

OIL & GAS PRODUCTION DATA ANALYSIS PROJECT

Microsoft Power BI

Nurul Fatihah Binti Rizwan

PROJECT OVERVIEW

Purpose of the project

- To analyze oil and gas production data across multiple states
- To identify high-output regions and data quality issues
- To provide actionable insights for operational optimization

Tools Used:

- Microsoft Excel
- Power BI (Power Query for Cleaning, Visualizations, and Mapping)

DATASET DESCRIPTION

- Multi-state oil & gas production data
- Fields include: State, County, FIPS Code, Commodity, Volume
- Contains missing values, negative/duplicate entries, and inconsistent formatting

DATA CLEANING SUMMARY

Steps performed:

1. State: Replaced blanks/nulls with “Offshore”
2. County: Replaced blanks/nulls with “Offshore”
3. FIPS Code: Replaced blanks/nulls with “Not Applicable”
4. Volume: Decimal type, missing values filled with median, flagged missing & outliers
5. Commodity: Validated & standardized, ready for pivot/aggregation
6. Duplicates: Identified and removed

KEY INSIGHTS

Crude Oil Leads Production

- Crude Oil is the top contributor (718,827 units), followed by Natural Gas (677,287 units) and Condensate (609,257 units)
- Production among the three commodities is relatively balanced

Texas is the Top Hub

- Texas leads in total volume (406,212 units), followed closely by California (389,771 units)

Concentrated County Output

- Production is concentrated in a few key counties: Harris, Cook, and Jefferson contribute the largest shares

Data Integrity Risks

- **Missing Data:** 24% of records (501 entries) lack State information
- **Anomalies:** 53 negative volume entries, likely errors or accounting reversals
- **Tracking Gaps:** 356,143 units labeled under “Unknown” counties, limiting local planning

RECOMMENDATIONS

- **Improve Data Entry Controls:**

Implement mandatory fields for State and County in the reporting system to eliminate “Unknown” values and reduce missing data issues.

- **Target High-Production Areas:**

Since Harris and Cook counties show the highest production volumes, prioritize infrastructure upgrades and safety inspections in these regions to minimize operational downtime.

- **Standardize FIPS Codes:**

Ensure all records contain valid FIPS codes to enable consistent geographic analysis.