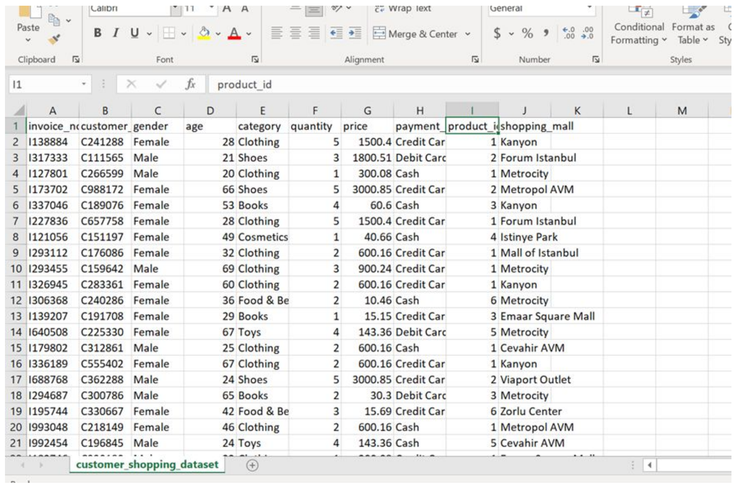
Enhancing Retail Data Management and Analysis with a MySQL Database

Maria Delia, Advised by Nada Al Sallami Worcester State University Computer Science

Background

Consider a retail company that is facing challenges in managing its customer and product data effectively. The current system lacks a comprehensive solution for recording purchase information, updating customer details, and analyzing sales data efficiently.

The aim of this project is to find and implement a database, the project seeks to update/simplify the recording of purchase information, enable updates to customer details, and facilitate effective analysis of sales data.



customer_shopping_dataset

invoice_no	customer_id	gender	age	category	quantity	price	payment_method	product_id	shopping_mall
I138884	C241288	Female	28	Clothing	5	1500.4	Credit Car	1	Kanyon
I317333	C111565	Male	21	Shoes	3	1800.51	Debit Carc	2	Forum Istanbul
I127801	C266599	Male	20	Clothing	1	300.08	Cash	1	Metrocity
I173702	C988172	Female	66	Shoes	5	3000.85	Credit Car	2	Metropol AVM
I337046	C189076	Female	53	Books	4	60.6	Cash	3	Kanyon
I227836	C657758	Female	28	Clothing	5	1500.4	Credit Car	1	Forum Istanbul
I121056	C151197	Female	49	Cosmetics	1	40.66	Cash	4	Istinye Park
I293112	C176086	Female	32	Clothing	2	600.16	Credit Car	1	Mall of Istanbul
I293455	C159642	Male	69	Clothing	3	900.24	Credit Car	1	Metrocity
I326945	C283361	Female	60	Clothing	2	600.16	Credit Car	1	Kanyon
I306368	C240286	Female	36	Food & Be	2	10.46	Cash	6	Metrocity
I139207	C191708	Female	29	Books	1	15.15	Credit Car	3	Emaar Square Mall
I640508	C225330	Female	67	Toys	4	143.36	Debit Carc	5	Metrocity
I179802	C312861	Male	25	Clothing	2	600.16	Cash	1	Cevahir AVM
I336189	C555402	Female	67	Clothing	2	600.16	Credit Car	1	Kanyon
I688768	C362288	Male	24	Shoes	5	3000.85	Credit Car	2	Viaport Outlet
I294687	C300786	Male	65	Books	2	30.3	Debit Carc	3	Metrocity
I195744	C330667	Female	42	Food & Be	3	15.69	Credit Car	6	Zorlu Center
I993048	C218149	Female	46	Clothing	2	600.16	Cash	1	Metropol AVM
I992454	C196845	Male	24	Toys	4	143.36	Cash	5	Cevahir AVM

Methodology

Database Initialization: Downloaded and initialized a new database from Kaggle, including initial data and table structures.

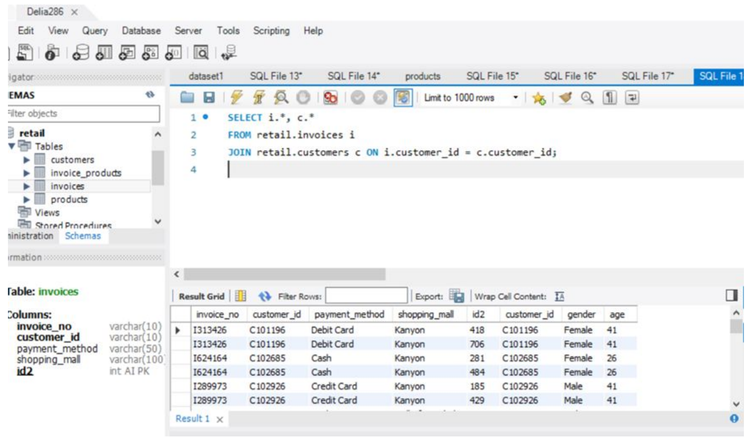
Table Creation: Organized and stored various data types by creating tables within the database.

One-to-Many Relationship: Established a one-to-many relationship between the 'customers' and 'invoices' tables, using a foreign key in 'invoices' to reference 'customers'.

Many-to-Many Relationship: Created a many-to-many relationship between 'invoices' and 'products' through a junction table named 'invoice_products', incorporating foreign keys linking both tables.

Query Execution: With the database structured, I'm set to run various SQL queries to select, aggregate, and join data, aiming to analyze and retrieve detailed insights.

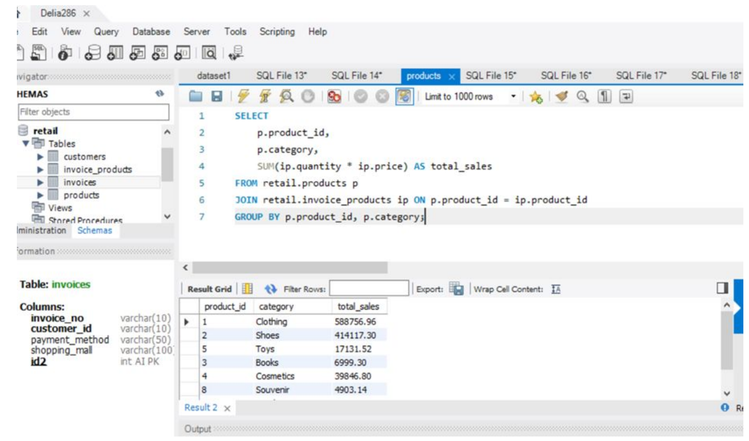
Data Analysis: Utilized queries to analyze customer behavior and product sales, providing insights into business patterns, trends, and metrics.



```
SELECT i.*, c.*
FROM retail.invoices i
JOIN retail.customers c ON i.customer_id = c.customer_id;
```

invoice_no	customer_id	payment_method	shopping_mall	id2	customer_id	gender	age
I313426	C101196	Debit Card	Kanyon	418	C101196	Female	41
I313426	C101196	Debit Card	Kanyon	706	C101196	Female	41
I624164	C102685	Cash	Kanyon	281	C102685	Female	26
I624164	C102685	Cash	Kanyon	484	C102685	Female	26
I289973	C102926	Credit Card	Kanyon	185	C102926	Male	41
I289973	C102926	Credit Card	Kanyon	429	C102926	Male	41

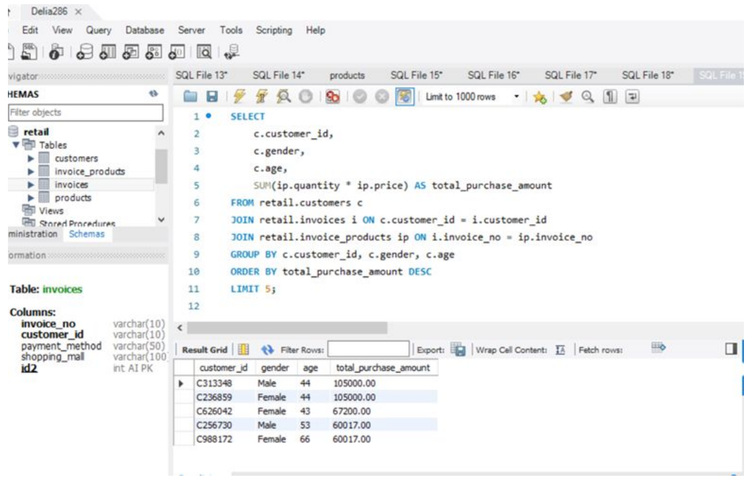
Query #1 Understanding Customer Preferences: Retrieve all invoices with customer details



```
SELECT
    p.product_id,
    p.category,
    SUM(ip.quantity * ip.price) AS total_sales
FROM retail.products p
JOIN retail.invoice_products ip ON p.product_id = ip.product_id
GROUP BY p.product_id, p.category;
```

product_id	category	total_sales
1	Clothing	588756.96
2	Shoes	414117.30
5	Toys	17131.52
3	Books	6999.30
4	Cosmetics	39846.80
8	Souvenir	4903.14

Query #2 Know Your Best-Sellers: Find the total sales for each product



```
SELECT
    c.customer_id,
    c.gender,
    c.age,
    SUM(ip.quantity * ip.price) AS total_purchase_amount
FROM retail.customers c
JOIN retail.invoices i ON c.customer_id = i.customer_id
JOIN retail.invoice_products ip ON i.invoice_no = ip.invoice_no
GROUP BY c.customer_id, c.gender, c.age
ORDER BY total_purchase_amount DESC
LIMIT 5;
```

customer_id	gender	age	total_purchase_amount
C313348	Male	44	105000.00
C236859	Female	44	105000.00
C626042	Female	43	67200.00
C256730	Male	53	60017.00
C988172	Female	66	60017.00

Query #3 Rewarding Our Best Customers: Identify customers with the highest total purchase amount

Future work

Integration with External Systems:

Explore opportunities for integrating my database with external systems or APIs. For example, I could integrate with an e-commerce platform, accounting software, or inventory management system to streamline business processes.

User Privileges:

I will try to establish different privileges for users to ensure secure access to sensitive data. For example, only authorized personnel should have the ability to update customer information or modify product details.

Challenges

Finding the right database and modifying it to fit the project's needs, such as shortening it and adjusting column formats to align with regional conventions, demanded careful consideration and flexibility. Additionally, the integration of new elements like the product_id column further highlighted the necessity of customization to achieve project objectives. Despite these hurdles, the experience served as a valuable lesson in problem-solving and the importance of meticulous attention to detail in database management projects.

Conclusion

In conclusion, this project offers a vital solution to the challenges faced by the retail industry in managing customer and product data effectively.

The establishment of efficient data relationships and the execution of targeted queries provide businesses with actionable intelligence to optimize operations, enhance customer engagement, and drive revenue growth. By understanding customer preferences, identifying top-selling products, and rewarding loyal customers, businesses can tailor strategies to meet market demands effectively.

In essence, this project underscores the transformative potential of data-driven solutions in revolutionizing retail management. By understanding the power of data, businesses can unlock new opportunities, drive innovation, and achieve sustainable success in today's dynamic marketplace.

References

https://www.kaggle.com/datasets/mehmettahiraslan/customer-shopping-dataset