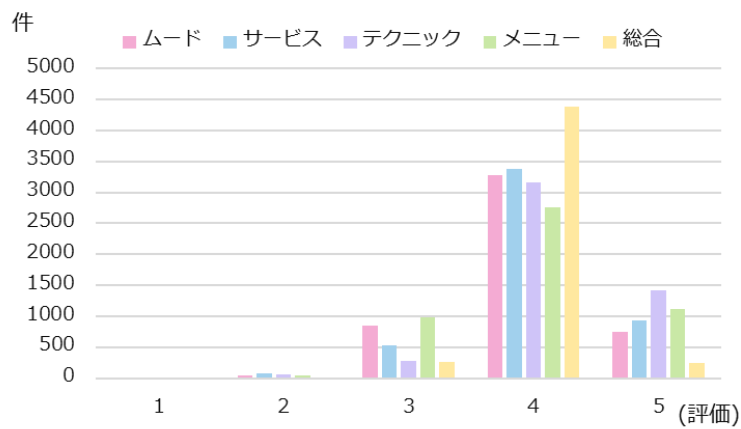
# Appendix 分析 クラスター 基礎集計

## ❖クラスターごとと評価件数

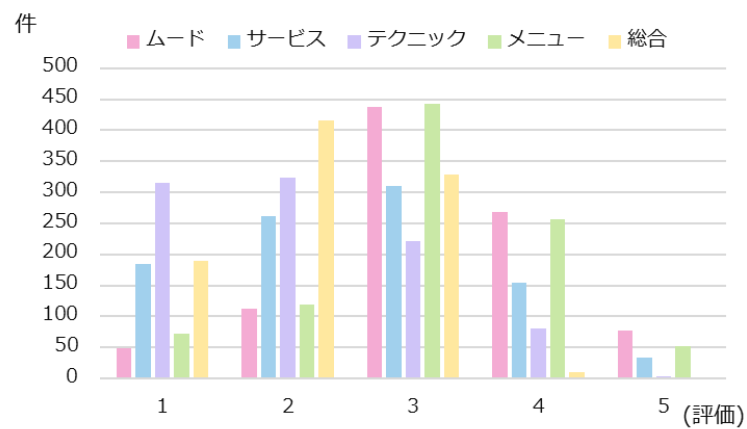
### ID1



### ID2



### ID3

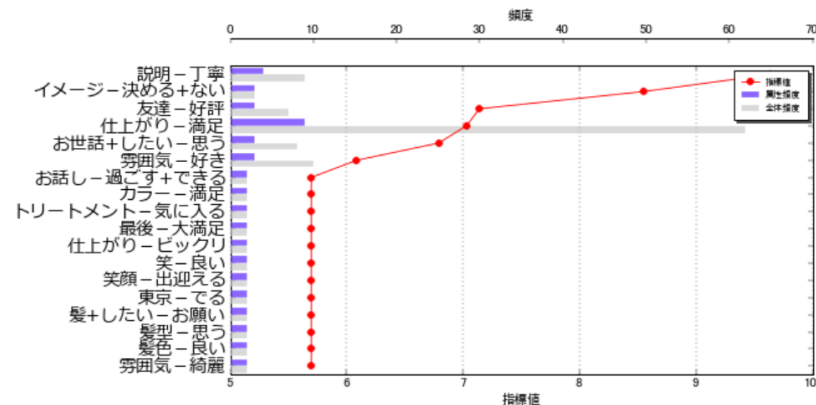




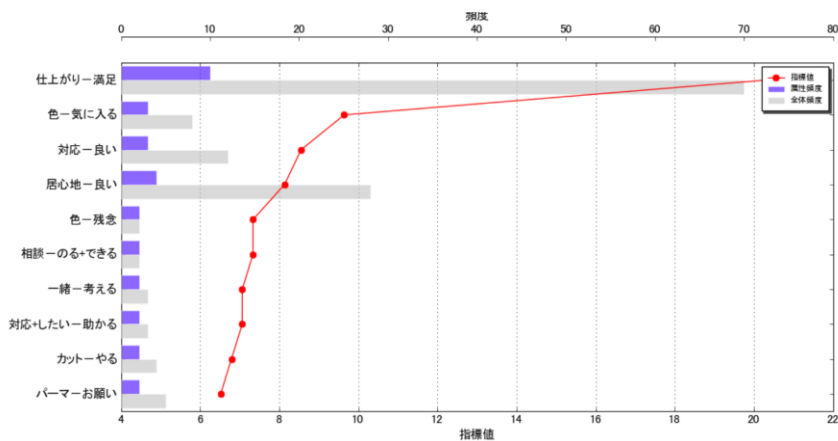
# Appendix 分析 テキストマイニング

## 10代以下 特徴表現抽出

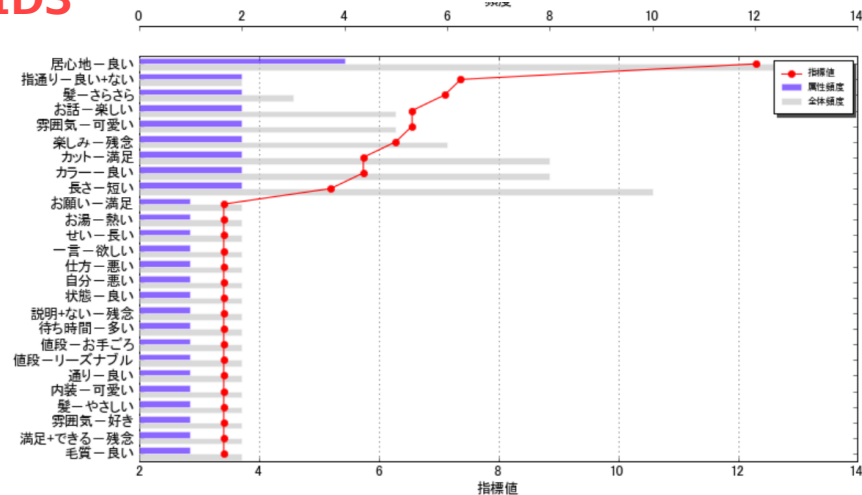
### ID1



### ID2



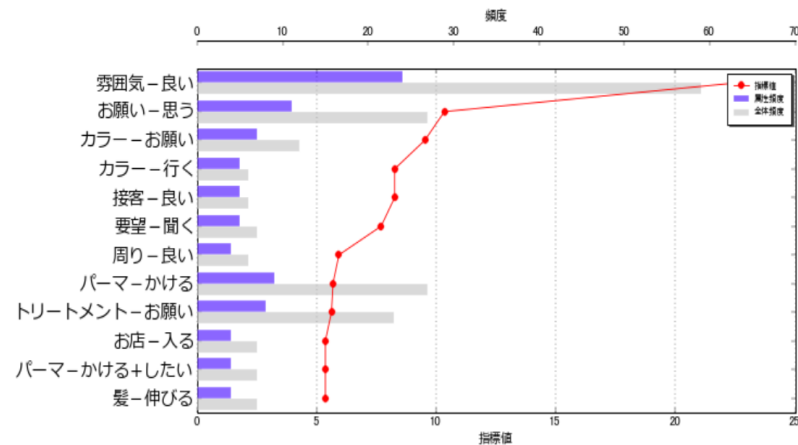
### ID3



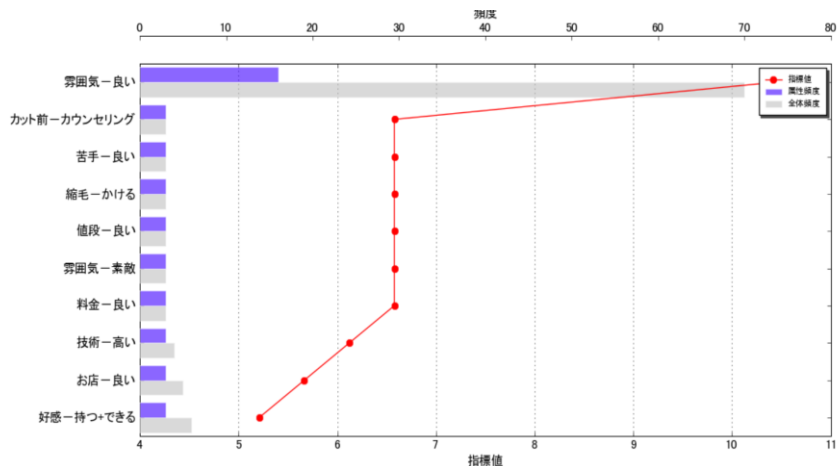
# Appendix 分析 テキストマイニング

## 20代前半 特徴表現抽出

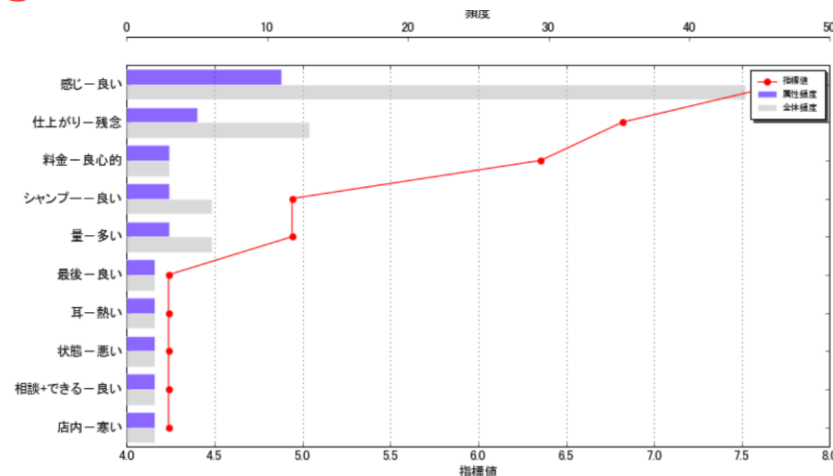
### ID1



### ID2



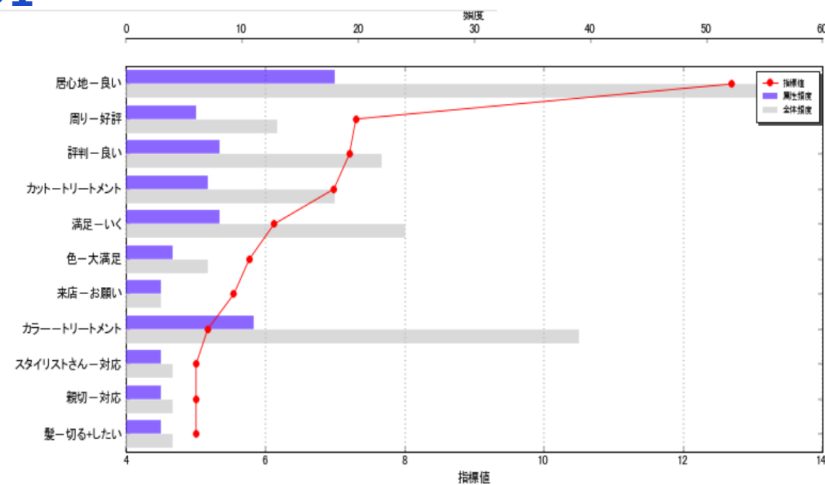
### ID3



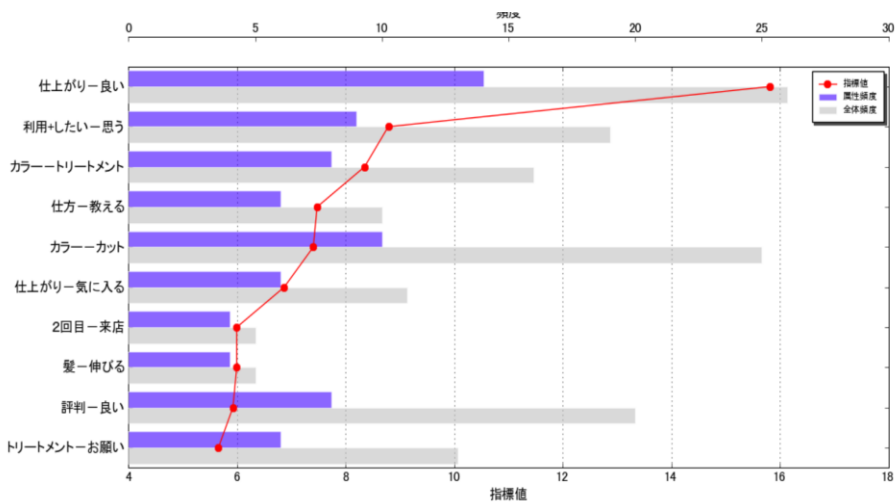
# Appendix 分析 テキストマイニング

## 20代後半 特徴表現抽出

ID1



ID2



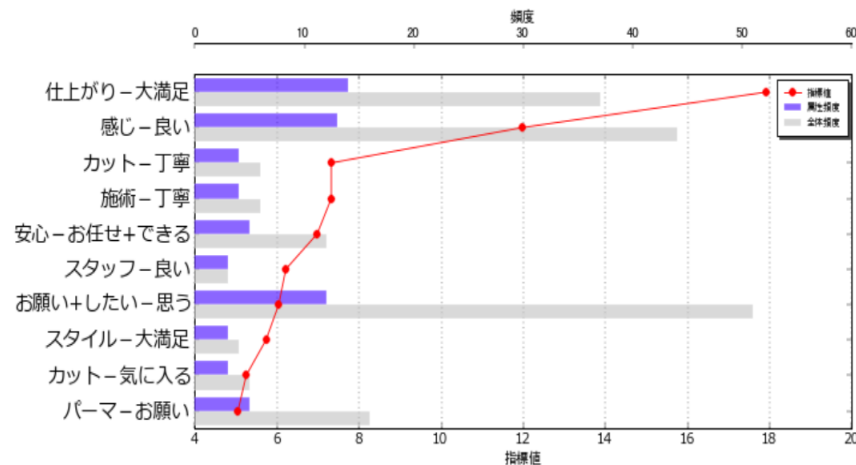
ID3

前

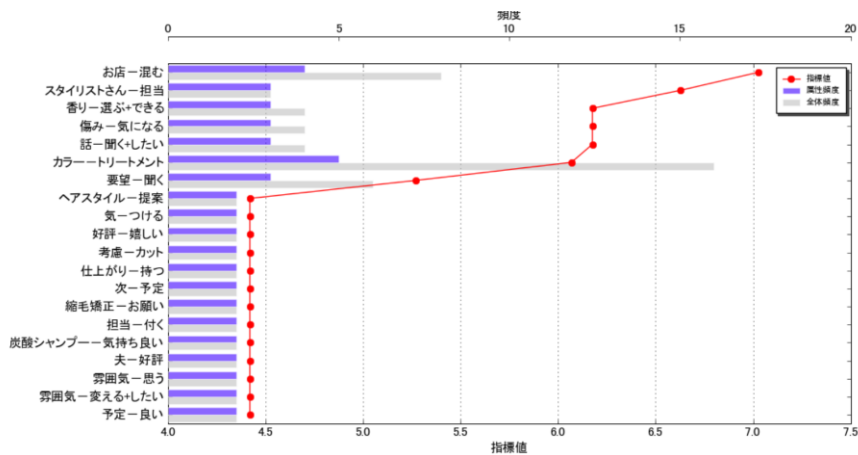
# Appendix 分析 テキストマイニング

## 30代前半 特徴表現抽出

### ID1



### ID2



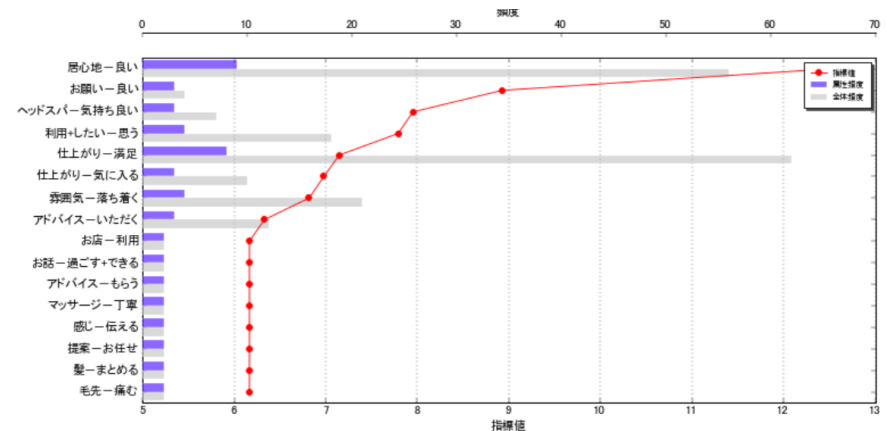
### ID3

前

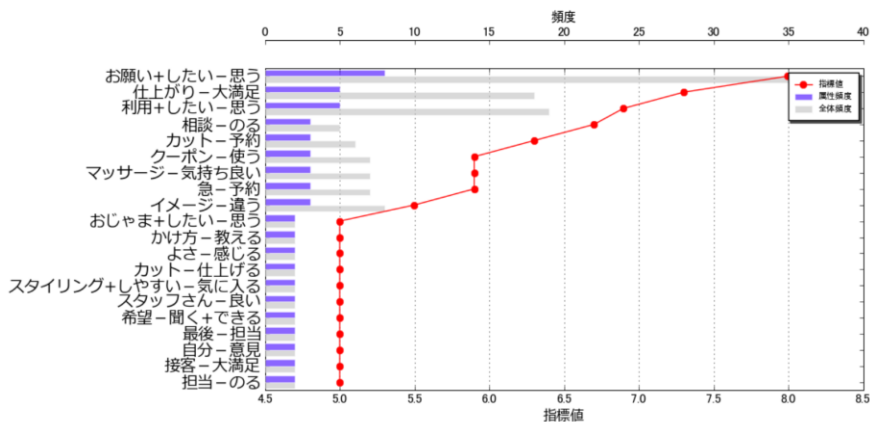
# Appendix 分析 テキストマイニング

## 30代後半 特徴表現抽出

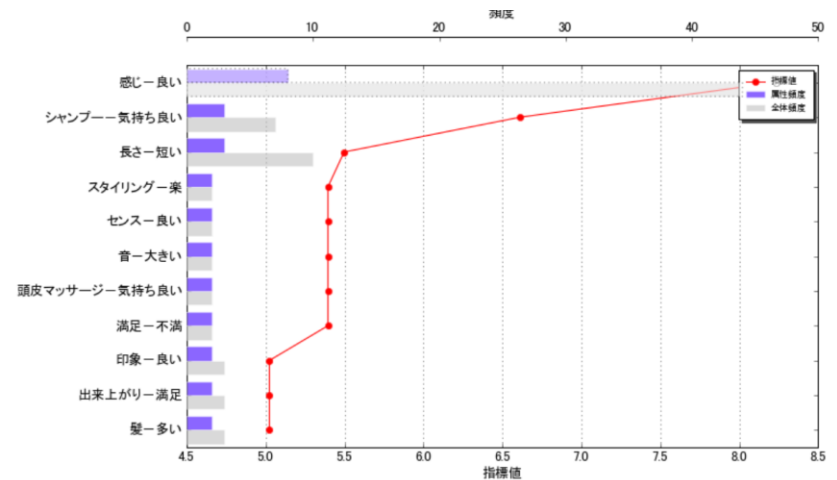
### ID1



### ID2



### ID3

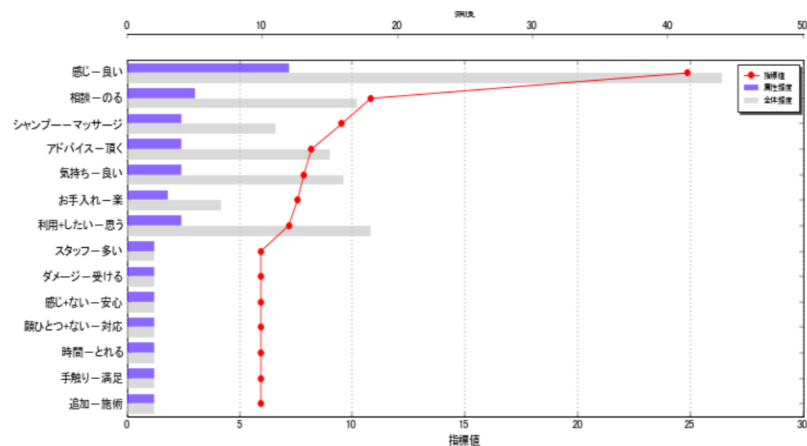


# Appendix 分析 テキストマイニング

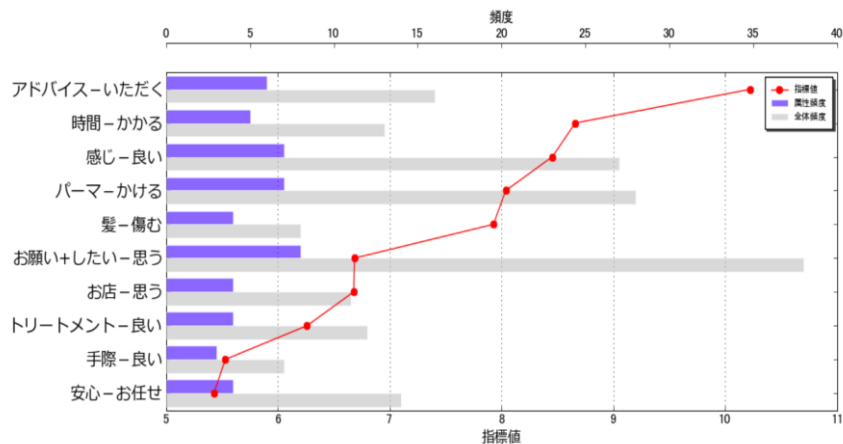
40代

特徴表現抽出

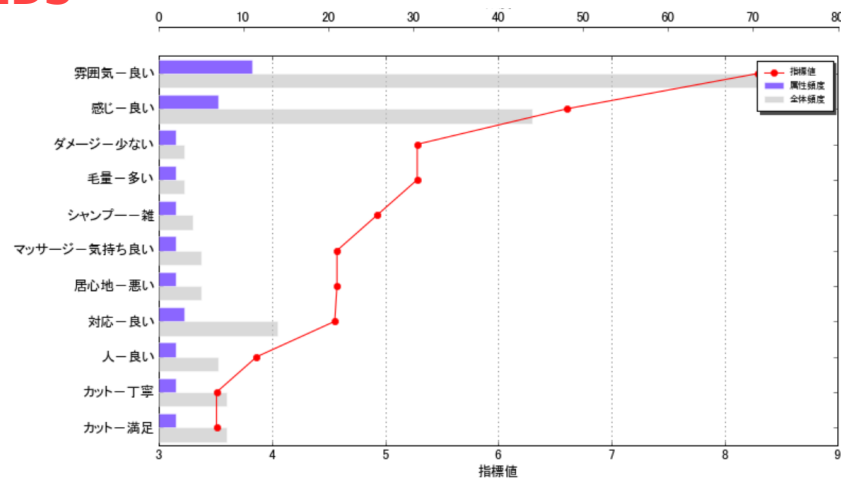
ID1



ID2



ID3

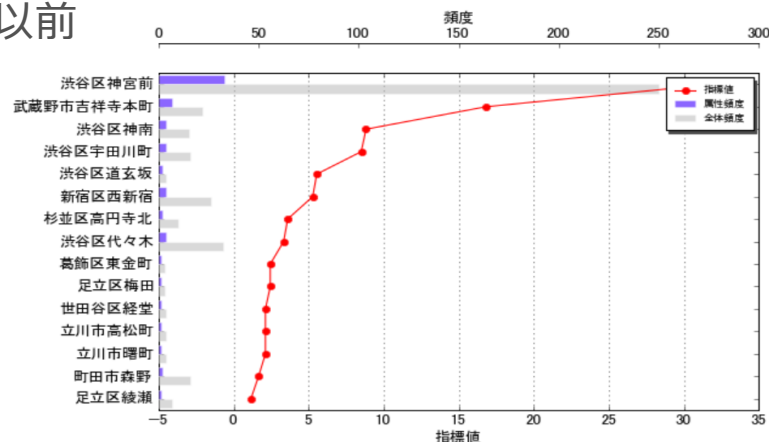




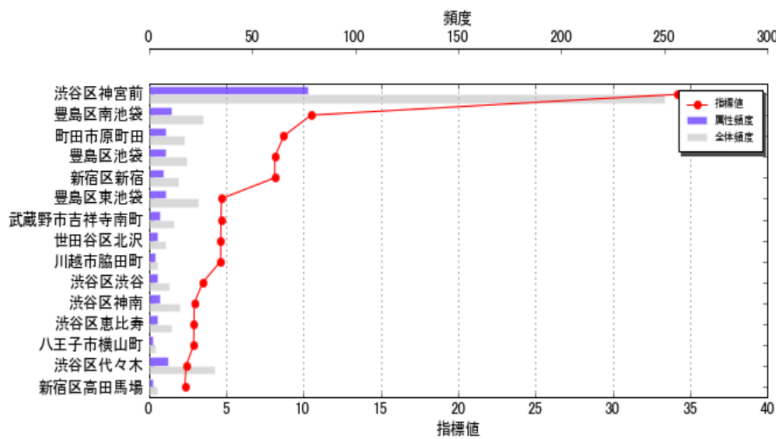
# Appendix 分析 テキストマイニング all5

## ◆特徴表現抽出 年代別

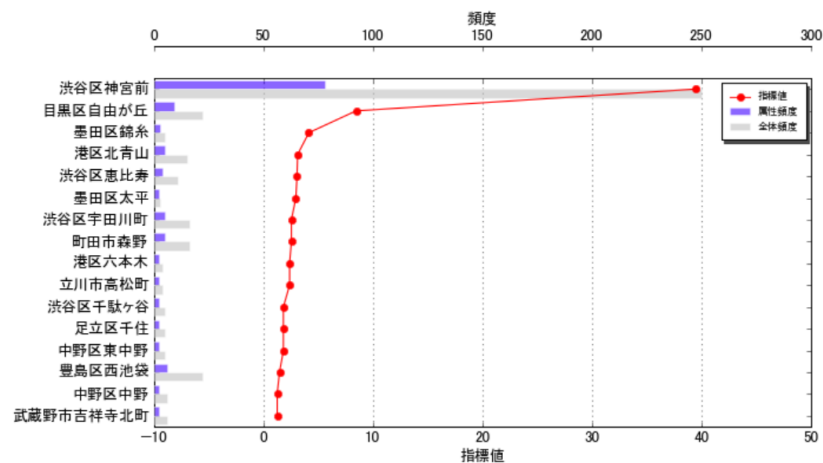
10代以前



20代前半



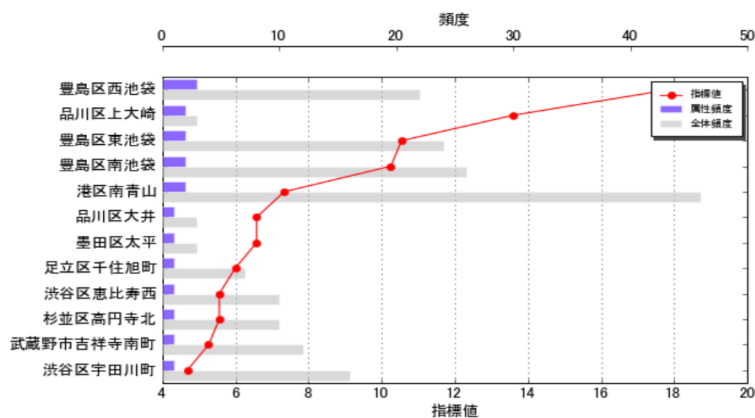
20代後半



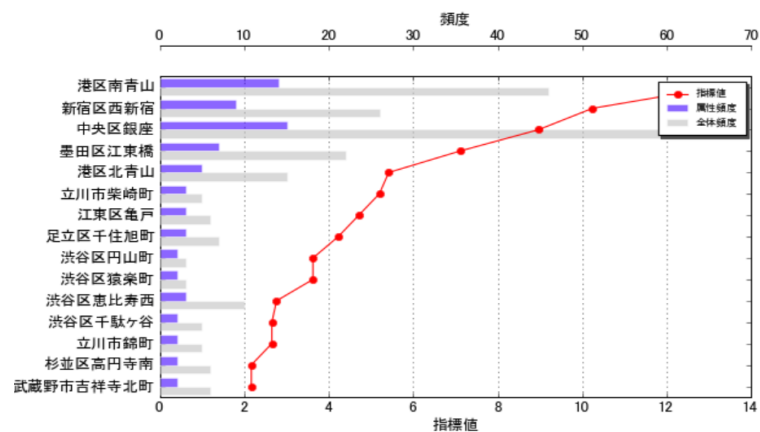


# Appendix 分析 テキストマイニング all5

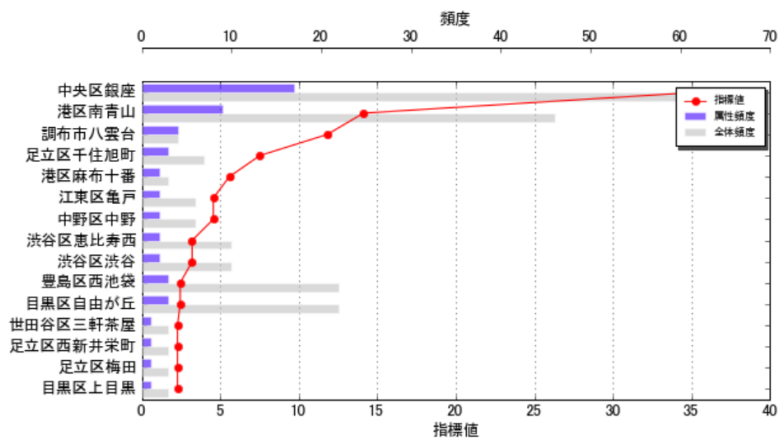
## 30代前半



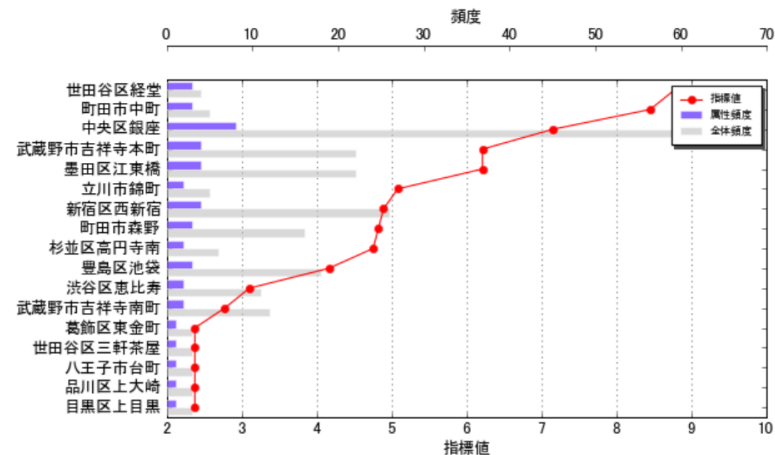
## 30代後半



## 40代



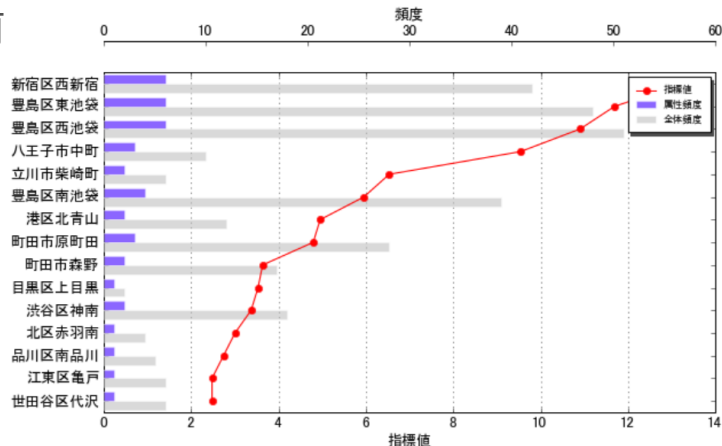
## 50代



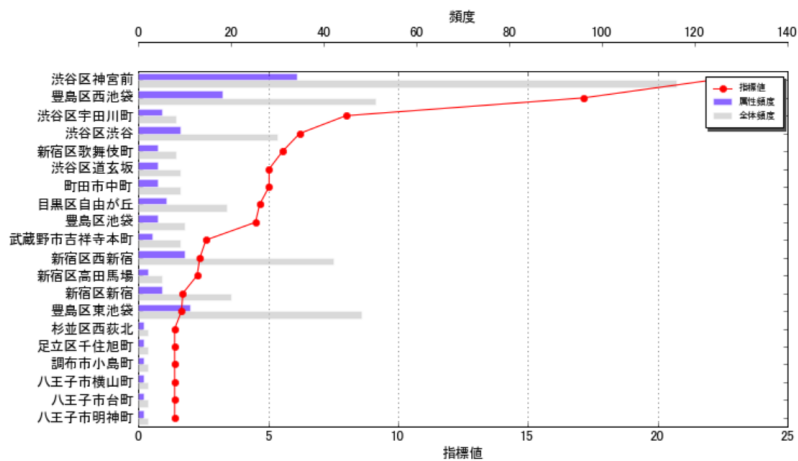
# Appendix 分析 テキストマイニング id3

## ◆特徴表現抽出 年代別

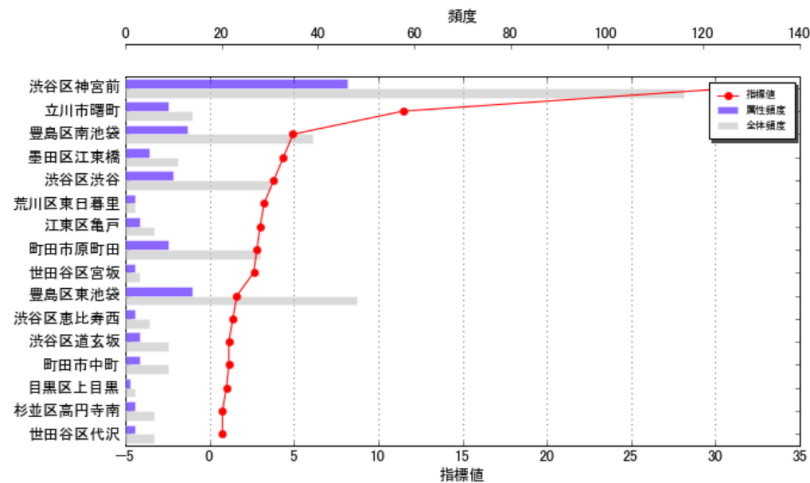
10代以前



20代前半

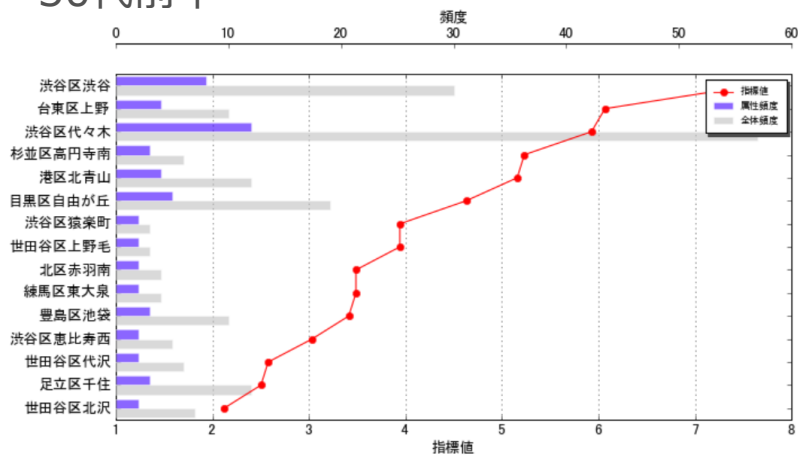


20代後半

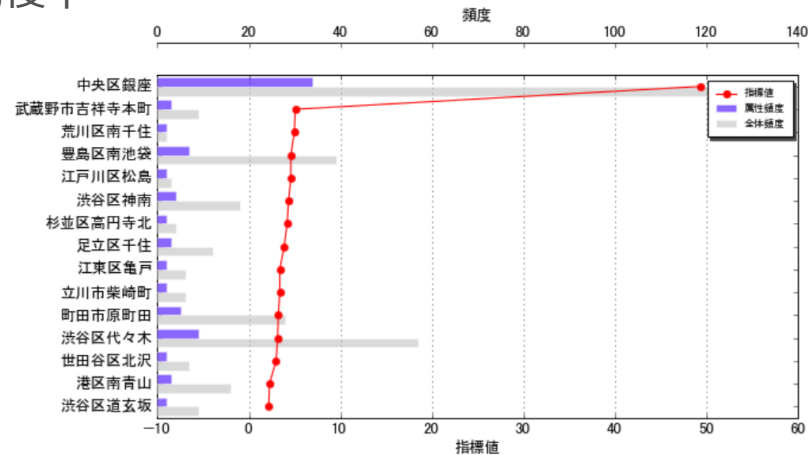


# Appendix 分析 テキストマイニング id3

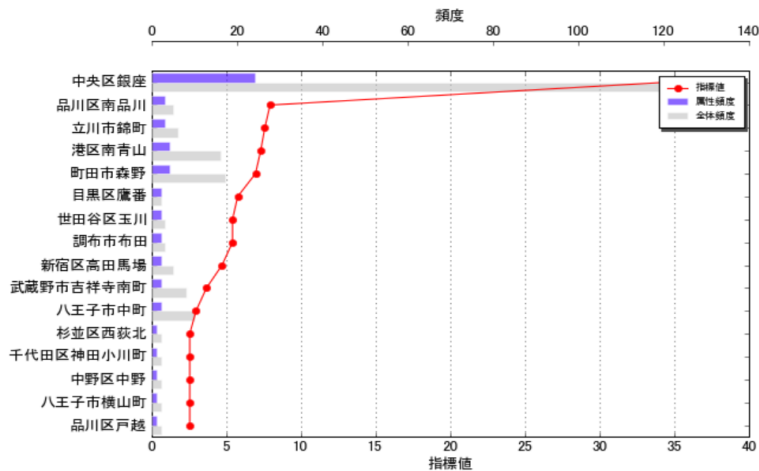
## 30代前半



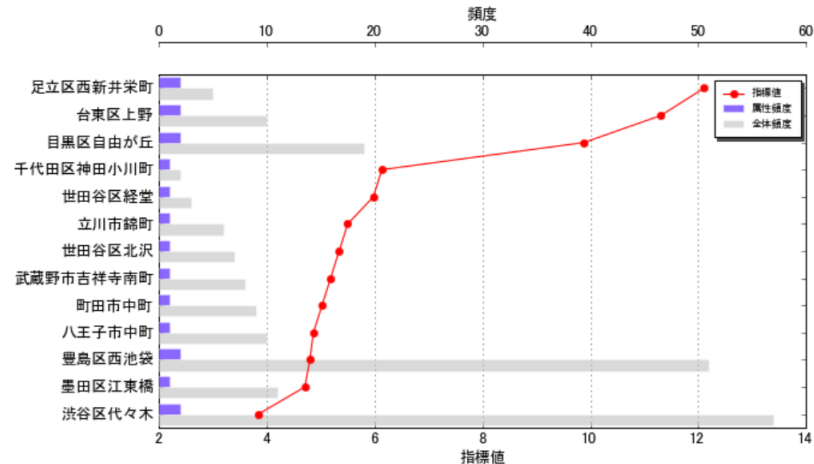
## 30代後半



## 40代



## 50代



# Appendix 分析 テキストマイニング

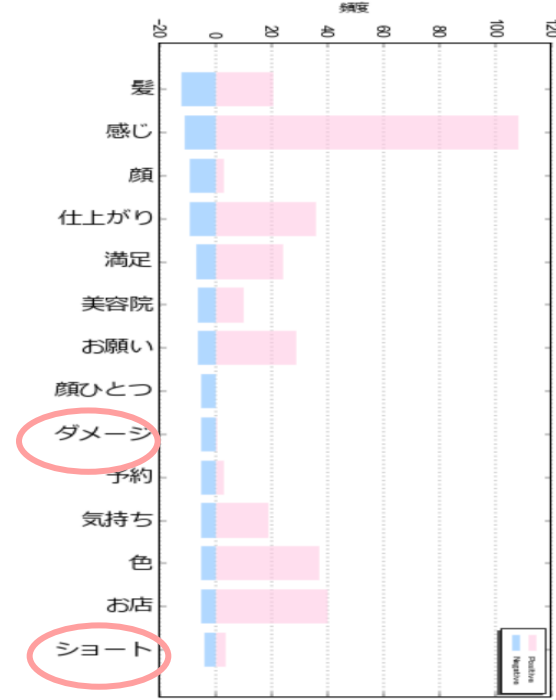
## ❖ 評判抽出 (id1)

※945件 抽出

### <好評語ランキング>



### <不評語ランキング>

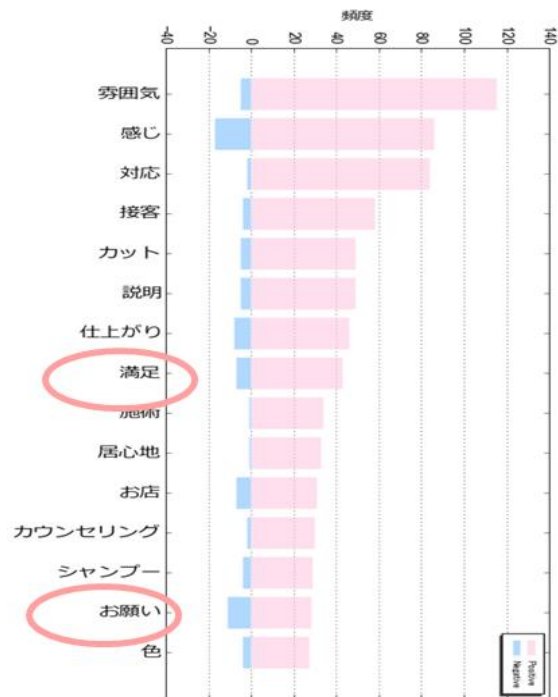


# Appendix 分析 テキストマイニング

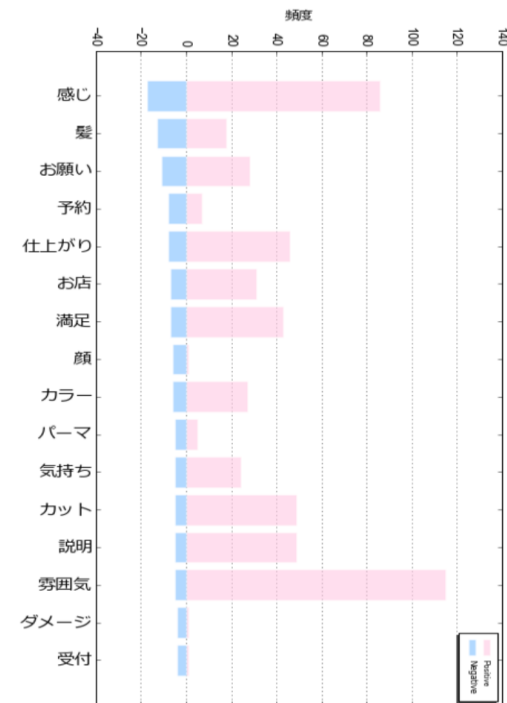
## ❖ 評判抽出 (id2)

※945件 抽出

### <好評語ランキング>



### <不評語ランキング>



# Appendix

