## 2.1: Intro to Data Visualization

**Step 1**

If you haven't done so already, install PostgreSQL and load the Rockbuster database using the instructions in the Exercise. Then download your Achievement 3 project brief (PDF) to get an idea of what each Exercise will cover.

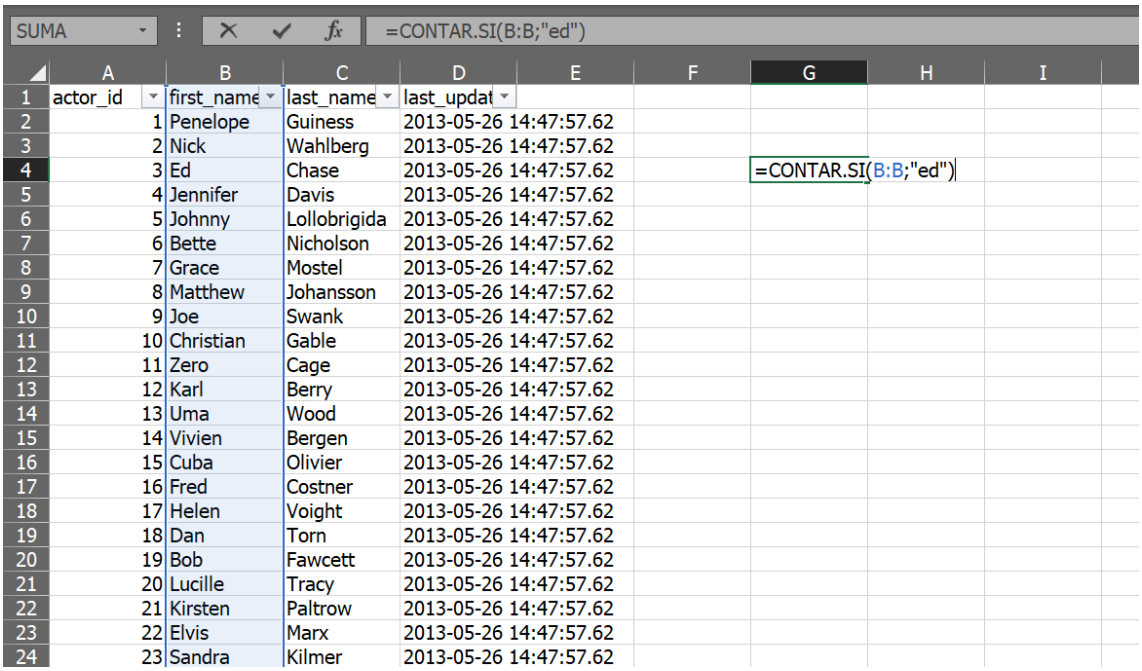
Already done

**Step 2**

Compare and contrast spreadsheets and databases by following the steps below:

Download the Rockbuster "actor.csv" file and open it in Excel.

Drawing on what you've learned in previous Achievements, use the appropriate functions in Excel to count all the actors whose first name is "Ed." Write down the result in a text document.

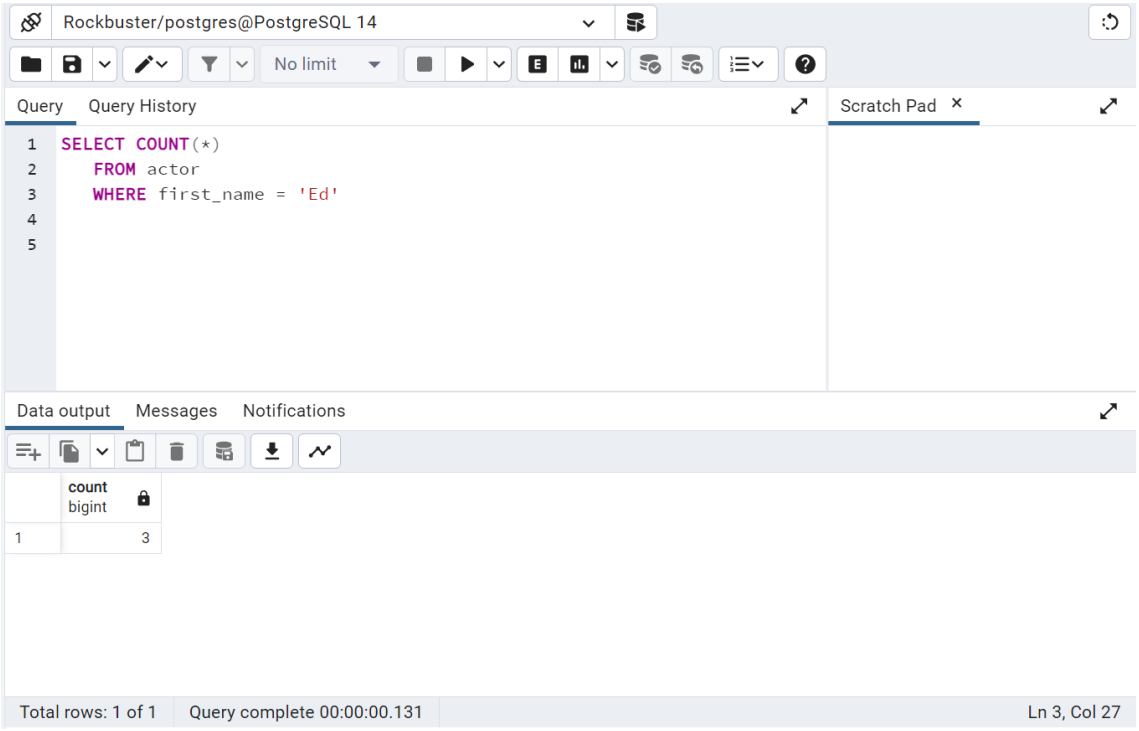



| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | actor_id | first_name | last_name | last_updat | | | | | |
| 2 | 1 | Penelope | Guiness | 2013-05-26 14:47:57.62 | | | | | |
| 3 | 2 | Nick | Wahlberg | 2013-05-26 14:47:57.62 | | | | | |
| 4 | 3 | Ed | Chase | 2013-05-26 14:47:57.62 | | | =CONTAR.SI(B:B;"ed") | | |
| 5 | 4 | Jennifer | Davis | 2013-05-26 14:47:57.62 | | | | | |
| 6 | 5 | Johnny | Lollobrigida | 2013-05-26 14:47:57.62 | | | | | |
| 7 | 6 | Bette | Nicholson | 2013-05-26 14:47:57.62 | | | | | |
| 8 | 7 | Grace | Mostel | 2013-05-26 14:47:57.62 | | | | | |
| 9 | 8 | Matthew | Johansson | 2013-05-26 14:47:57.62 | | | | | |
| 10 | 9 | Joe | Swank | 2013-05-26 14:47:57.62 | | | | | |
| 11 | 10 | Christian | Gable | 2013-05-26 14:47:57.62 | | | | | |
| 12 | 11 | Zero | Cage | 2013-05-26 14:47:57.62 | | | | | |
| 13 | 12 | Karl | Berry | 2013-05-26 14:47:57.62 | | | | | |
| 14 | 13 | Uma | Wood | 2013-05-26 14:47:57.62 | | | | | |
| 15 | 14 | Vivien | Bergen | 2013-05-26 14:47:57.62 | | | | | |
| 16 | 15 | Cuba | Olivier | 2013-05-26 14:47:57.62 | | | | | |
| 17 | 16 | Fred | Costner | 2013-05-26 14:47:57.62 | | | | | |
| 18 | 17 | Helen | Voight | 2013-05-26 14:47:57.62 | | | | | |
| 19 | 18 | Dan | Torn | 2013-05-26 14:47:57.62 | | | | | |
| 20 | 19 | Bob | Fawcett | 2013-05-26 14:47:57.62 | | | | | |
| 21 | 20 | Lucille | Tracy | 2013-05-26 14:47:57.62 | | | | | |
| 22 | 21 | Kirsten | Paltrow | 2013-05-26 14:47:57.62 | | | | | |
| 23 | 22 | Elvis | Marx | 2013-05-26 14:47:57.62 | | | | | |
| 24 | 23 | Sandra | Kilmer | 2013-05-26 14:47:57.62 | | | | | |

The answer is 3

Launch pgAdmin 4, open the Query Tool, copy-paste the SQL statement below into the Query Editor, and execute it.

```
SELECT COUNT(*)
FROM actor
WHERE first_name = 'Ed'
```




TIP

This statement will count all the instances of an actor with a first name “Ed” in the table actor. Copy the result that tells you the number of times the first name “Ed” appears in the “actor” table from the Data Output window into your text document from step 2b. Check that your answer matches your answer from step 2a. Was it easier to use Excel or the SQL statement and database to count the number of “Eds”? Provide an explanation for your answer in the same text document.

Easier in SQL, but kind of “blind” because we have not even seen the data we are working with

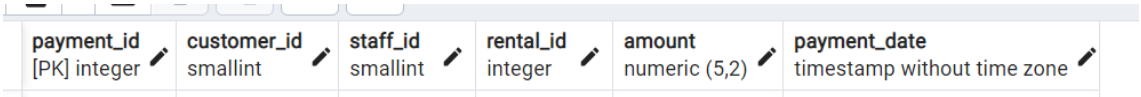
**Step 3**

To answer the next set of questions, you’ll be pasting the queries provided into the Query Editor in pgAdmin 4. Note down your answers in your running text document.

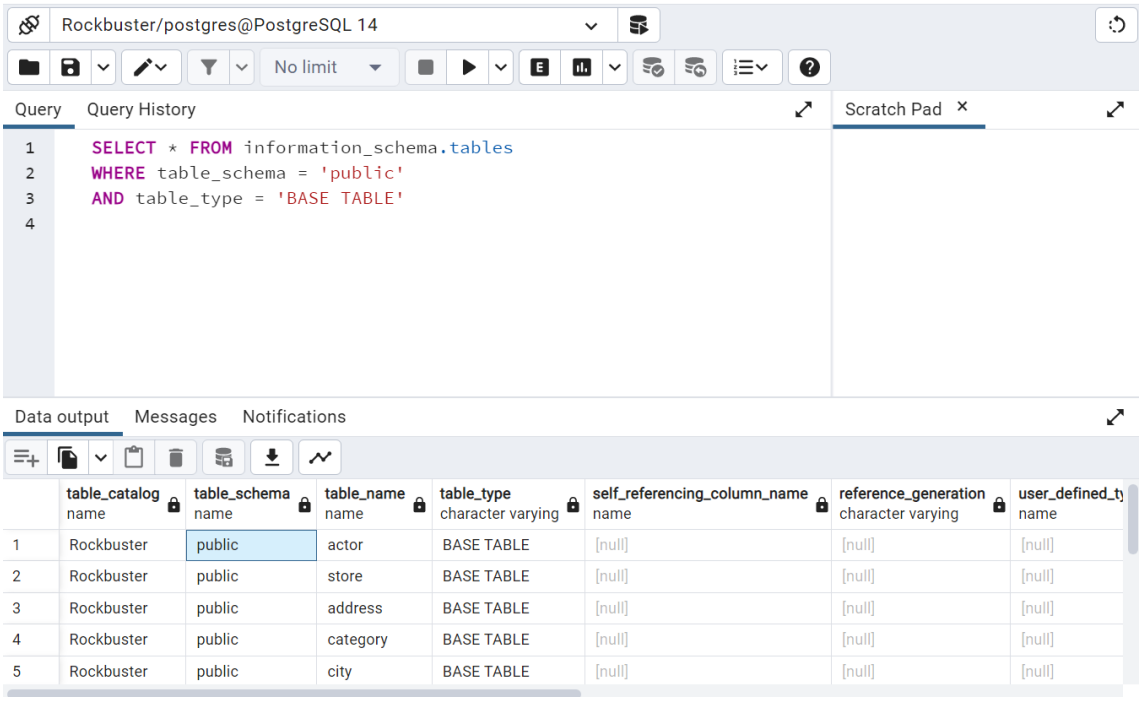
Execute the following query and list the names of the columns in the payment table.

```
SELECT * FROM payment LIMIT 10;
```

Under the “table_name” column, what are the names of the tables that are available in the Rockbuster database? (List all names.)




```
SELECT * FROM information_schema.tables
WHERE table_schema = 'public'
AND table_type = 'BASE TABLE'
```

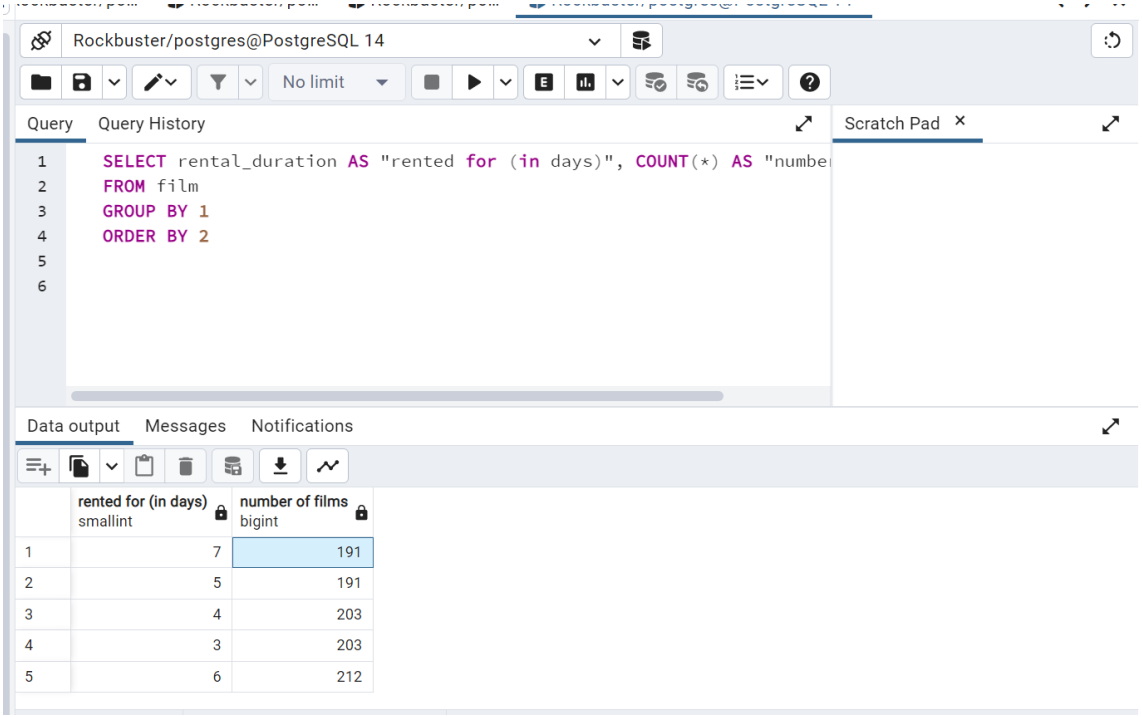



| | table_catalog name | table_schema name | table_name name | table_type character varying | self_referencing_column_name name | reference_generation character varying | user_defined_ty name |
|---|---|---|---|---|---|---|---|
| 1 | Rockbuster | public | actor | BASE TABLE | [null] | [null] | [null] |
| 2 | Rockbuster | public | store | BASE TABLE | [null] | [null] | [null] |
| 3 | Rockbuster | public | address | BASE TABLE | [null] | [null] | [null] |
| 4 | Rockbuster | public | category | BASE TABLE | [null] | [null] | [null] |
| 5 | Rockbuster | public | city | BASE TABLE | [null] | [null] | [null] |

Within the pgAdmin 4 console, can you think of another way to list all the table names in the database instead of the SQL statement above?

Analyze the rental duration distribution. How many days are most films rented for?

```
SELECT rental_duration AS "rented for (in days)", COUNT(*) AS "number of films"
FROM film
GROUP BY 1
ORDER BY 2
```




| | rented for (in days) smallint | number of films bigint |
|---|---|---|
| 1 | 7 | 191 |
| 2 | 5 | 191 |
| 3 | 4 | 203 |
| 4 | 3 | 203 |
| 5 | 6 | 212 |

**Step 4**

Think about who in Rockbuster Stealth might want to use an OLAP or OLTP system for their data needs; for example, the sales department, which is interested in sales trends, would likely use an OLAP system. Describe at least 2 situations for each type of system.

Purchase department would be also use an OLAP system to monitoring all the orders and follow estimated delivery dates. Also they could use it to build some dashboards of products/needs in order to search for alternative products in order to improve the margins

Finance department, and I know it well as I work in one of them could be interested on OLTP systems, as they may consume information (and an OLAP system would be enough), but also they may be interested in insert and update transactions. A simple example is an analytic P&L account, where they have to make expense distribution of cost centers between a lot of benefit center

**Step 5**

Rockbuster Stealth has received an invoice for the licenses for its new video collection.

Take a moment to familiarize yourself with data in the invoice, then note down the answers to the questions below.

Does the invoice contain structured or unstructured data? Write an explanation for your answer.

Structure

Organize and store the information on the invoice in a database. Step one will be to create a table in the text document you've started (you can insert a table if you're using MS Word or Google Docs, for example). Make sure your table contains columns with the appropriate labels, as well as the values from the invoice in each column. You're focusing, here, on a high-level structuring of your data.

| **Headboard/Positions** | **Master Data** | **Features** | |
|---|---|---|---|
| Headboard | Invoice # | | |
| | Customer | Name | |
| | | Code | |
| | | Adress | Street |
| | | | Postal Code |
| | | | City |
| | Company | Name | |
| | | Code | |
| | | Adress | Street |
| | | | Postal Code |
| | | | City |
| | Bank detail | Account name | |
| | | Account Number | |

| Positions | Material # | Material Description<br>Unit Price |
| --- | --- | --- |

Quantity and total amount of the invoice would not be a feature of the invoice as it has to come from the sale order/delivery note

**Step 6**

Save the text document containing your answers as a PDF and upload it here for your tutor to review. Don’t hesitate to contact your tutor or mentor if you have any questions!