Welcome to Intro to SQL & Databases

(Getting Started Guide)

Go to github.com/mvaleri12

Click on 'valeri-analytics-sql-bootcamp'

Click on 'Clone or download' then 'Download Zip'

Open zip file in 'Downloads' folder, then drag folder onto Desktop

Follow 'Read-Me' instructions at bottom of the page

amazon

T-Mobile

GoDaddy™

Expedia®

BIG FISH™

Mission

To help individuals of all backgrounds to become data literate.

And many more

What is SQL?

Code

ASK A QUESTION

Relational Database

Table

GET AN ANSWER

Related Tables

Sales *Products* *Stores*

Why do we have to use SQL?

COLLECT BILLIONS AND BILLIONS OF ROWS OF DATA….a day.

Excel Workbook
ROW Limit = 1 Million

CED .5 OZ	Ben E Keith	2 - FROZEN FOOD	0	0	0.00	$	-	0.00	$	-	0.00	$	-	0.00	$ -
CAESAR CREAMY	Ben E Keith	4 - GROCERY	0	0	0.00	$	-	0.00	$	-	0.00	$	-	0.00	$ -
E LIQUID OLEO	Ben E Keith	4 - GROCERY	0	0	0.00	$	-	0.00	$	-	0.00	$	-	0.00	$ -
UFFLE CLEAR	Ben E Keith	4 - GROCERY	0	0	0.00	$	-	0.00	$	-	0.00	$	-	0.00	$ -
SL SLOTTED	Ben E Keith	4 - GROCERY	0	0	0.00	$	-	0.00	$	-	0.00	$	-	0.00	$ -
16OZ 16/16	Ben E Keith	4 - GROCERY	0	0	0.00	$	-	0.00	$	-	0.00	$	-	0.00	$ -
	Ben E Keith	4 - GROCERY	0	0	0.00	$	-	0.00	$	-	1.00	$	5.79	0.00	$ -
epared	Ben E Keith	4 - GROCERY	512	fl oz	0.00	$	-	1.00	$	3.75	0.00	$	-	0.00	$ -
MESAN SHRED	Ben E Keith	4 - GROCERY	0	0	0.00	$	-	0.00	$	-	0.00	$	-	1.00	$ 13.
SHELL MED USDA AA	Ben E Keith	1 - PRODUCE	0	0	0.00	$	-	1.00	$	15.89	0.00	$	-	0.00	$ -
PPL CIDER 40GRAIN	Ben E Keith	4 - GROCERY	0	0	0.00	$	-	0.00	$	-	1.00	$	17.77	0.00	$ -
	Ben E Keith	1 - PRODUCE	12	ct	0.00	$	-	2.00	$	8.99	0.00	$	-	0.00	$ -
CED W/GREEN CHILES	Ben E Keith	4 - GROCERY	0	0	0.00	$	-	1.00	$	18.88	0.00	$	-	0.00	$ -
Vanilla Cr 3 Gal	Ben E Keith	6 - DAIRY	384	fl oz	0.00	$	-	0.00	$	-	0.00	$	-	0.00	$ -
ANCY 33% SOLIDS	Ben E Keith	4 - GROCERY	0	0	0.00	$	-	1.00	$	20.69	0.00	$	-	0.00	$ -
VI WHITE SMALL BUTTON	Ben E Keith	1 - PRODUCE	0	0	0.00	$	-	1.00	$	20.98	0.00	$	-	0.00	$ -
EASONED HOMESTYLE	Ben E Keith	4 - GROCERY	0	0	0.00	$	-	0.00	$	-	1.00	$	22.30	0.00	$ -
ISIANA RED HOT	Ben E Keith	4 - GROCERY	0	0	0.00	$	-	1.00	$	11.24	0.00	$	-	1.00	$ 11.
n Iceless W/Root	Ben E Keith	1 - PRODUCE	32	oz	0.00	$	-	1.00	$	8.29	1.00	$	8.29	0.00	$ -
WN LIGHT IN BAGS	Ben E Keith	4 - GROCERY	0	0	0.00	$	-	0.00	$	-	1.00	$	27.69	0.00	$ -
w Jumbo	Ben E Keith	1 - PRODUCE	800	oz	0.00	$	-	0.00	$	-	1.00	$	13.99	0.00	$ -

Database
ROW Limit = hold my beer

Where does SQL fit in the data flow?

What will I gain from learning SQL?

You are dangerous with SQL

Speak to analysts in their own language

Tell a client you have a great analytics foundation

Eliminate the analytics bottleneck

Today's Agenda

INTRODUCTION (10 min)	**10:00 – 10:10**
Part 1: Database Fundamentals (10 min)	**10:10 – 10:20**
Part 2: Datatype Fundamentals (10 min)	**10:20 – 10:30**
Part 3: Connecting to an RDB (15 min)	**10:30 – 10:45**
Part 4: The Basic Syntax of SQL	**11:00 – 11:30**
Part 5: Filtering with WHERE	**11:30 – 12:00**
Part 6: Aggregations and GROUP BY	**12:00 – 12:30**
LUNCH – (1 HOUR)	**12:30 – 01:30**
Part 7: Fix the SQL CODE	**01:30 – 02:15**
Part 8: Joining Tables	**02:15 – 03:15**
Part 9: Advanced SQL Concepts	**03:15 – 03:30**
Part 10: Conclusion and Next Steps	**03:30 – 03:45**
Q:A	**03:45 – 04:00**

Part 1

DATABASE FUNDEMENTALS

What is a relational database (RDB) and relational database management system (RDBMS)?

Relational Database Management System(s) – PGAdmin 4

Note: Different Dialects of SQL for different Relational Databases

Working with PostGreSQL today

SQL can…

- **SELECT data**
- **INSERT data**
- **CREATE data**
- **UPDATE data**
- **DELETE data**

Relational Database – PostGreSQL

Related Tables

Sales *Products* *Stores*

What are the other dialects of SQL databases?

PROS
Free
Scalable and can handle a lot of data

CONS
Spotty Documentation
Configuration can be confusing

PROS
Most Popular for organizations
Robust tools, cutting edge

CONS
Expensive, for small orgs
Significant resource maintenance

PROS
Free
Most popular for web-based apps
Can connect to Oracle

CONS
You need to pay for support

PROS
Fast and stable
Works well MS products

CONS
Expensive
Resource maintenance

How does an RDB consume raw data?

Random data sources

Raw disorganized data

Clean organized data

Will only consume what we tell it to consume

How is data organized in an RDB after it's consumed?

Unorganized Data

Schema

ALBUMS SCHEMA		
field/column	**data type**	**sample data**
row	integer	1
artist	text	Red Hot Chili Peppers
album	text	Californication
release_date	date	6/8/1999
genre	text	Rock
plays	integer	120
rating	numeric(10,2)	
org_price	money	
market_value	numeric(10,2)	
burned	boolean	
playable	boolean	

Organized Data in Tables

How is data organized in a table in a RDB?

ALBUMS SCHEMA

field/column	data type	sample data
row	integer	1
artist	text	Red Hot Chili Peppers
album	text	Californication
release_date	date	6/8/1999
genre	text	Rock
plays	integer	120
rating	numeric(10,2)	4
org_price	money	$11.99
market_value	numeric(10,2)	3.2
burned	boolean	FALSE
playable	boolean	TRUE

COLUMNS/FIELDS

ROWS/RECORDS

	row integer	artist text	album text	release_date date	genre text	plays integer	rating numeric (10,2)	org_price money	market_value numeric (10,2)	burned boolean	playable boolean
1	1	Red Hot Chili Peppers	Californication	1999-06-08	Rock	120	4.00	$11.99	3.20	false	true
2	2	Red Hot Chili Peppers	By the Way	2002-07-09	Rock	100	3.50	$11.99	4.00	true	false
3	3	Kanye West	College Dropout	2004-02-10	Rap	200	5.00	$10.99	5.00	false	true
4	4	Kanye West	Late Registration	2005-08-30	Rap	300	4.00	$9.99	7.00	false	true
5	5	Kanye West	Graduation	2007-09-11	Rap	250	4.00	$0.00	1.75	true	true
6	6	Papa Roach	Infest	2000-04-25	Rock	75	3.50	$11.99	0.50	false	false
7	7	Kid Cudi	Man on the Moon	2009-09-15	Rap	40	4.00	$10.99	6.00	false	false
8	8	Ratatat	Ratatat	2004-04-20	Electro...	60	5.00	$9.99	6.00	false	[null]
9	9	Ratatat	Classics	2006-08-22	Electro...	400	4.00	$0.00	12.99	[null]	false
10	10	Dragonforce	Sonic Firestorm	2004-05-11	Rock	500	5.00	$2.99	0.01	true	[null]
11	11	[null]	Summer Mix 08	[null]	[null]	1000	5.00	$0.00	0.00	true	true
12	12	[null]	Party Mix 07	[null]	[null]	4000	5.00	$0.00	0.00	true	true
13	13	Common	Be	2005-05-24	Rap	2000	4.50	$0.00	15.00	true	true
14	14	T.I.	Paper Trail	2008-09-26	Rap	300	4.00	$0.00	6.99	true	false
15	15	Children of Bodom	Children of Bodom Greate...	[null]	Metal	150	3.00	$0.00	0.00	[null]	false

How are the tables related to each other in an RDB?

valerianalytics database

ALBUMS SCHEMA

field/column	data type	sample data
row	integer	1
artist	text	Red Hot Chili Peppers
album	text	Californication
release_date	date	6/8/1999
genre	text	Rock
plays	integer	
rating	numeric(10,2)	
org_price	money	
market_value	numeric(10,2)	
burned	boolean	
playable	boolean	

NORTH_KOREAN_MISSLE_TESTS SCHEMA

field/column	data type	sample data
row_id	integer	5
date_launched	date	9/1/1984
date_entered	date	12/23/2016
launch_time_utc	date	(null)
missle_name	varchar	Scud-B
missle_type	varchar	SRBM
launch_authority	varchar	(null)
facility_name	varchar	Tonghae Satellite
facility_location	varchar	Hwadae County
facility_other_name	varchar	Musudan-ri
facility_latitude	float	40.8499966
facility_longitude	float	129.666664
landing_location	varchar	(null)
apogee	varchar	
distance_traveled	varchar	
confirmation_status	varchar	
test_outcome	varchar	
additional_info	varchar	

PRODUCTS SCHEMA

field/column	data type	sample data
item_no	integer	904616
category_name	text	TEQUILA
item_description	text	Jose Cuervo
vendor	integer	305
vendor_name	text	Mhw Ltd
bottle_size	integer	750
pack	integer	12
inner_pack	integer	1
age	text	(null)
proof	text	40
list_date	timestamp	2/11/2009
bottle_price	money	$9.77
shelf_price	numeric(10,2)	14.66
case_cost	numeric(10,2)	117.22

COUNTIES SCHEMA

field/column	data type	sample data
county	text	Adair
population	integer	7682

STORES SCHEMA

field/column	data type	sample data
store	integer	2106
name	text	Hillstreet News and Tobacco
store_status	text	A
store_address	text	2217 CollegeCedar Falls, IA
address_info	text	(null)

SALES SCHEMA

field/column	data type	sample data
date	timestamp	6/27/2014
convenience_store	text	Y
store	integer	4771
county_number	county_number	57
county	text	Linn
category	text	1081600
category_name	text	WHISKEY LIQUEUR
vendor_no	text	421
vendor	text	Sazerac Co. Inc.
item	integer	64858
description	text	Fireball Cinnamon Whiskey
pack	integer	1
liter_size	integer	3000
state_btl_cost	money	29.72
btl_price	money	44.58
bottle_qty	integer	1
total	numeric(10,2)	44.58

Database vs. Microsoft Excel

Row	Artist	Album	Release_Date	Genre	Plays	Rating	Org_Price	Mark
1	Red Hot Chili Peppers	Californication	6/8/1999	Rock	120	4	$11.99	
2	Red Hot Chili Peppers	By the Way	7/9/2002	Rock	100	3.5	$11.99	
3	Kanye West	College Dropout	2/10/2004	Rap	200	5	$10.99	
4	Kanye West	Late Registration	8/30/2005	Rap	300	4	$9.99	
5	Kanye West	Graduation	9/11/2007	Rap	250	4	$0.00	
6	Papa Roach	Infest	4/25/2000	Rock	75	3.5	$11.99	
7	Kid Cudi	Man on the Moon	9/15/2009	Rap	40	4	$10.99	
8	Ratatat	Ratatat	4/20/2004	Electronic	60	5	$9.99	
9	Ratatat	Classics	8/22/2006	Electronic	400	4	$0.00	
10	Dragonforce	Sonic Firestorm	5/11/2004	Rock	500	5	$2.99	
11		Summer Mix 08			1000	5	$0.00	
12		Party Mix 07			4000	5	$0.00	
13	Common	Be	5/24/2005	Rap	2000	4.5	$0.00	
14	T.I.	Paper Trail	9/26/2008	Rap	300	4	$0.00	
15	Children of Bodom	Children of Bodom Greatest Hits		Metal	150	3	$0.00	

Activity 1: Draw an entity relationship diagram between all the tables

SALES TABLE SCHEMA

field/column
date
convenience_store
store
county_number
county
category
category_name
vendor_no
vendor
item
description
pack
liter_size
state_btl_cost
btl_price
bottle_qty
total

STORES TABLE SCHEMA

field/column
store
name
store_status
store_address
address_info

COUNTIES TABLE SCHE

field/column
county
population

PRODUCTS TABLE SCHE

field/column
item_no
category_name
item_description
vendor
vendor_name
bottle_size
pack
inner_pack
age
proof
list_date
bottle_price
shelf_price
case_cost

Activity 1: Draw an entity relationship diagram between all the tables

SALES TABLE SCHEMA

field/column
date
convenience_store
store
county_number
county
category
category_name
vendor_no
vendor
item
description
pack
liter_size
state_btl_cost
btl_price
bottle_qty
total

STORES TABLE SCHEMA

field/column
store
name
store_status
store_address
address_info

COUNTIES TABLE SCHE

field/column
county
population

PRODUCTS TABLE SCHE

field/column
item_no
category_name
item_description
vendor
vendor_name
bottle_size
pack
inner_pack
age
proof
list_date
bottle_price
shelf_price
case_cost

Database Fundamentals: Key Takeaways

1. Not all tables in a database are related to each other
2. Investigate the column names to see which tables are related to what
3. Not all column/field names are a 1:1 match

Part 2

DATATYPE FUNDEMENTALS

What data types are supported in SQL tables?

DATA TYPE	BOOLEAN	CHARACTER	NUMERIC	TEMPORAL
DATA Example	TRUE / FALSE / (NULL) Or 1 / 0 / (NULL)	ABC / ABC123 / ABC 123 Or (NULL)	4 / 5.1 Or (NULL)	7/4/1776 Or 7/4/1776 23:00:12 Or 23:00:12 OR (NULL)
Field/Column Example	PLAYABLE	ARTIST	RATING	RELEASE DATE

Activity 2: Guess the data type

BOOLEAN	CHARACTER	NUMERIC	TEMPORAL
A	B	C	D

Activity 2: Guess the data type

BOOLEAN	CHARACTER	NUMERIC	TEMPORAL
A	B	C	D

FALSE

Activity 2: Guess the data type

BOOLEAN	CHARACTER	NUMERIC	TEMPORAL
A	B	C	D

FALSE

Activity 2: Guess the data type

BOOLEAN	CHARACTER	NUMERIC	TEMPORAL
A	B	C	D

311

Activity 2: Guess the data type

BOOLEAN	CHARACTER	NUMERIC	TEMPORAL
A	B	C	D

311

Activity 2: Guess the data type

BOOLEAN	CHARACTER	NUMERIC	TEMPORAL
A	B	C	D

4/29/1992

Activity 2: Guess the data type

BOOLEAN	CHARACTER	NUMERIC	TEMPORAL
A	B	C	D

4/29/1992

Datatype Fundamentals: Key Takeaways

1. Appearances can be deceiving – Find the table schema first, don’t guess data types.
2. Data type has influence on how we write and format our SQL code.
3. We can write SQL code to convert datatypes too…

Part 3

CONNECTING TO AN RDB

How do you connect to an RDB?

THIS CLASS

Handout 2: Connecting to 'valerianalytics' database

YOUR ORGANIZATION

ORACLE
Microsoft SQL Server
PostgreSQL
MySQL
SYBASE
IBM DB2.

Who can help me connect to an RDB at my organization?

YOUR ORGANIZATION

1. Find the Database Administrator (DBA) at your organization. Ask them for credentials

2. If you can't find them, find someone who publishes reports. They will probably know.

3. Remember, you're probably only looking for SELECT privileges in the RDB.

4. Don't GIVE UP, hardest step is just getting access.

How to connect to valerianalytics RDB

1. Download PGAdmin 4 (RDBMS)

https://www.pgadmin.org/download/

2. Create a server

3. Input credentials details*

(found in Handout 2: Connecting to 'valerianalytics' database)

Part 4

THE BASIC SYNTAX OF SQL

What is the basic syntax of SQL?

- ✔ **SELECT** Fields/columns you'd like to see in result-set
- ✔ **FROM** Table(s) where the fields/columns are located
- **WHERE** Rows/records you'd like to include/exclude in result-set
- **GROUP BY** Fields/columns you want your aggregate dimensions grouped by
- **HAVING** Include/exclude rows/records where aggregate values meet a condition
- **ORDER BY** Sort the rows/records of result-set by ASC/DESC for a column(s)
- **LIMIT** Limit the number of rows/records to include in result-set.

Guided Workshop: Basic SQL Syntax

	row integer	artist text	album text	release_date date	genre text	plays integer	rating numeric (10,2)	org_price money	market_value numeric (10,2)	burned boolean	playable boolean
1	1	Red Hot Chili Peppers	Californication	1999-06-08	Rock	120	4.00	$11.99	3.20	false	true
2	2	Red Hot Chili Peppers	By the Way	2002-07-09	Rock	100	3.50	$11.99	4.00	true	false
3	3	Kanye West	College Dropout	2004-02-10	Rap	200	5.00	$10.99	5.00	false	true
4	4	Kanye West	Late Registration	2005-08-30	Rap	300	4.00	$9.99	7.00	false	true
5	5	Kanye West	Graduation	2007-09-11	Rap	250	4.00	$0.00	1.75	true	true
6	6	Papa Roach	Infest	2000-04-25	Rock	75	3.50	$11.99	0.50	false	false
7	7	Kid Cudi	Man on the Moon	2009-09-15	Rap	40	4.00	$10.99	6.00	false	false
8	8	Ratatat	Ratatat	2004-04-20	Electro...	60	5.00	$9.99	6.00	false	[null]
9	9	Ratatat	Classics	2006-08-22	Electro...	400	4.00	$0.00	12.99	[null]	false
10	10	Dragonforce	Sonic Firestorm	2004-05-11	Rock	500	5.00	$2.99	0.01	true	[null]
11	11	[null]	Summer Mix 08	[null]	[null]	1000	5.00	$0.00	0.00	true	true
12	12	[null]	Party Mix 07	[null]	[null]	4000	5.00	$0.00	0.00	true	true
13	13	Common	Be	2005-05-24	Rap	2000	4.50	$0.00	15.00	true	true
14	14	T.I.	Paper Trail	2008-09-26	Rap	300	4.00	$0.00	6.99	true	false
15	15	Children of Bodom	Children of Bodom Greate...	[null]	Metal	150	3.00	$0.00	0.00	[null]	false

ASK A QUESTION

GET AN ANSWER

BASIC SQL Syntax: Key Takeaways

1. Any line, any case; SQL can be written on multiple lines or single line in any case for keywords.

2. Comma chameleon; Watch out for extra or missing commas in SELECT statement. Most common error in SQL

3. If you're having trouble combining multiple/columns fields, you might have to use CAST() to convert data types.

Part 5

FILTERING WITH WHERE

Filtering with WHERE: Outside SQL

Independent Workshop: Filtering with WHERE

	row integer	artist text	album text	release_date date	genre text	plays integer	rating numeric (10,2)	org_price money	market_value numeric (10,2)	burned boolean	playable boolean
1	1	Red Hot Chili Peppers	Californication	1999-06-08	Rock	120	4.00	$11.99	3.20	false	true
2	2	Red Hot Chili Peppers	By the Way	2002-07-09	Rock	100	3.50	$11.99	4.00	true	false
3	3	Kanye West	College Dropout	2004-02-10	Rap	200	5.00	$10.99	5.00	false	true
4	4	Kanye West	Late Registration	2005-08-30	Rap	300	4.00	$9.99	7.00	false	true
5	5	Kanye West	Graduation	2007-09-11	Rap	250	4.00	$0.00	1.75	true	true
6	6	Papa Roach	Infest	2000-04-25	Rock	75	3.50	$11.99	0.50	false	false
7	7	Kid Cudi	Man on the Moon	2009-09-15	Rap	40	4.00	$10.99	6.00	false	false
8	8	Ratatat	Ratatat	2004-04-20	Electro...	60	5.00	$9.99	6.00	false	[null]
9	9	Ratatat	Classics	2006-08-22	Electro...	400	4.00	$0.00	12.99	[null]	false
10	10	Dragonforce	Sonic Firestorm	2004-05-11	Rock	500	5.00	$2.99	0.01	true	[null]
11	11	[null]	Summer Mix 08	[null]	[null]	1000	5.00	$0.00	0.00	true	true
12	12	[null]	Party Mix 07	[null]	[null]	4000	5.00	$0.00	0.00	true	true
13	13	Common	Be	2005-05-24	Rap	2000	4.50	$0.00	15.00	true	true
14	14	T.I.	Paper Trail	2008-09-26	Rap	300	4.00	$0.00	6.99	true	false
15	15	Children of Bodom	Children of Bodom Greate...	[null]	Metal	150	3.00	$0.00	0.00	[null]	false

ASK A QUESTION

GET AN ANSWER

Filtering with WHERE: Key Takeaways

1. Many nuances; Syntax for filtering depends on the column data type you are working with.

2. You don’t need to SELECT columns you are filtering on in the WHERE clause

3. Use TO_CHAR() function to simplify dates when filtering. Use LIKE with ‘%’ to look for patterns.

Part 6

AGGREGATIONS AND GROUP BY

Aggregations and Group BY: An aggregation without a Group BY

9,495 plays

=

Pie Chart A: Total Plays

Legend

TOTAL

Aggregations and Group BY: What can we group our aggregations by?

9,495 plays

By ARTIST

By GENRE

By ANY COLUMN(s)

Aggregations and Group BY: Like a Pivot Table in Excel

By ARTIST

Row Labels	Sum of Plays
(blank)	5000
Common	2000
Kanye West	750
Dragonforce	500
Ratatat	460
T.I.	300
Red Hot Chili Peppers	220
Children of Bodom	150
Papa Roach	75
Kid Cudi	40
Grand Total	**9495**

By GENRE

Row Labels	Sum of Plays
Electronic	460
Metal	150
Rap	3090
Rock	795
(blank)	5000
Grand Total	**9495**

Guided Workshop: Aggregations and GROUP BY

	row integer	artist text	album text	release_date date	genre text	plays integer	rating numeric (10,2)	org_price money	market_value numeric (10,2)	burned boolean	playable boolean
1	1	Red Hot Chili Peppers	Californication	1999-06-08	Rock	120	4.00	$11.99	3.20	false	true
2	2	Red Hot Chili Peppers	By the Way	2002-07-09	Rock	100	3.50	$11.99	4.00	true	false
3	3	Kanye West	College Dropout	2004-02-10	Rap	200	5.00	$10.99	5.00	false	true
4	4	Kanye West	Late Registration	2005-08-30	Rap	300	4.00	$9.99	7.00	false	true
5	5	Kanye West	Graduation	2007-09-11	Rap	250	4.00	$0.00	1.75	true	true
6	6	Papa Roach	Infest	2000-04-25	Rock	75	3.50	$11.99	0.50	false	false
7	7	Kid Cudi	Man on the Moon	2009-09-15	Rap	40	4.00	$10.99	6.00	false	false
8	8	Ratatat	Ratatat	2004-04-20	Electro...	60	5.00	$9.99	6.00	false	[null]
9	9	Ratatat	Classics	2006-08-22	Electro...	400	4.00	$0.00	12.99	[null]	false
10	10	Dragonforce	Sonic Firestorm	2004-05-11	Rock	500	5.00	$2.99	0.01	true	[null]
11	11	[null]	Summer Mix 08	[null]	[null]	1000	5.00	$0.00	0.00	true	true
12	12	[null]	Party Mix 07	[null]	[null]	4000	5.00	$0.00	0.00	true	true
13	13	Common	Be	2005-05-24	Rap	2000	4.50	$0.00	15.00	true	true
14	14	T.I.	Paper Trail	2008-09-26	Rap	300	4.00	$0.00	6.99	true	false
15	15	Children of Bodom	Children of Bodom Greate...	[null]	Metal	150	3.00	$0.00	0.00	[null]	false

ASK A QUESTION

GET AN ANSWER

Aggregations and GROUP BY: Key Takeaways

1. Use GROUP BY when you want to understand aggregates (SUM/ COUNT/ etc.) by specific fields/columns. *I.e. Total sales by Country/Channel.*

2. Aggregate functions and aliases, **DO NOT** go in the GROUP BY.

3. You **CAN NOT** filter on aggregate functions in the WHERE clause, can only filter in HAVING clause.

4. You **CAN NOT** perform on aggregate on an aggregate (i.e. Avg(Sum(Plays)) by Genre, need to perform in two steps (more on this later).

LUNCH

(1 Hour)

Return at 1:30 PM

Part 7

FIX THE CODE

GROUP Workshop 4: Fix the Code

	row integer	artist text	album text	release_date date	genre text	plays integer	rating numeric (10,2)	org_price money	market_value numeric (10,2)	burned boolean	playable boolean
1	1	Red Hot Chili Peppers	Californication	1999-06-08	Rock	120	4.00	$11.99	3.20	false	true
2	2	Red Hot Chili Peppers	By the Way	2002-07-09	Rock	100	3.50	$11.99	4.00	true	false
3	3	Kanye West	College Dropout	2004-02-10	Rap	200	5.00	$10.99	5.00	false	true
4	4	Kanye West	Late Registration	2005-08-30	Rap	300	4.00	$9.99	7.00	false	true
5	5	Kanye West	Graduation	2007-09-11	Rap	250	4.00	$0.00	1.75	true	true
6	6	Papa Roach	Infest	2000-04-25	Rock	75	3.50	$11.99	0.50	false	false
7	7	Kid Cudi	Man on the Moon	2009-09-15	Rap	40	4.00	$10.99	6.00	false	false
8	8	Ratatat	Ratatat	2004-04-20	Electro...	60	5.00	$9.99	6.00	false	[null]
9	9	Ratatat	Classics	2006-08-22	Electro...	400	4.00	$0.00	12.99	[null]	false
10	10	Dragonforce	Sonic Firestorm	2004-05-11	Rock	500	5.00	$2.99	0.01	true	[null]
11	11	[null]	Summer Mix 08	[null]	[null]	1000	5.00	$0.00	0.00	true	true
12	12	[null]	Party Mix 07	[null]	[null]	4000	5.00	$0.00	0.00	true	true
13	13	Common	Be	2005-05-24	Rap	2000	4.50	$0.00	15.00	true	true
14	14	T.I.	Paper Trail	2008-09-26	Rap	300	4.00	$0.00	6.99	true	false
15	15	Children of Bodom	Children of Bodom Greate...	[null]	Metal	150	3.00	$0.00	0.00	[null]	false

ASK A QUESTION

GET AN ANSWER

Fix the Code: Key Takeaways

1. **Work backwards;** draw the result-output first

2. **Remove and add back;** Eliminate what doesn't work in the code by commenting out lines with '--' and '/* */'

3. **Ask for help;** collaborate with others

Part 8

JOINING TABLES

What is a JOIN?

Table A - Orders	
OrderID	CustomerID
12345	1
12346	2
12344	8
12355	NULL

Table B - Customers	
CustomerID	CustomerName
1	Chuckie. F
2	Tommy. P
3	Angelica P.
4	Dr. Liptshitz

Why use a JOIN? – We have a question we can't answer from a single table.

What is a JOIN?

Table A - Orders	
OrderID	CustomerID
12345	1
12346	2
12344	8
12355	NULL

LEFT/PRIMARY TABLE

+

Table B - Customers	
CustomerID	CustomerName
1	Chuckie. F
2	Tommy. P
3	Angelica P.
4	Dr. Liptshitz

RIGHT/SECONDARY TABLE

=

JOINED TABLE		
OrderID	CustomerID	CustomerName
BLAH	BLAH	BLAH
BLAH	BLAH	BLAH
BLAH	BLAH	BLAH

Same thing as a VLOOKUP in Excel

What is required to join tables in SQL?

Table A - Orders	
OrderID	CustomerID
12345	1
12346	2
12344	8
12355	NULL

Table B - Customers	
CustomerID	CustomerName
1	Chuckie. F
2	Tommy. P
3	Angelica P.
4	Dr. Liptshitz

1. At least 2 tables, with at least 1 column/field in common
2. Column/field(s) tables have in common, **MUST** be same data type
3. The column/field in Table B ***should*** be unique, otherwise duplication will occur *(more on this later)*

What is the JOIN syntax?

```
SELECT
A.ORDERID
,A.CUSTOMERID
,B.CUSTOMERNAME

FROM
ORDERS A

INNER JOIN CUSTOMERS B

ON  B.CUSTOMERID = A.CUSTOMERID

GROUP BY
A.ORDERID
,A.CUSTOMERID
,B.CUSTOMERNAME
```

1.**ASSIGN AN ALIAS** to each table and apply it to each distinct column.

2. **JOIN ON the columns the TWO tables have in common** (make sure same data type)

3. You need indicate what **STYLE of JOIN** *(more than 1 style)*

Table A - Orders

OrderID	CustomerID
12345	1
12346	2
12344	8
12355	NULL

Table B - Customers

CustomerID	CustomerName
1	Chuckie. F
2	Tommy. P
3	Angelica P.
4	Dr. Liptshitz

INNER JOIN RESULT SET (CustomerID)

OrderID	CustomerID	CustomerName
12345	1	Chuckie. F
12346	2	Tommy. P

How do we visualize an INNER JOIN?

```
SELECT <select_list>
FROM TableA A
INNER JOIN TableB B
ON A.Key = B.Key
```

JOIN → EVERYTHING TABLE A and B HAVE IN COMMON

Table A - Orders			Table B - Customers	
OrderID	CustomerID		CustomerID	CustomerName
12345	1		1	Chuckie. F
12346	2		2	Tommy. P
12344	8		3	Angelica P.
12355	NULL		4	Dr. Liptshitz

INNER JOIN RESULT SET (CustomerID)

OrderID	CustomerID	CustomerName
12345	1	Chuckie. F
12346	2	Tommy. P

How do we visualize a LEFT JOIN?

```
SELECT <select_list>
FROM TableA A
LEFT JOIN TableB B
ON A.Key = B.Key
```

LEFT JOIN → EVERYTHING THAT's in Table A regardless of whether it's in Table B

Table A - Orders			Table B - Customers	
OrderID	CustomerID		CustomerID	CustomerName
12345	1		1	Chuckie. F
12346	2		2	Tommy. P
12344	8		3	Angelica P.
12355	NULL		4	Dr. Liptshitz

LEFT JOIN RESULT SET		
OrderID	CustomerID	CustomerName
12345	1	Chuckie. F
12346	2	Tommy. P
12344	8	NULL
12355	NULL	NULL

What are the other types of JOINs?

Group Activity: Which JOIN do I use? (20 min)

Table A: ORDERS	
ORDER_ID	CUST_ID
12345	567
12346	457
12366	123
14567	(null)
19453	(null)
15678	234

Table B: CUSTOMERS	
CUST_ID	State
567	WA
367	SD
678	DE
457	MI
234	(null)

GROUP 1

GROUP 2

GROUP 3

GROUP 4

GROUP 5

Question: What is the count of customers and orders by state?

Step 1: Make a joined table of order_id, cust_id, and state. Draw circles to help you.

Step 2: Make a table of the count of orders and customers in each state. Remember, don't count NULL customer or order IDS.

Step 3: Tell us the total order and customer count overall.

Which JOIN do I use? (Most Common Use Cases)

"Give me a list of total orders/customers by customer's state."

INNER JOIN

GROUP 1

WILL EXCLUDE ALL ORDERS FOR CUSTOMERS THAT AREN'T IN TABLE B

Consequence: Order count underreported

LEFT JOIN

GROUP 2

WILL INCLUDE ORDERS FOR ALL CUSTOMERS, REGARDLESS OF EXISTENCE IN TABLE B.

Consequence: Some orders won't have a state assigned to them (will appear as NULL for State)

FULL OUTER JOIN

WILL INCLUDE ALL ORDERS AND CUSTOMERS FROM BOTH TABLES

Consequence: Some order's won't have a state assigned to them, and some customer's won't have an order assigned to them.

Which JOIN do I use? (Fringe Use Cases)

"Give me a list of total orders/customers by customer's state."

LEFT JOIN, WHERE CUST_ID IS NULL in TABLE B

USE CASE: You only want to see orders for customers that aren't in Table B.

Tell DBA to update Customer Attributes; Table B

FULL OUTER JOIN, WHERE CUST_ID IS NULL in TABLE A and B

USE CASE: You only want to see orders from customers that don't exist in Table B

AND

You only want to see customers that don't exist in Table A (don't have an order)

Can you join to more than 1 table?

```
SELECT c.field, a.field,
b.field, a.field,c.field
FROM table1 a
JOIN table2 b
ON a.field=b.field
JOIN table3 c
ON a.field=c.field
```

What if you want to join two tables that don't share the same column?

TABLE A: ORDERS

ORDER_ID	CUST_ID
1234	1

TABLE B: ACCOUNT_STATES

ACCOUNT_ID	STATE
ABC	WA

TABLE C: ACCOUNT_MAPPING

CUST_ID	ACCOUNT_ID
1	ABC

```sql
SELECT

 A.ORDER_ID
,A.CUST_ID
,B.STATE

FROM

ORDERS A

LEFT JOIN ACCOUNT_MAPPING C ON A.CUST_ID = C.CUST_ID

LEFT JOIN ACCOUNT_STATES B ON B.ACCOUNT_ID = C.ACCOUNT_ID
```

SOLUTION: DAISY CHAINING

What if you are getting duplicate records or rows after performing a join?

Table A: ORDERS

ORDER_ID	CUST_ID
12345	567
12346	457
12366	123
14567	(null)
19453	(null)
15678	234

Joining on CUST_ID

Table B: CUSTOMERS

CUST_ID	STATE	UPDATE_DATE
567	WA	1/2/2017
367	SD	1/2/2017
678	DE	1/2/2017
457	MI	1/2/2017
234	(null)	1/2/2017
457	OK	6/7/2018
567	MT	6/9/2018

CUST_ID is not unique in Table B

INNER JOIN

ORDER_ID	CUST_ID	STATE
12345	567	WA
12345	567	MT
12346	457	OK
12346	457	MI

LEFT JOIN

ORDER_ID	CUST_ID	STATE
12345	567	WA
12345	567	MT
12346	457	OK
12346	457	MI
12366	123	(null)
14567	(null)	(null)
19453	(null)	(null)
15678	234	(null)

REQUIREMENTS FOR A JOIN BETWEEN TWO TABLES

1. **At least 2 tables, with at least 1 column/field in common**
2. **Column/field(s) tables have in common, MUST be same data type**
3. **The column/field in Table B *should* be unique, otherwise duplication will occur *(more on this later)***

SOLUTIONS

Dedup TABLE B via Subquery or CTE.

Join on more than column, if possible.

Instructor Led: Write the SQL JOIN as a Class

Hey Analyst, Give me the top 10 stores in terms of sales for 2014. I want them listed by store name.

Framework for approach

1. Draw the result-output first. Determine if you need a join or not.
2. Make sure 3 JOIN requirements are met.
3. Follow 3 rules of the JOIN syntax.
4. Pick appropriate JOIN. Be ready to defend.

JOINING Tables: Key Takeaways

1. If you want to join two tables, they must have 1 column in common, and of the same data type. The rows/records in Table B should be unique, otherwise duplication. Check TABLE B with supplied code.

2. The JOIN style will determine the result-output. Think about what records/rows will be excluded. Left joins are safest.

3. Work backwards, draw result output first, then determine ingredients to writing code.

Part 9

ADVANCED SQL TOPICS

Writing Conditional Statements in SQL (CASE)

Question: Give me total sales for primary color items vs. non-primary color items…

Prod_ID	Color	Unit_Sold	Primary_Color
1	Red	3	?
2	Pink	2	
3	Yellow	5	
4	Blue	3	
5	Green	1	
6	Brown	6	

IF Statement

Color	Unit_Sold	Primary_Color
Red	3	=IF(OR(H21="Red",H21="Yellow",H21="Blue"),1,0)
Pink	2	0
Yellow	5	1
Blue	3	1
Green	1	0
Brown	6	0

CASE Statement

```
SELECT
prod_id
,color
,unit_sold
CASE WHEN COLOR IN ('Red','Yellow','Blue')
    THEN 1 ELSE 0 END AS PRIMARY_COLOR

FROM TABLE
```

CASE Statement Use Cases

- Case statements are useful for
  - Categorization
  - Inspiring action

```
CASE WHEN TOTAL_BIRTHS > TOTAL_DEATHS THEN 'WE HAVE ENOUGH BABIES'
     WHEN TOTAL_BIRTHS < TOTAL_DEATHS THEN 'MAKE MORE BABIES'
     WHEN TOTAL_BIRTHS = TOTAL_DEATHS THEN 'STEADY'
     END AS POPULATION_STATUS
```

Writing Subqueries in SQL – Use Cases

- **What is a subquery?**
  - It's a tool for performing operations in multiple steps. **It's a query within a query.** Use to return data that will be used in the main query.

- **When would you use?**
  - You want to take an aggregate of an aggregate
  - You want to de-duplicate a table to perform a JOIN
  - You want reference information from multiple tables
  - You want to use **WHERE to filter on an aggregate**

- **What is the order of operations for a subquery?**
  - From inside out (**Subquery runs first** → **Main query runs second**))

- **What are the requirements?**
  - Subqueries must be enclosed in ()
  - An alias must be supplied to them

RUNS 2nd

RUNS 1st

```
SELECT
 A.PLAYS
,A.GENRE

FROM

(SELECT

SUM(PLAYS) as PLAYS
,GENRE

FROM ALBUMS

GROUP BY GENRE) A

WHERE
A.PLAYS > 1000
```

Writing Subqueries in SQL – Example Flow

RUNS 2nd

RUNS 1st

```
SELECT
 A.PLAYS
,A.GENRE

FROM

(SELECT

SUM(PLAYS) as PLAYS
,GENRE

FROM ALBUMS

GROUP BY GENRE) A

WHERE
A.PLAYS > 1000
```

RUNS 1st – Creation of Table A

	plays bigint	genre text
1	5000	[null]
2	795	Rock
3	150	Metal
4	3090	Rap
5	460	Electro...

RUNS 2nd – Run SELECT STATEMENT, off Table A

Data Output | Explain

	plays bigint	genre text
1	5000	[null]
2	3090	Rap

Part 10
CONCLUSION

We extracted the data, using SQL. Now what do we do with it? Remember analytics workflow…

Hey Analyst,

People

What were tequila sales for 2014?

What month and liter sizes had the highest sales?

SQL

Dashboard | Properties | SQL

valerianalytics on valeri_teacher@Teacher Valeri Analytics - AWS

```sql
SELECT
        sum(total) as SALES
        ,to_char(date, 'YYYY') as YEAR
        ,category_name
        ,liter_size
        ,to_char(date, 'MM') as MONTH

        FROM SALES

        WHERE CATEGORY_NAME = 'TEQUILA'

        GROUP BY
        to_char(date, 'YYYY')
        ,category_name
        ,liter_size
        ,to_char(date, 'MM')
```

Tableau

Hey Boss,

People

Tequila sales were $19.3MM for 2014.

The highest sales month was Apr, with $1.7MM in sales.

The most popular liter size was 750 ml (55% of sales)

Identify problem → Extract data → Analyze data → Answer problem

We offer courses on this part of the chain as well..

Conclusion: Key Takeaways

1. The more you talk to a database, you more comfortable you'll get.

2. Today's class gave us a framework for not just talking to 'Red', but to any relational database.

3. Don't let anyone stop you from getting to know a database; push hard.

Example Email to Send To Your DBA to Get Access

Hello DBA,

I'd to request SELECT access to our relational database that stores (x) data. Could you supply me the necessary credentials or Wiki for gaining access?

Could you also let me know

1. *Which version of SQL our organization uses? (E.g. Oracle / MySQL / Hive / PostgreSQL / IBM BIG Insights / SQL Server / etc.)*
2. *What RDBMS (or GUI) we use as a query tool?*
3. *Any information regarding the schema structure of the tables in the database for reference.*

Thanks,

Your Name

Q&A

(15 min)