Amazon Book Sales Analysis

Project Overview

This project involves analyzing Amazon book sales data, with a clean dataset of 352 books after removing duplicates. The aim is to perform various tasks to better understand the sales, user ratings, and reviews of both fiction and non-fiction books on Amazon.

Dataset Description

The dataset consists of the following columns:

-Name: The title of the book (Data type: TEXT)
-Author: The author of the book (Data type: TEXT)
-User Rating: The average user rating (Data type: DECIMAL)
-Reviews: The total number of reviews (Data type: INTEGER)
-Price: The price of the book in dollars (Data type: CURRENCY)
-Year: The year of publication or sales (Data type: INTEGER)
-Genre: The genre of the book, categorized as either Fiction or Non Fiction (Data type: TEXT)

Project Tasks

1. Database Creation in SQL

Create a database schema in PostgreSQL for the dataset.

```
--Created a db in SQL about the Amazon dataset
-- Creted the table structure to import the Amazon_Book_Sales Data.
CREATE TABLE amazon_book_sales (
    book_name TEXT,
    author TEXT,
    user_rating DECIMAL,
    reviews INTEGER,
    price MONEY,
    year INTEGER,
    genre TEXT
```

Define appropriate data types for each column:

-Name: TEXT
-Author: TEXT
-User Rating: DECIMAL
-Reviews: INTEGER
-Price: CURRENCY
-Year: INTEGER
-Genre: TEXT

2. Categorize Books by Genre

After cleaning the data, there are 351 books in total.

Books are already categorized as Fiction & Non- Fiction with column name Genre.

Query to Count how many books are Fiction & Non-Fiction in Amazon_Book_sales datasets.

```
-- Categorize the 550 books to fiction and non-fiction.
-- After cleaning the datasets there is only 351 records of data is reamining.
-- below query to count the total numbers of " Fiction & Non-Fiction " Books.

SELECT genre, COUNT(*) AS book_count
FROM Amazon_Book_sales
WHERE genre IN ('Fiction', 'Non Fiction')
GROUP BY genre;
```

Result –

	genre text	book_count bigint
1	Fiction	160
2	Non Fiction	191

From The result is it clear that there are 160 fiction, and 191 non-fiction books are sales from Amazon between 2009 and 2019.

3. Query the Top 50 Best Sellers

Create an SQL query to retrieve the top 50 best-selling books based on the number of reviews.

Query

```
Select book_name, author, year, reviews
From amazon_book_sales
order By reviews DESC
Limit 50 ›
```

Result analysis-

Highest: "Where the Crawdads Sing" with 87,841 reviews

Lowest: "The Very Hungry Caterpillar" by Eric Carle with 19,546 reviews

2013 stands out as the year with the most bestsellers in this list with 9 books.

4. Query Highly Rated Books Released in 2019

Find all books with a rating greater than 4 that were released in 2019.

Query

```
--Query the books which have a rating greater than 4 released last 2019

SELECT book_name
FROM amazon_book_sales
WHERE year = '2019' AND user_rating > 4;
```

Result Analysis-

It is clear from the analysis that only 27 books fall under the rating of grater than 4 released in year 2019.

5. Query Books with Over 10K Reviews Released Before 2018

Find all books that have received more than 10,000 reviews and were released before 2018.

```
--Query the books which have greater than 10k reviews last 2018
SELECT book_name, reviews
FROM amazon_book_sales
WHERE reviews > 10000
AND year BETWEEN '2009' AND '2018';
```

Result Analysis

Total 101 Books have more than 10000 reviews published between 2009 to 2018.

3. Additional Queries

Which author has the most highly rated books?

```
--Which author has the most highly rated books?
SELECT author, user_rating
FROM amazon_book_sales
WHERE user_rating = (SELECT MAX(user_rating) FROM amazon_book_sales);
```

Result Analysis: 21 Authors have highest ratings in Books. In which Dav Pilkey have 6 books with highest rating , followed by J.K. Rowling with four books and Rush Limbaugh with two books

What are the average prices for books by genre?

```
--  What are the average prices for books by genre?
SELECT genre, AVG(price::numeric) AS average_price
FROM amazon_book_sales
GROUP BY genre;
```

Resullt :

	genre text	average_price numeric
1	Fiction	12.1562500000000000
2	Non Fiction	13.8481675392670157

Find the top 5 most expensive books

```
--Find the top 5 most expensive book

SELECT price, book_name, author
FROM amazon_book_sales
ORDER BY price DESC
LIMIT 5;
```

Result Analysis:

	price money	book_name text	author text
1	$105.00	Diagnostic and Statistical Manual of Mental Disorders, 5th Edition: DS...	American Psychiatric Association
2	$82.00	The Twilight Saga Collection	Stephenie Meyer
3	$54.00	Hamilton: The Revolution	Lin-Manuel Miranda
4	$53.00	The Book of Basketball: The NBA According to The Sports Guy	Bill Simmons
5	$52.00	Harry Potter Paperback Box Set (Books 1-7)	J. K. Rowling

The most expensive book is the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition: DSM-5 which is priced significantly higher than the others at $105.00. This is a specialized medical manual, often priced higher due to its niche market.

Popular fiction books like the twilight saga collection and Harry Potter.

Which year had the most books released?

```
--- Which year had the most books released?
SELECT year, COUNT(book_name) AS book_count
FROM amazon_book_sales
GROUP BY year
ORDER BY book_count DESC
LIMIT 1;
```

Result :

	year integer	book_count bigint
1	2009	50