

Proyecto Final – Modelos Predictivos

“Predicción de la productividad ciclo a ciclo de camiones de acarreo en minería usando modelos de regresión”

Presentación por Liz Pérez



Base de datos

Entendiendo la industria

+ 82mil datos

Se recopila información de cada ciclo recorrido por los camiones.

Nombre de la columna	Tipo de dato	Descripción	Unidades
OID	Numérica	Identificador único del registro	-
Cycle OID	Numérica	Identificador único del ciclo	-
Date	Fecha	Fecha del ciclo	-
Month	Numérica	Mes del ciclo	1-12
Year	Numérica	Año del ciclo	-
Secondary Machine Name	Catógórica	Nombre de la máquina secundaria (pala) (Siempre exc)	-
Secondary Machine Class Name	Catógórica	Clase de la máquina secundaria	-
Material Name	Catógórica	Nombre del material movido	-
Material Group	Catógórica	Grupo del material (Ore/Waste)	-
Cycle Duration	Numérica	Duración del ciclo	Minutos
Completed Cycle Duration Min	Numérica	Duración del ciclo completo	Minutos
Loading Duration (min)	Numérica	Tiempo de carga	Minutos
Travelling Empty Duration (min)	Numérica	Tiempo de viaje vacío	Minutos
Queuing At Source Duration (min)	Numérica	Tiempo de espera en la fuente	Minutos
Spotting At Source Duration (min)	Numérica	Tiempo de alineación en la fuente	Minutos
Wait For Load Duration (min)	Numérica	Tiempo de espera para cargar	Minutos
Payload Q	Numérica	Toneladas transportadas en el ciclo	Toneladas
Completed Cycle Duration	Numérica	Duración total del ciclo	Minutos





1. Que me motivo a desarrollar el estudio predictivo?



- **Aplicar a lo aprendido a mi día a día**
- **Poder realizar investigaciones más profundas en base a análisis más robustos**
- **Darle un toque más humano e innovador a la planificación en la industria**



2. Experiencia buscando la data

Buscar en Kaggle

Primordialmente intente buscar base de datos en kaggle



Luego caíen cuenta que tengo dudas internas en mi campo de estudio



Buscar data de mi área

Decidi buscar y hacer análisis para contestar las dudas que yo misma me hacía

1. Importante entender la data

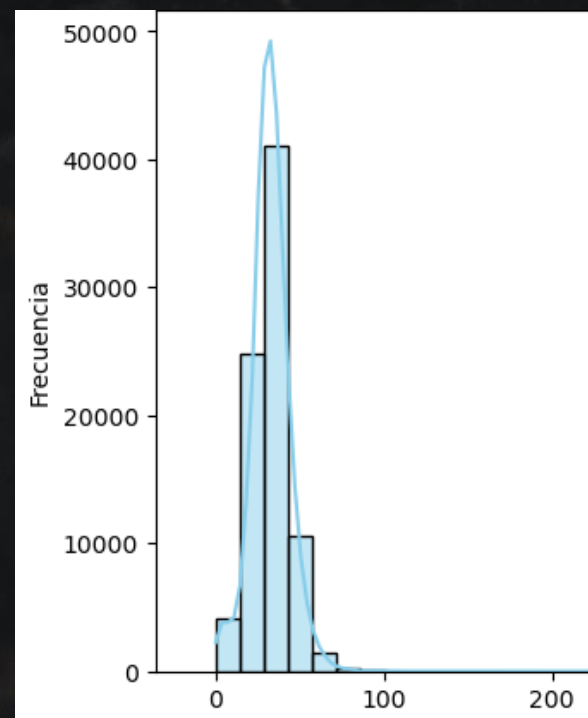
2. Saber que busco estudiar con el análisis

3. Muchas veces confiar en sistemas automatizados ayuda sin embargo hay que hacer pruebas de QA/QC para garantizar la validez de los datos.

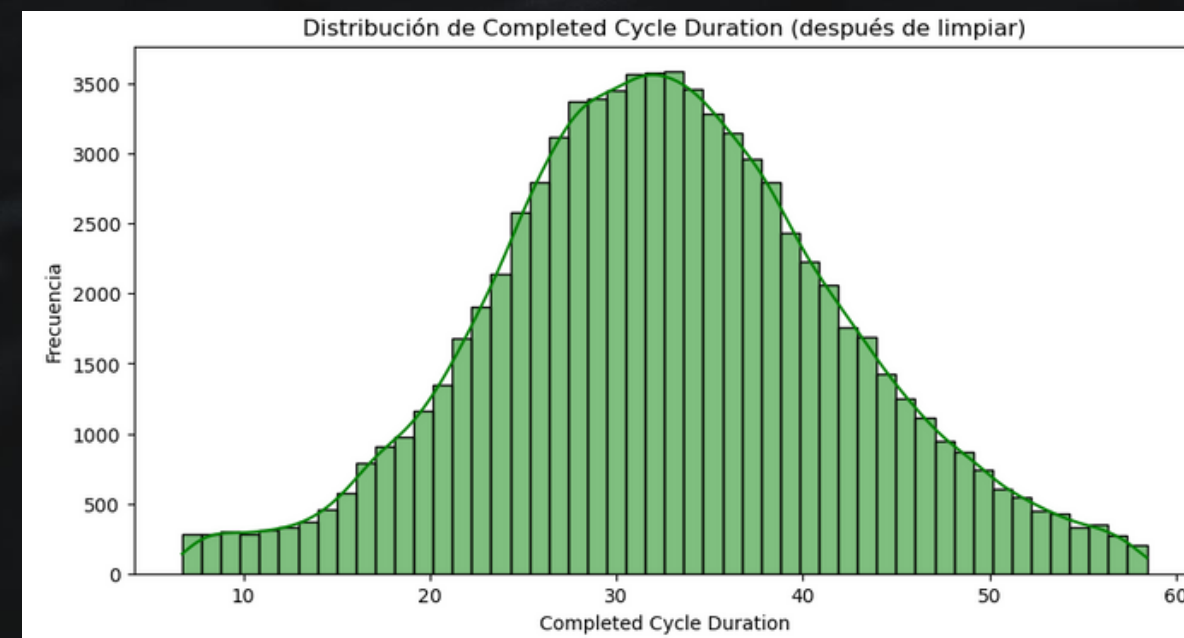
4. Muchos datos con valores nulos o extremos

3. Que lecciones aprendí al hacer análisis descriptivo?

Antes de la limpieza



Luego de la limpieza



4. Pasos para el análisis (predictivo y descriptivo)

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression, Ridge
from sklearn.ensemble import RandomForestRegressor, GradientBoostingRegressor
from sklearn.metrics import r2_score, mean_squared_error, mean_absolute_error

# Configuración opcional
pd.set_option('display.max_columns', None)

# Cargar datos
df = pd.read_excel("Truck Cycle Information for shovel.xlsx")
print(f"Forma del dataset: {df.shape}")
df.head()

# Filtrar solo los ciclos de la pala P&H 4100XPC y material Ore o Waste
df = df[(df['Secondary Machine Class Name'] == 'P&H 4100XPC') &
        (df['Material Group'].isin(['Ore', 'Waste']))].copy()

# Calcular productividad por ciclo
# Calcular productividad por ciclo en ton/hora
df['Productividad_avg'] = (df['Payload_Q'] / df['Completed Cycle Duration']) * 60

# Ver algunas filas
print(df[['Payload_Q', 'Completed Cycle Duration', 'Productividad_avg']].head())

# LIMPIEZA
# Revisar valores nulos
print(df.isnull().sum())

# Quitar filas con valores nulos en las variables clave
df.dropna(subset=['Payload_Q', 'Completed Cycle Duration', 'Productividad_avg'], inplace=True)

# Quitar filas con duración <= 0
df = df[df['Completed Cycle Duration'] > 0]

print(f"Dataset después de limpieza: {df.shape}")

# Estadísticas descriptivas de productividad
print("\n Estadísticas de Productividad (ton/hora):")
```

Importar dataset

Filtro

Crear columna variable

Crear la variable objetivo

Limpieza de datos

Análisis descriptivo

Seleccionar variables

Dividir set y entrenar modelo

Ciclos del equipo seleccionado

Duración total del ciclo

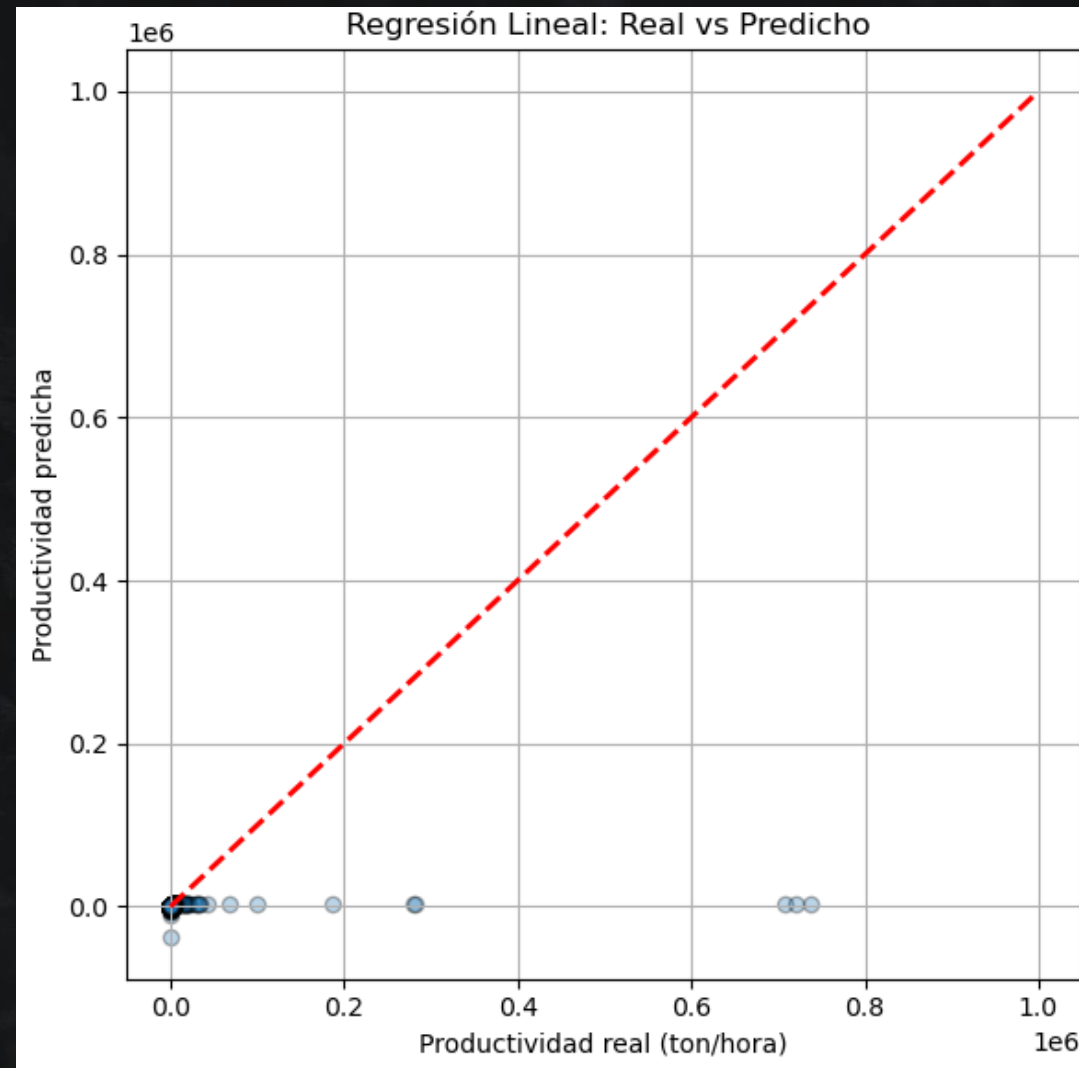
Productividad_avg = $\frac{\text{Payload_Q}}{\text{Duración total del ciclo}} * 60$

Valores nulos
Outliers
Ciclos del equipo seleccionado

Evaluar modelos

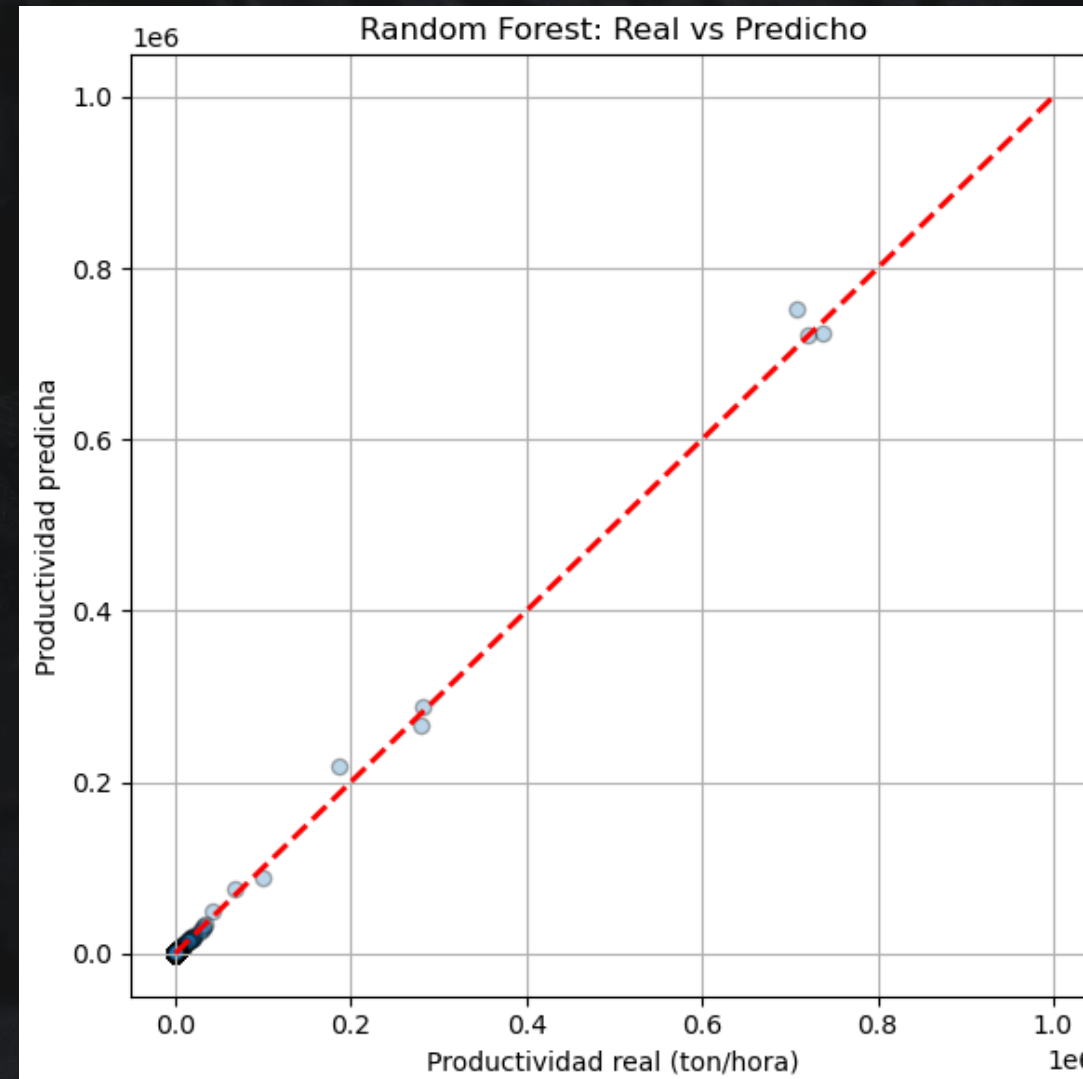


Regresión Lineal



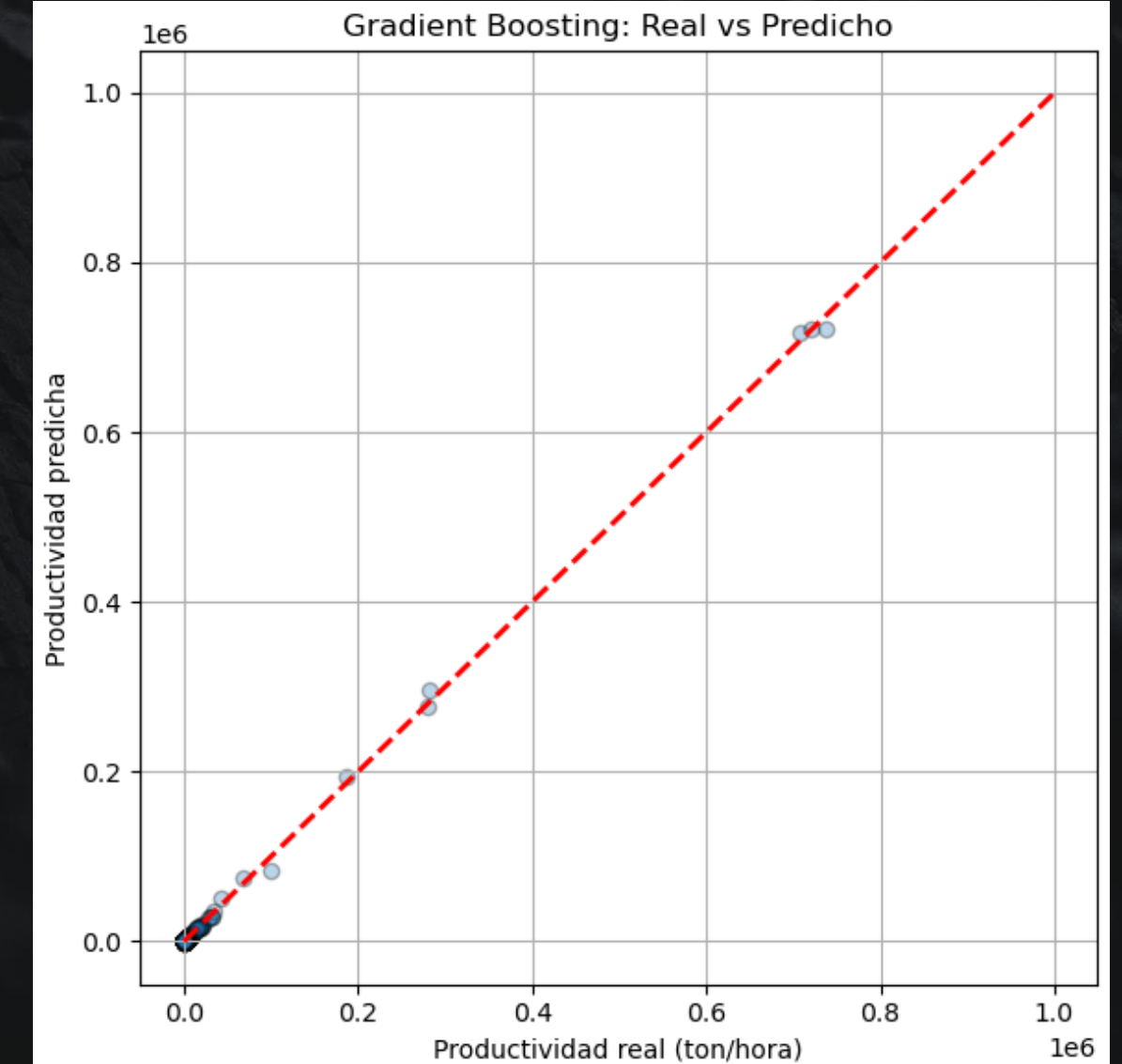
$R^2=0.006$
RMSE =
10371.75
MAE = 536.47

Regresión Random Forest



$R^2=0.998$
RMSE = 461.23
MAE = 9.99

Regresión Gradient Boosting



$R^2=0.999$
RMSE = 250.81
MAE = 21.05

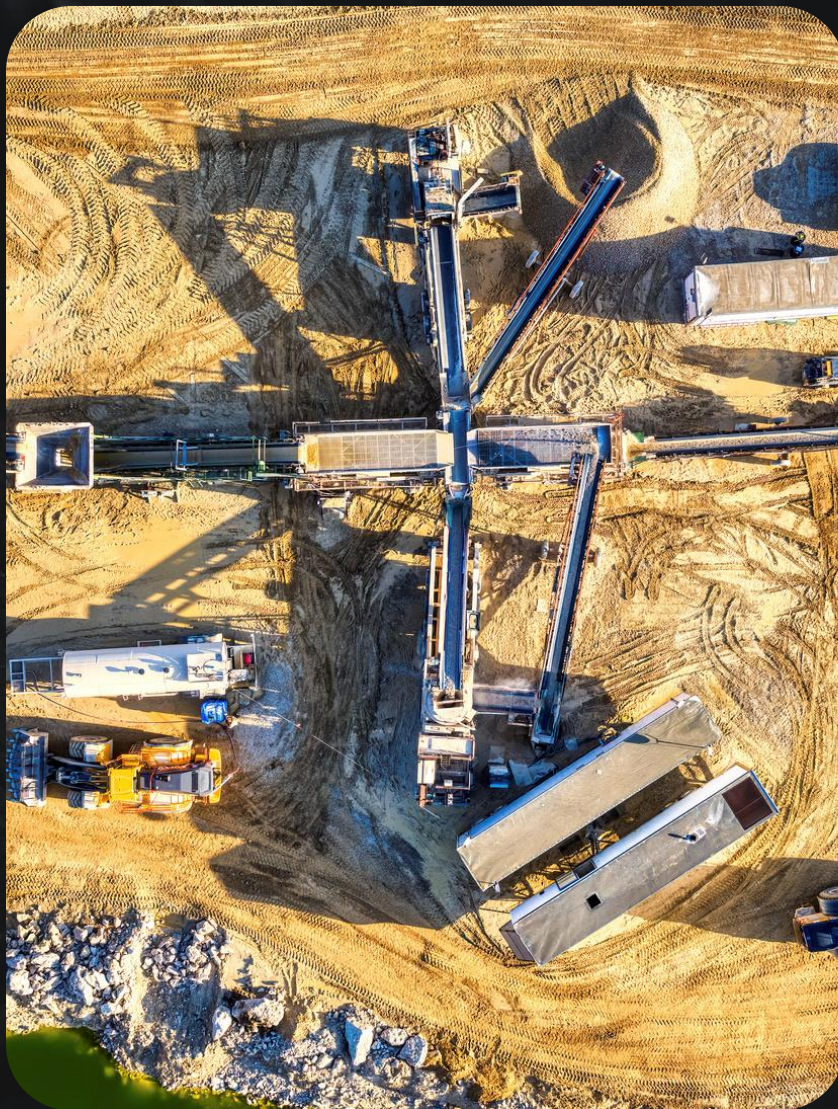
Análisis predictivo

5. Descubrimientos

La productividad de un camión no está influenciada de forma lineal con el payload y la duración de ciclo.

La productividad tiene una relación más compleja en relación a las variables.





6. Que haría mejor?

Utilizaría más variables

Por ejemplo distancia, condición climática, cantidad de retrasos en el viaje.

Convertir el análisis por serie en turnos

Seleccionar y especificar en que turno fueron los datos y evaluar si dependiendo del turno varían y que factor influye más.

Más información por meses y no solo 90 días

De tal manera que me permita pronosticar en base a una visión más estratégica y menos del día a día.



7. Que descubrí de mi

Descubrí que tengo la capacidad de aplicar herramientas y el aprendizaje para resolver problemas reales dentro del entorno minero. A pesar de no venir de una formación técnica en programación, logré estructurar un proyecto completo desde la limpieza de datos hasta la evaluación de modelos predictivos. Aprendí que puedo interpretar resultados estadísticos, tomar decisiones basadas en datos y que tengo un gran potencial para seguir profundizando en la analítica aplicada a operaciones industriales.

Además que hay mucho más por explorar y potencial de datos en la industria que me ayudarían a llegar al siguiente nivel.

The background of the image is a dark, almost black, highly textured surface. It appears to be a close-up of a material like charred wood or coal, showing a complex pattern of cracks, ridges, and small, irregular fragments. The lighting is dramatic, with some areas catching a bit of light, revealing a slightly glossy or ashy texture, while most of the surface is in deep shadow. In the center of the image, the word "Gracias" is written in a bold, yellow, sans-serif font. The text is clear and stands out prominently against the dark, busy background.

Gracias