**SQL Project: Udemy Course Data Analysis**

**Project Overview** This project focuses on analyzing Udemy course data using SQL. The dataset contains details about various courses, including course titles, pricing, subscription counts, and other metadata. The objective is to gain insights into course popularity, pricing trends, and user engagement based on the given attributes.

**Objectives**

* Understand the structure of the Udemy course dataset.
* Perform data cleaning and transformation.
* Extract insights using SQL queries.
* Identify trends in course pricing, subscriptions, and user engagement.

**Dataset Description** The dataset contains Udemy course details with the following attributes:

**Column Descriptions:**

* **course\_id**: Unique identifier for the course.
* **course\_title**: Title of the course.
* **url**: Link to the Udemy course page.
* **is\_paid**: Indicates if the course is paid (TRUE/FALSE).
* **price**: Cost of the course.
* **num\_subscribers**: Number of students enrolled in the course.
* **num\_reviews**: Number of reviews received for the course.
* **num\_lectures**: Total number of lectures in the course.
* **level**: Course difficulty level (Beginner, Intermediate, All Levels).
* **content\_duration**: Total duration of the course content (in hours).
* **published\_timestamp**: Date and time when the course was published.
* **subject**: Subject category of the course.

**Key Deliverables**

**1. Database & Table Creation**

* Create a database named udemy\_courses.

CREATE DATABASE udemy\_courses;

USE udemy\_courses;

* Create a table named courses with appropriate data types for each column.

CREATE TABLE courses(

course\_id INT PRIMARY KEY,

course\_title VARCHAR(255) NOT NULL,

url VARCHAR(500),

is\_paid BOOLEAN NOT NULL,

price DECIMAL(10,2),

num\_subscribers INT DEFAULT 0,

num\_reviews INT DEFAULT 0,

num\_lectures INT DEFAULT 0,

level VARCHAR(50),

content\_duration DECIMAL(10,2),

published\_timestamp DATETIME,

subject VARCHAR(100),

INDEX idx\_subject (subject),

INDEX idx\_level (level) );

**2. Data Cleaning & Preparation:**

-- 2. Data Cleaning & Preparation -- Update null or empty prices for free course\

UPDATE courses

SET price = 0

WHERE is\_paid = FALSE AND (price IS NULL OR price = '');

*-- Standardize level value*

UPDATE courses

SET level =

CASE

WHEN LOWER(level) LIKE '%begin%' THEN 'Beginner'

WHEN LOWER(level) LIKE '%interm%' THEN 'Intermediate'

WHEN LOWER(level) LIKE '%all%' THEN 'All Levels'

ELSE 'All Levels'

END

WHERE level IS NOT NULL;

*-- Standardize subject values*

UPDATE courses

SET subject = TRIM(INITCAP(subject))

WHERE subject IS NOT NULL;

*-- Convert empty strings to NULL for numerical fields*

UPDATE courses

SET num\_subscribers = NULLIF(num\_subscribers, ''),

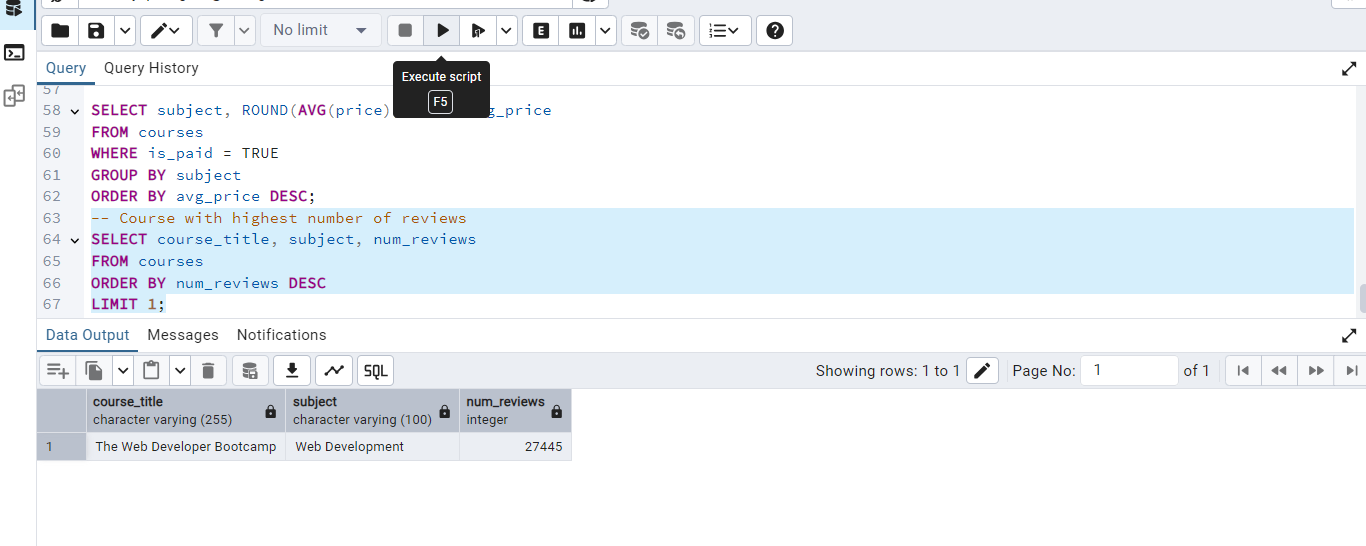
num\_reviews = NULLIF(num\_reviews, ''),

num\_lectures = NULLIF(num\_lectures, ''),

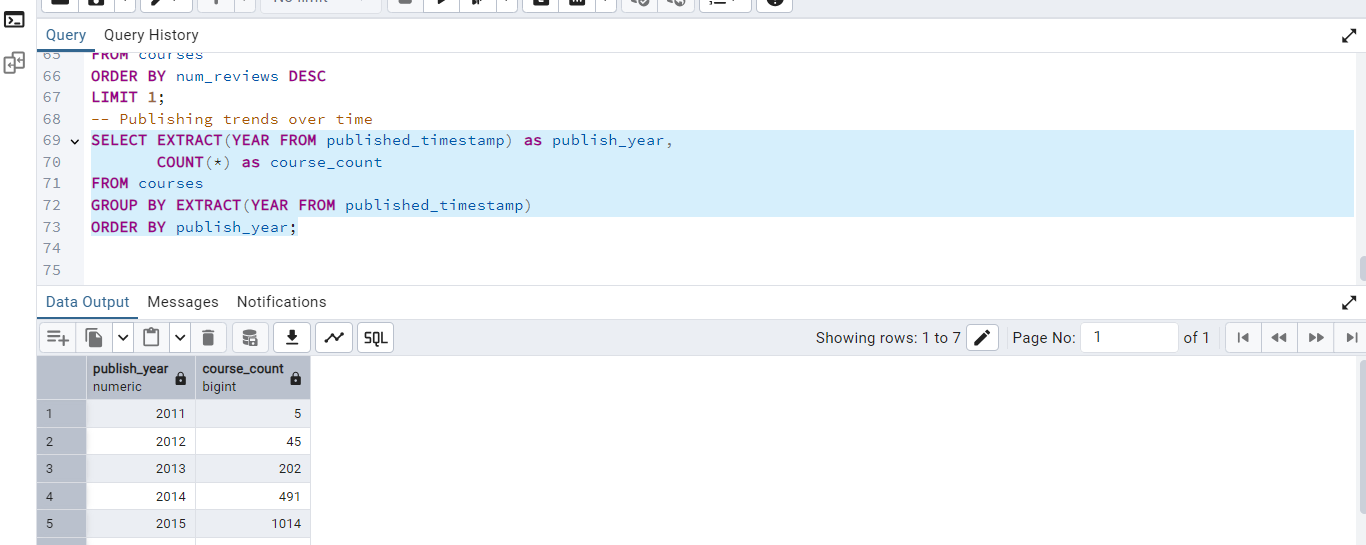
content\_duration = NULLIF(content\_duration, '');

**3. Exploratory Data Analysis (EDA) Using SQL**

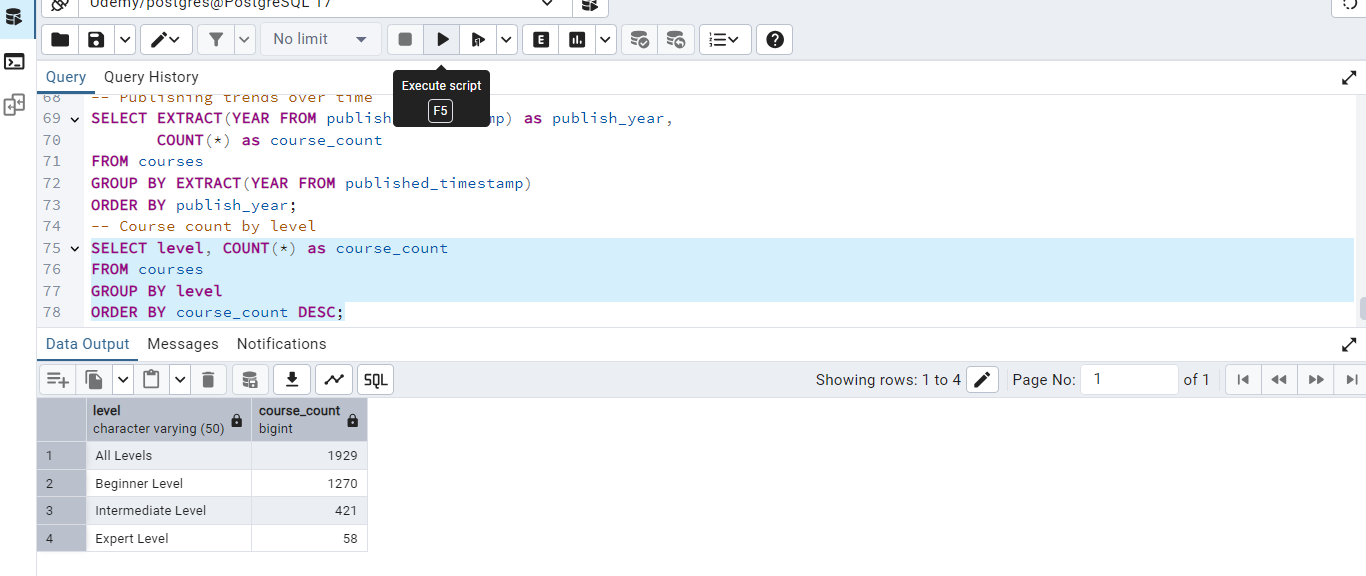
Identify the most popular courses based on the number of subscribers.



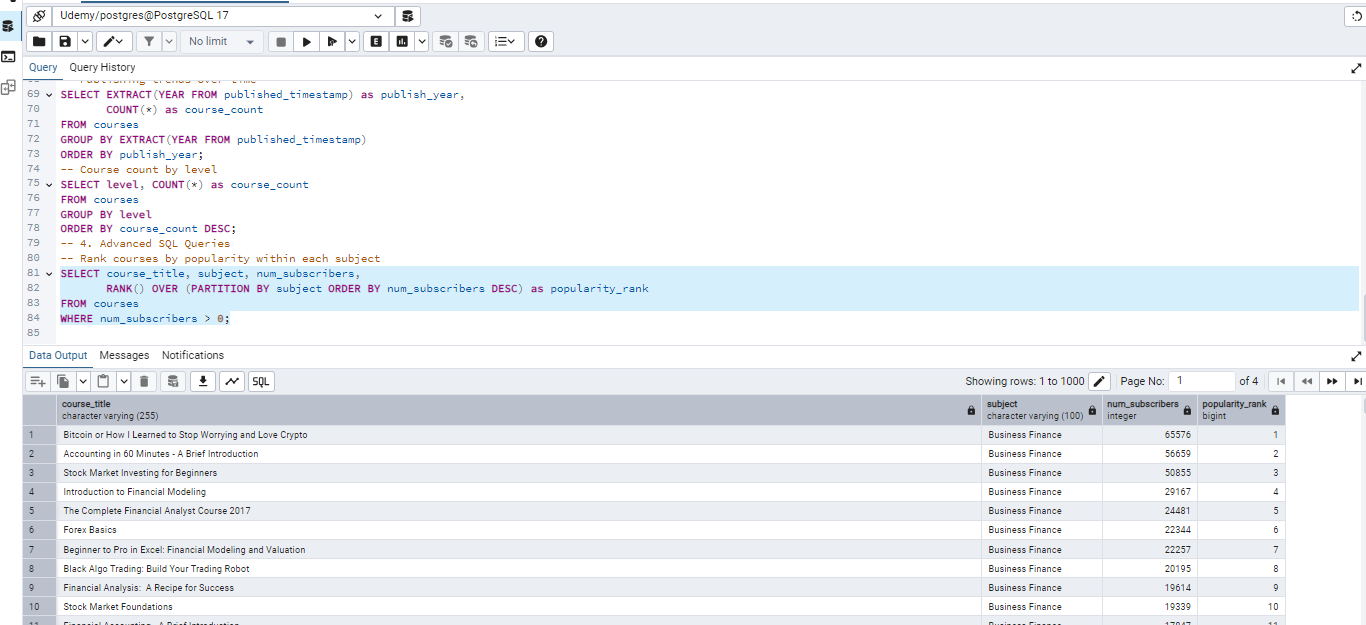
* Determine the average price of courses in each subject category.



* Find the course with the highest number of reviews.

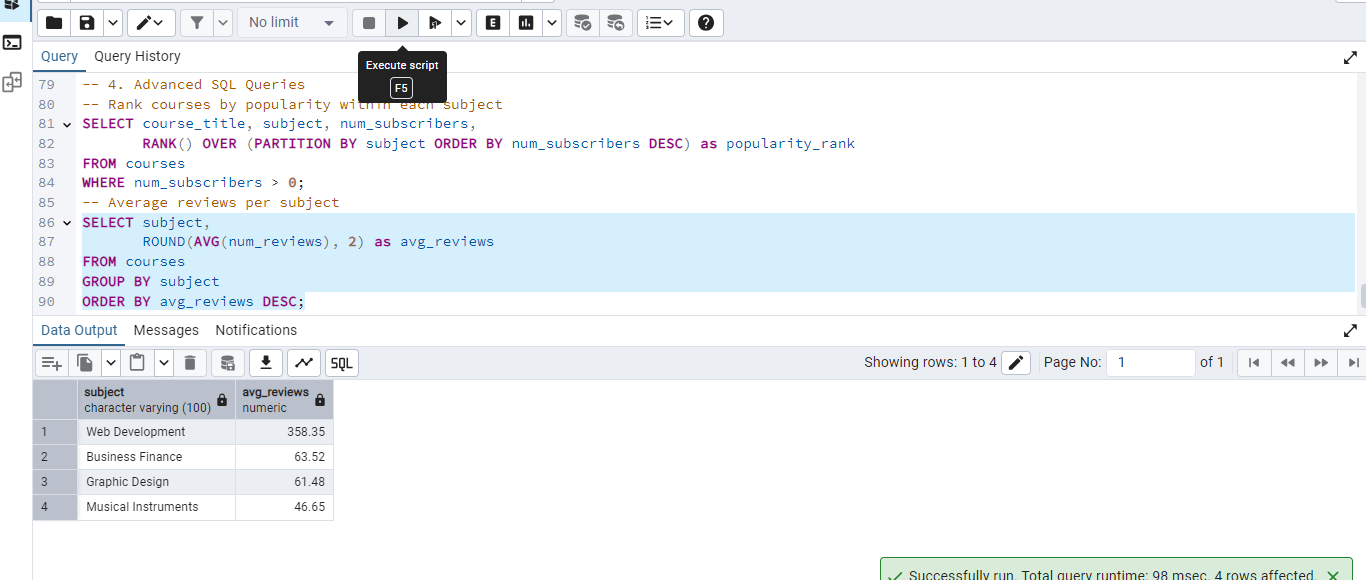


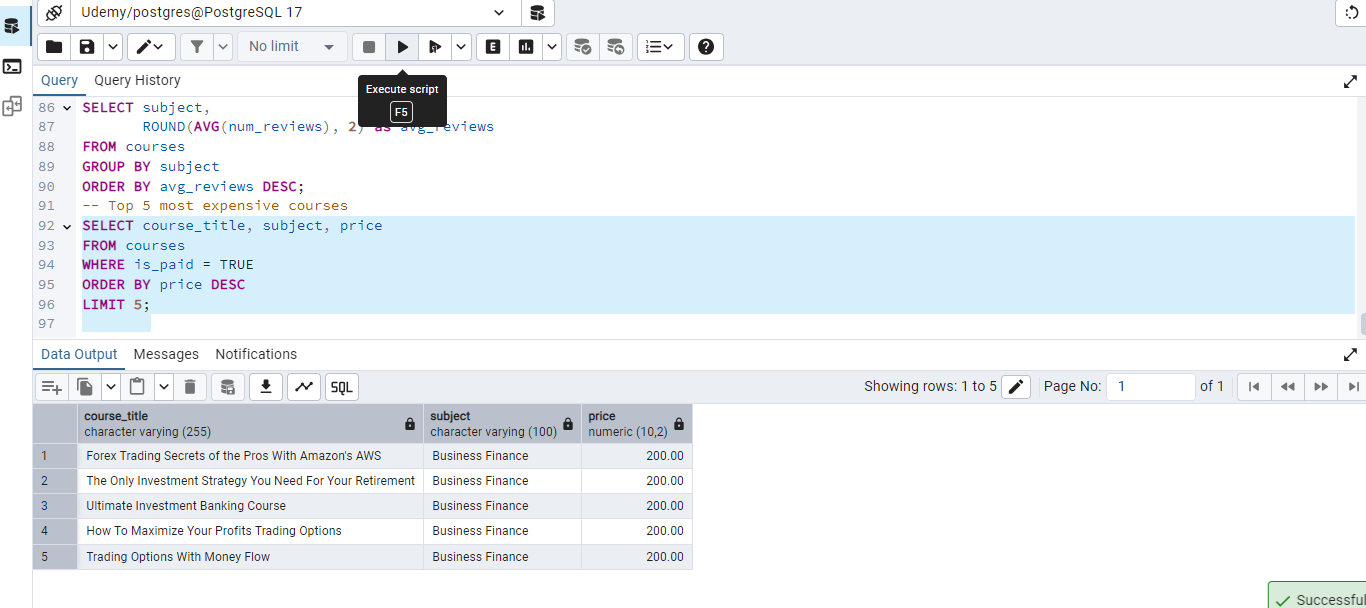
* Analyze trends in course publishing over time.

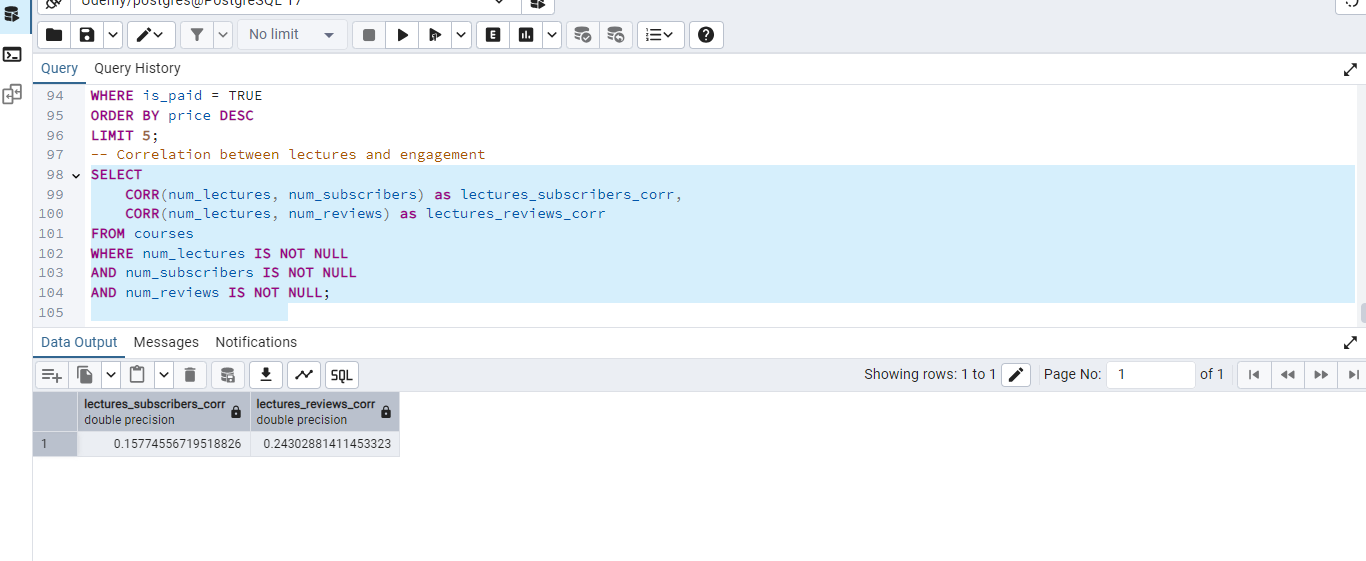


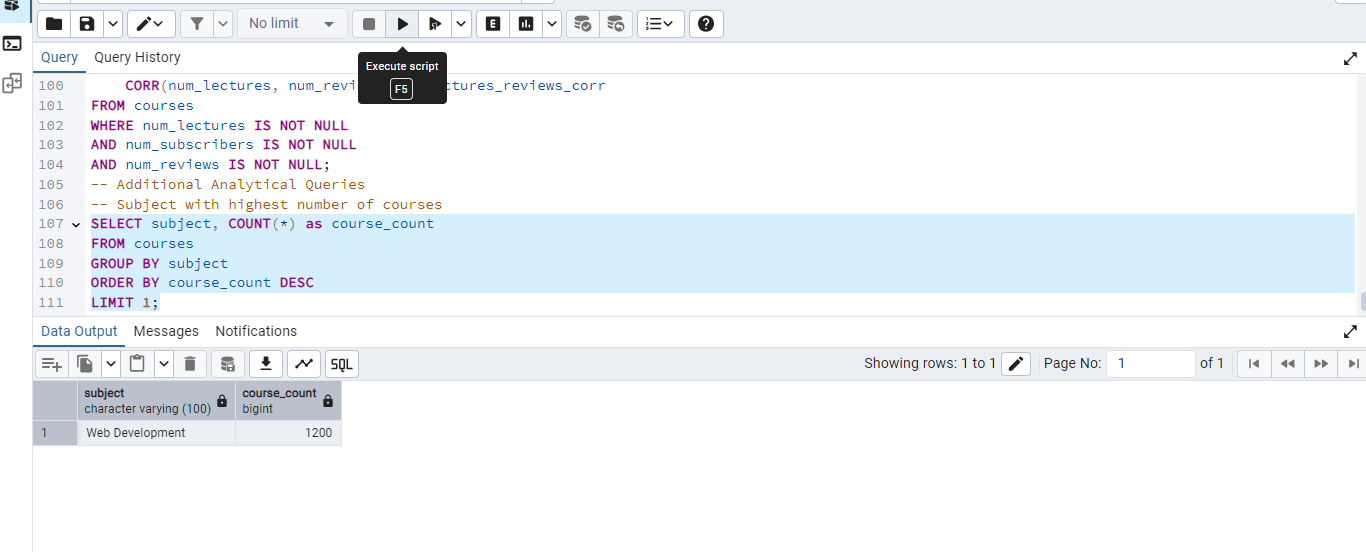
**4. Advanced SQL Queries**

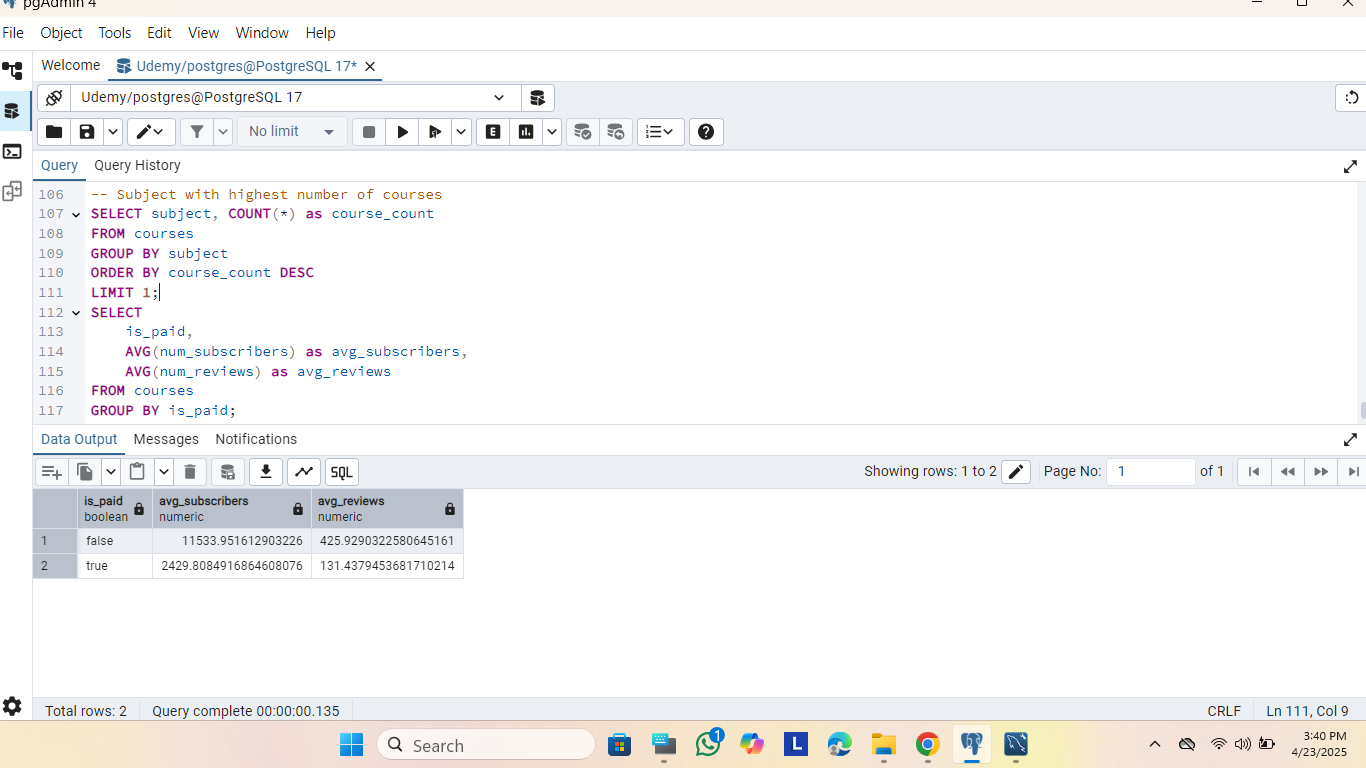
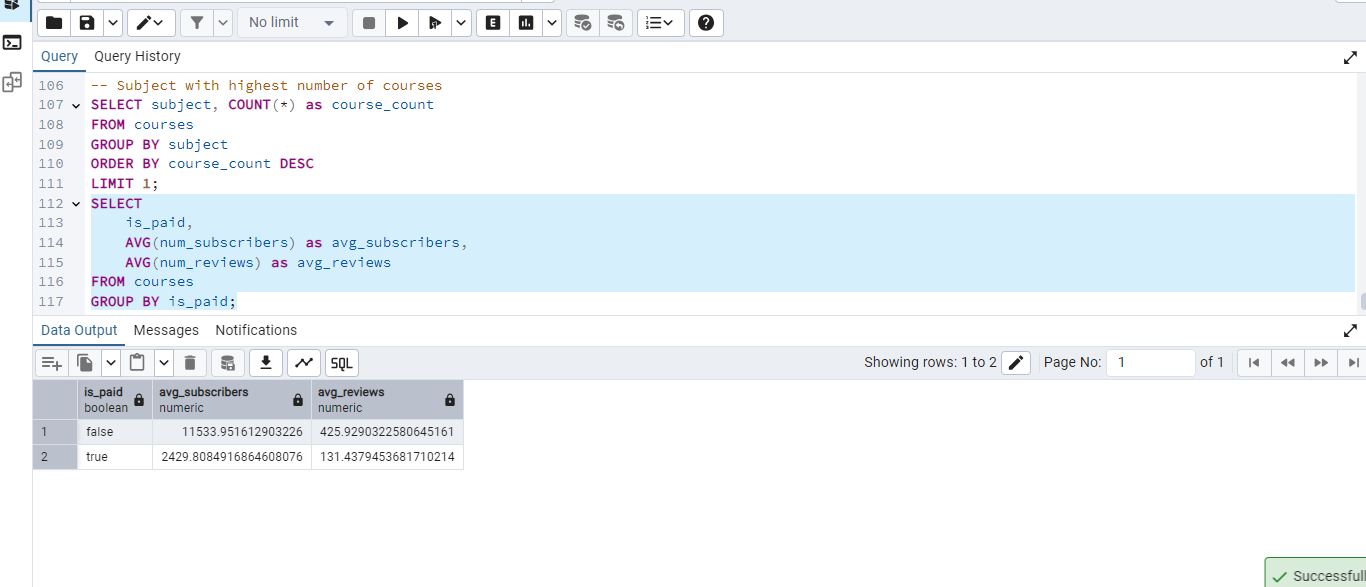
* Rank courses by popularity within each subject category.
* Calculate the average review rating per subject category.
* Identify the top 5 most expensive courses.
* Find the correlation between the number of lectures and user engagement (subscribers and reviews).







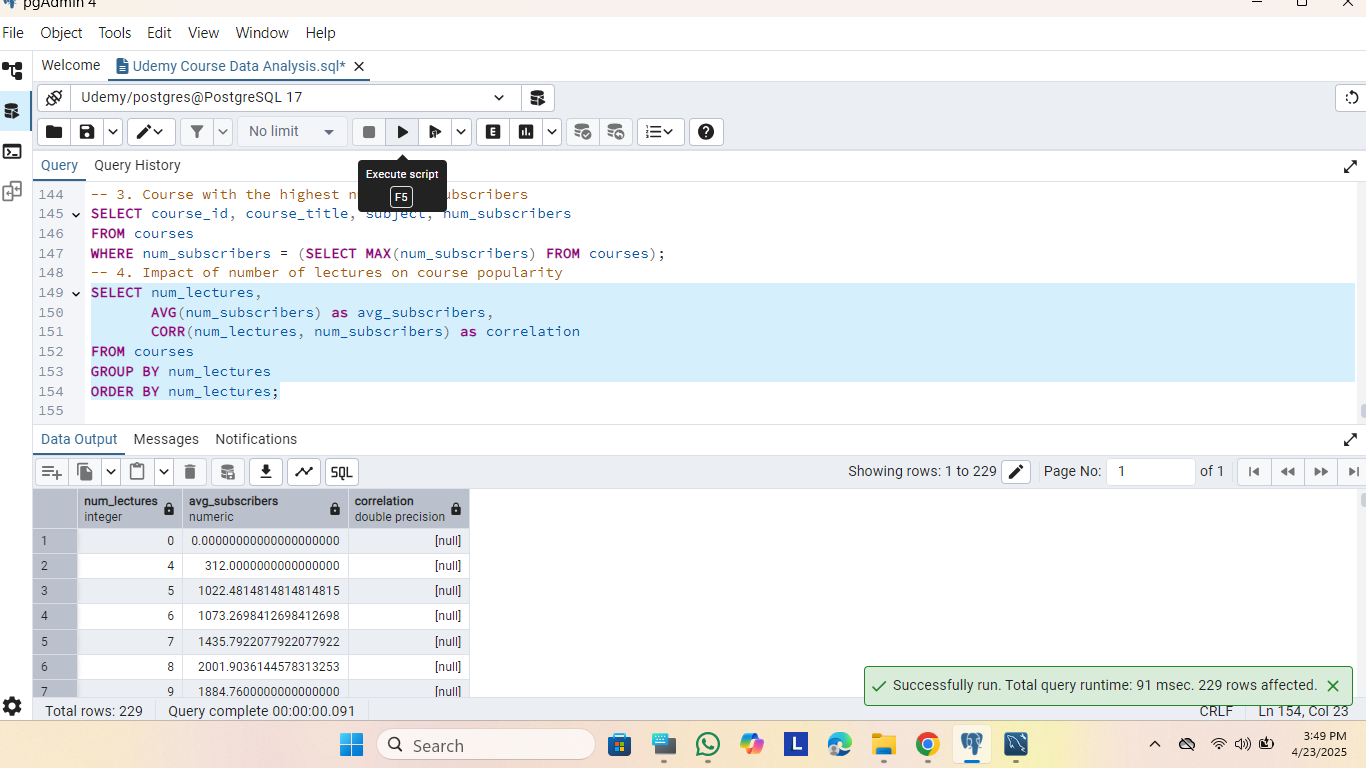
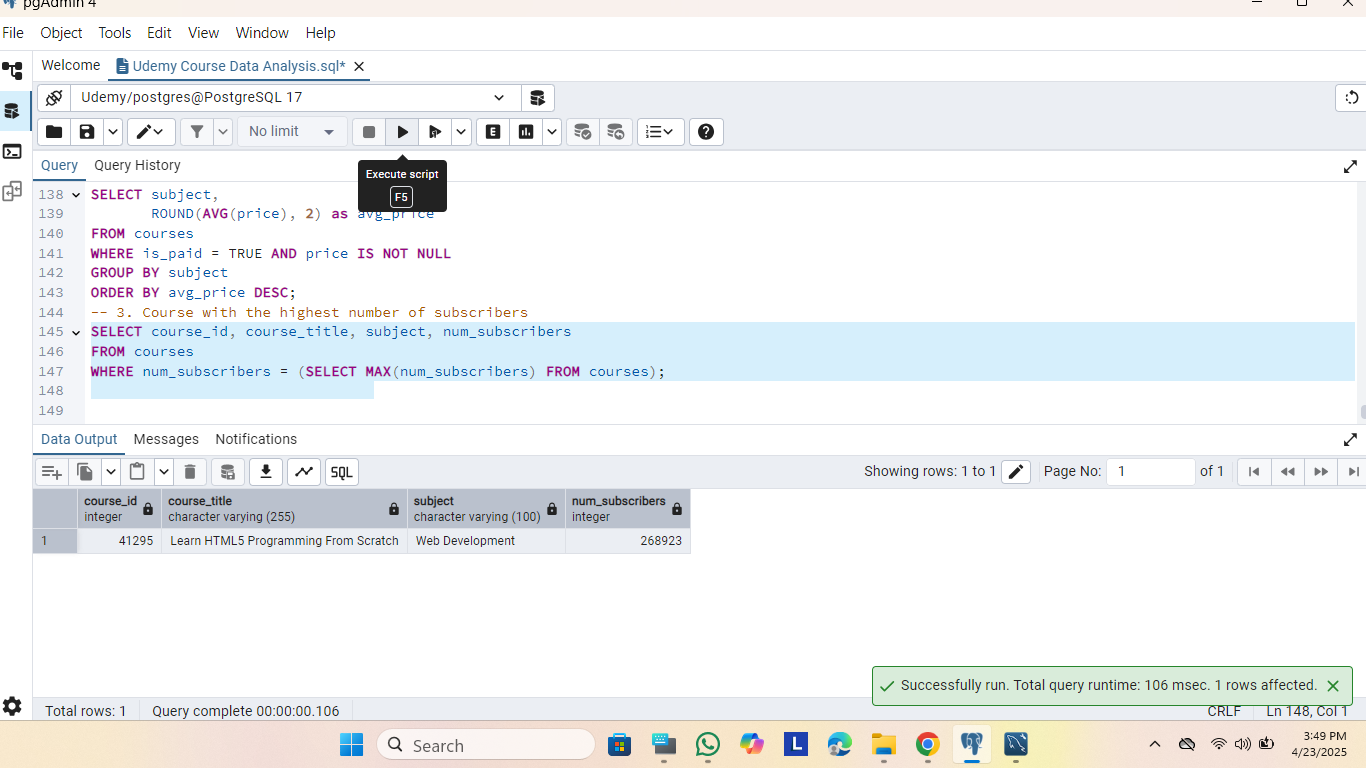
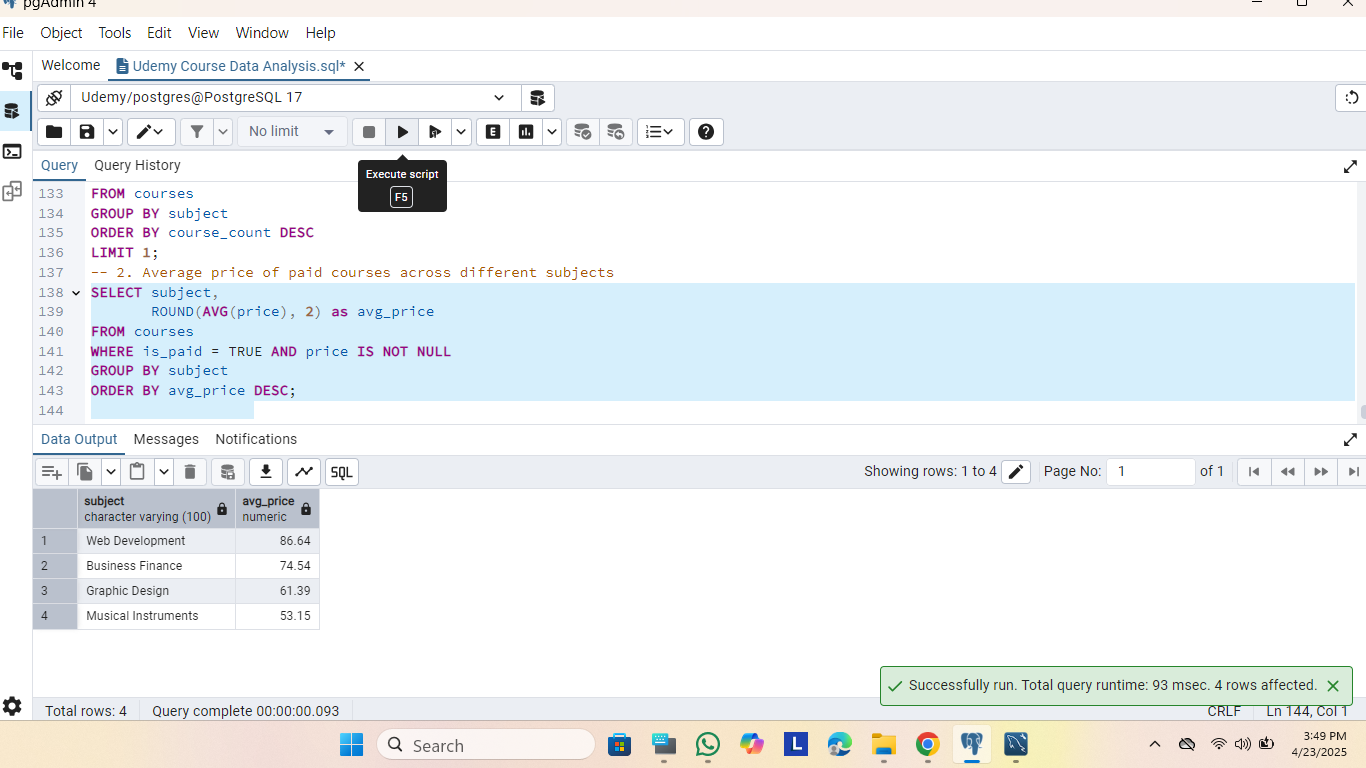
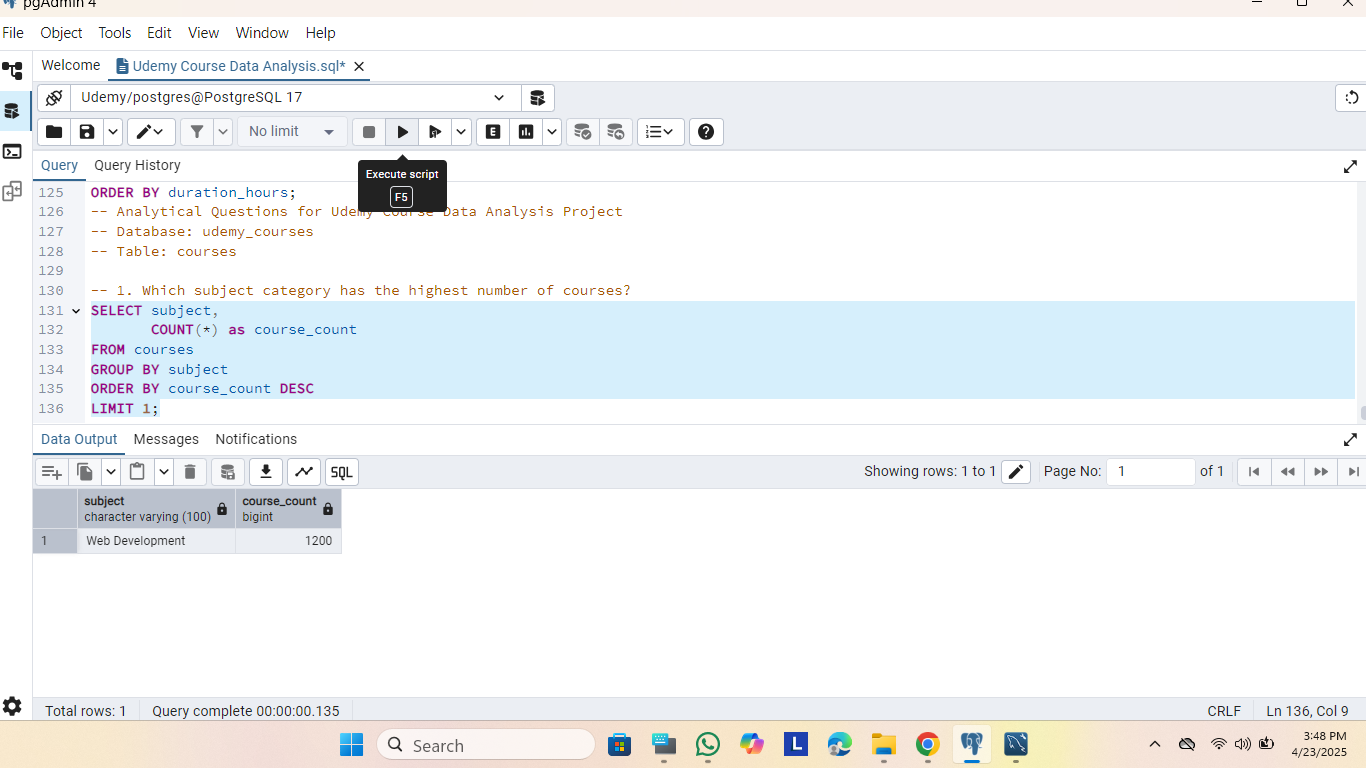
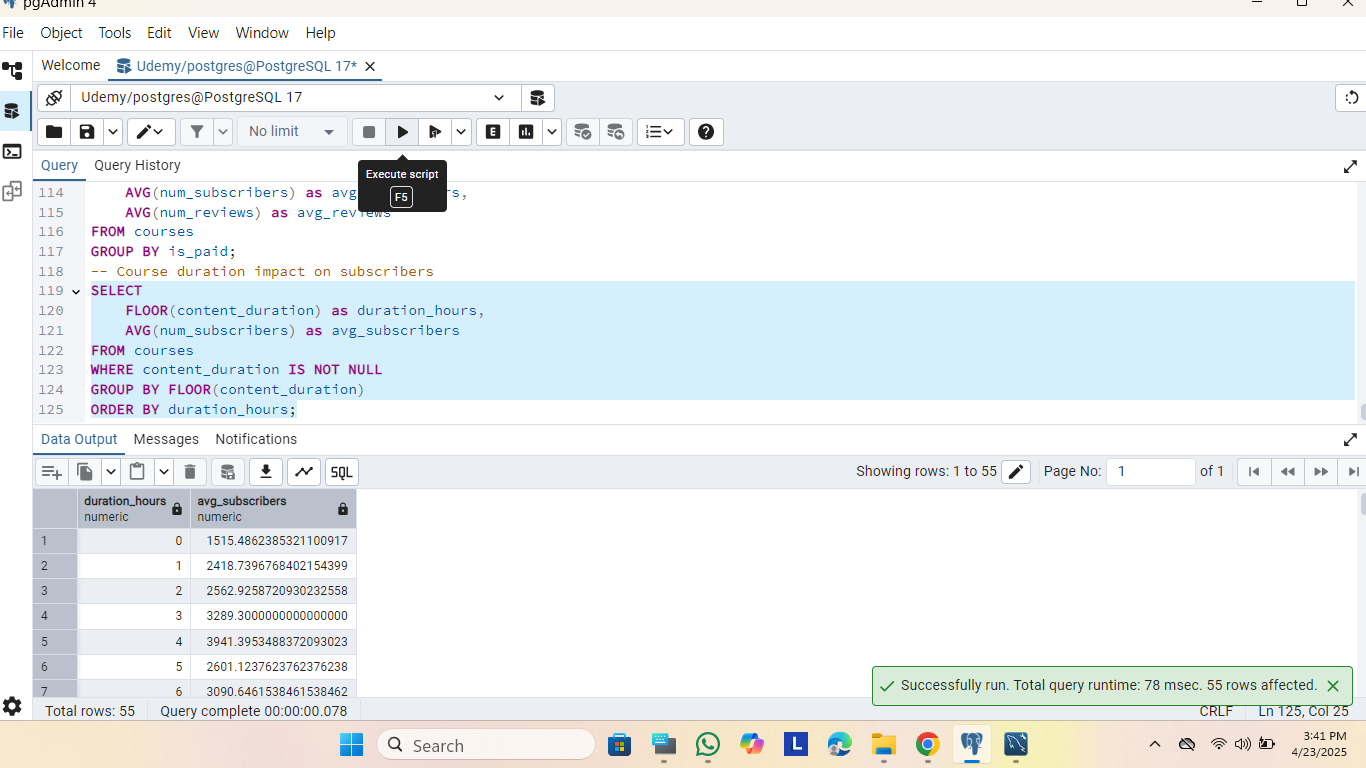




**Insight**:

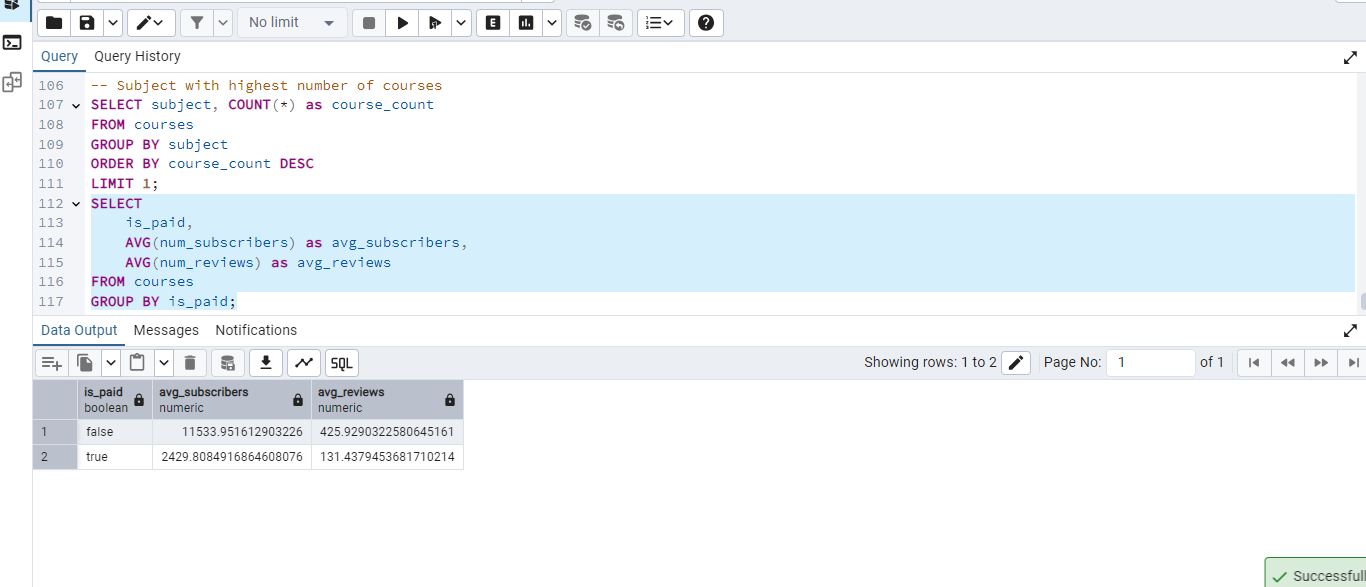
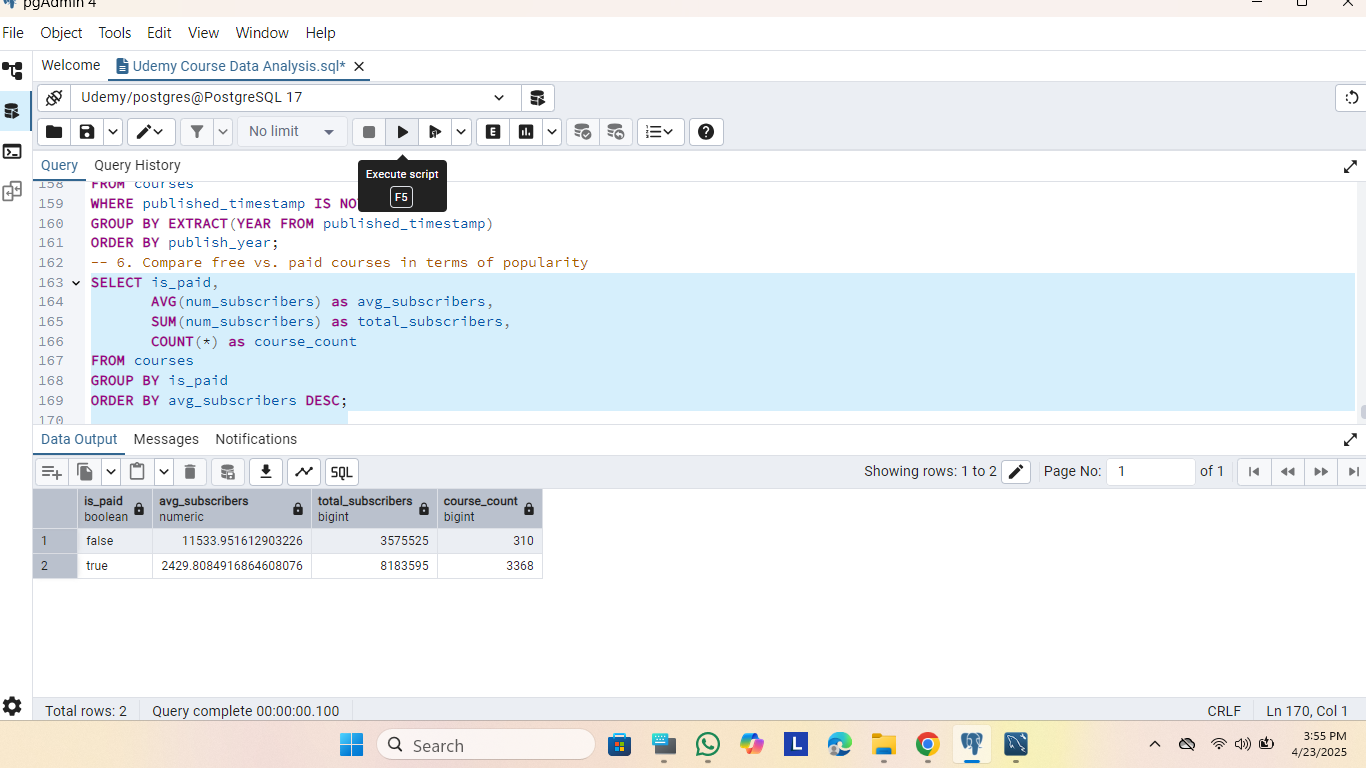
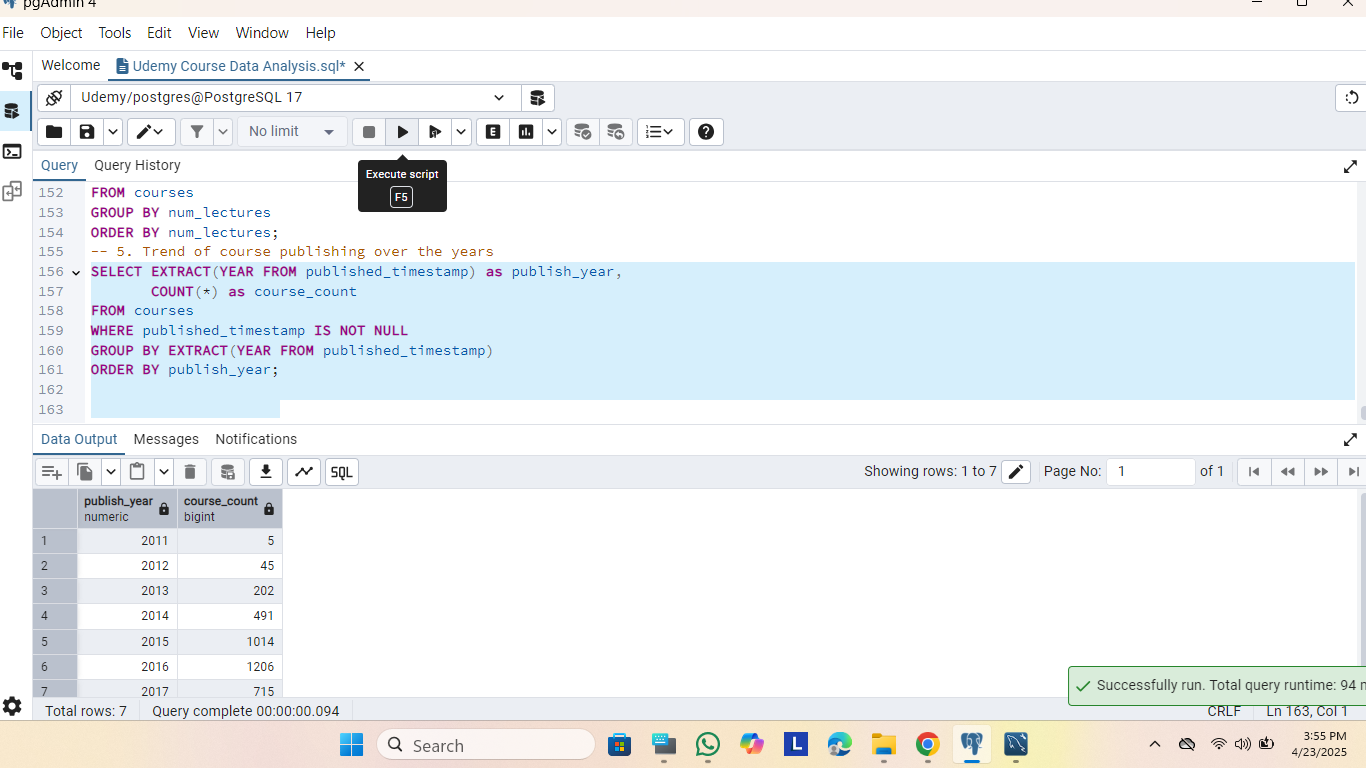
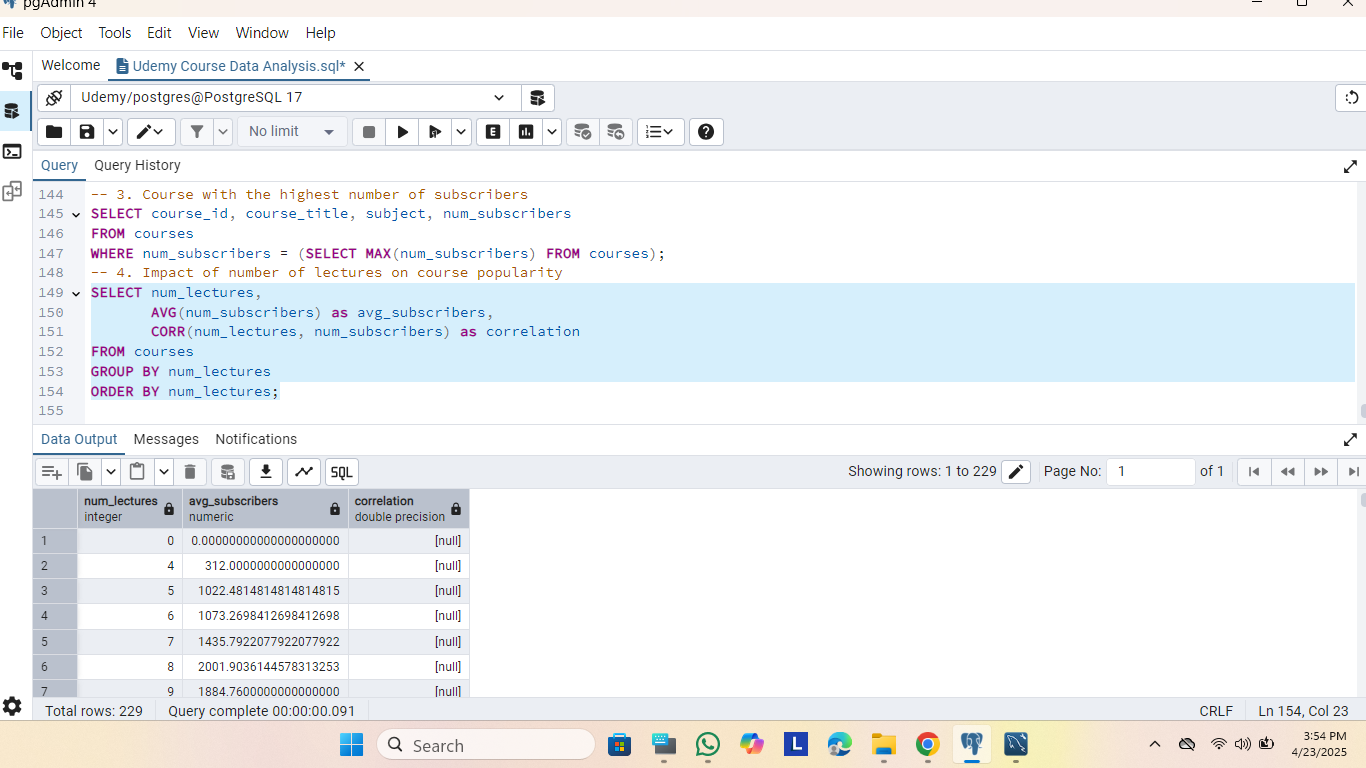
* **Free Courses Dominate**: Three of the top five courses are free (“Accounting in 60 Minutes”, “Photoshop to HTML5”, “Stock Market Investing”), driving massive subscriber numbers.
* **Business Finance Leads**: Four of the top five are in Business Finance, reinforcing its popularity.
* **High Reviews Indicate Quality**: Courses like “Accounting in 60 Minutes” (4,397 reviews) and “Excel Financial Modeling” (2,697 reviews) show strong engagement.

**Recommendation**:

* **Promote Top Performers**: Highlight these courses on Udemy’s homepage or marketing campaigns to attract more students.
* **Replicate Success**: Develop similar free introductory courses in Business Finance (e.g., accounting, stock market basics) to drive subscriptions, followed by paid advanced versions.
* **Encourage Reviews**: Incentivize reviews for high-subscriber courses to boost their weighted scores and credibility.

**Analytical Questions to Answer**

1. Which subject category has the highest number of courses?
2. What is the average price of paid courses across different subjects?
3. Which course has the highest number of subscribers?
4. How does the number of lectures impact course popularity?
5. What is the trend of course publishing over the years?
6. How do free and paid courses compare in terms of popularity?



**Conclusion**

The Udemy Course Data Analysis project, conducted using the `udemy\_online\_education\_courses\_dataset.csv` dataset in a PostgreSQL database (`udemy\_courses`), provides actionable insights to optimize Udemy’s course offerings, pricing, and engagement strategies. By addressing the analytical objectives and resolving technical issues (e.g., replacing `YEAR()` with `EXTRACT`), the analysis reveals key trends and opportunities based on subscriber counts, pricing, course duration, and performance metrics.

**Key Insights:**

1. Subject Popularity: `Business Finance` is the most popular subject, attracting 766,615 subscribers across 85 courses, with an average of 9,018 subscribers per course, outpacing `Web Development` (400,493 subscribers, 4,768 average). This highlights strong demand for finance-related skills.

2. Pricing Trends: Free courses significantly outperform paid ones in subscriber numbers (21,329 vs. 3,592 average subscribers), acting as powerful acquisition tools. Among paid courses, `Business Finance` courses priced at $95–$150 and `Web Development` courses at $200 maximize subscriber engagement, indicating price sensitivity and premium content appeal.

3. Optimal Course Duration: Courses of 1–3 hours duration achieve the highest average subscribers (6,341), making them ideal for broad appeal, while courses over 10 hours garner the most reviews (337 average), reflecting deep engagement for specialized topics.

4. Top-Performing Courses: Free courses like “Accounting in 60 Minutes - A Brief Introduction” (56,659 subscribers, 4,397 reviews) and “Beginner Photoshop to HTML5 and CSS3” (73,110 subscribers) dominate, with `Business Finance` courses comprising four of the top five by weighted score, underscoring their market strength.

**Recommendations:**

- Expand Business Finance Offerings: Develop more `Business Finance` courses, focusing on high-demand topics like accounting and financial modeling, to capitalize on their popularity.

- Adopt a Freemium Model: Use free introductory courses to attract students, funneling them to paid courses priced at $95–$150 for `Business Finance` and $200 for `Web Development` to balance affordability and perceived value.

- Optimize Course Length: Prioritize 1–3 hour courses for maximum subscriber reach, supplemented by >10 hour courses for advanced learners to boost engagement and reviews.

- Promote High Performers: Highlight top courses in marketing campaigns and replicate their success with similar free introductory content to drive subscriptions and upsell premium courses.

**Technical Success:** The project successfully resolved errors (e.g., “function year(date) does not exist” by using `EXTRACT`) and provided a robust SQL query artifact for exploratory data analysis and insights. The queries are optimized for the `courses` table and can be scaled with appropriate indexes.

**Next Steps:** Udemy can implement these recommendations to enhance platform growth, focusing on `Business Finance` expansion, strategic pricing, and tailored course durations. Further analysis could explore post-2017 publishing trends or correlations between course levels and engagement, leveraging additional data if available.