

Introduction

Can refer to this

本研究では、日本国内のグルメレビュー投稿サイトに掲載されている飲食店のレビュー情報を分析対象とする。具体的には、「評価点数が高い店舗」と「レビュー件数が多い店舗」の二つのグループを設定し、それぞれ 1,200 店舗ずつ、合計 2,400 店舗を抽出した。取得したデータは、各店舗のレビュー全文および評価点数で構成される。レビュー全文については、ストップワードの除去、記号・絵文字の削除、全角半角の統一などの前処理を行い、解析用のデータセットを作成する。本研究で用いるレビュー件数の総数は、表 1 に示すとおりである。評価点数上位店舗（1,200 店舗）のレビュー総数は 329,166 件、レビュー件数上位店舗（1,200 店舗）のレビュー総数は 613,453 件である。

表 1 分析対象の店舗数およびレビュー件数

分類	店舗数	総レビュー件数
評価点上位店舗	1,200	329,166
レビュー件数上位店舗	1,200	613,453

Research Steps

Stage	Step / Command	Key operations	Main output
1	Data collection (data_collection.py)	Scrape / import high-rating & most-commented CSV → save to data/raw/	*.csv
2	Clean & tokenise (main.py --stage clean)	NFKC → strip URL/emoji/digits → Janome tokenizer → keep 名詞・形容詞 → stop-words → join tokens	corpus_clean_text.parquet
3	TF-IDF vectorise (src/preprocessing.py)	TfidfVectorizer (1-2 gram, 20 k vocab) → sparse matrix X	—
4	K-means clustering (main.py --stage cluster)	Test k = 2-10, pick best Silhouette (k=9 or 10) → add "cluster" column	corpus_with_clusters.parquet + kmeans.pkl
5	Top terms per cluster (src/topic_inspection.py)	CountVectorizer → top-N uni/bi-grams (+ negative filter)	CSV / on-screen list
6	Rule-based sentiment (sentiment_analysis.py)	Add flag_negative column (keyword match)	updated parquet
7	8-topic LDA (main.py --stage lda)	Fit LDA → manual topic names → dominant_topic column	lda.pkl + updated parquet
8	Cluster × Topic heat-map (make_cluster_topic_heatmap.py)	Crosstab cluster vs dominant_topic → heat-map PNG	cluster_topic_heatmap.png

9	Corpus Δ analysis (analysis_summary.py)	Topic prevalence HR vs MC → topic_summary.c sv + Δ bar chart	delta_chart.png
10	Representative quotes (one-liner + summarise_cluster)	Random 3-5 reviews per cluster saved to TXT	representative_quotes_*.txt

High Rating Results

Silhouette vs k (can use line graph for this)

-silhouette result

k= 2 → silhouette=0.0041
k= 3 → silhouette=0.0047
k= 4 → silhouette=0.0052
k= 5 → silhouette=0.0055
k= 6 → silhouette=0.0057
k= 7 → silhouette=0.0060
k= 8 → silhouette=0.0060
k= 9 → silhouette=0.0063
k=10 → silhouette=0.0061

✓ k-means (k=9) saved → high_rating_kmeans_k9.pkl

-cluster result

Refer to high_rating_cluster_top_terms.csv

This is chatgpt explain d, means main theme for each cluster

Cluster (件数)	キーワード上位10	主なテーマ
C0	高い・残念・小さい・料理	価格／量ギャップ：値段は高いのに量が少ない不満
C1	時間・平日・待ち・遅い	行列・待機：平日でも並び、提供が遅い
C2	予約・電話・困難・時間	予約難易度＆遅延：電話が繋がらず、開始が遅い
C3	東京・たい・美味しい	高期待×高評価：首都圏来訪で品質検証
C4	人気・高い・予約	人気急騰店：点数・人気ワードが集中
C5	寿司・大将・美味しい	寿司コース体験：大将の握り・出汁への賛辞
C6	ラーメン・醤油・スープ	ラーメン比較：高級和食 vs B級評価
C7	ログ・評価・Award	Tabelog点数・受賞談義
C8	パン・シェフ・フレンチ	パン/洋食比較：シェフ名を上げて評価

Below is the example of reviews of each top words

==== Cluster 00 top words: 高い, ない, 残念, 料理, 訪問, コース, 美味しい, 評価, 今回, 小さい

- ・天気がみぞれから雪に変わるほど気温も天候も良くなかったこの日、3月の営業告知日の最終日だったことと、ホワイトデーということで家族から強い要望で買いに行きました。ですが、「あ、今日営業しているじゃん…
- ・2020年10月に新規開店した「焼肉割烹」の「雪月花銀座」へ。全席カウンターで調理の様子を見せてます。総額27500円で和牛を腹いっぱい食べる事ができて満足！聞いたら牛肉350g程度出した…
- ・ここ1週間で行ったすし屋3軒のうちのひとつMikami limited50さんは、いつも本当に美味しいです今回も会員さんに貸し切り会に連れてきてもらいました【鮪】【鰯】(き…

==== Cluster 01 top words: 時間, 高い, 平日, 訪問, 開店, 待ち, 到着, ない, 遅い, 残念

- ・2018年5月25日(金): 味玉鶏の中華そば + 和え玉こちらは2018年5月22日開店。開店から4日目の訪問です。場所はJR横浜線・大口駅西口から南西へ約100m。この日の到着は午後1時…
- ・朝10時少し前にお店に到着。既に20人以上の行列でした。11時から記帳開始との事でしたが、列が長くなった為10時半頃には記帳開始。ギリギリセーフで最終の14時40分が取れました。枠が取れなかつた方…
- ・麺屋さくら井味玉冷やしらあ麺お茶漬けセット(夏季限定)三鷹駅11時55分到着、14人待ち12時30分着席ごちそうさまでした！#ぐるめ#グルメ#グルメ好き…

==== Cluster 02 top words: 予約, 高い, ない, 訪問, 残念, 時間, コース, 料理, 電話, 遅い

- ・新聞記事で見て一度食べてみたい！と思った料理をいただきに3年振りにお伺いいたしました。電話予約の時から抜群の対応の良さは流石ですね。安心できるお店です。メニューを決める前にご挨拶の前菜・カンパ…
- ・渋谷のイベント前に、サクッと予約が取れたのでフレンチを食べに行きました。フレンチの有名店との事で、とても美味しかったです！スペシャリテの卵とサワラのやつが美味しかった！機会があ…

- ◊初訪問◊10月ランチ貸切会にてご予約やお声かけいただきたり、ご一緒出来た皆様どうもありがとうございました♡何年も気になってたお店やっと行けました場所は亀島方面...

==== Cluster 03 top words: 高い, たい, もの, 東京, ない, 料理, 残念, 美味しい, 訪問, 評価

- タンとせんまいが超美味しかった。いい焼肉屋さんって何食べても美味しい。ユッケも美味しかったしロースは優勝です。大好きな銀座の東京園は超えなくとも最高の焼肉屋さん。次回は一年半後の予約で気が遠くなります。笑
- ミルクとイチゴ。次回はピスタチオとナッツ系食べたい!
- 本日東京ランチは、低温で揚げる白いとんかつで有名な成蔵さんへとんかつ定食を食べに伺いました(*^ω^*)その日仕入れた銘柄豚の部位を自由に組み合わせ出来るとんかつ定食が、主力商品であり...

==== Cluster 04 top words: 人気, 高い, ない, 訪問, 予約, 人気高い, 時間, 残念, 評価, 料理

- 松阪駅より徒歩10~12分の場所にあるすき焼き屋さん和田金。松阪牛のすき焼きと言えば、こちらのお店無くして語れない老舗の人気店。2024年しゃぶしゃぶ・すき焼き百名店だけでなく、2017年~...
- 人気店のなかはらさん、初訪問! 焼肉が好きなので、色々なお店に食べに行っていますが、総合的に微妙。。コースはその日によって値段が変わるらしく、よく分からなかったので、単品で注文。...
- 北参道駅から歩いて5分の超人気スイーツ店。イートインは完全予約制で1日5組限定。スイーツのコースをいただけます。この日は栗づくしでした。・栗の水羊羹・栗とみるく・栗のタルト・栗のパフェ...

==== Cluster 05 top words: 寿司, 高い, ない, 予約, 残念, 訪問, 美味しい, 大将, 評価, 高い寿司

- ・カラソコロン カラソコロン心地よい下駄の音が鳴り響く店内祇園の名店に初訪問です。『八寸』の八寸! 焼物、揚げ物、お寿司、酢の物などが盛り込んである世界観は完成度が高く素敵でした。…
- ・福島県白河にある小判寿司さんに伺いました。小判寿司さんは、The Tabelog Award 2019、2018、2017 BRONZE 受賞店です。アクセスは、水郡線磐城棚倉駅または、中豊駅…
- ・常盤寿司さんの新発田店さんにお邪魔しました。カウンター11席ほどのお店でこの日は半数が外国の方でした。新潟の魚を使用した握りは勿論のこと、蟹玉をイメージしたズワイガニの料理など独創的なツマミも楽し…

==== Cluster 06 top words: ラーメン, 高い, 醤油, 訪問, ない, スープ, 評価, 高い ラーメン, 名店, 平日

- ・[場所]西日暮里駅 徒歩 10 分[時間・行列]2017/10/7(土) 19:00 行列なし[食べたもの・値段]醤油ラーメン 1200 円[感想]西日暮里のラーメン…
- ・平日、友人に連れて行ってもらいました。本来であれば多くを語りたいのですが、今回は詳細はあえて言いません(笑)12:00 の回で伺い醤油ラーメン、塩つけ麺をいただきました。もう…
- ・美味しい中華そばを求めて、北小金の名店『中華蕎麦 柳』に初来店です。13時過ぎ、7~8人並んでいましたが、10分程で入店できました。割と回転が早い! 煮干しと貝だしの『醤油と塩』 …

==== Cluster 07 top words: ログ, 高い, 評価, 評価高い, ログ 評価, 名店, 訪問, 点数, 料理, 予約

- ・《食べログブロンズランクイング》評価: 3.93 《2022 Tabelog Bronze Award》《2021 Tabelog フレンチ WEST 百名店》ミシュランの1星初めて日本…
- ・『食べログ 洋食 百名店』を何度も獲得されている「洋食の店 橋本」さんに伺いました。場所は熊本市中央区。市電「通町筋駅」から徒歩4分程の建物の2階にありました。(奥まったところを通るので分かり…

・銀座でランチのあとこちらのお店に。こちらのお店は、食べログスイツ百名店 TOKYO 2022 選出店です。この度は初選出おめでとうございます。店内のカフェスペースはお休みでしたので、店員…

==== Cluster 08 top words: パン, 高い, シェフ, 料理, ない, 残念, 美味しい, 訪問, フレンチ, 評価

・神楽坂にある、谷昇シェフのレストランル・マンジュ・トゥーへ牛込神楽坂駅から少し歩いて住宅街を入ると、おしゃれな外観に照明が点いていたのを発見しそこだ~と小走りになりました入口に入る…

・1,936 文字★01. ~シェフからの最初の贈り物~ 定点トマト餡（ミニトマト、ホーリーバジル、キプロス塩）02. 血を抜いて新生姜と蜂蜜で GV した跳子ツチクジラ 天然苦荷のマリネゴーヤの…

・大阪天満のミシュラン 2 つ星の日本料理店「お料理 宮本」。食べログも高評価だし、大阪のフレンチの 3 つ星店として有名な「HAJIME」のシェフが近隣でお薦めの料理店 3 つのうちのひとつとして紹介していたのを…

-lda result

Refer to high_rating_lda_top_terms.csv

LDA 8 トピック		
Topic	上位語	主題
0	料理・京都・ミシュラン	格式・地域料理
1	ない・訪問・美味しい	評価コメント混在
2	カレー・ケーキ・パン	デザート/カジュアル比較
3	ラーメン・待ち・行列	行列/時間ロス
4	ログ・シェフ・フレンチ	受賞・シェフ権威
5	予約・困難・時間	予約ストレス
6	店内・雰囲気・ランチ	雰囲気・立地
7	寿司・大将・握り	寿司コース

-insight of high rating

主要インサイト (High-Rating)	
観点	インサイト
① 価格×量ギャップ	C0 に「高い・小さい・残念」が集中。質より量を求める層が一定数。
② 行列・時間ロス	C1 + LDA Topic 3 で待機不満が顕著。人気店ゆえの課題。
③ 予約困難	C2 + Topic 5 が 予約 UX の弱さを示唆。
④ 権威ワード vs 体験評価	C7 (Award/点数) と C3 (純粋な味評価) が二極化。
⑤ ジャンル横断比較	ラーメン・カレー・パンと比較され「コスパ」話題へ拡散。

Cluster x topic modeling heatmap

Refer to high_rating_cluster_topic_heatmap.png

This is the explanation of the png

クラスタ ID	優勢トピック (比率>0.30)	主な意味合い
1	Topic 3 (キュー&待機) 0.85	高評価店でも行列・待機体験を詳細に語る層。
2	Topic 4 (受賞・話題性) 0.59	ミシュラン・シェフ・メディア露出など“権威・ハイブ”評価クラスタ。
3	Topic 1 (再訪意欲・味評価) 0.38	味・感動・「また来たい」系コメントが集中。
6	Topic 3 (キュー&待機) 0.80	C1 とは別の行列グループ：ロケーション違い？時間帯違い？
5	Topic 7 (寿司・握り) 0.36	寿司コースやネタの質へフォーカス。
4	Topic 2 0.21 + Topic 3 0.23	カレー・パン比較と待機ストレスが混在するハイブリッド層。
0 / 8	すべて 0.25 以下	価格・ポーション・雰囲気などが薄く分散した“その他”

-insight of lda vs topic high rating

主要インサイト (High-Rating コーパス)	
観点	インサイト
1. 価格 × 量ギャップ	クラスター 0 に「高い・小さい・残念」が集中。質より量を求める層が一定数存在し、★4.0 以上でもコスパ不満は残る。
2. 行列・時間コスト	C1 と C6 は Topic 3 (行列・待機) が 0.80 以上。高得点店でも待機体験がレビュー動機となり、感情の振れ幅が大きい。
3. 予約困難	C2 + Topic 5 が「予約電話・枠争奪」の不満を示す。UX の弱さがブランド価値を下げるリスク。
4. 権威ワード vs 体験評価の二極化	C2 (格式・受賞) と C3 (純粹味評価) が対照的：前者は賞歴・シェフ名指し、後者は味そのものを称賛。
5. ジャンル横断比較	寿司 (C5) ・ラーメン (C6) ・カレー／パン (C4) など、ジャンル別に「コスパ／話題性」論点が拡散。ラーメンと寿司は高評価でも議論が活発。

Most commented Results

Silhouette vs k (can use line graph for this)

-silhouette result

k= 2 → silhouette = 0.0047

k= 3 → silhouette = 0.0059

k= 4 → silhouette = 0.0066

k= 5 → silhouette = 0.0075

k= 6 → silhouette = 0.0077

k= 7 → silhouette = 0.0077

k= 8 → silhouette = 0.0086

k= 9 → silhouette = 0.0089

k=10 → silhouette = 0.0101

✓ k-means (k = 10) saved → most_commented_kmeans_k10.pkl-cluster result

Refer to most_commented_cluster_top_terms.csv

Cluster (件数)	キーワード上位10	主なテーマ
C0	高い・残念・小さい・料理	価格／量ギャップ：値段は高いのに量が少ない不満
C1	時間・平日・待ち・遅い	行列・待機：平日でも並び、提供が遅い
C2	予約・電話・困難・時間	予約難易度＆遅延：電話が繋がらず、開始が遅い
C3	東京・たい・美味しい	高期待×高評価：首都圏来訪で品質検証
C4	人気・高い・予約	人気急騰店：点数・人気ワードが集中
C5	寿司・大将・美味しい	寿司コース体験：大将の握り・出汁への賛辞
C6	ラーメン・醤油・スープ	ラーメン比較：高級和食 vs B級評価
C7	ログ・評価・Award	Tabelog点数・受賞談義
C8	パン・シェフ・フレンチ	パン/洋食比較：シェフ名を上げて評価

Below is the example of reviews of each top words

==== Cluster 00 top words: ブログ, かわいい, 那覇空港, 食堂, かわいい 土産, かわいい 見た目, ご飯, ご飯 今日, ひどい, ひどい ブログ

- 2020/8/5、こちらは関西では超有名店の尼崎にある“ぶたのほし”さんです。詳細は開店当初に訪店させていただいたときの私のブログをご確認ください☆彡とにかくお店がきれいなんです！！！（2...
- 【2015年115杯目】投稿したブログを振り返ってみて分かったが、1年10か月ぶりに東銀座の「函館らーめん船見坂」に訪問。(時間が経つのが早い……Orz)※過去の訪問記はコチラ。今...
- とまそんのブログ: <https://www.ramentabete.com/entry/2020/10/29/000000> とまそんのYouTube: <https://youtu.be/Cuj5kF...>

==== Cluster 01 top words: 餃子, 高い, 残念, 焼き, 注文, 遅い, 小さい, 浜松, 時間, 宇都宮

- 休日のお昼過ぎ。少し待ってから入店。2種ある焼き餃子のハーフ＆ハーフの定食と水餃子を頼む。石松餃子は普通のおいしい餃子。軽くっていっぱい食べられそう。肉餃子はちょっと味付けが変わってる感じ。水餃子はツ...
- 友人のすすめでこちらに。店内はこじんまりとしていて、隣のお客さんとの距離が近いです。私は麻婆豆腐と餃子を注文。麻婆豆腐は激辛との表記はなかったと思いましたが、かなりの辛口でした。...
- 平日 2235 待ち一名海老ワンタンそば(¥1,100)ネギ(+¥100)餃子 5個(¥500)ビアバーに飲みに行きコース料理を食べたのですが、〆はなく、こちらに流れてきました。夜遅い...

==== Cluster 02 top words: 東京, 高い, ない, ラーメン, 遅い, 訪問, 時間, 残念, 名店, 評価

- 息子の春休みに合わせて、ワタクシも久しぶりに平日休暇が取れたので東京浅草でランチすることにしました。平日でも浅草寺はかなりの混雑ですね。仲見世通りは人で溢れかえっていますが、通りを1, 2本ズ...
- [最寄駅] JR 東京駅丸の内南口→徒歩10分程度 [住所] 東京都千代田区丸の内2-6-1丸の内ブリックスクエア1F 今日は仕事帰りに有名すぎる「エシレ」さんに寄ってお土産を買って...

- ・東京駅内のラーメンストリートにあります。店内はカウンター席とテーブル席があります。看板メニューのとろり白湯塩玉らーめんを頂きました。スープが美味しかったです！ただその日は初夏を感じさせ…

==== Cluster 03 top words: ラーメン, 高い, スープ, ない, 評価, 訪問, 評価高い, 醤油, 残念, 時間

- ・ルクア大阪 B2 バルチカにあるラーメン屋名前の通り本店は銀座でこちらの店舗も食べログラーメン WEST 百名店に選出されている名店カウンター席×9、4名掛けテーブル席×3、2名掛けテーブル席×1 のレ…

- ・寒い日は味噌ラーメンしかも、極寒の日は赤味噌ラーメンと言うことで、「粹な一生」さんでランチスープは、塩味も甘味も丁度いい塩梅の味噌味…深紅の色が食欲を掻き立てるトロミも程よく、…

- ・六本木で博多ラーメン！ レビューが結構良かったので気になって行ってきた！ 昼下がりに入店。店内には2,3人のお客様。頼んだのはオーソドックスなラーメンと明太子ごはんのセット。らーめん75…

==== Cluster 04 top words: カレー, 高い, ない, 残念, 訪問, ランチ, 注文, 評価, 遅い, 人気

- ・Hot Spoon 金曜日はカレーの日なので30%オフ！濃厚牛スジ煮込みカレー大盛り 中辛五穀ご飯スープだけで6時間具材を煮込んでスパイスを作つてそこから更に8時…

- ・都営地下鉄「神保町駅」A5出口から徒歩5分 JR「御茶ノ水駅」御茶ノ水橋口から徒歩7分カレーバン名店だけでも11軒あり、その10番目に訪れたお店。普段、あまりスープカレーは手を出さない…

- ・気軽に美味しいカレーを頂けるポンガラカレーさん、3月中旬にリニューアルで「カントテグランデカレー」になるそうです。そこで遠征中の駆込み訪問となりました。◆選べる3種のカレ-プレ…

==== Cluster 05 top words: 高い, ない, 残念, 美味しい, 遅い, ランチ, 訪問, 小さい, 時間, 評価

- ・「oggeti さん、花は後から届いた自分より美しい花に嫉妬するのです。」＊＊＊皆さんは、思いを寄せている方へバラの切り花を贈ることはありますか？ 「いやー、私は花より団子だね…
- ・わざわざ北海道まで行ったのに、予定していたイベントが突然中止。さてさて、何をすればいいのでしょうか？？と露頭に迷っていた脳裏に、最初に思い浮かんだワードが「彩未」の二文字。前回がお初で、また戻って来たい…
- ・プチ旅行帰りに寄らせていただきました！ 学生時代から今に至るまで、いろいろな思い出があり本店の特製つけ麺を食べるたびに思い出されます…そんなことはさておき、特製つけ麺は絶品でおすすめです。そ…

==== Cluster 06 top words: カツ, 定食, ロース, 高い, 注文, 残念, ヒレ, ランチ, フライ, ロースカツ

- ・毎年百名店受賞しているすごいとんかつ屋さんはつけーーーん！！久しづりにとんかつ屋さんへ 18 時オープンで 10 分くらいに着きましたがすぐに満卓になってる！・ロースカツ定食注文¥1,70…
- ・・青森旅行の初日のランチに訪問。連休の 12 時少し前の到着で、30 分ほど並んで入店。・ホタテフライ定食とホタテの貝味噌煮定食と単品でホタテバター焼きをオーダー。・フライは甘みが一番出てホクホクで…
- ・予算:1500 円特選もちぶたロースカツ定食 130g ランチの開店時に行かないとすぐ売り切れてしまう特選。牡蠣の季節が終わったからチャンス！ と思ったらオープン即満員、並びも絶えずだった…

==== Cluster 07 top words: 訪問, 時間, 高い, 平日, 待ち, 遅い, 店内, 行列, 遅い時間, 到着

- ・大阪来るたびに寄っているお店です。時間帯によってはかなり待ちますが、今回は 2 回目でサクッとイートインに入れました。こちらの店は持ち帰りももちろん出来るのですが、まあどうせなら焼きたてのほうが私は好…
- ・手元の記録を見ると 35 回目、きっと、後 10 回くらいは行っているひよっこさんへ。今日はオープン直後の 11 時 10 分に到着、ちょうど満席でしたが食券購入と同時に一席空いて着席しました。提供待ちは私…

- 2025 年 2 月初訪問。北海道ツアーノの帰り、飛行機まで 2 時間あったので、1 軒だけ列びが凄い「海老そば一幻」さんへ。折り返した列びは 15 人を超えていた、どのくらい待つか不安になる。30 分程並んで…

==== Cluster 08 top words: 予約, 高い, ない, 時間, 訪問, 残念, 遅い, ランチ, コース, 人気

- 前に一度来たことあるなあと入口で思い出した。今回は先輩が 8,000 円の飲み放題付きコースを予約してくれています、着席してまずはビールをお願い。細かいことは忘れてしましたが、出てく…

- 遠方より来訪した知人への土産有名東京銘菓をと思い、電話予約して行きました。電話していなければ 1 時間強待たなければならない状態でしたが、直ぐ受け取ることができました。昔ながらの技法で 1 枚 1 枚丁…

- パンケーキが食べたい気分で、銀座の bills さんに行ってきました！ 前日に食べログで予約しようと思いましたが、15 時からしか空いておらず…コースの選択も必要だったので、直接お店に行くことに。…

==== Cluster 09 top words: うどん, 高い, カレー, ない, 遅い, カレー うどん, 訪問, 残念, 評価, 注文

- とり天うどん温 520 円かけうどん温 350 円五反田駅近くでさつと食べられるうどん百名店。夜遅く、飲み後でも空いてる安心感。シンプルイズベスト。染み渡る美味しさです。肉うど…

- 一際行列が目立つ人気のうどん屋さん。行列に惑わされずに、まずは券売機で食券を買いましょう。どこにも説明がないけど、買ってから並ぶシステムです。並ぶの無理って人はテイクアウトなども可能！ …

- 久しぶりに来店昼時なので相変わらずの行列客捌きマスターの下どんどん客は入れ替わり立ち替わり回転は早いので、トータル 15 分程度あります今日は稻庭うどん（ざる、中）にミニ…

-lda result

Refer to most_commented_lda_top_terms.csv

LDA 8 トピック		
Topic	上位語	主題
0	訪問・待ち・店内・満席	行列/待機
1	時間・行列・仕事	時間ロス
2	カレー・うどん・スパイス	カレー/粉物比較
3	大阪・中華そば・老舗	関西老舗ラーメン
4	スープ・ラーメン・濃厚	ラーメン品質
5	ランチ・メニュー・予約	ランチ&予約可否
6	店内・雰囲気・カフェ	雰囲気/テイクアウト
7	ラーメン・餃子・味噌	餃子×味噌ラーメン

Insight of most commented

主要インサイト (Most-Commented)	
観点	インサイト
① 価格×コスパ不満	C5 に「高い・残念・小さい・遅い」が集中。コメント総量の過半が“価格・量・提供速度”へのネガティブ反応。
② 行列・待機ストレス	C1 / C7 で“平日でも待ち” “行列で時間ロス” が突出。LDA Topic 0・1 も待機語で占有。
③ 予約 UX 不満	C8 に「予約・遅い・人気」—予約ページや電話の煩雑さ+開始遅延が多数報告。
④ 専門店比較 & コスパ競合	餃子 (C1)、ラーメン (C3)、カレー (C4)、とんかつ (C6)、うどん (C9) とジャンルごとに “名店 vs 本店” の値ざろ比較が盛ん。
⑤ 旅行/SNS 拡散型レビュー	C0 に “かわいい・那覇空港・土産・ブログ”—観光客が SNS 映え・手土産性を重視して投稿。
⑥ 雰囲気&テイクアウト	LDA Topic 6 で “店内・雰囲気・カフェ・テイクアウト”—空間体験やテイクアウト利用への言及が高評価店より多い。

Cluster x topic modeling heatmap

Refer to most_commented_cluster_topic_heatmap.png

This is the explanation of the png

クラスタID	優勢トピック (比率>0.30)	主な意味合い
0	Topic 3 (大阪・中華そば・昔ながら) 0.63	関西系ラーメン／中華そばの比較・語り合い。
1	Topic 7 (ラーメン+餃子) 0.66	ラーメンと餃子をセットで評価し価格・味を議論。
2	混合 (最大0.21)	明確な主題なし—軽微な不満や雑談の寄せ集め。
3	Topic 4 (スープ品質) 0.31 + Topic7 0.36	「スープ濃度」と「ラーメン+餃子」両方を論点。
4	Topic 2 (カレー／うどん／スパイス) 0.67	カレーライス・カレーうどんの値頃感や辛さ比較。
5	Topic 5 (ランチ&予約UX) 0.21 + Topic 6 (雰囲気) 0.20	ランチメニューのコスパと予約ストレスが共存。
6	Topic 4 (スープ品質) 0.35	こちらも“スープ出来”に集中するラーメン群。
7	Topic 0 (行列・待機) 0.67	平日でも行列・待ち時間に対する怒りが主題。
8	Topic 0 (行列・待機) 0.37	やや軽めの行列不満クラスタ。
9	Topic 2 (カレー／うどん) 0.63	Curry/Udon 論争その2。

-insight of lda vs topic most commented

主要インサイト (Most-Commented コーパス)	
観点	インサイト
1. 行列・待機ストレスが最大テーマ	C7・C8 で Topic 0 (行列/待機) が 0.37~0.67 と支配的。レビュー総数の 10 %以上 を占めており、コメント量を押し上げる主要因になっている。
2. ジャンル別コスパ論争が顕著	カレー／うどん系 (C4・C9) と ラーメン+餃子系 (C1・C3・C6) がそれぞれ Topic 2・Topic 7 で濃度 0.4~0.7。価格・味バランスの比較が盛んで、高評価コーパスで目立った「格式・権威」は希薄。
3. 予約 UX とランチ価格がセットで語られる	C5 では Topic 5 (ランチ&予約 UX) 0.21 + Topic 6 (雰囲気) 0.20。「ランチは手續だが予約が面倒」「遅延が発生」など、体験と UI/UX の両方が不満点として挙がる。
4. 雜多クラスタの存在	C2 はいずれのトピックも 0.21 以下。主題が薄いロングテールの“その他”コメントが集まっており、多様な軽微不満や雑談がここに吸収されている。

Search engine

-insight of lda vs topic most commented

Can create a search function to search reviews based on the words (like filter function)

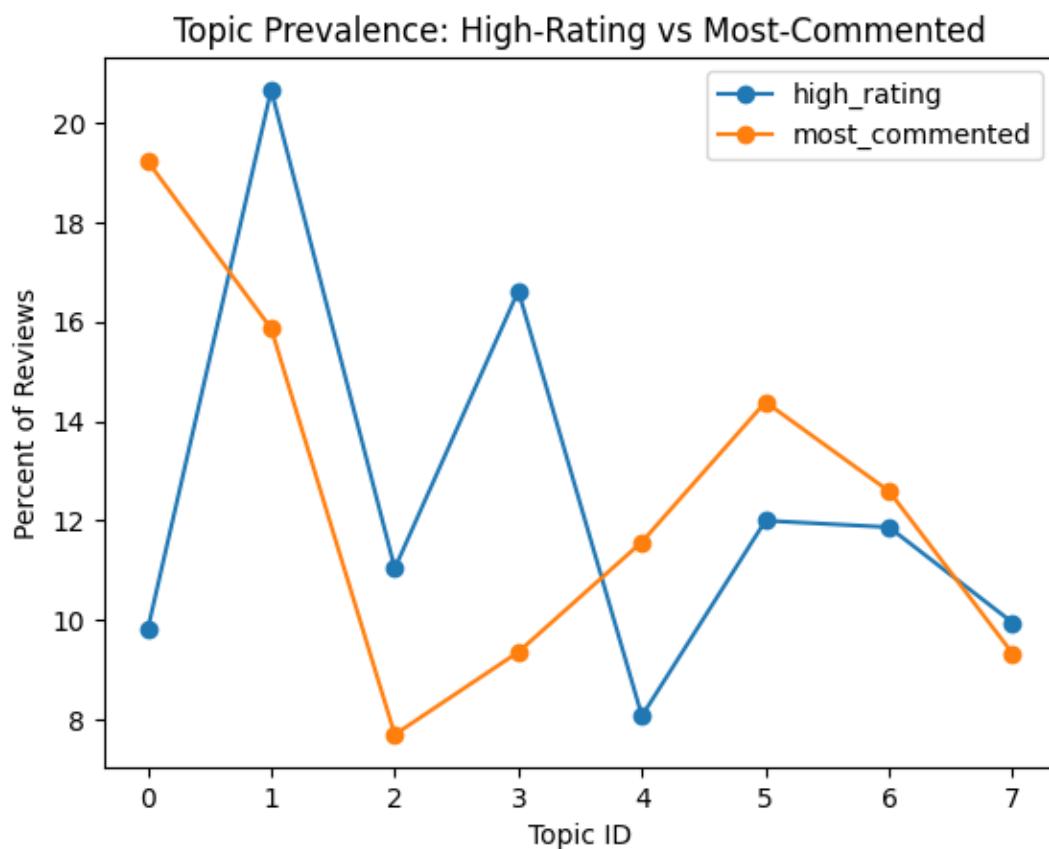
Can refer to query_reviews.py

Mixed topics

— Topic Prevalence: High-Rating vs Most-Commented —

Step 8 トピック出現率比較

高評価コーパス・最多コメントコーパスそれぞれについて、
dominant_topic を持つレビューの割合を算出（`value_counts(normalize=True)`）。
得られた8トピックの比率を Matplotlib で折れ線プロットし、
両コーパスの出現傾向を可視化した（Fig. X）。



Explanation of the graph

— Topic Prevalence: High-Rating vs Most-Commented —		
観点	読み取り結果	意味合い
1. トップ差分	Topic 0 Cuisine & Prestige が Most-Commented 19 % ▶ High-Rating 10 % (+9 pt)	コメントが多い店ほど「京都・ミシュラン・名店」など“肩書き／話題性”で語られる
2. 高評価側ピーク	Topic 1 Desire & Evaluation と Topic 3 Queue & Wait が High-Rating 21 % / 17 % ▶ Most 16 % / 9 %	高評価店のレビューは「また来たい・料理評価」に加え“行列体験”を詳細に記述
3. 予約・UX不満	Topic 5 Reservation Frustration Most-Commented 14 % ▶ High-Rating 12 %	コメント数を押し上げるのは予約ストレスの共有
4. Awards / Hype	Topic 4 Awards & Hype Most-Commented 11.6 % ▶ High-Rating 8 %	話題店の「受賞歴」「シェフ人気」がコメントを誘発
5. 残差の小さい領域	Topic 6 (Ambiance & Location) や Topic 7 (Sushi Variety) は両コーパスで差 < 1 pt	雰囲気・寿司コースの言及率は“評価”より“コメント量”に左右されにくい

-analysis summary

Topic ID	Theme	HighRating (%)	MostCommented (%)	Delta (%)
0	Cuisine & Prestige	9.8	19.2	-9.4
1	Desire & Evaluation	20.7	15.9	4.8
2	Dessert & Comparisons	11	7.7	3.3
3	Queue & Wait Times	16.6	9.4	7.3
4	Awards & Hype	8.1	11.6	-3.5
5	Reservation Frustration	12	14.4	-2.4
6	Ambiance & Location	11.9	12.6	-0.7
7	Menu Variety (Sushi)	9.9	9.3	0.6

This table refer to delta_chart.png

This is the insight of the table

主要インサイト（日本語）
<p>1. 行列・待機ストレスは高評価店に集中 高評価店ほど「2時間待った」「番号券がない」など待機体験への言及が多く、出現率で +7 pt。</p>
<p>2. 格式・受賞歴はコメント増には効くが評価点には直結しない Most-Commented 側では老舗／百名店／ミシュラン等の語が HR より ~10 pt高いが、星評価は伴わない。</p>
<p>3. 「味を語る」vs「UXを語る」軸が明確 HR：味・比較・評価の言語化が多い。 MC：予約・行列・価格 UX の体験共有が多い。</p>
<p>4. デザート・比較レビューは高評価店で厚み パン・ケーキ・カレーなど“こだわり勢”による味比較投稿が HR 側で多い。</p>

Conclusion

実務・提言	
テーマ	改善・活用アイデア
行列 / 待機	整理券発行、リアルタイム待ち時間表示でストレス緩和 → 星評価低下を防ぐ。
権威ワード活用	受賞やメディア露出時は SNS で話題を拡散。ただし予約案内を同時に改善し UX 苦情を回避。
予約 UX	Most-Commented で顕著な予約不満。自動返信・前払いシステム導入で行列発生を抑制。
レビュー誘導	「待ち時間」だけでなく味の感想を促すレビューキャンペーンで、評価質を向上。

観点	結論の要点
① 高評価店の特徴	味・クオリティへの評価語 (美味しい／比較) が多い一方、行列・待機ストレスも最も語られる。→「★が高い = 並ぶ価値はあるが、UX 改善余地が大きい」
② コメント数が多い店の特徴	格式・話題性 (老舗／百名店／ミシュラン) や予約困難の話題がコメント堆をけん引。→「パズリ要素は強いが、味そのものの評価語は相対的に少ない」
③ 両者の最大ギャップ	Queue & Wait Times (+7 pt) ★店のほうが待機不満を多く生みやすい。Cuisine & Prestige (-9 pt) コメント数店で格式ワードが圧倒的。
④ 共通する改善課題	予約 UX と行列管理。どちらのコーパスでも不満トピックが残る。
⑤ 実務示唆	- ★店：整理券／進捗表示で待機ストレスを減らし、星評価を維持。- コメント狙い店：受賞・SNS 露出に合わせて予約導線を強化し、ネガティブ爆発を防ぐ。