

PIZZA SALES REVENUE

About Dataset

Contents

This pizza sales dataset make up 12 relevant features:

- **order_id**: Unique identifier for each order placed by a table
- **order_details_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)
- **pizza_id**: Unique key identifier that ties the pizza ordered to its details, like size and price
- **quantity**: Quantity ordered for each pizza of the same type and size
- **order_date**: Date the order was placed (entered into the system prior to cooking & serving)
- **order_time**: Time the order was placed (entered into the system prior to cooking & serving)
- **unit_price**: Price of the pizza in USD
- **total_price**: $\text{unit_price} * \text{quantity}$
- **pizza_size**: Size of the pizza (Small, Medium, Large, X Large, or XX Large)
- **pizza_type**: Unique key identifier that ties the pizza ordered to its details, like size and price
- **pizza_ingredients**: ingredients used in the pizza as shown in the menu (they all include Mozzarella Cheese, even if not specified; and they all include Tomato Sauce, unless another sauce is specified)
- **pizza_name**: Name of the pizza as shown in the menu

Project Overview

Objective: To analyze pizza sales revenue, identify trends, and provide insights to improve sales.

Problem Statement

1. What days and times do we tend to be busiest?
2. How many pizzas are we making during peak periods?
3. What are our best and worst-selling pizzas?
4. And other information

Data Exploration and Cleaning:

data format :- CSV

Libraries Used:

- Pandas (pd) for data manipulation and exploration

1. Data Loading and Overview:

- The `pd.read_csv` function loaded the pizza sales data from a CSV file named "Data Model - Pizza Sales.xlsx - pizza_sales.csv"

2. Initial Examination:

- The `data.head(5)` method displayed the first five rows of the data to get a glimpse of its structure.
- The `data.info()` method provided information about the data types and existence of missing values in each column.
- The `data.shape` attribute revealed the dimensions of the data (number of rows and columns).
- The `data.isnull().sum()` method displayed the total count of missing values present in each column.
- The `data.columns` attribute listed all the column names in the dataset.

3. Data Cleaning:

- **Order Time:** The `order_time` column was converted to a string format using `astype('string')`.
- **Time Extraction:** The `str.split(":", expand=True)` function separated the "order_time" string into separate columns for Hour, Minute, and Second.
- **Date Conversion:** The `pd.to_datetime` function transformed the "order_date" column into a datetime format for further analysis based on dates.
- **Time Conversion:** Similar to the date, the "order_time" column was converted to a datetime format using `pd.to_datetime` with a specific format string `"%H:%M:%S"` to handle the time components (hours, minutes, seconds).

4. Discrepancy Check:

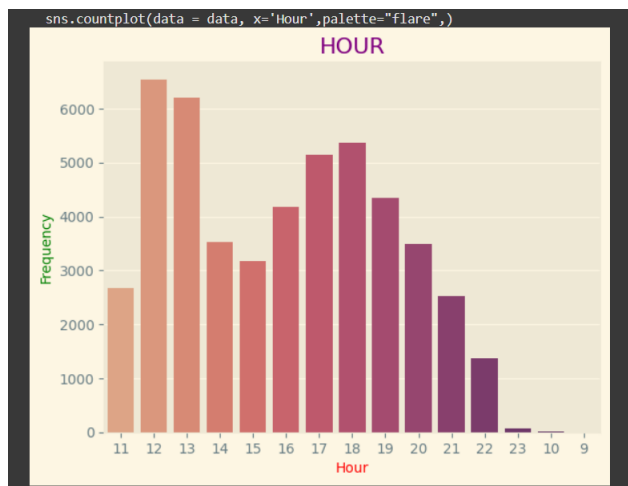
- A new column named "calculated_total" was created by multiplying "unit_price" and "quantity" to represent the expected total price based on individual item prices and quantities.
- The data was then filtered using boolean indexing to identify rows where the "calculated_total" differed from the recorded "total_price". This helped identify potential discrepancies in the total price calculation, highlighting `len(discrepancy)` number of inconsistencies.

Finally, the `data.describe()` method provided summary statistics like mean, standard deviation, minimum, and maximum values for numerical columns in the dataset.

Data Visualizations:

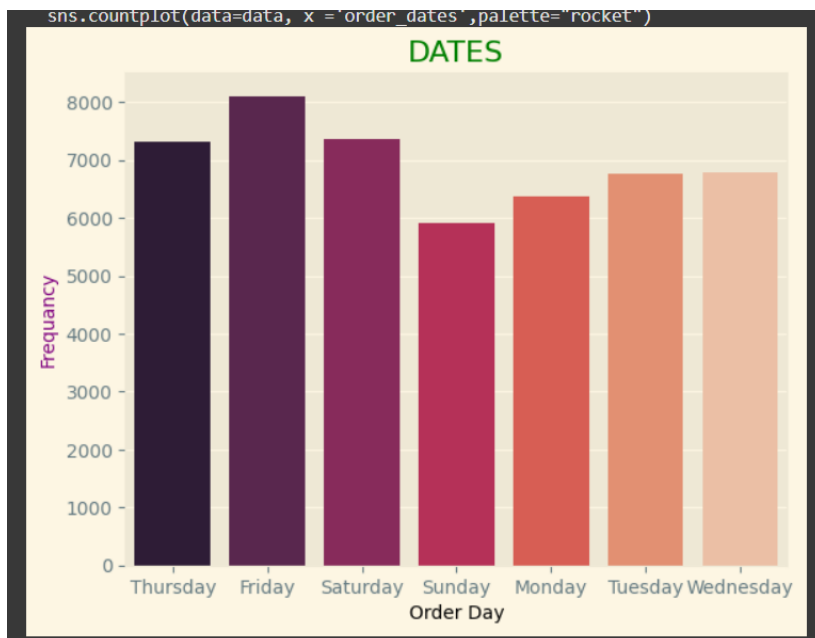
1.

Peak Hours: There are two distinct peak periods: one around lunchtime (12-14 hours) and another during dinner time (18-20 hours)

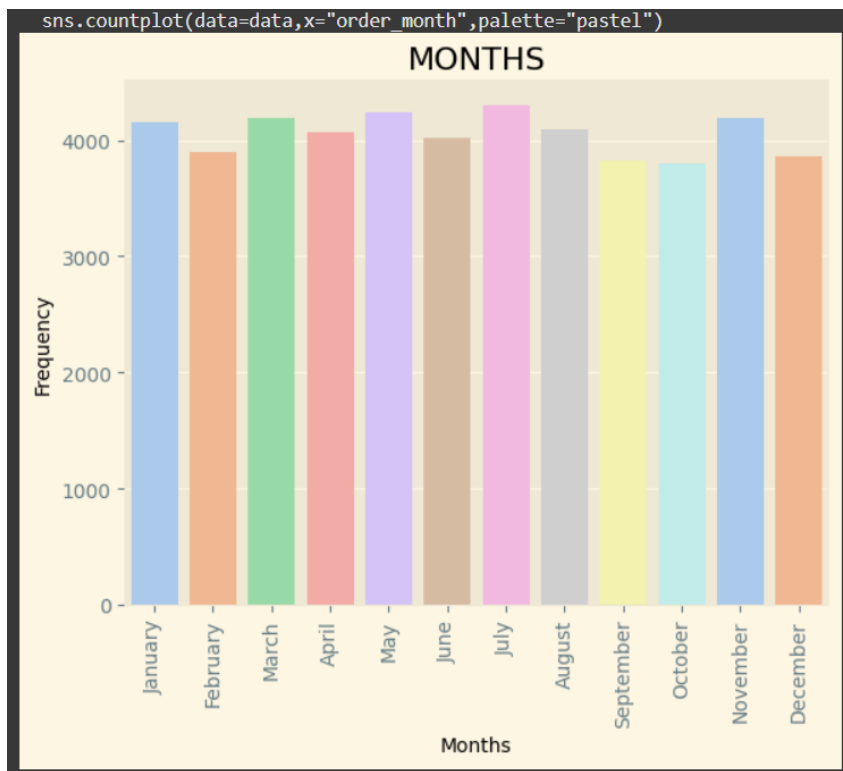


2.

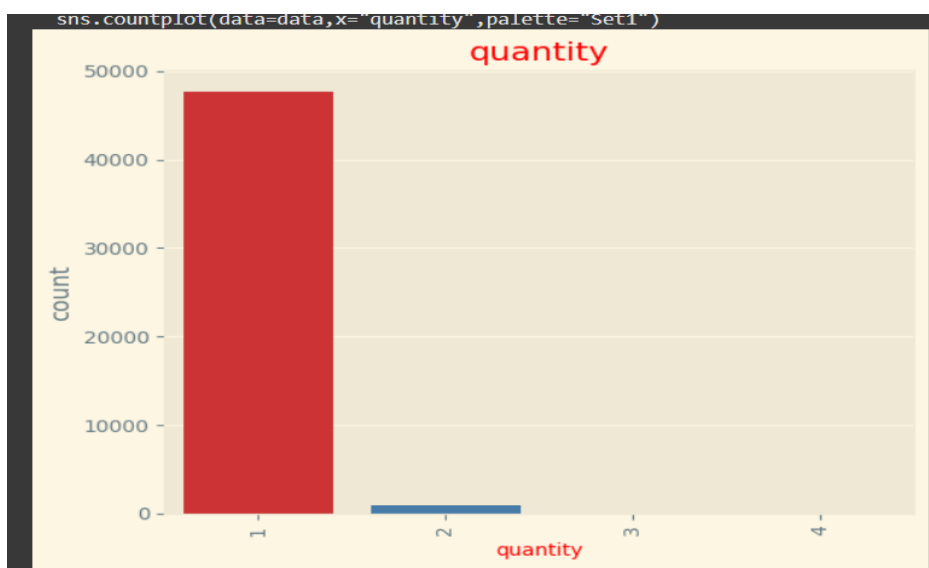
Peak Days: Friday and Saturday appear to be the busiest days for pizza orders, with significantly higher order frequencies compared to other days.



3. Peak Months: You can identify the months with the highest order frequency, which might be indicative of seasonal trends or special occasions.



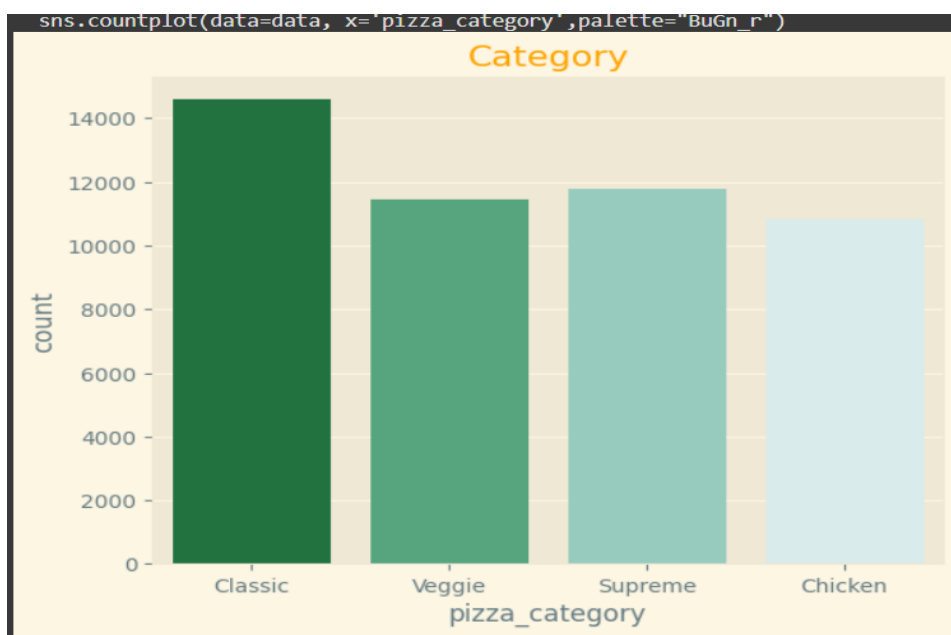
4. **Dominant Order Size:** The chart clearly shows that the majority of orders (represented by the tallest bar) involve purchasing just one pizza.



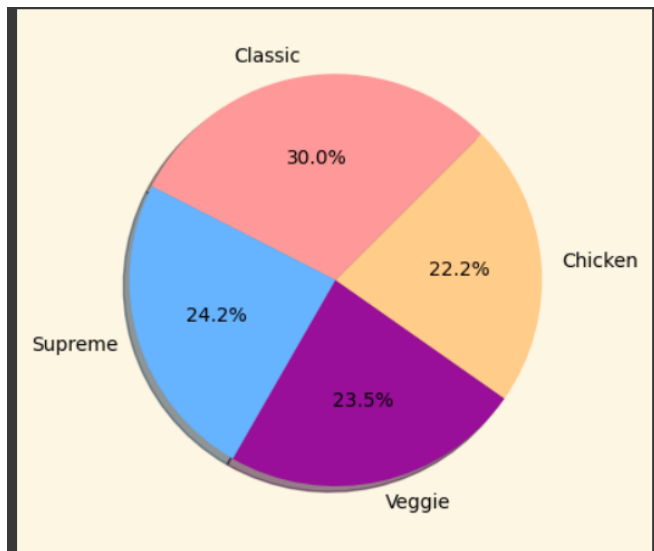
5. **Popular Sizes:** The chart indicates that **L** size is the most popular among customers.



6. **Most Popular Category:** The "Classic" pizza category appears to be the most popular, with the highest number of orders.

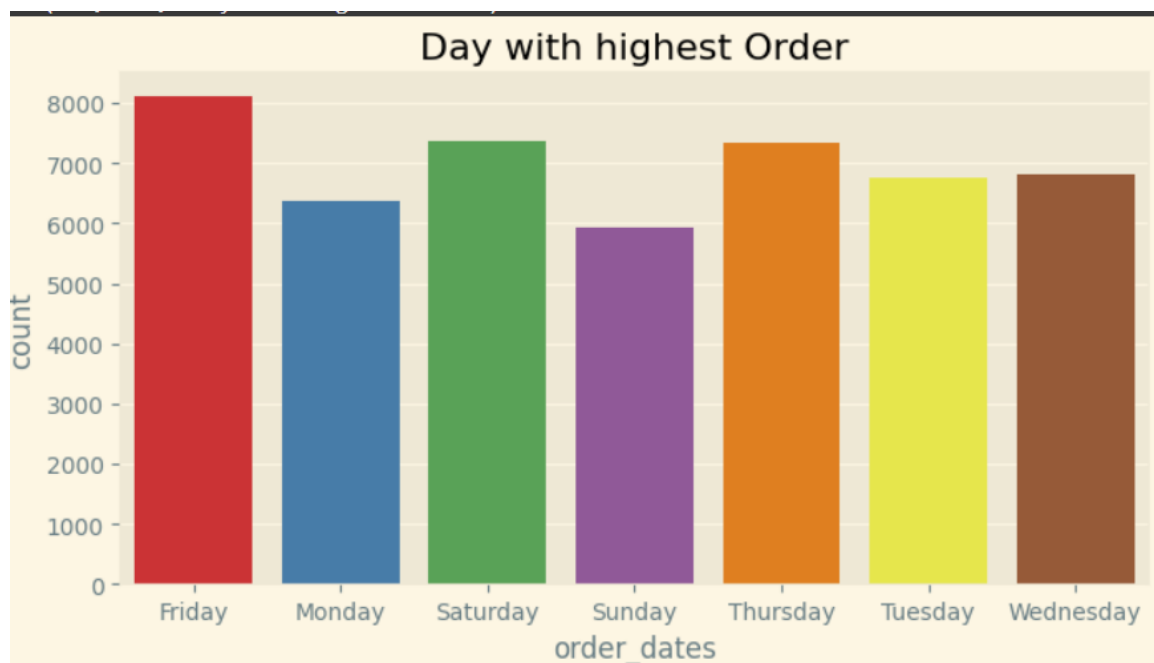


7. **Popular Categories:** The "Classic" category seems to be the most popular, followed by "Supreme".

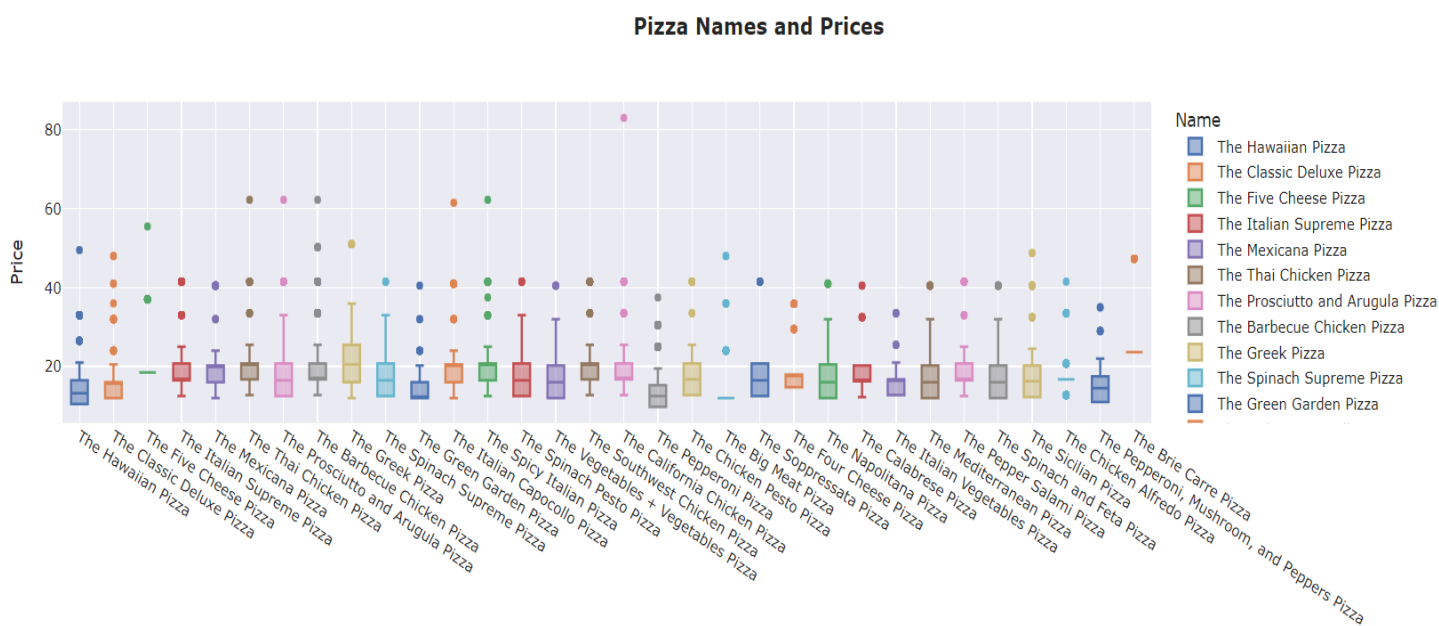


8. **Busiest Days:** Friday and Thursday appear to be the busiest days for pizza orders, with significantly higher order counts compared to other days.

Quietest Days: Monday and Sunday seem to be the slowest days, with lower order volumes.

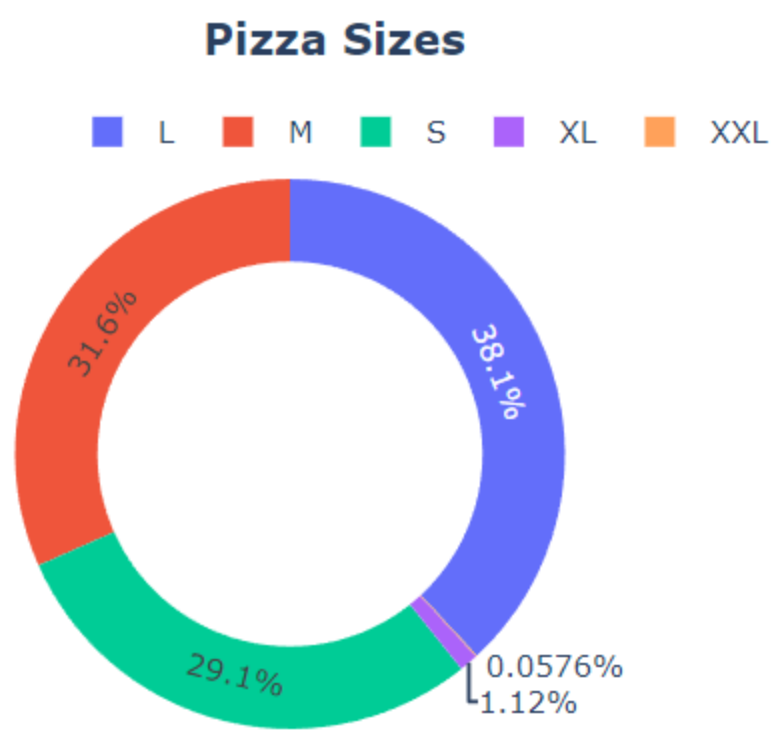


9. **Price Variation:** The chart shows that there is significant variation in prices across different pizza names. Some pizzas have a wider range of prices, while others have a narrower range.

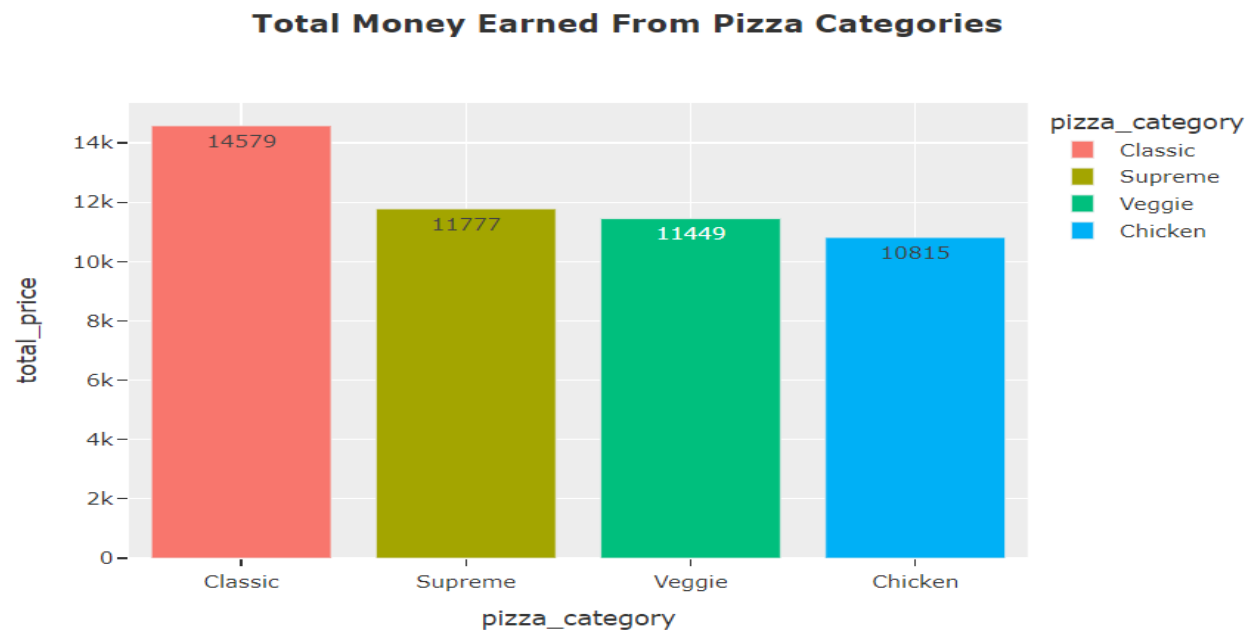


10. **Dominant Size:** The "L" size is the most popular, accounting for a significant portion of the orders.

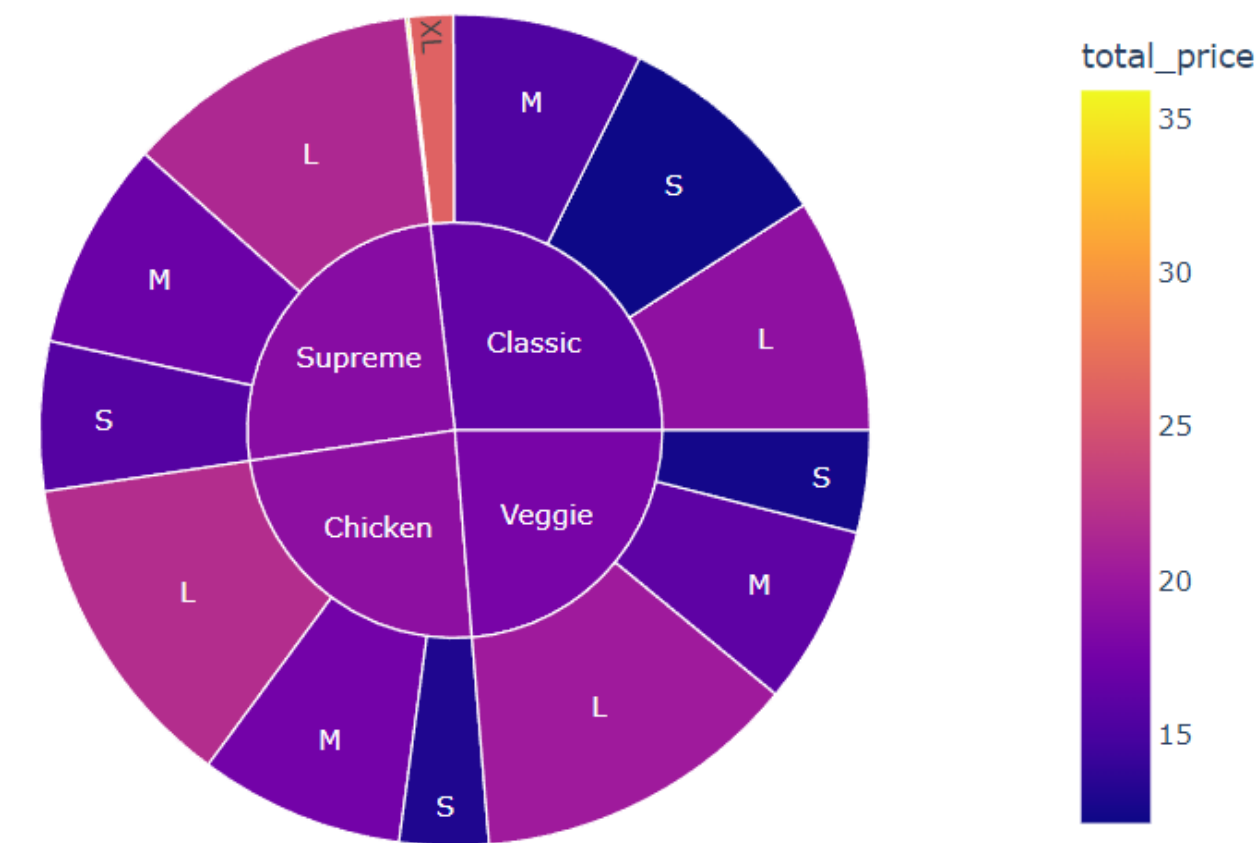
Less Popular Sizes: "XXL" and "XL" sizes have a very small percentage of orders, indicating that they might not be as popular among customers.



11. **Most Profitable Category:** The "Classic" category appears to be the most profitable, generating the highest revenue.

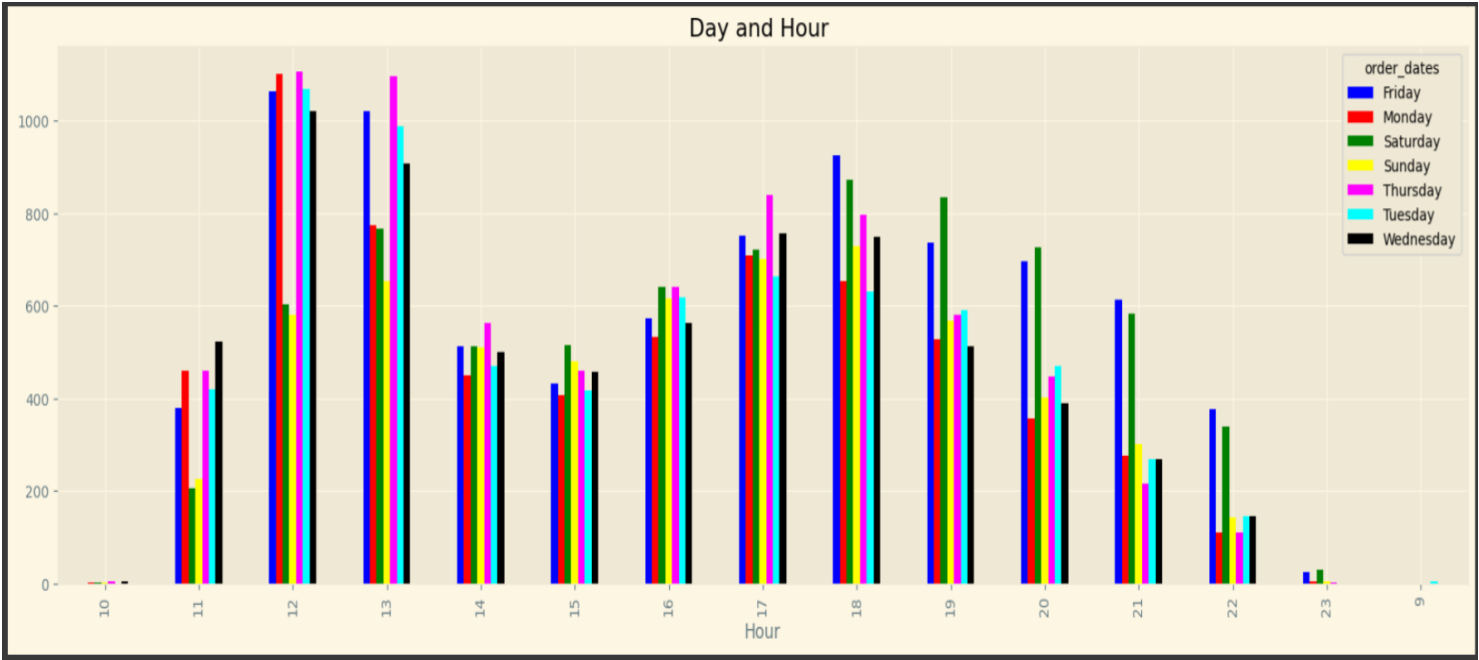


12. **Price Variation:** The color gradient provides insights into the price distribution for different combinations.



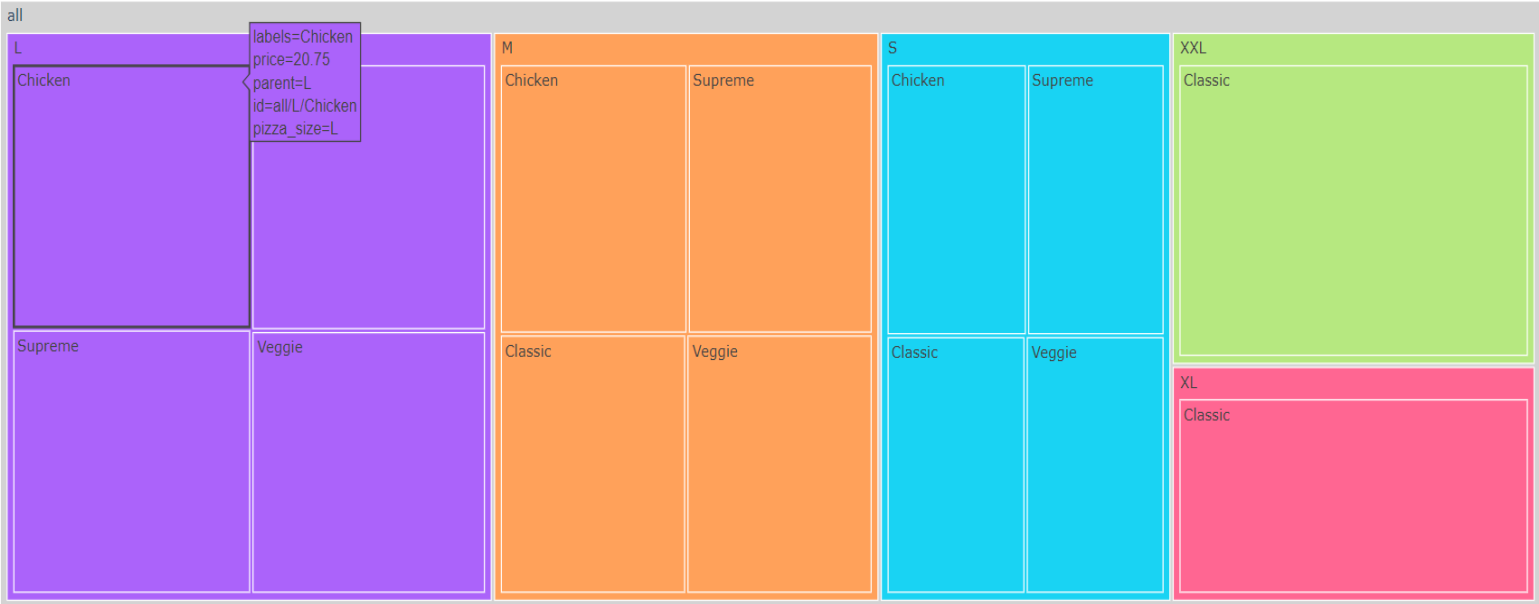
13. **Peak Hours:** You can identify the peak hours for pizza orders, which are typically in the evening (around 12-13).

Day-of-Week Variation: The chart shows that there are variations in order patterns across different days of the week. Some days might have higher order frequencies during specific hours compared to others.



14. Popular Categories: The larger rectangles likely represent more popular or profitable categories.

Popular Sizes: Within each category, the larger rectangles indicate more popular pizza sizes.



Summary of Pizza Sales Analysis

Key Findings:

- **Peak Hours:** The restaurant experiences peak demand during lunchtime and dinnertime.
- **Order Year:** The data analyzed is limited to orders placed in the year 2015.
- **Popular Day:** Fridays are the busiest days of the week for the restaurant.

- Top Seller: "Big meat s" is the most frequently ordered item.
- Group Orders: Single-item orders often indicate group orders.
- Size Preference: The majority of orders are for large-sized pizzas.
- Classic Appeal: Classic pizza flavors remain the top choice for customers.
- Weekend Surge: Order volumes see a noticeable increase on Fridays and Saturdays.
- Core Ingredients: Tomatoes, red peppers, and red onions are staples in most pizza orders.
- Morning Lull: Despite opening at 9 AM, there's minimal customer activity during the early morning hours.

Overall, the analysis provides valuable insights into customer behavior and preferences. The data suggests a strong demand for large-sized pizzas, particularly on weekends, with classic flavors being the most popular. Understanding these trends can help optimize operations, inventory management, and marketing strategies to maximize sales and customer satisfaction.

Data Analysis Tools and Libraries: - Jupyter Notebook, Python, Pandas, Visualization Libraries such as Matplotlib, Seaborn, Plotly.

Jupyter Notebook File:- <https://colab.research.google.com/drive/14Vxl09Lj0tI-doPUmGuRirfB4G9teXjM?usp=sharing>

Data Source:

<https://docs.google.com/spreadsheets/d/1Jps66JinPIDY3oO8Tzkh84fbVfaYTCVt/edit?gid=677877627#gid=677877627>