

Oil & Gas Well Data Quality Dashboard

Stakeholder Report

Project Summary

This report outlines the development and purpose of a data quality dashboard built to simulate and assess oil & gas well production data integrity. The dashboard is designed to profile common data issues such as missing values, negative volumes, and date anomalies, using SQL queries and visualization through Power BI.

The system provides a clear, interactive interface to track data quality trends and ensure clean, reliable reporting for downstream analytics and decision-making.

Project Objectives

- Detect and flag common data quality issues
- Provide visibility into data completeness and accuracy
- Create stakeholder-friendly visualizations and metrics
- Simulate metadata documentation (Alation-style)

Data Overview

Source: Simulated oil & gas production data stored in SQLite

Records: 5 well entries across multiple operators and states

Fields Captured: Well ID, Name, Location, Spud Date, Production Start/End, Volume Produced, Operator, Status

	Well_ID	Well_Name	Location	Spud_Date	Prod_Start	Prod_End	Volume_Produced	Operator	Status	
	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	
1	1001	Alpha-1	TX	2019-01-01	2019-04-15	2023-06-01	185000.0	Devon	Active	
2	1002	Beta-2	OK	2020-03-12	2020-05-22	2023-06-01	NULL	Unknown	Inactive	
3	1003	Gamma-3	NM	2021-07-08	2021-09-15	2021-09-01	200000.0	Conoco	Active	
4	1004	Delta-4	TX	2022-05-20	2022-07-01	2023-05-10	-500.0	Chevron	Inactive	
5	1005	Echo-5	CO	2022-09-10	2022-10-01	2022-10-01	130000.0	NULL	Active	

Data Quality Rules Applied

Rule Description	SQL Condition
Missing Volume or Operator	Volume_Produced IS NULL OR Operator IS NULL
Negative Production Volume	Volume_Produced < 0
Production End Before Start Date	Prod_End < Prod_Start
Duplicate Well IDs	GROUP BY Well_ID HAVING COUNT(*) > 1

Appendix A: SQL Data Quality Queries

1. Duplicate Well IDs

```
SELECT Well_ID, COUNT(*)  
FROM well_production  
GROUP BY Well_ID  
HAVING COUNT(*) > 1;
```

Results:

- 1002 | Beta-2 | OK | 2020-03-12 | 2020-05-22 | 2023-06-01 | Unknown | Inactive
- 1005 | Echo-5 | CO | 2022-09-10 | 2022-10-01 | 2022-10-01 | 130000.0 | Active

2. Negative Production Volume

```
SELECT * FROM well_production  
WHERE Volume_Produced < 0;
```

Result:

- 1004 | Delta-4 | TX | 2022-05-20 | 2022-07-01 | 2023-05-10 | -500.0 | Chevron | Inactive

3. Production End Before Start

```
SELECT * FROM well_production  
WHERE Prod_End < Prod_Start;
```

Result:

- 1003 | Gamma-3 | NM | 2021-07-08 | 2021-09-15 | 2021-09-01 | 200000.0 | Conoco | Active

4. Summary Statistics

```
SELECT  
    MIN(Volume_Produced) AS Min_Production,  
    MAX(Volume_Produced) AS Max_Production,  
    AVG(Volume_Produced) AS Avg_Production,  
    COUNT(*) AS Total_Records  
FROM well_production;
```

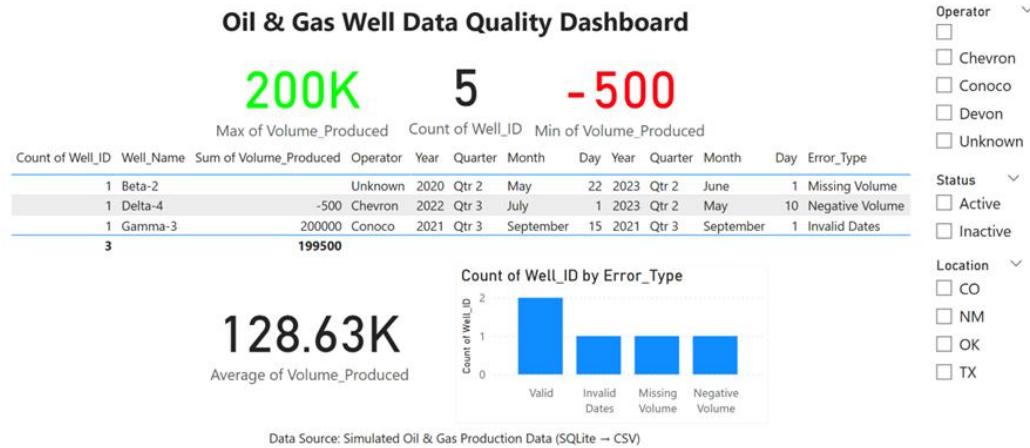
Results:

- Min Production: -500
- Max Production: 200,000
- Average Production: 128,625
- Total Records: 5

Dashboard KPIs

KPI	Value
Total Wells	5
Min Production Volume	-500

Max Production Volume	200,000
Average Production Volume	128,625
Records with Issues	3 (60%)



Simulated Metadata Documentation

Simulated Alation-style documentation was developed to represent how data stewardship might be captured in a production data catalog:

- Business Term Definitions
- Field Descriptions
- Data Flow Narrative
- Entry Validation Guidelines

Alation-Style Data Documentation

Business Terms

Term	Definition
Well_ID	Unique identifier for each well
Volume_Produced	Total oil or gas produced by a well (in bbls)
Spud_Date	Date when drilling began on the well
Prod_Start	Start date of production for the well
Prod_End	End date of production for the well
Operator	Company operating the well
Status	Current state of the well (e.g., Active)

Data Entry Guidelines

- `Volume_Produced` must be numeric and non-negative
- `Prod_End` must be on or after `Prod_Start`
- `Well_ID` must be unique
- `Operator` must not be null

Data Flow Diagram

ProCount (source) → SQLite → Power BI → Stakeholder Dashboard

Notes

This documentation simulates the type of metadata captured in Alation for business and technical stakeholders.

Conclusion

This dashboard provides clear, actionable insights into data quality issues for oil & gas well production. Its modular SQL logic and interactive visualizations make it adaptable for use in real-world data stewardship workflows.

Further improvements can include expanding the dataset, implementing automated profiling scripts, and incorporating additional KPIs or time-series metrics.

Contact

Will Thompson

✉️ Will.j.Thompson@outlook.com

🌐 github.com/Willthompson99