18ECP101L-MASSIVE OPEN ONLINE COURSE-I

SEMESTER VI

MONTH & YEAR: APR 2023

SQL FOR DATA SCIENCE

Report Submitted by
Kunal Keshan
[RA2011004010051]

Faculty in-charge
Sarada. V

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

COLLEGE OF ENGINEERING AND TECHNOLOGY

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

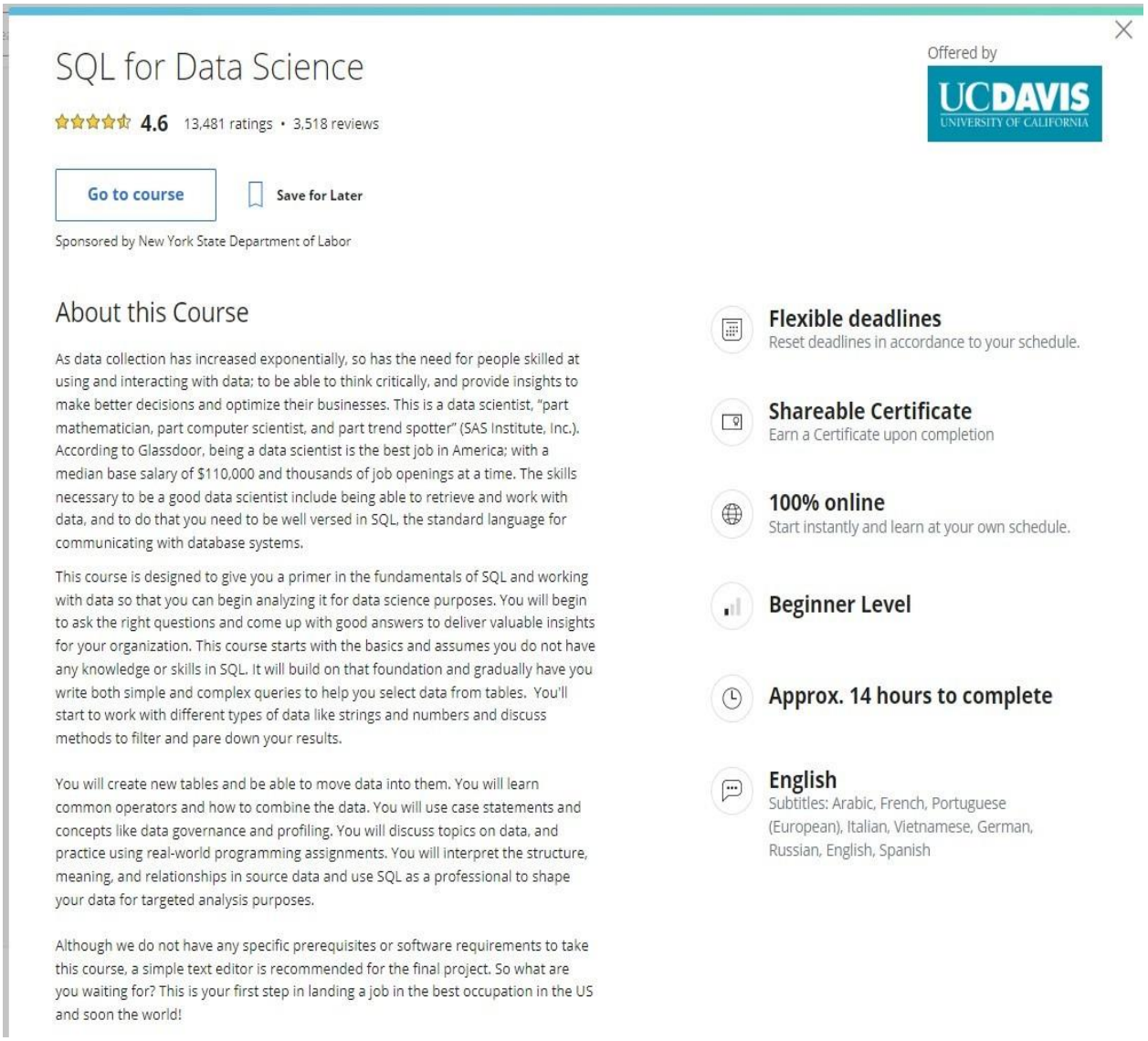
S.R.M. Nagar, Kattankulathur - 603203, Kancheepuram District

TABLE OF CONTENTS:

1.COURSE DETAILS

COURSE PLATFORM- Coursera
COURSE TITLE - SQL For Data Science
OFFERING UNIVERSITY - University of California Davis
COURSE DURATION- 4 Weeks

DASHBOARD:



2. INSTRUCTOR PROFILE

Sadie St. Lawrence
Founder and CEO Women in Data (WID)
University of California, Davis

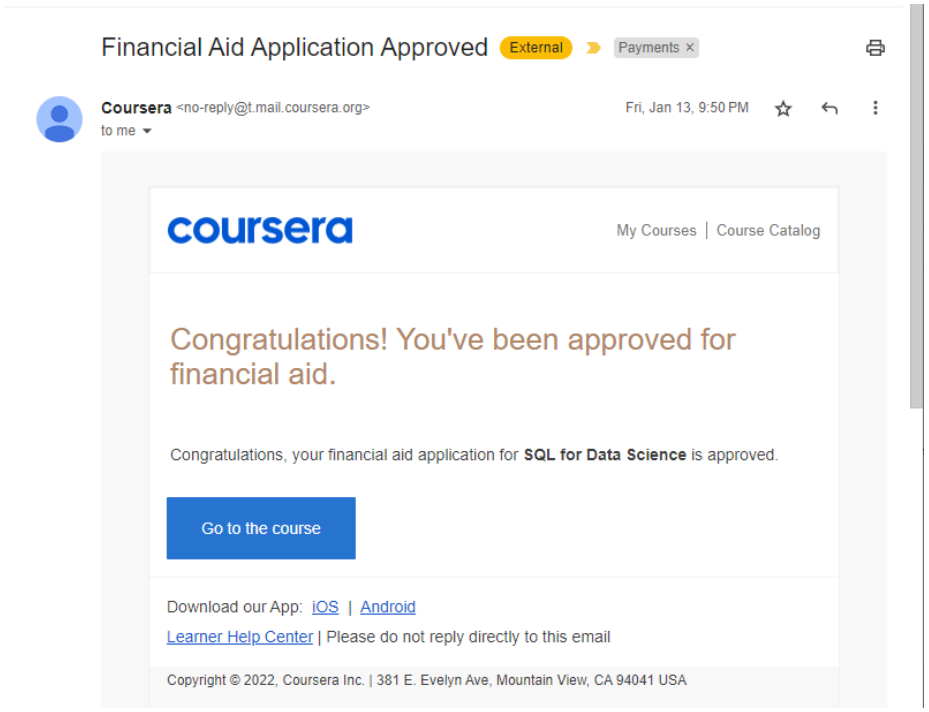
BIO:

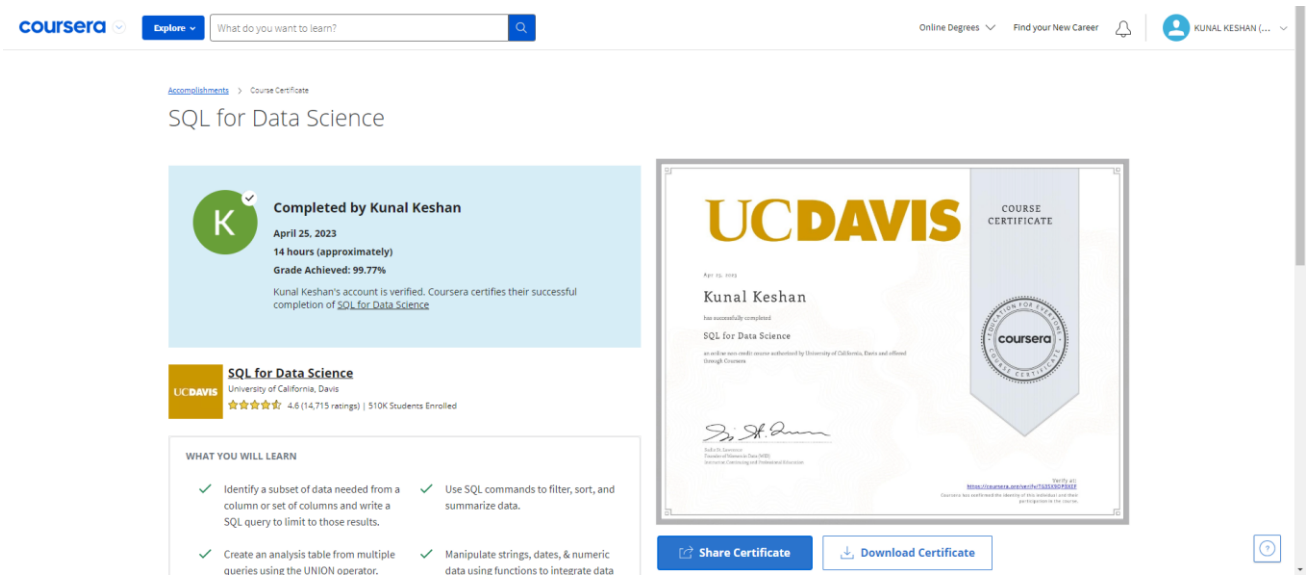
Sadie St. Lawrence is the Founder and CEO of Women in Data, a national non-profit organisation focused on increasing diversity, creating connections, and driving meaning from data.
Sadie brings a unique combination of technical expertise, analytics management experience and an ability to lead organisational change through compassion and problem-solving.
She has trained over 70,000 people in Data Science and is an advocate for democratising artificial intelligence and helping people transition into the 4th industrial revolution.
Sadie is on a mission to bring compassion to business and finds joy inhelping individuals become liberated and find their own personal journey to a more authentic and connected life.
In her free time, Sadie enjoys traveling, skiing, paddle boarding, yoga, and authentic conversations

3. COURSE TIMELINE

DATE OF ENROLMENT: 13th Jan 2023

DATE OF COMPLETION: 25th Apr 2023





4. SYLLABUS

Syllabus

WEEK 1

Getting Started and Selecting & Retrieving Data with SQL

In this module, you will be able to define SQL and discuss how SQL differs from other computer languages. You will be able to compare and contrast the roles of a database administrator and a data scientist, and explain the differences between one-to-one, one-to-many, and many-to-many relationships with databases. You will be able to use the SELECT statement and talk about some basic syntax rules. You will be able to add comments in your code and synthesize its

More

11 videos, 2 readings, 2 practice quizzes expand

Graded: Module 1 Quiz

Graded: Module 1 Coding Questions

WEEK 2

Filtering, Sorting, and Calculating Data with SQL

In this module, you will be able to use several more new clauses and operators including WHERE, BETWEEN, IN, OR, NOT, LIKE, ORDER BY, and GROUP BY. You will be able to use the wildcard function to search for more specific or parts of records, including their advantages and disadvantages, and how best to use them. You will be able to discuss how to use basic math operators, as well as aggregate functions like AVERAGE, COUNT, MAX, MIN, and others to begin analyzing our

More

9 videos, 1 reading, 1 practice quiz expand

Graded: Module 2 Quiz

Graded: Module 2 Coding Assignment

WEEK 3

Subqueries and Joins in SQL

In this module, you will be able to discuss subqueries, including their advantages and disadvantages, and when to use them. You will be able to recall the concept of a key field and discuss how these help us link data together with JOINs. You will be able to identify and define several types of JOINs, including the Cartesian join, an inner join, left and right joins, full outer joins, and a self join. You will be able to use aliases and pre-qualifiers to make your SQL code cleaner and efficient.

10 videos, 2 readings, 1 practice quiz **expand**

Graded: Module 3 Quiz

Graded: Module 3 Coding Assignment

WEEK 4

Modifying and Analyzing Data with SQL

In this module, you will be able to discuss how to modify strings by concatenating, trimming, changing the case, and using the substring function. You will be able to discuss the date and time strings specifically. You will be able to use case statements and finish this module by discussing data governance and profiling. You will also be able to apply fundamental principles when using SQL for data science. You'll be able to use tips and tricks to apply SQL in a data science context.

10 videos, 3 readings **expand**

Graded: Module 4 Quiz

Graded: Module 4 Coding Questions

Graded: Data Scientist Role Play: Profiling and Analyzing the Yelp Dataset

5. WEEK WISE CONTENTS

WEEK 1:

Video: Course Introduction

Video: Module Introduction

Video: What is SQL Anyway?

Video: Data Models, Part 1: Thinking About Your Data

Video: Data Models, Part 2: The Evolution of Data Models

Video: Data Models, Part 3: Relational vs. Transactional Models

Video: Retrieving Data with a SELECT Statement

Video: Creating Tables

Video: Creating Temporary Tables

Video: Adding Comments to SQL

Reading: SQL Overview

Reading: Data Modelling and ER Diagrams

MODULE 1 OBJECTIVES:

In this module, you will be able to define SQL and discuss how SQL differs from other computer languages. You will be able to compare the roles of a database administrator and a data scientist, and explain the differences between one-to-one, one-to-many, and many-to- many relationships with databases. You will be able to use the SELECT statement and talk about some basic syntax rules. You will be able to add comments in your code and synthesise its importance.

- Distinguish between use of SQL for data science applications and SQL for morecommon data management operations.
- Use an Entity Relationship diagram, describing the data elements, their relationships,and inter-dependencies and determine if the existent data is sufficient to address a business question.
- Retrieve one or more columns of data from a table that relates to the rese

- Identify a subset of data needed from a column or set of columns and write a SQLquery to limit to those results.
- Create an analysis environment and use INSERT to put data into a table.
- Add effective comments in your queries so that one, you can remember what you'redoing, and two, so others can review your work.

WEEK 2:

Video: Basics of Filtering with SQL

Video: Advanced Filtering: IN, OR, and NOT

Video: Using Wildcards in SQL

Video: Sorting withORDER

BY**Video:** Math Operations

Video: Aggregate Functions

Video: Grouping Data with

SQL **Video:** Putting it All

Together

Reading: SQL for Various Data Science Languages

MODULE 2 OBJECTIVES:

In this module, you will be able to use several newer clauses and operators including WHERE, BETWEEN, IN, OR, NOT, LIKE, ORDER BY, and GROUP BY. You will be able to use the wildcard function to search for more specific or parts of records, including their advantages and disadvantages, and how best to use them. You will be able to discuss how to use basic math operators, as well as aggregate functions like AVERAGE, COUNT, MAX, MIN, and others to begin analysing our data.

- Compare analytics tool and CPU time performance between a filtered andunfiltered dataset.
- Given a dataset analysis requirement, use WHERE, IN, NOT, AND, andOR alone or in combination to filter the dataset.
- Determine whether to use wildcards in a data filter or search situation.
- Use wildcards to search or filter data based on requirements. Use regularexpressions for text processing
- Use ORDER BY to sort data according to requirements for number ofcolumns in the sort, sort direction, and sort position.
- Create common math operation calculated fields and aliases forcalculated fields.
- Use AVG, COUNT, MAX, MIN, SUM to profile data.
- Summarise data according to one or more criterion using GROUP BYand HAVING clauses.

WEEK 3:

Video: Using Subqueries

Video: Subquery Best Practices and Considerations

Video: Joining Tables: An Introduction

Video: Inner Joins

Video: Aliases and Self Joins

Video: Advanced Joins: Left, Right, and Full Outer Joins

Video: Unions

Reading: SQL and Python

Reading: Union and Union All

MODULE 3 OBJECTIVES:

In this module, you will be able to discuss subqueries, including their advantages and disadvantages, and when to use them. You will be able to recall the concept of a key field and discuss how these help us link data together with JOINs. You will be able to identify and define several types of JOINs, including the Cartesian join, an inner join, left and right joins, full outer joins, and a self-join. You will be able to use aliases and pre-qualifiers to make your SQL code cleaner and efficient.

- Retrieve data from multiple tables using subqueries.
- Join tables using an Inner Join and table aliases.
- Filter a given data set using set theory by joining tables using Natural,Outer, and Self Joins.
- Assess the risk versus benefit of using a Cross Join or Cartesian Join on aset of data.
- Create an analysis table from multiple queries using the UNION operator.

WEEK 4:

Video: Working with Text Strings

Video: Working with Date and Time Strings

Video: Date and Time Strings Examples

Video: Case Statements

Video: Data Governance and Profiling

Video: Using SQL for Data Science, Part 1

Video: Using SQL for Data Science, Part 2

Reading: Welcome to Peer Review Assignments!

MODULE 4 OBJECTIVES:

In this module, you will be able to discuss how to modify strings by concatenating, trimming, changing the case, and using the substring function. You will be able to discuss the date and time strings specifically. You will be able to use case statements and finish this module by discussing data governance and profiling. You will also be able to apply fundamental principles when using SQL for data science. You'll be able to use tips and tricks to apply SQL in a data science context.

- Manipulate strings, dates, and numeric data using functions to integrate data from different sources into fields with the correct format for analysis.
- Use Case / When statements to recode a set of data for grouping at adifferent level (e.g., cities to regions).
- Use Views to simplify SQL operations
- Identify organisational, governance, business, and data conditions thatindicate use of a join to prepare data for analysis.
- Recite and implement the 3 rules for changing an analysis question into aSQL statement

6. Assignment/Quiz Details

Overall Grade: 99.77%

You have completed all of the assessments that are currently due.

You passed this course! Your grade is 99.77%.

Item	Status	Due	Weight	Grade
Module 1 Quiz Quiz	Passed	Apr 23 11:59 PM IST	5%	98%
Module 1 Coding Questions Quiz	Passed	Apr 23 11:59 PM IST	12.50%	100%
Module 2 Quiz Quiz	Passed	Apr 30 11:59 PM IST	5%	100%
Module 2 Coding Assignment Quiz	Passed	Apr 30 11:59 PM IST	12.50%	100%
Module 3 Quiz Quiz	Passed	May 7 11:59 PM IST	5%	97.50%
Module 3 Coding Assignment Quiz	Passed	May 7 11:59 PM IST	12.50%	100%
Module 4 Quiz Quiz	Passed	May 14 11:59 PM IST	5%	100%
Module 4 Coding Questions Quiz	Passed	May 14 11:59 PM IST	12.50%	100%
Data Scientist Role Play: Profiling and Analyzing the Yelp Dataset Submit your assignment and review 4 peers' assignments to get your grade.			30%	100%
Submit your assignment	Passed	May 14 11:59 PM IST		
Review 4 peers' assignments.	4/4 reviewed	May 17 11:59 PM IST		

7. COURSE OUTCOMES:

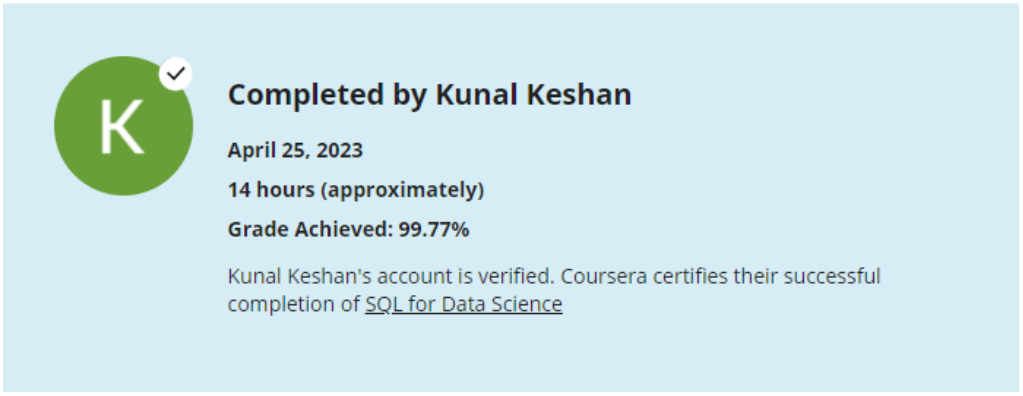
WHAT YOU WILL LEARN

- Identify a subset of data needed from a column or set of columns and write a SQL query to limit to those results.
- Use SQL commands to filter, sort, and summarize data.
- Create an analysis table from multiple queries using the UNION operator.
- Manipulate strings, dates, & numeric data using functions to integrate data from different sources into fields with the correct format for analysis.
-

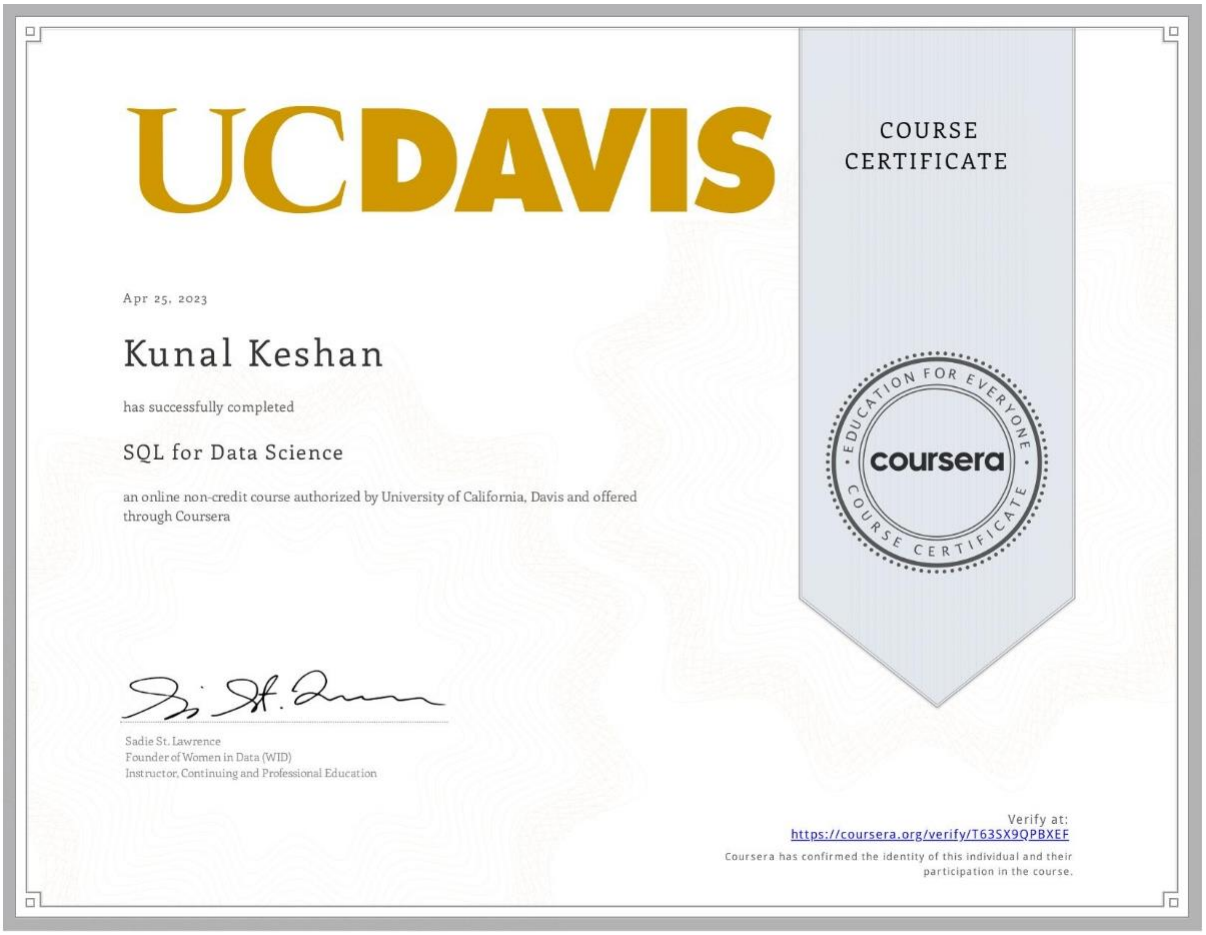
SKILLS YOU WILL GAIN

- Data Science
- Data Analysis
- SQLite
- SQL

8.PROOF OF COURSE COMPLETION:



9. COURSE COMPLETION CETRIFICATE:



UCDAVIS

COURSE CERTIFICATE

Apr 25, 2023

Kunal Keshan

has successfully completed

SQL for Data Science

an online non-credit course authorized by University of California, Davis and offered through Coursera

Sadie St. Lawrence
Founder of Women in Data (WID)
Instructor, Continuing and Professional Education

EDUCATION FOR EVERYONE
coursera
COURSE CERTIFICATE

Verify at:
https://coursera.org/verify/T63SX9QPBXEF

Coursera has confirmed the identity of this individual and their participation in the course.