Leveraging Pizza Sales Data for Business Optimization:

A Case Study by Angelo Gabriel Valenzuela

Abstract:

This paper presents a comprehensive analysis of pizza sales data aimed at optimizing business strategies for a pizza restaurant. Using sales data on Kaggle, we investigate key metrics such as revenue, customer preferences and operational efficiency. Our findings provide actionable insights for enhancing revenue, improving customer satisfaction, and streamlining operations in the pizza industry.

I.   Introduction

The pizza industry is highly competitive, with restaurants constantly seeking ways to improve sales and profitability. In this paper, we leverage sales data from a Greek-inspired pizza place in New Jersey to identify opportunities for business optimization. By analyzing factors such as revenue, customer preferences, and operational efficiency, we aim to provide actionable insights for enhancing business performance.

1. Tools used
- Google Sheets: Data Validation and Cleaning
- SQL: Querying and Grouping data
- Tableau: Data Visualization
- ChatGPT 3.5: DA Assistant / Subject Matter Expert

2. Pizza Restaurant Data

Classification	Field Name	Detail	Data Type	Sample
Dimension	*order_id*	Unique identifier for each order placed by a table	Integer	1
Dimension	*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)	Integer	1
Dimension	*pizza_id*	Unique key identifier that ties the pizza ordered to its details, like size and price	String	hawaiian_m
Metric	*quantity*	Quantity ordered for each pizza of the same type and size	Integer	3

Dimension	order_date	Date the order was placed (entered into the system prior to cooking & serving)	Date	1/13/2015
Dimension	order_time	Time the order was placed (entered into the system prior to cooking & serving)	Time	11:38:36
Metric	unit_price	Price of the pizza in USD	Float	13.25
Metric	total_price	unit_price * quantity	Float	39.75
Dimension	pizza_size	Size of the pizza (Small, Medium, Large, X Large, or XX Large)	String	M
Dimension	pizza_type	Unique key identifier that ties the pizza ordered to its details, like size and price	String	Classic
Dimension	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)	String	Sliced Ham, Pineapple, Mozzarella Cheese
Dimension	pizza_name	Name of the pizza as shown in the menu	String	The Hawaiian Pizza

II. Data Cleaning

1. Extraction

The data used for this study was extracted from Kaggle, the file size is 4MB and is formatted as an xlsx file. The dataset consists of a singular table that contains 12 columns and 48620 rows. The link to the dataset is located in the last page of this document.

2. Data Assessment:

- **Identify Data Types**
  - Used the "data" menu in Google Sheets to format columns with appropriate data types (e.g., dates, numbers, text).
  - Ensure consistent formatting throughout the dataset to avoid data type conflicts and inconsistencies.
- **Check for Duplicates**
  - Used the formula =COUNTIF(range, value) to count duplicate values within the unique identifiers.
- **Verify Data Completeness:**
  - Used the "data" menu to check column stats to detect empty rows.
  - This is applied to all columns to ensure completeness.

- **Ensure Data Consistency:**
  - Used data validation rules to restrict input to specific formats or values. Set data validation rules through the "Data" > "Data validation" menu.
- **Detect Outliers and Anomalies**
  - Used statistical functions such as AVERAGE, STDEV, MIN, and MAX to analyze numerical data for outliers and anomalies.

3. **Cleaning & Transforming**

   The step-by-step process of preparing data for analysis. The steps involved transforming tables and joining them in order to create a separate table for ingredients. Date and time were also reformatted to fit the preferences of Tableau.

- Split pizza_ingredient text to columns

Before

pizza_ingredient
Sliced Ham, Pineapple, Mozzarella Cheese

After

pizza_ingredient_1	pizza_ingredient_2	pizza_ingredient_3
Sliced Ham	Pineapple	Mozzarella Cheese

- Date and Time columns were joined together in a singular cell in order to follow Tableau's format.

Before

order_date	order_time
1/1/2015	11:38:36

After

order_datetime
1/1/2015 11:38:36

- Added a new column "*ingredient_quantity*" using =COUNTA(pizza_ingredient_1,...n)

- Uploaded Sales Table to SQL then transformed data to long data format to query the pizza name and ingredient.

SELECT DISTINCT pizza_name, ingredient
FROM (
*   SELECT pizza_name, pizza_ingredient_1 AS ingredient FROM sales*
*   UNION ALL*
*   SELECT pizza_name, pizza_ingredient_n AS ingredient FROM sales*
*   UNION ALL*
) AS ingredients
WHERE ingredient IS NOT NULL
*   ORDER BY pizza_name, ingredient;*

SQL Query

pizza_name	*ingredient*
hawaiian	Mozzarella Cheese
hawaiian	Pineapple

- Created a temporary table on excel "Ingredient Type" referenced from ChatGPT to classify them into different food groups.

Ingredient Type

ingredients	classification
Mozzarella Cheese	Dairy & Cheese
Pineapple	Fruit

- Created a new table "Inventory**"** then added 2 columns named ingredient_id and classification. The table consists of a total of 4 columns and 181 rows.

- Utilized =VLOOKUP to join the classification column on the "Inventory" table.

Inventory

ingredient_id	pizza_name	ingredient	classification
69	hawaiian	Mozzarella Cheese	Dairy & Cheese
70	hawaiian	Pineapple	Fruits

- Created a new table in Google Sheets "Customer Segmentation" that contains the order_id, total_spending and customer_segmentation. The table has 3 columns and 21,350 rows.

- SQL Query to fill columns order_id and total_spending

SELECT order_id, SUM(total_price) AS total_spending
FROM sales
GROUP BY order_id;

SQL Query

order_id	total_spending
1	13.25
2	92

- Created IF THEN statement on Google Sheets to group "Low Spender", "Medium Spender" and "High Spender"

IF(total_spending<=50,"Low Spender",IF(B2<=100,"Medium Spender","High Spender"))

Customer Segmentation

order_id	total_spending	customer_segmentation
1	13.25	Low Spender
2	92	Medium Spender

III. Data Modeling
1. **Dimensions:** Arranged from the most general level (Order Dimension) to more specific levels (Pizza Dimension and Order Details Dimension).
   a. **Order Dimensions**
      i. order_id
      ii. order_date
      iii. order_time
      iv. total_price
   b. **Pizza Dimensions**
      i. pizza_id
      ii. pizza_type
      iii. pizza_name
      iv. pizza_size
      v. pizza_ingredients
   c. **Ingredients Dimension**
      i. ingredient_id
      ii. ingredient_quantity
      iii. ingredient_classification
   d. **Order Details Dimension:**
      i. order_details_id
      ii. quantity
      iii. unit_price
   e. **Customer Segmentation Dimension:**
      i. order_id
      ii. customer_type

2. **Metrics**
   a. **Sales Metrics**
      i. Total Sales Revenue
      ii. Average Order Value
      iii. Sales Growth Rate
      iv. Sales by Date and Time: Hourly, Daily, Weekly and Monthly
   b. **Customer Metrics**
      i. Quantity of Customers by Date and Time: Hourly and Daily
      ii. Average Quantity of Pizzas / Table
   c. **Menu Performance Metrics**
      i. Top Performing Pizzas
      ii. Underperforming Pizzas
      iii. Seasonal Pizzas
   d. **Pricing Metrics**
      i. Average Cost by Pizza Size
      ii. Average Cost by Pizza Type

  e. Inventory Management Metrics
   i. Most Common Ingredients
   ii. Least Common Ingredient
   iii. Seasonal Ingredients
  f. Customer Segmentation Metrics
   i. Total Spending

IV. User Stories

  As the pizza shop owner I would only want to see the most important charts and graphs on my dashboard. There are several key metrics that would be crucial to monitor and to ensure the success and profitability of your business.

1. Insights
  a. Guide Questions
   i. What is the total revenue?
   ii. What is the average order value?
   iii. What are the most in-demand pizzas, sizes and ingredients?
   iv. What ingredients should the restaurant be careful about ordering?
   v. When is peak ordering time / season for the restaurant?
   vi. How can we streamline and optimize during peak times?
   vii. How frequently do customers come into the restaurant?
   viii. Who are our customers? (Vegetarians, Families, Generation)

2. Visualizations and Charts
  a. Score Cards
   i. Total Sales Revenue
   ii. Average Order Value
   iii. Average Quantity of Pizzas
   iv. Unique ingredients by Type

  b. Line Charts
   i. Sales by Date and Time: Daily and Monthly
   ii. Quantity of Customers by Date and Time: Hourly and Daily
  c. Bar Charts
   i. Top Performing Pizzas
   ii. Underperforming Pizzas
   iii. Most Demanded Ingredients
   iv. Least Demanded Ingredients
  d. Scatter Plot
   i. Price vs Quantity Sold Relationship:
  e. Heat Map
   i. Quantity of Orders: Hourly

V. Dashboard Documentation

Below are the metrics, dimensions and filters used in the dashboard. Using the variables below, the dashboard was made to be highly interactive and intuitive using text size, borders and color coded texts/graphs for the pizza manager to effortlessly inspect the information across many different levels.

1. Metrics

a. Sales

i. CM Sales (Current Monthly Sales)

IF DATEPART('month', [Order Datetime]) = [Select Month] THEN [Total Sales] END

ii. CM Volume (Current Monthly Volume)

IF DATEPART('month', [Order Datetime]) = [Select Month] THEN [Volume] END

iii. CQ Ingredients Qty. (Current Quarter Ingredients Quantity)

IF DATEPART('quarter', [Order Datetime]) = [Select Quarter] THEN [Ingredients Quantity] END

iv. CQ Sales (Current Quantity Sales)

IF DATEPART('quarter', [Order Datetime]) = [Select Quarter] THEN [Total Sales] END

v. CQ Volume (Current Quantity Volume)

IF DATEPART('quarter', [Order Datetime]) = [Select Quarter] THEN [Volume] END

vi. Ingredients Quantity
   - Ingredients x Quantity of Pizzas

vii. PM Sales (Previous Monthly Sales)

IF DATEPART('month', [Order Datetime]) = [Select Month]-1 THEN [Total Sales] END

viii.   PQ Sales (Previous Quarter Sales)

        IF DATEPART('quarter',[Order Datetime]) = [Select Quarter]-1 THEN [Total Sales] END

ix.   Total Sales
- Unit Price x Quantity

x.   Unit Price
- Price per pizza

xi.   Volume
- Quantity of Pizzas Sold

xii.   Sales (Count)
- Quantity of Sales

xiii.   Day (Order Datetime)
- Daily timeframe for dynamic charts

b. Inventory

i.   Ingredient Id
- Primary Key for Inventory Table

ii.   Inventory (Count)
- Number of Rows in the Table

c. Customer Segmentation

i.   Total Spending
- Total Amount Spent per Customer
- SQL Query Order_id x Total_Price

ii.   Customer Segmentation
- Customer Type
- Less than $50 = Low Spender
  Between $50 and $100 = Medium Spender
  Above $100 = High Spender

d. Others

i.   % Diff Sales
- Percentage comparison between previous month and current month

        (SUM([CM Sales]) - SUM([PM Sales])) / SUM([PM Sales])

ii.   AOV (Average Order Value)

        SUM([Total Sales]) / COUNTD([order_id])

    iii.   Min Max Monthly Sales

```
IF SUM([CM Sales]) = WINDOW_MAX(SUM([CM Sales]))
THEN
    SUM([CM Sales])
ELSEIF SUM([CM Sales]) = WINDOW_MIN(SUM([CM
Sales])) THEN
    SUM([CM Sales])
END
```

    iv.   Min Max Total Sales

```
IF SUM([Total Sales]) = WINDOW_MAX(SUM([Total
Sales])) THEN
    SUM([Total Sales])
ELSEIF SUM([Total Sales]) = WINDOW_MIN(SUM([Total
Sales])) THEN
    SUM([Total Sales])
END
```

2. Dimensions
a. Sales
    i.   Order Datetime
    ii.   Order Details Id
    iii.   Order_id
    iv.   Pizza Category
    v.   Pizza Name (Sales)
    vi.   Pizza Size

b. Inventory
    i.   Classification
    ii.   Ingredient
    iii.   Pizza Name

c. Customer Segmentation
    i.   Customer Type
    ii.   Order id
d. Others
    i.   KPI CM less PM

```
IF SUM([CM Sales]) < SUM([PM Sales]) THEN "▼"
ELSE ''
END
```

3. Filters

    a. Select Month

        i. Line Chart: Daily Sales

        ii. Heat Map: Hourly Sales

        iii. Bar Chart: Pizza Volume

    b. Select Quarter

        i. Bar Chart: Ingredient Demand

        ii. Bar Chart: Sales by Pizza Complexity

        iii. Scatter Plot: Volume vs. Sales by Size

4. Chart Placement

    a. Sales Dashboard

        i. Top Left

            - Scorecard: Key Performance Indicators

                a. Total Sales

                b. Pizza's Sold

                c. Average Order Value

        ii. Bottom Left

            - Bar Chart: Customer Segmentation Sales

                a. High Spender

                b. Medium Spender

                c. Low Spender

        iii. Left Side

            - Line Chart: Monthly Sales

                a. Total Sales for Current Month in Text

                b. % Difference Previous vs. Current Month

                c. Sales Daily Performance: Current Month

                d. Sales Daily Performance: Previous Month

        iv. Right Side

            - Bar Chart: Quantity Sold

                a. Pizza's Sold: Current Month

                b. Pizza's Sold: Previous Month

        v. Bottom Right

            - Heat Map: Weekly Sales

                a. Peak Hours

                b. Dead Hours

        vi. Top Right

            - Select Month Filter

            - Navigation Buttons

                a. Sales Dashboard

                b. Inventory Dashboard

b. Inventory Dashboard
- i. Top Left
  - Scorecard: Classification
    - a. Number of Unique Ingredients / Type
- ii. Left side
  - Bar Chart: Ingredient Demand
    - a. Unique Ingredients
    - b. Demand
- iii. Top Right
  - Scatter Plot: Quantity vs. Sales by Pizza Size
    - a. Pizza Name
    - b. Pizza Size
    - c. Gross Profit
    - d. Volume Sold
  - Navigation Buttons
    - e. Sales Dashboard
    - f. Inventory Dashboard
  - Quarterly Filter

- iv. Bottom Right
  - Bar Chart: Sales by Pizza Complexity
    - a. Pizza's Complexity Grouped by Color