

*SQL for Smart Folks\*\**

Instructor: [Michael Valeri](https://www.linkedin.com/in/mvaleri/), Analytics Practice Lead

|  |
| --- |
| ***\*GETTING STARTED GUIDE\****  *Step 1:* [*Download PG Admin 4*](https://www.pgadmin.org/download/)*. If trouble opening on Mac, see troubleshoot guide* [*here*](https://www.pgadmin.org/faq/)*.*  *Step 2: Connect to AWS server on PG Admin 4; credentials on Page 5*  *Step 3: Start brainstorming questions on introduction example, Page 2* |

[*Please Fill Out Survey at End*](https://www.surveymonkey.com/r/NHRDMQK)

*\*\*YOU*

1. **Introduction – Weekly Business Review**

*What questions can you expect to be generated from the above chart?*

|  |
| --- |
|  |

1. **The Analytics Framework – Where does SQL fit in?**

|  |
| --- |
|  |

1. **Relational Database Fundamentals – Entity Relationship Diagrams (ERDs)**

|  |
| --- |
|  |

1. **Datatype Fundamentals – Tables and Schemas**

*LAUNCHES (Table Schema)*

|  |  |
| --- | --- |
| **Field/Column** | **Data Type** |
| merchant\_id |  |
| merchant\_friendly\_name |  |
| merchant\_launch\_gl |  |
| account\_manager |  |
| launch\_date |  |
| marketplace\_id |  |
| launch\_offer\_count\_mfn |  |
| launch\_offer\_count\_fba |  |
| launch\_day\_gms |  |
| first\_launch |  |
| fba\_enrolled |  |

*MARKETPLACE\_DIM (Table Schema)*

|  |  |
| --- | --- |
| **Field/Column** | **Data Type** |
| marketplace\_id |  |
| marketplace\_name |  |

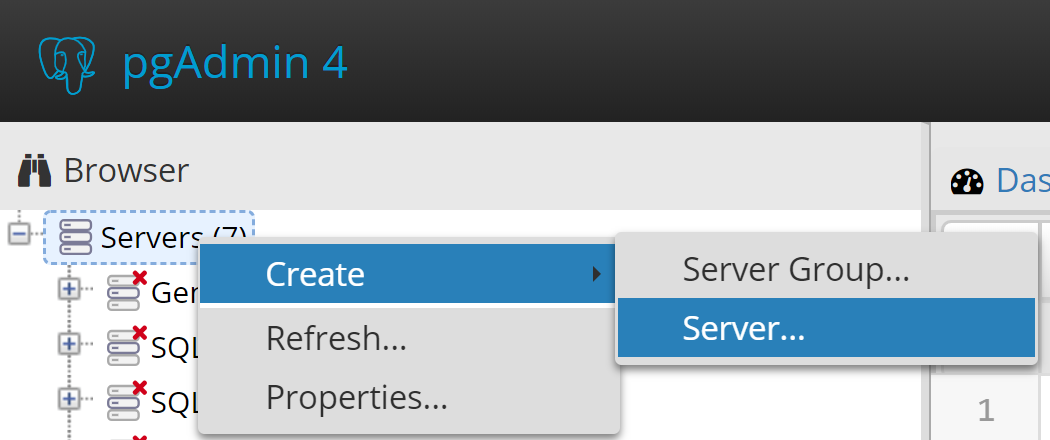
Takeaway: You should NEVER guess \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

1. **Connecting to an RDB (Red) – PostGreSQL Server Details**

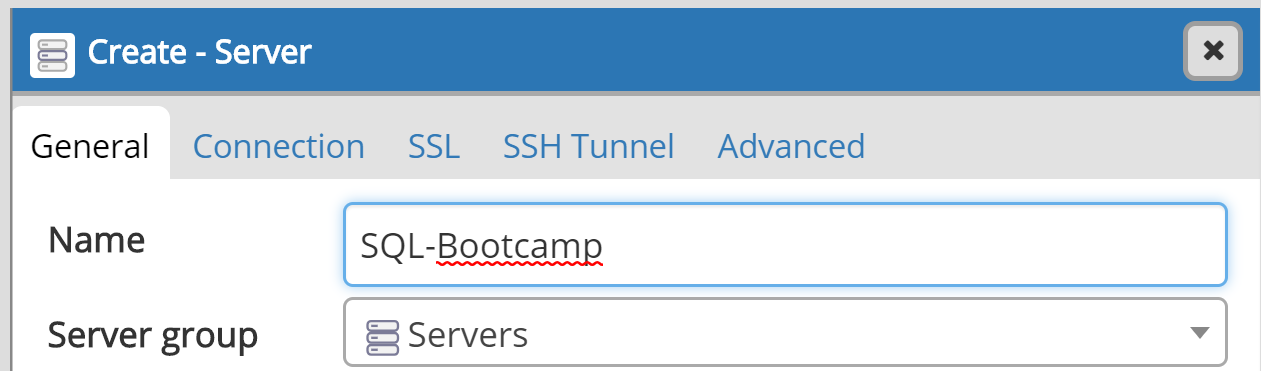
\*Make sure to DISCONNECT from VPN or Corporate Network\*

1. Open PG-Admin 4

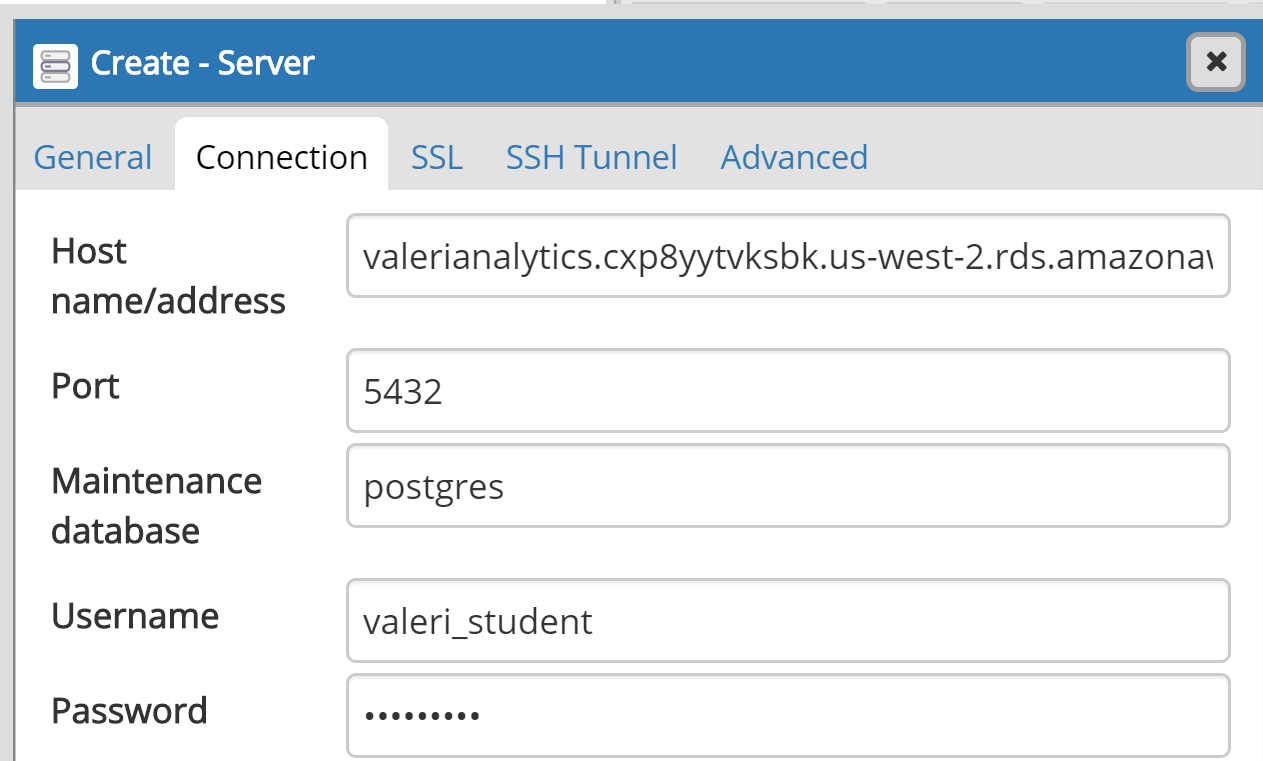
2. Right click “Servers” 🡪 “Create” 🡪 “Server”



3. “General” 🡪 Write “SQL-Bootcamp” in Name



4. “Connection” 🡪 Write below credentials in fields, then click “Save”



|  |  |
| --- | --- |
| *PRIMARY SERVER DETAILS*  Hostname/address: **valerianalytics.cxp8yytvksbk.us-west-2.rds.amazonaws.com**  Port: **5432**  Maintenance\_DB: **postgres**  Username: **valeri\_student**  Password: **shawshank** | *SECONDARY SERVER*  Hostname/address: **analyticsga.cuwj8wuu6wbh.us-west-2.rds.amazonaws.com**  Port: **5432**  Maintenance\_DB: **postgres**  Username: **analytics\_student**  Password: **analyticsga** |

1. **“Hey, Nice to Meet you Red…” – The Basic Syntax of SQL**

|  |  |  |
| --- | --- | --- |
| **SEQ.** | **KEYWORD** | **WHAT IT DOES** |
| 1 | SELECT |  |
| 2 | FROM |  |
| 3 | WHERE |  |
| 4 | GROUP BY |  |
| 5 | HAVING |  |
| 5 | ORDER BY |  |
| 6 | LIMIT |  |

Takeaways: SQL keywords are case \_\_\_\_\_\_\_\_\_\_\_; but the \_\_\_\_\_\_\_\_ matters.

1. **Asking Questions in SQL – SELECT, FROM, ORDER BY and LIMIT Keywords**

*Instructor Guided*

|  |  |
| --- | --- |
| **Question (English)** | **Question (SQL)** |
| I want to see ALL records in the *LAUNCHES* table… |  |
| I want to see a SAMPLE of 10 records in the *LAUNCHES* table… |  |
| I want to see a list of Merchant\_IDs in the LAUNCHES table… |  |
| I want to see a list of Merchant\_IDs in the LAUNCHES table, sorted from A-Z… |  |
| I want to see a list of Merchant IDs and Merchant Names in the LAUNCHES table, sorted by Name… |  |
| I want to see a DISTINCT list of Merchant\_IDs and Merchant Names in the Launches Table, sorted by Name |  |
| I want to see FBA Offers by Merchant Name by Day |  |
| I want to see Total Offers by Merchant Name by Day… |  |
| I want to see a list of Sellers ranked by the Largest Launch GMS by Day… |  |

*Independent Practice*

|  |  |
| --- | --- |
| **Question (English)** | **Question (SQL)** |
| *What Account Managers are in the Launches table sorted from A-Z?* |  |
| I want to see each Merchant’s share of GMS by Day and Marketplace (Merchant’s share = 30% of GMS) |  |
| I want to see a list of Account Mangers ranked by Smallest Total Offer Count by Merchant by Day |  |

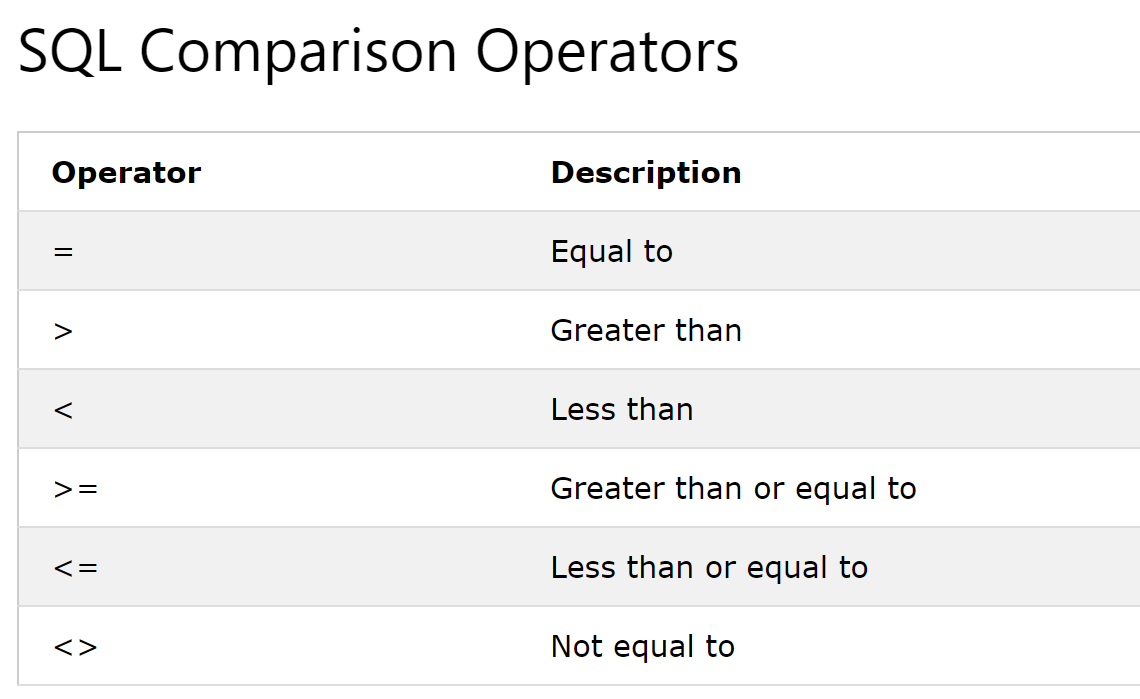
1. **Filtering data in SQL – WHERE keyword**

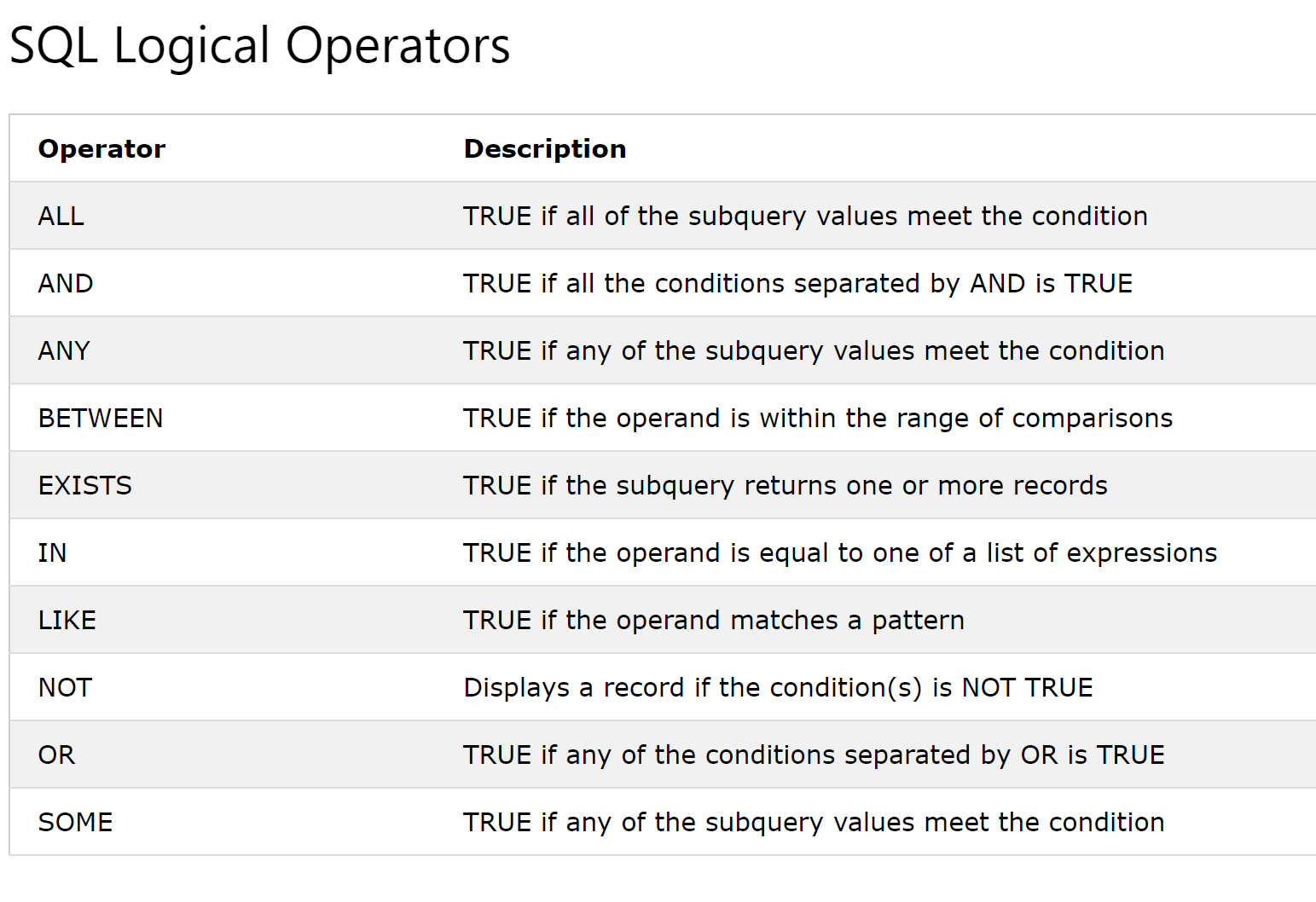
*Instructor Guided*

|  |  |
| --- | --- |
| **Question (English)** | **Question (SQL)** |
| Text Example: I want to see a list of Merchant’s launched by ‘Post Bologona’. |  |
| Integer Example: I want to see a list of Merchant’s launched in marketplace 6. |  |
| Multiple Conditions, 1 Column Example: I want to see a list of Merchant’s launched in marketplace 6 and 3240. |  |
| Multiple Conditions, Multiple Columns Example: I want to see a list of Merchant’s launched in marketplace 6 and 3240, that are FBA enrolled. |  |
| I want to see what Merchant’s launched between 10/30/18 and 11/1/18. |  |
| I want to see what Merchant’s launched in November 2018. |  |
| I want to see what Merchant’s names include *bag* in their name. |  |

*Independent Practice*

|  |  |
| --- | --- |
| **Question (English)** | **Question (SQL)** |
| Which Softline merchant(s) launched in Marketplace 3? |  |
| What Merchant(s)  launched in Oct-18 that have an MFN offer count over 400 or are in 3240 marketplace. |  |
| List all the merchant’s that have \_\_\_\_\_\_\_\_\_ in their name. |  |

Takeaways: There are many WHERE operators, we did not cover them all… (Look at Handout Sheet for comprehensive list!)



1. **Aggregating Data in SQL – SUM, COUNT, MIN, MAX, and AVG functions**

|  |  |
| --- | --- |
| **FUNCTION** | **WHAT DOES IT DO?** |
| SUM(COLUMN\_X) | Sums the values in entire column |
| COUNT(COLUMN\_X) | Counts the number of records in column (that aren’t null) |
| COUNT(DISTINCT COLUMN\_X) | Counts the number of distinct records in column (that aren’t null) |
| MIN(COLUMN\_X) / MAX(COLUMN\_X) | Finds min/max value in column |

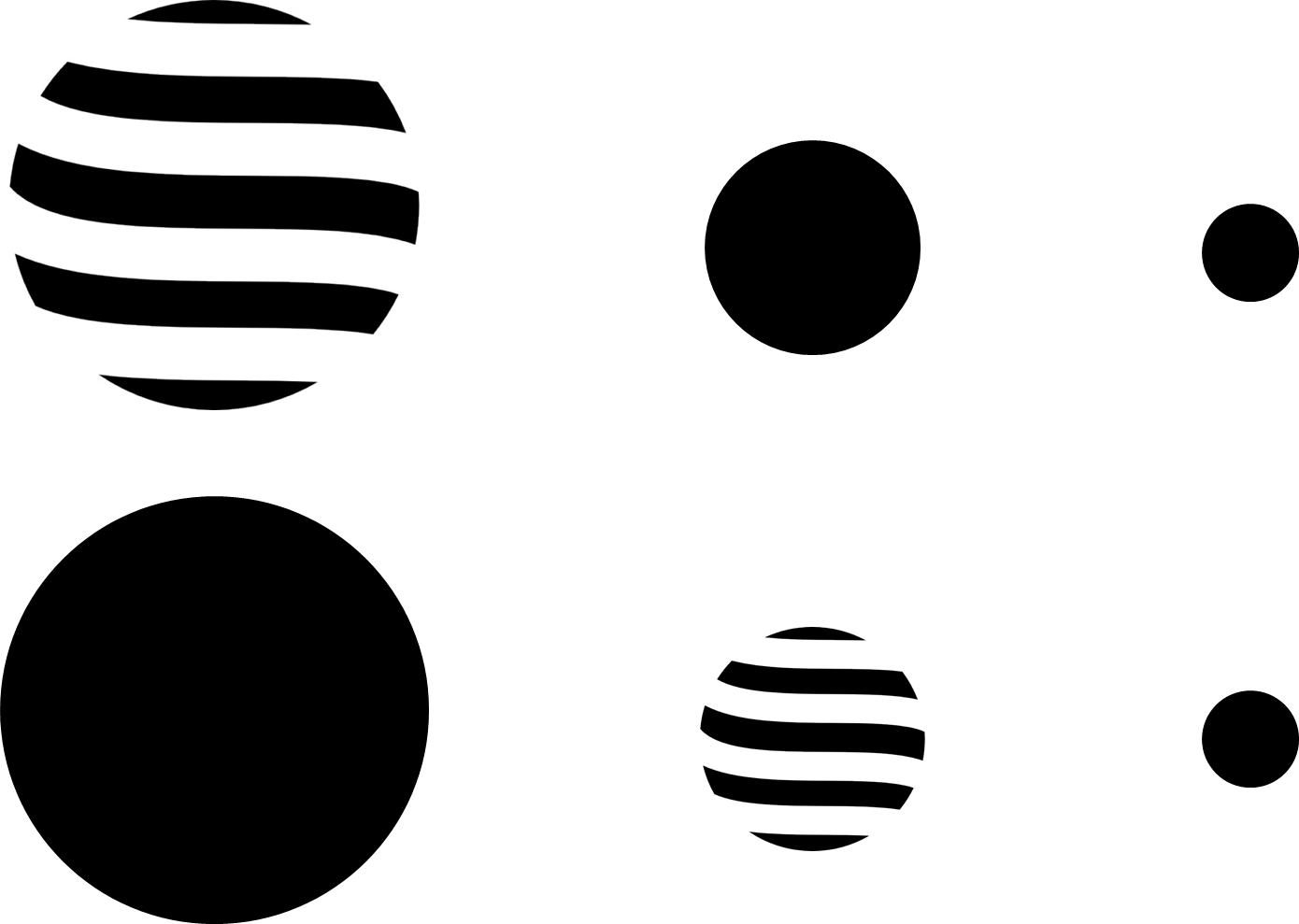
*Instructor Guided*

|  |  |
| --- | --- |
| **Question (English)** | **Question (SQL)** |
| What is the SUM of the Total Launch GMS in the table? |  |
| What is the COUNT of Total Merchants in the Table? |  |

Takeaways: Using an aggregate without a \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ is not helpful.

1. **Slicing Aggregates by Dimensions – Using GROUP BY in SQL**

How many rocks are in the box? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

****

*Rock Count by Size*

|  |  |
| --- | --- |
| **Rock Size** | **Count** |
| Big |  |
| Medium |  |
| Small |  |

*Rock Count by Stripes (T / F)*

|  |  |
| --- | --- |
| **IS\_STRIPED** | **Count** |
| TRUE |  |
| FALSE |  |

*Rock Count by Stripes and Size*

|  |  |  |
| --- | --- | --- |
| **Rock Size** | **IS\_STRIPED** | **Count** |
| Big | Y |  |
| Big | N |  |
| Medium | Y |  |
| Medium | N |  |
| Small | N |  |

*ROCKS\_IN\_THE\_BOX (Table Schema)*

|  |  |
| --- | --- |
| **Field/Column** | **Data Type** |
| rock\_id | text |
| is\_striped | boolean |
| rock\_size | text |

*Instructor Guided: Write an SQL statement that counts the number of rocks by Size and Stripe*

**SELECT**

**FROM** *ROCKS\_IN\_THE\_BOX*

**GROUP BY**

Takeaways: You use a GROUP BY when you want to understand an aggregate by \_\_\_\_\_\_\_\_\_\_\_\_\_\_. The GROUP BY and SELECT statement are identical; except you don’t include \_\_\_\_\_\_\_\_\_\_\_\_ or \_\_\_\_\_\_\_\_\_\_\_\_\_\_ in the GROUP BY.

*Independent Practice: Write a SQL statement that counts the number of merchants launched and GMS by day.*

*How could we further break this down? Modify the above code.*

1. **Relating Multiple Tables – Using JOINs in SQL**

*“I want to look at GMS for sellers by Marketplace country”*

*“I want to count the number of sellers Cardi Okra launched in Germany”*

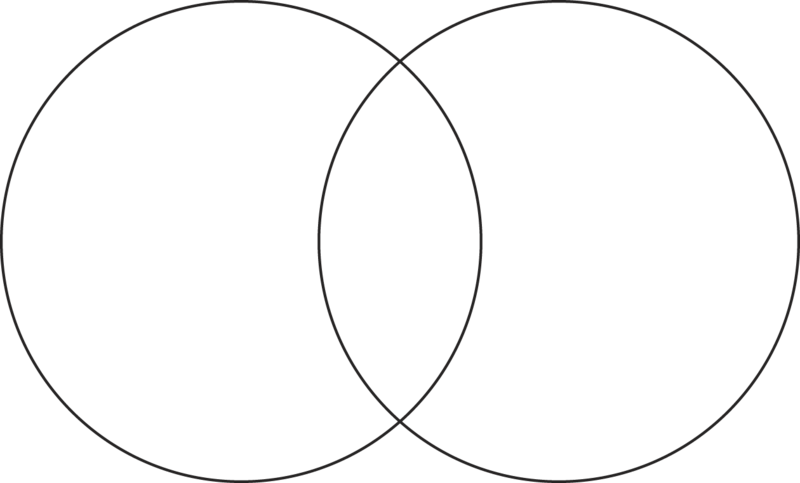
1. Why do we use a JOIN?

2. What is required to JOIN two tables?

3. What is a JOIN similar too?

4. How do we visualize a join

|  |  |
| --- | --- |
| **TABLE A:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** | **TABLE B:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** |
| Primary Table (most important) | Secondary Table (nice to have) |



*Write a query that JOINs the LAUNCHES* table to the *MARKETPLACE\_DIM* table.

Draw the difference between a LEFT JOIN and a JOIN