Appendix:

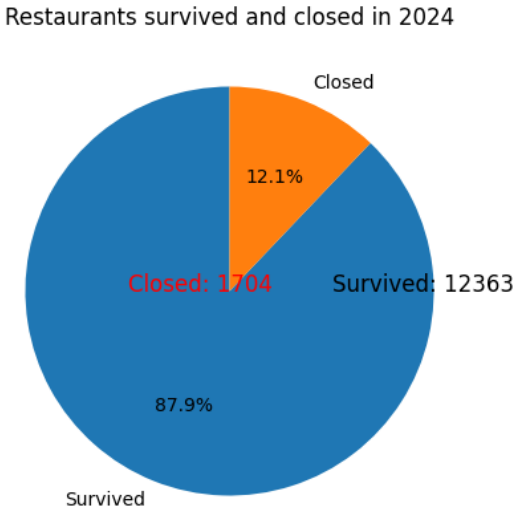
Part 1: Optimal Location Selection

1. Attributes of data in the master file:

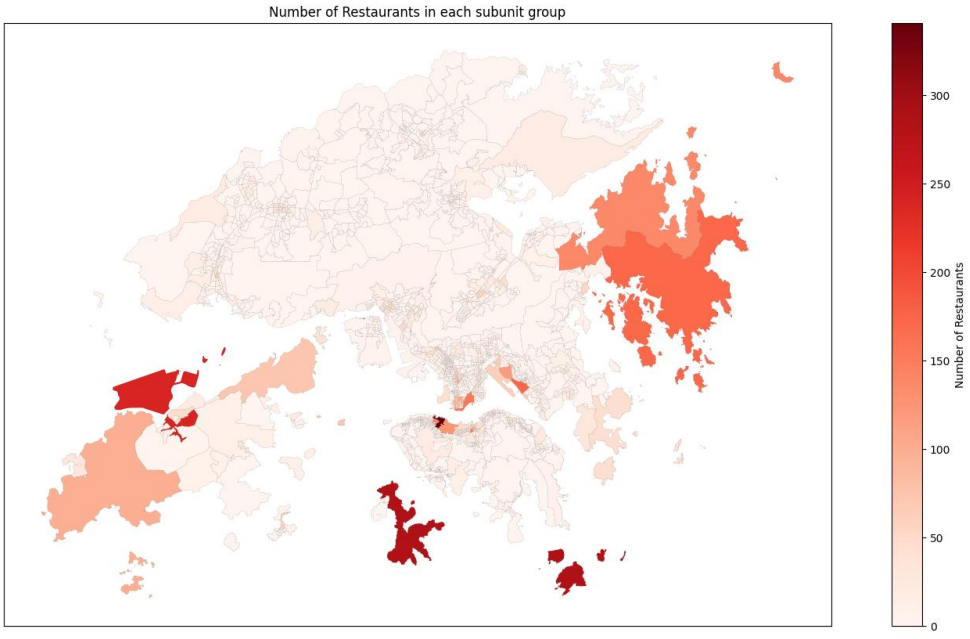
Column name	**Description**
TYPE	Interpretation of license type, RL represents General Restaurant Licence
SS	Name of restaurant
ADR	Operation address
lng	Longtitude of the restaurant
lat	Latitude of the restaurant
census_subnit	The subnit group in which the restaurant is located
number_of_nearby_restaurants	Number of restaurants 100m nearby
number_of_bus_stops	Number of bus stops 100m nearby
number_of_MTR_stations	Number of MTR stations 100m nearby
number_of_schools	Number of schools 100m nearby
number_of_commercial_facilities	Number of commercial facilities 100m nearby
number_of_government_facilities	Number of government offices 100m nearby
number_of_religious_facilities	Number of religious facilities (e.g. temples, churches) 100m nearby
number_of_medical_facilities	Number of medical facilities (e.g. hospital, clinics) 100m nearby
number_of_hotel_facilities	Number of hotels 100m nearby
mean_median_age	The median age of the subnit group in response to the restaurant
median_hs_income	The median household income of the subnit group in response to the restaurant
mean_hs_size	The median household size of the subnit group in response to the restaurant
labour_force_percent	The percentage of labour force of the subnit group in response to the restaurant
chinese_percent	The percentage of Chinese of the subnit group in response to the restaurant
pop_m_percent	The percentage of male of the subnit group in response to the restaurant
ten_oc_percent	The percentage of residents owning a property in the subnit group in response to the restaurant

rent_to_income_ratio	The rent to income ratio in the subnit group in response to the restaurant
survive	Whether the restaurant survived or not in 2024

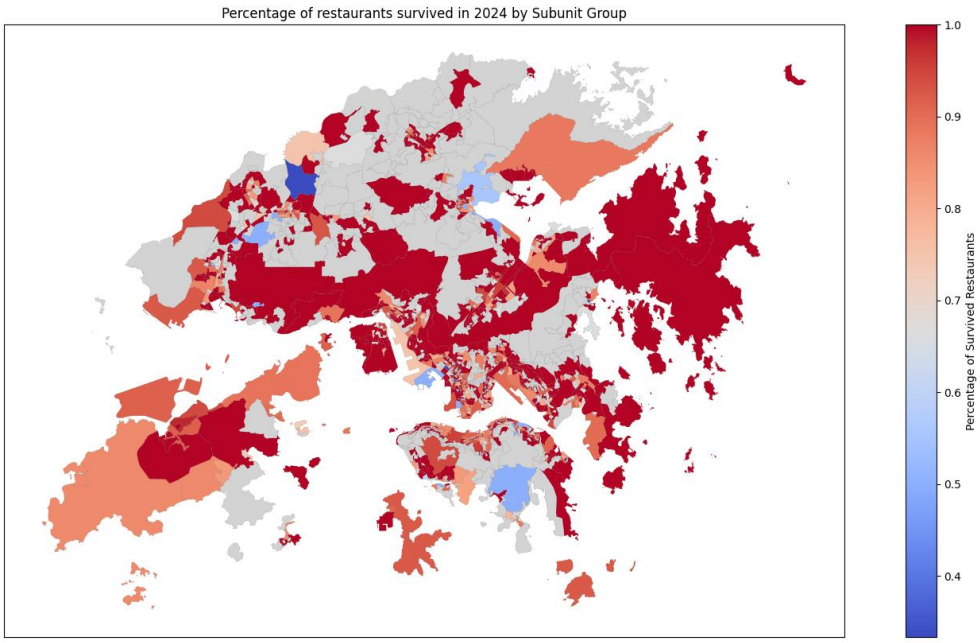
2. Restaurants survived and closed in 2024



3. Number of restaurants in each subunit group



4. Percentage of restaurants survived in 2024 by subunit group



Part 2: Pricing Strategy

1. Raw data (12cols, 48620rows)

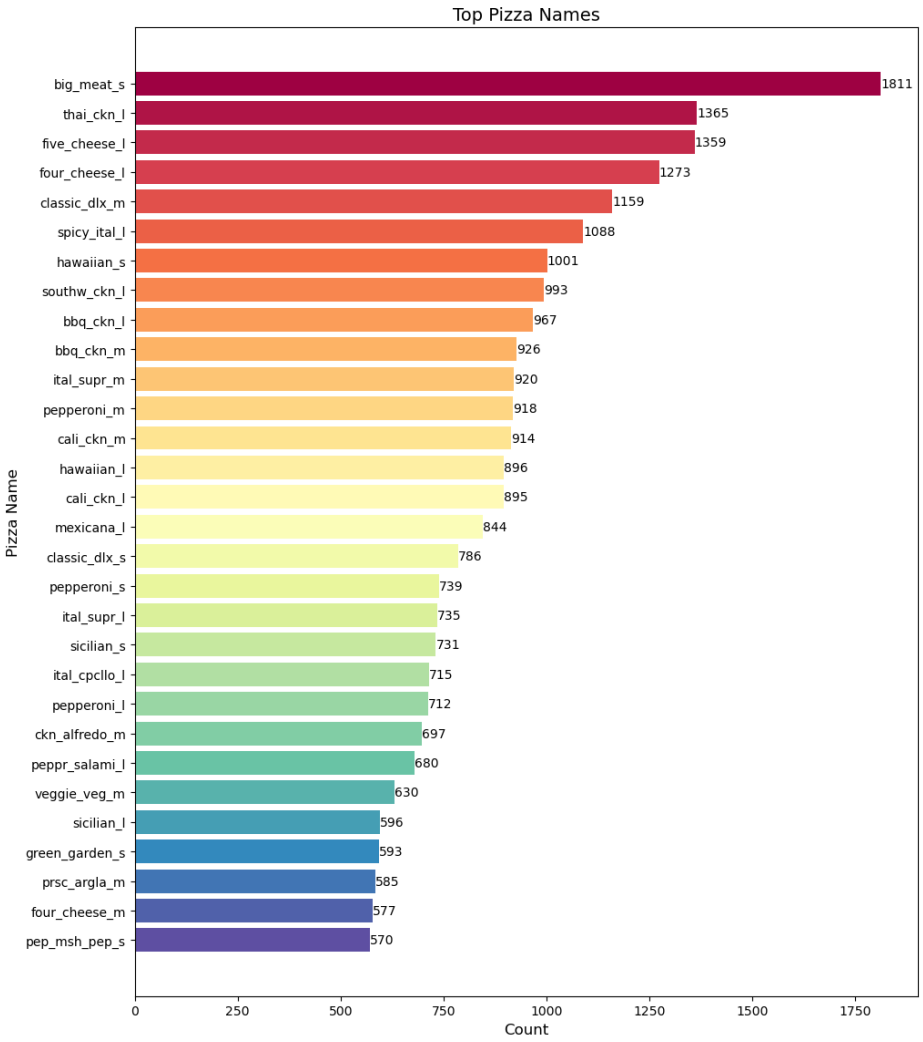
Id	Column Name	Description
1	order_details_id	Unique identifier for each order placed by a table
2	order_id	Unique identifier for each pizza placed within each order (pizzas of the same type and size are kept in the same row, and the quantity increases)
3	pizza_id	Unique key identifier that ties the pizza ordered to its details, like size and price
4	quantity	Quantity ordered for each pizza of the same type and size
5	order_date	Date the order was placed (entered into the system prior to cooking & serving)
6	order_time	Time the order was placed (entered into the system prior to cooking & serving)
7	unit_price	Price of the pizza in USD
8	total_price	unit_price * quantity

9	pizza_size	Size of the pizza (Small, Medium, Large, X Large, or XX Large)
10	pizza_category	Unique key identifier that ties the pizza ordered to its details, like size and price
11	pizza_ingredients	ingredients used in the pizza as shown in the menu (they all include Moz zarella Cheese, even if not specified; and they all include Tomato Sauce, unless another sauce is specified)
12	pizza_name	Name of the pizza as shown in the menu

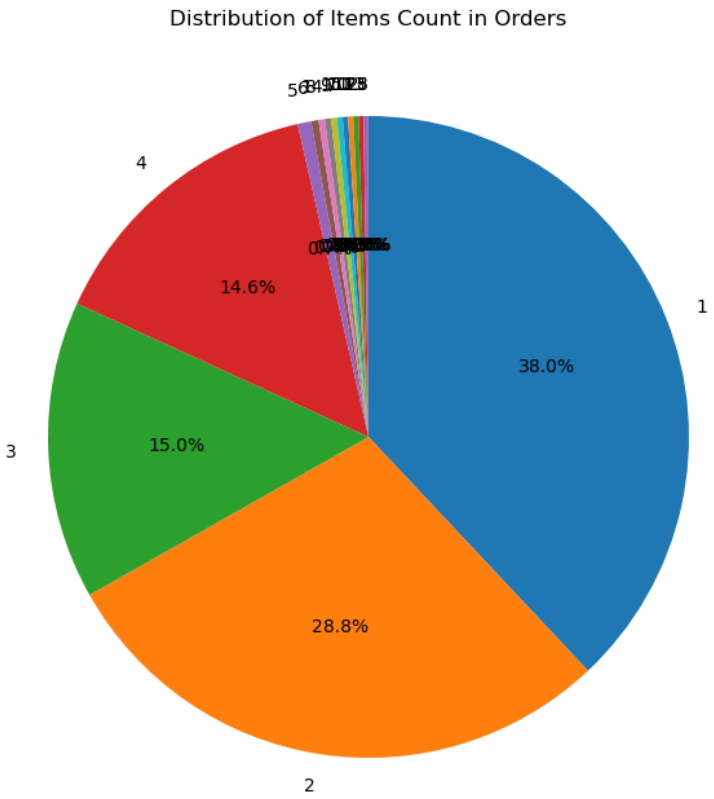
2. Columns obtained after feature engineering--Use data(12cols, 91rows)

Id	Column Name
1	unit_price
2	meat_count
3	vegetable_count
4	cheese_count
5	special_ingredients_count
6	fruit_ingredients_count
7	other_ingredients_count
8-12	size_L,size_M,size_S,size_XL,size_XXL

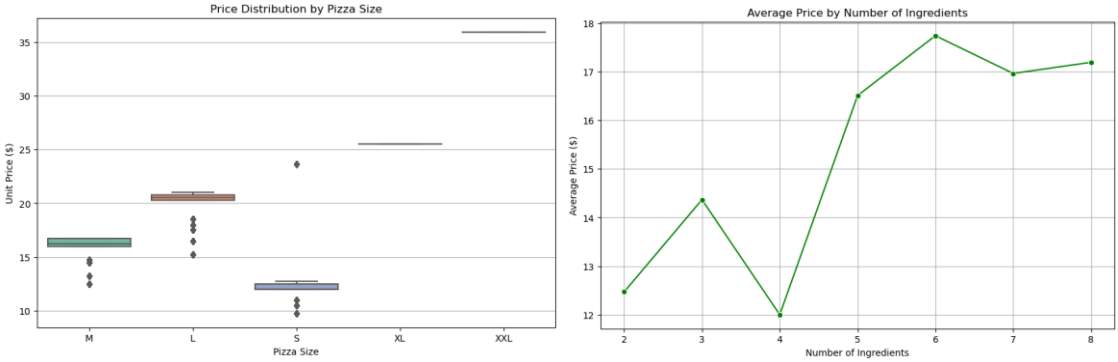
3. The 30 best-selling pizzas



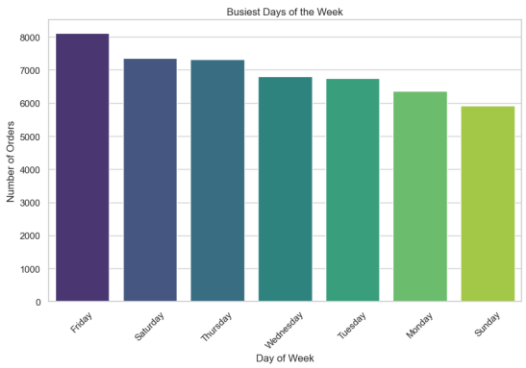
4. Most orders only purchase one type of pizza

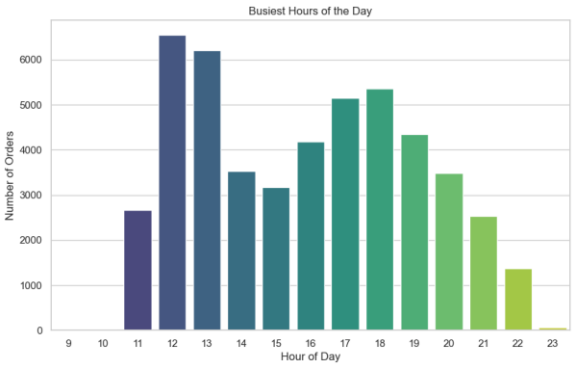


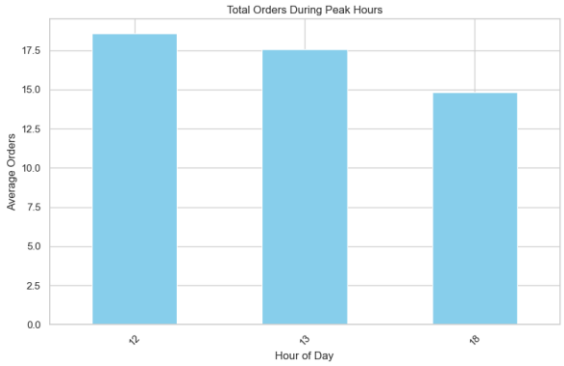
5. Pizza size and number of toppings have a certain impact on price

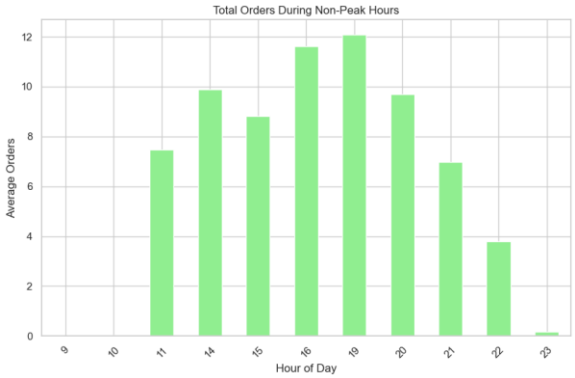


6. Peak times are Friday, Thursday and Saturday, especially lunchtime (12pm and 1pm) and dinner time (5pm to 7pm)

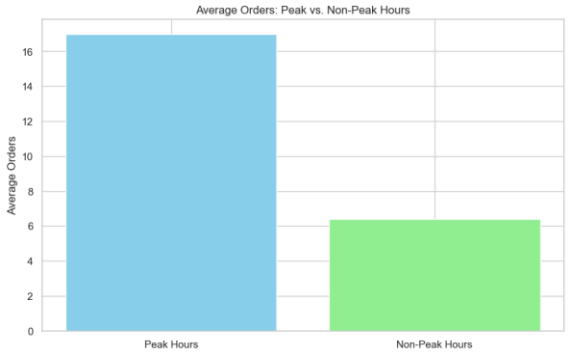


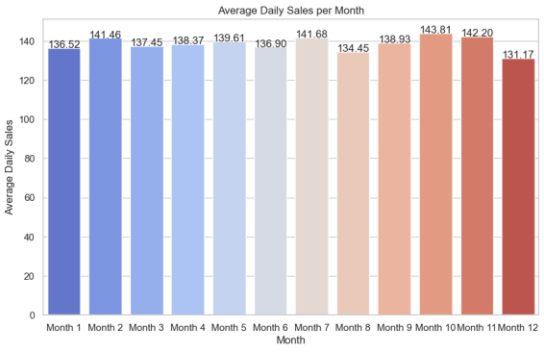






7. The average number of orders is about 17 orders/hour, the average number of orders during off-peak hours is about 6 orders/hour, and the estimated daily order volume is about 138 orders.





Part 3: Customer Concerns

1. Two word cloud visualizations to identify the main keywords in the restaurant review dataset for both positive and negative evaluations.