

## Aplicación DRILLING OPTIMIZATION

El objetivo de esta aplicación es mejorar el proceso de toma de decisiones en operaciones de perforación mediante el uso de herramientas de datascience, con el fin de:

- Reducir/Eliminar el ILT (invisible lost time), que se estima en 15% del tiempo total de drilling.
- Optimizar ROP (rate of penetration).
- Identificar patrones/tendencias a fin de minimizar el NPT (non productive time), planteando la operación de drilling como un todo (integrar datos del rig, formaciones, perfiles, antecedentes)

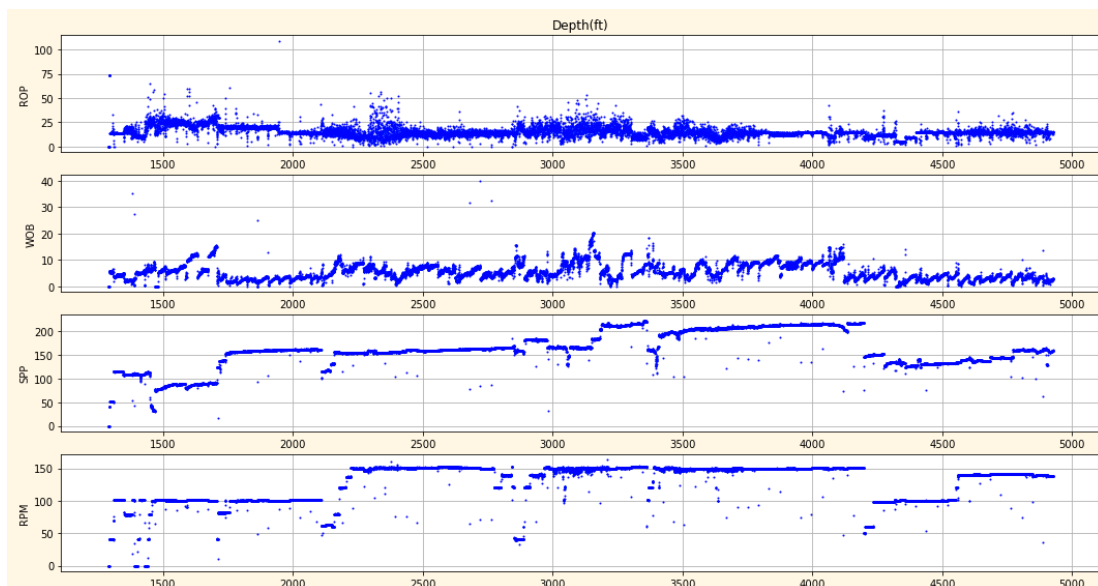
Para plantear la resolución de la oportunidad planteada, se utilizaron datos del repositorio público de datascience:

<https://www.kaggle.com/datasets/musaabyazan/mu-1x-drilling-data>

Dicho dataset corresponde a los parámetros típicos de perforación registrados en el pozo MU-1X, con profundidades comprendidas entre los 1,290 ft a 4,927 ft.

Los parámetros registrados corresponden las lecturas de ROP, WOB, RPM, GR entre otras, la estadística asociada al set de datos se muestra de esta manera:

	ROP	WOB	SPP	RPM	GR
Size	14,549.000	14,549.000	14,549.000	14,549.000	14,549.000
Min	0.000	0.000	0.000	0.000	4.300
Max	109.300	40.100	223.620	163.000	164.750
Mean	15.500	5.566	161.570	126.669	52.545
Std	5.778	3.058	37.801	30.393	30.796
Skew	2.224	1.141	-0.485	-1.311	-0.066
Kurtosis	16.516	3.484	0.314	1.682	-0.419
Percentile10	9.800	2.200	110.352	99.000	8.900
Percentile50	14.700	5.000	160.370	142.000	62.800
Percentile90	22.400	9.600	213.160	151.000	79.100



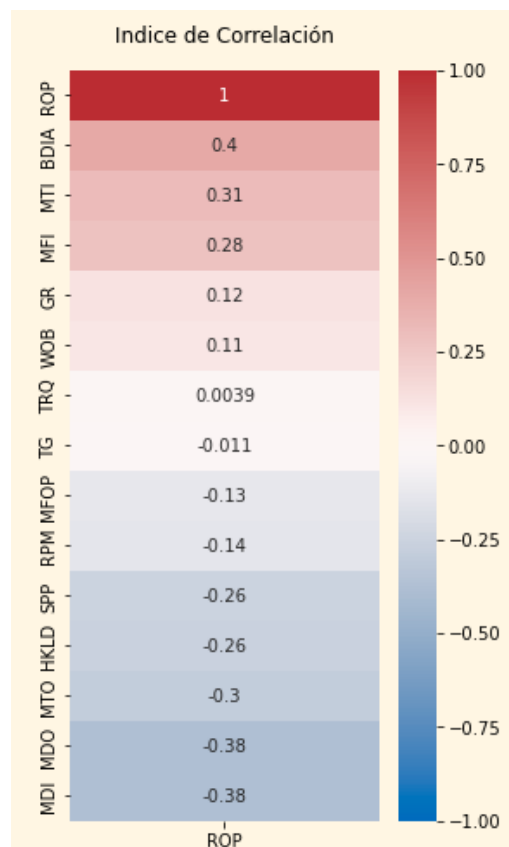
## Exploración de datos

En este apartado nos proponemos inspeccionar los datos disponibles, entender de qué manera están representados, cómo se distribuyen y de qué manera se relacionan.

Además, con la ayuda del heatmap de correlaciones comenzaremos a entender mejor de qué manera las distintas variables del proceso se relacionan entre sí, con el objetivo de las posibles conexiones que existientes entre las mismas.

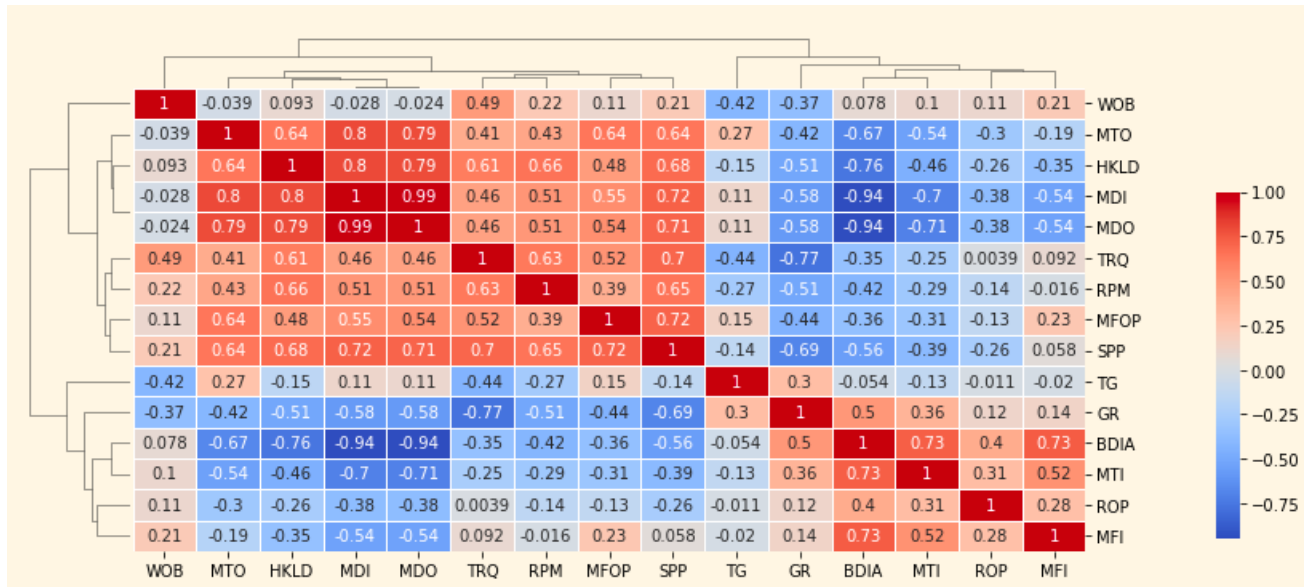
Como corolario planteamos el grado de correlación que existe entre todas las variables respecto al ROP, con algunos puntos a remarcar:

- GR (gamma ray) exhibe un grado de correlación menor, positivo, que insinúa la relevancia de la formación perforada.
- HKLD (hook load) evidencia un grado de correlación negativa, indicando a mayor ROP, menor carga en el hook.
- TRQ (torque) expresa un grado de correlación inexistente con ROP.



También es interesante remarcar que algunas consideraciones observadas en el dendograma del heat map:

- WOB (weight on bit) y ROP se presentan en extremos opuestos.
- TG (total gas) y GR se agrupan en un mismo cluster, manifestando el mismo origen entre ellos (variables de formación)

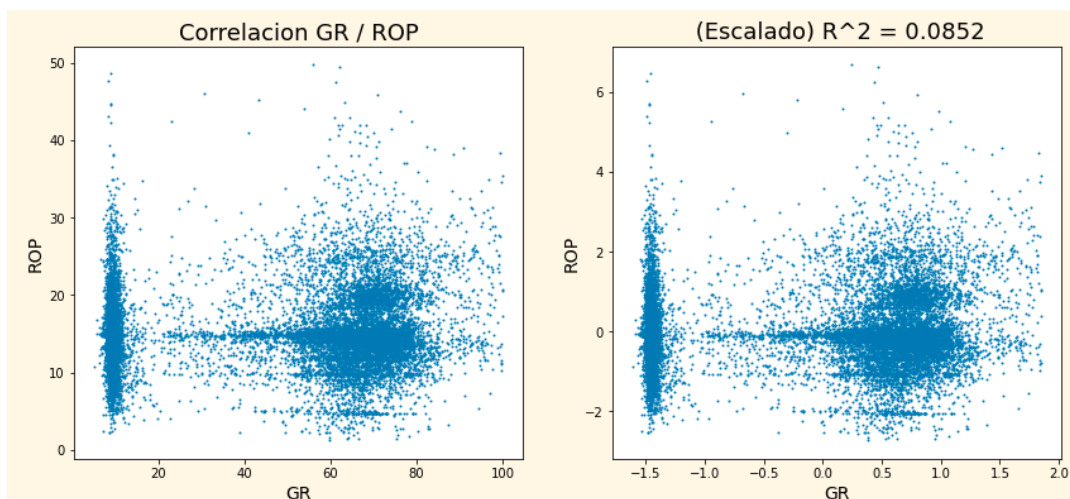


## Modelización I – Algoritmos No Supervisados

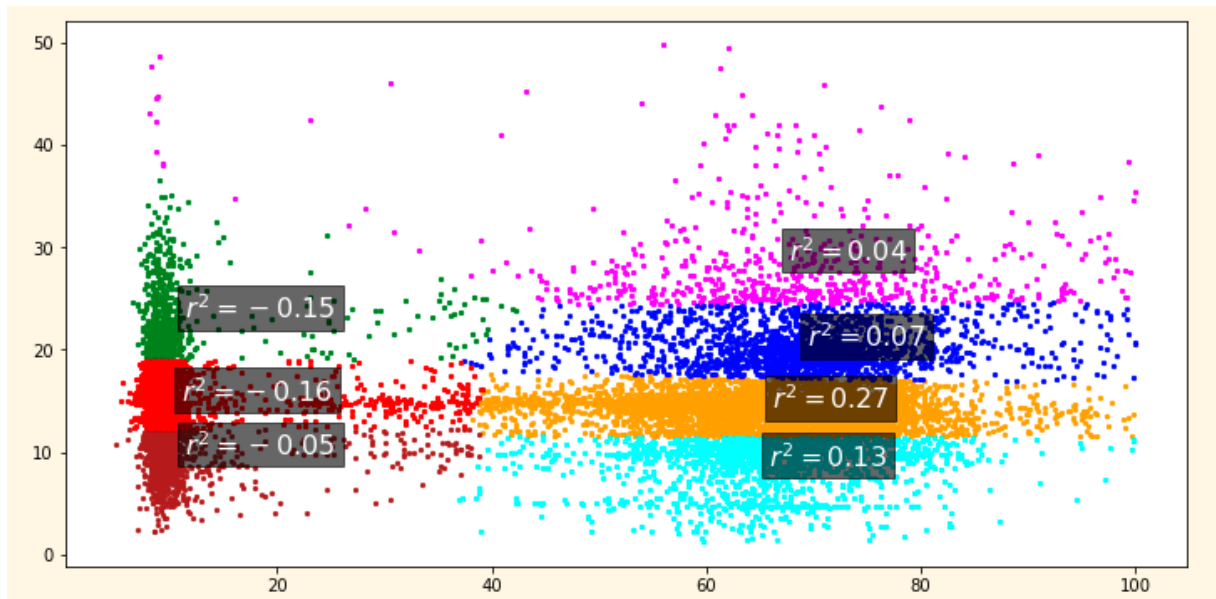
En esta etapa se focalizará principalmente en la aplicación de algoritmos de clustering no supervisados, de manera de identificar posibles patterns ocultos que al momento del análisis han permanecido desapercibidos.

En este ejemplo se intentará identificar alguna posible correlación entre GR (gamma ray) y ROP (rate of penetration), a fin de detectar algún posible comportamiento entre dichas variables.

La correlación existente entre GR / ROP es mínima (0.08) indicando que la tasa de penetración es transparente a la presencia de distintas formaciones, asumiendo que el GR es un indicativo de esta.



No obstante, si agrupamos en 7 clusters, podemos calcular nuevamente el índice de correlación, revelando dos grupos bien diferenciados (los de correlación positiva y los de correlación negativa) que están asociadas a las dos formaciones presentadas en el GR; además, el cluster 2 es el de mayor correlación, resultando ser:



- Cluster 0 – índice de correlación +0.04
- Cluster 1 – índice de correlación +0.07
- Cluster 2 – índice de correlación +0.27
- Cluster 3 – índice de correlación +0.13
- Cluster 4 – índice de correlación -0.15
- Cluster 5 – índice de correlación -0.16.
- Cluster 6 – índice de correlación -0.05

### Modelización II – Algoritmos Supervisados

En esta etapa nos focalizaremos en el uso de algoritmos supervisados para la optimización de los parámetros de perforación, los cuales estarán orientados a maximizar ROP.

Respecto al proceso de optimización, podemos mencionar que el máximo ROP podría ser alcanzado exigiendo al límite la capacidad del equipo y su conjunto, sin embargo, podremos setear constraints a fin de replicar las condiciones operativas lo más real posible.

La optimización del ROP será en función de la formación a perforar, para esto podemos asumir:

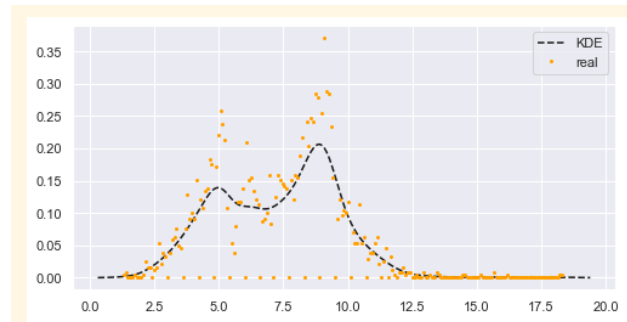
- indexo por depth, asumiendo que depth es representativo de la formación (alternativa implementada)
- indexo por clustering detectado sobre cada formación, y a partir de ahí conglomerado.
- indexo por GR o cambios en el mismo.

Definimos intervalos para modelar en base a la lectura de GR:

- intervalo1=1800:2800 ft
- intervalo2=3300:4000 ft
- intervalo3=4200:4900 ft

Para la simulación de los intervalos, consideramos lo siguiente:

- WOB: se distribuye según parámetros observados

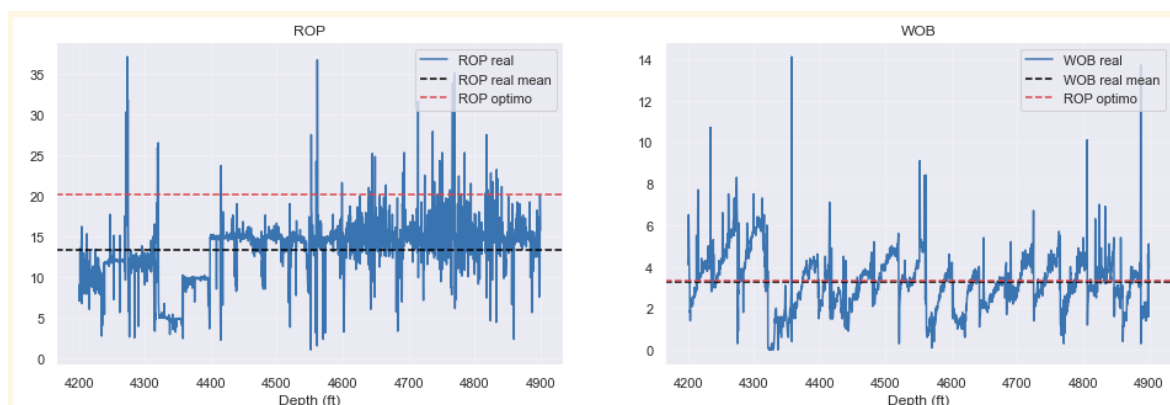


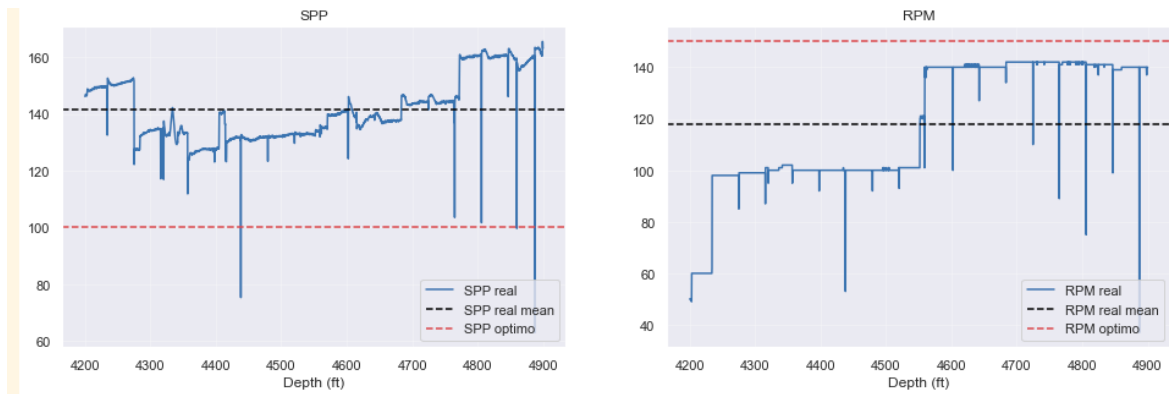
- RPM: se distribuye en 2 valores discretos: 100-150
- SPP : se distribuye en 3 valores discretos: 100-150-200

Para la simulación del intervalo3, los resultados son los siguientes:

ROP mean agrupado por RPM y SPP		
RPM	SPP	
100	100	18.909778
	150	14.352976
	200	17.428553
150	100	20.121373
	150	14.323100
	200	14.802950

Lo que indica que el ROP máximo de 20.12 se obtiene configurando un RPM de 150 y un SPP de 100.





## Conclusiones

Lo interesante de resolver el problema de optimización es que nos permite generar escenarios operativos alternativos, situación que en el experimento real (en campo) no se podría realizar (sesgos, cultura, restricciones, ...), y comprobar/validar cómo se comportaría el sistema en su conjunto.

Para los tres intervalos analizados, resultó ser:

- Intervalo 1800-2800 ft: ROP optimizado varía +15 % (20 ft/ht), esto se lograr reduciendo el SPP de 160 a 100 e incrementando los RPM de 130 a 150.
- Intervalo 3300-4000 ft: ROP optimizado varía +40 % (20 ft/ht), esto se lograr reduciendo el SPP de 2000 a 100 (RPM es constante en 150).
- Intervalo 4200-4900 ft: ROP optimizado varía +40 % (20 ft/ht), esto se lograr reduciendo el SPP de 140 a 100 e incrementando los RPM de 120 a 150.

En líneas generales para los tres intervalos se observa que una reducción del SPP (mantener en valores de 100) mejoraría la performance del ROP; y en menor medida un incremento del RPM (regímenes de operación de 150 RPM aproximadamente)

Los valores WOB (weight on bit) no varían desde del caso real al simulado ya que al ser generados con la misma función distribución, el valor mean es el mismo.

A modo de comentario final, el ejercicio es una implementación simplificada del pipeline de datascience, cuyo objetivo es ilustrar las bondades de dicho análisis.