

# 1.Introducción

Este reporte documenta el proceso de desarrollo de un modelo de detección de anomalías en el consumo de gas para Contugas. A partir de los datos históricos operacionales por cliente, se implementó una solución basada en segmentación de clientes y algoritmos de aprendizaje no supervisado. Se utilizaron modelos estadísticos paramétricos (ARIMA y Prophet) como también modelos no supervisados basados en árboles (Isolation Forest) con el objetivo de aprovechar su flexibilidad, precisión y escalabilidad.

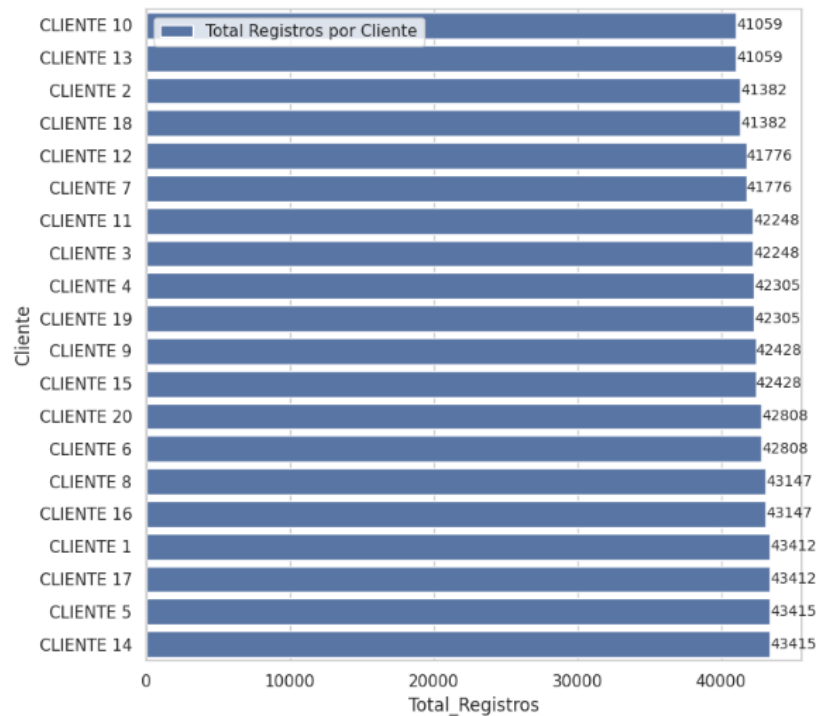
Como punto de partida, se definió el requerimiento analítico principal: identificar desviaciones atípicas en el comportamiento operativo de los clientes (volumen, presión y temperatura), con el fin de apoyar el mantenimiento preventivo, la validación de lecturas y la mejora de eficiencia operativa. El enfoque se diseñó para integrar análisis descriptivo, segmentación de perfiles y detección de anomalías, permitiendo capturar tanto patrones normales como desviaciones críticas de forma automatizada.

## 2.Análisis Exploratorio y Verificación de Supuestos

### 2.1 Descripción del Dataset

El dataset original contenía 847.960 registros y 5 columnas: fecha de lectura, presión, temperatura, volumen de consumo y cliente. A partir de la fecha de lectura, se generaron variables derivadas como mes, día de la semana y hora.

Se realizó una verificación del número de registros por cliente, encontrando una distribución similar entre ellos. Aunque existen pequeñas diferencias, la cantidad de datos es adecuada para modelar el comportamiento individual de cada cliente.



## 2.2 Verificación de calidad de datos

### *Duplicados:*

Se realizó una verificación de registros duplicados utilizando la función duplicated().

**Resultado:** 0 registros duplicados encontrados.

**Acción tomada:** No se encontraron duplicados.

### *Valores nulos:*

Se revisó la existencia de valores nulos en todas las columnas mediante la función isnull().sum().

**Resultado:** no se encontraron valores nulos.

**Acción tomada:** No fue necesaria; todos los registros están poblados, no hay nulos.

### *Datos faltantes:*

Al ser una serie de tiempo con una frecuencia o nivel de detalle horario, se debe validar si hay saltos de tiempo. Esto es importante porque muchos modelos no aceptan saltos temporales.

**Resultado:** Se encontraron vacíos o datos faltantes.

**Acción tomada:** Se modificó de tal manera que se agregaran los valores faltantes con un valor promedio entre sus valores consecutivos.

### *Tipos de datos:*

Se verificó que todas las variables relevantes para el modelado fueran de tipo numérico (float64 o int64), garantizando su adecuación para los algoritmos de machine learning.

```
#   Column      Non-Null Count  Dtype
---  -
0   Fecha       847960 non-null    datetime64[ns]
1   Presion      847960 non-null    float64
2   Temperatura  847960 non-null    float64
3   Volumen      847960 non-null    float64
4   Cliente      847960 non-null    object
dtypes: datetime64[ns](1), float64(3), object(1)
```

## 2.3 Análisis exploratorio

Estadísticas descriptivas:

Se calcularon medidas de tendencia central y dispersión (mean, std, min, max) para las principales variables numéricas.

	Fecha	Presion	Temperatura	Volumen
count	847960	847960.000000	847960.000000	847960.000000
mean	2021-07-08 22:08:19.659418112	16.072957	25.198239	62.328206
min	2019-01-14 00:00:00	2.934873	-5.257899	0.000000
25%	2020-04-13 23:00:00	17.097350	22.693027	0.000000
50%	2021-07-07 06:00:00	17.570449	25.379859	21.773567
75%	2022-10-06 02:15:00	17.694254	27.886244	99.319649
max	2023-12-31 23:00:00	20.307852	50.019853	577.413425
std	NaN	4.186408	3.790497	80.498112

A partir del análisis estadístico se concluyó:

*Sobre la variable volumen.*

- Se evidenció que algunos clientes presentan muchos registros con volumen igual a cero, lo cual resulta inusual.
- Aunque en ciertos casos podría estar asociado a horarios sin operación o pausas programadas, es necesario analizar más a fondo los patrones de consumo, por lo que se implementaron, como se describe más adelante, técnicas de clusterización.

*Sobre la variable temperatura.*

- Se encontraron registros con temperaturas negativas. Esto puede ser normal en sistemas de distribución de gas debido a:
  - Efecto Joule-Thomson: expansión del gas al pasar por válvulas de regulación puede generar enfriamiento y temperaturas bajo 0°C.
  - Ambiente externo: exposición del ducto a zonas frías sin aislamiento térmico adecuado.
  - Velocidad del flujo: Altas velocidades impiden que el gas se caliente por el entorno.
  - Falta de precalentamiento: Algunos puntos de entrega no cuentan con sistemas de calentamiento.
- Para los clientes 5, 7, 12 y 14 se observó una mayor dispersión en la temperatura, con rangos más amplios de variación.
- Se identificó una variación estacional clara: en los primeros meses del año las temperaturas son más altas y hacia fin de año disminuyen, en coherencia con los patrones climáticos de Perú.

*Sobre la variable Presión*

- Se evidenció que los clientes 3 y 11 presentan distribuciones de presión ligeramente diferentes, aunque en general los datos se encuentran centrados en un valor promedio similar al de los demás clientes.

Este análisis reafirma la necesidad de segmentar a los clientes antes de aplicar modelos de detección de anomalías, ya que el comportamiento operativo no es homogéneo entre ellos (es decir, hacer un análisis para cada una de las variables por cada cliente).

## 2.4 Distribución de variables:

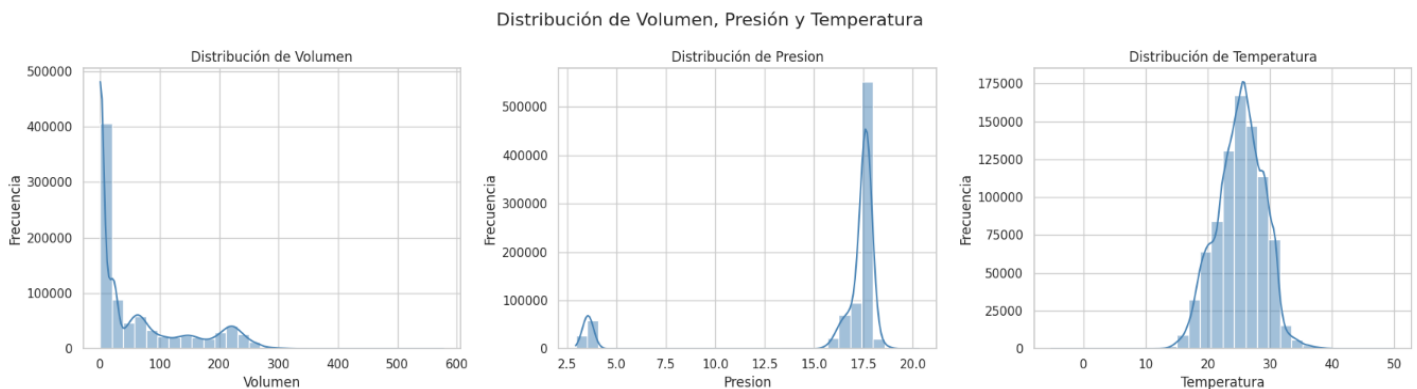
Se graficaron histogramas para las variables Volumen, Presión y Temperatura y se observó que:

**Volumen:** Presenta una distribución altamente asimétrica a la derecha, con una alta concentración de registros en valores bajos (cerca de 0) y una cola larga hacia

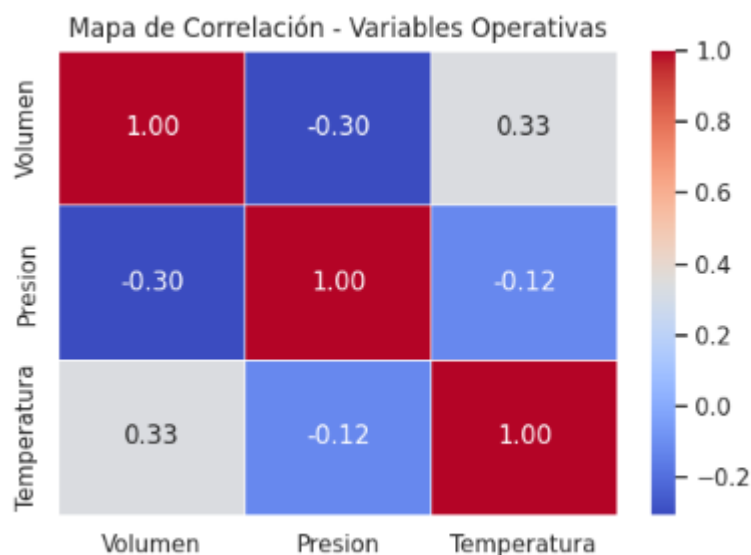
consumos altos. Esto sugiere que la mayoría de los consumos son pequeños, pero existen clientes con picos importantes de volumen.

**Presión:** Muestra una distribución con dos grupos bien diferenciados: Uno concentrado alrededor de los 17–18 psi y otro mucho más pequeño alrededor de 3.5 psi. Se observa que existen al menos dos tipos de condiciones operativas o redes en el sistema: una de presión alta (más común) y otra de presión baja (menos frecuente), lo que nos sugiere nuevamente que es necesario segmentar la población de clientes para diferenciar su comportamiento.

**Temperatura:** La temperatura presenta una distribución normal simétrica, centrada alrededor de 26–27 °C. Esto indica un comportamiento térmico estable y homogéneo en el tiempo.



## 2.5 Correlaciones



Se observa una correlación negativa moderada entre Volumen y Presión (-0.30), lo cual es consistente con principios físicos del comportamiento de gases: a mayor presión, el volumen tiende a disminuir si los demás factores se mantienen constantes.

La Temperatura presenta una correlación positiva leve con el Volumen (0.33), lo que puede sugerir que aumentos en temperatura están asociados con incrementos en el consumo o medición de volumen en ciertos contextos operativos.

Presión y Temperatura tienen una relación muy débil y negativa (-0.12), lo cual indica que su variación no es linealmente dependiente en este conjunto de datos.

Las correlaciones observadas permiten inferir que ninguna de las variables muestra alta correlación lineal, por lo que todas pueden incluirse en modelos multivariantes sin riesgo de multicolinealidad.

Estos comportamientos confirman la relevancia de usar en conjunto estas variables como entrada para el modelo de detección de anomalías, ya que contribuyen de forma complementaria en la identificación de comportamientos atípicos.

## 2.6 Verificación de Supuestos

Como parte del diseño metodológico, se consideraron inicialmente tres enfoques analíticos, agrupados en dos categorías:

- Modelos estadísticos paramétricos: ARIMA y Prophet.
- Modelos no supervisados basados en árboles: Isolation Forest.

Cada modelo fue evaluado considerando los siguientes criterios:

- Cumplimiento de supuestos (estacionariedad, estacionalidad, tipo de variables).
- Robustez frente a ruido o valores atípicos.
- Escalabilidad para aplicarse por cliente.
- Sensibilidad a parámetros.
- Nivel de complejidad técnica e interpretabilidad.

A continuación se presenta una tabla resumen de esta evaluación, que incluye los supuestos requeridos por cada modelo y el tratamiento aplicado en los datos

Aspecto / Supuesto	SARIMA	Prophet	Isolation Forest	DBSCAN	Autoencoder	Tratamiento aplicado
<b>Estacionariedad</b>	Obligatorio	Recomendado	No requerido	No requerido	No requerido	Evaluada con prueba ADF. Solo Volumen fue estacionaria en la mayoría de los clientes.
<b>Normalidad</b>	Recomendado	Recomendado	No requerido	No requerido	No requerido	Evaluada con Shapiro y D'Agostino. Ninguna variable la cumple.
<b>Linealidad</b>	Obligatorio	Recomendado	No requerido	No requerido	No requerido	No se asumió. Se puede optar por modelos que captan relaciones no lineales.
<b>Multilinealidad baja</b>	Recomendado	Recomendado	No requerido	Recomendado	No requerido	Se mantuvo una estructura multivariada.
<b>Formato de series temporales</b>	Obligatorio	Obligatorio	Recomendado	Recomendado	Recomendado	Se organizó por cliente con frecuencia horaria.
<b>Frecuencia temporal regular</b>	Obligatorio	Obligatorio	Recomendado	Recomendado	Recomendado	Se estandarizó la granularidad horaria.
<b>Tipo de variables</b>	Obligatorio	Obligatorio	Obligatorio	Obligatorio	Obligatorio	Se verificó que todas fueran numéricas.
<b>Transformación cíclica (fecha)</b>	No requerido	Obligatorio	Recomendado	Opcional	Recomendado	Se aplicó codificación seno-coseno a hora, día y mes.
<b>Escalado de variables</b>	Opcional	Opcional	Obligatorio	Obligatorio	Obligatorio	Se usó StandardScaler para homogeneizar magnitudes.

Este análisis incorporó modelos como ARIMA y Prophet, los cuales generan predicciones del comportamiento esperado a partir de patrones temporales y dependencias estadísticas del pasado (estacionariedad, estructura temporal). En conjunto, se empleó el modelo Isolation Forest, que permite identificar anomalías sin requerir ajustes específicos por cliente y puede aplicarse de manera eficiente sobre múltiples segmentos.

## 3. Selección de Modelos

### 3.1 Selección de modelos desde el enfoque técnico

Con el fin de seleccionar el enfoque más adecuado para la detección de anomalías en el consumo de gas, se evaluaron distintos modelos desde una perspectiva técnica y operativa. La tabla a continuación resume las principales características, ventajas y limitaciones de cada alternativa, considerando aspectos como escalabilidad, cumplimiento de supuestos estadísticos, robustez ante ruido, facilidad de implementación y nivel de interpretación:

Modelo	Tipo	Ventajas principales	Desventajas / limitaciones
<b>ARIMA</b>	Estadístico paramétrico (modelo lineal de series temporales)	<ul style="list-style-type: none"><li>- Bueno para series con estacionalidad</li><li>- Alta precisión si se cumplen los supuestos.</li></ul>	<ul style="list-style-type: none"><li>- Requiere series estacionarias.</li><li>- Difícil de escalar muchos clientes.</li><li>- Poca tolerancia a datos faltantes o eventos externos.</li></ul>
<b>Prophet</b>	Estadístico semi-paramétrico	<ul style="list-style-type: none"><li>- Fácil de usar e implementar.</li><li>- Maneja estacionalidades múltiples.</li><li>- Soporta cambios de tendencia.</li><li>- Tolerante con los datos faltantes.</li></ul>	<ul style="list-style-type: none"><li>- Supone estacionalidades regulares.</li><li>- Bajo desempeño si hay ruido.</li><li>- Aplica mejor para pronósticos que para detección de anomalías.</li></ul>
<b>Isolation Forest</b>	No supervisado, basado en árboles (detección de anomalías)	<ul style="list-style-type: none"><li>- No requiere etiquetas.</li><li>- Escalable a grandes volúmenes de datos.</li><li>- No asume estructura temporal.</li><li>- Detecta bien valores atípicos multivariados.</li><li>- No necesita normalidad ni estacionariedad.</li></ul>	<ul style="list-style-type: none"><li>- No modela la dinámica temporal directamente.</li><li>- Sensible a la calidad de los datos de entrada.</li><li>- Detecta anomalías pero no pronósticos.</li><li>- Sensible al parámetro de contaminación.</li></ul>

### 3.2 Selección de modelos desde el enfoque de negocio

El proceso de selección de modelos no solo consideró criterios técnicos, sino que también se orientó a resolver los problemas clave del negocio identificados en el caso de negocio de Contugas: la detección de anomalías y la sencillez de la solución (hará que sea más rápido).



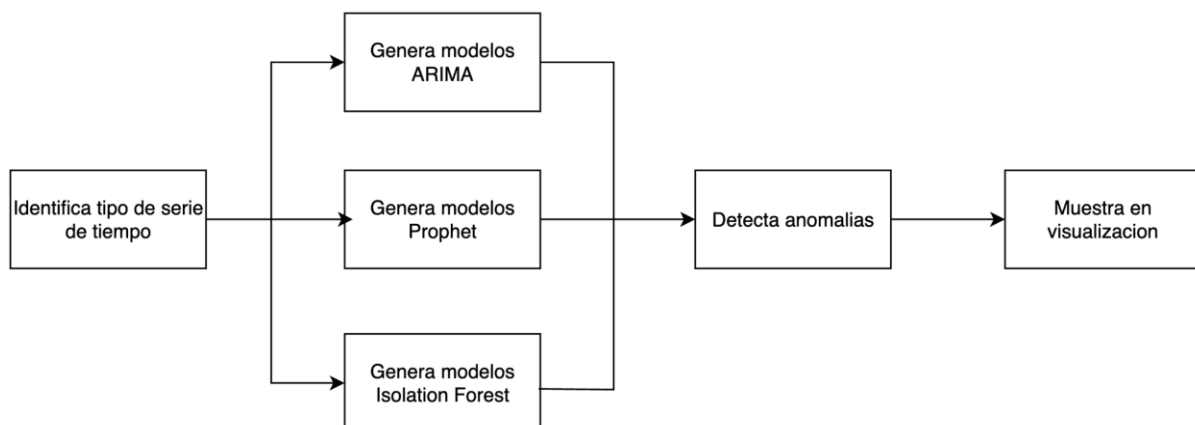
A partir del análisis de necesidades estratégicas, se identificaron las siguientes prioridades para la solución analítica:

- Monitoreo automático y en tiempo real, sin intervención manual constante.
- Escalabilidad por cliente, debido a la amplia red de usuarios con patrones distintos.
- Sensibilidad a desviaciones en variables clave (volumen, presión, temperatura).
- Capacidad de integración con el sistema de monitoreo actual de Contugas.
- Facilidad de mantenimiento y mejora continua del modelo.

## 4. Entrenamiento y Validación de Modelos

### 4. 1 Definición del flujo de trabajo

Para este ejercicio se definió una estrategia sencilla para la identificación del mejor algoritmo a implementar como también la identificación de los parámetros que mejor ajustan los modelos a cada una serie de tiempo:



#### 1. Identificar el tipo de serie de tiempo:

Con base en la exploración de los datos se logra detectar que cada cliente tiene 3 variables y que se encuentran definidas como series de tiempo. Las series de tiempo se pueden clasificar principalmente si son estacionarias y/o son estacionales. Con esta clasificación se puede agrupar las series de tiempo de una manera eficiente para la generación de los modelos ya que es más fácil definir qué tipo de algoritmo implementar en base a su clasificación.

- Prueba de Estacionalidad:

La estacionalidad es un patrón que se repite regularmente a lo largo del tiempo. Puede ser diaria, semanal, mensual, anual, etc. Para la detección de la estacionalidad en una serie de tiempo se logra a través de dos opciones: por medio visual de la serie temporal (bastante subjetivo y requiere experiencia) y por medio de la autocorrelación a distintos lags (ACF). Para la automatización de este proceso se decide tomar el análisis de ACF por su sencillez y adicionalmente que puede detectarse si hay auto correlación a varios niveles (por ejemplo, estacionalidades diarias, semanales, mensuales, trimestrales, etc)

- Prueba de Estacionariedad:

Una serie de tiempo es estacionaria si sus propiedades estadísticas se mantienen constantes en el tiempo. Es decir, deben cumplir con los siguientes supuestos:

1. Media constante: La serie no tiene tendencia creciente ni decreciente.
2. Varianza constante: La dispersión no cambia con el tiempo.
3. Autocorrelación estable: La relación entre un punto y sus rezagos no cambia con el tiempo.

Identificar la estacionariedad es importante porque para usar algoritmos como ARIMA requiere que se cumplan estos supuestos de estacionariedad. Para detectar la estacionariedad de una serie de tiempo, se realiza normalmente la prueba o test de Dickey-Fuller Aumentado (ADF) el cual evalúa la hipótesis:

- **H<sub>0</sub> (nula):** la serie NO es estacionaria (tiene raíz unitaria).
- **H<sub>1</sub> (alternativa):** la serie es estacionaria.

## 4.2 Implementación de modelos.

### 1. Modelos ARIMA:

Se usa el algoritmo ARIMA para generar los modelos sobre las series de tiempo que fueron detectadas como Estacionarias ya que cumplen con los supuestos estadísticos y como bien se sabe estos algoritmos tienen muy buen desempeño cuando se cumplen dichos supuestos. ARIMA significa:

- A: AutoRegresivo (AR) → usa valores pasados para predecir el presente.
- I: Integrado (I) → diferencia los datos para hacerlos estacionarios.
- MA: Media Móvil (MA) → modela el error como combinación de errores pasados.

### 2. Modelos Prophet:

Prophet es una herramienta (open source creada por Meta) pensada para hacer pronósticos automáticos y precisos de series de tiempo que presentan:

- Tendencia.
- Estacionalidad.

- Festividades o eventos especiales

Es muy fácil de usar ya que no requiere de mucho ajuste de parámetros con lo cual la automatización es mucho más sencilla.

Este modelo se aplicará solo para las series que fueron detectadas como estacionales.

### 3. Modelos Isolation Forest:

Es un algoritmo de machine learning no supervisado, basado en árboles de decisión, que busca aislar anomalías en un conjunto de datos. A diferencia de otros métodos que modelan el comportamiento normal, Isolation Forest se enfoca en aislar las observaciones inusuales.

Es muy bueno detectando anomalías y puede implementarse en series de tiempo, la única desventaja es que no tiene en cuenta características como la estacionalidad y estacionariedad.

Este tipo de algoritmos se pueden implementar en series que no son estacionales ni estacionarias, sin necesidad de complicar el algoritmo de detección de anomalías.

## 4.3 Calibración de parámetros.

### 1. Calibración modelo ARIMA

Se diseñó una función en Python que automatiza el proceso de ajuste de modelos ARIMA sobre series de tiempo univariadas. Esta función permite encontrar el mejor modelo posible dentro de un rango definido de parámetros ( $p$ ,  $d$ ,  $q$ ), utilizando el algoritmo `auto_arima` de forma eficiente y sin necesidad de intervención manual.

El flujo de la función incluye los siguientes pasos:

1. **Limpieza de datos:** Se eliminan valores nulos para asegurar que el ajuste del modelo sea robusto.
2. **Selección automática del modelo:** Se emplea un proceso de búsqueda escalonada para identificar el mejor modelo ARIMA en función de criterios estadísticos como AIC.
3. **Predicción dentro de la muestra:** Una vez ajustado el modelo, se generan predicciones sobre el mismo rango temporal de los datos de entrenamiento para evaluar su precisión.
4. **Cálculo y análisis de residuos:** Se comparan los valores reales con las predicciones para obtener los residuos, y se evalúan posibles anomalías utilizando un umbral basado en desviaciones estándar. Esto permite detectar puntos que el modelo no logra explicar adecuadamente. Se consideran anómalos aquellos puntos cuyo error (residuo) está a más de 3 desviaciones estándar del promedio.

## 2. Calibración modelo PROPHET

Se desarrolló una función en Python que automatiza el entrenamiento del modelo Prophet, incorporando múltiples componentes estacionales detectados previamente en la serie de tiempo. El flujo es el siguiente:

1. **Preparación de la serie:** La serie se transforma al formato requerido por Prophet.
2. **Filtrado inteligente de estacionalidades:** Se aplica una lógica para descartar periodos redundantes o múltiplos entre sí (los periodos estacionales se encontraron con la prueba de estacionalidad de cada serie). Esto evita añadir estacionalidades innecesarias que puedan causar sobreajuste.
3. **Ajuste del modelo Prophet:** Se inicializa un modelo Prophet con el ancho de intervalo de confianza deseado (inicialmente el 95%). Luego, para cada periodo detectado y filtrado, se añade una componente estacional personalizada utilizando un número de términos de Fourier que varía en función de la duración del ciclo (más largo, mayor complejidad permitida). Este es el ajuste paramétrico más importante dentro del modelo.
4. **Entrenamiento y evaluación:** El modelo se ajusta a la serie original y se generan predicciones dentro del mismo rango temporal. Se calculan los residuos para evaluar la precisión del modelo.
5. **Detección de anomalías:** Se consideran anómalos aquellos puntos cuyo error (residuo) está a más de 3 desviaciones estándar del promedio.

## 3. Calibración modelo Isolation Forest

Se desarrolló una función en python que automatiza la creación del modelo Isolation Forest:

**1. Limpieza inicial:** La serie de tiempo es procesada para eliminar valores faltantes, asegurando una entrada limpia para el análisis.

**2. Construcción de características:** Dado que Isolation Forest es un algoritmo basado en árboles y no entiende el comportamiento de una serie temporal, se extraen varias características numéricas:

- Media móvil: representa la tendencia local.
- Desviación estándar: indica la variabilidad local.
- Valores rezagados (lags): se incluyen uno o más pasos anteriores como referencia.

**3. Entrenamiento del modelo Isolation Forest:** Se entrena el modelo con las características creadas. Isolation Forest identifica los puntos que pueden aislarse rápidamente con pocas divisiones en un bosque aleatorio de árboles de decisión. Esos puntos son considerados anómalos.

El parámetro contamination (parámetro que puede ser ajustado) indica qué proporción estimada del total de datos se considera potencialmente anómala (en la librería sklearn puede detectarlo automáticamente).

**4. Predicción de anomalías:** El modelo asigna una etiqueta a cada punto: normal (1) o anómalo (-1).

#### 4.4 Métricas usadas

##### 5.4.1 ARIMA

**Métrica usada para selección del modelo:**  
AIC (Akaike Information Criterion): Se minimiza automáticamente con auto\_arima para elegir el orden óptimo (p, d, q).

**Métrica para evaluar ajuste:**

- Residuos (errores) en la predicción dentro de la muestra.
- Detección de anomalías por desviación estándar: se marcan como anómalos los puntos cuyo residuo es mayor a  $3\sigma$  (por defecto).

##### 5.4.2 Prophet

- Residuos (errores) entre valores reales y predicción.
- Se usa la métrica RMSE internamente para definir el mejor ajuste del modelo.

##### 5.4.3 Isolation Forest

El modelo no usa una métrica explícita para su ajuste, pero internamente basa sus decisiones en selección de anomalos:

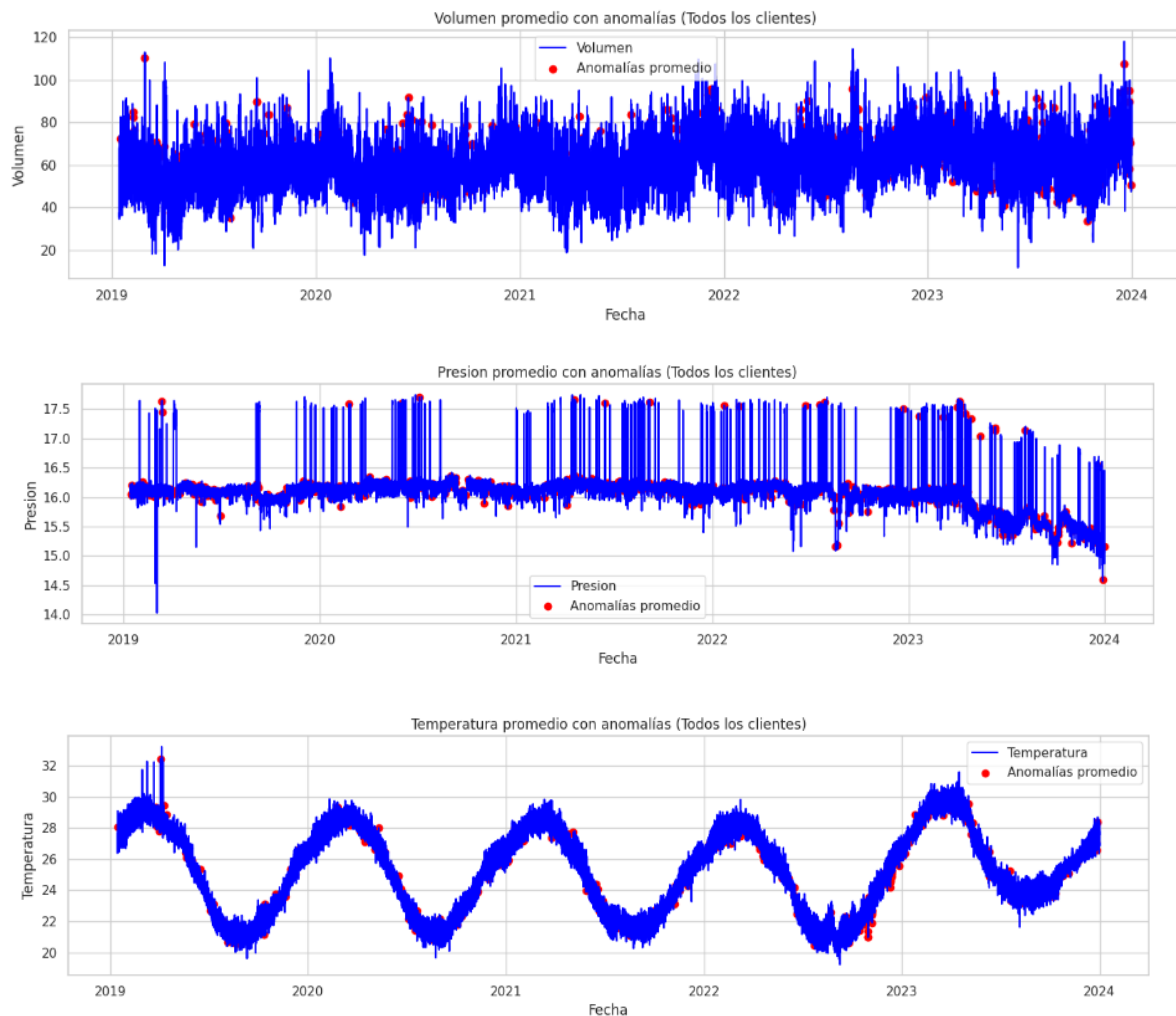
- Número de divisiones necesarias para aislar un punto.
- Cuanto más rápido se aísla, más anómalo se considera.

## 6.Resultados y Análisis

Con el fin de evaluar el desempeño del modelo implementado, se realizó un análisis integral de los resultados generados. Inicialmente, se exploraron los patrones de comportamiento y las anomalías detectadas a nivel agregado para todas las métricas y clientes, con el objetivo de identificar tendencias generales, comportamientos estacionales y eventos atípicos. Posteriormente, se seleccionaron casos específicos por cliente para realizar experimentos detallados que permitieran validar la robustez del modelo en distintos contextos operativos y profundizar en la interpretación técnica de las anomalías detectadas.

**Visualización de anomalías detectadas:**

A partir de las siguientes gráficas se pueden observar los resultados que está generando la ejecución del modelo, donde se observa que el modelo implementado demostró capacidad para capturar tanto estacionalidades como tendencias generales en los datos de presión, temperatura y volumen. Prophet, en particular, facilitó el modelado de patrones múltiples con estacionalidades detectadas automáticamente, mientras que SARIMA fue efectivo en series estacionarias.



Las anomalías detectadas (marcadas en rojo) se superponen sobre las curvas de tendencia promedio, lo que permite visualizar claramente desvíos inusuales tanto en picos como en caídas.

- **Temperatura:** La estacionalidad anual es evidente y muy estable a lo largo del tiempo, lo que facilita un entrenamiento más robusto del modelo y una detección precisa de rupturas en ese patrón. Las anomalías aparecen, en su mayoría, cuando hay desviaciones significativas respecto al comportamiento cíclico esperado.
- **Presión:** Aunque muestra una tendencia general ligeramente decreciente, la serie está marcada por frecuentes interrupciones y picos negativos, lo que genera una

mayor cantidad de falsos positivos. Sin embargo, estos mismos eventos permiten alertar sobre posibles caídas sistemáticas hacia el final del periodo analizado.

- **Volumen:** Se trata de la serie más errática y ruidosa entre las tres. A pesar de ello, el modelo logró identificar múltiples picos anómalos, particularmente en los extremos de la distribución. Estas anomalías pueden estar asociadas a eventos operacionales específicos o a comportamientos atípicos de consumo que requieren revisión.

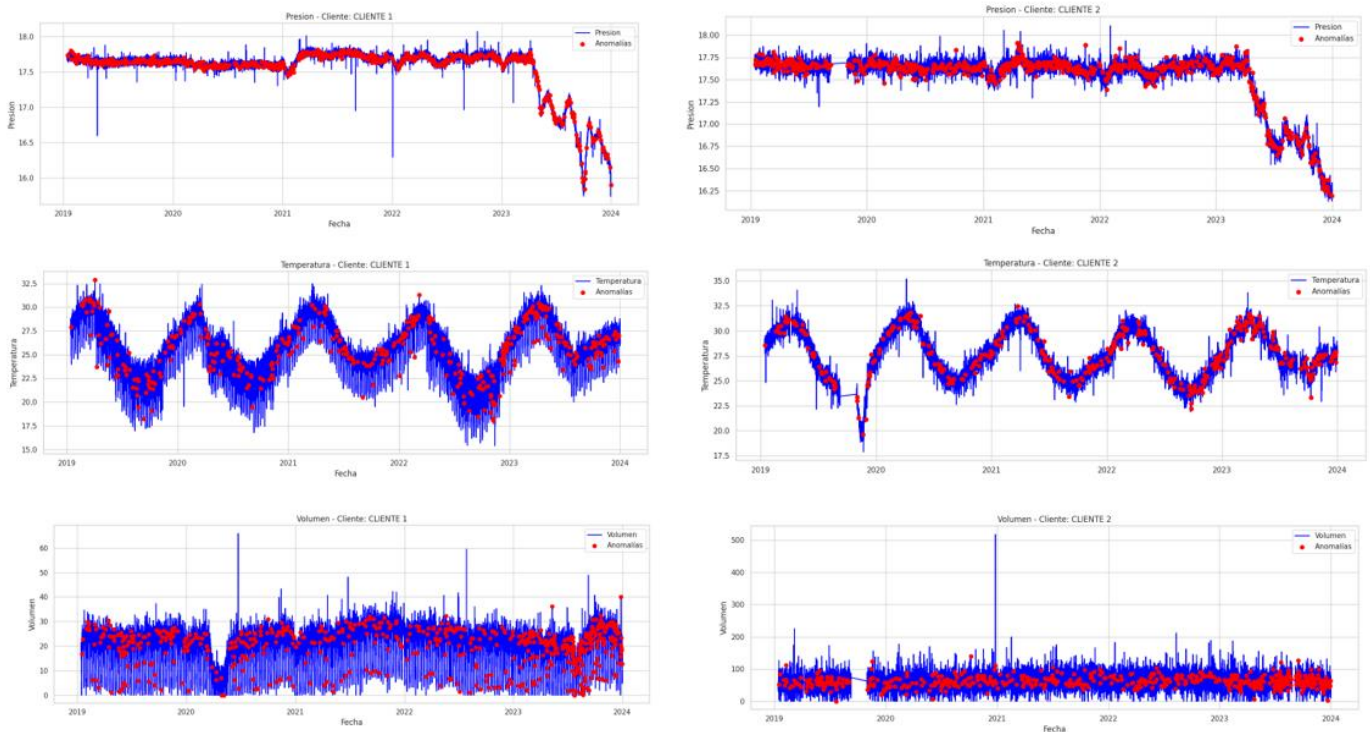
### Experimentos con casos (clientes específicos):

Para validar la efectividad del modelo implementado, se revisó en detalle el comportamiento de las detecciones de anomalías en algunos clientes específicos, en la siguiente tabla se resumen los hallazgos por cada métrica analizada:

Cliente	Métrica	Patrón General	Análisis de Anomalías	Conclusión
CLIENTE 1	Presión	Tendencia estable hasta 2023, luego caída progresiva	Alta concentración de anomalías en fase de descenso	Revisar posible degradación operativa o cambio estructural
CLIENTE 1	Temperatura	Estacionalidad anual clara con ciclos consistentes	Anomalías en quiebres del patrón estacional	Modelo robusto para estacionalidad; seguir monitoreo
CLIENTE 1	Volumen	Serie errática con múltiples picos y caídas	Anomalías en los extremos; algunas falsas alarmas posibles	Requiere validación para evitar falsos positivos
CLIENTE 2	Presión	Tendencia estable hasta 2023, luego descenso sistemático	Anomalías marcan claramente el cambio estructural en presión	Indicador de evento sostenido; investigar causas sistémicas
CLIENTE 2	Temperatura	Estacionalidad clara con algunas interrupciones	Detecciones alineadas con distorsiones del patrón	Buen desempeño del modelo para detección de ruptura estacional
CLIENTE 2	Volumen	Serie altamente ruidosa con valores extremos puntuales	Identifica múltiples valores atípicos en rangos extremos	Atípicos detectados deben contrastarse con eventos operativos para evitar interpretaciones



				erróneas
--	--	--	--	----------



En resumen, los resultados permiten afirmar que el modelo implementado es capaz de detectar patrones anómalos con buena precisión, especialmente en métricas con comportamiento estacional. No obstante, para afinar su utilidad operativa, es clave contrastar estos hallazgos con conocimiento de campo, extender el análisis a más clientes o regiones, y definir umbrales o alertas automáticas para facilitar la toma de decisiones.

## 7. Plan de Implementación del Artefacto

### 7.1 Funcionalidades implementadas

- **Preparación y validación de datos:** Se realizó validación de calidad de datos, transformación de variables temporales (mes, día, hora) y verificación de tipo de datos.



- **Verificación de supuestos estadísticos:** Se evaluó la estacionalidad y estacionariedad de las series por cliente y variable, para orientar la selección de modelos.
- **Modelos seleccionados y ajustados por cliente y variable:** Se implementaron modelos ARIMA, Prophet e Isolation Forest según la naturaleza de cada serie, con calibración automatizada de parámetros.
- **Detección automática de anomalías:** Cada modelo marca los puntos anómalos según criterios estadísticos definidos.
- **Dashboard preliminar:** Se desarrolló una visualización que permite explorar el comportamiento por variable (presión, temperatura, volumen), aplicar filtros por cliente y fecha, y ver las alertas de anomalías en contexto.

## 7.2 Pendientes por implementar

- **Visualización web de alertas en entorno productivo:** Migrar el dashboard actual a una solución accesible vía web con autenticación y acceso controlado.
- **Automatización del pipeline analítico:** Desarrollar un flujo de actualización automatizado que permita ejecutar el modelo con nueva data cada hora o en la frecuencia deseada.
- **Notificación de anomalías:** Diseñar un mecanismo de alertas cuando se detecten desviaciones críticas.
- **Documentación técnica y operativa:** Completar manuales de uso, configuración de parámetros y lineamientos para mantenimiento evolutivo del sistema.