1. Record Nr. UNINA9910555033103321

Autore	Teate Renee M.
Titolo	SQL for data scientists : a beginner's guide for building datasets for analysis / / Renee M. Teate
Pubbl/distr/stampa	Hoboken, New Jersey : , : John Wiley & Sons, Inc., , [2021] 2021
ISBN	1-119-66939-1 1-119-66938-3 1-119-66937-5 9781119669364 9781119669371
Descrizione fisica	1 online resource (291 pages)
Disciplina	005.756
Soggetti	SQL (Computer program language) software data science information analysis text and data mining programming language
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	Cover -- Title Page -- Copyright Page -- About the Author -- About the Technical Editor -- Acknowledgments -- Contents at a Glance -- Contents -- Introduction -- Who I Am and Why I'm Writing About This Topic -- Who This Book Is For -- Why You Should Learn SQL if You Want to Be a Data Scientist -- What I Hope You Gain from This Book -- Conventions -- Reader Support for This Book -- Companion Download Files -- How to Contact the Publisher -- How to Contact the Author -- Chapter 1 Data Sources -- Data Sources -- Tools for Connecting to Data Sources and Editing SQL -- Relational Databases -- Dimensional Data Warehouses -- Asking Questions About the Data Source -- Introduction to the Farmer's Market Database -- A Note on Machine Learning Dataset Terminology -- Exercises -- Chapter 2

The SELECT Statement -- The SELECT Statement -- The Fundamental Syntax Structure of a SELECT Query -- Selecting Columns and Limiting the Number of Rows Returned -- The ORDER BY Clause: Sorting Results -- Introduction to Simple Inline Calculations -- More Inline Calculation Examples: Rounding -- More Inline Calculation Examples: Concatenating Strings -- Evaluating Query Output -- SELECT Statement Summary -- Exercises Using the Included Database -- Chapter 3 The WHERE Clause -- The WHERE Clause -- Filtering SELECT Statement Results -- Filtering on Multiple Conditions -- Multi-Column Conditional Filtering -- More Ways to Filter -- BETWEEN -- IN -- LIKE -- IS NULL -- A Warning About Null Comparisons -- Filtering Using Subqueries -- Exercises Using the Included Database -- Chapter 4 CASE Statements -- CASE Statement Syntax -- Creating Binary Flags Using CASE -- Grouping or Binning Continuous Values Using CASE -- Categorical Encoding Using CASE -- CASE Statement Summary -- Exercises Using the Included Database -- Chapter 5 SQL JOINs -- Database Relationships and SQL JOINs -- A Common Pitfall when Filtering Joined Data -- JOINs with More than Two Tables -- Exercises Using the Included Database -- Chapter 6 Aggregating Results for Analysis -- GROUP BY Syntax -- Displaying Group Summaries -- Performing Calculations Inside Aggregate Functions -- MIN and MAX -- COUNT and COUNT DISTINCT -- Average -- Filtering with HAVING -- CASE Statements Inside Aggregate Functions -- Exercises Using the Included Database -- Chapter 7 Window Functions and Subqueries -- ROW NUMBER -- RANK and DENSE RANK -- NTILE -- Aggregate Window Functions -- LAG and LEAD -- Exercises Using the Included Database -- Chapter 8 Date and Time Functions -- Setting datetime Field Values -- EXTRACT and DATE_PART -- DATE_ADD and DATE_SUB -- DATEDIFF -- TIMESTAMPDIFF -- Date Functions in Aggregate Summaries and Window Functions -- Exercises -- Chapter 9 Exploratory Data Analysis with SQL -- Demonstrating Exploratory Data Analysis with SQL -- Exploring the Products Table -- Exploring Possible Column Values -- Exploring Changes Over Time -- Exploring Multiple Tables Simultaneously -- Exploring Inventory vs. Sales -- Exercises -- Chapter 10 Building SQL Datasets for Analytical Reporting -- Thinking Through Analytical Dataset Requirements -- Using Custom Analytical Datasets in SQL: CTEs and Views -- Taking SQL Reporting Further -- Exercises -- Chapter 11 More Advanced Query Structures -- UNIONs -- Self-Join to Determine To-Date Maximum -- Counting New vs. Returning Customers by Week -- Summary -- Exercises -- Chapter 12 Creating Machine Learning Datasets Using SQL -- Datasets for Time Series Models -- Datasets for Binary Classification -- Creating the Dataset -- Expanding the Feature Set -- Feature Engineering -- Taking Things to the Next Level -- Exercises -- Chapter 13 Analytical Dataset Development Examples -- What Factors Correlate with Fresh Produce Sales? -- How Do Sales Vary by Customer Zip Code, Market Distance, and Demographic Data? -- How Does Product Price Distribution Affect Market Sales? -- Chapter 14 Storing and Modifying Data -- Storing SQL Datasets as Tables and Views -- Adding a Timestamp Column -- Inserting Rows and Updating Values in Database Tables -- Using SQL Inside Scripts -- In Closing -- Exercises -- Appendix Answers to Exercises -- Chapter 1: Data Sources -- Answers -- Chapter 2: The SELECT Statement -- Answers -- Chapter 3: The WHERE Clause -- Answers -- Chapter 4: CASE Statements -- Answers -- Chapter 5: SQL JOINs -- Answers -- Chapter 6: Aggregating Results for Analysis -- Answers -- Chapter 7: Window Functions and Subqueries -- Answers -- Chapter 8: Date and Time Functions -- Answers -- Chapter 9: Exploratory Data

	Analysis with SQL -- Answers -- Chapter 10: Building SQL Datasets for Analytical Reporting -- Answers -- Chapter 11: More Advanced Query Structures -- Answers -- Chapter 12: Creating Machine Learning Datasets Using SQL -- Answers -- Chapter 14: Storing and Modifying Data
Sommario/riassunto	SQL for Data Scientists: A Beginner's Guide for Building Datasets for Analysis is a resource that’s dedicated to the Structured Query Language (SQL) and dataset design skills that data scientists use most. Aspiring data scientists will learn how to how to construct datasets for exploration, analysis, and machine learning. You can also discover how to approach query design and develop SQL code to extract data insights while avoiding common pitfalls. You may be one of many people who are entering the field of Data Science from a range of professions and educational backgrounds, such as business analytics, social science, physics, economics, and computer science. Like many of them, you may have conducted analyses using spreadsheets as data sources, but never retrieved and engineered datasets from a relational database using SQL, which is a programming language designed for managing databases and extracting data. This guide for data scientists differs from other instructional guides on the subject. It doesn’t cover SQL broadly. Instead, you’ll learn the subset of SQL skills that data analysts and data scientists use frequently. You’ll also gain practical advice and direction on "how to think about constructing your dataset." In this book, author Renee Teate shares knowledge gained during a 15-year career working with data, in roles ranging from database developer to data analyst to data scientist. She guides you through SQL code and dataset design concepts from an industry practitioner’s perspective, moving your data scientist career forward!