

Definición del problema de negocio

El problema que se aborda en este proyecto es la baja eficiencia operativa en el ciclo de cargue de los camiones de acarreo de material estéril en la empresa Drummond Ltd., cuya duración mayor a 30 minutos (*valor objetivo*) genera pérdidas significativas de productividad y afecta el indicador clave **Truck Factor (TF)**. De acuerdo a un análisis preliminar, se espera que las paradas improductivas dentro del ciclo, especialmente las mayores a 6 minutos, ocasionan una pérdida mensual cercana a 140 horas de camiones, limitando el número de viajes posibles por jornada y reduciendo la capacidad efectiva de transporte. Por tanto, el desafío central consiste en reducir los tiempos improductivos dentro del ciclo de acarreo mediante el uso de analítica de datos que permita identificar, clasificar y predecir los eventos que prolongan innecesariamente los tiempos de ciclo.

Justificación y relevancia del problema

El **Truck Factor (TF)** mide la eficiencia con la que la flota convierte el tiempo disponible en producción efectiva. Una mejora en este indicador tiene un impacto directo en la rentabilidad de la operación, ya que permite aumentar la producción sin necesidad de ampliar la flota.

Reducir el ciclo promedio en un 5 % (**de 30 a 28,5 minutos**) implica un beneficio económico estimado de **30.000 USD** mensuales, equivalente a un aumento de la capacidad efectiva de acarreo de material estéril.

El problema es además prioritario porque:

- Representa más del 60 % del costo operativo total de la mina
- Tiene datos medibles, continuos y confiables provenientes de sistemas corporativos (VisionLink de Caterpillar y Dispatch).
- Se alinea con los objetivos estratégicos corporativos: productividad, sostenibilidad, seguridad y transformación digital.

Modelos propuestos

- Modelo A (Regresión): Predecir eventos de en los que los camiones se detienen un tiempo ≥ 6 o ≥ 10 minutos, generando alertas tempranas para reducir pérdidas. (En este modelo consideraremos los tiempos menores a 6 minutos como condiciones normales del proceso).
- Modelo B (clasificación): etiquetar ciclos o rutas como *eficientes/no eficientes* para monitoreo.

- Modelo C (No supervisado/geoespacial): identificar rutas y turnos con mayor tiempo detenido, priorizando intervenciones operativas.

Pregunta de negocio

¿Cómo puede la analítica de datos optimizar el ciclo de cargue de los camiones de roca para reducir las paradas improductivas y mejorar el Truck Factor (TF) en la operación minera?

Alcance del proyecto

Operativo

- Cobertura inicial: 4 camiones y 1 mes de datos de VisionLink.
- Variables base: tiempos por fase (Empty, Load, Travel), paradas por rango, distancia, carga útil, combustible, turno y ruta.
- Áreas involucradas: Operaciones, Despacho, Mantenimiento de Vías, Planeación y Mejora Continua

Analítico

- Se desea finalizar los 3 modelos planteados anteriormente.

Organizacional

- Integración con los stakeholders: validación continua con Operaciones, Despacho y Mejora Continua para asegurar la aplicabilidad de los resultados.
- Entregable final: tablero analítico que muestre rutas críticas, eficiencia por turno y alertas de paradas prolongadas.
- Escalabilidad: posterior extensión a toda la flota de acarreo.

Valor esperado

- Reducción del 20 % de horas en paradas ≥ 6 min.
- Reducción del 5 % en tiempo de ciclo promedio (de 30 a 28.5 min).
- Incremento del TF en +3 puntos porcentuales (MTD).
- Aumento del 15 % en la proporción de ciclos eficientes.

Descripción del conjunto de datos

Fuente

Los datos provienen del sistema VisionLink (Caterpillar), plataforma corporativa utilizada para el monitoreo telemático de la flota de camiones en la operación minera.

Este sistema registra, minuto a minuto, variables relacionadas con el ciclo de acarreo y cargue de los camiones de roca, incluyendo tiempos de desplazamiento, paradas, carga útil, consumo de combustible y distancias recorridas

Estructura de los datos

Variable	Descripción	Tipo de dato
Equipment	Identificador del camión (ej. M02782)	Categórica
Date and Time	Fecha y hora del ciclo	Temporal
Payload Weight (t)	Peso del material transportado	Numérica (continua)
Fuel Burned (L)	Combustible consumido durante el ciclo	Numérica (continua)
Total Cycle Time (hh:mm:ss)	Duración total del ciclo de acarreo	Temporal
Total Distance Travelled (km)	Distancia total recorrida	Numérica
Empty Travel Time (hh:mm:ss)	Tiempo de desplazamiento con camión vacío	Temporal
Empty Travel Distance (km)	Distancia recorrida en vacío	Numérica
Empty Stop Time (hh:mm:ss)	Tiempo detenido con camión vacío	Temporal
Load Time (hh:mm:ss)	Duración del proceso de cargue	Temporal
Load Stop Time (hh:mm:ss)	Tiempo detenido durante el cargue	Temporal
Load Travel Time (hh:mm:ss)	Tiempo de desplazamiento cargado	Temporal
Load Travel Distance (km)	Distancia recorrida cargado	Numérica
Start Latitude, Longitude	Coordenadas del punto de inicio del ciclo	Geográfica
Destination Latitude, Longitude	Coordenadas del destino del ciclo	Geográfica

Observaciones:

- Cada registro corresponde a un ciclo completo de acarreo (inicio-fin) ejecutado por un camión determinado.
- Los datos incluyen tiempos, distancias, consumo y coordenadas, permitiendo una visión integral del rendimiento operativo.

- Ausencia de coordenadas GPS específicas de las paradas (solo coordenadas inicio–fin).
- Clasificar los ciclos en eficientes/no eficientes
- Predecir la probabilidad de paradas largas (≥ 6 min).

Nota Importante: El análisis exploratorio de los datos se encuentra en el notebook realizado y que está en el repositorio del proyecto, allí se puede ver todo el EDA realizado. El notebook requiere de un archivo “tiempos.xlsx” el cual no se encuentra en el repositorio en remoto porque se le realiza el tracking al .dvc de la data, por tal razón, el notebook tiene todas las celdas ejecutadas.

Control de versiones con Git y DVC

Durante el desarrollo del proyecto Proyecto_DSA, se implementó un esquema de control de versiones tanto para el código como para los datos, utilizando Git y DVC (Data Version Control).

Esta configuración permite mantener un historial ordenado de los cambios realizados, garantizar la trazabilidad de los datos y facilitar la colaboración y reproducibilidad de los análisis.

Git – Control de versiones del código

El repositorio del proyecto fue inicializado con Git y sincronizado con un repositorio remoto en GitHub, lo que permite registrar y gestionar cada cambio en el código fuente, notebooks, scripts y configuraciones.

Se estableció una estructura de carpetas organizada (data/, notebooks/, src/, models/) y se incorporó un archivo .gitignore para evitar el rastreo de archivos temporales, entornos virtuales y datos pesados.

De esta manera, Git almacena únicamente el código relevante, manteniendo el repositorio limpio y eficiente.

DVC – Control de versiones de los datos

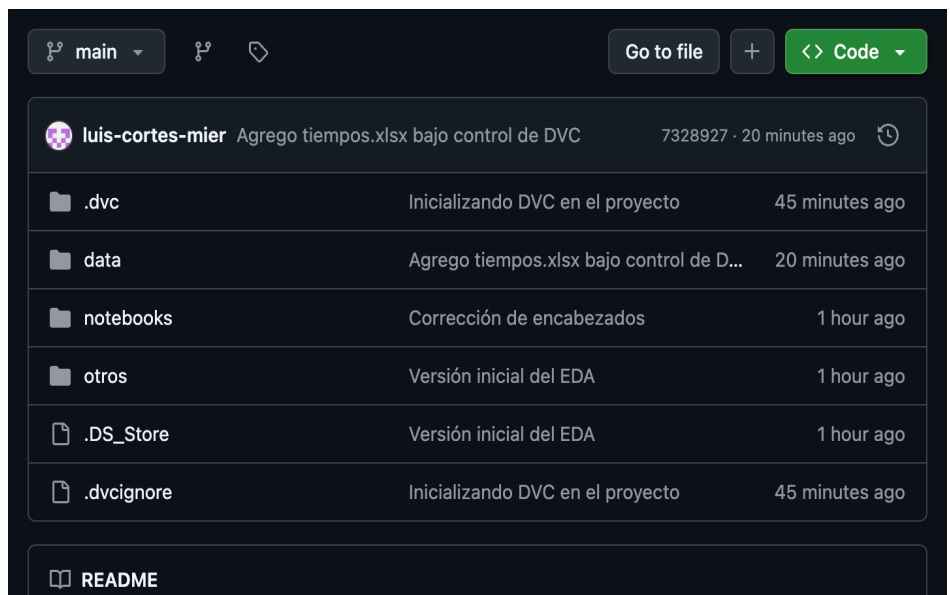
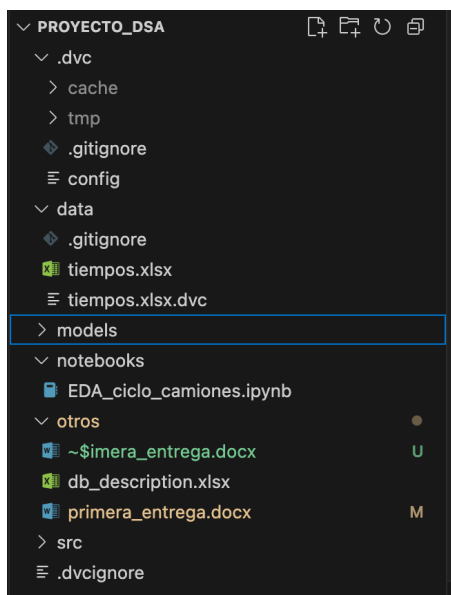
Para complementar a Git y manejar los datos se configuró DVC.

El archivo tiempos.xlsx fue agregado bajo control de DVC, generando tiempos.xlsx.dvc, el cual contiene la información necesaria para reconstruir o recuperar el dataset cuando sea necesario.

El Excel real fue excluido del rastreo de Git, asegurando que el repositorio permanezca liviano y evitando conflictos o duplicaciones de datos

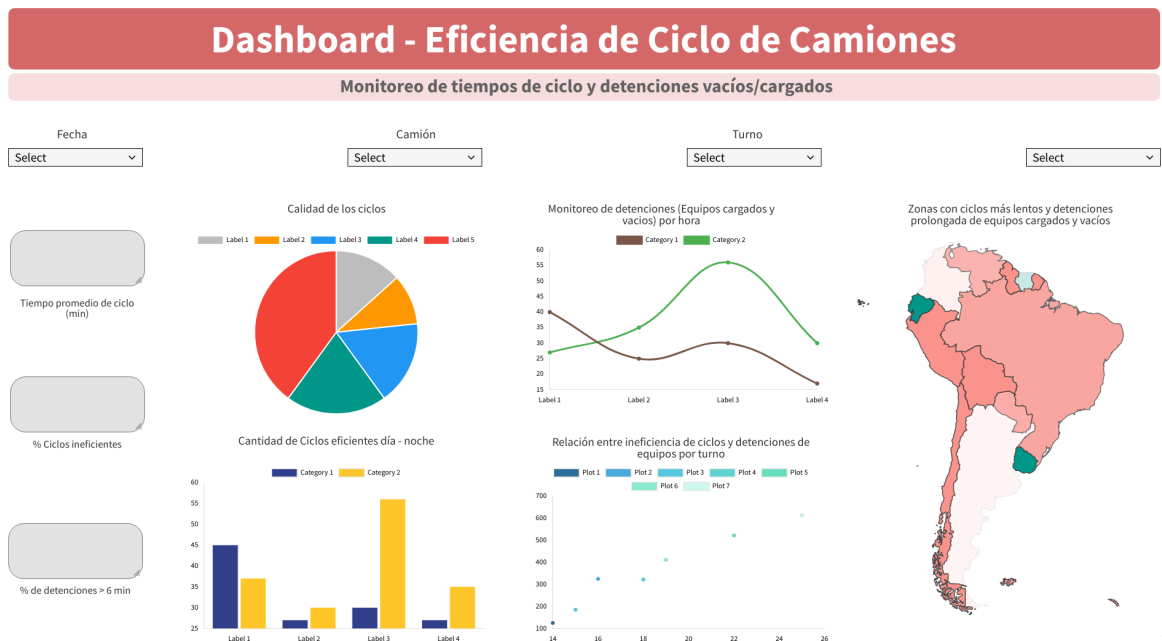
El flujo de trabajo verificado fue el siguiente:

- Inicialización de Git y conexión con GitHub (git init, git add, git commit, git push).
- Inicialización de DVC y adición del dataset (dvc init, dvc add data/tiempos.xlsx).
- Verificación del estado (git status y dvc status) confirmando que el proyecto se encuentra sincronizado.
- DVC permite restaurar los datos en cualquier momento mediante dvc pull, garantizando la reproducibilidad total del proyecto.



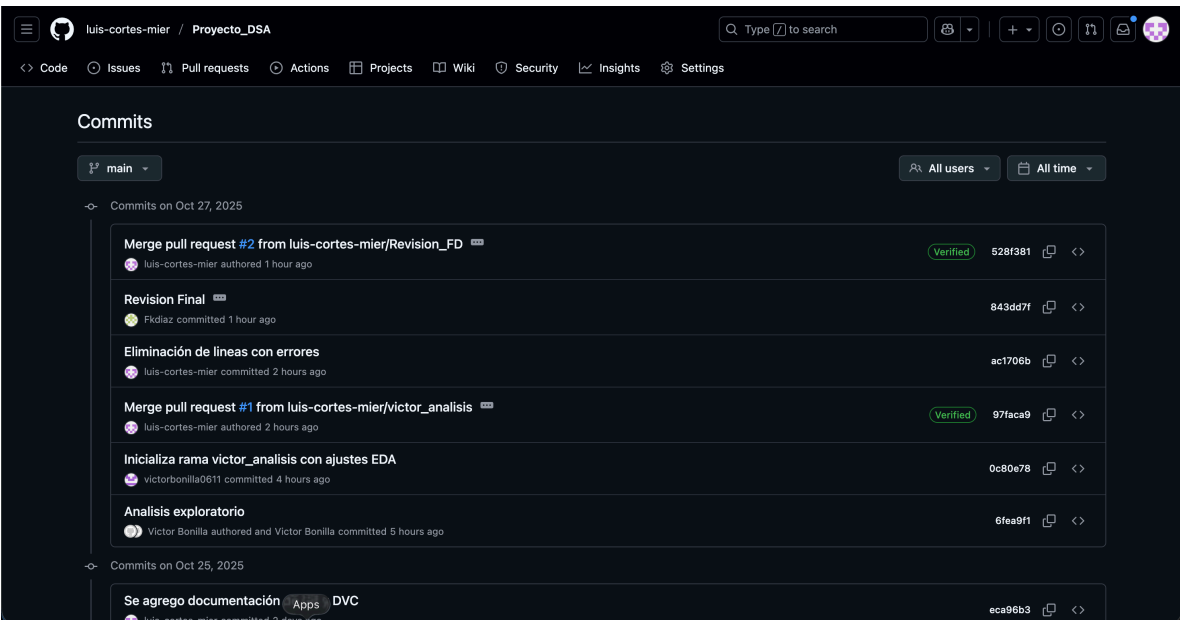
Link del repositorio: https://github.com/luis-cortes-mier/Proyecto_DSA.git

Primera versión del mockup



Trabajo Colaborativo

A continuación, se puede ver los pull request realizado por los compañeros del equipo del proyecto.



Roles de cada uno:

Nuestro equipo está conformado por tres integrantes, y aunque cada uno tiene un enfoque principal, todos participamos activamente en las diferentes etapas del proyecto, compartiendo responsabilidades y apoyándonos mutuamente para lograr los objetivos planteados.

- Luis Cortes, se encarga del manejo de todo el entorno de MLOps. Esto incluye garantizar que el proyecto tenga un flujo de trabajo colaborativo, con un adecuado control de versiones del código y de los datos. Para ello, implementé herramientas como Git, GitHub y DVC, que permiten mantener el trabajo organizado y reproducible.
- Víctor Bonilla es el principal responsable del modelado de las soluciones propuestas. Él desarrolla y prueba diferentes enfoques analíticos y modelos que permitan dar respuesta efectiva a la pregunta de negocio, asegurando que las soluciones sean técnicamente sólidas y alineadas con los objetivos del proyecto.
- Por su parte, Frank Díaz lidera el diseño y construcción del tablero final (dashboard), que tiene como propósito presentar de manera clara e intuitiva los resultados obtenidos, facilitando la interpretación y la toma de decisiones por parte de los usuarios finales.

A pesar de estas divisiones, todos trabajamos de manera conjunta en cada fase: desde la exploración de los datos hasta la evaluación de resultados. Esta colaboración constante nos ha permitido mantener coherencia en el desarrollo del proyecto.