Project Title: E-commerce Database Analysis

- Description: Analyze an e-commerce database to identify sales trends and customer behavior.
- Data Collection: Use open datasets from platforms like Kaggle, or simulate data by generating sample transactions and user profiles.
- Skills Demonstrated: Database design, SQL queries using PostgreSQL, subqueries and data aggregation.

Title: Analyzing an e-commerce dataset to identify sales trends and customer behavior.

Dataset: https://www.kaggle.com/datasets/carrie1/ecommerce-data

Details about the dataset: This dataset contains all the transactions which occurred between 01/12/2010 and 09/12/2011 for a UK –based and registered non-store online retail. This dataset is present in the UCI Machine Learning repository.

The below query shows how to create the table that is required for the analysis.

Query Query History

```sql
CREATE TABLE sales (
    invoice_no VARCHAR(20),
    customer_id VARCHAR(20),
    category VARCHAR(50),
    quantity INT,
    price NUMERIC(10, 2),
    invoice_date DATE,
    shopping_mall VARCHAR(50)
);
```

This syntax below helps in loading the data to the table that was just created using the SQL command CREATE TABLE sales.

```sql
COPY sales FROM 'C:\Users\Andrea\Downloads\sales_data-2.csv' WITH CSV HEADER;
SELECT * FROM sales;
```

Data Output Messages Notifications

	invoice_no character varying (20)	customer_id character varying (20)	category character varying (50)	quantity integer	price numeric (10,2)	invoice_date date	shopping_mall character varying (50)
1	I138884	C241288	Clothing	5	1500.40	2022-08-05	Kanyon
2	I317333	C111565	Shoes	3	1800.51	2021-12-12	Forum Istanbul
3	I127801	C266599	Clothing	1	300.08	2021-11-09	Metrocity
4	I173702	C988172	Shoes	5	3000.85	2021-05-16	Metropol AVM
5	I337046	C189076	Books	4	60.60	2021-10-24	Kanyon
6	I227836	C657758	Clothing	5	1500.40	2022-05-24	Forum Istanbul
7	I121056	C151197	Cosmetics	1	40.66	2022-03-13	Istinye Park
8	I293112	C176086	Clothing	2	600.16	2021-01-13	Mall of Istanbul
9	I293455	C159642	Clothing	3	900.24	2021-11-04	Metrocity
10	I326945	C283361	Clothing	2	600.16	2021-08-22	Kanyon
11	I306368	C240286	Food & Beverage	2	10.46	2022-12-25	Metrocity

Total rows: 1000 of 99457 Query complete 00:00:02.377

1.Calculate the total sales for each month?

```sql
--Calculate the total sales for each month?
SELECT DATE_TRUNC('month',invoice_date) AS monthly_sales, SUM(price) AS total_sales FROM sales GROUP BY DATE_TRUNC('month',invoice_date) ORDER BY monthly_sales;
```

Data Output Messages Notifications

	monthly_sales timestamp with time zone	total_sales numeric
1	2021-01-01 00:00:00-05	2656422.78
2	2021-02-01 00:00:00-05	2358636.34
3	2021-03-01 00:00:00-05	2618434.14
4	2021-04-01 00:00:00-04	2558825.62
5	2021-05-01 00:00:00-04	2662369.93
6	2021-06-01 00:00:00-04	2547239.73
7	2021-07-01 00:00:00-04	2802468.58
8	2021-08-01 00:00:00-04	2632303.32
9	2021-09-01 00:00:00-04	2530305.88
10	2021-10-01 00:00:00-04	2782418.40
11	2021-11-01 00:00:00-04	2547152.35
12	2021-12-01 00:00:00-05	2619727.56
13	2022-01-01 00:00:00-05	2656149.96

Total rows: 27 of 27 Query complete 00:00:00.547 Ln 19, Col 1

2.Which category of products have the highest sales ?

```
--Which category of products have the highest sales ?
SELECT category, sum(price * quantity) AS total_sales FROM sales GROUP BY category ORDER BY total_sales DESC;
```

Data Output Messages Notifications

	category character varying (50)	total_sales numeric
1	Clothing	113996791.04
2	Shoes	66553451.47
3	Technology	57862350.00
4	Cosmetics	6792862.90
5	Toys	3980426.24
6	Food & Beverage	849535.05
7	Books	834552.90
8	Souvenir	635824.65

Total rows: 8 of 8 Query complete 00:00:00.166 Ln 18, Col 1

3.Which customers have made the most purchases?

```
--Which customers have made the most purchase?--
SELECT customer_id, quantity FROM sales
ORDER BY quantity DESC, customer_id ASC LIMIT 50;
```

Data Output Messages Notifications

	customer_id character varying (20)	quantity integer
1	C100004	5
2	C100012	5
3	C100095	5
4	C100110	5
5	C100157	5
6	C100159	5
7	C100168	5
8	C100181	5
9	C100182	5
10	C100204	5

Total rows: 50 of 50 Query complete 00:00:00.165 Ln 20, Col 1

4.Average spending per customer

```sql
--average spending per customer--
SELECT AVG(total_spent) , customer_id FROM (SELECT SUM(price * quantity) AS total_spent, customer_id
FROM sales GROUP BY customer_id) AS customer_spent
GROUP BY customer_id;
```

Data Output Messages Notifications

	avg numeric	customer_id character varying (20)
1	4801.2800000000000000	C317795
2	1016.5000000000000000	C568047
3	4801.2800000000000000	C303567
4	2700.7200000000000000	C284246
5	83.6800000000000000	C103726
6	1200.3200000000000000	C204395
7	1050.0000000000000000	C734492
8	650.5600000000000000	C552872
9	9602.7200000000000000	C167970
10	20.9200000000000000	C160350
11	1050.0000000000000000	C290823
12	600.1700000000000000	C100691
13	11.7300000000000000	C148710

Total rows: 1000 of 99457 Query complete 00:00:00.521 Ln 8, Col 16

5.Which shopping malls have the highest number of transactions?

Showing the top 5 shopping malls with the highest number of transactions.

```sql
--Which shopping malls have the highest number of transactions?
SELECT DISTINCT shopping_mall, COUNT(invoice_no) AS transaction_count FROM sales
GROUP BY shopping_mall ORDER BY transaction_count DESC LIMIT 5;
```

Data Output Messages Notifications

	shopping_mall character varying (50)	transaction_count bigint
1	Mall of Istanbul	39886
2	Kanyon	39646
3	Metrocity	30022
4	Metropol AVM	20322
5	Istinye Park	19562

Total rows: 5 of 5 Query complete 00:00:00.178 Ln 6, Col 64

6.What is the average order value for each shopping mall?

```sql
--What is the average order value for each shopping mall?
SELECT shopping_mall, AVG(total_sales) AS average_order_value FROM (
SELECT invoice_no, shopping_mall, SUM(price* quantity) AS total_sales FROM sales
GROUP BY invoice_no, shopping_mall) AS s1
GROUP BY shopping_mall;
```

Data Output Messages Notifications

	shopping_mall character varying (50)	average_order_value numeric
1	Emaar Square Mall	5157.3894367075452089
2	Kanyon	5100.5630933763809716
3	Viaport Outlet	5096.1903622303622304
4	Forum Istanbul	4974.2960339599757429
5	Metrocity	4970.0602664712544134
6	Istinye Park	5034.0103629485737655
7	Mall of Istanbul	5101.7882645539788397
8	Metropol AVM	4995.5542151363054817
9	Zorlu Center	5084.1591408866995074
10	Cevahir AVM	5067.1761971548787818

Total rows: 10 of 10 Query complete 00:00:00.788

7. How does the quantity of products sold vary across different shopping malls?

```sql
--How does the quantity of products sold vary across different shopping malls?
SELECT * FROM sales;

SELECT shopping_mall, AVG(quantity), MAX(quantity), MIN(quantity), SUM(quantity) as total_quantity_per_mall FROM sales
GROUP BY shopping_mall;
```

Data Output Messages Notifications

	shopping_mall character varying (50)	avg numeric	max integer	min integer	total_quantity_per_mall bigint
1	Cevahir AVM	2.9951913444199559	5	1	29898
2	Emaar Square Mall	3.0141342756183746	5	1	29002
3	Forum Istanbul	3.0022235698403073	5	1	29704
4	Istinye Park	3.0124731622533483	5	1	58930
5	Kanyon	2.9993946425868940	5	1	118914
6	Mall of Istanbul	3.0142907285764429	5	1	120228
7	Metrocity	2.9907401239091333	5	1	89788
8	Metropol AVM	3.0046255289833678	5	1	61060
9	Viaport Outlet	2.9947089947089947	5	1	29432
10	Zorlu Center	3.0017733990147783	5	1	30468

Total rows: 10 of 10 Query complete 00:00:00.323 Ln 19, Col 1

8. Which customer has made the most purchases based on the number of invoices?

```sql
--Which customer has made the most purchases based on the number of invoices?

SELECT * FROM sales;

SELECT COUNT(invoice_no) as total_count,customer_id FROM sales GROUP BY customer_id ORDER BY total_count DESC LIMIT 5;
```

Data Output Messages Notifications

	total_count bigint	customer_id character varying (20)
1	2	C317795
2	2	C303567
3	2	C330428
4	2	C446935
5	2	C279585

Total rows: 5 of 5 | Query complete 00:00:00.537 | Ln 23, Col 21

9. What is the average price of items sold in each product category?

```sql
---What is the average price of items sold in each product category?
SELECT * FROM sales;
SELECT AVG(price) as average_price, category FROM sales
GROUP BY category
ORDER BY average_price DESC ;
```

Data Output Messages Notifications

	average_price numeric	category character varying (50)
1	3156.9355484387510008	Technology
2	1807.3885678692445685	Shoes
3	901.0840212253892771	Clothing
4	122.4486255547459760	Cosmetics
5	107.7331852879944483	Toys
6	45.5686207588837583	Books
7	34.8943448689737948	Souvenir
8	15.6719484298863021	Food & Beverage

Total rows: 8 of 8 | Query complete 00:00:00.185

10. How does revenue change over different months?

Query Query History

```sql
SELECT EXTRACT(YEAR FROM invoice_date) AS year,
EXTRACT(month FROM invoice_date) AS month,
SUM(price*quantity) AS total_sales
FROM sales
GROUP BY year, month
ORDER BY year, month;
```

Data Output Messages Notifications

	year numeric	month numeric	total_sales numeric
13	2022	1	19528622.28
14	2022	2	16688223.84
15	2022	3	19973370.32
16	2022	4	18652288.88
17	2022	5	19895148.26
18	2022	6	19295007.90
19	2022	7	20135205.90
20	2022	8	19303411.18
21	2022	9	19215258.58
22	2022	10	20564150.74
23	2022	11	17883169.32
24	2022	12	19739770.96
25	2023	1	18971199.66
26	2023	2	19017325.92
27	2023	3	5028293.58

Total rows: 27 of 27 Query complete 00:00:00.174

11. Which shopping mall sells the most expensive items on average?

```sql
--Which shopping mall sells the most expensive items on average?
SELECT * FROM sales;

SELECT AVG(price) as average_price, shopping_mall FROM sales
GROUP BY shopping_mall
ORDER BY average_price DESC LIMIT 1;
```

Data Output Messages Notifications

	average_price numeric	shopping_mall character varying (50)
1	704.7200810642278113	Emaar Square Mall

12. What is the total quantity sold per month?

```sql
---What is the total quantity sold per month?
SELECT EXTRACT(year FROM invoice_date) AS year,
EXTRACT(month FROM invoice_date) AS month,
SUM(quantity) as total_quantity FROM sales
GROUP BY month, year
ORDER BY month, year ASC;
```

Data Output Messages Notifications

	year numeric	month numeric	total_quantity bigint
14	2021	6	22704
15	2022	6	22644
16	2021	7	23972
17	2022	7	23426
18	2021	8	22410
19	2022	8	23432
20	2021	9	21802
21	2022	9	22394
22	2021	10	23422
23	2022	10	23532
24	2021	11	22710
25	2022	11	22462
26	2021	12	23302
27	2022	12	23002

Total rows: 27 of 27 Query complete 00:00:00.152

13. Which month had the highest sales in terms of revenue?

Query Query History

```sql
--Which month had the highest sales in terms of revenue?
SELECT EXTRACT(month FROM invoice_date) as month,
EXTRACT (YEAR FROM invoice_date) as year,
SUM(price*quantity) AS total_revenue
FROM sales
GROUP BY month, year
ORDER BY total_revenue DESC
LIMIT 1;
```

Data Output Messages Notifications

	month numeric	year numeric	total_revenue numeric
1	7	2021	20622239.36

14. How does the quantity of products sold vary across different categories over time?

```sql
--How does the quantity of products sold vary across different categories over time?
SELECT EXTRACT(YEAR from invoice_date) AS year,
EXTRACT(month from invoice_date) AS month,
AVG(quantity) as average_quantity,
SUM(quantity) as total_quantity, category
FROM sales
GROUP BY year, month, category
LIMIT 10;
```

Data Output Messages Notifications

	year numeric	month numeric	average_quantity numeric	total_quantity bigint	category character varying (50)
1	2021	1	3.0796019900497512	1238	Books
2	2021	1	2.9699021820917983	7894	Clothing
3	2021	1	3.0288288288288288	3362	Cosmetics
4	2021	1	2.9892473118279570	3336	Food & Beverage
5	2021	1	2.9951807228915663	2486	Shoes
6	2021	1	3.0000000000000000	1224	Souvenir
7	2021	1	3.0106951871657754	1126	Technology
8	2021	1	2.9792746113989637	2300	Toys
9	2021	2	2.8727272727272727	948	Books
10	2021	2	3.0033698399326032	7130	Clothing

Total rows: 10 of 10 Query complete 00:00:00.255

15. What is the average quantity of items purchased per transaction?

```sql
--What is the average quantity of items purchased per transaction?
SELECT AVG(total_quantity) as average_quantity
FROM (SELECT invoice_no, sum(quantity) as total_quantity FROM sales GROUP BY invoice_no) AS s1
```

Data Output Messages Notifications

	average_total_quantity numeric
1	6.0068572347848819

16. Which category has the highest average transaction value?

```sql
--Which category has the highest average transaction value?

SELECT category, AVG(total_transaction_value) AS average_transaction_value
FROM(SELECT category, invoice_no, SUM(price * quantity) AS total_transaction_value FROM sales GROUP BY category, invoice_no) AS s1
GROUP BY category
ORDER BY average_transaction_value DESC
LIMIT 1;
```

Data Output Messages Notifications

	category character varying (50)	average_transaction_value numeric
1	Technology	23163.470776621297

17.What is the top-selling products by total revenue generated?

```sql
---What is the top-selling products by total revenue generated?
SELECT category, SUM(price*quantity) AS total_sales
FROM sales
GROUP BY category
ORDER BY total_sales DESC
LIMIT 5;
```

Data Output Messages Notifications

	category character varying (50)	total_sales numeric
1	Clothing	227993582.08
2	Shoes	133106902.94
3	Technology	115724700.00
4	Cosmetics	13585725.80
5	Toys	7960852.48

18. Which customer has spent the most money overall?

```sql
-- Which customer has spent the most money overall?
SELECT customer_id, SUM(price*quantity) as total_sales
FROM sales
GROUP BY customer_id
ORDER BY total_sales DESC
LIMIT 1
```

Data Output Messages Notifications

	customer_id character varying (20)	total_sales numeric
1	C248633	52500.00

19. What is the distribution of sales across different shopping malls?

```sql
19--What is the distribution of sales across different shopping malls?
SELECT shopping_mall,AVG(total_sales) as average_sales
FROM (SELECT shopping_mall, SUM(price*quantity) AS total_sales FROM sales GROUP BY shopping_mall) AS s1
GROUP BY shopping_mall
ORDER BY average_sales DESC;
```

Data Output Messages Notifications

	shopping_mall character varying (50)	average_sales numeric
1	Mall of Istanbul	101744963.360000000000
2	Kanyon	101108462.200000000000
3	Metrocity	74605574.660000000000
4	Metropol AVM	50759826.380000000000
5	Istinye Park	49237655.360000000000
6	Zorlu Center	25802107.640000000000
7	Cevahir AVM	25290276.400000000000
8	Viaport Outlet	25042679.440000000000
9	Emaar Square Mall	24812200.580000000000
10	Forum Istanbul	24607842.480000000000

Total rows: 10 of 10 Query complete 00:00:00.190

20. How many unique customers have made purchases at each shopping mall?

```sql
20 -- How many unique customers have made purchases at each shopping mall?
SELECT Count(customer_id) as count_of_customers, shopping_mall FROM sales
GROUP BY shopping_mall
ORDER BY count_of_customers DESC;
```

Data Output Messages Notifications

	count_of_customers bigint	shopping_mall character varying (50)
1	39886	Mall of Istanbul
2	39646	Kanyon
3	30022	Metrocity
4	20322	Metropol AVM
5	19562	Istinye Park
6	10150	Zorlu Center
7	9982	Cevahir AVM
8	9894	Forum Istanbul
9	9828	Viaport Outlet
10	9622	Emaar Square Mall