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．A screenshot of a computer

AI-generated content may be incorrect.

**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.csv + 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

A screenshot of a black screen

AI-generated content may be incorrect.

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 】【 鱚（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。枠が取れなかった方...

• 麺屋 さくら井味玉冷やしらぁ麺 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ったきました。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りたり、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、色ん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

• カランコロン カラン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キング》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、食べログスイーツ百名店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を入る...

• 1,936文字★01.～シェフからの最初の贈り物～ 定点 トマト飴（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した銚子ツチクジラ 天然茗荷のマリネ 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

A screenshot of a black screen

AI-generated content may be incorrect.

-insight of high rating

A screenshot of a black screen

AI-generated content may be incorrect.

Cluster x topic modeling heatmap

Refer to high\_rating\_cluster\_topic\_heatmap.png

This is the explanation of the png

A screenshot of a black screen

AI-generated content may be incorrect.

-insight of lda vs topic high rating

A screenshot of a computer

AI-generated content may be incorrect.

**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

A screenshot of a black screen

AI-generated content may be incorrect.

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を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名がけテーブル席×3、2名がけテーブル席×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、「粋な一生」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人のお客さん。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ってそこから更に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分カレー百名店だけでも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？？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っけーーーん！！久しぶ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ょっとこさんへ。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人を超えている、どのくらい待つのか不安になる。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ぁ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して、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いましたが、出てく...

• 遠方より来訪した知人への土産に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っ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分程度でありつけます今日は稲庭うどん（ざる、中）にミニ...

-lda result

Refer to most\_commented\_lda\_top\_terms.csv

A screenshot of a black screen

AI-generated content may be incorrect.

Insight of most commented

A screenshot of a computer

AI-generated content may be incorrect.

Cluster x topic modeling heatmap

Refer to most\_commented\_cluster\_topic\_heatmap.png

This is the explanation of the png

A screenshot of a black screen

AI-generated content may be incorrect.

-insight of lda vs topic most commented

A screenshot of a computer

AI-generated content may be incorrect.

**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）。

A graph with blue and orange lines

AI-generated content may be incorrect.

Explanation of the graph

A screenshot of a black screen

AI-generated content may be incorrect.

-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

A screenshot of a computer

AI-generated content may be incorrect.

**Conclusion**

A screenshot of a black screen

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.