

# Modelo de Regresión para la Predicción del Consumo Energético en la Industria Siderúrgica

## INTEGRANTES

- ANGELO CRISTOBAL AGUI
- ELVIS BRAVO SÁNCHEZ
- JOSUE MORENO CIÑA
- PAUL CÁCERES

# CONTENIDO

- |           |                     |           |              |
|-----------|---------------------|-----------|--------------|
| <b>01</b> | CASO DE ESTUDIO     | <b>07</b> | TUNING       |
| <b>02</b> | OBJETIVOS           | <b>08</b> | RESULTADOS   |
| <b>03</b> | EDA                 | <b>09</b> | ESTRATEGIAS  |
| <b>04</b> | FEATURE ENGINEERING | <b>10</b> | CONCLUSIONES |
| <b>05</b> | MODELING            |           |              |
| <b>06</b> | FEATURE SELECTION   |           |              |

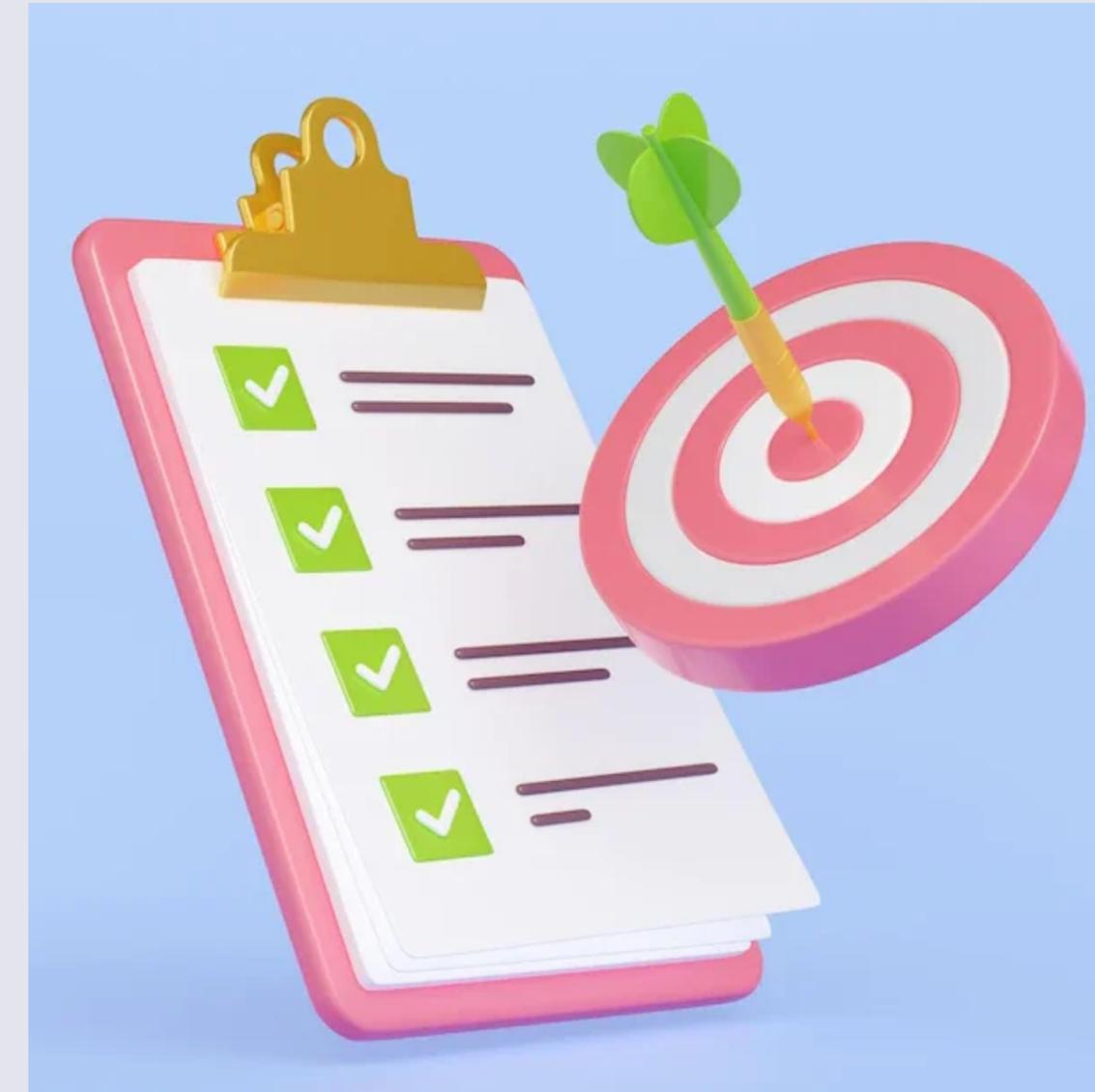
# CASO DE ESTUDIO

El caso de estudio consiste en predecir el **consumo energético** en una planta siderúrgica inteligente de pequeña escala. Se utiliza un conjunto de características operativas registradas en tiempo real por sensores industriales, como consumo energético, factores de potencia y condiciones de operación. Es una tarea de **regresión**, donde el objetivo es estimar la variable continua **Usage\_kWh**.



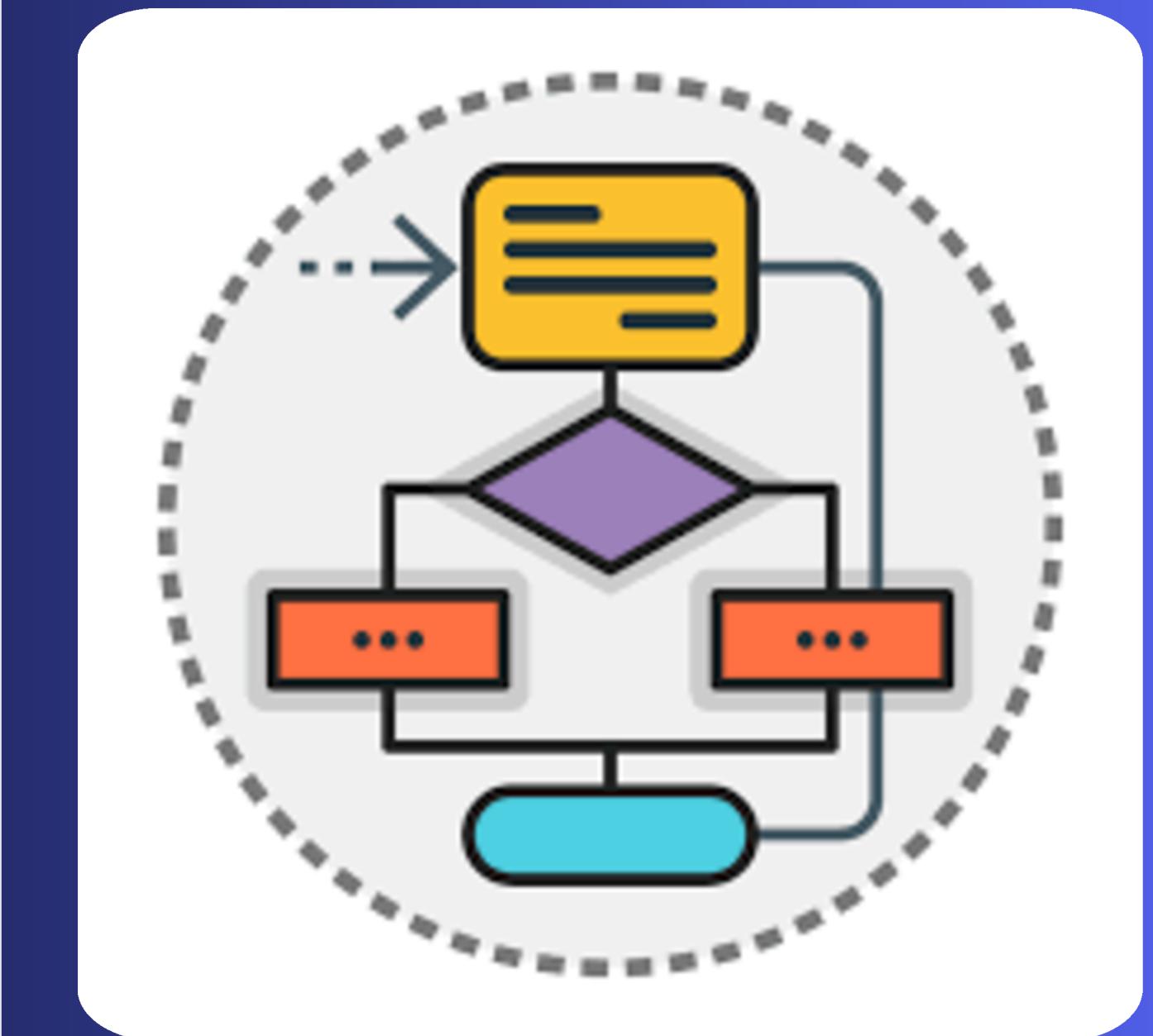
# OBJETIVOS

- Entrenar un modelo de **regresión** que, basado en las características operativas de la planta, sea capaz de **predecir** con precisión el **consumo energético** (Usage\_kWh)
- **Optimizar la gestión energética** mediante la aplicación de modelos predictivos, contribuyendo a la eficiencia y sostenibilidad

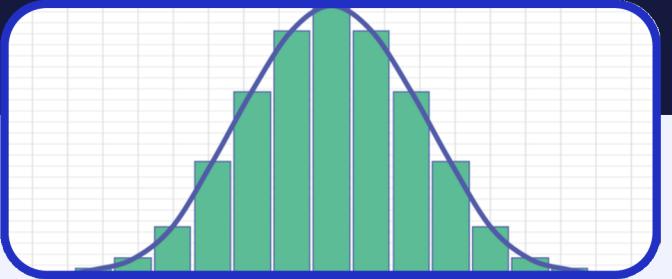


# METODOLOGÍA

Se realizó un **análisis exploratorio de datos (EDA)** para comprender la distribución y relaciones entre variables, utilizando visualizaciones y detección de anomalías (no se encontraron anomalías significativas o valores faltantes) Luego, se desarrolló el **modelamiento**, aplicando técnicas de ingeniería de características (extracción de variables temporales, codificación) y probando 4 algoritmos de regresión. Se evaluó la estabilidad de los modelos mediante validación cruzada. Finalmente, los modelos fueron **evaluados con métricas** como MAE y RMSE

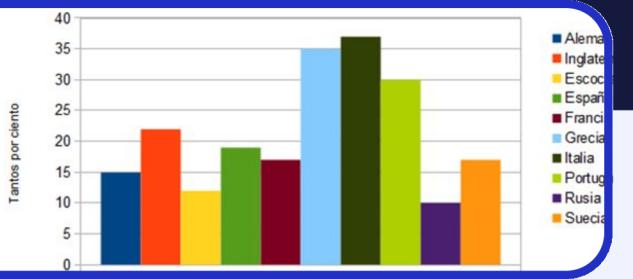
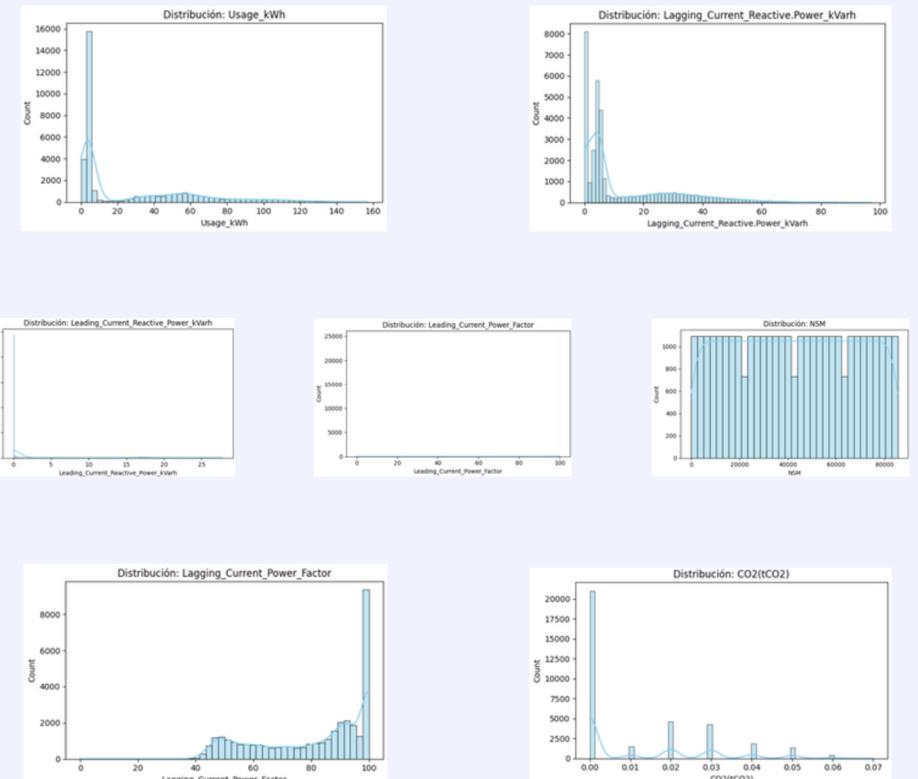


# 1. EDA



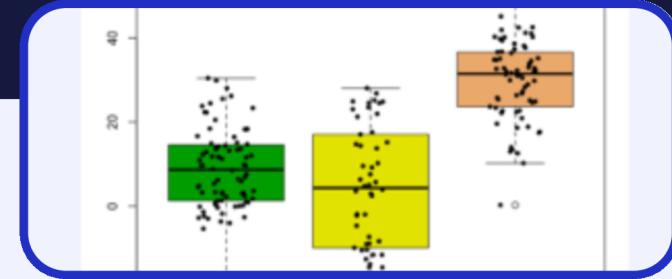
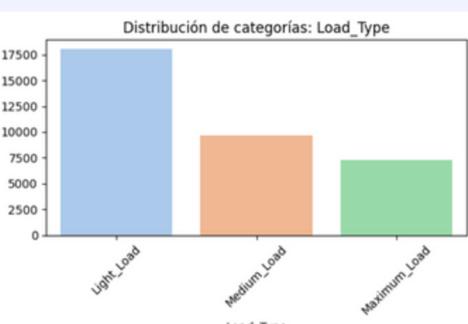
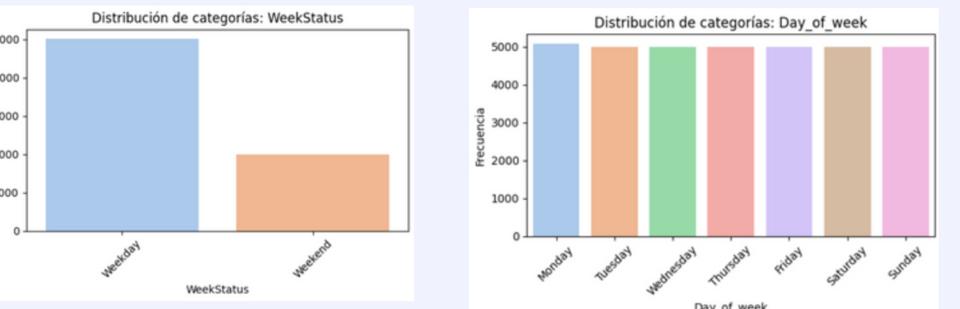
## HISTOGRAMA DE DATOS NUMERICOS

- Usage\_kWh
- Lagging\_Current\_Reactive.Power\_kVarh
- Leading\_Current\_Reactive\_Power\_kVarh
- CO2(tCO2)
- Lagging\_Current\_Power\_Factor
- Leading\_Current\_Power\_Factor
- NSM



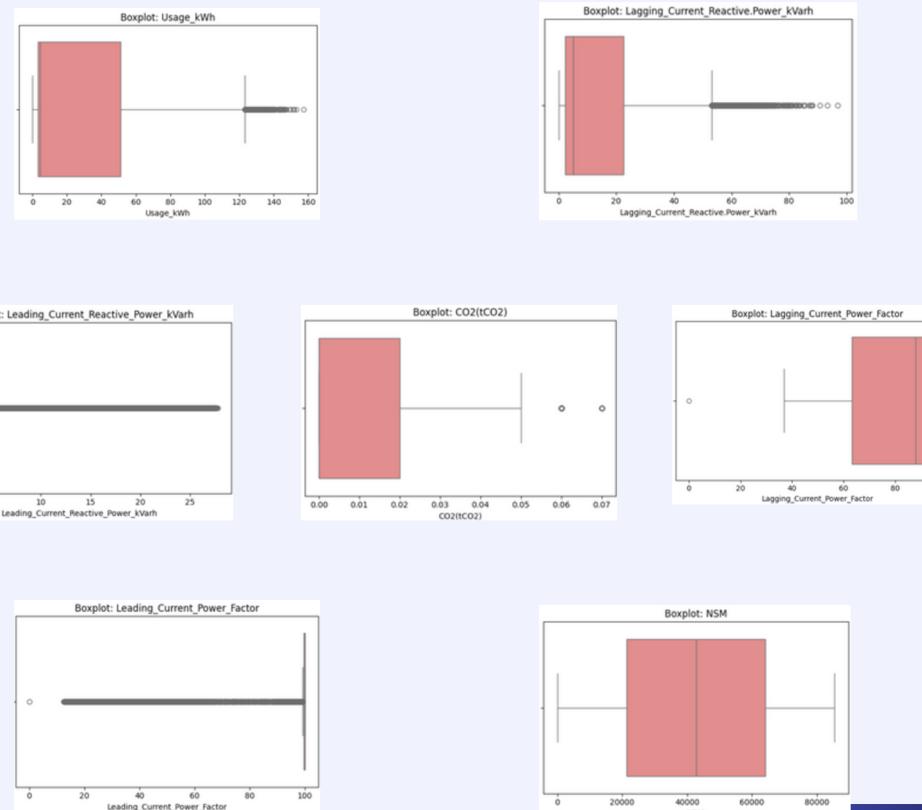
## GRAFICOS DE BARRAS DE DATOS CATEGORICOS

- WeekStatus
- Day\_of\_week
- Load\_Type



## BOXPLOT: DIAGRAMA DE CAJA DE BIGOTES

Con el fin de eliminar outliers antes de realizar el análisis de datos vacíos

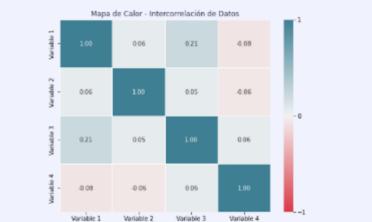
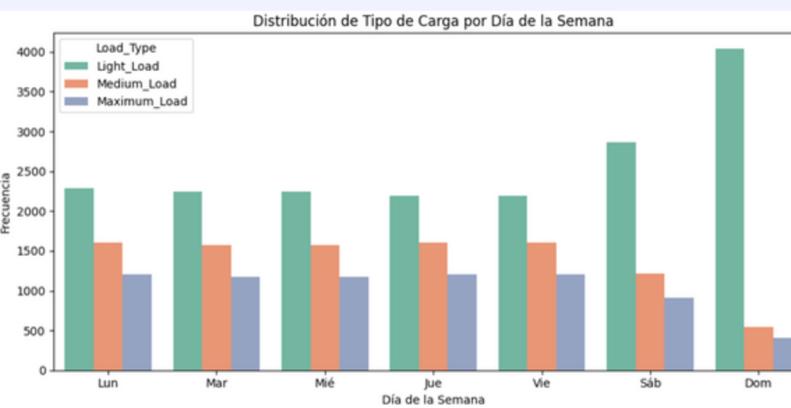
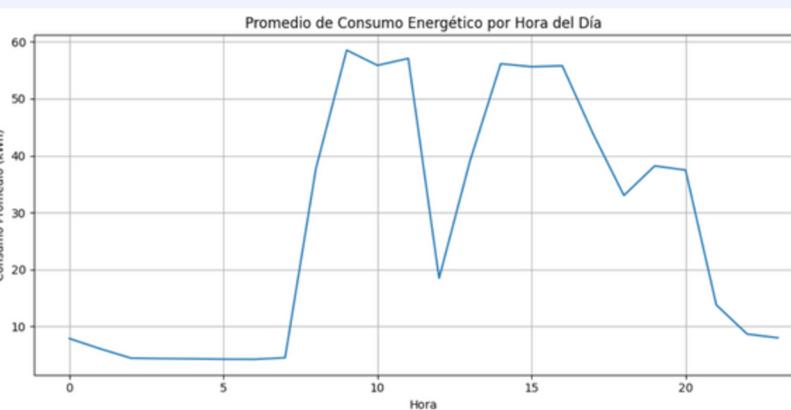


# 1. EDA



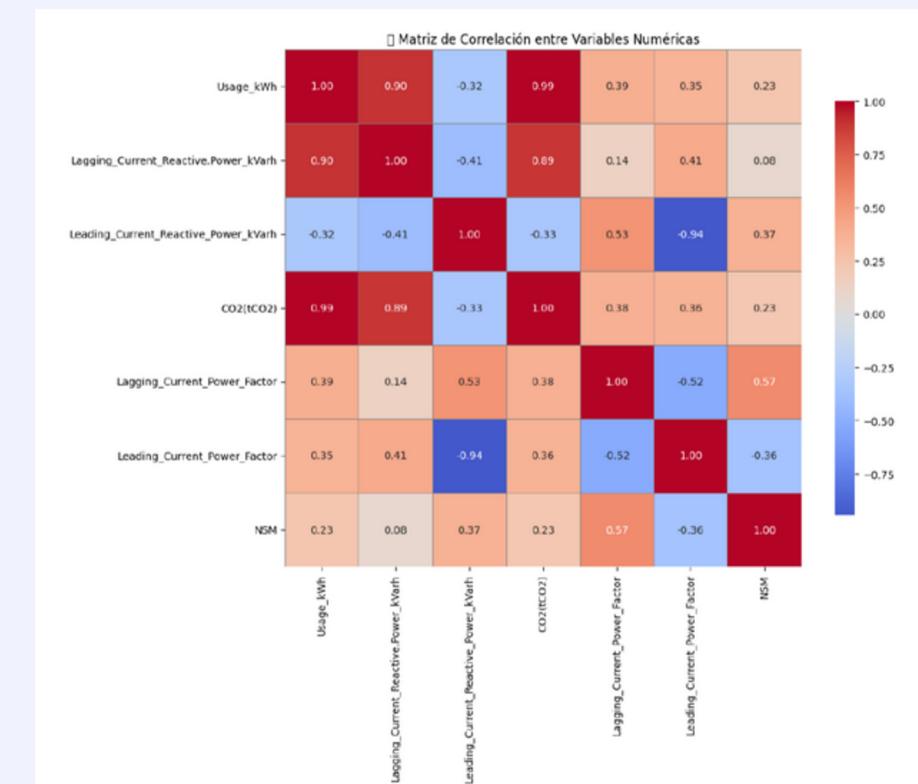
## ANALISIS TEMPORAL

Se realizó un análisis temporal para evaluar patrones de consumo energético a lo largo del día, la semana y el mes.

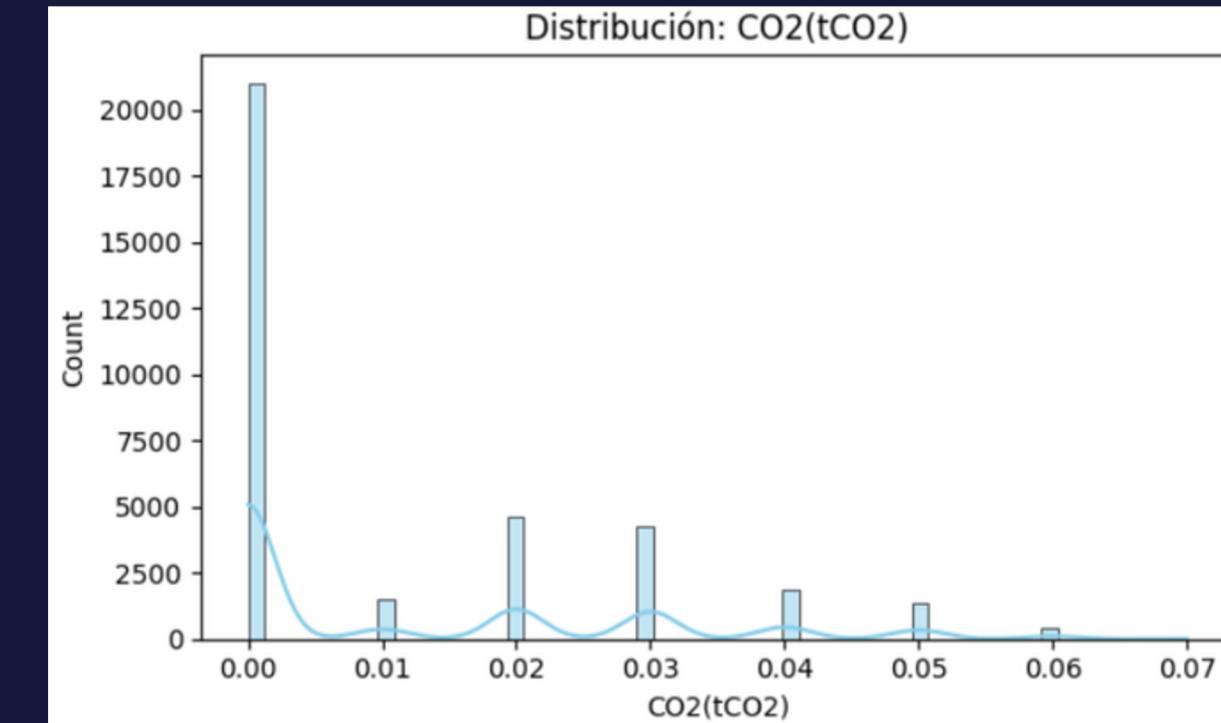
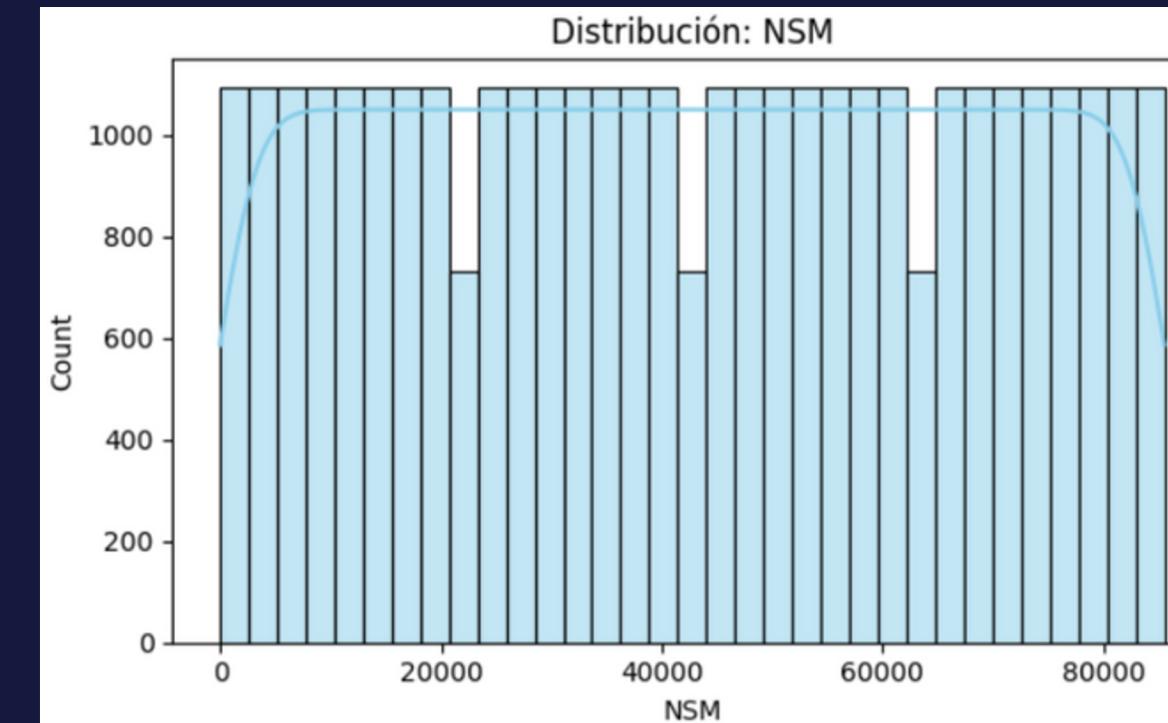
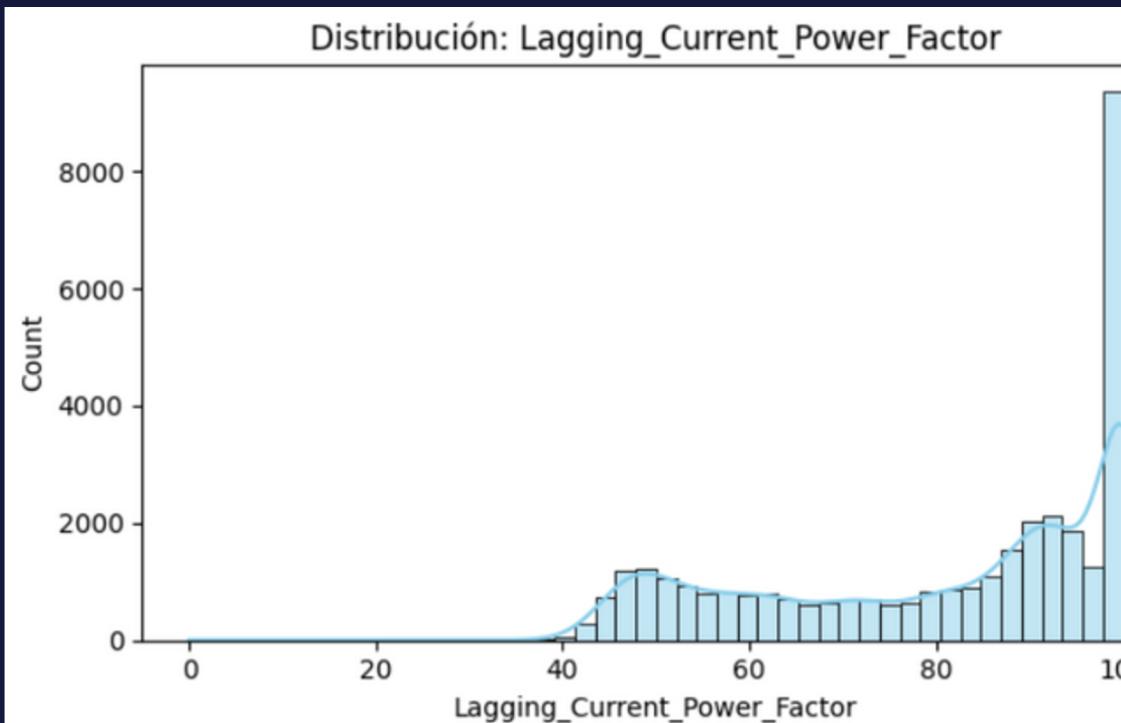
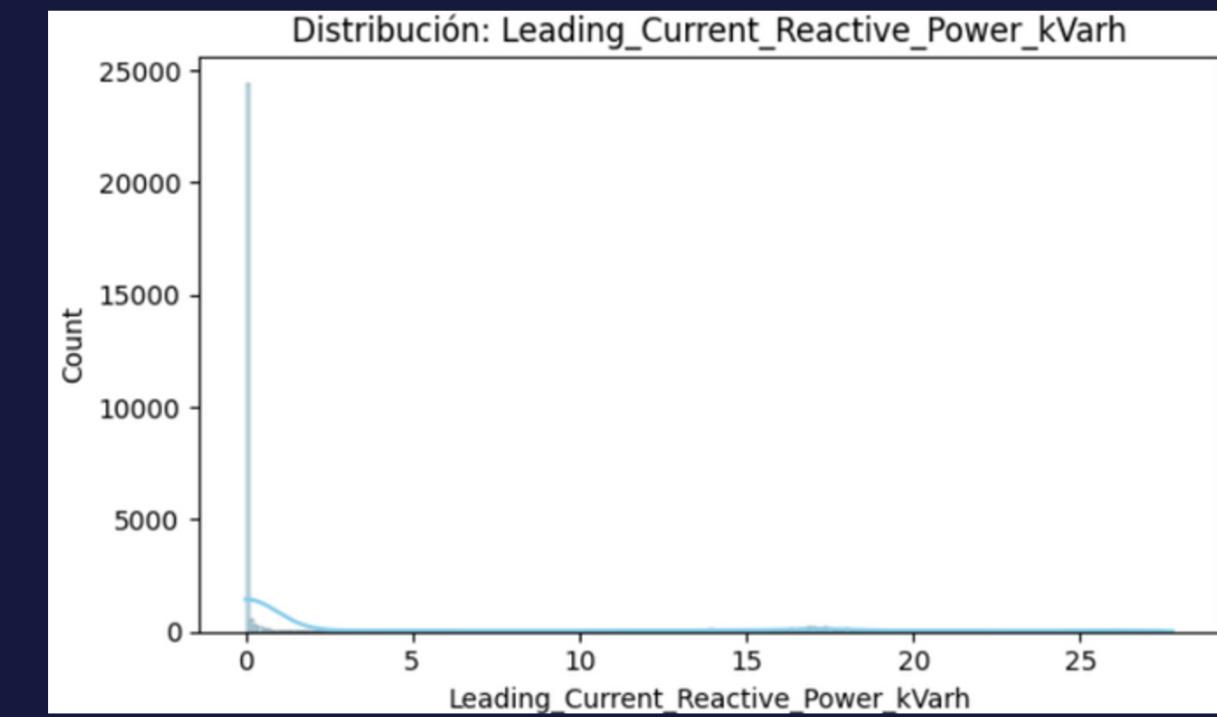
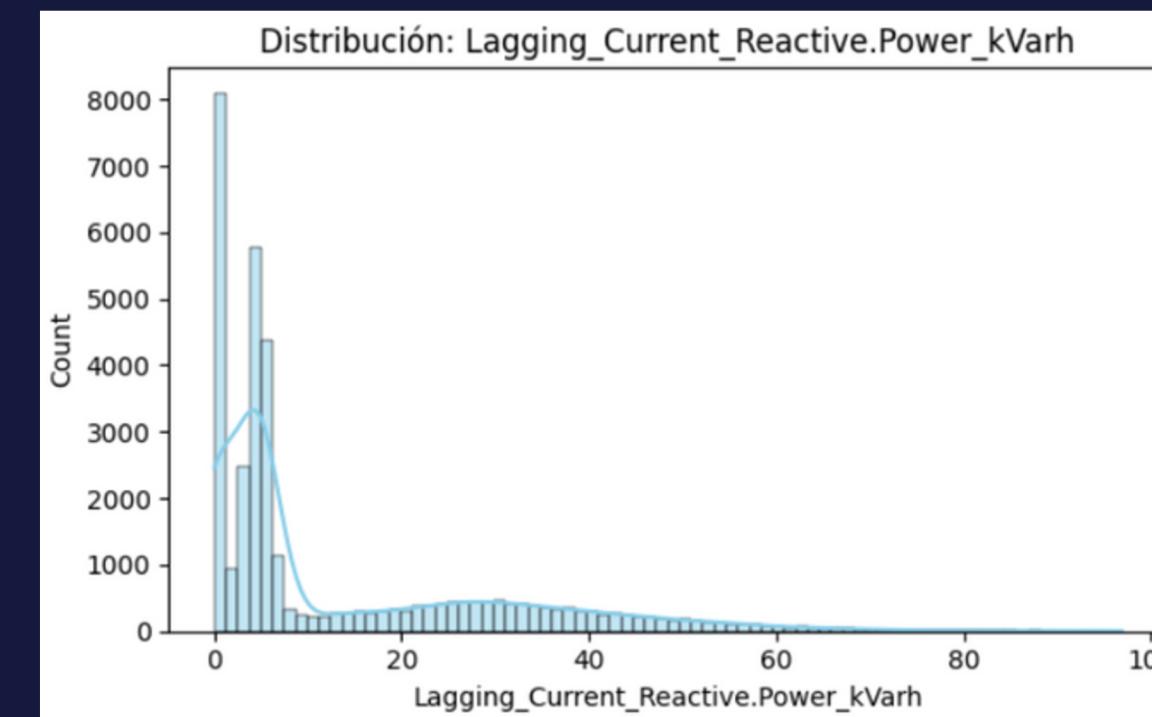
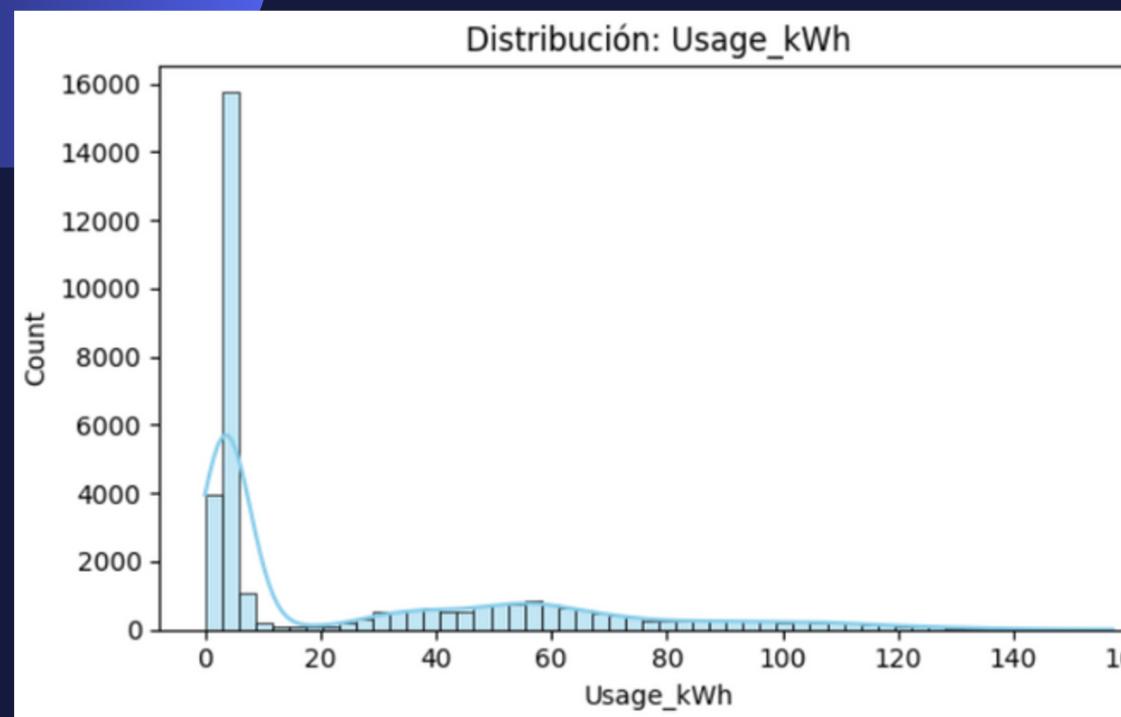


## MATRIZ DE CORRELACION

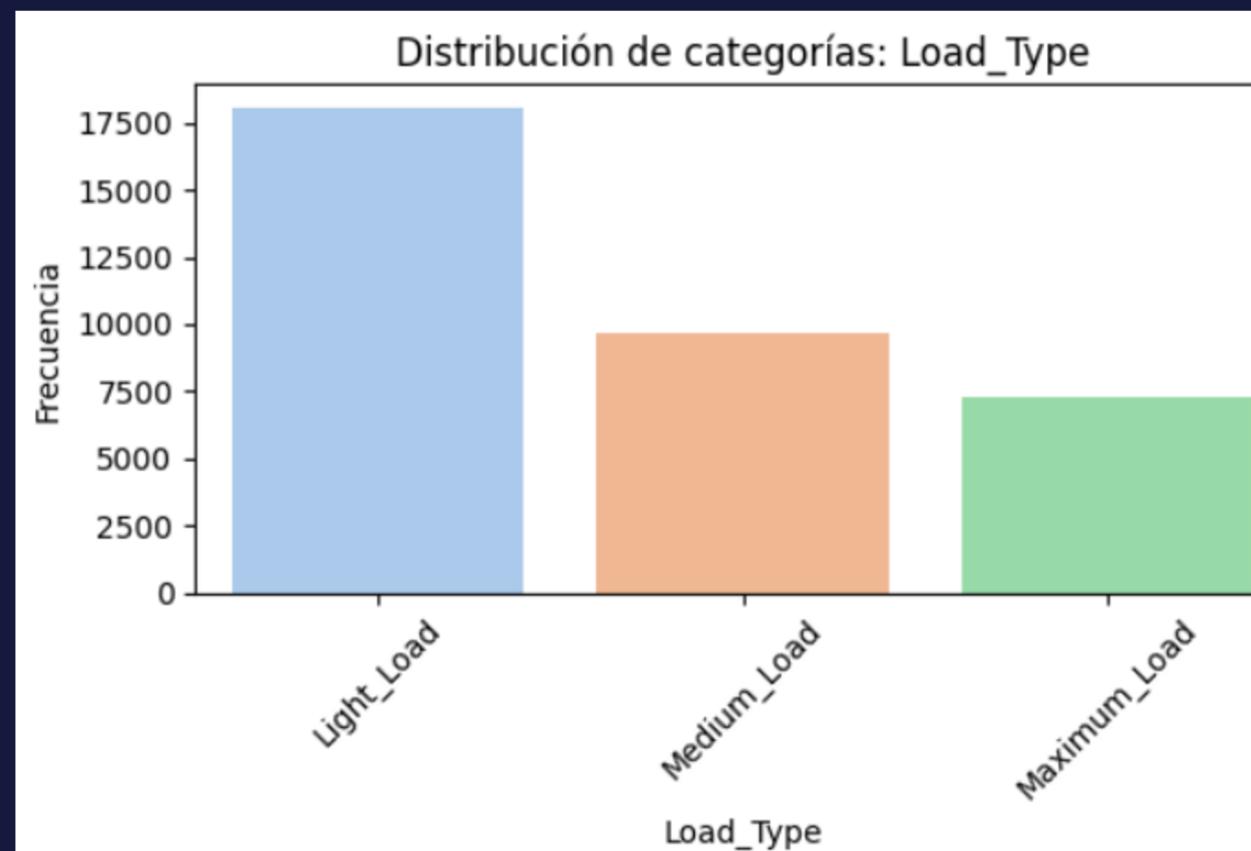
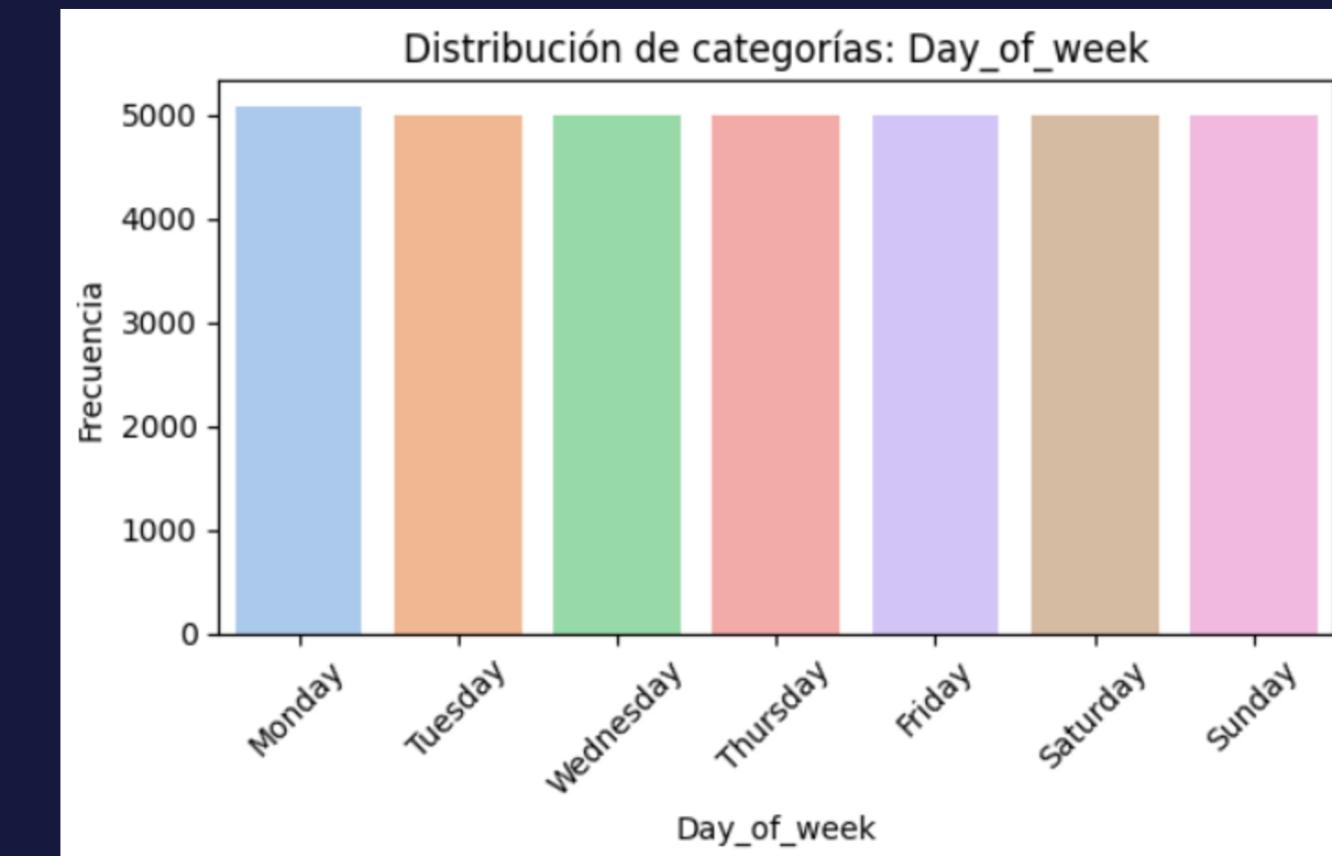
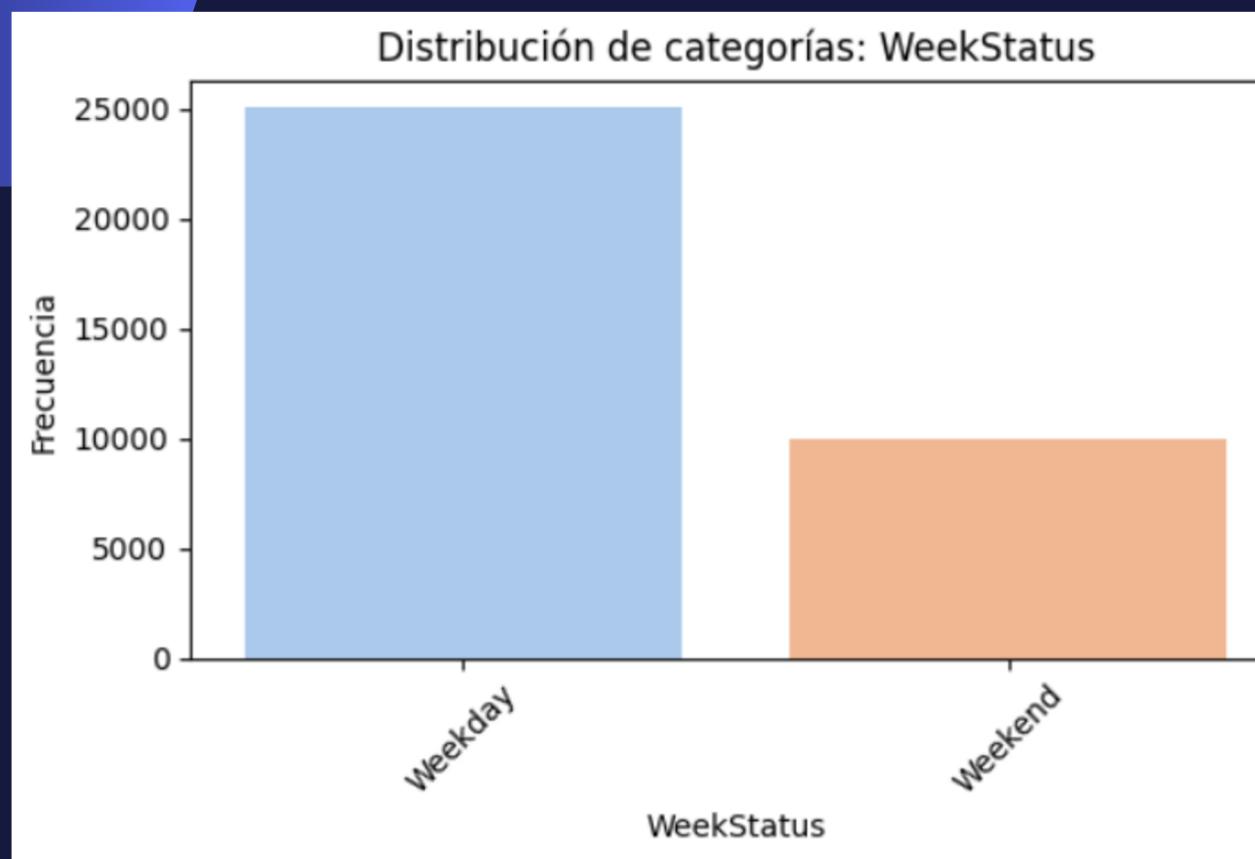
Se examinó la correlación entre variables numéricas para identificar relaciones fuertes y posible colinealidad



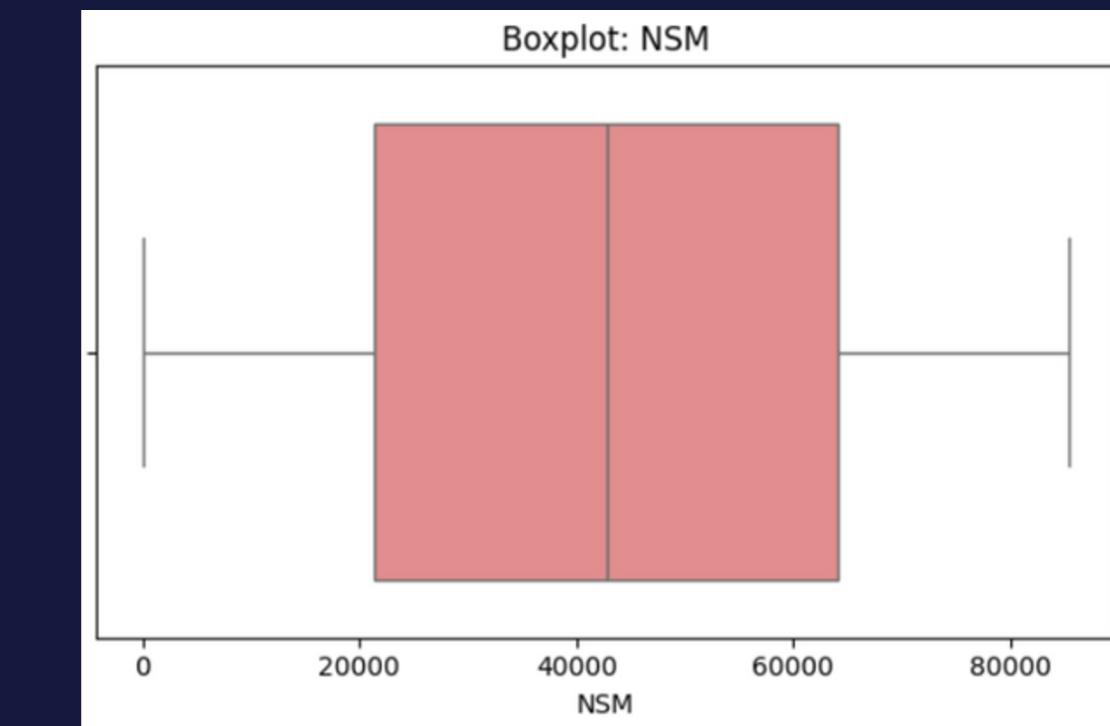
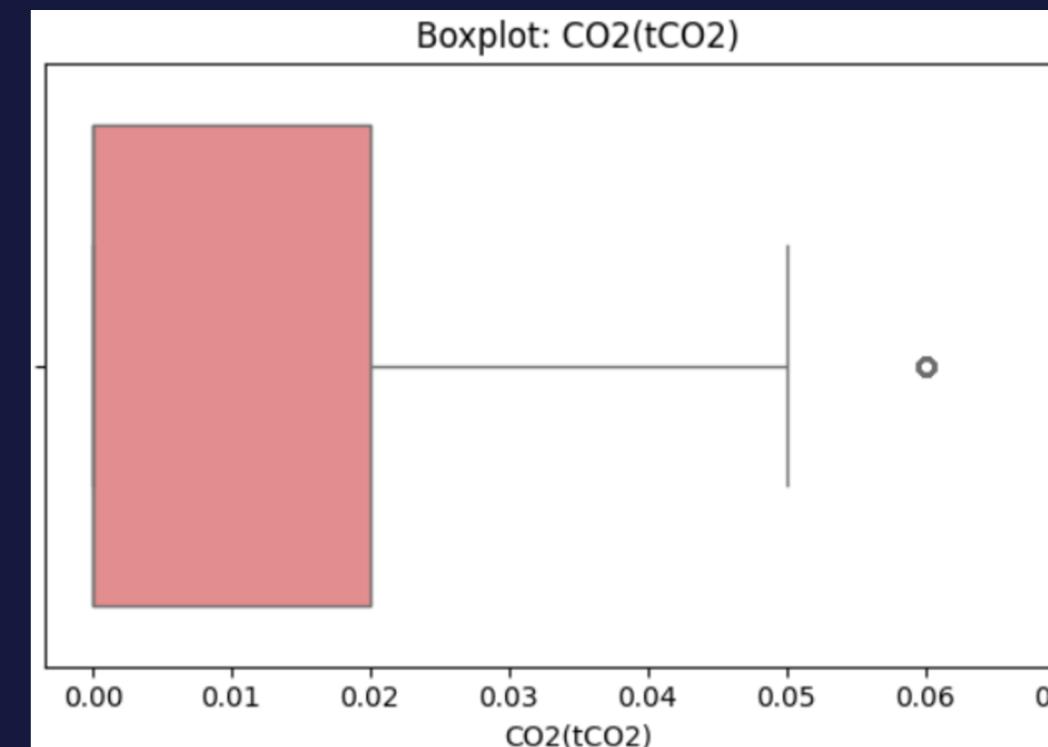
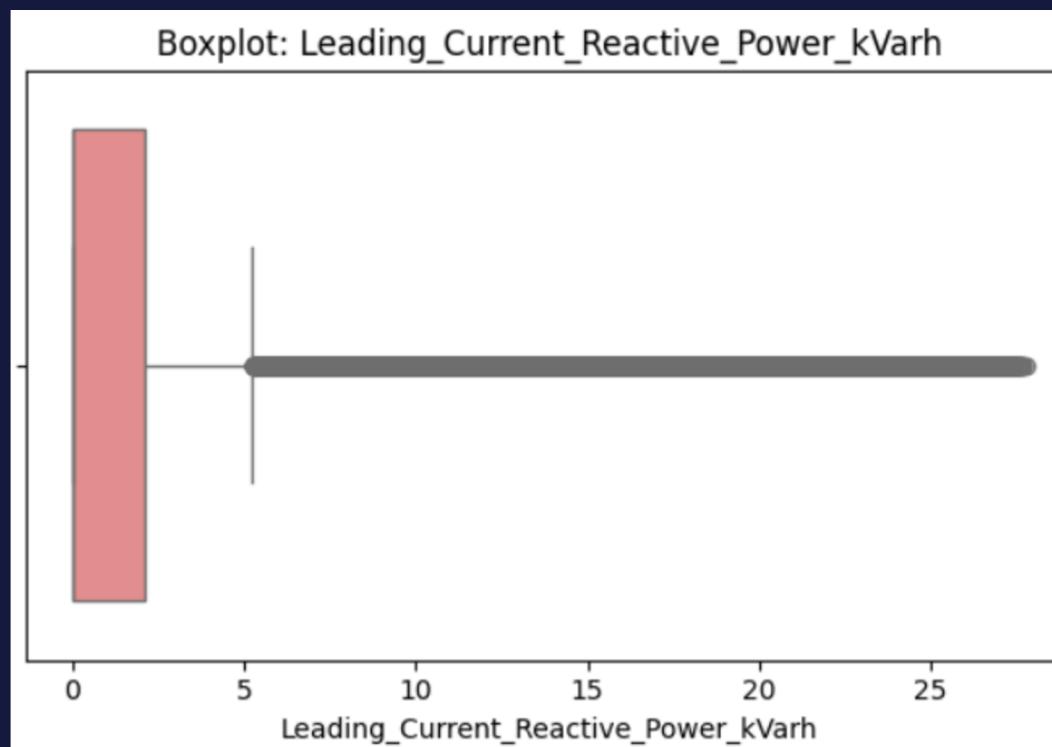
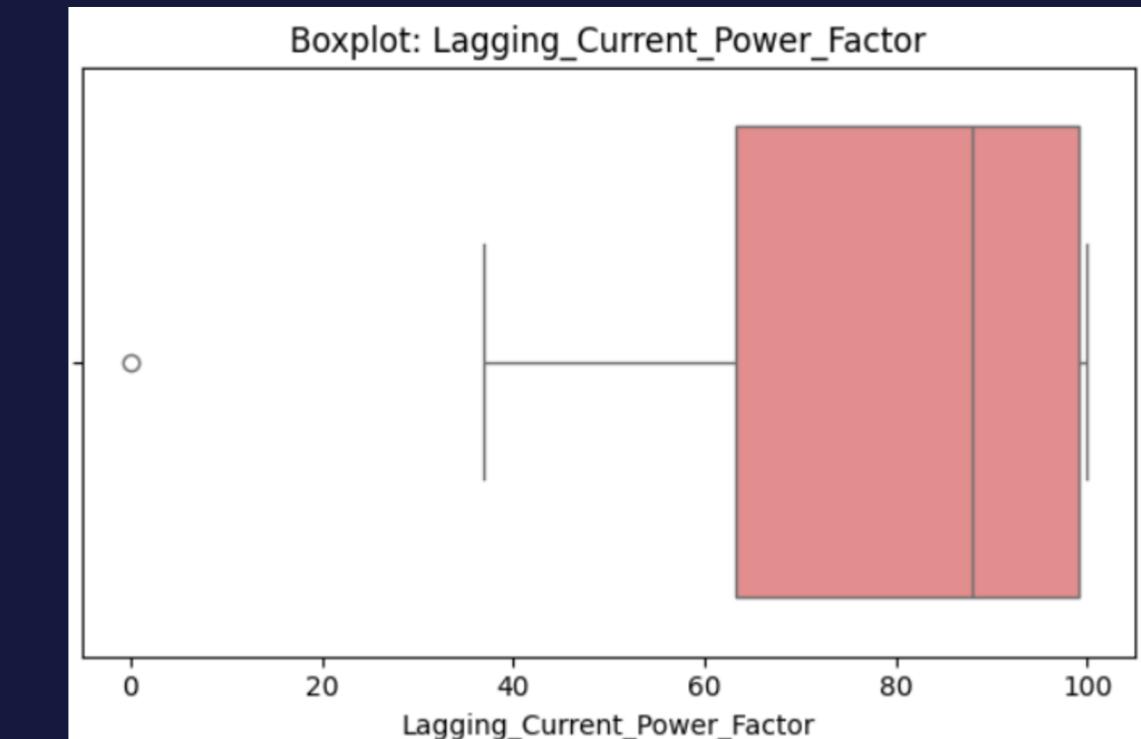
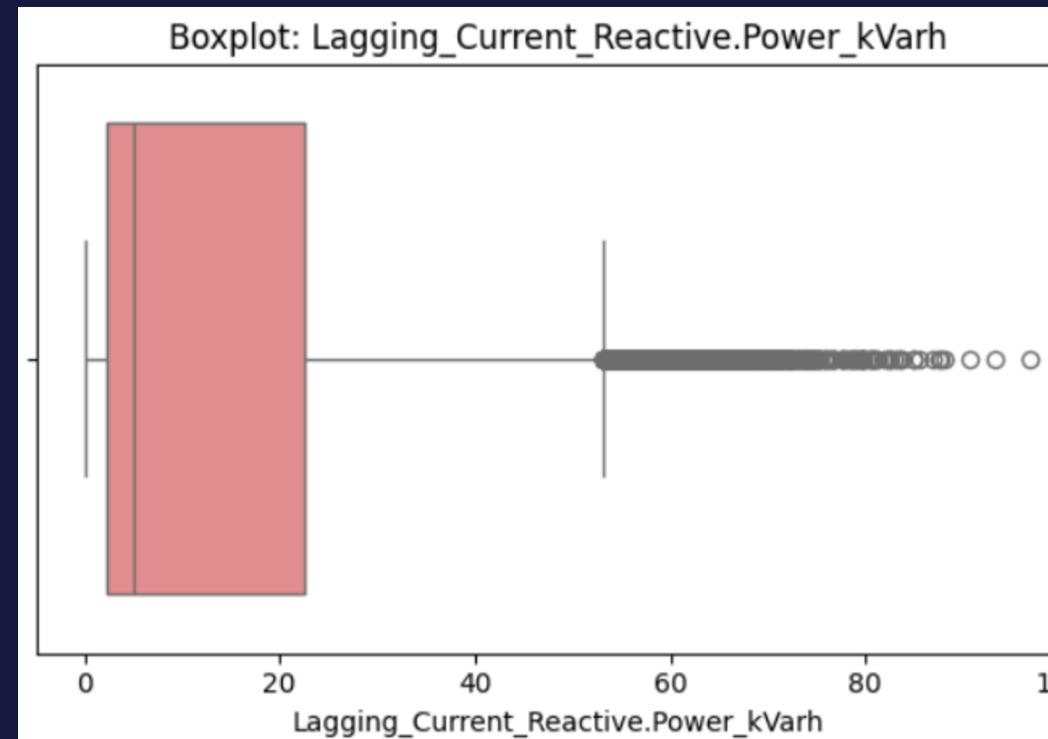
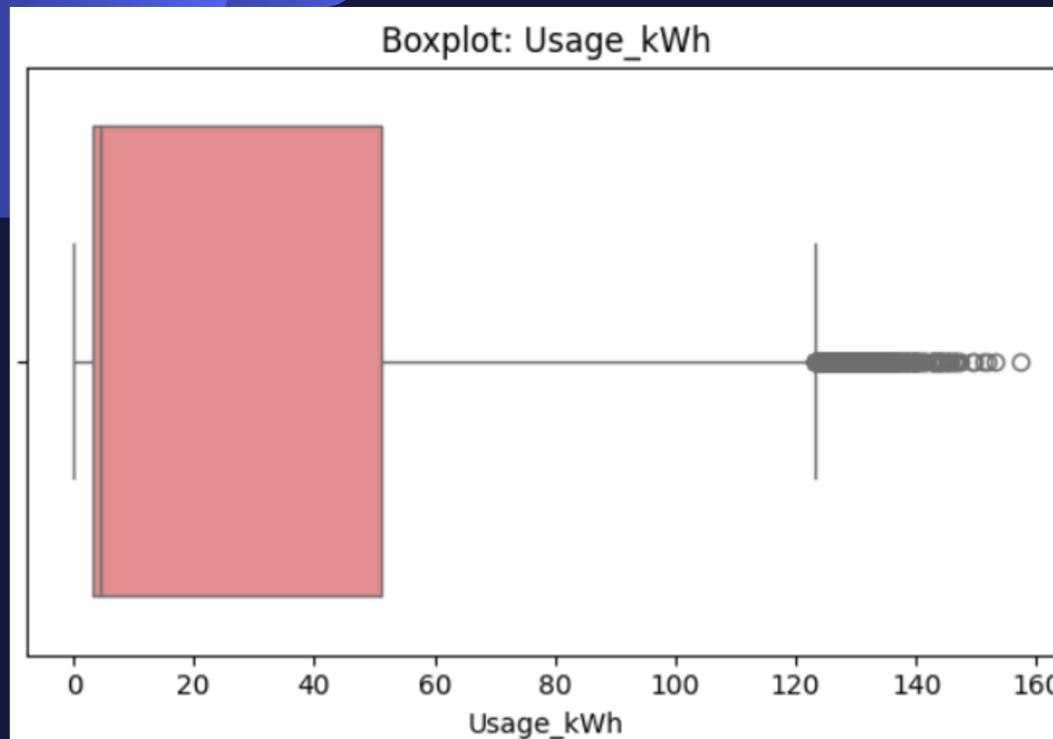
# HISTOGRAMA DE DATOS NUMÉRICOS



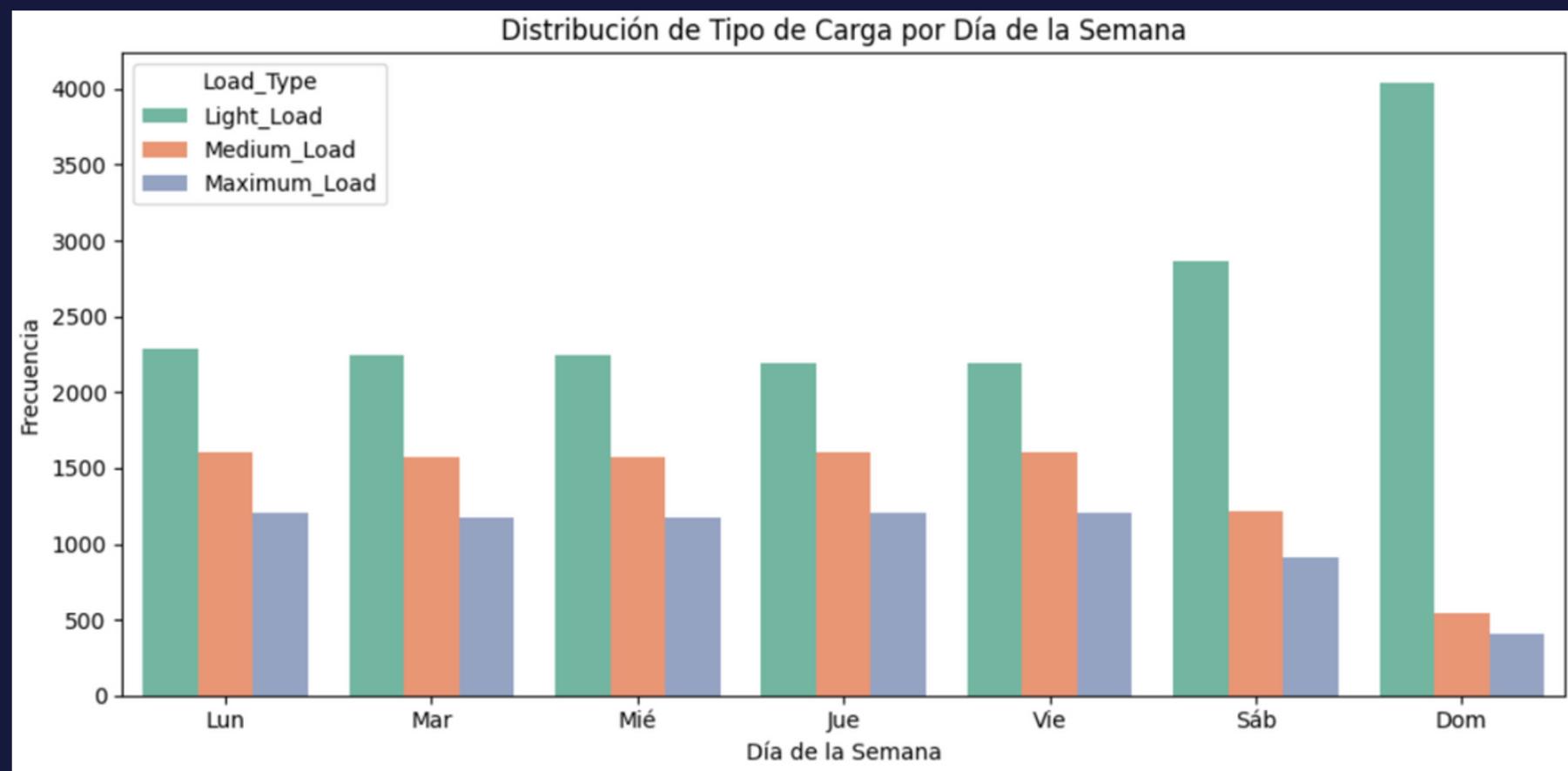
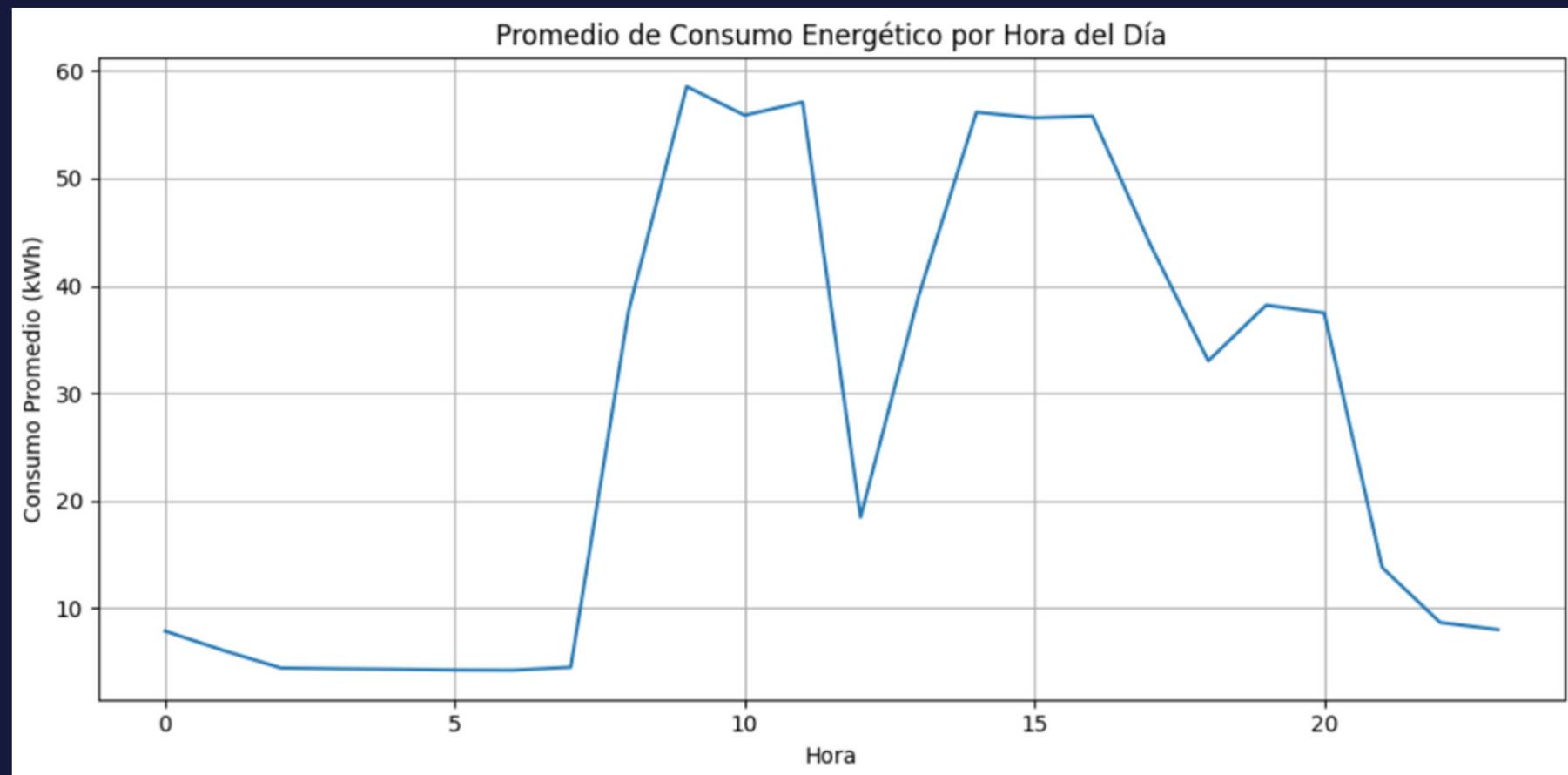
# GRÁFICO DE BARRAS DE DATOS CATEGÓRICOS



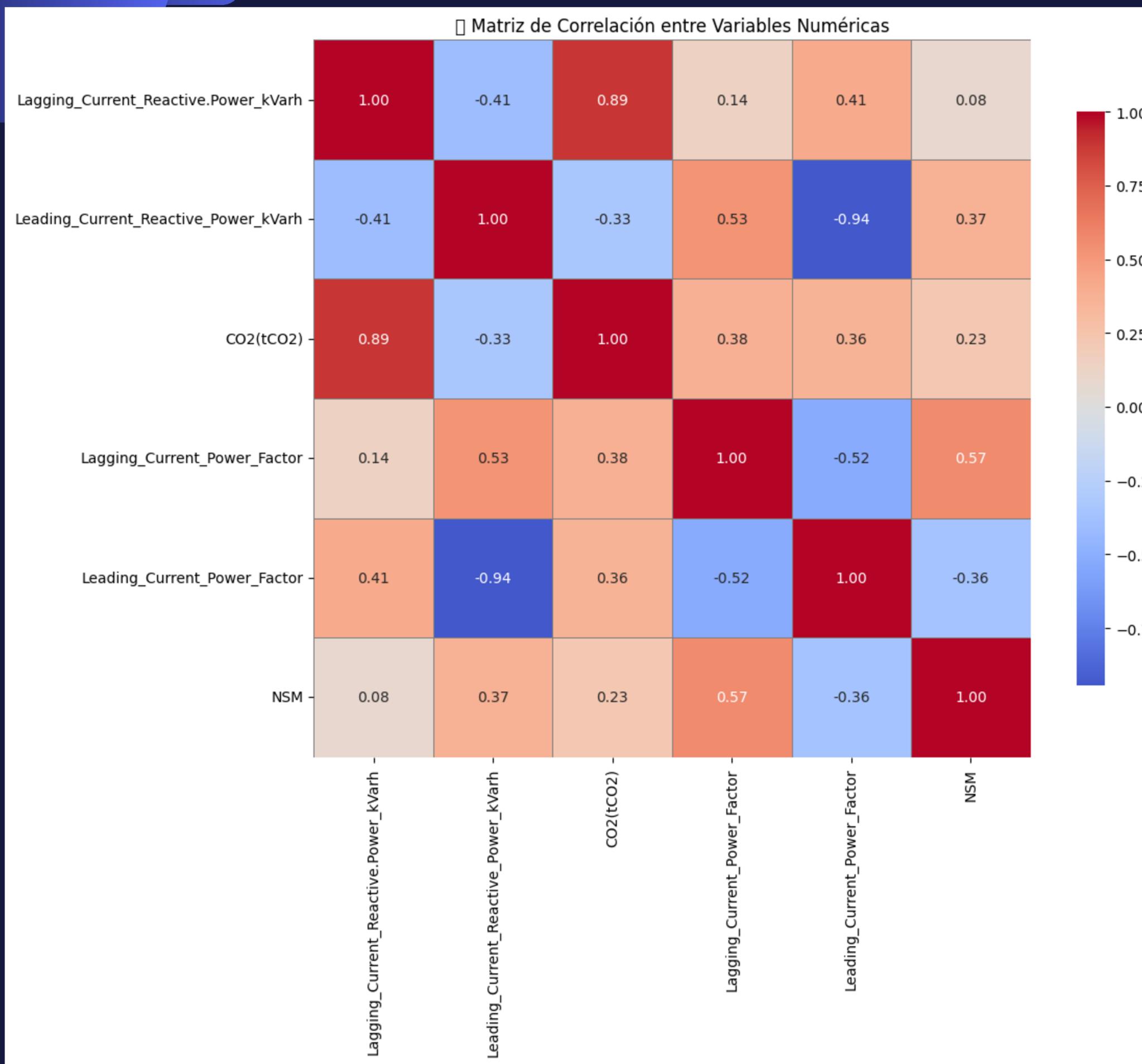
# DIAGRAMA DE CAJA Y BIGOTES



# ANÁLISIS TEMPORAL



# MATRIZ DE CORRELACIÓN



Se excluyeron variables altamente correlacionadas como **Lagging\_Current\_Reactive.Power\_kVarh** y **Leading\_Current\_Reactive\_power\_KVarh\_Factor** para reducir la redundancia y mejorar la eficiencia del modelo.

# 2. MODELAMIENTO

## PROCESO DE PREPARACIÓN DEL DATASET



## PROCESO DE EVALUACIÓN POR CADA MODELO



# 2. MODELAMIENTO

## RESUMEN DE MODELOS UTILIZADOS

Modelo utilizado	Parámetros para optimizar	Parámetros finales	Resultados de la validación cruzada
Random Forest Regressor	<ul style="list-style-type: none"><li>• max_depth=20</li><li>• min_samples_leaf</li><li>min_samples_split</li><li>• n_estimators</li></ul>	max_depth=20, min_samples_leaf=1, min_samples_split=2, n_estimators=200	RMSE: $1.2411 \pm 0.0883$
XGBoost Regressor	<ul style="list-style-type: none"><li>• learning_rate</li><li>• max_depth</li><li>• n_estimators</li><li>• subsample</li></ul>	learning_rate= 0.1 max_depth=6 n_estimators=200 subsample=0.8	RMSE: $1.3236 \pm 0.0582$
Perceptrón multicapa	<ul style="list-style-type: none"><li>• capas intermedias</li><li>unidades por capa</li></ul>	capas intermedias = 1 unidades por capa = 64	RMSE: $0.9286 \pm 0.0885$
Ridge Regression	<ul style="list-style-type: none"><li>• alpha</li></ul>	alpha=0.01	RMSE: $4.8101 \pm 0.1703$

# 4. RESULTADO

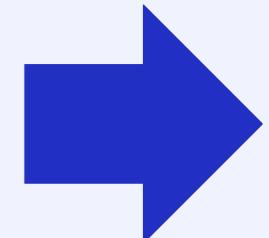
## ENTRENAMIENTO Y PREDICCIONES

```
# Random Forest
rf_model.fit(X_train, y_train)
rf_val_preds = rf_model.predict(X_val)
rf_test_preds = rf_model.predict(X_test)

# XGBoost
xgb_model.fit(X_train_scaled, y_train)
xgb_val_preds = xgb_model.predict(X_val_scaled)
xgb_test_preds = xgb_model.predict(X_test_scaled)

# Ridge Regression
ridge_model.fit(X_train_scaled, y_train)
ridge_val_preds = ridge_model.predict(X_val_scaled)
ridge_test_preds = ridge_model.predict(X_test_scaled)

# Red Neuronal (MLP)
mlp_model.fit(X_train_scaled, y_train)
mlp_val_preds = mlp_model.predict(X_val_scaled)
mlp_test_preds = mlp_model.predict(X_test_scaled)
```



## EVALUACION

Models	Opt. Parameters	Training		Testing	
		RMSE	MAE	RMSE	MAE
XGBoost	max_depth = 6	1.103	1.191	0.541	0.00
RF	max_depth = 20	1.151	0.382	1.134	0.400
Ridge	alpha = 0.01	4.478	2.611	4.222	2.612
MLP	HiddenL = (1,64)	0.914	0.448	0.902	0.448

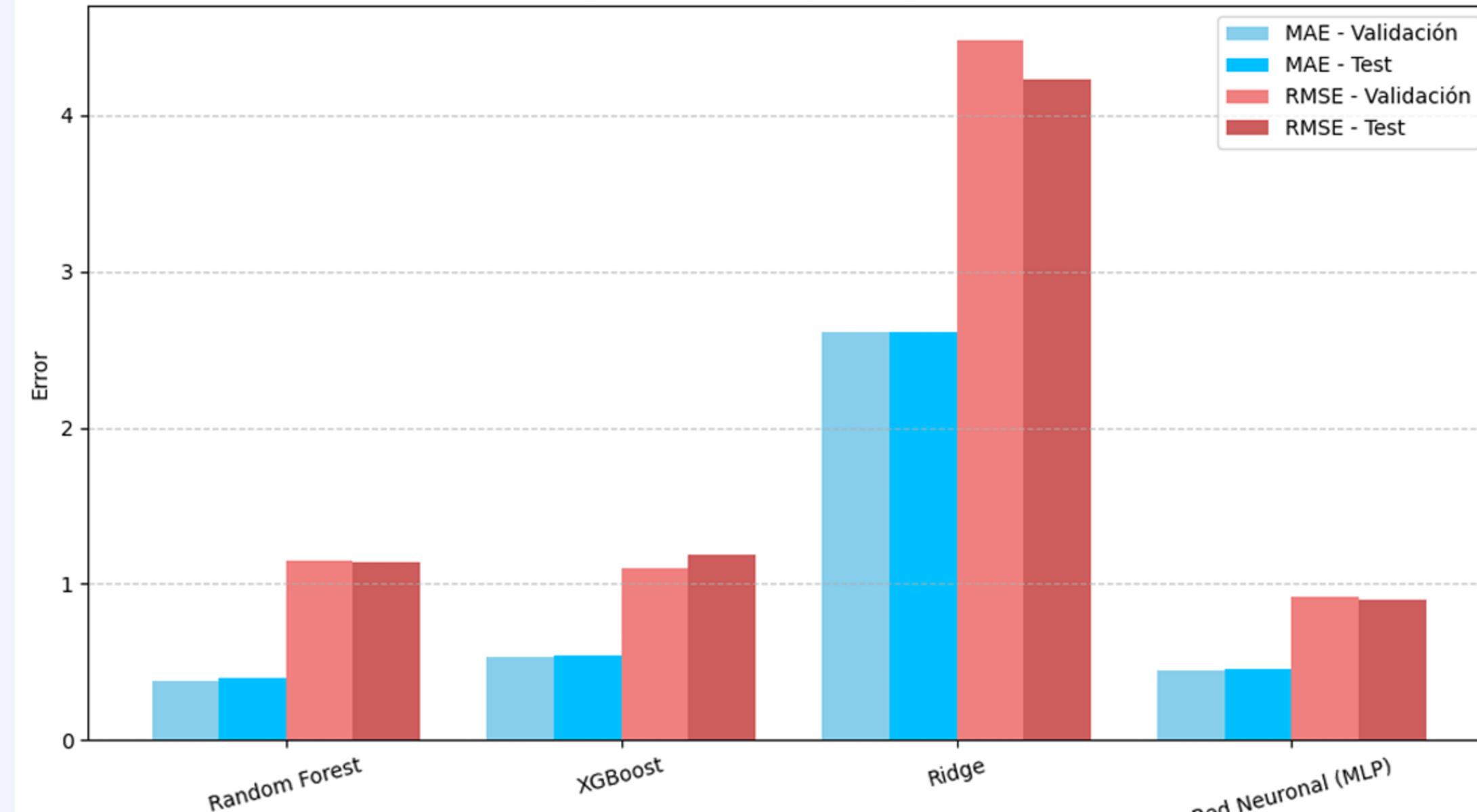
# 4. RESULTADO

Models
XGBoost
RF
Ridge
MLP

A pesar de que la red neuronal presentó métricas ligeramente superiores en términos de MAE y RMSE, se optó por el modelo Random Forest debido a su mayor **interpretabilidad** y **facilidad de ajuste**

## EVALUACION

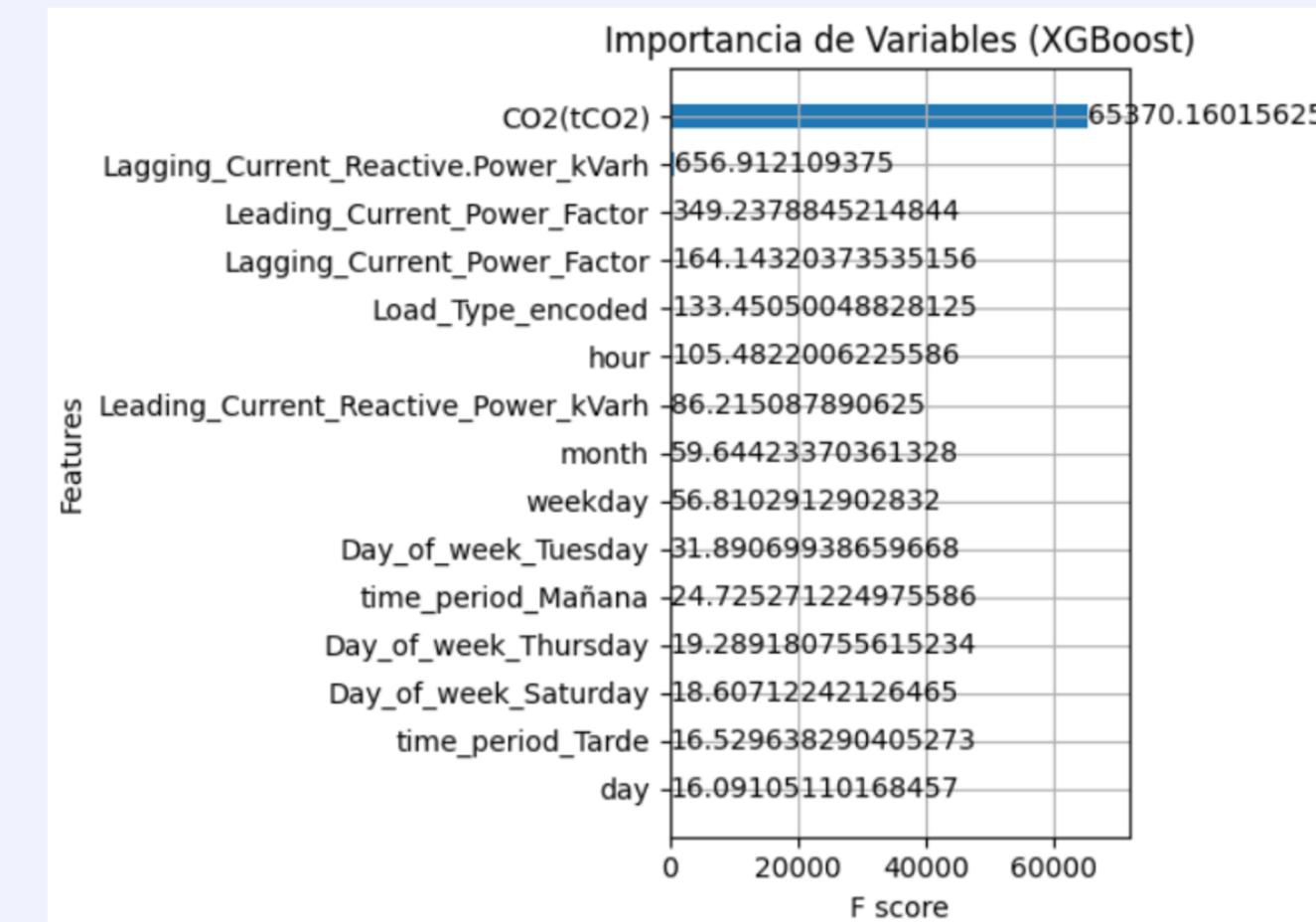
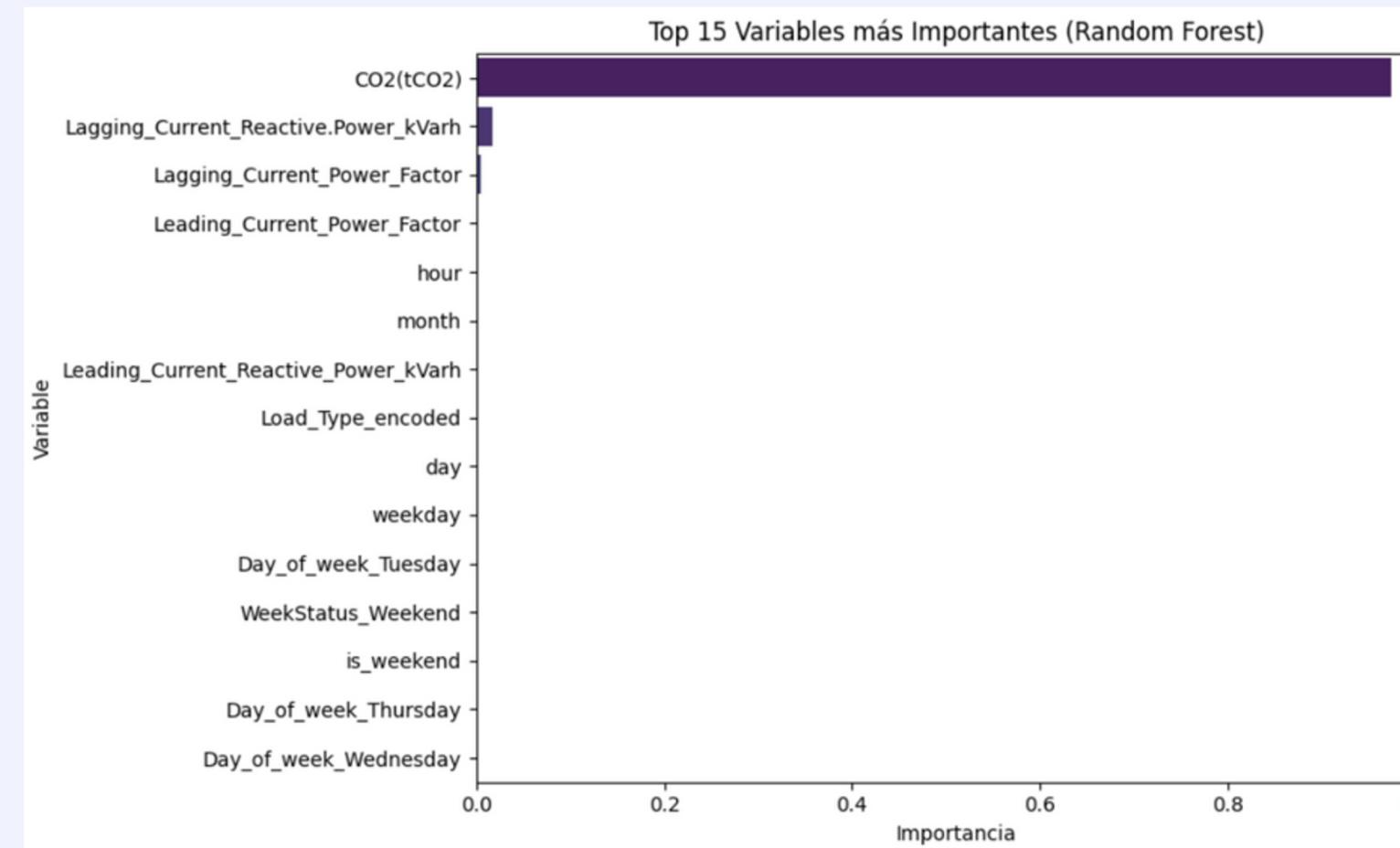
Comparación de MAE y RMSE por modelo



# 4. RESULTADO

## EVALUACIÓN E INTERPRETACIÓN

### Importancia de variables



# 6. ESTRATEGIA DE NEGOCIO

## RESULTADOS

A pesar de que la red neuronal presentó métricas ligeramente superiores en términos de MAE y RMSE, se optó por el modelo Random Forest debido a su mayor interpretabilidad y facilidad de ajuste. El modelo XGBoost se descartó al momento de realizar la validación cruzada, por tener un RMSE de 1.3236 mayor que el de Random Forest 1.24.

## CARACTERÍSTICAS IMPORTANTES

La principal característica que determina el consumo energético es el CO<sub>2</sub> emitido ya es un subproducto directo de los procesos de combustión y reacciones químicas en la producción de acero, que a su vez están intrínsecamente ligados al consumo energético. Una alta emisión de CO<sub>2</sub> generalmente implica un alto consumo de combustibles fósiles o energía eléctrica intensiva.

## PROGRAMAS DE RESPUESTA BAJO DEMANDA

Utilizar las predicciones horarias de Usage\_kWh para identificar períodos de alto consumo de la planta. Si existen programas de "Respuesta a la Demanda" ofrecidos por la compañía eléctrica (donde se incentiva reducir el consumo durante picos de la red), la planta puede planificar la reducción temporal de cargas no críticas durante esos eventos.

## PLANIFICACIÓN DE PRODUCCIÓN BASADA EN COSTO ENERGÉTICO PREDICHO:

Usar predicciones de Usage\_kWh para estimar el costo energético horario/diario. Programar procesos productivos flexibles y de alto consumo energético en períodos de bajo costo energético

## 7. CONCLUSIONES

La **solución desarrollada es robusta y cumple con los objetivos del estudio**: permite predecir el consumo energético con alta precisión y ofrece información accionable para la optimización operativa en la planta siderúrgica. Las variables clave identificadas, como CO2(tCO2) y factores eléctricos, aportan un valor estratégico significativo, sentando las bases para estrategias de gestión energética más eficientes y sostenibles

- 01
- 02
- 03
- 04

El modelo Random Forest seleccionado demostró una capacidad buena para predecir el Usage\_kWh validando así su efectividad para estimar el consumo energético en entornos industriales complejos

El análisis de importancia de características reveló que CO2(tCO2) es el predictor más influyente (superando el 90% de importancia), seguido de otras variables eléctricas

Altos niveles de potencia reactiva inductiva indican un uso intensivo de motores y hornos, lo que se traduce directamente en mayor consumo. La hora refleja los ciclos de producción con mayor demanda.

LA importancia de Leading\_Current\_Power\_Factor sugiere que la corrección del factor de potencia (mediante bancos de capacitores) está activa y tiene un impacto medible