

thesis result

high rating

-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

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

-lda result

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

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「コスバ」話題へ拡散。

most comment

-kmeans

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

Cluster (件数)	キーワード上位10	主なテーマ
C0 (1 116)	ブログ・かわいい・那覇空港	旅行ブログ系・可愛い土産紹介
C1 (15 068)	餃子・浜松・宇都宮・高い・遅い	餃子専門店の価格&待ち時間不満
C2 (27 655)	東京・高い・ラーメン・名店	首都圏の有名店比較／コスバ論争
C3 (73 960)	ラーメン・スープ・醤油	ラーメン品質・評価談義
C4 (28 550)	カレー・ランチ・注文	カレー専門店の価格&行列
C5 (319 130)	高い・残念・遅い・小さい	価格／量／待ち 最大の不満塊
C6 (12 856)	カツ・定食・ロース・ヒレ	とんかつ系：部位×値段比較
C7 (90 630)	訪問・時間・平日・行列	行列・待機ストレス集中クラスター
C8 (23 138)	予約・高い・遅い	予約煩雑+遅延への不満
C9 (21 350)	うどん・カレーうどん	カレーうどん行列／値段話題

-lda result

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

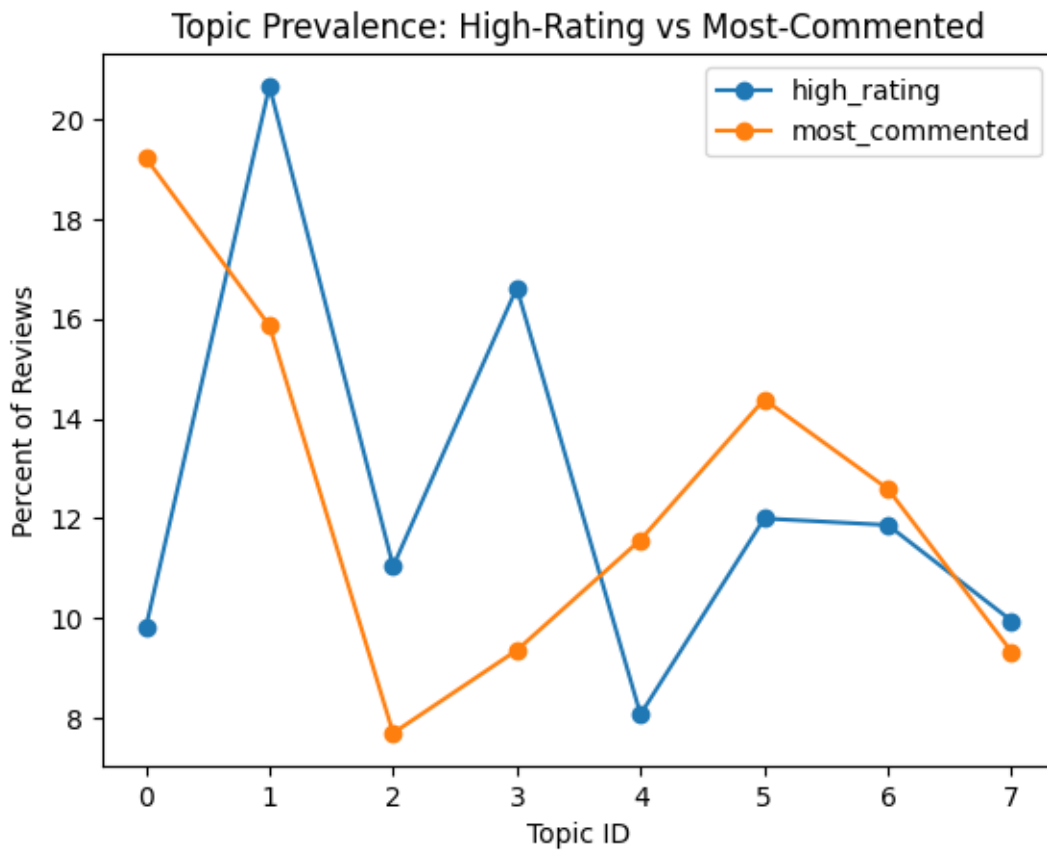
-insight

主要インサイト (Most-Commented)	
観点	インサイト
● 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。
● 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。
● 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。
● 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」の値ごろ比較が盛ん。
● 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。
● 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。

— 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）。



— 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

/Users/kevinsoon/Documents/GitHub/thesis_split/data/processed/
high rating cluster topic heatmap.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"

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 以上でもコスバ不満は残る。
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」の不满を示す。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 (純粋味評価) 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「コスバ／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。

[/Users/kevinsoon/Documents/GitHub/thesis_split/data/processed/
most commented cluster topic heatmap.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バ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。

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. ジャンル別コスト競争が顕著	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。「ランチは手頃だが予約が面倒」「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. 雑多クラスタの存在	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。

-analysis summary
Markdown Table:

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

トピックテーマ	High-Rating 出現率	Most-Commented 出現率	Δ = HR - MC	読み取りポイント
Cuisine & Prestige (格式・名声)	9.8 %	19.2 %	-9.4 pt	コメント数が多い店では「老舗」「メディア露出」など格式ワードが大幅に多い。
Desire & Evaluation (行きたい・評価語り)	20.7 %	15.9 %	+4.8 pt	高評価コーパスでは「美味しい／行きたい／感想評価」が多く語られる。
Dessert & Comparisons (デザート・他店比較)	11.0 %	7.7 %	+3.3 pt	高評価店でパン・ケーキなど趣味性の高い比較レビューが目立つ。
Queue & Wait Times (行列・待ち時間)	16.6 %	9.4 %	+7.3 pt	★最も差が大きい。高評価店は待機ストレスの話題が多い＝“並ぶほど旨い”構造。
Awards & Hype (受賞・話題性)	8.1 %	11.6 %	-3.5 pt	Most-Commented で百名店／ミシュラン等の権威語が多い。
Reservation Frustration (予約トラブル)	12.0 %	14.4 %	-2.4 pt	コメント数の多い店で「予約取れない・電話繋がらない」不満が強い。
Ambiance & Location (雰囲気・立地)	11.9 %	12.6 %	-0.7 pt	両コーパスでほぼ同水準。
Menu Variety (Sushi) (寿司ネタ多様性)	9.9 %	9.3 %	+0.6 pt	差はごく小さい。

主要インサイト（日本語）

- 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。
- 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。
- 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。
- 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。

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増をけん引。→「バズり要素は強いが、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。