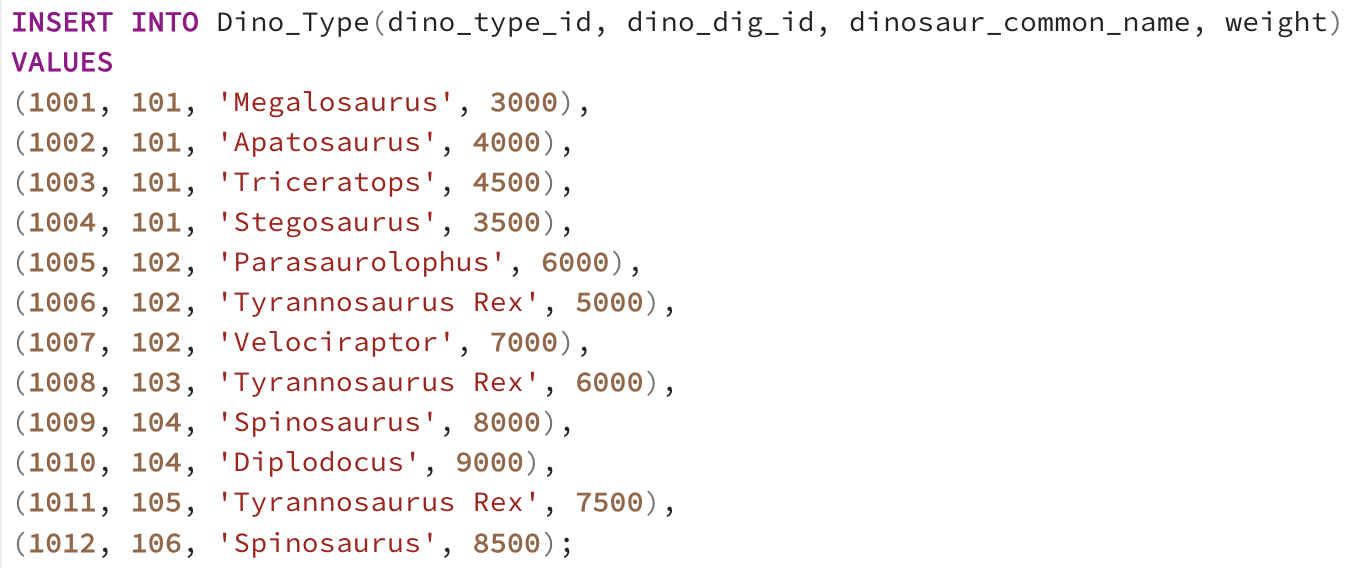
***Section One – Aggregating Data***

1. *Creating Table Structure and Data* – Create the tables in the schema, including all of their columns, datatypes, and constraints, and populate the tables with data. Most but not all the data is given to you in the table below; ***you should also insert information for one additional dinosaur discovery of your choosing.*** Although the data is in flattened representation below, you will need to insert the data relationally into the schema with foreign keys referencing the appropriate primary keys. You may choose any primary key values you would like for each table. We will learn in a later lab how to automatically generate primary key values.

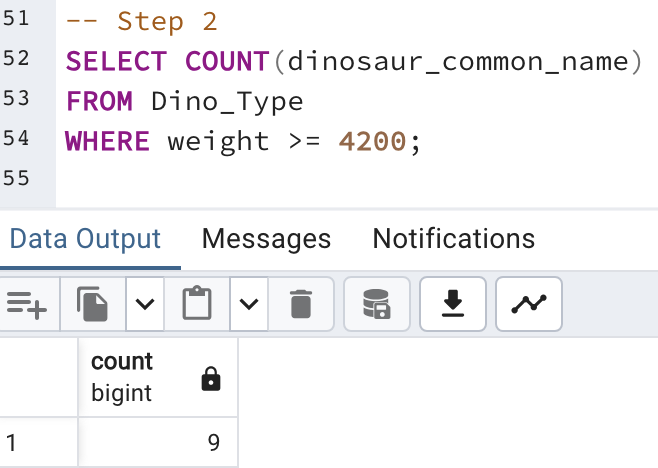
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Location** | **Dig Name** | **Dig Cost** | **Dinosaur Common Name** | **Weight**  **(in pounds)** | **Paleontologist** |
| Stonesfield | Great British Dig | $8,000 | Megalosaurus | 3000 | William Buckland |
| Stonesfield | Great British Dig | $8,000 | Apatosaurus | 4000 | William Buckland |
| Stonesfield | Great British Dig | $8,000 | Triceratops | 4500 | William Buckland |
| Stonesfield | Great British Dig | $8,000 | Stegosaurus | 3500 | William Buckland |
| Utah | Parowan Dinosaur Tracks | $10,000 | Parasaurolophus | 6000 | John Ostrom |
| Utah | Parowan Dinosaur Tracks | $10,000 | Tyrannosaurus Rex | 5000 | John Ostrom |
| Utah | Parowan Dinosaur Tracks | $10,000 | Velociraptor | 7000 | John Ostrom |
| Arizona | Dynamic Desert Dig | $3,500 | Tyrannosaurus Rex | 6000 | John Ostrom |
| Stonesfield | Mission Jurassic Dig |  | Spinosaurus | 8000 | Henry Osborn |
| Stonesfield | Mission Jurassic Dig |  | Diplodocus | 9000 | Henry Osborn |
| Stonesfield | Ancient Site Dig | $5,500 | Tyrannosaurus Rex | 7500 | Henry Osborn |

Note that the Dig Cost for “Mission Jurassic Dig” is null (has no value).

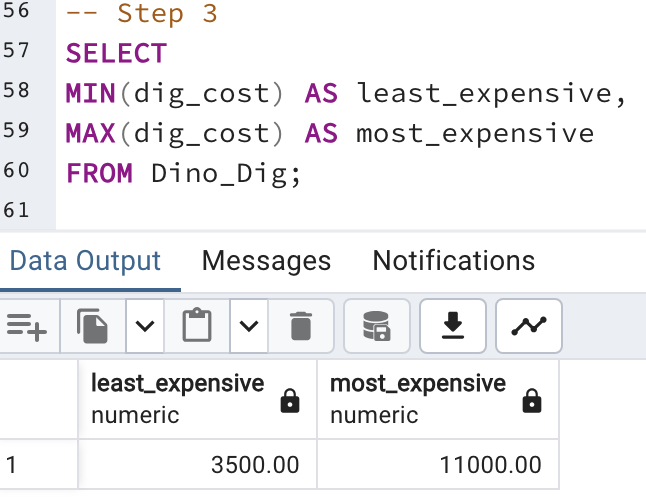




1. *Counting Matches* – A museum wants to know how many dinosaur discoveries weigh at least 4,200 pounds. Write a single query to fulfill this request.

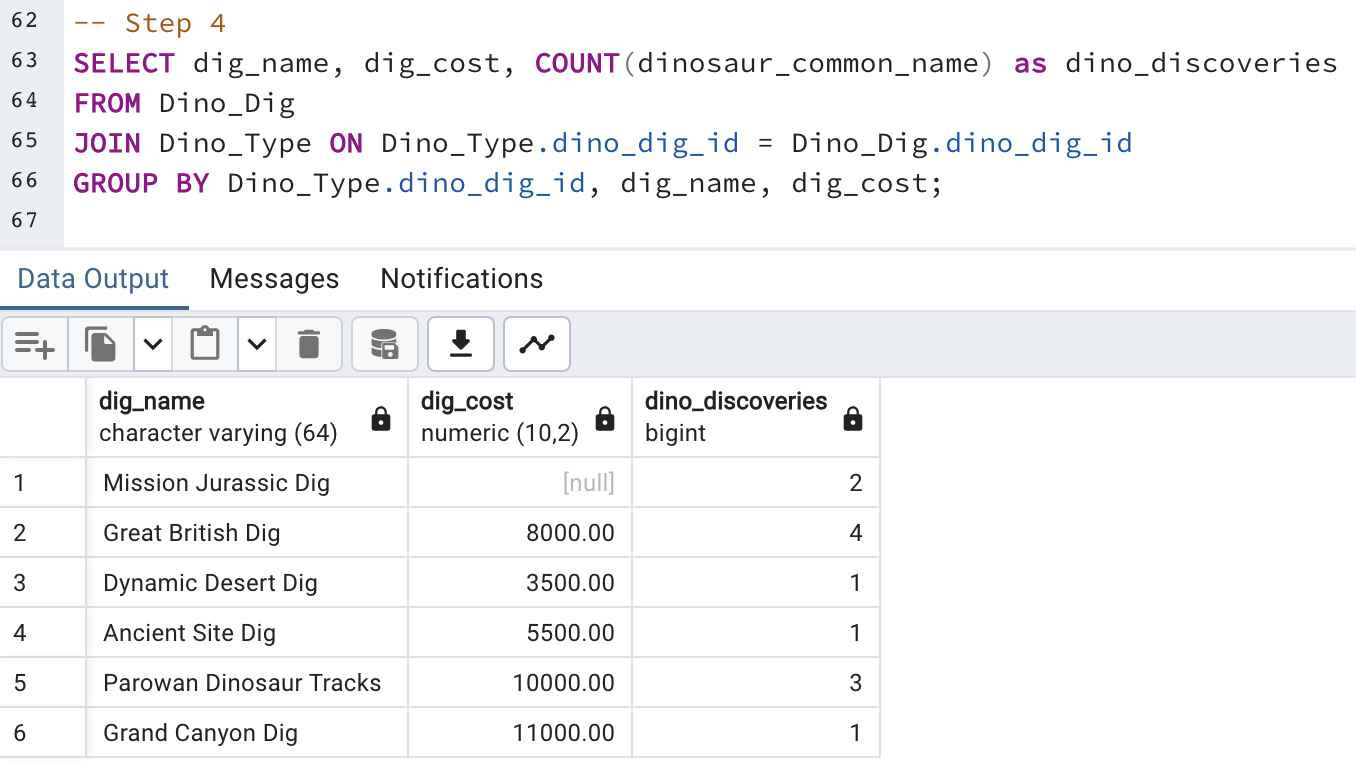


1. *Determining Highest and Lowest –* The same museum needs to know the cost of the most expensive and least expensive dinosaur digs. Write a single query to fulfill this request. Explain how the SQL processer treated the dig costs for the “Mission Jurassic Dig” differently than the other cost values.



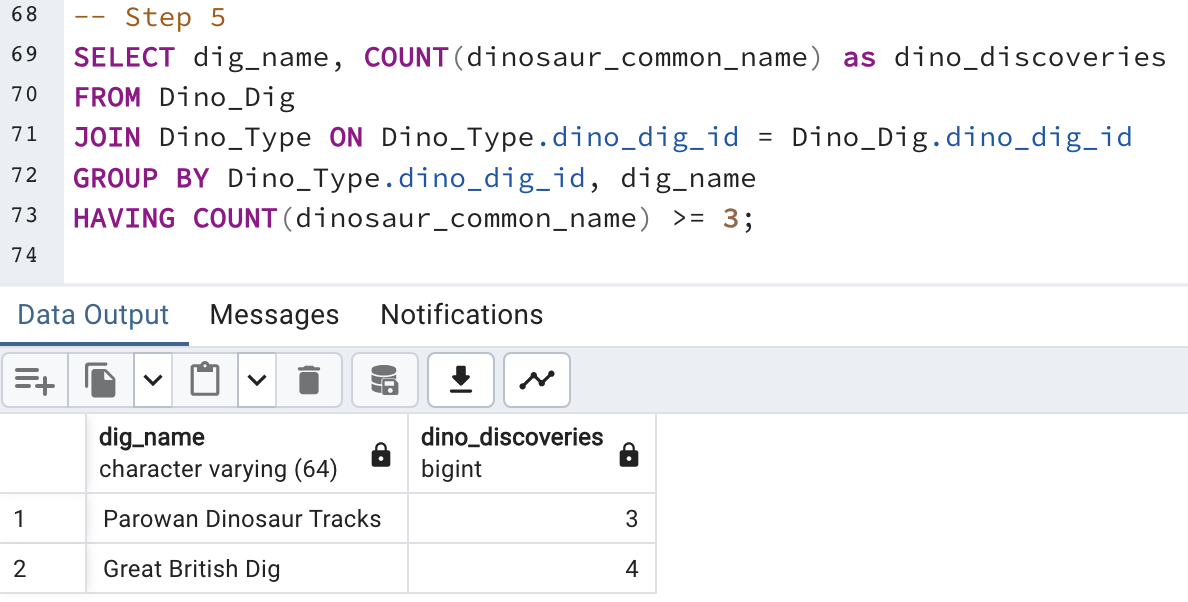
The SQL processor ignores null values so the cost for the “Mission Jurassic Dig” was ignored as it’s cost was input as null.

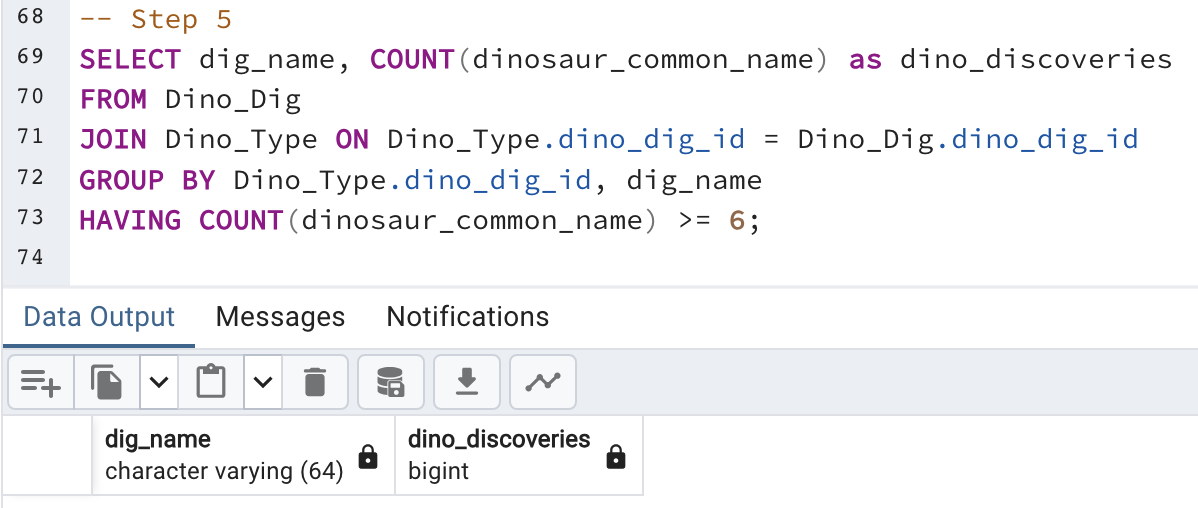
1. *Grouping Aggregate Results –* A museum is considering supporting their own paleontological expedition and needs to know the dig site name and cost, along with the number of dinosaur discoveries at each site. Write a single query to fulfill this request.



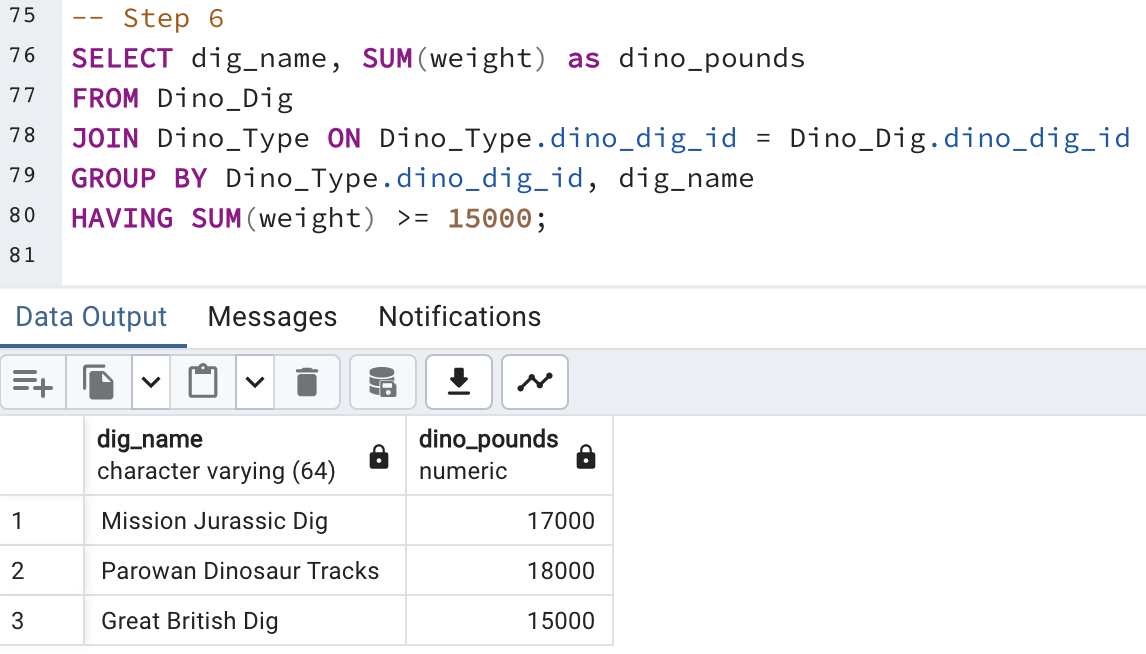
1. *Limiting Results by Aggregation* – A paleontologist, looking to dig at a location ripe with discoveries, wants to search for locations with at least 6 dinosaur discoveries. Write a single query to fulfill this request.

I wrote this for 3 discoveries so that I can show it works and then again for 6.

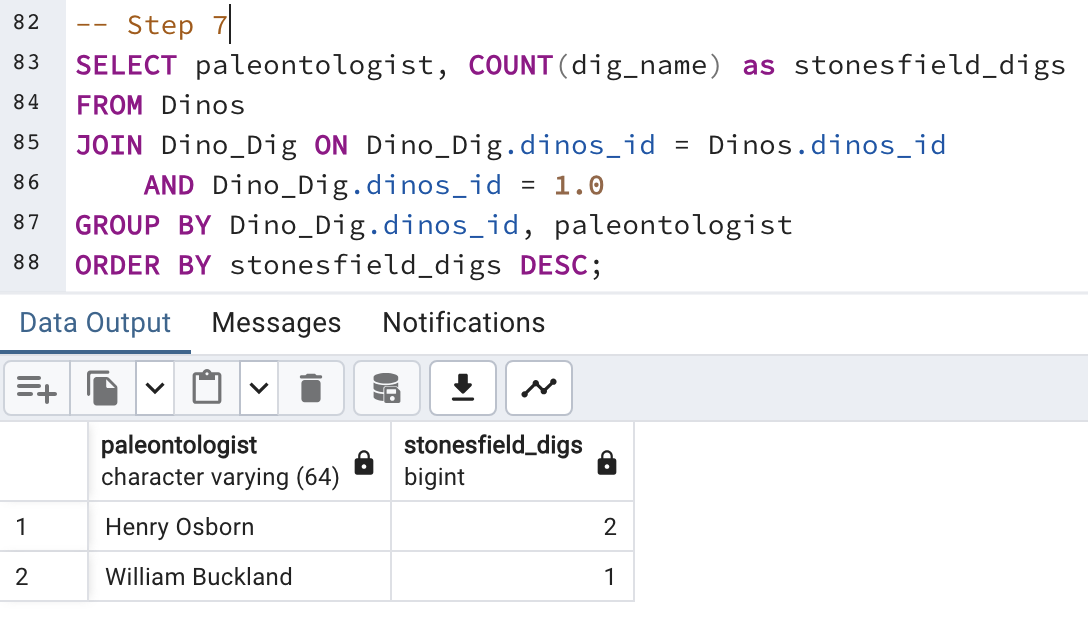




1. *Adding Up Values –* A museum needs to know which digs (not locations) had at least 15,000 pounds of discovered dinosaur remains. Write a single query that gives this information, with useful columns.



1. *Integrating Aggregation with Other Constructs –* A research institution wants the names of all paleontologists. For each paleontologist, the institution also wants the number of digs they participated in at the "Stonesfield" location, even if the number is 0. The institution wants the list to be ordered from most to least; the paleontologist who discovered the most Stonesfield dinosaurs will be at the top of the list, and the one with the least will be at the bottom. Write a single query that gives this information, with useful columns.



***Section Two – Data Visualization***

1. *Visualizing Data with One or Two Measures –* Use the SQL results obtained for Step #4 to address the following.  
   1. Create a bar chart with the dig name as one axis, and the dig cost as another axis. Explain the story this visualization describes.

This tells us how much more expensive certain digs are than others. This emphasizes how cheap the Dynamic Desert Dig was.

* 1. Create a scatterplot with the dig cost as one axis, and the number of dinosaurs found as another axis. Ensure that each dig name is labeled with its name, either directly or with a legend. Explain the story this visualization describes.

This shows how much more expensive digs are when you find more dinosaurs. From this data we can gather that there isn’t much correlation between cost and dinosaurs found.

1. *Another Data Visualization –* Create a visualization of your choosing for data in the Dinosaur schema. The visualization should tell a useful story. If you find that you need more dinosaurs in the schema to tell the story well, feel free to add them. Make sure to explain the data story, and to explain why you chose that particular chart or visualization.

This graph shows us how successful each site was in finding dinosaur remains. Pounds is the best measure of this because counting the dinosaurs does not show how much of the dinosaur they were able to get.