



DEPARTAMENTO
DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

Trabajo Práctico 02

Informe Gerencial

Calidad de Datos

Grupo : LutSaRod

Integrante	LU	Correo electrónico
Said, Tomás Uriel	170/23	saidtomasur@gmail.com
Tag, Lucio	876/22	luciotag2011@gmail.com
Mizrahi, Rafael	282/22	rafamizrahi30@gmail.com



Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

Ciudad Universitaria - (Pabellón I/Planta Baja)

Intendente Güiraldes 2610 - C1428EGA

Ciudad Autónoma de Buenos Aires - Rep. Argentina

Tel/Fax: (+54 +11) 4576-3300

<http://www.exactas.uba.ar>

Resumen

Este informe resume el análisis realizado sobre los datos de producción de petróleo y gas en Argentina, con foco en la calidad de los datos, la detección de anomalías y la consistencia entre los pozos convencionales y no convencionales. El objetivo principal fue identificar y corregir problemas que afectan la confiabilidad del dataset y proponer reglas de validación para análisis futuros.

Decisiones y Cambios Claves

- **Limpieza de datos:** Se detectaron y corrigieron valores negativos en las producciones reportadas y registros sin correspondencia geográfica válida, garantizando la coherencia y exactitud de los datos.
- **Validación de clasificación de pozos:** Identificamos intersecciones significativas entre pozos clasificados como convencionales y no convencionales. Esta duplicación genera valores de producción inflados, lo que podría llevar a interpretaciones erróneas de la producción total.
- **Detección de anomalías:** Aplicamos métodos como K-medias e *Isolation Forest* para identificar casos con comportamientos atípicos en la producción. Estas técnicas permitieron marcar pozos con incrementos de producción inesperados o inconsistencias en su registro temporal.
- **Análisis geoespacial:** Utilizamos datos geográficos para validar que los pozos están correctamente ubicados dentro de Argentina y vinculados a provincias y áreas de producción relevantes.

Datos Claves

- Se registró una producción total de petróleo de **33,313,653.11 litros** en pozos convencionales y **18,630,057.34 litros** en pozos no convencionales. Sin embargo, la producción total ajustada (evitando duplicados) fue de **10,123,474.34 litros**, destacando la relevancia de tratar correctamente las intersecciones.
- En el caso del gas, la producción total de pozos convencionales fue de **42,317,680.84 Cf**, mientras que la de pozos no convencionales alcanzó los **27,284,972.41 Cf**. Nuevamente, se detectaron duplicaciones significativas entre ambos tipos de pozos.
- Más de **500 registros** iniciales parecían estar fuera de los límites argentinos, pero un análisis visual confirmó que estaban correctamente ubicados dentro del país.

Conclusiones

Este análisis dejó en evidencia la importancia de manejar adecuadamente las intersecciones entre pozos convencionales y no convencionales para evitar inflar los valores de producción. También resaltó la relevancia de validar los datos geográficos y corregir inconsistencias en los registros temporales y clasificaciones.

Los métodos de detección de anomalías empleados permitieron identificar patrones atípicos en la producción, brindando una base sólida para futuras investigaciones y políticas de mejora de la calidad de los datos. Estos hallazgos son fundamentales para garantizar análisis confiables en un sector estratégico para el país y mejorar la toma de decisiones en el ámbito energético.