



## 楽天データ公開

楽天の持つ様々なデータを、大学・公的研究機関の研究用に公開しています。

### Rakuten

- ・全商品データ (約2億8,300万商品)
- ・商品レビューデータ (約7,000万件)
- ・ショッピングレビューデータ (約2,250万件)

### Rakutenレシピ

- ・レシピデータ (約80万件)
- ・レシピ画像 (約80万件)
- ・Pickupレシピ (1,854件)
- ・ディリシャスニュース (362件)

### Rakuten Travel

- ・施設データ (約2.9万施設)
- ・レビューデータ (約656万件)

### Rakuten GORA

- ・ゴルフ施設データ (1,669施設)
- ・レビューデータ (約32万件)

### アノテーション付きデータ

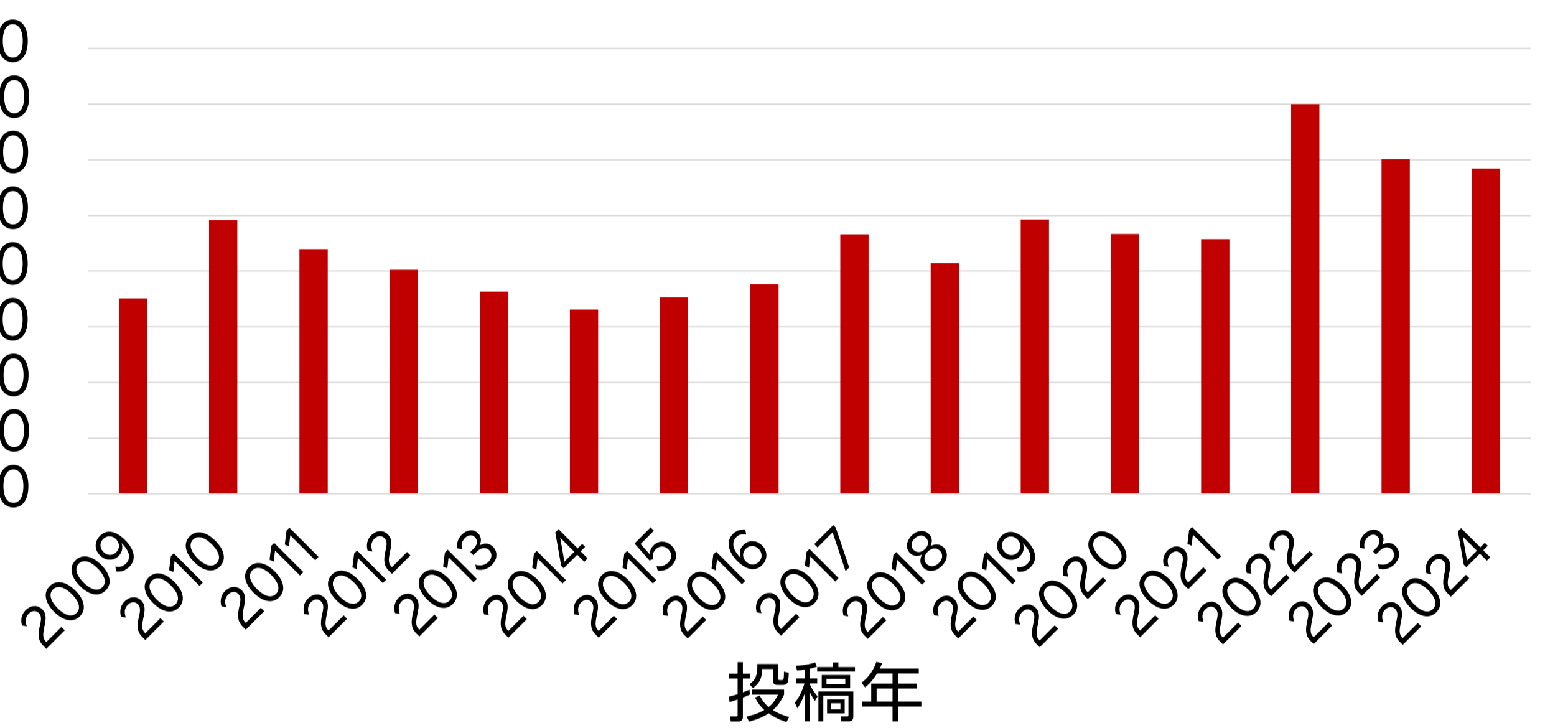
- ・筑波大学文単位評価極性タグ付きコーパス (TSUKUBAコーパス)
- ・カテゴリラベル付き商品画像データセット
- ・文字領域アノテーション画像
- ・楽天不動産間取り図と壁ラベル
- ・Rakuten France ユーザ評価/レビュー有効性情報/書籍データ/著者名のアノテーション付きデータ
- ・Rakuten France マルチモーダルプロダクトデータセット
- ・楽天ブックス: 著者名の曖昧性解消実験用書誌データ
- ・楽天トラベル: レビューアスペクト・センチメントタグ付きコーパス

## 楽天トラベルのレビューデータを更新します！！ (近日、提供開始予定) [1]

- ・昨今のトレンドの移り変わりや、COVID-19前・中・後のデータ分析を可能にします
- ・機械学習モデルを、昨今のデータで学習・評価を行いたい場合の元データとしてご活用できます

- ✓ データドリフトにより、頑健・実用的なモデルを構築するためには、新しいデータを使用したモデルの再学習・再評価が重要

レビュー投稿件数



	更新前	更新後
投稿年の期間	2004-2019	2009-2024
レビュー件数	約656万件	約729万件

### 有意水準1%で文書頻度に有意差のある単語n-gram

要因	Word n-gram (1<=n<=3)	文書頻度		オッズ比
		2024年	2019年	
COVID-19	コロナ	1152	20	48.79
	コロナ禍	530	0	∞
	リモートワーク	99	2	41.78
	非接触	115	0	∞
	QRコード	819	20	31.50
	モバイル-オーダー	55	0	∞
	ホカンス	68	0	∞
SDGs	SDGs	253	1	196.73
	フードロス	85	6	11.96
	環境-配慮	151	38	3.35
	アメニティ-環境-配慮	22	0	∞
新機能・トレンド	ReFa	900	0	∞
	ナノバブル	29	0	∞
	ドームテント	128	0	∞
	ととのう	177	15	9.96
インバウンド	インバウンド	1337	263	4.30
	インバウンド-影響	102	10	8.61
スラング	エモい	23	0	∞
インフレ	高騰	1286	173	6.12
	価格高騰	126	3	35.45
	物価高	249	1	210.28

### 2019年と2024年のレビュー例

#### 意味ドリフト

2019年: コロナの施設が下位にあり、楽しむことが出来ました。

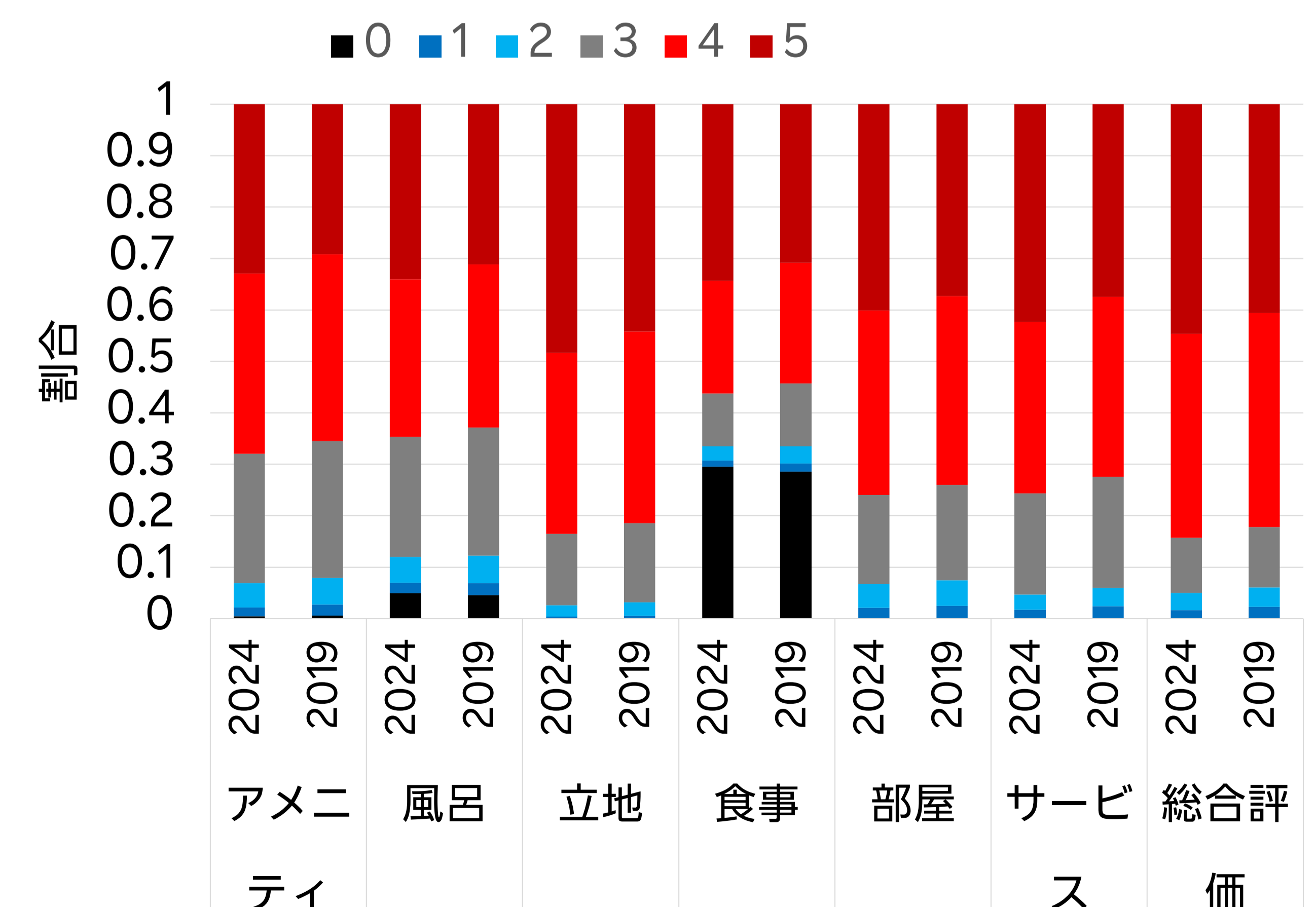
#### 評価観点の変化

2024年: アメニティも全て紙や木を使ったもので環境に配慮されていて、かつオシャレでした。

#### スラングの創出

2024年: クラブラウンジは、とても静かでチルい!  
2024年: 沼るカレーです。

### 評価点の分布



大きいカテゴリでは傾向に違いは見られない  
⇒細かい粒度でカテゴリごとにどのような違い (観点、理由など) があるのかを、明らかにしていただくと幸いです

[1] Yuki Nakayama, Koki Hikichi, Yun Ching Liu, Yu Hirate.  
Rakuten Data Release: A Large-Scale and Long-Term Reviews Corpus for Hotel Domain (arXivにて、近日公開予定)