VIDEO 1 OF 8

SQL for Data Analysis

with Google BigQuery

Getting Started + Understanding the Dataset

Instructor: Uttam Deb
Akaademy Ltd.

Basic to Intermediate

What You'll Learn in This Course

- Write SQL queries from scratch in Google BigQuery
- Filter, sort, and summarize real business data
- Create calculated metrics (profit margins, averages)
- Analyze sales trends over time
- Detect data quality issues and duplicates
- Use intermediate SQL: CTEs and Window Functions

We use a real-world-style sales dataset throughout — no toy examples.

What is SQL?

SQL = Structured Query Language

The language used to talk to databases and ask questions about data.

Who uses SQL?

- Data Analysts — pull reports, find trends
- Product Managers — track metrics, user behavior
- Engineers — build features backed by data
- Business Teams — answer revenue, growth questions

"If data lives in a database, SQL is how you ask it questions."

Why Google BigQuery?

- Serverless — no installation, no server management
- Free tier — 1 TB queries/month, 10 GB storage
- Handles billions of rows effortlessly
- Standard SQL — skills transfer to any database
- Used by real companies for production analytics
- Start in 2 minutes with a Google account

No setup headaches. Just open your browser and start querying.

BigQuery Interface Overview

LIVE DEMO

Key areas to know:

- Navigation Panel — projects, datasets, tables (left sidebar)
- Query Editor — where you write SQL (top center)
- Results Panel — query output appears here (bottom center)
- Schema Tab — column names and data types
- Preview Tab — peek at data without running a query

Key Concepts: Project > Dataset > Table

Project

Your Google Cloud project
(billing unit)

Think: "Company"

Dataset

A folder for related tables
(container)

Think: "Department"

Table

The actual data
(rows + columns)

Think: "Spreadsheet"

```
-- How you reference a table in BigQuery:
SELECT * FROM `project_id.dataset_name.table_name`
```

Importing the CSV into BigQuery

LIVE DEMO

1. Open BigQuery Console
2. Create a new dataset (name it "sql_course")
3. Click "Create Table" inside that dataset
4. Source: Upload → Select the CSV file
5. Table name: "sales"
6. Schema: check Auto-detect
7. Click "Create Table" → Done!

Always check: Did BigQuery detect the column types correctly?

Meet the Dataset: Sales Data

Column	Type	Role
Order ID	STRING	Identifier
Amount	INTEGER	Revenue
Profit	INTEGER	Profitability
Quantity	INTEGER	Volume
Category	STRING	Product group
Sub-Category	STRING	Product detail
PaymentMode	STRING	Payment type
Order Date	STRING*	Date of order
CustomerName	STRING	Customer
State	STRING	Location
City	STRING	Location detail
Year-Month	STRING	Time shortcut

* Order Date may import as STRING — we'll fix this in the Dates video

Dimensions vs Metrics

METRICS (Numbers you measure)

- Amount — revenue
- Profit — profitability
- Quantity — volume

DIMENSIONS (Categories you group by)

- Category, Sub-Category
- State, City
- PaymentMode, CustomerName

The Analysis Formula:

Pick a METRIC + Group by a DIMENSION = Business Insight

Example: Total Sales (metric) by State (dimension) → "Which state earns the most?"

What Does One Row Represent?

Each row = one line item in an order

- One Order ID can have multiple rows (multiple products)
- Example: Order B-26776 has both Electronic Games and Printers
- If you count rows, you're counting line items, not unique orders
- This is common in real business data

> ⚠ **Important Nuance**
>
> Some Order IDs appear with different customers and different dates. We'll investigate this in the Data Quality video.

First Look at the Data

LIVE DEMO

```sql
-- Preview first 10 rows
SELECT *
FROM `sql_course.sales`
LIMIT 10;
```

```sql
-- How many rows total?
SELECT COUNT(*) AS total_rows
FROM `sql_course.sales`;
```

```sql
-- What categories exist?
SELECT DISTINCT Category
FROM `sql_course.sales`;
```

```sql
-- What payment modes exist?
SELECT DISTINCT PaymentMode
FROM `sql_course.sales`;
```

Don't worry about syntax yet — we'll break down every keyword in the next video.

BigQuery Pricing: What You Should Know

✗ **Expensive — scans ALL columns**

```
SELECT *
FROM sales;
```

✓ **Cheaper — scans only 2 columns**

```
SELECT Amount, Profit
FROM sales;
```

- BigQuery charges based on data scanned, not rows returned
- Free tier: 1 TB of queries/month (plenty for learning)
- Good habit: only SELECT the columns you actually need

Check Your Data Types After Import

LIVE DEMO

Click on your table → "Schema" tab → Verify:

- Amount, Profit, Quantity → should be INTEGER or FLOAT ✓
- Category, State, City, CustomerName → should be STRING ✓
- Order Date → might be STRING instead of DATE ⚠
- Year-Month → will be STRING (expected) ✓

Why this matters:

Wrong types = wrong results. You can't do date math on a STRING. We'll fix this in Video 5.

Business Questions This Course Will Answer

- Which category generates the most revenue?
- Which states are most profitable?
- Which payment mode is most popular?
- What are the monthly sales trends?
- Are there data quality issues we should worry about?
- Which sub-categories are high revenue but low profit?
- Who are the top customers?
- How do we rank and compare performance across segments?

These are the exact questions a product manager or business lead would ask you.

What's Next

Video 2: SQL Basics — SELECT, WHERE, ORDER BY & Calculated Columns

Now that you know the workspace and the data, we'll start writing real queries.

1. Setup + Dataset	2. SELECT / WHERE / ORDER BY	3. Aggregates + GROUP BY	4. HAVING + CASE WHEN
5. Dates + Trends	6. Data Quality Audit	7. Subqueries + CTEs	8. Window Functions

Practice Task

Do this before watching Video 2 (takes ~5 minutes):

- [ ] Import the CSV into your own BigQuery project
- [ ] Run: SELECT * FROM sql_course.sales LIMIT 10
- [ ] Check the Schema tab — note STRING vs INTEGER columns
- [ ] Identify which columns are dimensions and which are metrics
- [ ] Count total rows using COUNT(*)
- [ ] List unique values for Category, State, and PaymentMode