

Slide 1: Using SQL for Data Science, Part 1

Using SQL for Data Science, Part 1

Sadie St. Lawrence, MSA
Data Scientist, VIP

UCDAVIS
EXTENSION

Slide 2: Learning Objectives

Learning Objectives

Discuss importance of understanding your data when starting a new problem

Discuss importance of understanding business needs before beginning data analysis

Slide 3: Working Through a Problem from Beginning to End

Working Through a Problem from Beginning to End

Data Understanding

Test

Business
Understanding

Format & Comment

Review

Profiling

Start with SELECT

Slide 4: Data Understanding

Data Understanding

Most important step

Understanding relationships in your data

NULL values

String values

Dates and times

Slide 5: Subject Area Understanding

Subject Area Understanding

Until you gain business understanding, writing queries may take more time

Understanding where data joins are

Differentiating integers from strings

Investing time to understand your subject will help later during data analysis

Slide 6: Business Understanding

Business Understanding

Ask question about business problem you are solving

Hard to separate data and business understanding



Slide 7: Beware of the Unspoken Need

Beware of the Unspoken Need

“We want to predict whether or not a customer is likely to buy our product.”

Which customers?

What product?

What is/should be excluded?

What is/should be counted from past?

Slide 8: Moving Between Data and Business Understanding

