The Great Pizza Analytics Challenge

Transforming Raw Sales Data into Business Intelligence using SQL

Presenter: Tharun Pranav T
Email: tharunpranavt2606@gmail.com
GitHub: https://github.com/TharunPranav2007
LinkedIn: www.linkedin.com/in/tharunpranavt

Acknowledgment: Special thanks to **Indian Data Club** and **DPDZero**.

Project Overview: Unleashing Pizza Potential

Our mission was to act as the **Lead Data Analyst for IDC Pizza**, transforming raw sales data into actionable business intelligence. This involved building a robust relational database and extracting critical insights to empower data-driven decision-making.

The Challenge

Convert raw CSV data into a functional database.

Input

4 Raw CSV datasets: Orders, Order Details, Pizzas, Types.

Tool

MySQL Workbench for database management.

Output

Actionable insights on sales, inventory, and pricing.

Key Skills Utilized: Data Modeling, Data Cleaning, Complex SQL Joins, and Aggregations.

Database Architecture: The Blueprint

We designed a **normalized relational schema** to ensure data integrity and efficient querying, forming the backbone of our analytics system.

Orders

Stores Date & Time of transactions.

Order Details

Links Orders to specific Pizzas and quantities.

Pizza Types

Stores Names, Categories, and Ingredients.

Pizzas

Stores Price and Size data.

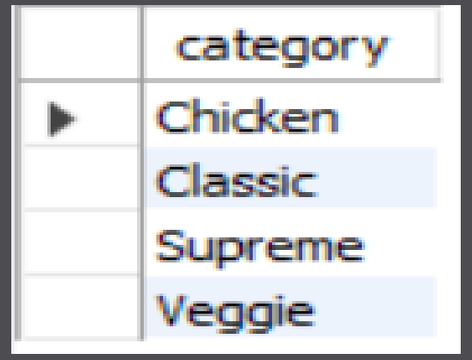
Relationships: All tables are interconnected via Primary Keys (PK) and Foreign Keys (FK) to maintain robust data integrity.

Phase 1: Foundation & Data Cleaning

Our initial objective was to ensure data quality and establish a clear understanding of the dataset structure before diving into deeper analysis.

- Q1: List all unique pizza categories.

```
SELECT DISTINCT category FROM pizza_types;
```



	category
▶	Chicken
	Classic
	Supreme
	Veggie

- Q2: Display pizza details & handle missing ingredients.

```
SELECT pizza_type_id, name, COALESCE(ingredients, 'Missing Data') AS ingredients FROM pizza_types LIMIT 5;
```

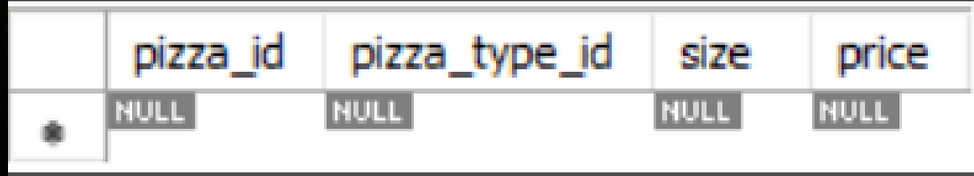
Technique: Used `COALESCE()` to replace NULL values, ensuring no critical data points are overlooked.

	pizza_type_id	name	ingredients
▶	bbq_ckn	The Barbecue Chicken Pizza	Barbecued Chicken, Red Peppers, Green Peppe...
	big_meat	The Big Meat Pizza	Bacon, Pepperoni, Italian Sausage, Chorizo Sau...
	brie_carre	The Brie Carre Pizza	Brie Carre Cheese, Prosciutto, Caramelized Oni...
	calabrese	The Calabrese Pizza	'Nduja Salami, Pancetta, Tomatoes, Red Onions...
	cali_ckn	The California Chicken Pizza	Chicken, Artichoke, Spinach, Garlic, Jalapeno P...

- Q3: Check for data errors (Missing Prices).

```
SELECT * FROM pizzas WHERE price IS NULL;
```

Insight: Identifying and rectifying missing values is crucial for accurate financial analysis.



	pizza_id	pizza_type_id	size	price
●	NULL	NULL	NULL	NULL

Phase 2: Filtering & Exploration (Part 1)

This phase focused on extracting specific subsets of data by applying various filtering criteria, allowing us to pinpoint critical operational moments and product offerings.

- Q4: Orders placed on New Year's Day (2015-01-01).

```
SELECT * FROM orders WHERE
date = '2015-01-01';
```

Insight: Understanding peak demand days helps in staffing and inventory management.

- Q5: List pizzas by price (High to Low).

```
SELECT * FROM pizzas ORDER
BY price DESC;
```

Insight: Reveals premium offerings and helps in pricing strategies.

- Q6: Filter for 'L' or 'XL' sized pizzas.

```
SELECT * FROM pizzas WHERE size IN ('L', 'XL');
```

Insight: Identifies popular sizes for inventory optimization.

order_id	date	time
1	2015-01-01	11:38:36
2	2015-01-01	11:57:40
3	2015-01-01	12:12:28
4	2015-01-01	12:16:31
5	2015-01-01	12:21:30

pizza_id	pizza_type_id	size	price
the_greek_xxl	the_greek	XXL	35.95
the_greek_xl	the_greek	XL	25.50
brie_carre_s	brie_carre	S	23.65
ital_veggie_l	ital_veggie	L	21.00
bbq_ckn_l	bbq_ckn	L	20.75
soppressata_l	soppressata	L	20.75
southw_ckn_l	southw_ckn	L	20.75
spicy_ital_l	spicy_ital	L	20.75
peppr_salami_l	peppr_salami	L	20.75
spin_pesto_l	spin_pesto	L	20.75

pizza_id	pizza_type_id	size	price
bbq_ckn_l	bbq_ckn	L	20.75
big_meat_l	big_meat	L	20.50
calabrese_l	calabrese	L	20.25
cali_ckn_l	cali_ckn	L	20.75
ckn_alfredo_l	ckn_alfredo	L	20.75
ckn_pesto_l	ckn_pesto	L	20.75
classic_dlx_l	classic_dlx	L	20.50
five_cheese_l	five_cheese	L	18.50
four_cheese_l	four_cheese	L	17.95
green_garden_l	green_garden	L	20.25

Phase 2: Advanced Filtering (Part 2)

We leveraged advanced SQL capabilities like pattern matching and complex logical operators to uncover more nuanced insights from the data.

- Q7: Find pizzas priced between $15 and $17.

```
SELECT * FROM pizzas WHERE
price BETWEEN 15.00 AND 17.00;
```

Technique: The `BETWEEN` operator provides an efficient way to filter within a range.

- Q8: Find all pizzas containing "Chicken".

```
SELECT * FROM pizza_types
WHERE name LIKE '%Chicken%';
```

Technique: Used the `LIKE` operator with wildcards for flexible pattern matching.

- Q9: Orders on Feb 15th OR after 8:00 PM.

```
SELECT * FROM orders WHERE date = '2015-02-15' OR time > '20:00:00';
```

Insight: Helps identify specific high-activity periods for targeted promotions or operational adjustments.

pizza_id	pizza_type_id	size	price
bbq_ckn_m	bbq_ckn	M	16.75
big_meat_m	big_meat	M	16.00
calabrese_m	calabrese	M	16.25
cali_ckn_m	cali_ckn	M	16.75
ckn_alfredo_m	ckn_alfredo	M	16.75

pizza_type_id	name	category	ingredients
bbq_ckn	The Barbecue Chicken Pizza	Chicken	Barbecued Chicken, Red Peppers, Green Peppe...
cali_ckn	The California Chicken Pizza	Chicken	Chicken, Artichoke, Spinach, Garlic, Jalapeno P...
ckn_alfredo	The Chicken Alfredo Pizza	Chicken	Chicken, Red Onions, Red Peppers, Mushrooms...
ckn_pesto	The Chicken Pesto Pizza	Chicken	Chicken, Tomatoes, Red Peppers, Spinach, Garl...
southw_ckn	The Southwest Chicken Pizza	Chicken	Chicken, Tomatoes, Red Peppers, Red Onions, ...
thai_ckn	The Thai Chicken Pizza	Chicken	Chicken, Pineapple, Tomatoes, Red Peppers, T...
NULL	NULL	NULL	NULL

order_id	date	time
60	2015-01-01	20:05:16
61	2015-01-01	20:08:43
62	2015-01-01	20:50:16
63	2015-01-01	20:51:42
64	2015-01-01	20:52:08
65	2015-01-01	21:16:00

Phase 3: Sales Metrics - Quantifying Success

This crucial phase involved calculating high-level financial and volume metrics to assess overall business performance and identify key trends.

- Q10: Total quantity of pizzas sold.

```
SELECT SUM(quantity) AS total_quantity FROM order_details;
```

Metric: A fundamental indicator of overall sales volume.

- Q11: Average price of a pizza.

```
SELECT AVG(price) AS average_price FROM pizzas;
```

Metric: Provides insight into the average value of a single pizza.

- Q12: Total Revenue per Order.

```
SELECT o.order_id, SUM(od.quantity * p.price) AS total_order_value FROM orders o JOIN order_details od ON o.order_id = od.order_id JOIN pizzas p ON od.pizza_id = p.pizza_id GROUP BY o.order_id;
```

Technique: Joined **three tables** (`orders`, `order_details`, `pizzas`) and aggregated to calculate revenue per transaction, demonstrating complex data integration.

total_quantity
49574

average_price
16.440625

order_id	total_order_value
1	13.25
2	92.00
3	37.25
4	16.50
5	16.50
6	24.75
7	12.50
8	12.50
9	143.25
10	41.00
11	73.50
12	70.75
13	20.25
14	12.00
15	63.25

Phase 3: Category Performance - Identifying Bestsellers

Understanding how different pizza categories perform is vital for menu optimization, marketing efforts, and resource allocation.

- Q13: Total quantity sold per category.

```
SELECT pt.category, SUM(od.quantity) AS total_quantity FROM
pizza_types pt JOIN pizzas p ON pt.pizza_type_id = p.pizza_type_id
JOIN order_details od ON p.pizza_id = od.pizza_id GROUP BY
pt.category;
```

Insight: Highlights top-performing categories, guiding future menu development and promotions.

- Q14: High-Volume Categories (> 5,000 sold).

```
SELECT pt.category, SUM(od.quantity) AS total_quantity FROM
pizza_types pt JOIN pizzas p ON pt.pizza_type_id = p.pizza_type_id
JOIN order_details od ON p.pizza_id = od.pizza_id GROUP BY
pt.category HAVING SUM(od.quantity) > 5000;
```

Technique: Used the `HAVING` clause to filter aggregated results, isolating the most popular pizza categories.

	category	total_quantity
▶	Chicken	11050
	Classic	14888
	Supreme	11987
	Veggie	11649

	category	total_quantity
▶	Chicken	11050
	Classic	14888
	Supreme	11987
	Veggie	11649

Phase 3: Inventory & Pricing Strategy - Optimizing Resources

Beyond sales, analyzing inventory and pricing structures helps identify inefficiencies and opportunities for improved profitability.

- Q15: Pizzas that have NEVER been ordered.

```
SELECT p.pizza_id FROM pizzas p LEFT JOIN order_details od ON p.pizza_id = od.pizza_id WHERE od.order_details_id IS NULL;
```

Technique: A `LEFT JOIN` combined with checking for `NULL` values effectively identifies unpurchased items, vital for inventory management.

pizza_id
big_meat_l
big_meat_m
five_cheese_m
five_cheese_s
four_cheese_s

- Q16: Price difference between sizes.

```
SELECT p1.pizza_type_id, (p1.price - p2.price) AS price_diff FROM pizzas p1 JOIN pizzas p2 ON p1.pizza_type_id = p2.pizza_type_id WHERE p1.pizza_id != p2.pizza_id;
```

Technique: Employed a **SELF JOIN** to compare different sizes of the same pizza type, aiding in strategic pricing adjustments.

pizza_type_id	size_1	price_1	size_2	price_2	price_difference
bbq_ckn	L	20.75	M	16.75	4.00
bbq_ckn	L	20.75	S	12.75	8.00
bbq_ckn	M	16.75	L	20.75	-4.00
bbq_ckn	M	16.75	S	12.75	4.00
bbq_ckn	S	12.75	L	20.75	-8.00
bbq_ckn	S	12.75	M	16.75	-4.00
big_meat	L	20.50	M	16.00	4.50
big_meat	L	20.50	S	12.00	8.50
big_meat	M	16.00	L	20.50	-4.50
big_meat	M	16.00	S	12.00	4.00
big_meat	S	12.00	L	20.50	-8.50
big_meat	S	12.00	M	16.00	-4.00
calabrese	L	20.25	M	16.25	4.00
calabrese	L	20.25	S	12.25	8.00
calabrese	M	16.25	L	20.25	-4.00

Business Insights & Conclusion: A Data-Driven Future

Our comprehensive analysis has yielded actionable insights that can significantly enhance IDC Pizza's operational efficiency and strategic planning.

Operations

Sales spike significantly after 8 PM, indicating a critical need for late-shift staffing and optimized resource allocation.

Menu Strategy

"Chicken" pizzas are a primary revenue driver, suggesting further investment in chicken-based offerings and promotions.

Inventory Optimization

Identified specific pizza sizes as dead stock (zero sales), highlighting opportunities for menu streamlining and inventory cost reduction.

Impact: The newly created database now serves as a **single source of truth** for all operational decisions at IDC Pizza, empowering truly data-driven growth.