



SQL Masterclass for Data Scientists

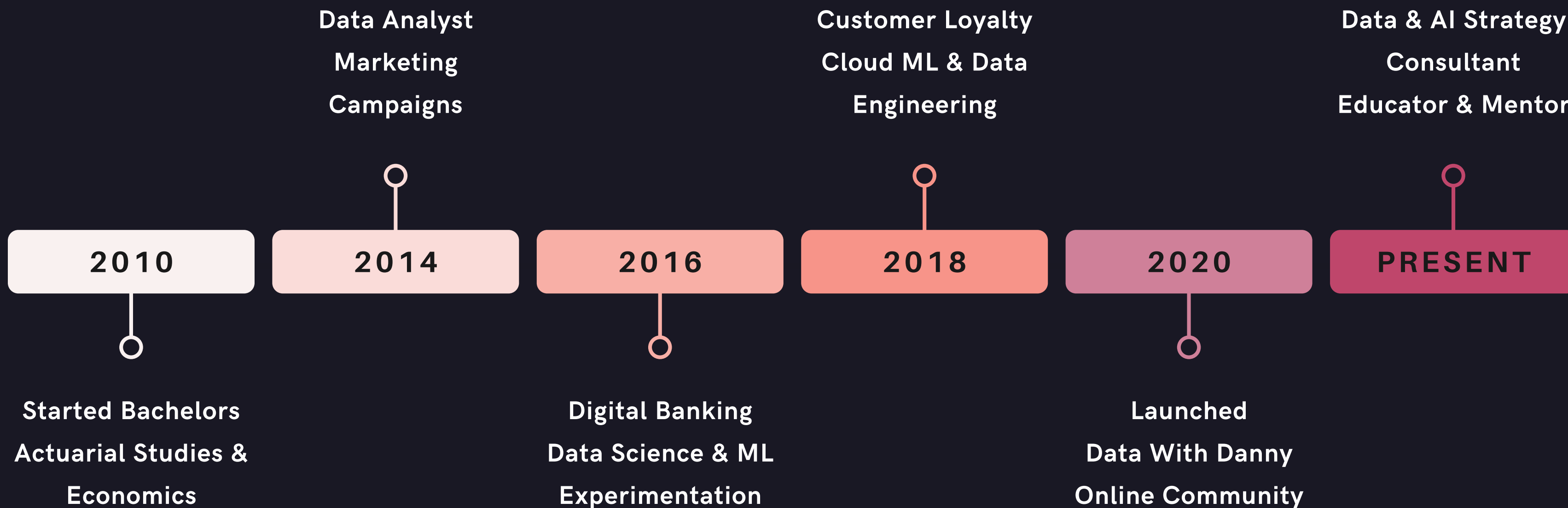
Danny Ma
Sept 2021



My Data Science Journey

(so far...)

03





DATA WITH DANNY

VIRTUAL DATA APPRENTICESHIP

How can I find
out more about
Data With Danny?



bit.ly/dwd-info

COMMON SQL MYTHS

(CREATED BY DATA SCIENTISTS)

"SQL is such an old language, I only want to learn modern tools so it looks better on my resume"

"I can do everything in Python, I don't need SQL for any of my data science tasks"

"SQL is so slow and poor performing, I can't rely on it because sometimes the database can't handle my queries"

"The only SQL I need is `SELECT * FROM` because I always need the entire dataset from the database"

"SQL is going away soon and I don't want to waste my time learning it - everything is moving to NoSQL anyway"

"SQL is only for data analysts, as a data scientist - I don't need to use SQL anymore because it's below me"

SQL FACTS

07

SQL appears in over 40% of all data job listings posted on Indeed in 2021

SQL queries can be used with many languages and frameworks via objects and APIs to interact with databases

SQL is a standardised language making it a very versatile skill used across different domains and tech environments

65% of data scientists and data analysts use SQL according to a 2020 Stack Overflow survey

The modern data stack is moving to SQL-centric tools such as Snowflake, Amazon Redshift and Google BigQuery

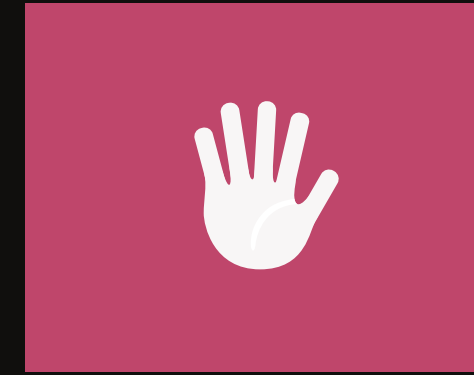
SQL databases can also store semi-structured data like JSON objects with expanding support for images and other newer data sources

THE PLAN FOR TODAY

08



**SQL
Masterclass**



**General
Q&A**

Where can I find
today's slides &
code examples?



bit.ly/odsc-sql



SQL Masterclass



General Q&A