Table Of Contents

1.Introduction

I grew up in my family's independently owned Mexican restaurant in California, I saw first hand the heart, hustle, and hurdles that define mom and pop hospitality. These independent businesses often compete with big chains, without the access to the same tools. That experience inspired me to uncover how AI can help independently owned restaurants in London better understand their customers and personalize their offerings. This report aims to show how even the smallest restaurants can harness approachable data and AI support to achieve a lasting impact.

2.Synthetic Data Generation

Link to data set: Mom-and-Pop_Restaurant_Customers__Synthetic_Data_

To simulate realistic insights, I used Chat GPT to generate a synthetic dataset of 500 unique and diverse customers. I defined all the variables based on relevance to the industry and refined the outputs to reflect London specific conditions (such as diverse customer motivations and higher meal costs). The data set includes demographic, behavioral, transactional, and psychographic variables. It also includes a short review of each customer's recent visits. Synthetic data generation using generative AI enables marketers to simulate behavior in data-scarce contexts, especially for small businesses without access to large CRM systems. This method allowed me to analyze patterns that would otherwise be invisible in limited or standalone datasets. These refinements helped the final dataset have diverse values, ordering behaviors, and dining motivations, laying the groundwork for the targeting and recommendations later on.

I chose my demographic variables to be ages 18-70. As well as household income categories (low, medium, and high). Home occupancy type (single, family with kids, couple), which is important as it influences budget, timing, size of group. For my behavioral variables I put eating out frequency, meal time preference, ordering style as they can help shape business logistics and service hours. Transactional data points have: average spend in a visit, loyalty program usage, and total visits in the past 6 months. This is important for understanding customer retention and revenue impact. Lastly, psychographic variables were selected for segmentation. This includes: dining motivations, food values, openness to new cuisines. This is good to know as they speak to how and why people eat out on a deeper level. Crafted prompt engineering can simulate informative datasets in data scarce environments which can be crucial for marketers without access to real time first party data. Synthetic data generation via strategic prompts can be helpful as they reflect real world diversity, cost structures, psychographic variation, etc.

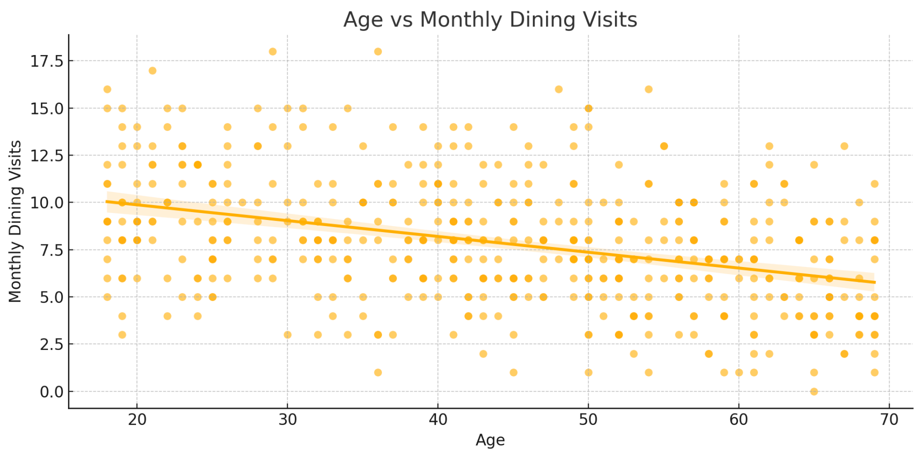
3.Consumer Insights

To reveal the actionable insights from the synthetic data set, Chat GPT and I conducted a multi layered analysis with correlation testing, behavioral interpretation, and review theme mining. The objective was to identify the most meaningful patterns that could help inform segmentation and personalization strategies for the businesses.

3.1 Dining Frequency by Age Group

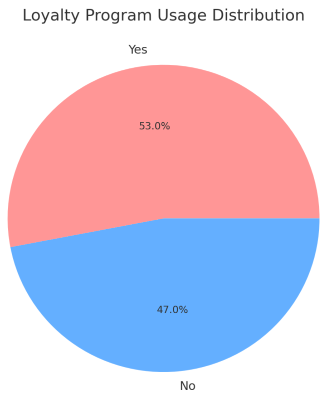
The first analysis was to test the relationship between dining frequency and age.

The correlation between age and restaurant visits was weak ($r = +0.037$). Younger adults 18-24 years old, were the most frequent diners as they averaged 7.32 visits per month, followed by adults aged 55-64 with 6.64 visits per month. It is likely that these averages stem from social spontaneity in younger groups and habitual comfort in older ones. Restaurants would benefit from targeting groups differently. Perhaps with mobile friendly offers for younger audiences and cozy, dine in focused bundles for older patrons. The figure below shows how younger customers tend to dine out more frequently than older ones.



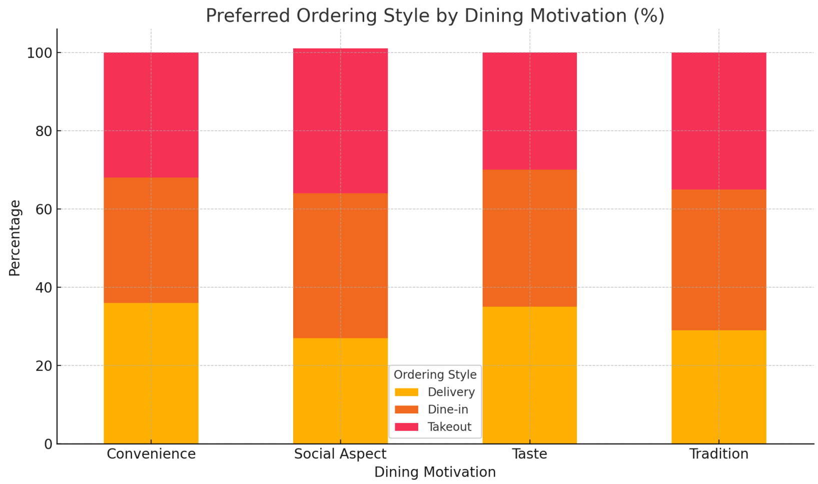
3.2 Loyalty Program and Visit Frequency

The second insight explored whether loyalty program usage was associated with a higher visit frequency. The results revealed that program members visited average 20.2 times over the past six months, as opposed to average 18.96 visits for non members, so there is an increase of +1.24 visits for loyalty program members. The correlation was slight (r = +0.018), however the result is positive because this shows how incremental improvements in retention can equate to increase in revenue. This suggests that perhaps, personalized loyalty incentives (ex: secret menu items, member only deals) could be a worthwhile addition.



3.3 Spending Patterns By Food Values

The third insight was focused on the relation between customer spending and food values. The health conscious patrons (those who prioritize sustainability), had the highest average spend per visit at £33.57. This might come down to pricier ingredients or simply a greater willingness to pay for what's seen as higher quality. This doesn't prove causation, but it shows an important psychographic subgroup with strong revenue potential.



3.4 Themes from Customer reviews

In addition to quantitative insights, I requested realistic customer reviews and asked to identify themes or emerging patterns. These reviews revealed three dominant themes: service & atmosphere, food quality & taste, and surprise or novelty.

Theme	***Mentions***	***Common Phrases***	***Example Verbatim***	***Segment Connection***
Service & Atmosphere	198	“friendly,” “cozy,” “comforting”	“Felt like I was eating at a friend’s house. So comforting.”	Tradition- and social-driven segments; aligns with high dine-in preference
Food Quality & Taste	189	“authentic,” “homemade,” “flavorful”	“The portions were generous and the flavors authentic.”	Taste-motivated customers; matches higher spend per visit
Surprise & Novelty	101	“pleasantly surprised,” “not my usual spot,”	“Not my usual spot, but I was pleasantly	Experimental segment; reflects

		“tried something new”	surprised.”	openness to new cuisines and supported by psychographic trait patterns

When customer motivations were cross tabulated with their unique preferred ordering styles, clear patterns were revealed. Convenience driven diners favored delivered food (36%), while socially and traditionally motivated patrons preferred dine in and takeout (37% and 35% respectively). This highlights how dining psychology should guide menu presentation or channel strategies.

3.5 Correlation vs. Causation

The insights that emerged did imply useful patterns, yet it's important to understand correlation from causation. An example is how loyalty members may dine out more not due to the program, but because frequent diners just so happen to be more likely to join. Also, higher spending among health conscious patrons may reflect a willingness to pay for perceived quality, not just cost of healthy food. Patterns like these can inform hypotheses and guide campaign strategies, but should not be treated as definitive. A/B testing could help validate which methods could drive behavior change. Marketers must avoid misinterpreting correlation as causation and instead use findings to guide hypothesis formation for A/B tests or campaign variants.

4.Customer Segmentation

K-means clustering methodology was used to help find distinct customer groups within the data set. This is a form of unsupervised learning. Supervised models like teachable machines rely on pre labeled outcomes, yet unsupervised clustering detects natural groups in data without predefined labels. This makes it optimal for revealing personas based on shared motivations and characteristics. Building on the variables I selected, the recommendation system applies segment traits alongside peer based and situational insights to show more personalized offerings. I chose these variables as they not only show how customers behave, but also why they choose certain restaurants to dine. This helps to offer more actionable insights for real world targeting.These variables were chosen to reflect who customers are, how they behave, and why they dine out ensuring that segmentation goes beyond surface traits.

4.1 Segment Profiles and Behavior Patterns

Segment Name	***%***	***Traits***	***Psychographics***	***Key Behavior***	***Primary Marketing Opportunity***
On-the-Go Gourmets	29%	Aged 18–24, single, low–mid income	Motivated by convenience & bold flavors	Takeout, 8.5 visits/month	App-based “Midnight Bites Club” + late-night loyalty perks
Wholesome Homebodies	24%	Aged 35–50, families, higher income	Tradition & social bonding, health-focused	Dine-in, 5–6 visits/month	“Sunday Supper Series” – locally sourced family meals

Cultural Curators	18%	Aged 28–40, couples, high income	Taste-first, adventurous, story-driven	High spend, global flavors	“Chef’s Table Passport” with rotating menus
Routine Relishers	29%	Aged 55–64, couples, moderate income	Tradition, routine, comfort food	Dine-in, early meals, low loyalty use	“Neighborhood Nosh” punch card for weekday regulars

Certain segments are less suitable for specific strategies. For example, routine relishers are unlikely to engage with adventurous tasting menus, while on the go gourmets may not respond well to slow paced dine in experiences.

5.Recommendation System Logic

To help independently owned restaurants in London deliver smarter, more personalized dining experiences, I developed a hybrid recommendation system that blends three AI filtering techniques: content-based, collaborative, and context-aware. Drawing on the customer segments developed in Section 3, this system mirrors the hybrid structure introduced in class to maximize relevance, reduce cold-start friction, and adapt to situational needs like dining mode or time of day. It also reflects the principles of real time personalization and user centric design from Session 4, particularly Slide 15’s call for AI to enhance customer experience without overwhelming small teams or budgets.

5.1 Curated Product Recommendations

These seven offerings were designed with segment preferences, psychographics, and ordering modes in mind. They include both dine-in and takeout-friendly options, appealing to a broad range of motivations from convenience to cultural exploration.

Product Name	***Cuisine***	***Dine in/Takeout***	***Target Segments***
Slow-Braised Lamb w/ Apricots	Moroccan	Dine-in	Routine Relishers, Wholesome Homebodies
Grilled Halloumi & Quinoa Bowl	Mediterranean	Both	Wholesome Homebodies, Value-Conscious Foodies
Tandoori Chicken Wrap	Indian	Takeout	On-the-Go Gourmets, Experience Seekers
Courgette & Chickpea Tagine	North African	Both	Value-Conscious Foodies, Cultural Curators
Vegan Mac & Cheese Bake	Plant-Based Comfort	Both	Ethical eaters across all segments
Shoreditch Mezze Box	Middle Eastern	Takeout	On-the-Go Gourmets, Cultural Curators

Sunday Roast Grain Bowl	Modern British	Both	Routine Relishers, Wholesome Homebodies

5.2 Decision Tree Design

To make product recommendations, the hybrid recommendation system takes into account situational considerations (context-aware), peer behavior (collaborative), personal history (content-based), and segment assignment (by K-means clustering). For instance, a vegan in that section may be presented with the Courgette & Chickpea Tagine, while an On the Go Gourmet perusing at night might be presented with the Tandoori Chicken Wrap or Shoreditch Mezze Box. Health conscious eaters are directed to the Halloumi & Quinoa Bowl, while Wholesome Homebodies strolling on a Sunday may be served the Sunday Roast Grain Bowl. The Tagine or a seasonal Chef's Tasting Menu may be given to cultural curators who are highly open and affluent.

5.3 Why It's Hybrid

This system combines collaborative filtering (Session 4, Slides 12&13), context aware filtering during (Session 4, Slide 15), and content based filtering (Session 4, Slide 11). By preventing filter bubbles, resolving cold start problems, and accommodating real world use cases, these strategies collectively solve the drawbacks of single method systems. This is consistent with the hybrid logic that we investigated in Session 4, Slide 14. Personal traits = content-based. Peer trends = collaborative. Time & place = context aware. Each branch uses a different lens to boost recommendation relevance.

Balancing Personalization & Discovery

The system features shifting recommendations (such as seasonal foods) aimed at Experience Seekers and inquisitive patrons in order to prevent over personalization. To keep the experience interesting, these recommendations are based on review trends and psychographic data.

5.4 Measuring Effectiveness

Key parameters including click-through rate (CTR), conversion rate from recommendation to order, average order value (AOV) by segment, frequency of repeat purchases, and adoption of new or seasonal meals can be used to monitor the effectiveness of this hybrid system. A/B testing offers a strong framework for evaluating filter combinations, and menu level statistics assist in determining whether tactics genuinely increase engagement and income. This aligns with Session 4, Slides 23 & 25, where we discussed metrics like CTR, CLV, diversity, and serendipity.

6. Conclusion

This report shows how even small, independently owned restaurants in London can harness the power of AI to better understand their customers and offer more thoughtful, tailored dining experiences. By pairing smart tools with human insights, local restaurants can better meet their customers' needs and achieve long term success.

Appendix

Section 1 Synthetic Data Generation

Prompt: Hello, I am working on a project that involves synthetic data generation, customer segmentation and recommendation systems and I am interested in doing the hospitality industry/ more specifically mom and pop restaurants. Can you please help me to generate a synthetic data set of about 500 customers that dine at independently owned restaurants in London. You are a marketing data analyst (Please have the data set in a table format, CSV-style), with the following variables: Demographic variables- -Age (numbered between 18-70) -Household Income (categories like low, medium, high) -Household occupancy type (categories like single, couple, family with kids, etc) Behavioral Variables: -Eating out frequency monthly (numeric from 0-13 visits) -Preferred Meal Time (categories: breakfast, lunch, dinner, late night) -Ordering Style (categories: dine-in, delivery, takeout) Transactional Variables: -Average spend per visit (numeric, in british pounds) -Loyalty program usage (categories: yes or no) -Total visits in the last 6 months (numeric) Psychographic Variables: -Motivation For Dining (categories: convenience, taste, tradition, social aspect) -Values in Food choices (categories: healthy, locally sourced, comfort food, experimental) -Openness to new cuisines (categories: low, medium, or high)

Refinements:

- I do have just a few refinements to reflect realistic data for dining in London. Can you please ensure that (for average spending) is ensured average spending costs for London prices, as London can be quite costly compared to other parts of UK.
- Please also ensure that there is more of a distribution across food values. That way, it will enhance segmentation quality and to avoid biased customers, etc.
- Can you also please ensure more of a variation across customer reviews that weigh on different motivations and tones. This will help to understand customer feelings and behavioral patterns.

Thanks!

Section 2 Consumer Insights

Prompts used:

- Can you help me to detect statistical Insights from the data or any meaningful patterns or trends? (Note: Once my AI provided me with a list of insights, I picked three that I found to be most important/useful)

- I am curious about the correlation between the age of customers and the amount that they dine out. What age groups tend to dine out the most?
- Now, I am curious about if there is a positive correlation between total visits and loyalty program usage. This is important to know to measure the efficiency of the loyalty programs. Can you please tell me if there is a significant difference in total visits between customers who do use the loyalty programs and those who do not. Please run the averages of those groups so I can see, thanks!

- Now I would like some insight into how varying psychological factors influence behaviors. Can you please tell me how different dining motivations (the categories were convenience, tradition, social, taste,etc) correlate with the preferred ordering styles (dining in, delivery, takeout)?

- Customer review: perhaps a 1-2 sentence review of their most recent visit at a restaurant. Please help me to generate realistic and diverse responses. please ensure variation across the variables and with no blank fields. Let me know how else I can refine this request if needed, thank you

- Now I'm going to request some data visualizations to help showcase the data. Can you please generate a scatter plot showing the relationship between customer age and their monthly dining visit frequency? Each point will represent one customer. Please include a trendline to highlight any overall pattern, and label the axes clearly age on the x-axis, monthly dining visits on the y-axis). This will help visualize how dining habits vary across age groups

- Can you generate a pie chart that reflects the amount of loyalty program users and nn loyalty program users?

- For my next data visual, can you please create a bar chart that shows how ordering styles (dine in, delivery, takeout) varies by dining motivation (taste, tradition, convenience, social). This can help me see how psychological factors influence behavior

- Can you please analyze the customer reviews that are in my data set and identify any recurring themes. I'm looking for more insight into how customers describe their experience. Perhaps if they mention the price, service, taste, etc. Can you please show any frequent themes along with some example quotes? Thank you

Section 3 Customer Segmentation

Prompts used:

- Using the data set that you generated, can you please generate 3 to 5 distinct customer segments using unsupervised learning logic (like clustering or groups on similarity) For each segment please have a descriptive, unique name. % of customers who fall into the group. common demographic traits (household type, income, age range). Behavioral traits (meal time preference, frequency, ordering style). Transactional traits (spend, loyalty). psychographic traits (dining motivation, food values, openness to new cuisines). Please make this as realistic as possible, making sure the segments reflect variation across all variables I listed. You can use key patterns like: younger customers dining out more often, loyalty program members who spend a bit more, and healthy food valuers spending the most per visit. thanks

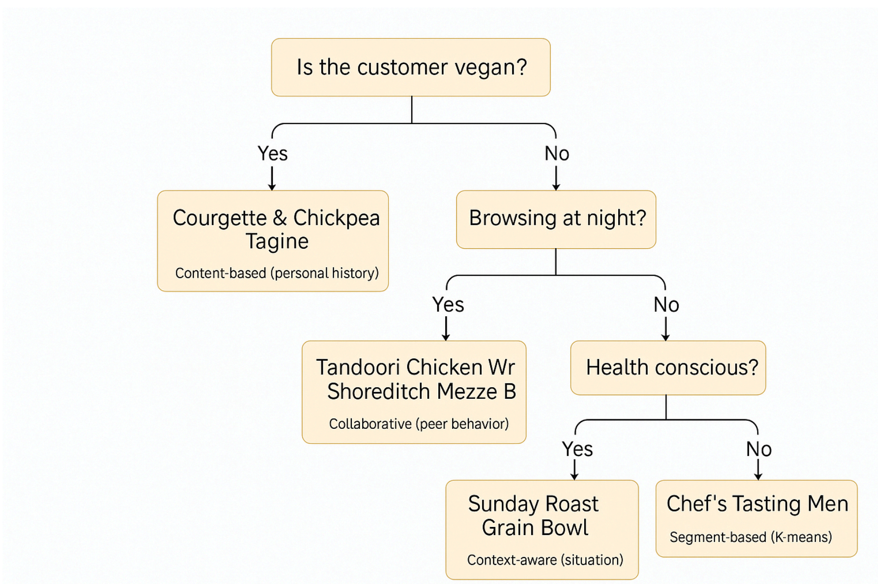
- I am building some recommendations for the locally owned restaurants in London and need to define about five to eight products (dishes or offerings) that could be found at these establishments. The products can reflect the diversity in restaurants and align with varying customer preferences and their motivations (taste, tradition, health). Please include both dine in and takeout friendly options. please have each product include a name, very short description, type of cuisine, dine or take out suited, idea customer segment it might appeal to.
- **Refinement needed**: Can you please refine these product recommendations? Maybe sprinkle in some healthy options, also can you substitute the chefs mole of the month for something else please? Can you also change the name of the east London mezze box to something that rolls off the tongue better? Also, the Sunday roast ramen sounds a little outlandish, can you make it Sunday roast theme but still realistic? In fact, can you make sure all of these offerings are more realistic?

Section 4 Recommendation Logic

Prompts used:

- Using the customer segments we previously made, can you please design a hybrid recommendation system that combines content-based, collaborative, and context-aware filtering? I’d like 5–8 specific food offerings, each matched to relevant segments and contextual logic like dine-in vs takeout, time of day, or dietary preferences. Please also provide a simple decision tree showing how the system would match customers to recommendations.

Hybrid Recommendation Logic (Decision Tree)



START

├── **IF** customer data available (segment, food values, past orders, time of day):

│ └── **ASSIGN** to customer segment:

│ ├── IF attributes match → **On-the-Go Gourmet**

│ ├── IF attributes match → **Wholesome Homebody**

│ └── IF attributes match → **Cultural Curator**

├── **CHECK** Personal Preferences (Content-Based Filtering):

│ ├── IF dietary preference = vegan/flexitarian → Recommend plant-based dishes

│ ├── IF prefers healthful meals → Recommend nutrient-rich options

│ └── IF repeat orders = high → Prioritize familiar favorites

├── **CHECK** Contextual Factors (Context-Aware Filtering):

│ ├── IF time = Sunday afternoon → Boost traditional offerings

│ ├── IF time = evening & mobile order → Prioritize quick takeout meals

│ └── IF weather = cold → Recommend warm, hearty dishes

├── **APPLY** Collaborative Filtering:

│ ├── IF similar users in segment ordered Item A → Recommend Item A

│ ├── IF Item B is trending in segment → Recommend Item B

│ └── IF new customer with sparse data → Use top-rated dish from matched segment

├── **Final Recommendation Logic**:

│ ├── IF user = returning:

│ │ └── Weight: 60% collaborative + 30% content-based + 10% context-aware

│ └── ELSE IF user = new:

│ └── Use default segment recommendations + trending items

└── **ALWAYS INCLUDE**:

├── 2–3 personalized product suggestions

└── 1–2 exploratory items based on season or popularity

END

Prompt: Can you include a product table that lists cuisine, dine-in/takeout availability, and target segments for each menu item? Please use segment names we created, and make sure the offerings reflect real variety (some plant-based, some hearty, some grab then go).

Note: This helped make the system more actionable and visually useful. I needed the recommendations to feel grounded in realistic restaurant offerings that reflected segment preferences and food values from the dataset.

Refinement: Please reformat the logic into a clear, written decision tree that shows how a customer is routed to a specific dish based on their segment, personal preferences (e.g. health or novelty), and contextual triggers like day of week or time as we learned about contextual factors (such as weather) in class. Aligned with Session 4 content on hybrid filtering approaches and contextual triggers (Slides 11–15).

Why it was needed:

The first version lacked structure. I wanted the system to clearly show how all three filtering techniques (content-based, collaborative, and context-aware) were integrated