PROJECT: Why Learn Data Modeling and SQL in Data Science

**1. Importance of Structured Data**

- Structured data ensures clean, reliable pipelines for analysis and model training.  
- It supports efficient transformations like filtering, grouping, and aggregation.  
**2. Role of Data Modeling**

- Data modeling provides blueprints (conceptual, logical, physical) for databases.  
- It helps in reducing redundancy, improving data quality, and preparing data for machine learning.

**3. Relational Databases in Real Projects**

- RDBMS (**Relational Database Management System**.) ensures data integrity and scalability using ACID properties.

ACID stands for the **four key properties** that guarantee reliable processing of database transactions in a **Relational Database Management System (RDBMS)**:

|  |  |
| --- | --- |
| Property | Meaning |
| A – Atomicity | All operations in a transaction are completed, or none are. No partial changes. |
| C – Consistency | The database moves from one valid state to another, maintaining all rules and constraints. |
| I – Isolation | Transactions are processed independently and transparently — one doesn't affect another. |
| D – Durability | Once a transaction is committed, it will remain so — even in case of power loss or crash. |

- Used in real-world systems like data warehouses and production analytics systems.

**4. Why SQL Remains Foundational**

- SQL is declarative, efficient, and more reproducible than manual coding in Pandas.  
- SQL integrates seamlessly with Python and is used for preprocessing before ML

**5. Real-World Use of SQL Before ML**

- SQL is used to calculate metrics like customer lifetime value, CTR(**Click-Through Rate**.), and session duration.  
- These metrics are later used as input features in machine learning pipelines.

**6. Reflection**

- Course assignments involving ER diagrams and SQL mirror real-world modeling practices.  
- Understanding the why behind modeling and querying makes you a better data professional.

**References**

1. https://www.dasca.org/world-of-data-science/article/what-is-data-modeling-and-why-do-you-need-it

2. https://www.zucisystems.com/blog/what-is-data-modeling-and-why-is-it-important/

3. https://www.sap.com/products/data-cloud/datasphere/what-is-data-modeling.html

4. https://en.wikipedia.org/wiki/Data\_modeling

5. https://odsc.medium.com/5-reasons-why-sql-is-still-the-most-accessible-language-for-new-data-scientists-7c9ac1419b47

6. https://hackernoon.com/why-are-we-teaching-pandas-instead-of-sql

7. https://www.kdnuggets.com/using-pandas-and-sql-together-for-data-analysis

8. https://www.reddit.com/r/Python/comments/1byhpm7/do\_folks\_ever\_use\_pandas\_when\_they\_should\_use\_sql/