



Bôkan Restaurant Rebranding
Data Science Consulting Project – March 2020

OUTLINING YOUR GOAL



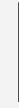
→ Rebrand the restaurant

- Innovation in the choice of cuisines and the menu
- Space rebranding based on analysis of data collected



→ Increase profit margin

- Breakeven Estimation
- Expansion of Client base
- Securing higher performance



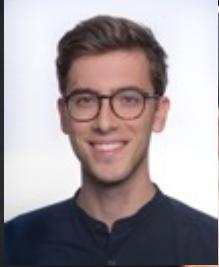
TEAM



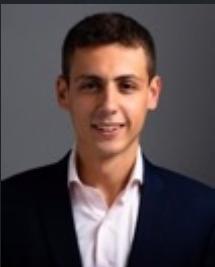
TOMMY TRAN



ISMAIL MAJJAD



THOMAS DE
MAREUIL



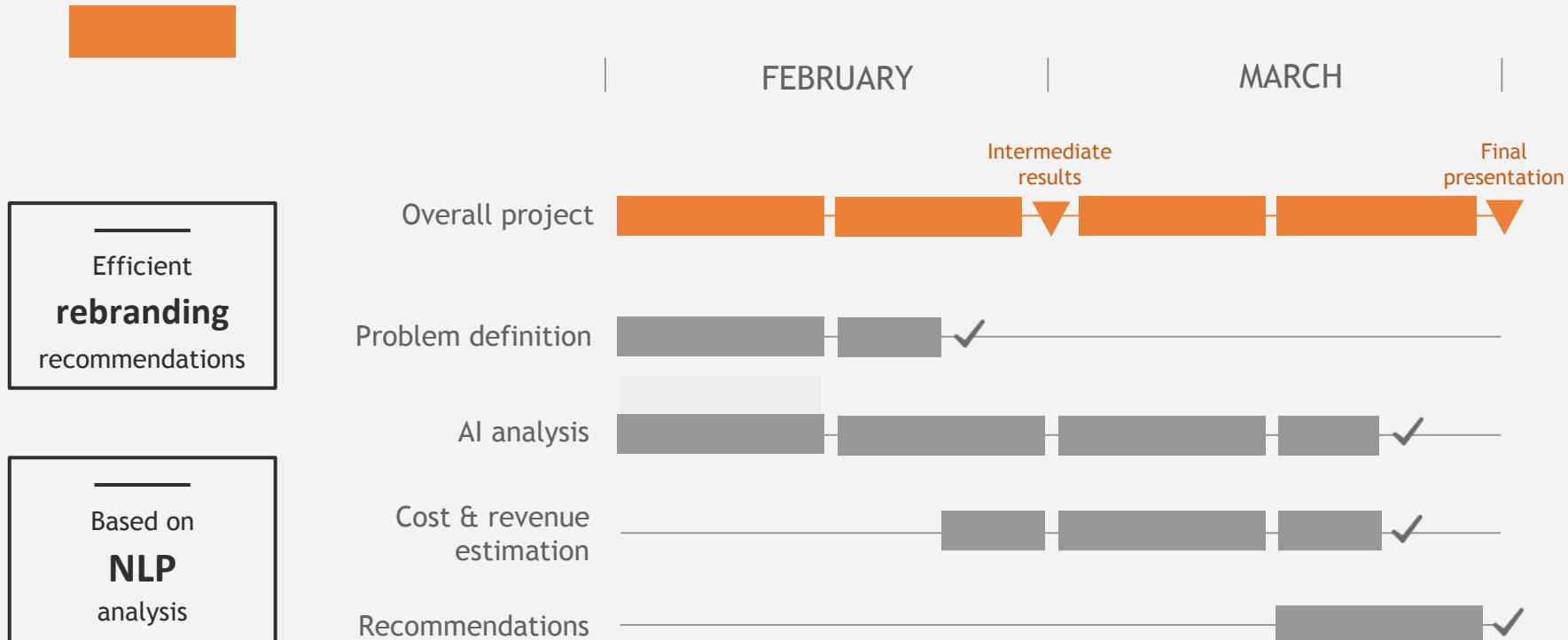
BADR EL-IDRISSI
MOKDAD



AKHILA VANGARA



PROJECT TIMELINE



AGENDA

01

REGULAR BUSINESS
APPROACH TO
YOUR PROBLEM

02

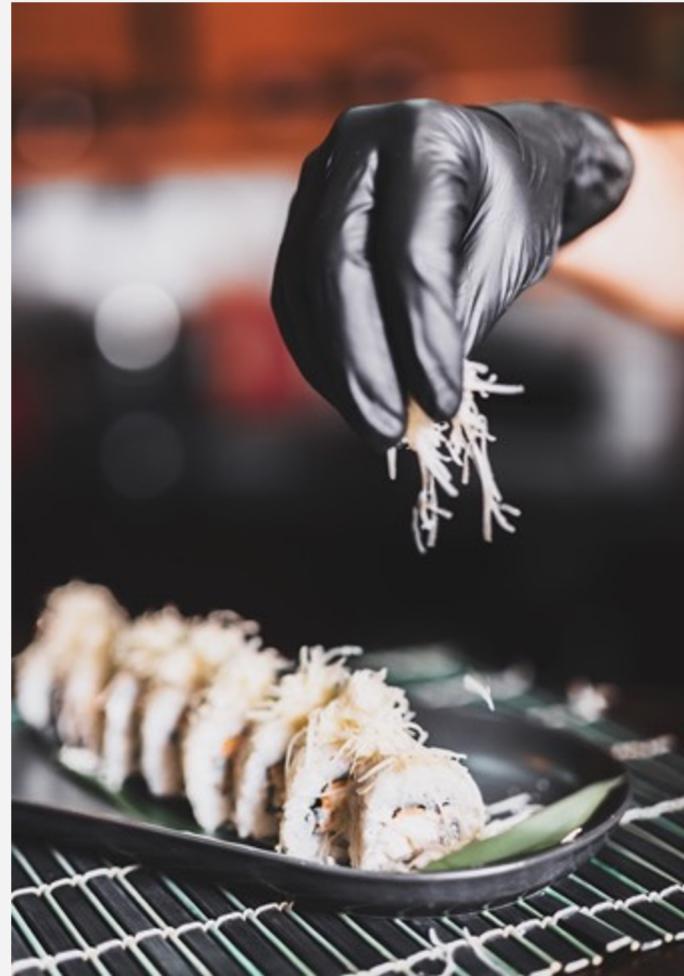
OUR ADDED-VALUE :
TECHNICAL ANALYSIS

03

OUR
RECOMMENDATIONS

04

AFTER THE
REBRANDING



01

REGULAR BUSINESS APPROACH

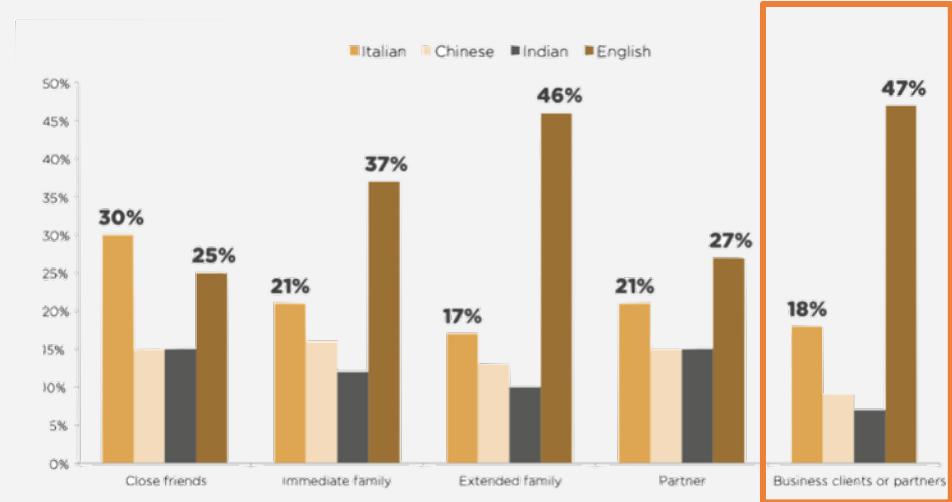
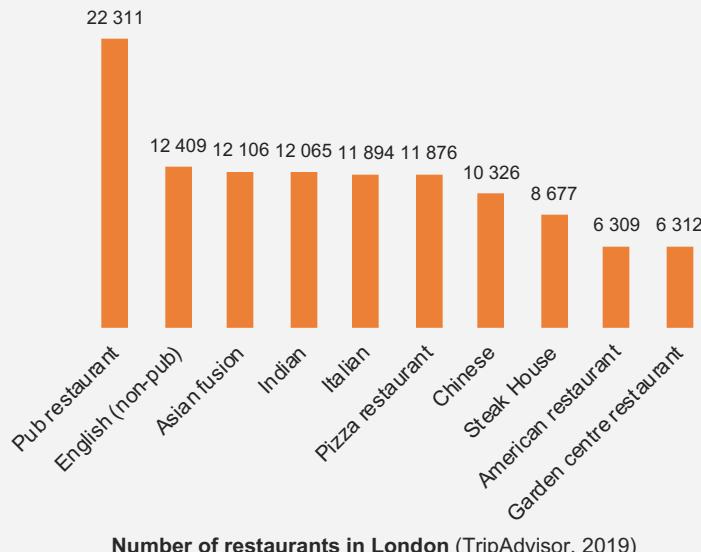
Market research
Marketing analysis



MARKET TRENDS – Pub restaurants and English cuisine are the preferred choice for business meals



- As presented below, our research suggests that **Pub restaurants** and **English cuisine** are the most popular restaurant types in London - they are both the **most numerous venues**, and the **preferred type for business meals**.

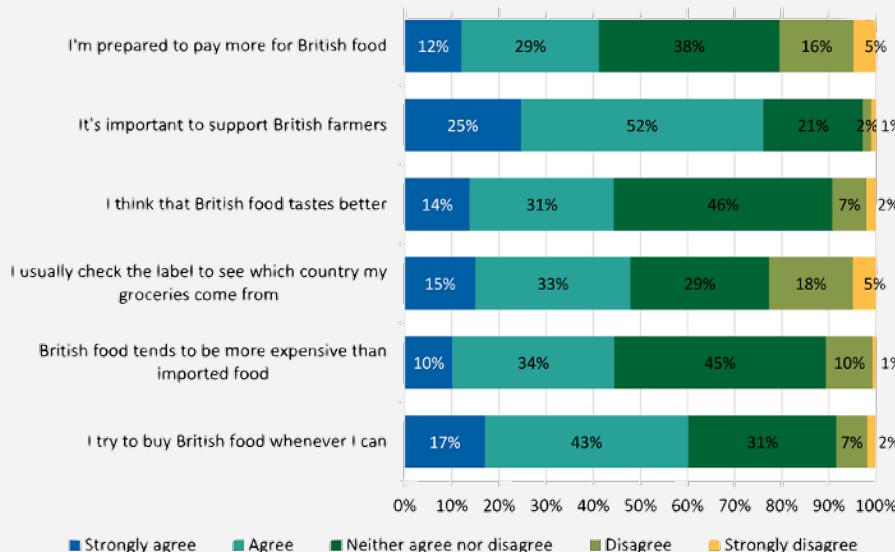


Number of restaurants in London (TripAdvisor, 2019)

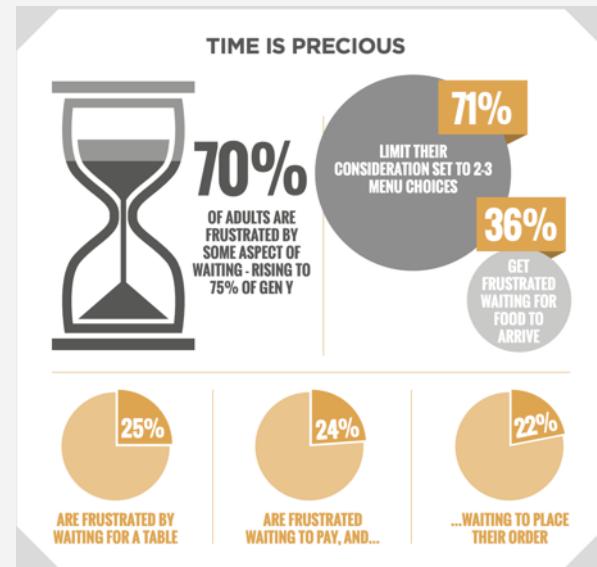
Preferred restaurant type depending on situation (London Evening Standard, 2018)

MARKET TRENDS – Clients value primarily local food and efficient service

- We also found out that clients particularly value locally-sourced food (the first figure even suggests that 42% of them are ready to pay more for British food), and that waiting time was the main pain point in London restaurants.



Customer attraction for local food (inew.co.uk, 2019)



London client preferences, infographic (TimeOut, 2017)

COMPETITION – Main competitors in Canary Wharf have a strong brand identity

- The restaurant market in Canary Wharf has specific features related to its location:



Restaurants focused on corporate dining, bars focused on afterwork drinks



≈ 120,000 people working in Canary Wharf daily (2019), and thousands more coming as visitors



Strong competition, with main competitors on the high-end segment being Roka and Gaucho



Local restaurants can source some of the finest catches at Billingsgate fish market, the UK's largest inland fish market, right next to Canary Wharf

- We identified 2 restaurants as Bokan's main competitors in the area. They both have a very strong brand identity:



Japanese-style dishes in a modern setting



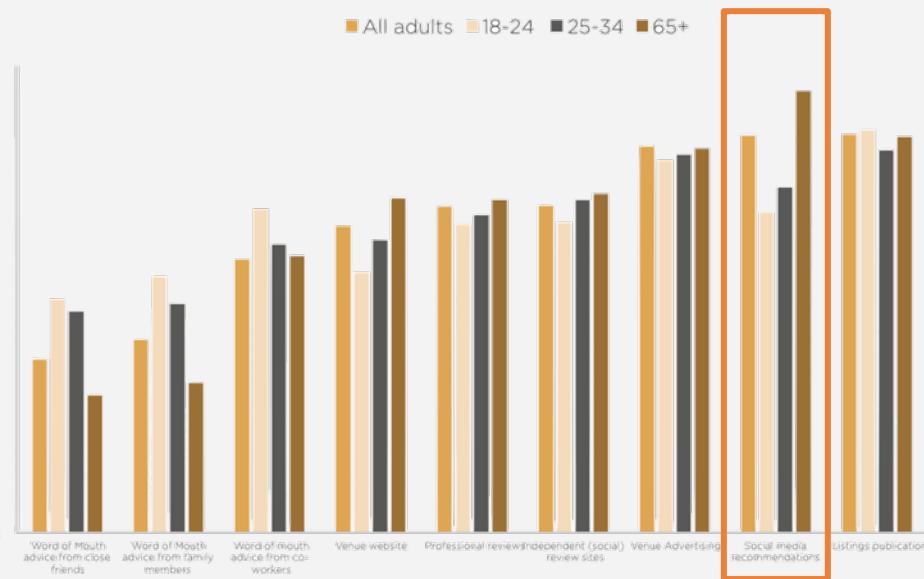
Argentinian cuisine with premium beef and wine



Branded as “European cuisine”: less strong identity

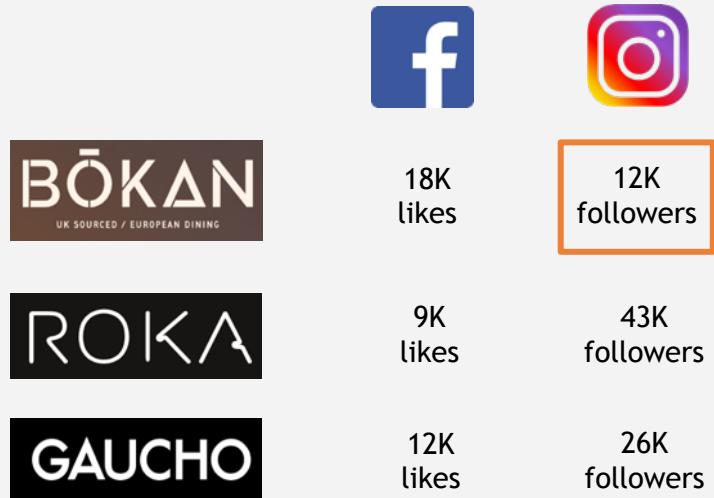
MARKETING STRATEGY – Bokan is behind competitors for its presence on Instagram

- Social media is among the most important channels for restaurant choice, in all age groups:



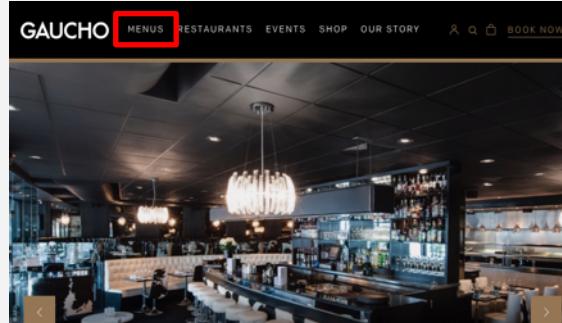
Preferred communication channels for decision-making (*London Evening Standard*, 2018)

- We identified 2 restaurants as Bokan's main competitors in the area. They both have a very strong brand identity:



ONLINE USER EXPERIENCE – Bokan's UX is less fluid than competitors

- We analysed Bokan's website, with in mind the famous “15 second rule”: when looking for information, customers usually don't spend more than 15 seconds on your website
- To find the restaurant menu on Bokan's website, you have to scroll through 3 different pages, vs. a single click on the homepage for Roka and Gaucho (see red boxes on images)

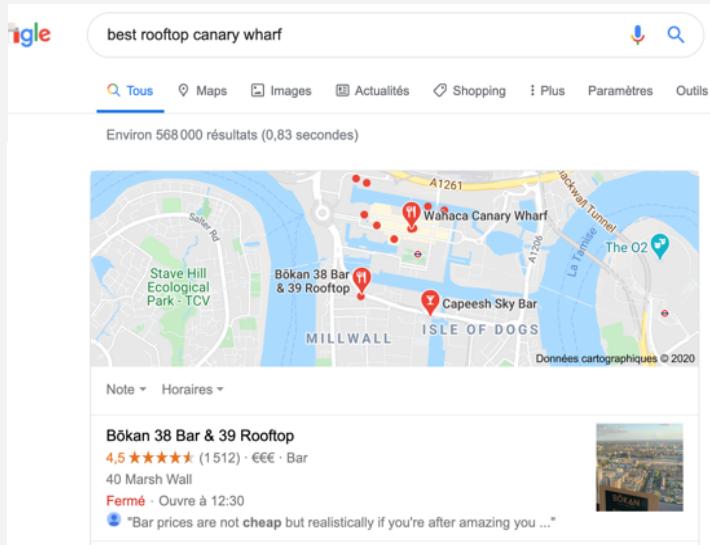


After a stint working at a private members club, Guillame joined the Michelin-starred L'Atelier de Joël Robuchon where he met Aurélie Altemaire. Two years later, Guillame joined Aurélie at Bokan and has not looked back since!

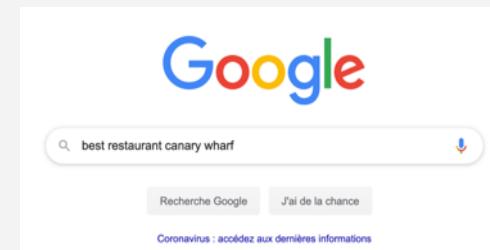
*Menus are subject to change as Chef Guillaume and his brigade offer you only the best seasonal produce to your table.

SEARCH ENGINE OPTIMIZATION – Bokan is listed by Google as best rooftop in Canary Wharf

- When searching for “best rooftop canary wharf”, Bokan is the first result:



- When searching for the best restaurant in Canary Wharf, Bokan is behind Roka:



- Bokan 37: 3 occurrences on first page of results

- Roka: 8 times on first page



- Bokan 38: 5 times on first page

- Roka: 5 times on first page

UPDATED SWOT ANALYSIS (1/2)

- Based on this new business research, we were able to update the SWOT analysis from our intermediate presentation:

Strengths



Flexibility: the menu changes regularly and offers local food, which was an important point in our findings



Client experience: the live music events and the breathtaking rooftop view are strong competitive advantages



Staff quality: experienced French chef Guillaume Gillan is in charge of the kitchen

Weaknesses



Brand identity: compared to top competitors, the restaurant isn't focused on a specific cuisine and has a more common 'modern' atmosphere



Online presence and experience: presence on Instagram and user website experience are behind competition



Access means: the restaurant & bar are on the 37th and 38th floor of a hotel, which requires additional marketing efforts so that clients find them easily

UPDATED SWOT ANALYSIS (2/2)

Opportunities



Location: location in an international hotel in Canary Wharf allows for a solid & predictable client base year-round



Pricing power: business clients are less sensitive to prices - focus on location, experience, quality



Unoccupied niche: no restaurant in the area focuses on what our analysis identified as the preferred food type in London: British / (high-end) pub cuisine

Threats



Competition from new business: growing trend over new types of restaurant such as fancy fast-foods or vegan restaurants



Competition from existing business: clients might want to stick to landmarks with strong identity such as Roka & Gaucho



Slowdown in activity with Brexit: banks moving offices to France or other EU countries might impact Canary Wharf.



In this first section we identified several issues, using a traditional business approach (market research, SWOT). We will complement them with a technical analysis, and propose solutions in the Recommendations section.

02

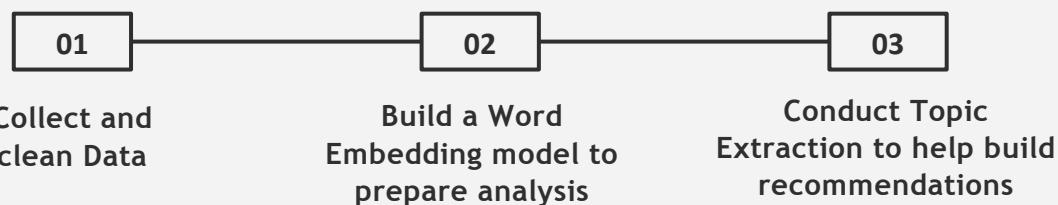
OUR ADDED-VALUE : TECHNICAL ANALYSIS

Web scrapping
Review embeddings
Topic extraction



OUR ADDED VALUE – Natural Language Processing (NLP) to complement business findings

- To bring value on top of the regular business approach, we conducted an AI-based technical analysis, using reviews from *Tripadvisor.com* for restaurants in London.
- We built a NLP model to understand what makes restaurants successful (or not!) in London, and provide more accurate recommendations for Bokan's rebranding.
- Our technical analysis went through the following steps:



DATA COLLECTION & DATA CLEANING

STEP

01

Web Scrapping

Collect as many reviews on the Internet as possible (several thousands)



Tokenization

Break a stream of text into words, phrases, or other meaningful elements called tokens



Lemmatization

Reduce inflection in words to their root forms ensuring that the root word belongs to the language



REVIEW EMBEDDING

STEP

02

Embedding

Representation in which a word is associated to a numeric vector. We can then perform calculations and operations on embedding matrices



LSI

Mathematical method for modeling the meaning of words and passages by analysing a corpus of texts

&

Word2Vec

Embedding method which embarks a neural net with context and target word notions



TOPIC EXTRACTION

STEP

03

Topic extraction

Method to identify the latent topics in the corpus of reviews and the words most related to each topic



LDA

A statistical model that views each review as a mixture of topics and each word as related to one of the topics



&

GSDMM

A variant of the LDA model to extract topics through clustering of shorter texts



TOPIC EXTRACTION – Most relevant topics in high reviews confirm market trends observed in our market research

- We extracted the topics most relevant to high reviews (4-5 stars) in London:

Some topics reveal the preferred types of food in London (Italian, pub food, burgers, etc.):



Italian food - pizza, pasta, wine



Main course meals - Fish & chips, burgers, steak, sushi, meat



Smaller meals - tea, breakfast, eggs, sandwich

Some topics reveal the elements most valued by clients in good reviews (short wait time, live music, etc.) :



Service details - recommendations, wait time, order time, events



Atmosphere - pub, drinking, live music, garden, atmosphere



Simple location-related topic - street, venue, London, restaurant

Additional details in Appendix

➤ These results complement and confirm our previous findings concerning the preferred types of food in London (Italian, English/pub cuisine) and the most relevant items for client satisfaction (wait time, atmosphere/identity).

TOPIC EXTRACTION – Most relevant topics in low reviews also highlight the main pain points

- We used a Word Cloud to reveal the most occurring words in 1-star reviews:



- Putting aside the common words such as ‘restaurant’ or ‘food’, we can see that **relevant words include vocabulary related to service** ('staff', 'order', 'time' 'rude', 'service', 'manager') **and to food quality** ('cold', 'price', 'disappointed')
- Again, this confirms our finding that service and waiting time issues are of particular importance for customer satisfaction

→ Thanks to a technical analysis using AI methods, we were able to confirm and complement the ‘business’ findings of the first section. Using these results, we crafted rebranding recommendations, presented in the next section.

03

OUR RECOMMENDATIONS

Rebranding & prototype
Other strategic advice



OUR VISION



Bokan, the new “Gastronomic pub”
reference in London

A CONTEMPORARY INTERPRETATION OF BRITISH AUTHENTICITY

- Both our market/business research and our technical analysis show that English and pub cuisine are the top choices in London. We present in the next slides the 3 axes we recommend for Bokan's rebranding.
- The “gastro pub” term might not be the fanciest one, but it's very trendy and represents a **clear opportunity** for Bokan. What's more, Bokan **already has most elements in place** - rebranding could be quick and reasonably costly.



Interior design

- Build on existing decoration to enhance the experience of contemporary British authenticity
- Keep Bokan's name for coherence / existing customers



New menu

- Rebrand the restaurant as high-end pub cuisine
- Keep local sourcing and regularly changing menu
- Apply slight price increase (see next slides)



Marketing strategy

- Creative agency to redesign website and print material
- Social media campaign targeted to London area
- Marketing in specialized magazines to support change

FINE-TUNE INTERIOR DESIGN AND ATMOSPHERE

- Typical gastro-pub features include: comfortable leather sofas, thick wood, industrial materials (brick and metal), references to game hunting, Victorian-style flowers (less modern, not recommended here)
- Fortunately, Bokan already has much of these elements (see on right): deep seats with cow skin, thick wood tables, brick wall (left wall). It mixes it with a **modern industrial design** of metal & glass



Pushing further in current direction would give Bokan a strong identity and offer clients a **contemporary pub experience**

Our main inspiration for decoration is the Culpeper, one of the most renowned gastronomic pubs in London (see on right), with similar industrial fittings (glass and metal), a little more green features (climber plants), and a daily-changing menu of modern British dishes



Bokan 37 (top) and Culpeper, London (bottom)



BUILD A NEW MENU – Prototype

- Our new menu prototype includes **typical British / pub cuisine** (red bliss potatoes, chocolate marble...)
- It also includes **food that can be locally sourced** (organic chicken, market vegetables...) and **seasonally changed**. Bokan can **keep its current ‘UK sourcing’ engagement and existing supply channels**, which **reduces transformation costs**
- Along with menu changes, we recommend Bokan to apply a ≈ **+10% price increase on dinner prices** (see revenue estimation in next section)
- The sleek menu **design** (indicative - to be reviewed by a creative agency) remains coherent with Bokan’s **authentic but contemporary atmosphere**



LAUNCH A NEW MARKETING & ONLINE STRATEGY



- Bokan is behind competitors regarding presence on **Instagram** (cf. marketing analysis in the 1st section), which is increasingly important to drive **new customers**
- The user experience on Bokan's website is **less fluid** than main competitors, with several clicks required just to find the menu
- A **marketing campaign**, both print (specialized magazines) and online (social media), will be required to broadcast Bokan's new brand image and attract new customers
- A **creative agency** will also be required to redesign the website, visual identity, menu and print material according to the new brand orientation



Bokan has all what's needed to reinvent the high-end British pub and become a reference restaurant in London. In the next (and last) section, we estimate revenue post-rebranding, rebranding costs and breakeven, and key KPIs.

04

AFTER THE REBRANDING

Revenue estimation
Costs & breakeven estimation
Key KPIs to monitor



REVENUE ESTIMATION POST-REBRANDING – Restaurant

- Based on our recommendations, we updated the revenue estimation from our intermediate presentation
- We estimated that the rebranding could help reach a **+10% in revenue and +15% in profits at a 2-year horizon**, which would bring **total revenue at c. £7.5M per year**. This is not an engagement, but an estimation, which can depend on multiple factors. We listed the most relevant KPIs in the last section of this report. We detail our estimations post-rebranding below:

RESTAURANT

1- BRUNCH	2- DINNER
	70 seats 2 services
	50% Occupancy rate
	Menu (food+bev) at £52
	Menu (food+bev) at £52
	70 people eating
	£52 per person
£3,700 per brunch	
	70 seats 2 services
	80% Occupancy rate
	100% Fri/Sat 80% Sun 70% Other
	+5%
	110 people eating
	£50 per person +10%
	£25 per person +10%
	£75 per person
	£8,200 per dinner

REVENUE ESTIMATION POST-REBRANDING – Bar / Rooftop

BAR / ROOFTOP

	1- 12h–15h	2- 15h–19h	3- 19h–00h
	50 seats	50 seats	50 seats
	Winter: 20% Occupancy rate Summer: 50% Occupancy rate	Winter: 50% Occupancy rate +10% Summer: 90% Occupancy rate +5%	Winter: 60% Occupancy rate Summer: 90% Occupancy rate +5%
	£20 per person	£12 per person	£12 per person
	£20 per person	£30 per person +5%	£40 per person +10%
Sub-total:	Winter: £600 Summer: £1,500	Winter: £2,100 Summer: £3,800	Winter: £3,900 Summer: £5,800
Total:	£6,600 per day in winter and £11,100 per day in summer		

REVENUE ESTIMATION POST-REBRANDING – Key takeaways

Restaurant / Brunch

- £3,700 per lunch
- 355 days

1.3 M£ per Year

Restaurant / Dinner

- £8,200 per dinner
- 355 days

3.1 M£ per Year

Bar & Rooftop / Winter

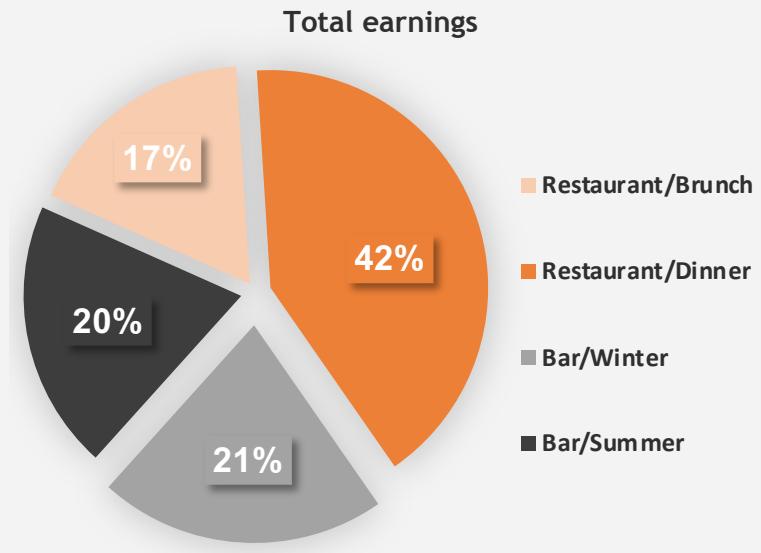
- £6,600 per day
- 8 months

1.6 M£ per Year

Bar & Rooftop / Summer

- £11,100 per day
- 4 months

1.5 M£ per Year



Total Revenues: £7.5M per year

COSTS & BREAK EVEN ESTIMATION – Fixed and variable costs

- We evaluated Bokan's different costs, in order to provide a break-even estimate:

Cost Of Goods Sold

- COGS represent **~30-35% of Sales** for fine-dining restaurants (market research)
- Local sourcing brings costs up, while efficient inventory management brings them down
- Bokan already has a local sourcing strategy with an existing network of local suppliers
- Therefore we would recommend Bokan to target COGS at **30% of Sales**

→ Total: **£2.3M per year**

Other variable costs

- Labour costs: 25%
 - Utilities: 2%
 - Marketing costs: 3%
- Total: **£2.3M per year**

Fixed costs

- Rent: 5% of Sales (0% if the building is fully-owned by Novotel)
 - Other fixed costs (overheads, licensing, IT, kitchen material maintenance): 10%
- Total: **£1.1M per year**

Total variable costs: £4.6M (→ **Gross Margin ≈ 38%**)

Total fixed costs: £1.1M

COSTS & BREAK EVEN ESTIMATION – Rebranding and break even

Rebranding investments

- As most of the current decoration already fits the contemporary pub atmosphere, **new decoration**: $\approx \text{£30K}$
- New branding** (creative agency for website, print, etc.): $\approx \text{£30K}$
- Additional marketing campaigns** (social media, specialized magazines): $\approx \text{£40K}$

Rebranding costs (capex): £100K

Breakeven point

- Based on our revenue estimate of £7.5M per year, sales are $\approx \text{£100K per week}$, which means that **rebranding investments are absorbed in ≈ 1 week**
- Our breakeven formula is:

$$\text{Breakeven point} = \frac{\text{Total Fixed Costs}}{\text{Gross Margin ratio}}$$

- This formula gives us a **break even at $\approx \text{£51K per week}$** (i.e. weekly revenue $> \text{£51K}$ directly contributes to profits)



With a **break even point** well under its weekly revenue estimate (post-rebranding), and with relatively reasonable **rebranding investments** (absorbed in ≈ 1 week of sales), Bokan is bound to be **significantly profitable**.

Key Performance Indicators (I)

➤ Lastly, in order to prepare the ground for performance monitoring after the project, we identified several important KPIs to follow:

Sales

- **Breakeven point** - Making sure profitability is reached back quickly after rebranding investments
- **Gross Profit ($Sales - COGS$)** - Keeping an eye on the cost of goods sold to maintain profits

Social Media

- **Engagement** - Number of likes, shares, retweets or comments for each post
- **Reach** - Social media audience: number of followers, traffic data, etc.

Customer Feedback

- **Customer Satisfaction** - Monitoring satisfaction through reviews and surveys
- **Customer Retention Rate (%)** - Following the proportion of new clients vs. repetitive clients and build a loyal client base to secure long-term revenue.

Key Performance Indicators (II)

Employee Engagement

- **Employee Turnover** - Keeping experienced employees, that clients know, can help increase customer loyalty and satisfaction.
- **Employee Satisfaction** - can be measured through team meetings and surveys, and help secure efficient employees in the flexible UK market

Front house management

- **Table turnover rate** - to monitor to ensure adequate wait times and maximize sales
- **Revenue per available seat hour** (= total rev / [available seats x opening hours]) - to monitor to optimize labour scheduling, food purchasing and table turnover
- **Sales per guest** - to help restructure menu, e.g. based on time of the day
- **Menu item profit and popularity** - allows for strategic pricing of items on the menu

 *Based on these monitoring tools and our previous recommendations, we believe that Bokan could go through a successful and valuable rebranding. We remain at your disposal for any further information.*

MANY THANKS FOR
YOUR ATTENTION



APPENDIX

Additional market research material
Technical analysis supplement



MARKET TRENDS – Most trending types of cuisine in London

- This other visualization graph confirms our previous findings
- The most trending types of cuisine in London are British and Italian



Topic Extraction – LDA output

- This is the output of the LDA model. Each topic has distinct words associated, and words are captured along with their associated probabilities to be in a given topic:

```
0: 0.022*"chip" + 0.020*"fish" + 0.016*"order" + 0.015*"chicken" + 0.013*"cook" + 0.013*"tast" + 0.011*"food" + 0.011*"good" + 0.010*"dish" + 0.010*"main" + 0.010*"sauc" + 0.009*"like" + 0.009*"starter" + 0.008*"steak" + 0.008*"portion" + 0.008*"burger" + 0.008*"fri" + 0.008*"meal" + 0.006*"dri" + 0.006*"meat"

1: 0.044*"pizza" + 0.024*"wine" + 0.014*"delici" + 0.014*"pasta" + 0.012*"italian" + 0.012*"roast" + 0.010*"love" + 0.009*"dessert" + 0.009*"salad" + 0.009*"cl  
eess" + 0.008*"good" + 0.007*"lunch" + 0.007*"top" + 0.007*"bread" + 0.007*"nice" + 0.007*"glass" + 0.007*"also" + 0.007*"place" + 0.007*"made" + 0.007*"great"

2: 0.044*"restaur" + 0.042*"london" + 0.027*"best" + 0.020*"visit" + 0.020*"food" + 0.018*"place" + 0.017*"one" + 0.015*"tri" + 0.011*"time" + 0.010*"street" +  
0.009*"area" + 0.009*"hotel" + 0.009*"find" + 0.008*"stay" + 0.008*"indian" + 0.008*"look" + 0.008*"mani" + 0.008*"ever" + 0.007*"found" + 0.007*"first"

3: 0.062*"pub" + 0.027*"bar" + 0.023*"drink" + 0.021*"beer" + 0.013*"lunch" + 0.013*"great" + 0.012*"music" + 0.011*"staff" + 0.010*"room" + 0.010*"visit" + 0.  
010*"night" + 0.010*"garden" + 0.010*"work" + 0.009*"friend" + 0.009*"area" + 0.009*"old" + 0.008*"live" + 0.008*"local" + 0.008*"atmospher" + 0.008*"sunday"

4: 0.036*"tabl" + 0.026*"us" + 0.019*"ask" + 0.014*"drink" + 0.013*"book" + 0.012*"wait" + 0.011*"arriv" + 0.010*"came" + 0.010*"minut" + 0.010*"one" + 0.010  
*"seat" + 0.009*"staff" + 0.009*"serv" + 0.009*"bar" + 0.009*"went" + 0.008*"said" + 0.008*"even" + 0.008*"order" + 0.008*"told" + 0.008*"waiter"

5: 0.028*"food" + 0.022*"servic" + 0.016*"staff" + 0.016*"us" + 0.015*"restaur" + 0.014*"even" + 0.014*"love" + 0.014*"meal" + 0.013*"great" + 0.013*"visit" +  
0.012*"recommend" + 0.012*"friend" + 0.012*"night" + 0.011*"went" + 0.011*"time" + 0.011*"amaz" + 0.011*"thank" + 0.010*"dinner" + 0.010*"birthday" + 0.010*"e  
cel"

6: 0.028*"food" + 0.026*"order" + 0.022*"time" + 0.012*"place" + 0.012*"go" + 0.011*"restaur" + 0.010*"take" + 0.010*"never" + 0.010*"year" + 0.009*"get" + 0.  
09*"servic" + 0.008*"use" + 0.008*"would" + 0.008*"review" + 0.007*"wait" + 0.007*"eat" + 0.007*"away" + 0.007*"good" + 0.007*"one" + 0.007*"back"

7: 0.022*"chicken" + 0.022*"dish" + 0.019*"food" + 0.014*"delici" + 0.014*"good" + 0.013*"curri" + 0.012*"burger" + 0.012*"tri" + 0.011*"tasti" + 0.010*"order"  
+ 0.010*"lamb" + 0.010*"rice" + 0.010*"love" + 0.010*"realili" + 0.010*"sushi" + 0.009*"fresh" + 0.009*"also" + 0.008*"meat" + 0.007*"place" + 0.007*"recommend"

8: 0.014*"breakfast" + 0.013*"tea" + 0.011*"servic" + 0.010*"menu" + 0.010*"price" + 0.009*"food" + 0.009*"coffe" + 0.009*"charg" + 0.008*"egg" + 0.007*"one" +  
0.007*"bit" + 0.006*"2" + 0.006*"full" + 0.006*"offer" + 0.006*"cold" + 0.006*"bill" + 0.006*"good" + 0.006*"pay" + 0.006*"get" + 0.006*"sandwich"

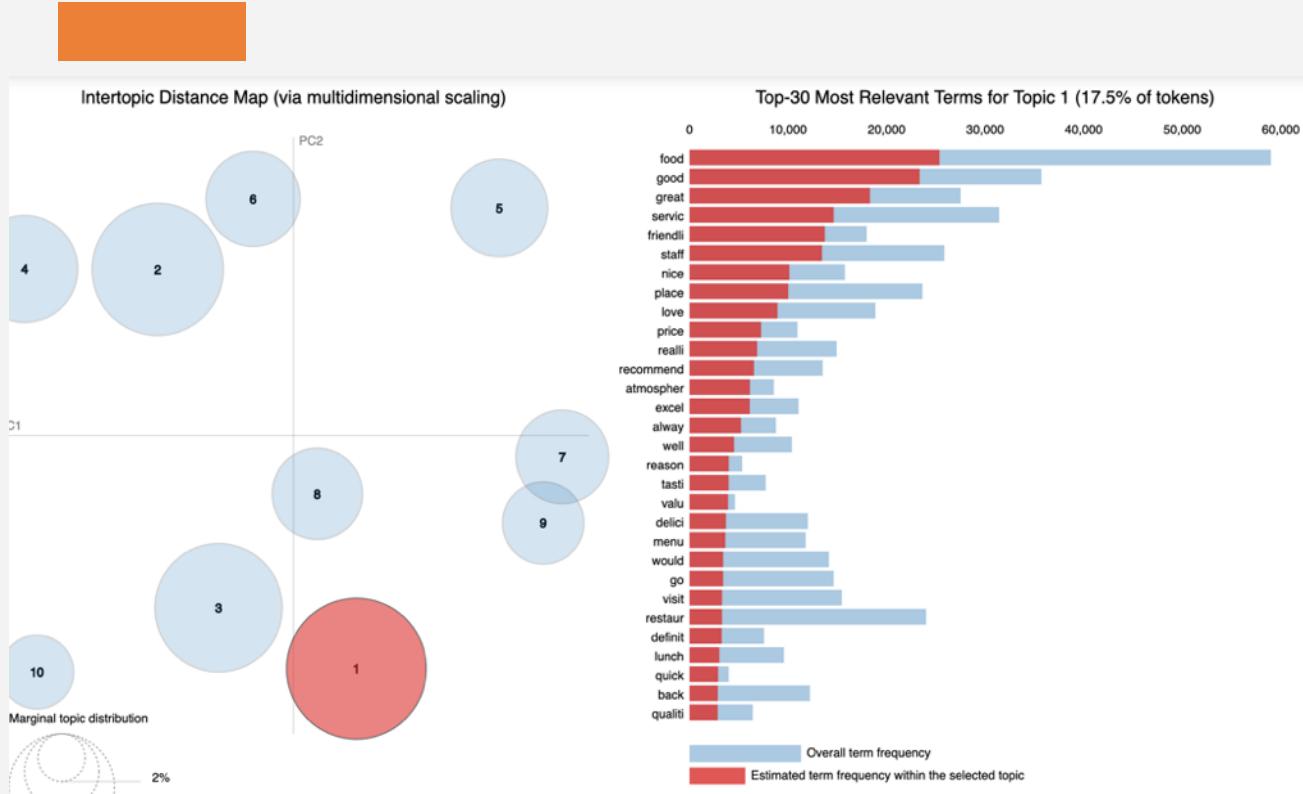
9: 0.058*"food" + 0.053*"good" + 0.042*"great" + 0.033*"servic" + 0.031*"friendli" + 0.031*"staff" + 0.023*"nice" + 0.023*"place" + 0.020*"love" + 0.016*"pric  
e" + 0.016*"realili" + 0.015*"recommend" + 0.014*"atmospher" + 0.014*"excel" + 0.012*"alway" + 0.010*"well" + 0.009*"reason" + 0.009*"tasti" + 0.009*"valu" + 0.  
008*"delici"
```

Topic Extraction – GSDMM output

- These are the text clusters created by the GSDMM model. It's a dimension reduction technique: clusters are classified by order of importance (i.e. highest number of documents and of words related to each topic cluster):

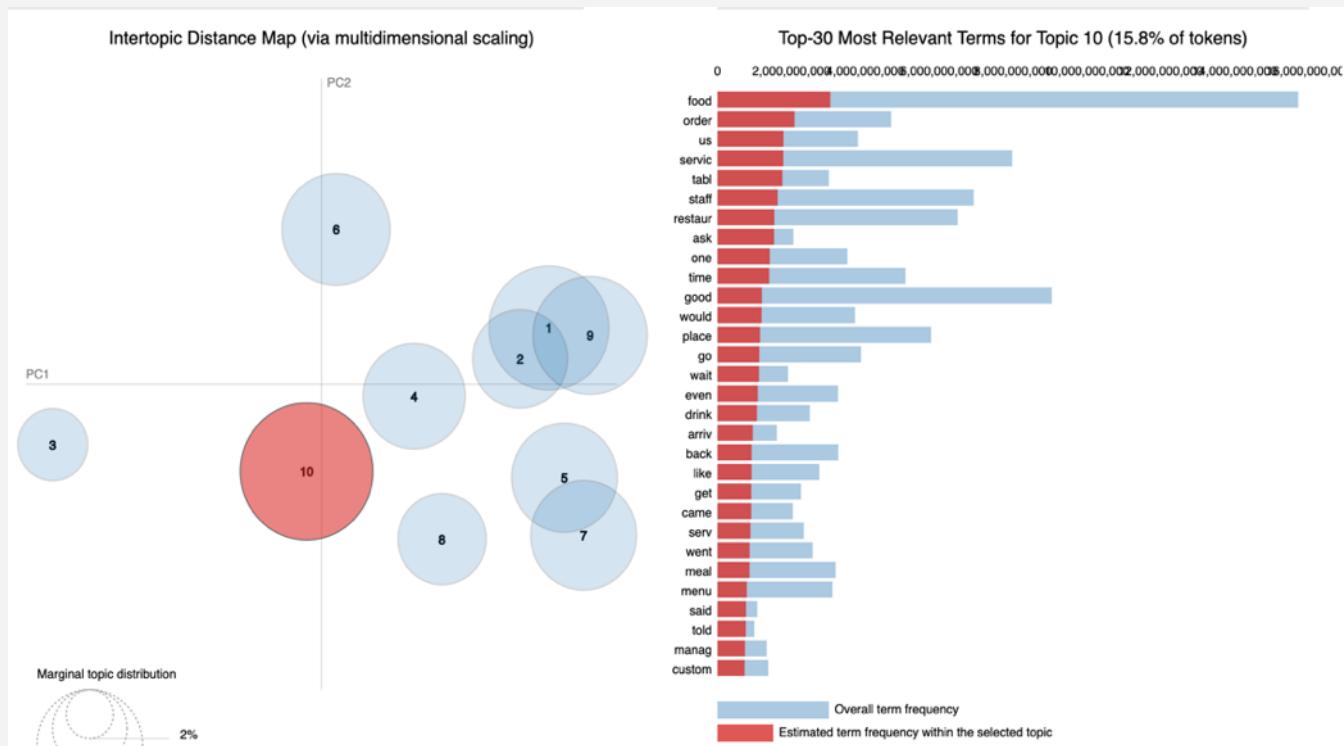
```
Number of documents per topic : [15348 7207 9756 7716 9507 11388 6990 12051 6741 13214]
*****
Most important clusters (by number of docs inside): [0 9 7 5 2 4 3 1 6 8]
*****
Cluster 0 : [('food', 12556), ('good', 7379), ('great', 6907), ('servic', 6306), ('staff', 6060), ('friendli', 5577), ('love', 4039), ('recommend', 3657), ('place', 3531), ('nice', 3059), ('restaur', 2967), ('alway', 2907), ('realli', 2817), ('price', 2625), ('excel', 2572), ('visit', 2412), ('time', 2396), ('atmospher', 2233), ('would', 2084), ('definit', 1989), ('delici', 1969), ('go', 1889), ('amaz', 1788), ('tasti', 1686), ('back', 1618), ('well', 1586), ('valu', 1496), ('highli', 1478), ('menu', 1470), ('reason', 1469)]
*****
Cluster 9 : [('food', 9416), ('order', 7972), ('servic', 5175), ('us', 5154), ('tabl', 5015), ('restaur', 4653), ('time', 4531), ('staff', 4428), ('ask', 4314), ('one', 3899), ('wait', 3722), ('would', 3645), ('place', 3291), ('arriv', 3262), ('good', 3236), ('even', 3218), ('go', 3200), ('drink', 2891), ('back', 2774), ('get', 2752), ('told', 2733), ('went', 2521), ('came', 2506), ('said', 2489), ('minut', 2487), ('meal', 2479), ('like', 2383), ('custom', 2377), ('serv', 2297), ('menu', 2248)]
*****
Cluster 7 : [('good', 5045), ('food', 4839), ('pub', 4104), ('great', 3845), ('place', 3118), ('staff', 3049), ('friendli', 2717), ('nice', 2544), ('servic', 2284), ('restaur', 2173), ('drink', 2124), ('visit', 2109), ('love', 2058), ('bar', 1934), ('beer', 1762), ('time', 1594), ('atmospher', 1474), ('look', 1376), ('go', 1371), ('area', 1368), ('london', 1353), ('realli', 1351), ('tabl', 1342), ('even', 1298), ('well', 1284), ('price', 1272), ('lunch', 1176), ('night', 1162), ('stay', 1147), ('excel', 1130)]
*****
Cluster 5 : [('food', 6731), ('great', 4288), ('servic', 4225), ('staff', 3988), ('love', 3309), ('us', 3105), ('even', 2437), ('friendli', 2369), ('restaur', 2330), ('visit', 2317), ('time', 2312), ('amaz', 2163), ('recommend', 2132), ('birthday', 2056), ('thank', 2018), ('good', 2010), ('friend', 1906), ('excel', 11851), ('night', 1847), ('realli', 1826), ('meal', 1737), ('back', 1702), ('went', 1651), ('made', 1535), ('place', 1469), ('would', 1424), ('famili', 1411), ('definit', 1378), ('delici', 1378), ('help', 1372)]
*****
Cluster 2 : [('food', 6303), ('restaur', 3515), ('good', 3413), ('servic', 2758), ('time', 2748), ('place', 2433), ('year', 2042), ('visit', 1941), ('go', 1813), ('alway', 1756), ('staff', 1493), ('great', 1390), ('tri', 1286), ('one', 1224), ('price', 1186), ('order', 1172), ('review', 1155), ('use', 1124), ('quali', 1041), ('best', 1003), ('would', 994), ('mani', 972), ('realli', 970), ('eat', 966), ('love', 936), ('never', 936), ('take', 922), ('last', 899), ('get', 895), ('takeaway', 893)]
*****
Cluster 4 : [('food', 4965), ('good', 4766), ('servic', 3388), ('great', 2797), ('restaur', 2789), ('menu', 2561), ('love', 2410), ('delici', 2405), ('staff', 2124), ('meal', 2099), ('nice', 2048), ('excel', 2003), ('dish', 1975), ('friendli', 1923), ('wine', 1911), ('lunch', 1773), ('realli', 1736), ('visit', 1721), ('main', 1654), ('well', 1637), ('starter', 1629), ('recommend', 1564), ('order', 1541), ('us', 1508), ('would', 1475), ('even', 1448), ('cook', 1437), ('enjoy', 1437), ('place', 1400), ('time', 1398)]
*****
Cluster 3 : [('coffe', 3087), ('breakfast', 3038), ('good', 3017), ('food', 2603), ('great', 2221), ('love', 2087), ('place', 2008), ('friendli', 1916), ('caf', 1904), ('staff', 1882), ('tea', 1602), ('servic', 1578), ('nice', 1465), ('cake', 1225), ('delici', 1214), ('visit', 1213), ('realli', 1085), ('egg', 1074), ('sandwich', 1057), ('go', 1048), ('time', 1045), ('lunch', 1037), ('price', 932), ('one', 866), ('serv', 840), ('fresh', 834), ('well', 834), ('littl', 824)]
```

INTER-TOPIC DISTANCE (LDA)



- The intertopic distance was plotted to ensure that the topics extracted are in fact independent from each other
- pyLDAvis was used to generate these graphs
- The topics are plotted against Principal Components - breaking down the data in topics aids in dimension reduction analysis
- Most relevant terms per topic and overall were studied to engineer recommendations

INTER-TOPIC DISTANCE (GSDMM)



- Interdistance map is not as well spread out as that for LDA model.
- These represent more in details the most important words in each cluster and their frequency of occurrence
- It helped us fine-tune our recommendations

SENTIMENT ANALYSIS

STEP

03 bis

Sentiment analysis

We used sentiment analysis as a “bonus” step in our technical analysis. It helped us observe how high ratings are related to positive sentiments



Flair

Flair is a library developed by Zalando on top of PyTorch. It combines different word embeddings to beat state-of-the art results in sentiment analysis.

&

VADER

VADER is a model used for sentiment analysis that is sensitive to both polarity (positive/negative) and intensity (strength) of emotion.

