Poject on:

A Data-Driven Analysis of Bookstore Sales

CompetitorLeveraging SQL to Uncover Key Insights.

Name:

Parmar Arpitaba M.

Date:

10-08-2025

INTRODUCTION & PROJECT GOAL

The Story of Our Data:

- Every book sold, every customer, and every transaction generates a piece of data. This project is about bringing that data to life.
- Our mission is to transform raw sales records into a clear narrative of our business performance.

Our Mission:

- Every book sold, every customer, and every transaction generates a piece of data. This project is about bringing that data to life.
- Our mission is to transform raw sales records into a clear narrative of our business performance.

.

Data & tool

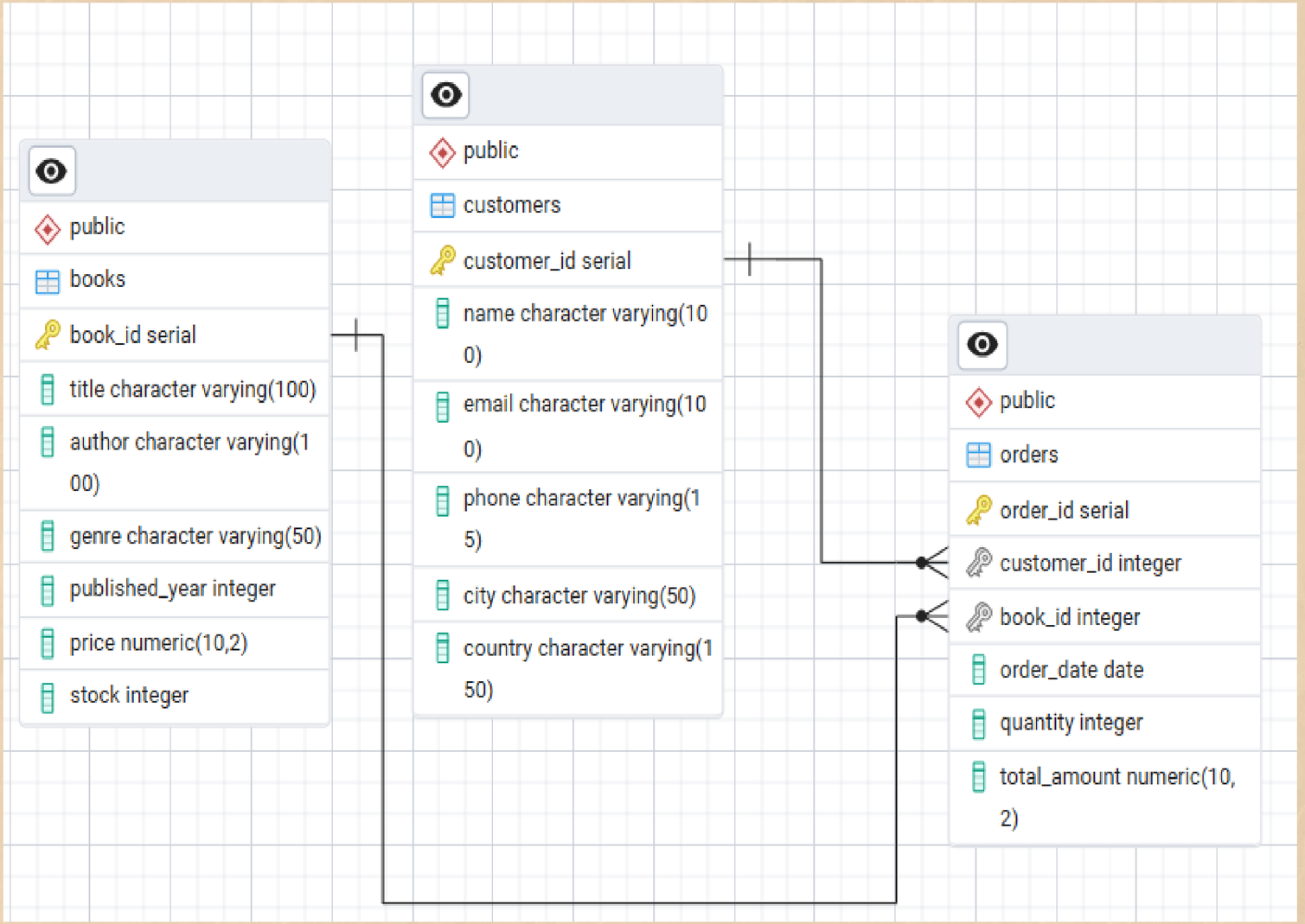
➢ Key tables and data point:

- **books:** book_id,title,author,genre,published_year,price,stock
- **customers:** customer_id,name,email,phone,city,country
- **orders:** order_id,customer_id, book_id,order_date,quantity,total_amount

➢ Tools used:

- **Database:** PostgreSQL,MySQL,or SQLite
- **Query language:** SQL

E-r diagram


public
books
book_id serial
title character varying(100)
author character varying(100)
genre character varying(50)
published_year integer
price numeric(10,2)
stock integer
public
customers
customer_id serial
name character varying(100)
email character varying(100)
phone character varying(15)
city character varying(50)
country character varying(150)
public
orders
order_id serial
customer_id integer
book_id integer
order_date date
quantity integer
total_amount numeric(10,2)

Key questions & analysis

- **Basic:**

1. Retrieve all books in the "Fiction" genre:
2. Find books published after the year 1950:
3. List all customers from the Canada:
4. Show orders placed in November 2023:
5. Retrieve the total stock of books available:
6. Find the details of the most expensive book:
7. Show all customers who ordered more than 1 quantity of a book:
8. Retrieve all orders where the total amount exceeds $20:
9. List all genres available in the Books table:
10. Find the book with the lowest stock:
11. Calculate the total revenue generated from all orders:

Key questions & analysis

- **Advanvced:**

1. Retrieve the total number of books sold for each genre:
2. Find the average price of books in the "Fantasy" genre:
3. List customers who have placed at least 2 orders:
4. Find the most frequently ordered book:
5. Show the top 3 most expensive books of 'Fantasy' Genre:
6. Retrieve the total quantity of books sold by each author:
7. List the cities where customers who spent over $30 are located:
8. Find the customer who spent the most on orders:
9. Calculate the stock remaining after fulfilling all orders:

Key questions & analysis

- **Retrieve all books in the "Fiction" genre:**

```
-- 1) Retrieve all books in the "Fiction" genre:
select *from books
where genre='Fiction'
```

- **Output:**

	book_id [PK] integer	title character varying (100)	author character varying (100)	genre character varying (50)	published_year integer	price numeric (10,2)	stock integer
1	4	Customizable 24hour product	Christopher Andrews	Fiction	2020	43.52	8
2	22	Multi-layered optimizing migration	Wesley Escobar	Fiction	1908	39.23	78
3	28	Expanded analyzing portal	Lisa Coffey	Fiction	1941	37.51	79
4	29	Quality-focused multi-tasking challenge	Katrina Underwood	Fiction	1905	31.12	100
5	31	Implemented encompassing conglomerati...	Melissa Taylor	Fiction	2010	21.23	44

Key questions & analysis

- **Find books published after the year 1950:**

```
-- 2) Find books published after the year 1950:
select *from books
where Published_Year>1950;
```

- **Output:**

	book_id [PK] integer	title character varying (100)	author character varying (100)	genre character varying (50)	published_year integer	price numeric (10,2)	stock integer
1	2	Persevering reciprocal knowledge user	Mario Moore	Fantasy	1971	35.80	19
2	4	Customizable 24hour product	Christopher Andrews	Fiction	2020	43.52	8
3	5	Adaptive 5thgeneration encoding	Juan Miller	Fantasy	1956	10.95	16
4	6	Advanced encompassing implementation	Bryan Morgan	Biography	1985	6.56	2
5	8	Persistent local encoding	Troy Cox	Science Fiction	2019	48.99	84

Key questions & analysis

- **List all customers from the Canada:**

```
-- 3) List all customers from the Canada:
select *from customers
where country='Canada'
```

- **Output:**

	customer_id [PK] integer	name character varying (100)	email character varying (100)	phone character varying (15)	city character varying (50)	country character varying (150)
1	38	Nicholas Harris	christine93@perkins.com	1234567928	Davistown	Canada
2	415	James Ramirez	robert54@hall.com	1234568305	Maxwelltown	Canada
3	468	David Hart	stokesrebecca@gmail.com	1234568358	Thompsonfurt	Canada

Key questions & analysis

- **Show orders placed in November 2023:**

```
-- 4) Show orders placed in November 2023:
select *from orders
where order_date between '2023-11-01' and '2023-11-30'
```

- **Output:**

	order_id [PK] integer	customer_id integer	book_id integer	order_date date	quantity integer	total_amount numeric (10,2)
1	4	433	343	2023-11-25	7	301.21
2	19	496	60	2023-11-17	9	316.26
3	75	291	375	2023-11-30	5	170.75
4	132	469	333	2023-11-22	7	194.32
5	137	474	471	2023-11-25	8	363.04

Key questions & analysis

- **Retrieve the total stock of books available:**

```sql
-- 5) Retrieve the total stock of books available:
select sum(stock) as total_stock
from books
```

- **Output:**

	total_stock bigint
1	25056

Key questions & analysis

- **Find the details of the most expensive book:**

```
-- 6) Find the details of the most expensive book:

select *from books
order by price desc
limit 1
```

- **Output:**

	book_id [PK] integer	title character varying (100)	author character varying (100)	genre character varying (50)	published_year integer	price numeric (10,2)	stock integer
1	340	Proactive system-worthy orchestration	Robert Scott	Mystery	1907	49.98	88

Key questions & analysis

- **Show all customers who ordered more than 1 quantity of a book:**

```
-- 7) Show all customers who ordered more than 1 quantity of a book:
select *from orders
where quantity>1
```

- **Output:**

	order_id [PK] integer	customer_id integer	book_id integer	order_date date	quantity integer	total_amount numeric (10,2)
1	1	84	169	2023-05-26	8	188.56
2	2	137	301	2023-01-23	10	216.60
3	3	216	261	2024-05-27	6	85.50
4	4	433	343	2023-11-25	7	301.21
5	5	14	431	2023-07-26	7	136.36

Key questions & analysis

- **Retrieve all orders where the total amount exceeds $20:**

```
-- 8) Retrieve all orders where the total amount exceeds $20:
select *from orders
where total_amount>20
```

- **Output:**

	order_id [PK] integer	customer_id integer	book_id integer	order_date date	quantity integer	total_amount numeric (10,2)
1	1	84	169	2023-05-26	8	188.56
2	2	137	301	2023-01-23	10	216.60
3	3	216	261	2024-05-27	6	85.50
4	4	433	343	2023-11-25	7	301.21
5	5	14	431	2023-07-26	7	136.36

Key questions & analysis

- **List all genres available in the Books table:**

```sql
-- 9) List all genres available in the Books table:
select distinct genre
from books
```

- **Output:**

	genre character varying (50)
1	Romance
2	Biography
3	Mystery
4	Fantasy
5	Fiction
6	Non-Fiction
7	Science Fiction

Key questions & analysis

- **Find the book with the lowest stock:**

```
-- 10) Find the book with the lowest stock:
select *from books
order by stock
limit 1
```

- **Output:**

	book_id [PK] integer	title character varying (100)	author character varying (100)	genre character varying (50)	published_year integer	price numeric (10,2)	stock integer
1	44	Networked systemic implementation	Ryan Frank	Science Fiction	1965	13.55	0

Key questions & analysis

- **Calculate the total revenue generated from all orders:**

```sql
-- 11) Calculate the total revenue generated from all orders
select sum(total_amount) as total_revenue
from orders
```

- **Output:**

	total_revenue numeric
1	75628.66

Key questions & analysis

- **Retrieve the total number of books sold for each genre:**

```sql
-- 1) Retrieve the total number of books sold for each genre:
select b.genre,sum(o.quantity)
from books b
join orders o
on b.book_id=o.book_id
group by genre
```

- **Output:**

	genre character varying (50)	sum bigint
1	Romance	439
2	Biography	285
3	Mystery	504
4	Fantasy	446
5	Fiction	225
6	Non-Fiction	351
7	Science Fiction	447

Key questions & analysis

- **Find the average price of books in the "Fantasy" genre:**

```
-- 2) Find the average price of books in the "Fantasy" genre:
select avg(price)
from books
where genre='Fantasy'
```

- **Output:**

	avg numeric
1	25.981690140845070400

Key questions & analysis

- **List customers who have placed at least 2 orders:**

```sql
-- 3) List customers who have placed at least 2 orders:
select c.name,c.customer_id,count(o.order_id)
from orders o
join customers c
on o.customer_id=c.customer_id
group by c.name,c.customer_id
having count(o.order_id)>=2
```

- **Output:**

	name character varying (100)	customer_id [PK] integer	count bigint
1	Richard Mclaughlin	184	2
2	Carl Smith	272	3
3	Stacey Adams	22	3
4	Victoria Dixon	173	2
5	Jason Bell	189	2

Key questions & analysis

- **Find the most frequently ordered book:**

```
-- 4) Find the most frequently ordered book:
select b.book_id,b.title,count(o.book_id) as booked
from orders o
join books b
on o.book_id=b.book_id
group by b.book_id,b.title
order by booked desc
limit 1
```

- **Output:**

	book_id [PK] integer	title character varying (100)	booked bigint
1	273	Devolved zero administration process improvement	4

Key questions & analysis

- **Show the top 3 most expensive books of 'Fantasy' Genre :**

```
-- 5) Show the top 3 most expensive books of 'Fantasy' Genre :
select *from books
where genre='Fantasy'
order by price desc
limit 3
```

- **Output:**

	book_id [PK] integer	title character varying (100)	author character varying (100)	genre character varying (50)	published_year integer	price numeric (10,2)	stock integer
1	240	Stand-alone content-based hub	Lisa Ellis	Fantasy	1957	49.90	41
2	462	Innovative 3rdgeneration database	Allison Contreras	Fantasy	1988	49.23	62
3	238	Optimized even-keeled analyzer	Sherri Griffith	Fantasy	1975	48.97	72

Key questions & analysis

- **Retrieve the total quantity of books sold by each author:**

```sql
-- 6) Retrieve the total quantity of books sold by each author:
select b.author,sum(o.quantity) as total_book_sold
from orders o
join books b
on o.book_id=b.book_id
group by b.author
```

- **Output:**

	author character varying (100)	total_book_sold bigint
1	Jared Cortez	10
2	Tracy Parker	11
3	Taylor Wang	9
4	Cathy Knight	6
5	Bianca Matthews	3

Key questions & analysis

- **List the cities where customers who spent over $30 are located:**

```sql
-- 7) List the cities where customers who spent over $30 are located:
select distinct c.city,o.total_amount
from customers c
join orders o
on c.customer_id=o.customer_id
where o.total_amount>30
```

- **Output:**

	city character varying (50)	total_amount numeric (10,2)
1	Taylorfurt	189.45
2	Leeport	141.39
3	Port Jasonview	149.12
4	Port Aaronstad	145.44
5	Matthewfurt	328.50

Key questions & analysis

- **Find the customer who spent the most on orders:**

```
-- 8) Find the customer who spent the most on orders:
select c.name,sum(o.total_amount) as amount
from customers c
join orders o
on c.customer_id=o.customer_id
group by c.name
order by amount desc
limit 1
```

- **Output:**

	name character varying (100)	amount numeric
1	Kim Turner	1398.90

Key questions & analysis

- **Calculate the stock remaining after fulfilling all orders:**

```
--9) Calculate the stock remaining after fulfilling all orders:

SELECT b.book_id, b.title, b.stock, COALESCE(SUM(o.quantity),0) AS Order_quantity,
    b.stock- COALESCE(SUM(o.quantity),0) AS Remaining_Quantity
FROM books b
LEFT JOIN orders o ON b.book_id=o.book_id
GROUP BY b.book_id ORDER BY b.book_id;
```

- **Output:**

	book_id [PK] integer	title character varying (100)	stock integer	order_quantity bigint	remaining_quantity bigint
1	1	Configurable modular throughput	100	3	97
2	2	Persevering reciprocal knowledge user	19	0	19
3	3	Streamlined coherent initiative	27	5	22
4	4	Customizable 24hour product	8	0	8
5	5	Adaptive 5thgeneration encoding	16	8	8


Thank You