

Ejercicio:

Amelia Herrera Briceño, [melinnicri@gmail.com](mailto:melinnicri@gmail.com),

Noviembre, 2024.



### Descripción del Problema

Se dispone de una base de datos que registra el rendimiento diario de diferentes palas y camiones de transporte de mineral. La información con la que se cuenta es la **cantidad de tonelaje movido durante un día de operación y el número de paladas**.

El objetivo principal es **identificar los equipos de mejor rendimiento en términos de tonelaje extraído por día, así como los principales factores que influyen en esta métrica**.

## Análisis Estadístico Descriptivo:

### Datos Iniciales:

Camiones: 47

Palas: 4

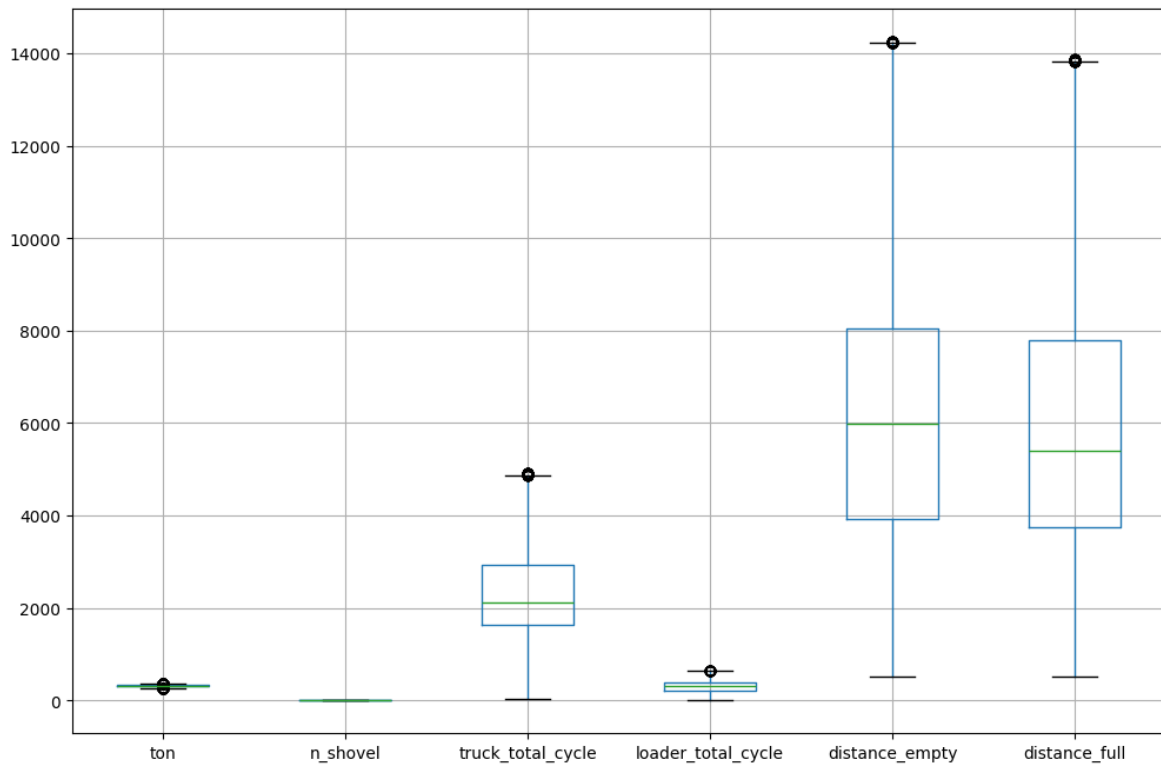
Datos nulos: Ninguno

Datos faltantes: Ninguno

Datos atípicos: Sí, pero resultaron ser menos del 1% y fueron eliminados. Estos no desvían significativamente los promedios, en un principio.

### Gráfico 1: Caja con Bigotes (Variabilidad)

**Gráfico 1:** gráfico con bigotes de los datos iniciales:



**Abreviaturas:** ton: toneladas de material con que se carga un camión; n\_shovel: número de paladas con que se llenan un camión; truck\_total\_cycle: ciclo total del camión (en segundos); loader\_total\_cycle: ciclo total de paladas (en segundos); distance\_empty: distancia recorrida de los camiones vacíos (en metros); distance\_full: distancia recorrida de los camiones llenos (en metros).

**Abreviaturas:**

*ton*: Toneladas de material con que se carga un camión

*n\_shovel*: Número de paladas con que se llena un camión

*truck\_total\_cycle*: Ciclo total del camión (en segundos)

*loader\_total\_cycle*: Ciclo total de paladas (en segundos)

*distance\_empty*: Distancia recorrida de los camiones vacíos (en metros)

*distance\_full*: Distancia recorrida de los camiones llenos (en metros)

**Estadísticas Descriptivas Sin Valores Atípicos:**

**Registros totales:** 385 181 (aproximadamente 2.7% de datos eliminados, de 395 680

Registros iniciales).

Medias: Muy cercanas a las originales.

Desviaciones estándar: Ligeramente disminuidas.

Valores mínimos y máximos: Más concentrados.

**Interpretación:**

Rendimiento General: Más consistente y predecible sin outliers.

Decisiones Basadas en Datos: Menos susceptibles a variaciones extremas.

Visualización y Reportes: Más claros y representativos.

**Requerimientos del Análisis:**

Los Ránking de Eficiencia por cada nueva variable, se pueden verificar tanto en el notebook (**examen.ipynb**) como en el **Dashboard, Tabla 1**, con date = fecha(Enero-2023 a Mayo 2024); truck = camión; loader = pala; eficiencia\_camion = eficiencia de camión; eficiencia\_pala = eficiencia de pala; tonelaje\_por\_paladas = tonelaje por paladas; eficiencia\_distancia = eficiencia de distancia.

**Ranking de Desempeño:** Elaborar un ranking del desempeño de los equipos con base en su eficiencia en el transporte de tonelaje.

**Cálculo de Eficiencias (sin considerar outliers):**

1.- Eficiencia del Camión:  $\text{ton} / \text{truck\_total\_cycle}$

2.- Eficiencia de Pala:  $\text{ton} / \text{loader\_total\_cycle}$

3.- Productividad de Operación de Carga:  $\text{ton} / n_{\text{shovel}}$

4.- Eficiencia por Distancia:  $\text{distance\_full} / \text{distance\_empty}$

**Tabla 1: Ranking de Eficiencias:**

Ranking	Eficiencia de Camión	Eficiencia de Palas	Tonelaje por Paladas	Eficiencia por Distancia (mayor a 1)	Eficiencia por Distancia (menor a 1)
	0.1639	1.1673	102.5775	1.4430	0.8934

**Interpretaciones:**

Eficiencia de Camión (0.1639): Margen para mejorar.

Eficiencia de Pala (1.1673): Buen uso de palas.

Eficiencia Tonelaje por Paladas (102.5775): Buen rendimiento de carga.

Eficiencia por Distancia (mayor a 1) (1.4430): Optimización en la entrega de material.

Eficiencia por Distancia (menor a 1) (0.8934): Optimización en el retorno.

**Acciones que tomar:**

Optimizar rutas de camiones.

Mantener el buen uso de palas.

Aumentar la productividad de carga.

Continuar optimizando la eficiencia de distancia.

Monitorear y corregir ineficiencias.

Mejorar prácticas operativas y mantenimiento.

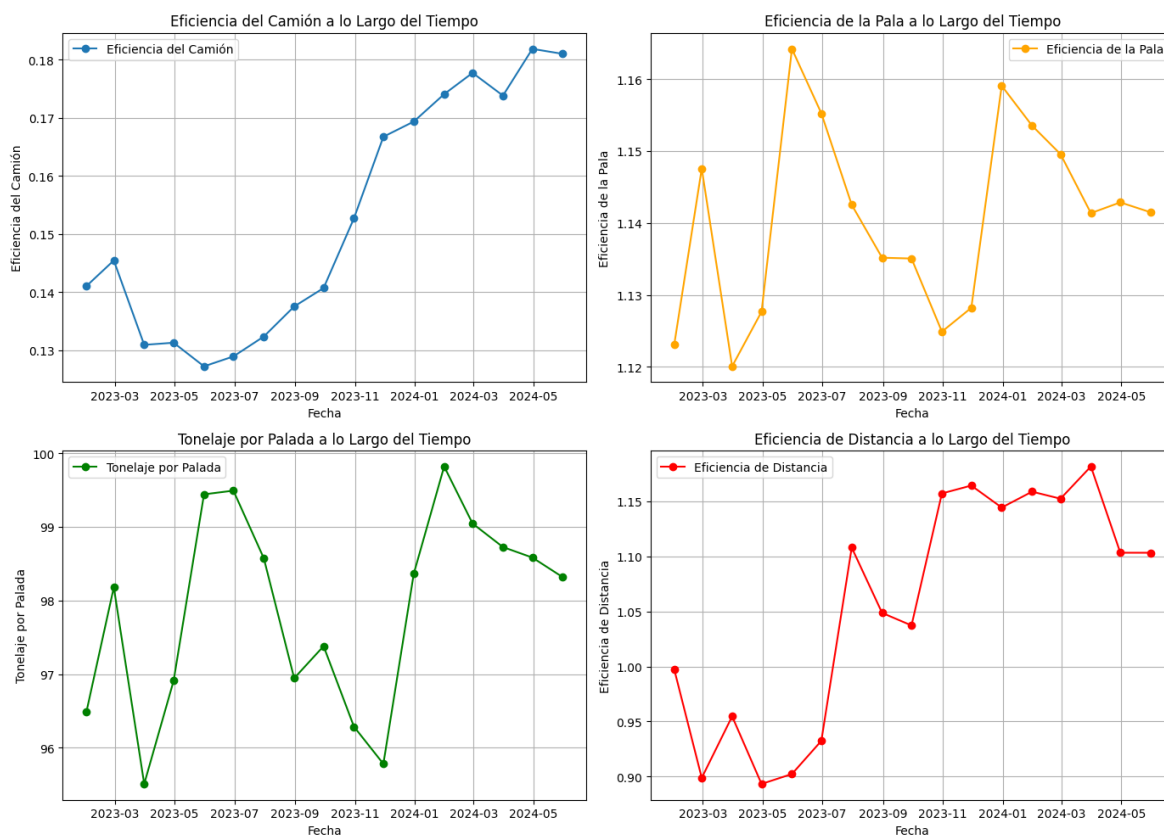
Capacitación para operadores.

### Identificación de Factores Críticos:

Determinación de las variables que más influyen en el rendimiento de los equipos mediante una matriz de correlación.

En cuanto a factores críticos, se realizó una matriz de correlación, donde se cruzaron las variables, como se muestran en el **Gráfico 2**:

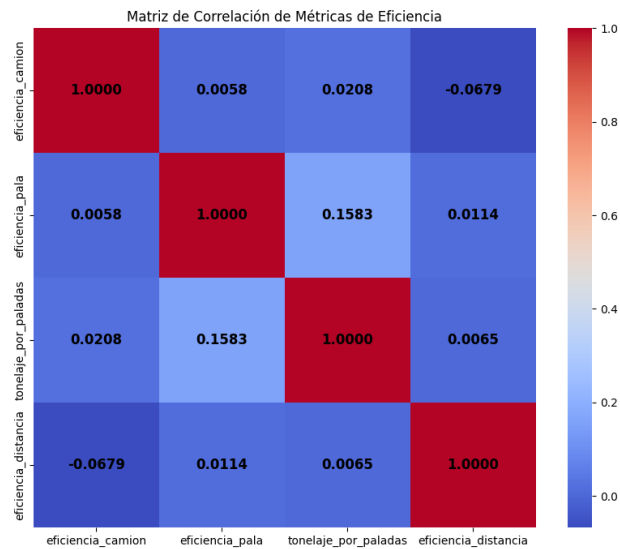
**Gráfico 2:** Variables en el tiempo (Enero del 2023 – Mayo del 2024)



Eficiencia de camión, Eficiencia de pala, Tonelaje por palada, Eficiencia de distancia, entre Enero del 2023 a Mayo del 2024.

El **Gráfico 3**, muestra los resultados de la Matriz de correlación de las variables mostradas anteriormente.

**Gráfico 3, Matriz de correlación:**



Eficiencia de camión, Eficiencia de pala, Tonelaje por palada, Eficiencia de distancia, entre Enero del 2023 a Mayo del 2024.

De acuerdo con los resultados, de la Matriz de correlación, en general, hay una baja o débil correlación en general.

#### **Eficiencia del Camión:**

- **Con Eficiencia de la Pala (0.005791):** Hay una correlación muy débil y positiva entre la eficiencia del camión y la eficiencia de la pala. Esto sugiere que no hay una relación significativa entre estas dos métricas.
- **Con Tonelaje por Palada (0.020797):** También hay una correlación muy débil y positiva con el tonelaje por palada, indicando nuevamente una relación insignificante.
- **Con Eficiencia de Distancia (-0.067893):** La correlación es negativa y muy débil, indicando una relación muy pequeña y negativa entre la eficiencia del camión y la eficiencia de distancia.

### **Eficiencia de la Pala:**

- **Con Tonelaje por Palada (0.158310):** Hay una correlación positiva débil entre la eficiencia de la pala y el tonelaje por palada. Esto sugiere que a medida que aumenta el tonelaje por palada, la eficiencia de la pala también tiende a aumentar, aunque esta relación es débil.
- **Con Eficiencia de Distancia (0.011406):** Hay una correlación muy débil y positiva, lo que indica prácticamente ninguna relación significativa entre estas métricas.

### **Tonelaje por Palada:**

- **Con Eficiencia de Distancia (0.006519):** La correlación es extremadamente débil y positiva, indicando casi ninguna relación entre el tonelaje por palada y la eficiencia de distancia.

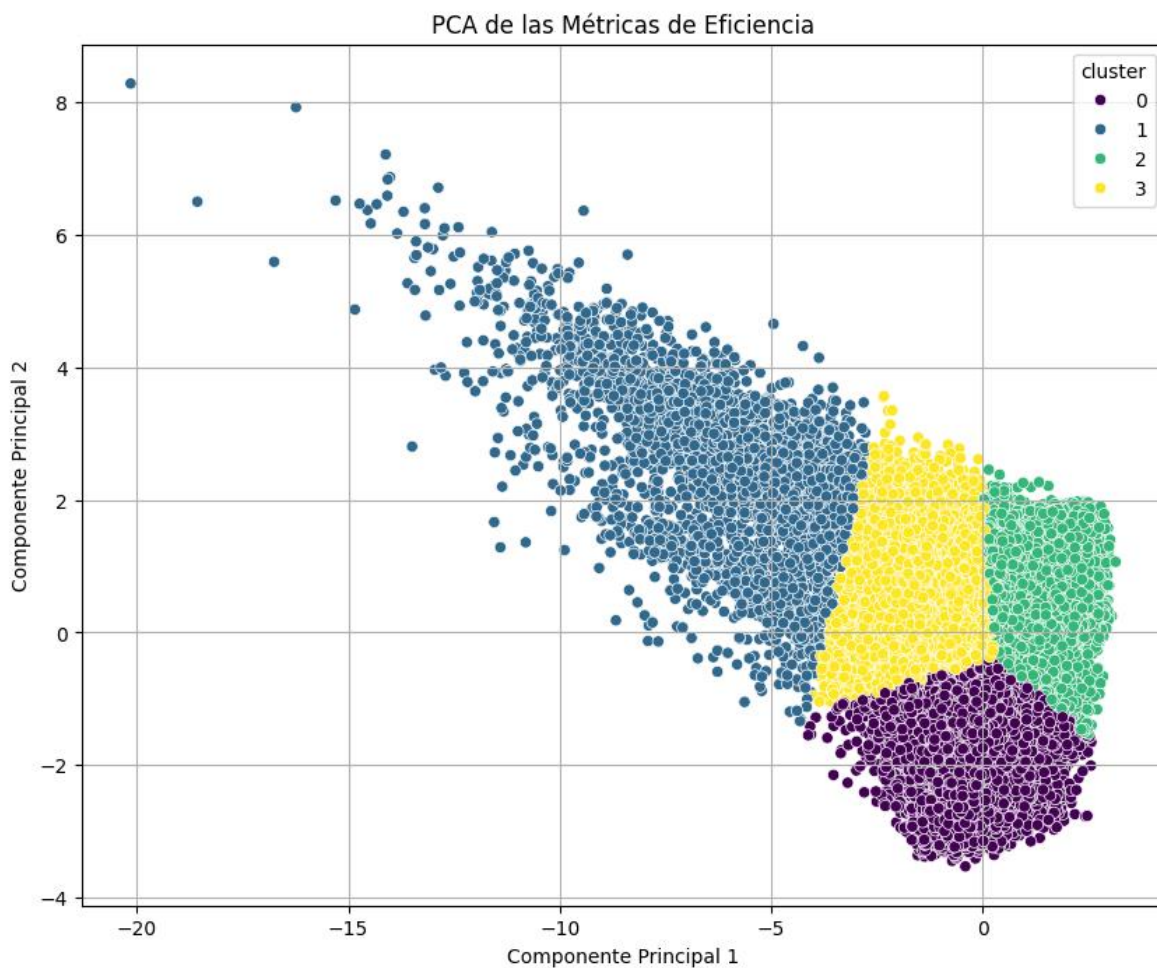
### **Eficiencia de Distancia:**

- Como se mencionó anteriormente, todas las correlaciones con la eficiencia de distancia son muy débiles, indicando poca o ninguna relación significativa con las otras métricas.

### **Conclusiones Generales:**

- **Correlaciones Muy Débiles:** Las correlaciones son en su mayoría muy débiles, lo que sugiere que las métricas no están fuertemente relacionadas entre sí. Esto podría indicar que cada métrica captura diferentes aspectos de la eficiencia operativa y no están directamente influidas por las otras métricas.
- **Relación Débil:** La única correlación que muestra una relación débil es entre la eficiencia de la pala y el tonelaje por palada, lo que podría indicar que mejorar el tonelaje por palada puede tener un pequeño impacto positivo en la eficiencia de la pala.

**Como la correlación era débil, se aplicó Modelo PCA por separado** para camiones y para palas, relacionando las otras variables, Eficiencia de camión, Eficiencia de pala, Tonelaje por palada, Eficiencia de distancia, entre Enero del 2023 a Mayo del 2024, en 4 clústers, tal como muestra el **Gráfico 4:**

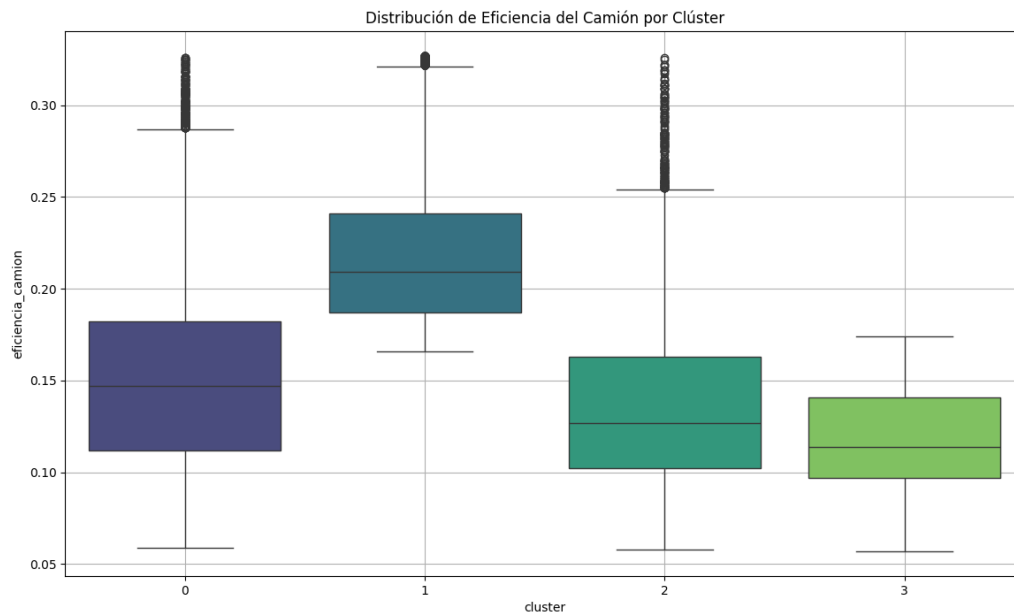
**Gráfico 4:** PCA de dos componentes.

La separación visual entre los clústeres indica que el PCA ha logrado reducir la dimensionalidad de los datos mientras preserva la variabilidad y las relaciones importantes. Cada clúster representa un grupo de camiones y palas con patrones de eficiencia similares.

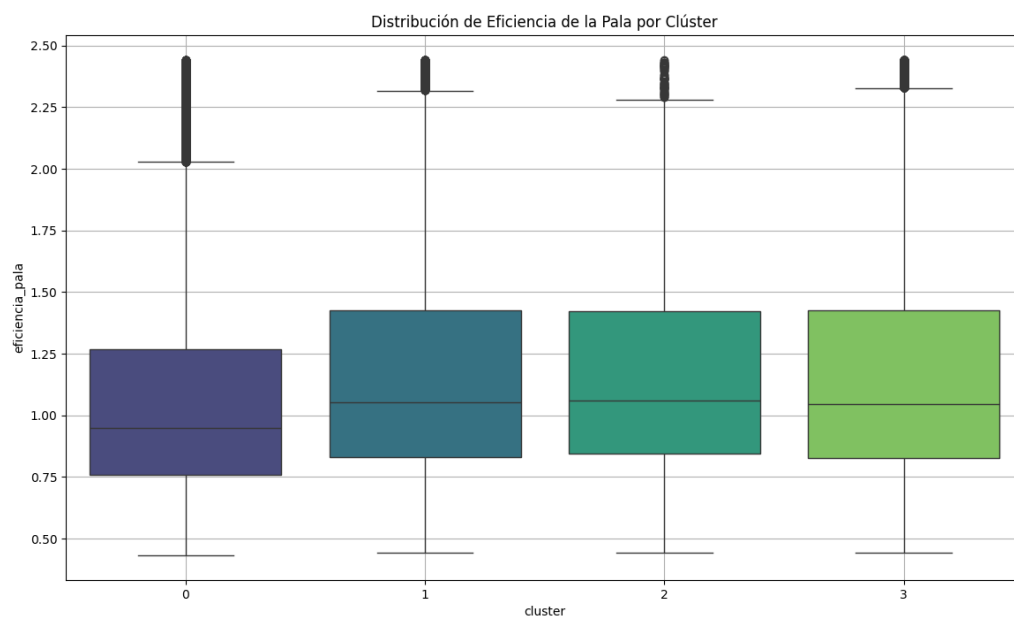


En las siguientes gráficas (**Gráfico 5** y **Gráfico 6**) se puede visualizar la distribución de los puntos en los componentes principales mostrando cómo los diferentes clústeres se relacionan con las métricas de eficiencia. Los clústeres están bien diferenciados, lo que sugiere que hay diferencias claras en las eficiencias operativas de los grupos.

**Gráfico 5:** Distribución de Eficiencia de camión por clúster.



**Gráfico 6:** Distribución de Eficiencia de pala por clúster.



### **Interpretación del Gráfico PCA Componentes Principales:**

Componente Principal 1 (PC1) y Componente Principal 2 (PC2) representan las direcciones de mayor variabilidad en los datos. Estos componentes son combinaciones lineales de las variables originales (eficiencia\_camion, tonelaje\_por\_paladas, eficiencia\_distancia).

#### **Clústeres:**

Los puntos en el gráfico están coloreados según los clústeres (agrupaciones) a los que pertenecen. En este caso, hay cuatro clústeres, identificados por los colores morado, azul, verde y amarillo.

### **¿Qué Significan los Clústeres?**

#### **Clúster 0 (Morado):**

Este grupo de puntos comparte características similares en términos de eficiencia del camión, tonelaje por palada y eficiencia de distancia. Puede ser útil investigar las características específicas de este grupo para entender qué las hace similares.

#### **Clúster 1 (Azul):**

Similarmente, los puntos en este clúster comparten características comunes. Compararlos con otros clústeres puede revelar diferencias clave que podrían ser explotadas para mejorar la eficiencia operativa.

#### **Clúster 2 (Verde):**

Este grupo puede representar un conjunto de camiones y palas con características de eficiencia particulares que los diferencian de los otros clústeres. Podría ser interesante investigar si hay prácticas operativas o condiciones específicas que influyen en su desempeño.

#### **Clúster 3 (Amarillo):**

Este clúster agrupa otro conjunto de datos con características similares. Al igual que los otros clústeres, investigar las propiedades comunes puede ofrecer insights sobre su desempeño.

### **Interpretación de los Resultados por Clúster (gráficas de distribución):**

#### **Clúster 0:**

Fecha Promedio: 16 de septiembre de 2023

Camión Promedio: 23.79 (codificado numéricamente)

Loader (pala) Promedio: 1.58 (codificado numéricamente)

Eficiencia del Camión: 0.150339

Eficiencia de la Pala: 1.056925

Tonelaje por Palada: 79.063127

Eficiencia de Distancia: 0.990484

- Este grupo tiene una eficiencia del camión moderada y una eficiencia de la pala más baja en comparación con otros clústeres. El tonelaje por palada también es el más bajo, y la eficiencia de distancia está cerca de 1.

### **Clúster 1:**

Fecha Promedio: 3 de noviembre de 2023

Camión Promedio: 23.46 (codificado numéricamente)

Loader (pala) Promedio: 1.72 (codificado numéricamente)

Eficiencia del Camión: 0.139137

Eficiencia de la Pala: 1.175720

Tonelaje por Palada: 99.073852

Eficiencia de Distancia: 7.377063

- Este grupo presenta una baja eficiencia del camión, pero una eficiencia de la pala relativamente alta. El tonelaje por palada es alto y la eficiencia de distancia es muy alta, lo que sugiere mucha variabilidad en la distancia recorrida.

### **Clúster 2:**

Fecha Promedio: 27 de agosto de 2023

Camión Promedio: 23.39 (codificado numéricamente)

Loader (pala) Promedio: 1.79 (codificado numéricamente)

Eficiencia del Camión: 0.118032

Eficiencia de la Pala: 1.169723

Tonelaje por Palada: 104.649197

Eficiencia de Distancia: 1.008742

- Este grupo tiene la eficiencia del camión más baja pero una eficiencia de la pala alta. El tonelaje por palada es el más alto, y la eficiencia de distancia está cerca de 1, indicando poca variabilidad en la distancia.

**Clúster 3:**

Fecha Promedio: 3 de diciembre de 2023

Camión Promedio: 23.29 (codificado numéricamente)

Loader (pala) Promedio: 1.16 (codificado numéricamente)

Eficiencia del Camión: 0.217199

Eficiencia de la Pala: 1.171333

Tonelaje por Palada: 104.381247

Eficiencia de Distancia: 0.956280

- Este grupo tiene la eficiencia del camión más alta, y una eficiencia de la pala alta. El tonelaje por palada es alto, similar al Clúster 2, y la eficiencia de distancia es la más baja, sugiriendo menos variabilidad en la distancia recorrida.

**Conclusiones:**

**Clúster 0:** Indica operaciones con eficiencia de camión moderada y menor tonelaje por palada. Podría ser un enfoque para optimización (menos eficiencia).

**Clúster 1:** Presenta una eficiencia de distancia muy alta y variabilidad en la distancia, posiblemente indicando rutas largas o cambios en las condiciones operativas.

**Clúster 2:** Tiene la menor eficiencia de camión, pero el mayor tonelaje por palada, sugiriendo posibles mejoras en la operación de los camiones.

**Clúster 3:** Muestra la mejor eficiencia de camión, similar a Clúster 2 en tonelaje por palada, pero con una menor eficiencia de distancia.

Esto proporciona una base sólida para investigar y optimizar diferentes aspectos operacionales según los clústeres (se puede revisar cada clúster y su composición en el Dashboard adjunto).

**Acciones Sugeridas finales:****Optimización de Rutas (Clúster 2):**

Se podría investigar las rutas de los camiones en el clúster 2 para entender por qué la eficiencia por distancia es tan alta. Podría haber rutas subóptimas o desviaciones innecesarias.

**Mejorar la Eficiencia de Carga (Clúster 0):**

Evaluar las operaciones de carga en el clúster 0 para aumentar el tonelaje por paladas y la eficiencia de camión. Esto podría incluir capacitación adicional para operadores de palas o ajustes en la logística de carga.

**Comparar Buenas Prácticas (Clúster 1):**

Examinar las operaciones en el clúster 1, donde la eficiencia es alta, para identificar mejores prácticas que se puedan implementar en otros clústeres.

**Revisión de Operaciones (Clúster 3):**

Aunque la eficiencia de pala y el tonelaje por paladas son altos, la baja eficiencia de camión en el clúster 3 sugiere que podría haber cuellos de botella o ineficiencias en el uso del camión que deben ser resueltas.

**Observaciones sobre la Comparación entre Clústeres:**

La posición de las cajas y las líneas de mediana ayudan a comparar rápidamente la eficiencia de las palas entre diferentes clústeres.

Por ejemplo, si una caja está más alta en el eje Y, indica que ese clúster tiene una mayor eficiencia de pala en promedio.

**Variabilidad dentro de los Clústeres:**

La altura de las cajas muestra la variabilidad de la eficiencia de las palas dentro de cada clúster.

Clústeres con cajas más altas tienen una mayor variabilidad en la eficiencia de las palas.

**Outliers (posteriores):**

Los puntos individuales fuera de los bigotes indican outliers, lo que podría sugerir operaciones anómalas o excepcionales en esos clústeres.

**Tendencias Generales:**

Se puede ver si hay clústeres que consistentemente tienen una eficiencia de pala más alta o más baja, lo que puede ser útil para identificar mejores prácticas operativas o áreas de mejora.

**Acciones Sugeridas Extras:**

- **Identificación de Mejores Prácticas:** Examinar los clústeres con alta eficiencia de pala para entender qué prácticas operativas pueden estar contribuyendo a su éxito.
- **Investigación de Outliers:** Investigar los outliers para determinar si representan oportunidades de mejora o si son casos atípicos que requieren atención especial.
- **Optimización de Operaciones:** Utilizar esta información para ajustar las operaciones de los clústeres con menor eficiencia, enfocándose en aumentar la eficiencia de las palas y mejorar la uniformidad de las operaciones.

**Archivos adjuntos:**

- Notebook donde se trabajó el archivo de datos .csv, se realizó ETL y EDA:  
[EjercicioML/examen.ipynb at main · melinnicri/EjercicioML](#)
- Notebook donde se trabajó el nuevo DATAFRAME para hacerse el ML:  
[EjercicioML/MLmineria.ipynb at main · melinnicri/EjercicioML](#)
- Notebook donde se hicieron las listas con los camiones, palas y sus eficiencias:  
[EjercicioML/cluster\\_cam\\_pal.ipynb at main · melinnicri/EjercicioML](#)
- Panel en PowerBI: [Dashboard](#),
- [Imagen Dashboard](#).//...\_@v