

Slide 1: Using SQL for Data Science, Part 2

Using SQL for Data Science, Part 2

Sadie St. Lawrence, MSA
Data Scientist, VIP

UC DAVIS
EXTENSION

Slide 2: Learning Objectives

Learning Objectives

Determine and map out data elements needed for a query

Discuss strategies to use to write complex queries

Explain common troubleshooting techniques

Slide 3: Profiling Data

Profiling Data

Get into the **details** of your data

Create a **data model** and map the fields and tables you need

Consider **joins** and **calculations**

Understand any **data quality** or **format issues**

Slide 4: Start with SELECT

Start with SELECT

Start **simply**

Any query begins with **SELECT** statement

Add in **special formatting**

If using subqueries, start with the inner-most query and **work outward**

Slide 5: Test and Troubleshoot

Test and Troubleshoot

Do not wait until the end to test queries

Test after each join or filter

Are you getting the results you expect?

Start small and go step-by-step when troubleshooting a query

Slide 6: Format and Comment

Format and Comment

Use correct formatting and indentation

Comment strategically

Clean code and comments help when you revisit or hand off code

Slide 7: Review

Review

Always review old queries

Business rules

Date changes

Date Indicators

Work the problem for beginning to end