**Pizza sales Analysis**

Author: Nityam Jain

Date:12-07-2024

**Description:**

The purpose of this document is to discuss the sales data of a pizza restaurant and its analyses in details. The first purpose is to analyse pizza revenue and recognize best-selling varieties together with the overall elevated revenue throughout the years. In terms of data analysis tools, the SQL language is used for extracting information from the database while Excel is used for further data analysis and visualization.

**Project Components:**

**SQL**: SQL is the language that is applied at the surface to communicate with the database and to query the data to solve smart queries. SQL scripts are used to define tables, data insertions, and execute general, specific, aggregates and calculations.

**Excel**: Microsoft Excel is helpful in visual analysis of the data collected in that it is used to display the data in the form of charts or graphs, depending on the trends to be analyzed. It also has an advantage of conducting other data analysis and summarize the results.

**Schema** **Creation**: Tables for different entities of the application and their relations are designed in the MySQL database with proper naming conventions. This involves setting up relationships between these tables, for instance, the orders table, order details table, pizzas table and pizza types of tables.

**Insight** **Extraction**: It is made by creating SQL queries to extract several useful information points: the distribution of the revenue per type of pizza and the growth of the total revenue over time in percentage.

**Visualization**: Deploying Excel helps in converting the extracted insights into graphic forms for easier interpretation of data.

This analysis will help in making informed business decisions, stock management, and the promotion of goods and services to improve sales and corporate revenues. This keeps one from focusing on the shortcomings of sales and the overall performance of the restaurant in equal measure.

**Data Source:**

The current data set for this analysis was obtained from Kaggle uploaded by “Mysar Ahemed Bhat”, which is a platform for datasets. It has usability of 10.00 so we don’t have to perform data cleaning operation.

It consists of all the information about orders and order details and pizza and type of pizza. The given set of data covers essential transactional details like the order date, type of pizza, number of pizzas sold, and the price, which will be useful for evaluating the sales trends daily.

This dataset was chosen because it is related to the restaurant industry and the analysis of the revenue and customers’ preferences in this field is quite possible when using this data.

Data set link:

<https://www.kaggle.com/datasets/mysarahmadbhat/pizza-place-sales>

A screenshot of a computer

Description automatically generated

In further process we locally download dataset form Kaggle read description and meta data about whole dataset and identify the whole scenarios to better understand the primary and foreign keys , columns ,datatype of columns etc

**Data Schema:**

The data schema has a significant function in defining the pattern or structure in which the imported dataset is stored within the MySQL database. Here’s an outline of how the schema was designed for this Pizza Sales Analysis project:

Here’s an outline of how the schema was designed for this Pizza Sales Analysis project:

The dataset consists of 4 Table which are:

1.order\_details 2. orders

|  |  |
| --- | --- |
| **Columns** | **Datatype** |
| order\_details | int |
| order\_id | int |
| pizza\_id | text |
| quantity | int |

|  |  |
| --- | --- |
| **Columns** | **Datatype** |
| order\_id | int |
| date | datetime |
| time | text |
| 21350 X 3 |  |

48620 X 4

3.pizza\_types 4. pizzas

|  |  |
| --- | --- |
| **Columns** | **Datatype** |
| pizza\_type\_id | text |
| name | text |
| category | text |
| ingredients | text |

|  |  |
| --- | --- |
| **Columns** | **Datatype** |
| pizza\_id | text |
| pizza\_type\_id | text |
| size | text |
| price | float |

32 X 4 96 X 4

The arrows represent the Connection between different tables With primary keys and foreign keys which further use to join different table to find different insights

**Data Loading process:**

For perform SQL quarries in order to find desire tables or insight we use MySQL Database, First we create a Database in MySQL by a this Quary:

Blue text on a white background

Description automatically generated

A group of white objects with black text

Description automatically generatedIt will create a database name pizzadb

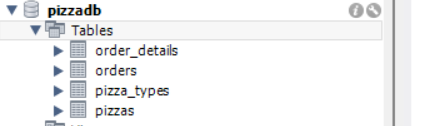
To import the CSV file, we can use the “Table data import wizard” option in MySQL to directly load and create the table from the CSV file. However, this process when applied on large files may be cumbersome and may have some performance concern. When it comes to handling and processing the import process as efficiently as possible, we can use the Jupyter Notebook alongside Python scripts where you receive full access to the powerful pandas for data processing along with MySQL-connector-python for the database connection and the os module for working with files.

Here, we can use pandas to load the given CSV into a Data Frame, which gives convenience for invoking multiple data operation functions. After washing and pre-processing the data, the next logical step is to import the data from a Data Frame into MySQL where we can use MySQL-connector-python.

Here is the python script which we use to load data into MySQL which also create tables according to csv:

A screenshot of a computer

Description automatically generatedpreview



once all the csv files are imported in MySQL in form of tables we can perform different SQL queries on them to find the specific insights.

**Business insights/questions:**

From the pizza dataset, it is therefore possible to answer business questions that will be informative to the overall progress of the business. This is achieved through the assessment of past and current rates of sale, the performance of the regions, customer information, promotions and the feedback received to determine where improvements and expansions need to be made.

The business question is in below link:

<https://github.com/Nityamjain/Pizza-sales-analysis/blob/919dc0a7376135d89d5e9a2b26370923aa0c5a63/Business%20Questions.txt>

Note: the Credit of source of this question goes to Ayushi Jain: [https://github.com/Ayushi0214/Ecommer...](https://www.youtube.com/redirect?event=video_description&redir_token=QUFFLUhqbUdMZFlub19oX183V0xCeWhBMzlXd2xTclFEd3xBQ3Jtc0trVjFvaUpqWXlLLTFFRjQybXZ1Sk5MSXd2clZJYzd4c21mTHFJeUhDWTFKbF9weFAxbWFzRWhBcnZDZTZVejRjbEVnOGJEYnRDRkZjV3dDdUVtdHliVEhTUjFFQVdQb19Tb1ZORWF6N0ttRUVsNHkyOA&q=https%3A%2F%2Fgithub.com%2FAyushi0214%2FEcommerce_SQL&v=2w-M4z-iQpk)

**Answering Business question through queries**:

**1.Retrieve the total number of orders placed.**

Ans: 48620

A screenshot of a computer

Description automatically generatedQuery:

**2.Calculate the total revenue generated from pizza sales**

Ans : 817860

A screen shot of a computer

Description automatically generatedQuary:

**3.Identify the highest-priced pizza.**

Ans:



Quary:

A computer screen shot of a computer code

Description automatically generated

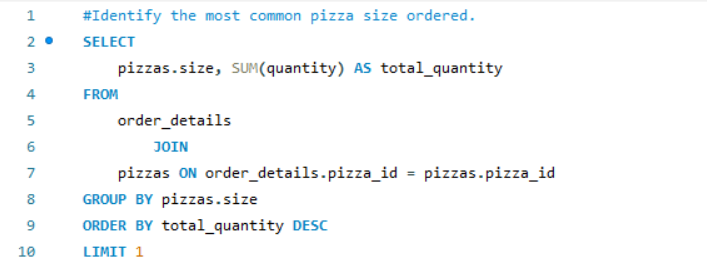
**4.Identify the most common pizza size ordered**.

Ans;

A screenshot of a computer

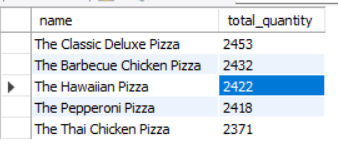
Description automatically generated

Quary;



**5.List the top 5 most ordered pizza types along with their quantities.**

Ans:



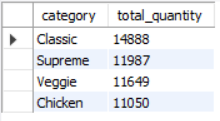
Query:

A computer screen shot of a computer code

Description automatically generated

**6.Find the total quantity of each pizza category ordered.**

Ans:



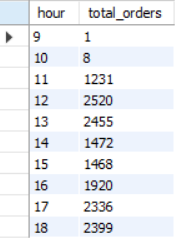
Query :

A screenshot of a computer code

Description automatically generated

**7.Determine the distribution of orders by hour of the day.**

Ans :



Query:

A screen shot of a computer

Description automatically generated

**8. Find the category-wise distribution of pizzas.**

Ans:

A screenshot of a computer

Description automatically generated

Query:

A screenshot of a computer

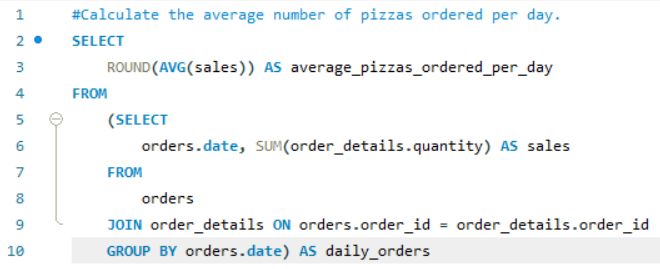
Description automatically generated

**9. Calculate the average number of pizzas ordered per day.**

Ans:



Query:



**10. Determine the top 3 most ordered pizza types based on revenue.**

Ans:

A screenshot of a computer

Description automatically generated

Query:

A computer screen shot of a computer code

Description automatically generated

**11. Calculate the percentage contribution of each pizza type to total revenue.**

Ans:

A screenshot of a computer

Description automatically generated

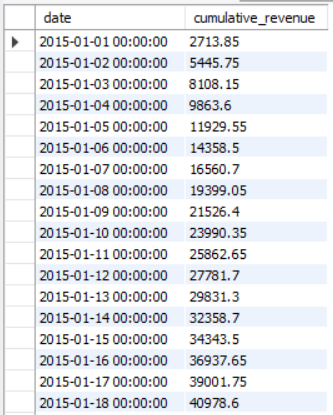
Query:

A screenshot of a computer

Description automatically generated

**12. Analyse the cumulative revenue generated over time.**

Ans:



Query:

A screenshot of a computer code

Description automatically generated

**13. Determine the top 3 most ordered pizza types based on revenue for each pizza category.**

Ans:

A screenshot of a computer

Description automatically generated

Query:

A screenshot of a computer program

Description automatically generated

All the SQL queries are available at my GitHub repositories:

<https://github.com/Nityamjain/Pizza-sales-analysis/tree/919dc0a7376135d89d5e9a2b26370923aa0c5a63/queries>

**SQL concepts used in above queries are:**

**1.Subqueries**: **Subqueries**: query nested inside another query. It is used to perform intermediate calculations or retrieve data that will be used by the main query.

**2.Joins:** used to combine rows from two or more tables based on a related column between them.

**3. Aggregation**: functions like SUM, AVG, COUNT, MIN, and MAX are used to perform calculations on multiple rows of a table to return a single value.

**4.GROUP BY**: The GROUP BY clause groups rows that have the same values in specified columns into summary rows, often used with aggregation functions.

**5.Window Functions**: perform calculations across a set of table rows related to the current row. They are often used for ranking, running totals, moving averages, etc.

**6.Filtering**: involves using conditions to restrict the rows returned by a query, typically done with the WHERE clause.

**Other SQL Concepts**

1. **SELECT:**
   * **Definition:** The SELECT statement retrieves data from a database. It is the most commonly used SQL command**.**
   * **Usage:** Used to specify the columns and data to be retrieved.
2. **ORDER BY:**
   * **Definition**: The ORDER BY clause is used to sort the result set of a query by one or more columns.
   * **Usage:** Helps in ordering the retrieved data, such as sorting products by price or customers by name.
3. **DISTINCT:**
   * **Definition:** The DISTINCT keyword is used to return only distinct (different) values.
   * **Usage:** Useful for removing duplicates from the result set.

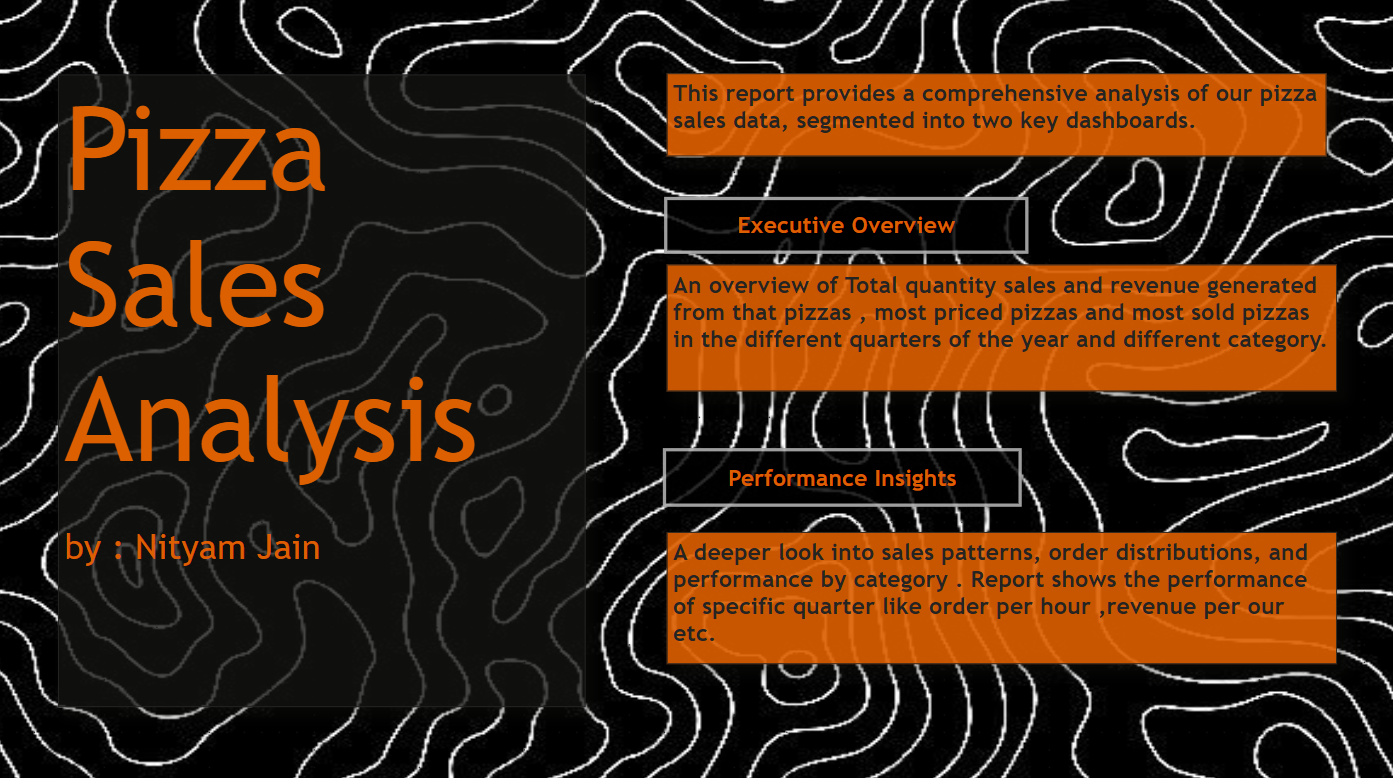
**Dashboard Representation:**

**Introduction:**

To represent the answers to the above queries, we utilized the visualization tool Power BI in this project. Power BI enables us to create interactive and comprehensive dashboards, providing clear and actionable insights from the data analysis. We imported some MySQL queries into Power BI to create precise and interactive visuals and designed a schema that supports all filter changes.

In this power bi report, we have a cover page and two dashboards which are:

**1.Cover page:**



Cover pages include Title of the report and navigation button for both the dashboard

With their basic descriptions.

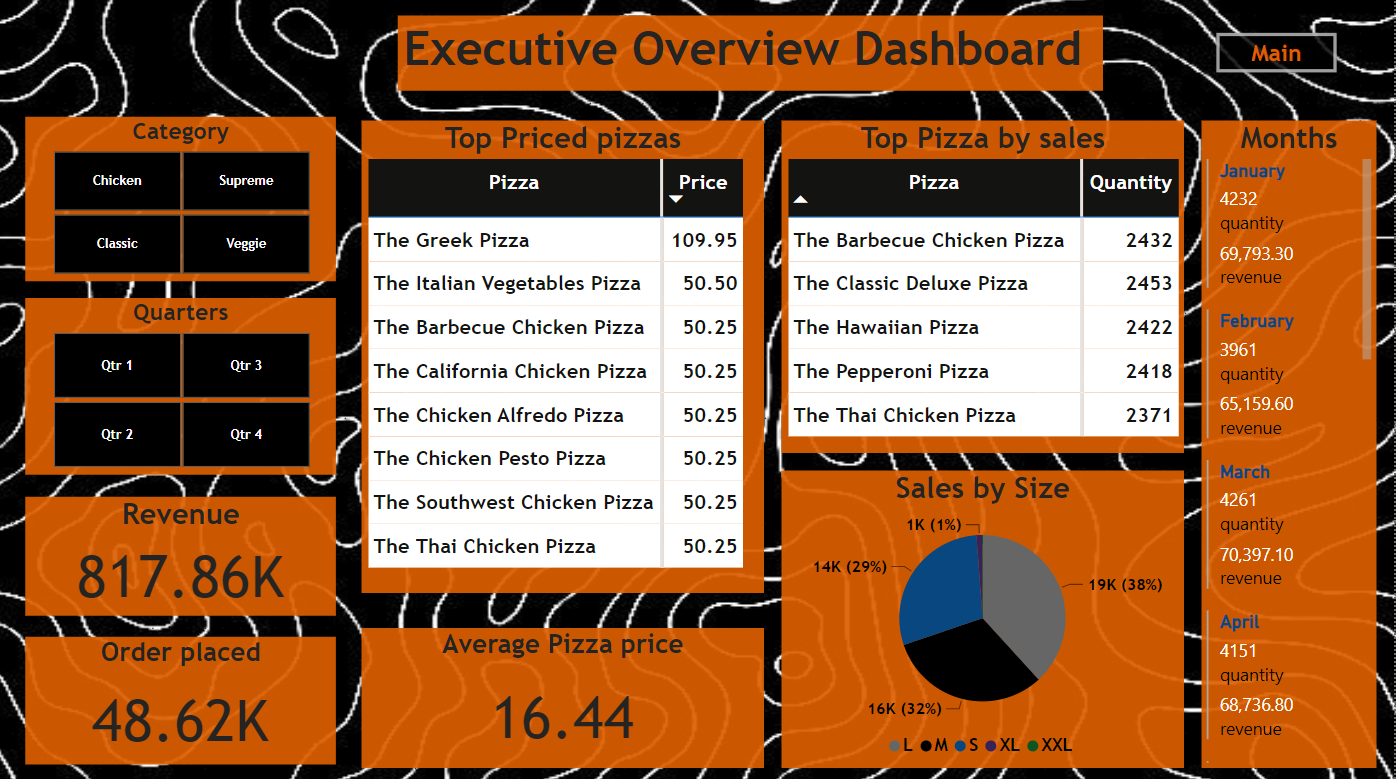
**2.Executive Overview Dashboard:**

**Description:**

* **Revenue**: Displays the total revenue generated within the selected period.
* **Order placed**: Shows the total number of orders placed.
* **Average pizza price** shows the average price of pizza sales in particular period
* **Top pizza by sales**: Lists the pizzas with the highest sales.
* **Top priced pizzas** show the highest price pizza and its price**.**
* **Sales by size:** percentage distribution of pizza size sales**.**
* **Months:** give the information of sales and revenue of Months.
* **Slicers**:

**Category:** Allows filtering the visuals based on different pizza categories, providing a focused view of specific product segments.

**Quarter:** Enables filtering the data by fiscal quarters, allowing for the analysis of seasonal trends and performance over different time periods.



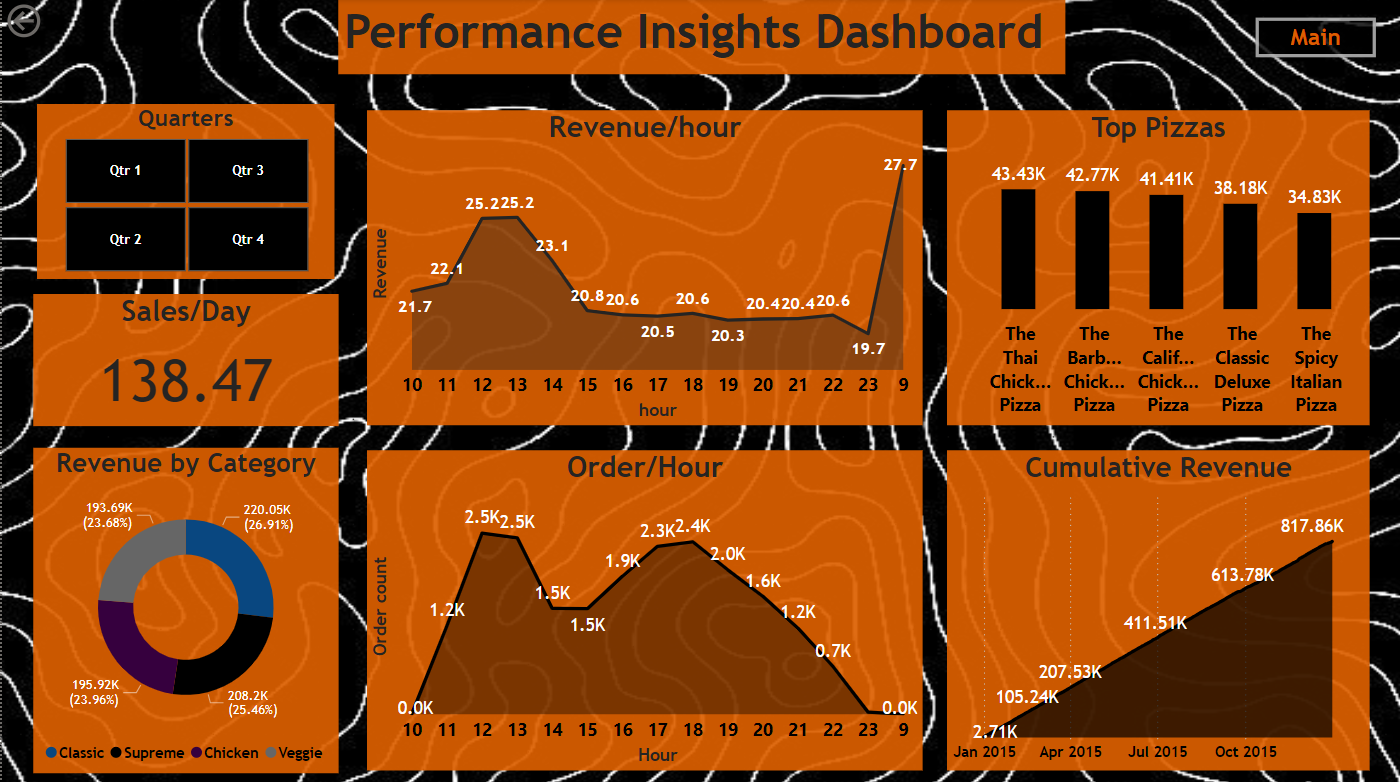
**Insights and Interpretations:**

* **Revenue and Order Placed:** It gives a broad picture of the total sales picture and customer activity going on at the given organizational level.
* **Average Pizza Price:** Shows the customers’ spending trends and checking the marks on the pricing strategies
* **Top Pizza by Sales:** Shows the most frequently sold articles, helping to decide on the amount of each product in the store and the advertising campaigns.
* **Top Priced Pizzas:** Indicates clients are willing to purchase high quality goods and services influencing menu prices and special offers.
* **Sales by Size:** It explains the size popularity for pizzas, helping modify menu offerings.
* **Monthly Sales and Revenue:** Recognises the fiscal seasons and the peak selling period which is crucial when planning for offers as well as stock control.
* **Slicers:**

**Category:** Enables one to hone down his/her scrutiny to certain product segments.

**Quarter:** It can assist in finding out seasonal trends as well as planning for performance for the current quarter.

**3. Performance insights Dashboard:**

****

**Description:**

* **Sales per Day:** Used to present daily sale figures, which assists in the determination of days with big or low sales.
* **Revenue by Category:** Subtotals revenue by the various pizza types offered to show which type of pizza brings in the most revenue**.**
* **Revenue per Hour:** Reflects the level of sales throughout a day in order to distinguish the increase and decrease of revenues**.**
* **Orders per Hour:** Enables one to understand how many orders were placed in each hour to notice periods of high traffic**.**
* **Top Pizzas:** These are the pizzas with high ratings: important products sold under this category are included.
* **Cumulative Revenue:** Records the growth of revenue within a certain period of time and on this basis demonstrates the general tendencies.
* **Slicer:**

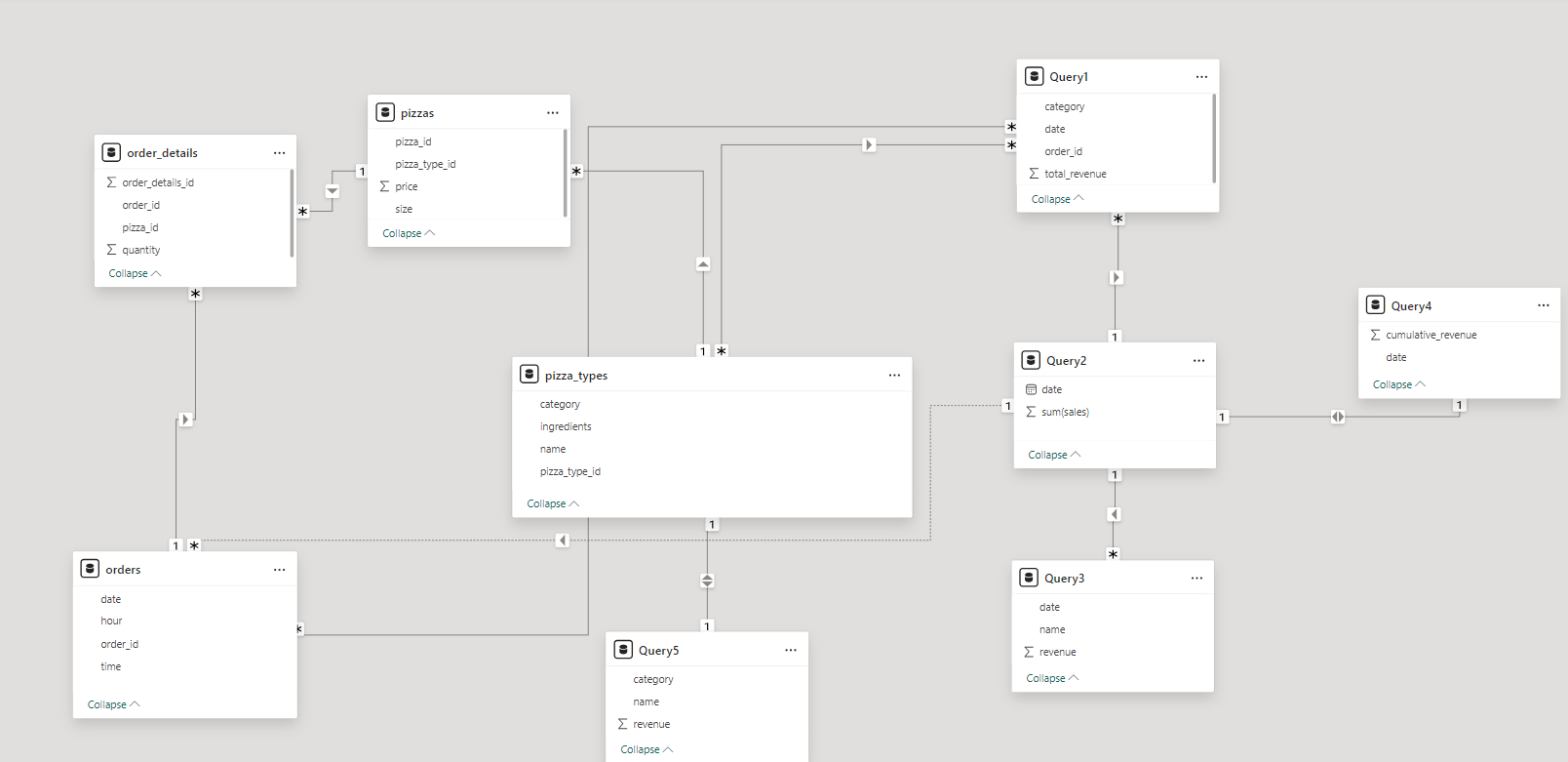
**Quarter:** Enables analysis of the data based on the fiscal quarters which make it possible to compare the performance in different periods**.**

**Insights and Interpretations:**

* **Sales per Day:** It also helps in identifying the daily sales which assist in staffing and flow of stocks well in the business.

* **Revenue by Category**: Identifies the most attractive pizza segments to help to direct the attention to the most profitable areas of menu offerings and advertising.
* **Revenue per Hour:** Unearths specific peak and other non-peak times to help in Staffing and patterning the busy business period.
* **Orders per Hour:** Displays when the business is busy so as to better organize orders and serve customers.
* **Top Pizzas:** It identifies goods that have the most demand so that it can order the goods and more importantly design promotion strategies**.**
* **Cumulative Revenue:** Records the revenue increase with time as it helps to evaluate the overall performance of the business, as well as the future tendencies**.**
* **Quarter Slicer:** Abstract and general information is helpful for comparison of the results by fiscal quarters, budgeting and other planning activities**.**

**Power BI Schema**

****

To ensure comprehensive data analysis, the Power BI schema was designed to integrate both the original CSV data and query-based tables imported from MySQL. This approach allows for seamless filtering and interactivity across all visuals.

**Schema Description:**

* **Original CSV Data**:
  + Imported directly into Power BI to retain raw data integrity.
  + Includes tables such as orders, order\_details, pizzas, and pizza\_types.
* **Query-Based Tables from MySQL**:
  + Custom queries were run on the MySQL database to generate aggregated and derived data tables.
  + These tables were imported into Power BI to enhance the analytical capabilities and support complex visualizations.
  + Examples include tables for cumulative revenue, revenue by category, and hourly sales.

**Schema Structure**:

* **Tables**:
  + **Orders**: Contains all order-related data.
  + **Order\_Details**: Detailed information for each order.
  + **Pizzas**: List of available pizzas with their attributes.
  + **Pizza\_Types**: Categorization of pizzas.
  + **Revenue\_By\_Category**: Aggregated revenue data grouped by pizza category.
  + **Hourly\_Sales**: Revenue and order counts aggregated by hour.
* **Relationships**:
  + **Orders** to **Order\_Details**: Linked by order\_id.
  + **Order\_Details** to **Pizzas**: Linked by pizza\_id.
  + **Pizzas** to **Pizza\_Types**: Linked by pizza\_type\_id.

By combining original CSV data with MySQL query-based tables and transform them by adding new columns, the Power BI schema facilitates dynamic and detailed analysis, providing valuable insights into pizza sales performance.

**Conclusion**

In conclusion, this pizza sales analysis project incorporated a holistic research design to allow for the identification of pertinent findings using SQL and Power BI. We made a dynamic schema using Kaggle data and custom MySQL queries in order to produce detailed and interactive views on data.

Through the two main dashboards:

**Executive Overview Dashboard**: Others included a brief annual executive summary of measures like total revenue, total orders, average price of pizza and popularity of the different pizzas available for sale and detailed category and quarter slicers.

**Performance insights Dashboard**: They provided advanced and detailed information on daily sales performance, daily revenue per category and hours of the day, order frequency, types of pizzas, and overall sales; with options to add a quarter of the day for further evaluation of performance.

These dashboards enable the management to make informed decisions, improve organizational efficiency, increase customer satisfaction and thus increase revenues. Indeed, the neat schema design and the integration of the Power BI schema guarantee that all filter changes are supported for a broad analysis and interaction.

**Key Takeaways:**

Panels of Power BI with SQL tools for data preparation give strong data visualization.

Workers should also integrate data successfully and design schema for efficient interactive and detailed analysis.

Knowledge about the sales patterns, customers’ needs, and performance indicators can improve business strategy and processes.

This paper acts as a reference to the process and knowledge obtained from the pizza sales analysis where the presentation of results relies on both accurate computations combined with simple graphical interface.

**Tools**:

Database and querying: MySQL

Visualization: Power BI

Data source: Kaggle and GitHub

Data inspection: Microsoft Excel

Documentation: Microsoft Word

**Acknowledgments**

We would like to thank the following resources and individuals for their contributions to this project:

* **Kaggle**: For providing the raw dataset containing detailed information on pizza sales, orders, and related attributes.
* **Development Communities**: The teams behind MySQL, Power BI, and Microsoft Excel for their powerful tools that made this analysis possible.
* **WS Cube Tech YouTube Channel**: Special thanks to Ayushi Jain for her insightful videos that guided us through the data analysis process. Her tutorials on the WS Cube Tech YouTube channel were invaluable resources for this project.

<https://youtu.be/2w-M4z-iQpk?si=qg6pL4C43ybsUJVf>

Check out My GitHub profile for whole project sources:

<https://github.com/Nityamjain>

**Thank you !**