

スケジュール

day 2

- 講義 – 自然言語処理の最新動向
- 講義 – テキストマイニングの手順
- 講義&演習 – データ理解

day 3

- 講義&演習 – 演習環境の準備
- 講義&演習 – テキスト解析 (1)
- 講義&演習 – テキスト解析 (2)

day 4

- 講義&演習 – テキスト分析 (1)
- 演習 – テキスト分析 (実践編)

day 5

- 演習 – テキスト分析 (実践編)

自然言語処理の最新動向

テキストマイニングの手順

テキストマイニング

- 驚異的な大量の文書データに記述されている多種多様な内容を対象として、その相関関係や出現傾向などから新たな知識を発見する [那須川,1999]
- 市場調査や販売戦略の立案、製品やサービス改善、顧客対応の改善に役立てたい
 - アンケート、レビューサイトのクチコミ、ツイートなど
- 最近では、報道番組などで Twitter 分析を取り上げることも多い
 - 震災、選挙、新型コロナウィルスなど

クチコミ分析の例 — コックroach

- パッケージ描かれたイラストが嫌 → 変更後、前年比2倍の出荷



出典: http://www.kincho.co.jp/seihin/insecticide/go_aerosol/gokiburi_u_spray/index.html

クチコミサイトの例 — 楽天トラベル

● ホテルのクチコミ数: 1,393万件 ※年間約60~80万

The screenshot shows the Rakuten Travel website at <https://travel.rakuten.co.jp/review/>. The main heading is 'お客様の声' (Customer Reviews) with the number '13,929,531'. Below it is a search bar for reviews and a section for '新着! 最新のクチコミ' (Newest reviews). There are also icons for domestic and overseas hotels, and a sidebar with reviews for CANDEO HOTELS and BIO-RESORT HOTEL & SPA OSE.

経年変化:

- 780万件 (2015)
 - 836万件 (2016)
 - 900万件 (2017)
 - 973万件 (2018)
 - 1,042万件 (2019)
 - 1,098万件 (2020)
 - 1,165万件 (2021)
 - 1,237万件 (2022)
 - 1,325万件 (2023)
 - **1,393万件 (今回)**
- ※ 2024/5/19現在

Rakuten Hotel Review Page

楽天宿・航空券・ツアー予約

国内旅行 国内ツアー レンタカー 高速バス 海外旅行 海外ツアー 海外航空券 海外ホテル 割引クーポン 懸賞広場 観光案内

楽天トラベルトップ > 全国 > 千葉県 > 外房(鴨川・勝浦・御宿・茂原) > 鴨川温泉 > 鴨川シーウールドホテル クチコミ・感想・情報

鴨川シーウールドホテル

★★★★★ 4.12 クチコミ・お客さまの声(886件) この宿泊施設をお気に入りに追加 メルマガ 幹事さん機能 友達登録 シェアする 3 ツイート フacebook リン

日程からプランを探す

- 国内宿泊
- ANA 航空券+宿泊
- JAL 航空券+宿泊
- 日帰り・ディユース
- 日付未定
- チェックイン 2015/06/21
- チェックアウト 2015/06/22
- ご利用部屋数 1 部屋
- ご利用人数 1部屋目：大人(1人)子供(0人)
- 金額(1部屋1泊あたり消費税込) 下限(制限なし) 上限(制限なし)
- 検索

地図

最近見た宿泊施設

11軒の閲覧履歴があります [ページ1/6]

鴨川シーウールドホテル 痞めぐらの宿 茂みさき もっと見る

施設間違情報

- 鴨川シーウールドホテル
- ★トップページ★
- 鴨川シーウールドホテル
- ★鴨シニュース★

楽天カード入会で2,000ポイントプレゼント カード GORA 楽天市場

楽天トラベルの使い方 サイトマップ ヘルプ Languages 30%以上ポイントバック!

ようこそ、楽天トラベルへ 会員登録 ログイン 予約の確認・キャンセル

鴨川シーウールドホテルのクチコミ・感想・情報

総合評価 ★★★★★ 4.12 アンケート件数: 886件

施設紹介 プラン一覧 フォトギャラリー(76) 地図・アクセス お客様の声(886) クーポン一覧 プレゼント

鴨川シーウールドホテルのクチコミ・お客さまの声

●ホテル・旅行のクチコミTOP

総合評価 ★★★★★ 4.12 アンケート件数: 886件

評価内訳

5点	236件	サービス ★★★★★ 4.11
4点	302件	立地 ★★★★★ 4.61
3点	47件	部屋 ★★★★★ 3.53
2点	15件	設備・アメニティ ★★★★★ 3.62
1点	9件	風呂 ★★★★★ 3.53
		食事 ★★★★★ 4.10

項目別評価

サービス	立地	部屋	設備・アメニティ	風呂	食事	サニタリ
★★★★★ 4.11	★★★★★ 4.61	★★★★★ 3.53	★★★★★ 3.62	★★★★★ 3.53	★★★★★ 4.10	★★★★★ 3.53

最新の投稿順 評価が高い順 (総合 | サービス | 立地 | 部屋 | 設備・アメニティ | 風呂 | 食事)

772件中 1~20件表示 [1 | 2 | 3 | 4 | 5 | ... 全 39 ページ] ●次の20件

総合 ★★★★★ 4 RENDEZ_VOUSさんの 鴨川シーウールドホテル のクチコミ (感想・情報)

RENDÉZ_VOUSさん (3件) [30代/女性] 2015年06月17日 19:20:22

入園パスポート付き、バイキングの夕食と朝食付きでとてもお得な価格で泊まれました。

2日間ともシーウールドに立園出来るのでとても便利です。バイキングも種類が豊富でおいしく楽しくいただきました。温泉は湯船ひとりづきしかないので少しうつりませんでしたが、メインはシーウールドなので仕方ないかなと。

お部屋は古くて他の部屋の物音(子供が走り回る足音など)が気になりました。

量の上に座ったら立ちあちこちかゆくなりました。オーシャンビューで部屋の前の海が海なりうれしいのですが、波の音が大き聞こえて、なかなか眠れませんでした。設備が古くから仕方ないのかな。

総評すると部屋はちょっと微妙ですがお安いですし満足しています。シーウールドのシャチのショーや金曜日と土曜日に見ましたが、全然レベルが違った(笑)。

平日は練習程度のレベルでした。土曜日にもう一度見てよかったです。シャチ以外のショーは平日でもしっかりしていました。

レビューを評価してください このレビューは参考になりましたか?

旅行の目的 … レジャー 同伴者 … 家族 宿泊年月 … 2015年06月

ご利用の宿泊プラン いい値! パリュープラン

ご利用のお部屋 [wa5]シーウールドが見える特別室禁煙 [洋室]

不適切なレビューを報告する [はい] [いいえ]

Rakuten Hotel Review Page

鴨川シーウールドホテル

総合 ★★★★★ 2 投稿者さんの 鴨川シーウールドホテル のクチコミ (感想・情報)

投稿者さん 2015年06月11日 17:03:57

良かったところ

- ・部屋からの景色(朝日最高でした)
- ・食事(品数が多く、朝夕とも良かったです)
- ・フロントの方の対応(お姉さんがとても頑張っていました)以上。

掃除が行き届いているとの口コミを多く見ましたが、そろは思いませんでした。

気にかかる事は多々ありましたが、フロントのお姉さんが一生懸命で、その笑顔に救われた思いです。

フロントスタッフへのお言葉、誠にありがとうございます。

モチベーションアップに繋がりますので、お客様からの声として、スタッフと共有させて頂きます。

機会がございましたら、またご利用をお待ちしております。

旅行の目的 … レジャー 同伴者 … 家族 宿泊年月 … 2015年06月

ご利用の宿泊プラン [洋室 禁煙・特別室] お部屋からシャチやイルカも見える シーウールドと海一望宿泊プラン

ご利用のお部屋 [wa5]シーウールドが見える特別室禁煙 [洋室]

総合 ★★★★★ 4 投稿者さんの 鴨川シーウールドホテル のクチコミ (感想・情報)

投稿者さん 2015年06月11日 07:33:49

夫、2歳半と5ヶ月の子どもの4人で宿泊しました。

【立地】当たり前ですが鴨川シーウールドにとても近く、ゆっくり館内を見学できました。

【部屋】至って普通です。(古いからか、勝手の声は少し聞こえます。) トイレ掃除などはしっかりとされていました。清淨機などもTEL一本ですぐに届けて下さいました。

【食事】夜朝共にバイキング。イスですが子ども用イス、エプロン、ベビーベッドを用意して下さっています。キッズスペースも食事時間中に専門のスタッフの方がおりゆっくり食事ができます。

【風呂】小さな子ども(赤ちゃん)用のグッズ(ベビーベッド、コーナー、バス、おもちゃ、泡ソーパ、支えのあるイス)が揃っていました。お子さん連れも多く気兼ねなく楽しめました。しかしお風呂がひとつしかないのに、温泉を楽しむという雰囲気ではなく、銭湯のお湯が温泉という感じです。

また、23時頃にお風呂に行くと、アメニティやシャンプーが空だったのは少し残念でした。

【サービス】受付スタッフの皆さんとても親切、丁寧です。チェックアウト後に子どもの薬を冷蔵庫にいれておいて欲しいとダメ元で頼むと快く入

いい値! パリュープラン

【最安料金 (目安)】 10,186円~ (消費税込11,000円~)

【当日15:50からアシカと一緒に写真】笑ラジカと一緒にパチリ付プラン

【最安料金 (目安)】 10,278円~ (消費税込11,100円~)

【13:40~エコ・アーチームコミュニケーションタイム 1日3組限定】

【最安料金 (目安)】 10,278円~ (消費税込11,100円~)

【夜の水族館探検】 3月~10月の火・木曜日限定プラン

【最安料金 (目安)】 10,278円~ (消費税込11,100円~)

【14:50からイルカと一緒にパチリ 2室限】鴨川シーウールド体験プラン

【最安料金 (目安)】 10,463円~ (消費税込11,300円~)

今しかない!★アビ料理付シーウールド入園バスポート付で大満足♪5月・6月の月~木曜日限定プラン

【最安料金 (目安)】 10,926円~ (消費税込11,800円~)

【便利な赤ちゃんグッズ付】初お泊りはお母さんも嬉しい★赤ちゃんなつ得プラン

【最安料金 (目安)】 10,926円~ (消費税込11,800円~)

お子様にも大好評! オーシャンビュープラン

【最安料金 (目安)】 11,112円~ (消費税込12,000円~)

【80cmのジャンボサイズ】海の王者シャチぬいぐるみ付プラン

【最安料金 (目安)】 11,204円~ (消費税込12,100円~)

房総2大テーマパーク満喫「マザーパークチケット」付プラン

【最安料金 (目安)】 11,389円~ (消費税込12,300円~)

【14:50~イルカ

鴨川シーワールドホテルのクチコミ・お客様の声

[●ホテル・旅行のクチコミTOPへ](#)

総合評価

4.12

アンケート件数：886件

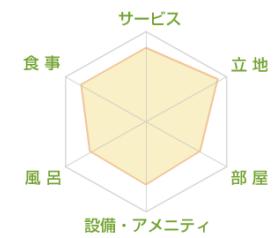
評価内訳

- 5点
- 4点
- 3点
- 2点
- 1点

236件
302件
47件
15件
9件

項目別の評価

サービス	4.11
立地	4.61
部屋	3.53
設備・アメニティ	3.62
風呂	3.53
食事	4.10



総合 2

投稿者さんの 鴨川シーワールドホテル のクチコミ (感想・情報)



投稿者さん

2015年06月11日 17:03:57

良かったところ

- ・部屋からの景色（朝日最高でした）
- ・食事（品数が多く、朝夕とも良かったです）
- ・フロントの方の対応（お姉さんがとても頑張っていました）以上。

掃除が行き届いているとの口コミを多く見ましたが、それは思いませんでした。

気にかかることは多々ありましたが、フロントのお姉さんが一生懸命で、その笑顔に救われた思いです。

評価

... 総合 2

- | | |
|----------|---|
| サービス | 2 |
| 立地 | 4 |
| 部屋 | 4 |
| 設備・アメニティ | 2 |
| 風呂 | 2 |
| 食事 | 4 |

旅行の目的

... レジャー

同伴者

... 家族

宿泊年月

... 2015年06月



鴨川シーワールドホテル

2015年06月11日 19:32:50

この度は、ご利用頂きまして誠にありがとうございます。

客室内清掃の件、大変申し訳

重要改善として、早急に対応いたします。

今後は、この様な事の無いように、清掃・点検を強化いたします。

フロントスタッフへのお言葉
誠にありがとうございます。

セラベーションアップに繋がる
お客様からの声として、
スタッフと共有させて頂きます。

機会がございましたら、またご利用をお待ちしております。

数値評価

テキストマイニングの手順

- データをよく知る
 - データ件数や構成比を集計 → データを理解する
 - 旅行目的別の人気エリアは?
 - 同伴者別の人気エリアは?
 - 数値評価による人気エリアの差異は?
- テーマを設定する
 - 解決すべき課題を決める → 分析目的を明確にする
 - 数値評価が低い原因是?
 - 高評価の施設に学ぶ改善点は?
- テキスト分析に取り組む
 - これら課題を解決するために、テキスト分析を実施

データ理解

実習用のデータ (Webサイトクローリング)

● 楽天トラベル のクチコミデータ

- 収集期間は 2019-2020 および 2022-2023(～GW明け) の 2セット
- 以下の 10 エリアごと同数に 1,000件ずつ ランダムサンプリング
- データ件数は 1万件 × 2セット

レジャー	5エリア	登別、草津、箱根、道後、湯布院	1,000件 × 10エリア
ビジネス	5エリア	札幌、名古屋、東京、大阪、福岡	= 計10,000件

実習用のデータ (Webサイトクローリング)

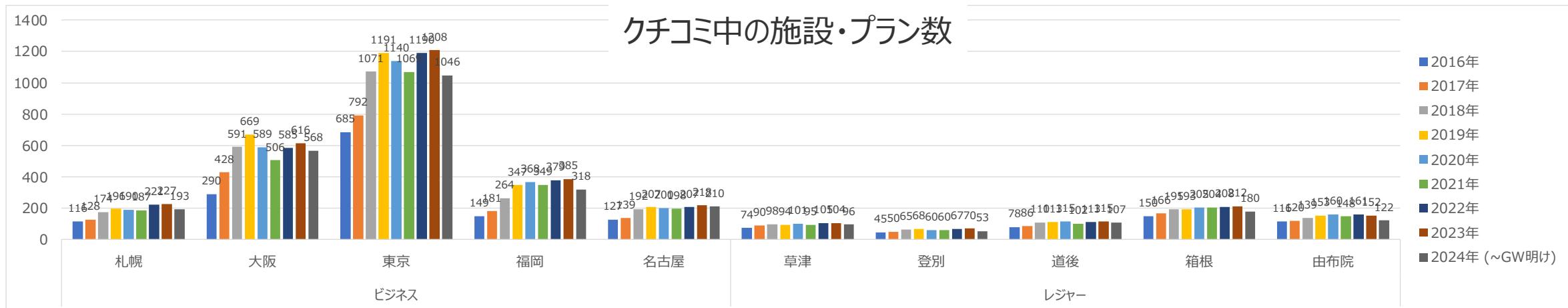
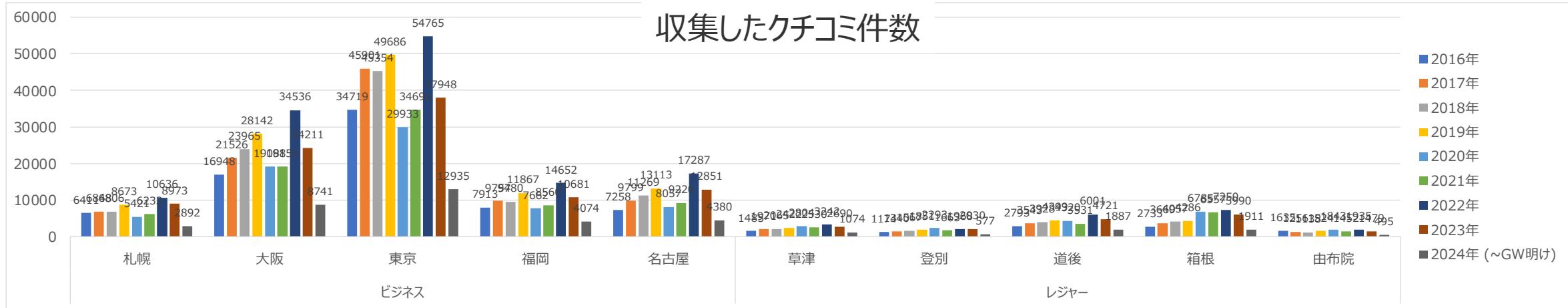
● 楽天トラベル のクチコミデータ

- 収集期間は 2019-2020 および 2022-2023(～GW明け) の 2セット
- 以下の 10 エリアごと同数に 1,000件ずつ ランダムサンプリング
- データ件数は 1万件 × 2セット
- データ項目は 18項目 (テキスト1項目+その他の属性17項目)

施設情報	4項目 カテゴリ, エリア, 施設番号, 施設名
口コミ	1項目 コメント (テキスト)
ユーザー評価	7項目 総合, サービス, 立地, 部屋, 設備・アメニティ, 風呂, 食事
その他の分類	2項目 旅行の目的, 同伴者
宿泊日	1項目 宿泊年月
ユーザー情報	3項目 ユーザー, 年代, 性別

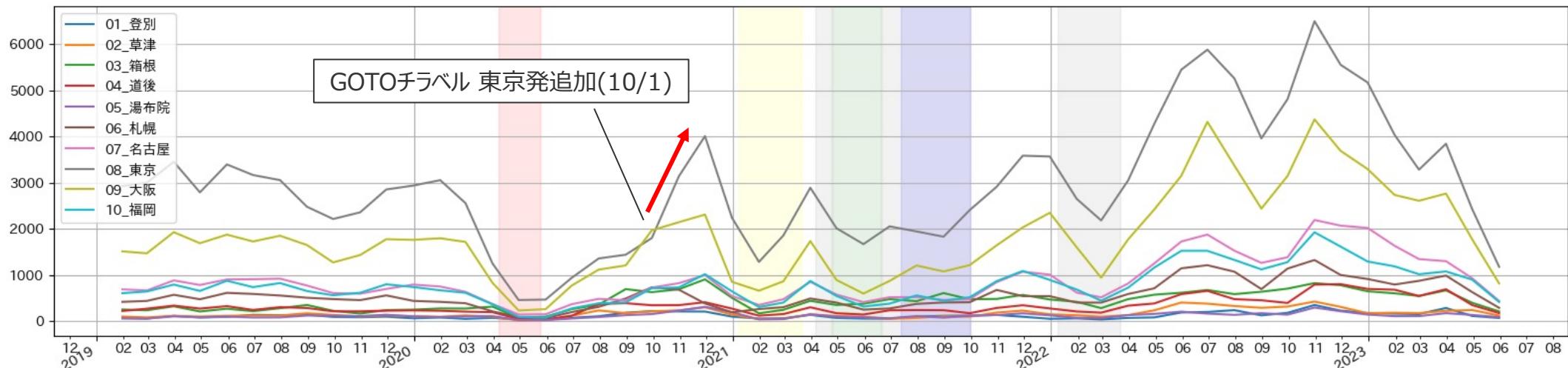
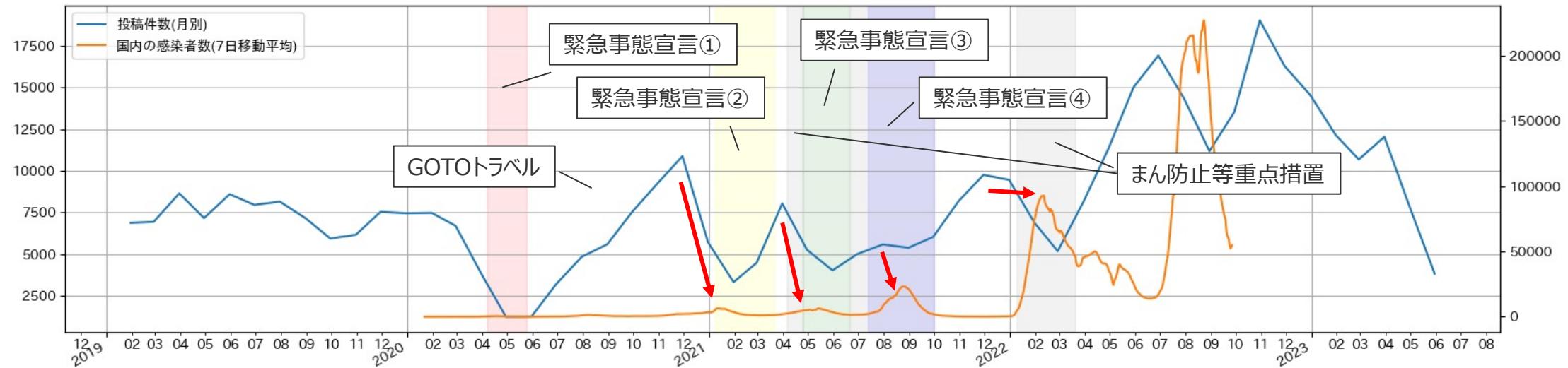
参考 — Webサイトクローリング

- 全量では 166.4万件、2021-2022は19.4万件、2023-2024は15.1万件



(参考) COVID-19 の影響

- クチコミの件数と感染者数の増減が連動 → クチコミ件数が一定の人流を反映している



実習用データ — ファイル一覧

● 実習用データは以下の通り → 主に「**rakuten-1000-2023-2024.xlsx**」を使用する

ファイル名	件数	データセット	備考
<u>rakuten-1000-2023-2024.xlsx.zip</u>	10,000	<ul style="list-style-type: none">・レジャー+ビジネスの 10エリア・エリアごと 1,000件 (ランダムサンプリング)・期間: 2023/1~2024 GW明け	本講義の全体を通して使用する
<u>rakuten-1000-2021-2022.xlsx.zip</u>	10,000	<ul style="list-style-type: none">・レジャー+ビジネスの 10エリア・エリアごと 1,000件 (ランダムサンプリング)・期間: 2021/1~2022/12	演習用 (年度で比較する場合など)
<u>rakuten-all-2023-2024-tsv.zip</u>	150,537	<ul style="list-style-type: none">・レジャー+ビジネスの 10エリア・サンプリング前の全データ・期間: 2023/1~2024 GW明け	参考用
<u>rakuten-all-2021-2022-tsv.zip</u>	193,721	<ul style="list-style-type: none">・レジャー+ビジネスの 10エリア・サンプリング前の全データ・期間: 2021/1~2022/12	参考用
<u>rakuten-all-tsv.zip</u>	1,664,301	<ul style="list-style-type: none">・レジャー+ビジネスの 10エリア・サンプリング前の全データ・期間: 2009/3~2024 GW明け	参考用

テキストマイニングの手順

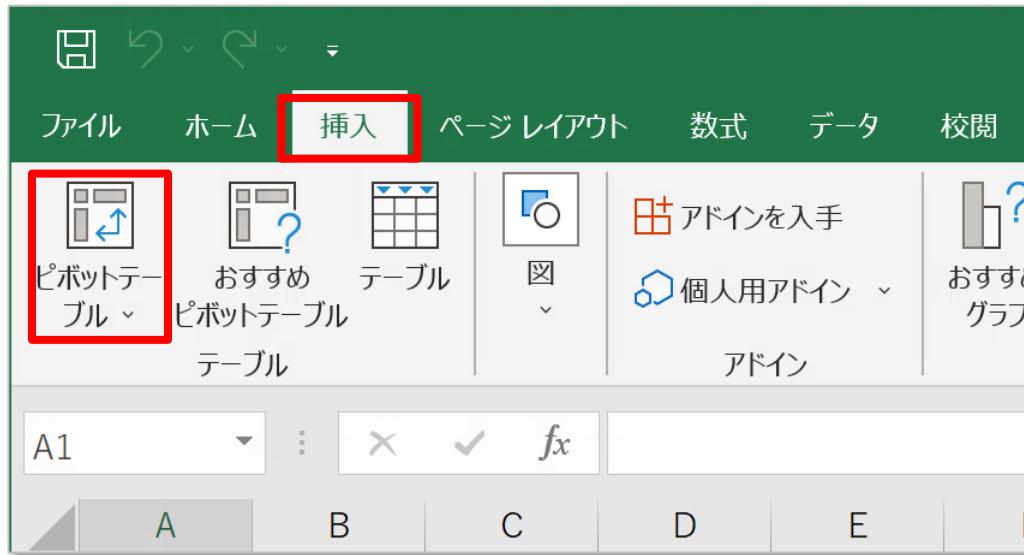
- データをよく知る
 - データ件数や構成比を集計 → データを理解する
 - 旅行目的別の人気エリアは?
 - 同伴者別の人気エリアは?
 - 数値評価による人気エリアの差異は?
- テーマを設定する
 - 解決すべき課題を決める → 分析目的を明確にする
 - 数値評価が低い原因是?
 - 高評価の施設に学ぶ改善点は?
- テキスト分析に取り組む
 - これら課題を解決するために、テキスト分析を実施

参考 — EXCEL を使った集計

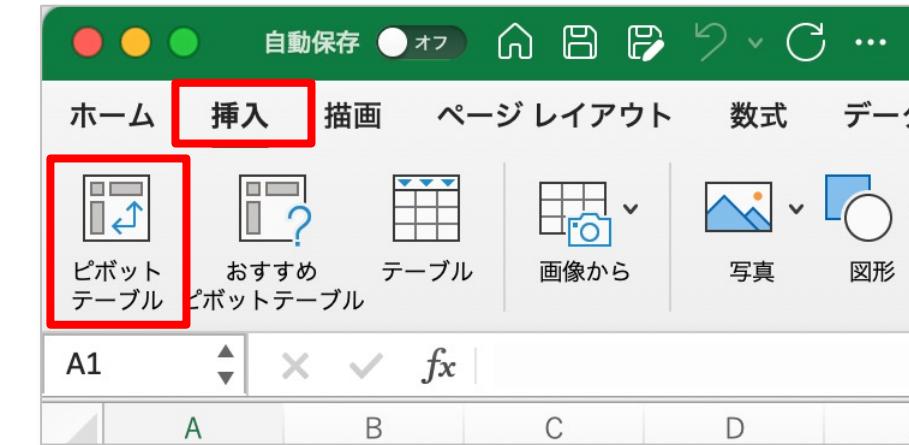
● EXCEL のピボットテーブルを使ってデータを集計する

- ① ファイル **rakuten-1000-2023-2024.xlsx** を開く
- ② A～R 列を選択し、ピボットテーブルを作成する
- ③ [挿入] タブ [テーブル] グループの [ピボットテーブル] ボタンをクリックする

Windows



Mac



データ理解 — 集計例

①件数 (エリア別)

行ラベル	個数 / コメン
■ A_レジャー	5000
01_登別	1000
02_草津	1000
03_箱根	1000
04_道後	1000
05_湯布院	1000
■ B_ビジネス	5000
06_札幌	1000
07_名古屋	1000
08_東京	1000
09_大阪	1000
10_福岡	1000
総計	10000

②投稿者の傾向 (年代別x性別)

行ラベル	個数 / コメン	列ラベル			
		男性	女性	na	総計
10代		0.00%	0.05%	0.00%	0.05%
20代		0.70%	1.16%	0.00%	1.86%
30代		1.94%	2.60%	0.00%	4.54%
40代		4.54%	3.76%	0.00%	8.30%
50代		7.95%	4.17%	0.00%	12.12%
60代		6.34%	1.96%	0.00%	8.30%
70代		1.38%	0.36%	0.00%	1.74%
80代		0.07%	0.06%	0.00%	0.13%
na		0.00%	0.00%	62.95%	62.95%
120代		0.00%	0.01%	0.00%	0.01%
総計		22.92%	14.13%	62.95%	100.00%

③投稿者の傾向 (性別xカテゴリ別)

行ラベル	個数 / コメン	列ラベル		
		A_レジャー	B_ビジネス	総計
男性		22.52%	23.32%	22.92%
女性		15.98%	12.28%	14.13%
na		61.50%	64.40%	62.95%
総計		100.00%	100.00%	100.00%

データ理解 — 集計例

④投稿者の傾向 (性別xカテゴリーエリア別)

個数 / コメント	列ラベル	A_レジャー 集計										B_ビジネス 集計			総計
		01_登別	02_草津	03_箱根	04_道後	05_湯布院	06_札幌	07_名古屋	08_東京	09_大阪	10_福岡				
男性	A_レジャー	24.10%	22.80%	16.20%	27.10%	22.40%	22.52%	26.40%	25.30%	21.30%	20.50%	23.10%	23.32%	22.92%	
女性	A_レジャー	16.00%	16.00%	17.20%	11.40%	19.30%	15.98%	12.70%	12.90%	11.70%	11.80%	12.30%	12.28%	14.13%	
na	A_レジャー	59.90%	61.20%	66.60%	61.50%	58.30%	61.50%	60.90%	61.80%	67.00%	67.70%	64.60%	64.40%	62.95%	
総計		100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	

⑤投稿者の傾向 (年代別xカテゴリーエリア別)

個数 / コメント	列ラベル	A_レジャー 集計										B_ビジネス 集計			総計
		01_登別	02_草津	03_箱根	04_道後	05_湯布院	06_札幌	07_名古屋	08_東京	09_大阪	10_福岡				
10代	A_レジャー	0.10%	0.00%	0.10%	0.00%	0.00%	0.04%	0.00%	0.00%	0.10%	0.20%	0.00%	0.06%	0.05%	
20代	A_レジャー	0.80%	3.40%	2.60%	1.60%	1.90%	2.06%	2.00%	2.50%	0.70%	1.60%	1.50%	1.66%	1.86%	
30代	A_レジャー	4.20%	5.00%	4.40%	4.00%	6.30%	4.78%	3.40%	4.90%	4.80%	4.30%	4.10%	4.30%	4.54%	
40代	A_レジャー	8.30%	9.00%	7.30%	6.70%	10.20%	8.30%	8.20%	8.80%	7.10%	9.20%	8.20%	8.30%	8.30%	
50代	A_レジャー	13.40%	11.60%	8.70%	13.30%	11.60%	11.72%	15.10%	12.00%	11.80%	9.80%	13.90%	12.52%	12.12%	
60代	A_レジャー	10.70%	8.10%	7.60%	10.20%	9.90%	9.30%	9.00%	8.50%	7.10%	6.10%	5.80%	7.30%	8.30%	
70代	A_レジャー	2.30%	1.60%	2.30%	2.70%	1.50%	2.08%	1.20%	1.50%	1.40%	1.00%	1.90%	1.40%	1.74%	
80代	A_レジャー	0.30%	0.10%	0.40%	0.00%	0.30%	0.22%	0.10%	0.00%	0.00%	0.10%	0.00%	0.04%	0.13%	
120代	A_レジャー	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.10%	0.00%	0.00%	0.00%	0.00%	0.02%	0.01%	
na	A_レジャー	59.90%	61.20%	66.60%	61.50%	58.30%	61.50%	60.90%	61.80%	67.00%	67.70%	64.60%	64.40%	62.95%	
総計		100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	

データ理解 — 集計例

⑥投稿者の傾向 (同行者別xカテゴリ-エリア別)

個数 / コメント 行ラベル	列ラベル	A_レジャー 集計										B_ビジネス 集計			総計
		01_登別	02_草津	03_箱根	04_道後	05_湯布院	06_札幌	07_名古屋	08_東京	09_大阪	10_福岡				
一人		24.80%	13.70%	14.00%	44.40%	13.10%	22.00%	57.90%	64.30%	65.60%	57.70%	53.50%	59.80%		40.90%
家族		63.10%	64.30%	66.10%	42.70%	69.60%	61.16%	30.90%	24.10%	23.40%	29.40%	32.10%	27.98%		44.57%
恋人		4.80%	14.60%	11.20%	4.90%	8.30%	8.76%	4.40%	4.20%	4.10%	4.40%	3.90%	4.20%		6.48%
友達		5.30%	5.70%	7.10%	5.50%	7.80%	6.28%	4.40%	4.40%	4.80%	6.80%	7.60%	5.60%		5.94%
仕事仲間		1.50%	0.70%	0.50%	1.80%	0.60%	1.02%	1.90%	2.50%	1.30%	1.40%	2.40%	1.90%		1.46%
その他		0.50%	1.00%	1.10%	0.70%	0.60%	0.78%	0.50%	0.50%	0.80%	0.30%	0.50%	0.52%		0.65%
総計		100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%		100.00%

⑦投稿者の傾向 (年代別xカテゴリ-エリア別)

個数 / コメント 行ラベル	列ラベル	A_レジャー 集計										B_ビジネス 集計			総計
		01_登別	02_草津	03_箱根	04_道後	05_湯布院	06_札幌	07_名古屋	08_東京	09_大阪	10_福岡				
5		41.90%	48.40%	48.90%	49.60%	67.90%	51.34%	41.70%	36.70%	36.90%	41.90%	38.50%	39.14%		45.24%
4		41.30%	36.10%	36.90%	36.50%	22.50%	34.66%	41.90%	47.30%	41.70%	41.10%	40.90%	42.58%		38.62%
3		9.90%	9.90%	7.80%	9.20%	4.90%	8.34%	11.60%	12.00%	14.30%	11.70%	12.60%	12.44%		10.39%
2		4.30%	3.40%	4.30%	2.90%	3.20%	3.62%	3.20%	2.70%	4.60%	3.60%	5.00%	3.82%		3.72%
1		2.60%	2.20%	2.10%	1.80%	1.50%	2.04%	1.60%	1.30%	2.50%	1.70%	3.00%	2.02%		2.03%
総計		100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%		100.00%

• レジャーの高評価5は、湯布院が多く、登別が少ない

データ理解 — 集計例

⑧-a 数値評価の平均 (エリア別×数値評価別)

行ラベル	平均 / サービス	平均 / 立地	平均 / 部屋	平均 / 設備・アメニ	平均 / 風呂	平均 / 食事	平均 / 総合
■ A_レジャー	4.25	4.25	4.13	4.05	4.29	4.29	4.30
01_登別	4.07	4.21	3.95	3.90	4.34	4.08	4.16
02_草津	4.23	4.22	4.07	3.97	4.32	4.20	4.25
03_箱根	4.24	4.12	4.18	4.05	4.29	4.33	4.26
04_道後	4.19	4.41	4.07	4.00	4.03	4.19	4.29
05_湯布院	4.51	4.28	4.37	4.35	4.46	4.61	4.52
■ B_ビジネス	3.98	4.30	4.01	3.88	3.74	4.05	4.13
06_札幌	4.05	4.30	4.09	3.93	3.79	4.12	4.19
07_名古屋	4.00	4.25	4.04	3.89	3.75	4.03	4.15
08_東京	3.93	4.38	3.94	3.82	3.70	3.99	4.06
09_大阪	4.01	4.35	4.05	3.93	3.82	4.06	4.18
10_福岡	3.93	4.24	3.96	3.84	3.64	4.01	4.07

⑧-b 数値評価の平均 (カテゴリ別×数値評価別)

行ラベル	平均 / サービス	平均 / 立地	平均 / 部屋	平均 / 設備・アメニ	平均 / 風呂	平均 / 食事	平均 / 総合
A_レジャー	4.25	4.25	4.13	4.05	4.29	4.29	4.30
B_ビジネス	3.98	4.30	4.01	3.88	3.74	4.05	4.13

データ理解 — 集計例

⑨-a 数値評価の平均 (20~30代, 性別)

行ラベル	平均 / サービス	平均 / 立地	平均 / 部屋	平均 / 設備・アメニ	平均 / 風呂	平均 / 食事	平均 / 総合
■ A_レジャー	4.48	4.32	4.35	4.33	4.42	4.46	4.46
男性	4.38	4.23	4.30	4.18	4.38	4.48	4.44
女性	4.55	4.37	4.38	4.42	4.45	4.44	4.47
■ B_ビジネス	4.18	4.36	4.19	4.10	3.93	4.41	4.28
男性	4.16	4.33	4.13	4.10	3.93	4.35	4.23
女性	4.19	4.39	4.24	4.10	3.93	4.45	4.32

⑨-b 数値評価の平均 (40~50代, 性別)

行ラベル	平均 / サービス	平均 / 立地	平均 / 部屋	平均 / 設備・アメニ	平均 / 風呂	平均 / 食事	平均 / 総合
■ A_レジャー	4.29	4.33	4.15	4.10	4.32	4.35	4.34
男性	4.23	4.29	4.10	4.05	4.29	4.33	4.32
女性	4.38	4.38	4.21	4.17	4.36	4.37	4.36
■ B_ビジネス	4.00	4.30	4.07	3.91	3.79	4.09	4.17
男性	3.93	4.27	4.00	3.85	3.73	4.07	4.12
女性	4.13	4.36	4.19	4.02	3.89	4.13	4.26

⑨-c 数値評価の平均 (60~80代, 性別)

行ラベル	平均 / サービス	平均 / 立地	平均 / 部屋	平均 / 設備・アメニ	平均 / 風呂	平均 / 食事	平均 / 総合
■ A_レジャー	4.23	4.24	4.08	3.98	4.29	4.26	4.31
男性	4.21	4.20	4.05	3.92	4.26	4.27	4.27
女性	4.27	4.36	4.16	4.12	4.38	4.26	4.39
■ B_ビジネス	3.95	4.30	4.01	3.80	3.78	3.87	4.14
男性	3.91	4.24	3.98	3.77	3.75	3.79	4.13
女性	4.19	4.59	4.15	3.96	3.94	4.23	4.21

データ理解 — 集計結果の整理

観点	データの特徴	テキスト分析時に注意すべき点
年代別・性別	<ul style="list-style-type: none">約60%が年代や性別を表明していない・・	<ul style="list-style-type: none">レビュー観点がある年代や性別に偏っている可能性・・
目的別	<ul style="list-style-type: none">レジャーは家族が多い、ビジネスは一人が多い・・	<ul style="list-style-type: none">レビューの観点が性別によって偏っている可能性・・
数値評価 (総合)	<ul style="list-style-type: none">旅行目的によらず評価は高め・・	<ul style="list-style-type: none">コメントが好評価に偏っている可能性・・
数値評価 (項目ごと)	<ul style="list-style-type: none">レジャーの評価は、風呂や食事 > 設備や部屋・・	<ul style="list-style-type: none">旅行目的によって評価の観点や重みが異なっている可能性・・
全体	<ul style="list-style-type: none">・・・	

データ理解 — 集計例

①件数 (エリア別)

行ラベル	個数 / コメン
■ A_レジャー	5000
01_登別	1000
02_草津	1000
03_箱根	1000
04_道後	1000
05_湯布院	1000
■ B_ビジネス	5000
06_札幌	1000
07_名古屋	1000
08_東京	1000
09_大阪	1000
10_福岡	1000
総計	10000

②投稿者の傾向 (年代別x性別)

行ラベル	個数 / コメン	列ラベル	男性	女性	na	総計
10代			0.00%	0.05%	0.00%	0.05%
20代			0.70%	1.16%	0.00%	1.86%
30代			1.94%	2.60%	0.00%	4.54%
40代			4.54%	3.76%	0.00%	8.30%
50代			7.95%	4.17%	0.00%	12.12%
60代			6.34%	1.96%	0.00%	8.30%
70代			1.38%	0.36%	0.00%	1.74%
80代			0.07%	0.06%	0.00%	0.13%
na			0.00%	0.00%	62.95%	62.95%
120代			0.00%	0.01%	0.00%	0.01%
総計			22.92%	14.13%	62.95%	100.00%

③投稿者の傾向 (性別xカテゴリ別)

行ラベル	A_レジャー	B_ビジネス	総計
男性	22.52%	23.32%	22.92%
女性	15.98%	12.28%	14.13%
na	61.50%	64.40%	62.95%
総計	100.00%	100.00%	100.00%

- 男性の投稿者が多い(女性の倍程度) → 男性の観点によるコメントが多い

- 無回答(na)の中の分布が、表明した層と異なる(ある年代や性別に偏っている)可能性もある

データ理解 — 集計例

④投稿者の傾向 (性別xカテゴリーエリア別)

個数 / コメント	列ラベル	A_レジャー 集計					B_ビジネス 集計					総計	
行ラベル		01_登別	02_草津	03_箱根	04_道後	05_湯布院	06_札幌	07_名古屋	08_東京	09_大阪	10_福岡		
男性	A_レジャー	24.10%	22.80%	16.20%	27.10%	22.40%	22.52%	26.40%	25.30%	21.30%	20.50%	23.10%	22.92%
女性	A_レジャー	16.00%	16.00%	17.20%	11.40%	19.30%	15.98%	12.70%	12.90%	11.70%	11.80%	12.30%	14.13%
na		59.90%	61.20%	66.60%	61.50%	58.30%	61.50%	60.90%	61.80%	67.00%	67.70%	64.60%	62.95%
総計		100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

- 男女差は、レジャーに比べてビジネスが大きい
- 男女差がレジャーで大きいのは道後(次いで登別や草津も大きい)

⑤投稿者の傾向 (年代別xカテゴリーエリア別)

個数 / コメント	列ラベル	A_レジャー 集計					B_ビジネス 集計					総計		
行ラベル		01_登別	02_草津	03_箱根	04_道後	05_湯布院	06_札幌	07_名古屋	08_東京	09_大阪	10_福岡			
10代	A_レジャー	0.10%	0.00%	0.10%	0.00%	0.00%	0.04%	0.00%	0.00%	0.10%	0.20%	0.00%	0.06%	0.05%
20代	A_レジャー	0.80%	3.40%	2.60%	1.60%	1.90%	2.06%	2.00%	2.50%	0.70%	1.60%	1.50%	1.66%	1.86%
30代	A_レジャー	4.20%	5.00%	4.40%	4.00%	6.30%	4.78%	3.40%	4.90%	4.80%	4.30%	4.10%	4.30%	4.54%
40代	A_レジャー	8.30%	9.00%	7.30%	6.70%	10.20%	8.30%	8.20%	8.80%	7.10%	9.20%	8.20%	8.30%	8.30%
50代	A_レジャー	13.40%	11.60%	8.70%	13.30%	11.60%	11.72%	15.10%	12.00%	11.80%	9.80%	13.90%	12.52%	12.12%
60代	A_レジャー	10.70%	8.10%	7.60%	10.20%	9.90%	9.30%	9.00%	8.50%	7.10%	6.10%	5.80%	7.30%	8.30%
70代	A_レジャー	2.30%	1.60%	2.30%	2.70%	1.50%	2.08%	1.20%	1.50%	1.40%	1.00%	1.90%	1.40%	1.74%
80代	A_レジャー	0.30%	0.10%	0.40%	0.00%	0.30%	0.22%	0.10%	0.00%	0.00%	0.10%	0.00%	0.04%	0.13%
120代	A_レジャー	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.10%	0.00%	0.00%	0.00%	0.00%	0.02%	0.01%
na		59.90%	61.20%	66.60%	61.50%	58.30%	61.50%	60.90%	61.80%	67.00%	67.70%	64.60%	64.40%	62.95%
総計		100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

- あくまでも投稿者の傾向であって、旅行者の実態と一致するは限らない

データ理解 — 集計例

⑥投稿者の傾向 (同行者別)

- レジャーの中で一人が多いのは道後 →道後はもはや仕事で行く場所 (性別でも男性が多い)

- レジャーは家族が多く、ビジネスは一人が多い →出張は複数より単独が多い

個数 / コメント 行ラベル	A_レジャー 集計												B_ビジネス 集計			総計
	01_登別	02_草津	03_箱根	04_道後	05_湯布院	06_札幌	07_名古屋	08_東京	09_大阪	10_福岡						
一人	24.80%	13.70%	14.00%	44.40%	13.10%	22.00%	57.90%	64.30%	65.60%	57.70%	53.50%	59.80%				40.90%
家族	63.10%	64.30%	66.10%	42.70%	69.60%	61.16%	30.90%	24.10%	23.40%	29.40%	32.10%	27.98%				44.57%
恋人	4.80%	14.60%	11.20%	4.90%	8.30%	8.76%	4.40%	4.20%	4.10%	4.40%	3.90%	4.20%				6.48%
友達	5.30%	5.70%	7.10%	5.50%	7.80%	6.28%	4.40%	4.40%	4.80%	6.80%	7.60%	5.60%				5.94%
仕事仲間	1.50%	0.70%	0.50%	1.80%	0.60%	1.02%	1.90%	2.50%	1.30%	1.40%	2.40%	1.90%				1.46%
その他	0.50%	1.00%	1.10%	0.70%	0.60%	0.78%	0.50%	0.50%	0.80%	0.30%	0.50%	0.52%				0.65%
総計	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%				100.00%

⑦投稿者の傾向 (年代別xカタ

- 数値評価は、目的によらず高め →好評価しか投稿しない偏りがあるの可能性にも注意

- 高評価5は、レジャーがビジネスよりもやや多い

個数 / コメント 行ラベル	A_レジャー 集計												B_ビジネス 集計			総計
	01_登別	02_草津	03_箱根	04_道後	05_湯布院	06_札幌	07_名古屋	08_東京	09_大阪	10_福岡						
5	41.90%	48.40%	48.90%	49.60%	67.90%	51.34%	41.70%	36.70%	36.90%	41.90%	38.50%	39.14%				45.24%
4	41.30%	36.10%	36.90%	36.50%	22.50%	34.66%	41.90%	47.30%	41.70%	41.10%	40.90%	42.58%				38.62%
3	9.90%	9.90%	7.80%	9.20%	4.90%	8.34%	11.60%	12.00%	14.30%	11.70%	12.60%	12.44%				10.39%
2	4.30%	3.40%	4.30%	2.90%	3.20%	3.62%	3.20%	2.70%	4.60%	3.60%	5.00%	3.82%				3.72%
1	2.60%	2.20%	2.10%	1.80%	1.50%	2.04%	1.60%	1.30%	2.50%	1.70%	3.00%	2.02%				2.03%
総計	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%				100.00%

- レジャーの高評価5は、湯布院が多く、登別が少ない

- ビジネスの高評価5は、札幌・大阪が多く、名古屋・東京がやや少ない

データ理解 — 集計例

⑧-a 数値評価の平均 (エリア別×数値評価別)

行ラベル	平均 / サービス	平均 / 立地	平均 / 部屋	平均 / 設備・アメニ	平均 / 風呂			
■ A_レジャー	4.25	4.25	4.13	4.05	4.29	4.29	4.29	4.30
01_登別	4.07	4.21	3.95	3.90	4.34	4.08	4.16	
02_草津	4.23	4.22	4.07	3.97	4.32	4.20	4.25	
03_箱根	4.24	4.12	4.18	4.05	4.29	4.33	4.26	
04_道後	4.19	4.41	4.07	4.00	4.03	4.19	4.29	
05_湯布院	4.51	4.28	4.37	4.35	4.46	4.61	4.52	
■ B_ビジネス	3.98	4.30	4.01	3.88	3.74	4.05	4.13	
06_札幌	4.05	4.30	4.09	3.93	3.74	3.99	4.06	
07_名古屋	4.00	4.25	4.04	3.89	3.74	4.06	4.18	
08_東京	3.93	4.38	3.94	3.82	3.70	3.99	4.06	
09_大阪	4.01	4.35	4.05	3.93	3.82	4.06	4.18	
10_福岡	3.93	4.24	3.96	3.84	3.64	4.01	4.07	

- レジャーは、風呂や食事が、設備や部屋に比べて高評価

- 湯布院は、レジャーの中で、軒並み高評価が多い

⑧-b 数値評価の平均 (カテゴリ別×数値評価別)

行ラベル	平均 / サービス	平均 / 立地	平均 / 部屋	平均 / 設備・アメニ	平均 / 風呂	平均 / 食事	平均 / 総合
A_レジャー	4.25	4.25	4.13	4.05	4.29	4.29	4.30
B_ビジネス	3.98	4.30	4.01	3.88	3.74	4.05	4.13

- レジャーもビジネスも立地が評価される
- ビジネスは、立地がその他に比べて高評価

データ理解 — 集計例

⑨-a 数値評価の平均 (20~30代, 性別)

行ラベル	平均 / サービス	平均 / 立地	平均 / 部屋	平均 / 設備・アメニティ	平均 / 風呂	平均 / 食事	平均 / 総合
■ A_レジャー	4.48	4.32	4.35	4.33	4.22	4.46	4.46
男性	4.38	4.23	4.30	4.18	4.38	4.48	4.44
女性	4.55	4.37	4.38	4.42	4.45	4.44	4.47
■ B_ビジネス	4.18	4.36	4.19	4.10	3.93	4.41	4.28
男性	4.16	4.33	4.13	4.10	3.93	4.35	4.23
女性	4.19	4.39	4.24	4.10	3.93	4.45	4.32

- 20~50代はレジャーに対する風呂や食事、サービスの評価が概ね高い

⑨-b 数値評価の平均 (40~50代, 性別)

行ラベル	平均 / サービス	平均 / 立地	平均 / 部屋	平均 / 設備・アメニティ	平均 / 風呂	平均 / 食事	平均 / 総合
■ A_レジャー	4.29	4.33	4.15	4.10	4.32	4.35	4.34
男性	4.23	4.29	4.10	4.05	4.29	4.33	4.32
女性	4.38	4.38	4.21	4.17	4.36	4.37	4.36
■ B_ビジネス	4.00	4.30	4.07	3.91	4.00	4.33	4.28
男性	3.93	4.27	4.00	3.89	4.00	4.33	4.28
女性	4.13	4.36	4.19	4.02	4.02	4.37	4.32

- 年齢が高くなるに連れてレジャー・ビジネスとも評価が厳しくなる

⑨-c 数値評価の平均 (60~80代, 性別)

行ラベル	平均 / サービス	平均 / 立地	平均 / 部屋	平均 / 設備・アメニティ	平均 / 風呂	平均 / 食事	総合
■ A_レジャー	4.23	4.24	4.08	3.98	4.29	4.26	4.31
男性	4.21	4.20	4.08	3.98	4.27	4.27	4.27
女性	4.27	4.36	4.15	3.98	4.26	4.39	4.39
■ B_ビジネス	3.95	4.30	3.98	3.77	3.75	3.79	4.14
男性	3.91	4.24	3.98	3.77	3.75	3.79	4.13
女性	4.19	4.59	4.15	3.96	3.94	4.23	4.21

- 女性は立地に対する評価が高めで、60~80代で顕著（→期待が高い）

- グループワーク (~20:40)
 - データ集計によって発見した、データセットに関する特徴や傾向、テキスト分析時に注意すべき点について、グループ内で討論する
 - 前ページの表を参考に、集計結果から得られた知見を整理する

(再掲) 数値評価で違いを見るのは難しい

【再掲】⑧-a 数値評価の平均 (エリア別×数値評価別)

行ラベル	平均 / サービス	平均 / 立地	平均 / 部屋	平均 / 設備・アメニ	平均 / 風呂			
■ A_レジャー	4.25	4.25	4.13	4.05	4.29	4.29	4.29	4.50
01_登別	4.07	4.21	3.95	3.90	4.34	4.08	4.16	
02_草津	4.23	4.22	4.07	3.97	4.32	4.20	4.25	
03_箱根	4.24	4.12	4.18	4.05	4.29	4.33	4.26	
04_道後	4.19	4.41	4.07	4.00	4.03	4.19	4.29	
05_湯布院	4.51	4.28	4.37					4.52
■ B_ビジネス	3.98	4.30	4.01					4.13
06_札幌	4.05	4.30	4.09					4.19
07_名古屋	4.00	4.25	4.04	3.89	3.75			4.15
08_東京	3.93	4.38	3.94	3.82	3.70	3.99		4.06
09_大阪	4.01	4.35	4.05	3.93	3.82	4.06		4.18
10_福岡	3.93	4.24	3.96	3.91	3.64	4.01		4.07

- ユーザーの8割が4~5の評価、1~2をつけない→本音が見えない

- 同じ点数でもテキストを見れば差異があるかも

- すべての項目に回答する→どこに注目しているかよくわからない

【再掲】⑧-b 数値評価の平均 (カテゴリ別×数値評価別)

行ラベル	平均 / サービス	平均 / 立地	平均 / 部屋	平均 / 設備・アメニ	平均 / 風呂	平均 / 食事	平均 / 総合
A_レジャー	4.25	4.25	4.13	4.05	4.29	4.29	4.30
B_ビジネス	3.98	4.30	4.01	3.88	3.74	4.05	4.13

辻井康一 and 津田和彦「テキストマイニングを用いた宿泊レビューからの注目情報抽出方法」, デジタルプラクティス 3.4 (2012): 289-296.

【再掲】⑧-b 数値評価の平均 (カテゴリ別×数値評価別)

行ラベル	平均 / サービス	平均 / 立地	平均 / 部屋	平均 / 設備・アメニ	平均 / 風呂	平均 / 食事	平均 / 総合
A_レジャー	4.25	4.25	4.13	4.05	4.29	4.29	4.30
B_ビジネス	3.98	4.30	4.01	3.88	3.74	4.05	4.13

● 数値評価のみから違いを見つけるのは難しい！！

- ・ ユーザーの 8割が 4~5 の評価, 1~2をつけない
- ・ ユーザーは 注目の有無に関係なくすべての項目に回答

→ レジャーとビジネスでは、評価すべき項目も異なることを確認した

→ テキストと対応付ければ、同じ点数でも差異があることを確認した

レポート課題

- 以下を PDF ファイルで提出してください
 - データ集計により作成した「集計表」のキャプチャ (P.XX~XX) ※ページ番号は各スライド右下に記載
 - 作成した「集計結果の整理」の表 (P.XX) ※ページ番号は各スライド右下に記載
- ※ 何らかの事情で上記2つを提出できない場合、本日の講義の感想を文章で記述してください

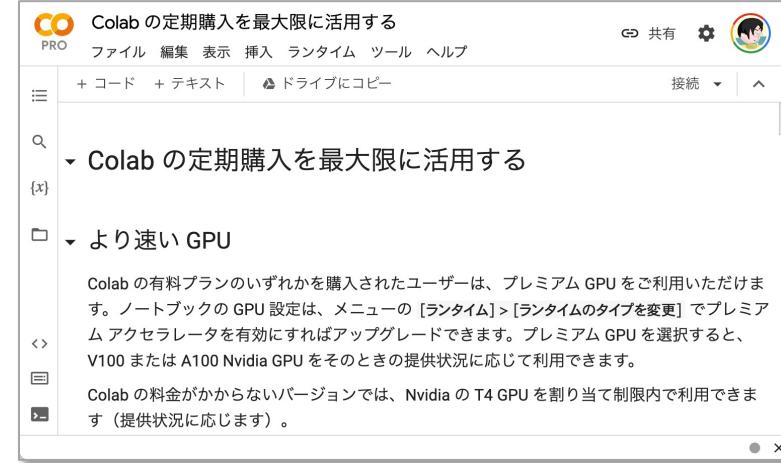
レポート形式	提出先	期限
PDF	manaba	次回～18:20

(再掲) 無償で利用できる機械学習環境

- 近年、機械学習の教育・研究を目的とした研究用ツールが相次いで登場

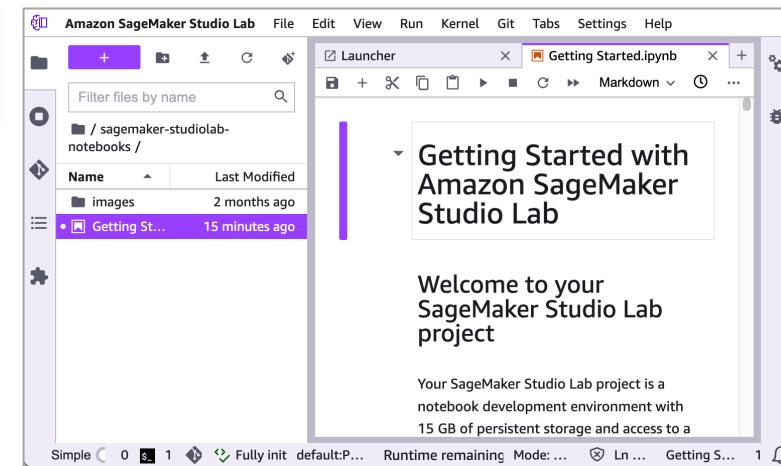
 Colaboratory

<https://colab.research.google.com>



 Amazon SageMaker Studio Lab

<https://studiolab.sagemaker.aws/>



演習で使用
↓

	Colab(無償版)	Studio Lab
GPU	T4(16GB)	T4(16GB)
最長実行時間	12時間	CPU:12時間 GPU:4時間
メモリ	12GB	15GB
ディスク	CPU:100GB GPU:78GB	15GB (永続化)
ターミナル	×	○
ランタイムの保存と再開	×	○
費用	無償	無償
その他	Googleアカウントが必要	AWSアカウントは不要 (クレカ不要)

(再掲) SageMaker Studio Lab のアカウント作成

- <https://studiolab.sagemaker.aws/> にログインして、アカウントを作成してください

amazon
SageMaker Studio Lab

Sign in Request account

Request account

Request a free Amazon SageMaker Studio Lab account.

Enter your email*

Enter your first name

Enter your last name

Select your country

Enter your company or organization name

Select your occupation

Why are you interested in Amazon SageMaker Studio Lab?

Enter referral code **XXX-XXXX**

Submit request

アカウント作成手順

1. [アカウント作成フォーム](#)からアカウントの申し込みを行う
注意: リファラルコードをアカウント作成フォームに忘れずに入力ください (**受講者限定**です)
2. 「Account request confirmed ...」のメールを受信し、メール内のリンクからアカウントを作成する
→ リクエストの受付はすぐにメールが届きます
3. 「Verify your email ...」のメールを受信し、メール内のリンクからメールアドレスを認証する
→ リファラルコードを利用している場合は2~3分以内に結果が届きます
4. 「Your account is ready ...」のメールを受信する
→ これで「Sign in」できます

※ リファラルコードの有効期間: 2024/6/27 ~ 2024/7/12