Why Learn Data Modeling and SQL in Data Science

Adham Al Sinaidi

Data Science

Structured Data

Structured data means data stored in a clear format, like tables with rows and columns. It's easy for people and computers to read. Unstructured data, like text or images, has no clear format and takes extra work to prepare.

When data is structured, it's stored in a consistent way. For example, dates stay as dates, and numbers stay as numbers. This stops errors and confusion during analysis. It's also easier to join data from different places when the structure matches. A tidy table makes it simple to run statistics or build machine learning models. Each column holds one variable, and each row is one observation.

Clean, structured data helps machine learning work well. Models learn patterns best when data is organized. For example, a table of customer transactions with set columns feeds into a model right away. But loose receipts or mixed formats would take hours to sort out first.

In finance, structured data helps detect fraud because you can look for patterns across many records. In healthcare, structured records help algorithms make better diagnoses. Across industries, structured data is the backbone that makes data science possible.

How Data Modeling Prepares Data for Analysis

Data modeling designs how data should look and connect. It defines tables, fields, and links between data. A good model gives data a clear, logical shape. This helps clean and prepare the data for use.One method is "tidy data," where each column is one variable and each row is one observation. Tidy data makes it easy to analyze. If messy data comes in, data modeling organizes it into neat tables. This makes errors easier to spot and fix. For example, if text shows up in a date field, a well-defined schema will reject it.

Data modeling also helps transform raw data into useful form. A popular approach is the "medallion architecture," used in some Gulf companies. Data moves through layers: raw data in Bronze, cleaned data in Silver, and final data ready for reporting in Gold. Each step applies rules so data stays consistent and reliable.Data modeling cuts down on duplication. In a good database, each piece of information is stored only once. For instance, if a customer's email changes, you update it in one place. Without this structure, the same data might appear in many places, causing errors.

In short, data modeling builds a solid base for analysis. It gives structure, enforces rules, and keeps data clean.

Why Relational Databases Support Clean Data Workflows

Relational databases store data in linked tables. Each table holds related data, and keys connect the tables. This keeps data organized and accurate.

Relational databases enforce rules like unique IDs and data types. These rules prevent mistakes and keep data consistent. They also let analysts combine data from many tables using joins. For example, a single query can pull customer data, sales records, and marketing campaigns into one report.

Unlike spreadsheets, relational databases handle large volumes of data. Excel caps at about one million rows, but databases manage billions. Multiple people can work with a database at once without conflicts. If someone updates data, everyone sees the new version right away.

A good example in the Gulf is Majid Al Futtaim, which runs Carrefour stores. They moved to a central data warehouse. This system stores all their data in one place with clear rules and models. Their reports run faster, and the company works from a single, trusted source of truth.

Relational databases also help retrieve data fast. Indexes and built-in optimization mean you can search huge tables quickly. Compared to working with text files, databases save time and reduce errors. For clean, reliable data science work, relational databases remain essential.

Why SQL Remains Key in Data Science

SQL, or Structured Query Language, has been around for decades. But it's still a core skill for data science. Most business data lives in relational databases. SQL is the main way to get that data.

SQL lets you write precise queries to pull exactly the data you need. Without SQL, analysts might have to rely on others or use tools with limited features. SQL gives control and flexibility.

SQL also handles large data better than Python for many tasks. Python tools like pandas are great, but they can slow down or run out of memory with big data. Databases run SQL fast because they're built to handle large, complex queries.

Often, data work uses both SQL and Python. You might use SQL to filter or summarize data, then analyze it in Python. Many Python tools can run SQL directly, showing that the two work together rather than replace each other.

Skipping SQL can limit your work. Databases often have rules and security that block direct access to raw data. Without SQL, you may be stuck with prepared reports or spreadsheets. Also, many modern tools, from Spark to cloud databases, use SQL or similar languages. Learning SQL opens doors across many systems.

In short, SQL remains a must-have skill because it gives direct, fast, and reliable access to the data that powers data science.

Real-World Examples in the Gulf

- **Majid Al Futtaim (Carrefour)**
  In the UAE, Majid Al Futtaim needed to analyze growing retail data. Their old systems were slow. They built a new data warehouse using IBM Netezza. This centralized all their data, made queries three times faster, and ensured everyone worked from the same data set.
- **Middle East Omni-Channel Retailer**
  A large retailer moved from legacy systems to a cloud platform using Databricks. They used a medallion architecture with Bronze, Silver, and Gold layers. This kept data clean and ready for machine learning. Processing times dropped, and business users could access reliable data.
- **Open Data Portals**
  Gulf governments now share large data sets online. Saudi Arabia's portal offers over 10,000 datasets, and the UAE's Bayanat has more than 3,000. These datasets cover topics like health, transport, and the economy. They are structured for easy analysis, showing how governments see data as a tool for growth.
- **SQL in Telecom Analytics**
  In telecom, analysts use SQL to study churn. Before building a churn model, they run SQL queries to check call times, support tickets, or payments. This helps build features for machine learning. Similar SQL steps help find fraud patterns or build recommendation systems.

These examples show how data modeling, structured data, and SQL work together. In the Gulf, they help both businesses and governments turn data into clear insights.

Personal Reflection

This research made what I've learned in class feel more real. Before, data modeling seemed abstract. Now I see it's key to keeping data clean and usable. Learning how companies in the Gulf manage data shows these skills are needed everywhere, not just in tech giants abroad.

I was surprised how important SQL still is. I thought newer tools might replace it. Instead, SQL runs under many modern tools and remains central to getting data. I see now that knowing SQL helps me work faster and with fewer limits.

Seeing Gulf projects made me realize there are big opportunities close to home. Governments and businesses here invest in structured data and analytics. It makes me eager to keep building these skills, knowing they're in demand.

References

- ClearPeople. (2023). *Importance of structured data in AI*. ClearPeople. https://www.clearpeople.com/blog/importance-of-structured-data-in-ai
- IBM Case Study, “How Fresh Data Can Drive Better Decision-Making,” 2022. https://www.ibm.com/case-studies/capital-bank
- Javier Canales Luna, “SQL vs Python: Which Should You Learn?” DataCamp Blog, 2023. https://www.datacamp.com/blog/sql-vs-python-which-to-learn
- Kenton Rambsy, “Why is Structured Data Important?” The Data Notebook (UTA Pressbooks), 2019. https://uta.pressbooks.pub/datanotebook/chapter/1-3-structured-data/
- Corey Ginsberg, “How are Relational Databases Used in Data Analytics?” Noble Desktop, 2025. https://www.nobledesktop.com/classes-near-me/blog/relational-databases-in-data-analytics
- “Database vs Spreadsheet: Full Comparison,” 365 Data Science, 2020. https://365datascience.com/tutorials/sql-tutorials/database-vs-spreadsheet/
- Siddharth Yadav & Mahdi Ghuloom, “From Innovation to Participation: Why Open Data Matters in the Gulf,” ORF Middle East, 2025. https://orfme.org/research/gulf-digital-transformation-open-data/
- Prescience Decision Solutions, “Middle-Eastern Omni-Channel Retailer Migrates to a New Data Platform,” 2024. https://prescienceds.com/case-studies/a-middle-east-based-multinational-company-migrates-to-a-new-data-intelligence-platform/