

Detección de anomalías en el consumo de gas de clientes industriales de la filial Contugas.

Mauricio Gonzalez Caro

Eva Karina Diaz Gavalo

Juan Felipe Padilla Sepúlveda

Andrés Eduardo Quiñones Ortiz

Universidad de los Andes

Maestría en Inteligencia Analítica de Datos

Gerencia de Proyectos de Analytics

2024

Continuando con el proyecto desarrollado para Contugas, procederemos a abordar aspectos relacionados con los datos. La información suministrada por Contugas proviene de medidores de gas industriales que miden tres variables clave: volumen [m³], temperatura [°C] y presión [bar]. Estos medidores están instalados en los puntos de consumo de gas de cada cliente industrial y recopilan información de manera continua, proporcionando datos detallados y en tiempo real sobre el consumo de gas. La captura de datos se realiza a nivel horario, lo que permite un monitoreo preciso de las fluctuaciones en el consumo de gas y en las condiciones operativas. Para este proyecto en particular se cuenta con datos estáticos, desde 2019 hasta 2023.

La importancia de estos datos radica en que reflejan no solo los patrones normales de operación de cada cliente, sino también posibles comportamientos anómalos que podrían ser indicativos de fallas en el sistema, irregularidades en los equipos o problemas operativos. Estas anomalías pueden manifestarse de diferentes maneras, dependiendo de la variable afectada:

- **Presión:** Una caída abrupta podría indicar fugas, mientras que un aumento inesperado podría señalar bloqueos o malfunciones. Las fluctuaciones normales son esperadas debido a cambios en la demanda, pero las anomalías representan desviaciones significativas de estos patrones.
- **Temperatura:** La temperatura puede estar influenciada por factores externos como el clima. Sin embargo, un aumento en la temperatura sin una variación correspondiente en la presión podría deberse a condiciones ambientales en lugar de problemas operativos. Si tanto presión como temperatura aumentan de manera considerable y simultánea, esto podría indicar una obstrucción o expansión inusual del gas.
- **Volumen (Consumo):** Un consumo anómalo podría reflejar alteraciones en el medidor, fugas, o lecturas defectuosas. Por ejemplo, un consumo bajo o nulo durante horas operativas esperadas podría señalar problemas en el flujo o fallas en el equipo.

Características de calidad de los datos

Para evaluar la calidad de los datos suministrados por Contugas, se deben reconocer algunas de sus dimensiones de valor, entre estas se encuentran:

- **Granularidad:** La información se encuentra a nivel horario, es decir por día se deberían encontrar 24 registros, uno por cada hora; esto permite un análisis detallado del consumo y de las posibles anomalías, por lo que esta granularidad es beneficiosa para el objetivo del análisis
- **Fidelidad y exactitud:** Contugas es garante de que los medidores que capturan la información, son precisos y deben estar al día con los mantenimientos preventivos para asegurar la confiabilidad de los registros.
- **Edad:** si bien los datos contemplan 5 años de registros, datos más recientes podrían ser relevantes para detectar tendencias actuales

A continuación se describirán los aspectos relacionados con las características de los datos:

1. **Formato:** Se validaron los tipos de datos por cada columna y se logra evidenciar que los datos son apropiados para el tipo de datos de cada variable, en donde las fechas son tipo *datetime64*, lo cual resulta apropiado para analizar series de tiempo y facilita el manejo de datos por rangos específicos de tiempo. Las variables numéricas como volumen, temperatura y presión están almacenadas como tipo *float*, es decir, números con decimales. Finalmente la columna Cliente es de tipo *object*.

#	Column	Non-Null Count	Dtype		#	Column	Non-Null Count	Dtype
0	Fecha	847960 non-null	datetime64[ns]		0	Fecha	847960 non-null	datetime64[ns]
1	Presion	847960 non-null	float64		1	Presion	847960 non-null	float64
2	Temperatura	847960 non-null	float64		2	Temperatura	847960 non-null	float64
3	Volumen	847960 non-null	float64		3	Volumen	847960 non-null	float64
4	Cliente	847960 non-null	object		4	Cliente	847960 non-null	category

De acuerdo a lo anterior, la mayoría de las variables cuentan con un tipo adecuado de datos, únicamente en el caso de Cliente se transformará a tipo category para optimizar el uso de memoria en operaciones de agrupamiento, el resto de variables permanecerán con su formato original.

2. **Compleitud:** Se evaluaron las fechas mínimas y máximas de los registros, que abarcan desde el 14 de enero de 2019 a las 00:00:00 hasta el 31 de diciembre de 2023 a las 23:00:00, cubriendo un total de 1813 días, teniendo en cuenta que el 2020 fue año bisiesto. Según los años y las horas establecidas, cada cliente debería tener 43512 registros para cubrir los cinco años de datos. En la siguiente tabla se presenta un resumen anual de los días y registros esperados:

Año	Min.	Máx.	Días	Registros
2019	14/01/2019	31/12/2019	352	8448
2020	01/01/2020	31/12/2020	366	8784
2021	01/01/2021	31/12/2021	365	8760
2022	01/01/2022	31/12/2022	365	8760
2023	01/01/2023	31/12/2023	365	8760
Total			1813	43512

Al comparar lo esperado con lo realmente almacenado se obtuvo lo siguiente:

- Total de registros actuales: 847960
- Total de registros faltantes: 22280
- Porcentaje de faltantes general: 2.56%

Este porcentaje de faltantes se extrajo también a nivel de clientes, para tener un nivel de detalle mayor. Algunos clientes como 10 y 13 presentan un gran porcentaje de faltantes, del 5.64%, lo cual podría ser indicio de anomalías. Por otra parte, se debe tener en cuenta que estos registros faltantes son diferentes a los registros en 0. Los registros en 0 son mediciones acertadas, mientras que, los registros faltantes son por fallas en los medidores, de acuerdo a lo comunicado por los stakeholders. Estos registros faltantes serán tratados en la sección de limpieza de datos.

3. **Consistencia:** Según lo indicado por Contugas, las unidades del conjunto de datos son grados Celsius para la temperatura, bar para la presión y metros cúbicos para el volumen. Esto asegura que las variables ya estén en unidades consistentes, sin necesidad de transformaciones adicionales, sin embargo, se revisarán los rangos entre los que se encuentran estas variables; para tal fin se verificaron tanto boxplots por variable y por cliente de tal forma que se identifiquen valores sobresalientes, asimismo se consideró como anomalía inicial (solo para efectos de este análisis) aquellos valores por fuera de los percentiles 1 y 99, analizando por año, cliente y variable:
 - Algunos clientes como el 03 y el 11 presentan valores considerablemente bajos en la variable presión (por debajo de 5 bar), lo cual no se considera razonable en condiciones operativas estándar.
 - En cuanto a la temperatura, clientes como el 07 y el 12 muestran valores cercanos a 0 o incluso negativos, los cuales reflejan posibles errores en la recolección de datos.
 - En volumen, clientes como el 04 y el 09 presentan un porcentaje significativo de registros en valores cercanos a 0, lo que podría ser un patrón de consumo particular o fallos en los medidores.
 - Se analizó el mínimo, máximo, media y desviación estándar por cliente. Clientes como el 06 en volumen y el 07 en temperatura presentan un rango de valores considerablemente más amplio que otros clientes, lo que podría sugerir comportamientos operativos únicos o problemas en el registro de datos.

Con base en lo anterior, se concluye que existen inconsistencias en algunos de los datos recolectados para ciertas variables y clientes que deben ser revisadas. Es fundamental verificar si los rangos están alineados con las

condiciones operativas reales, y determinar si estas inconsistencias son errores en los medidores o patrones válidos.

4. **Claridad:** La claridad de los datos está asegurada por la consistencia en sus unidades de medición y el etiquetado de las variables. Las unidades empleadas —grados Celsius, bar y metros cúbicos— cumplen con los estándares requeridos para el monitoreo del gas, eliminando la necesidad de realizar transformaciones adicionales. Además, los datos presentan un formato estructurado, con las fechas en formato datetime64 y las variables numéricas como float, facilitando tanto el análisis como las transformaciones requeridas para modelar series temporales. De igual forma, esto se puede corroborar a través de la implementación de técnicas de análisis estadísticas presentes en este informe y en el repositorio que contiene el pre-procesamiento de datos y análisis exploratorio. Posteriormente se abordarán los insights asociados al conjunto de datos en general y de manera particular por cliente.

En conclusión, si bien deben ser procesados, los datos son concordantes con el problema del negocio y de acuerdo a lo identificado en esta fase se procederá a realizar la limpieza y/o imputación pertinente para revisar con mayor profundidad las inconsistencias presentadas.

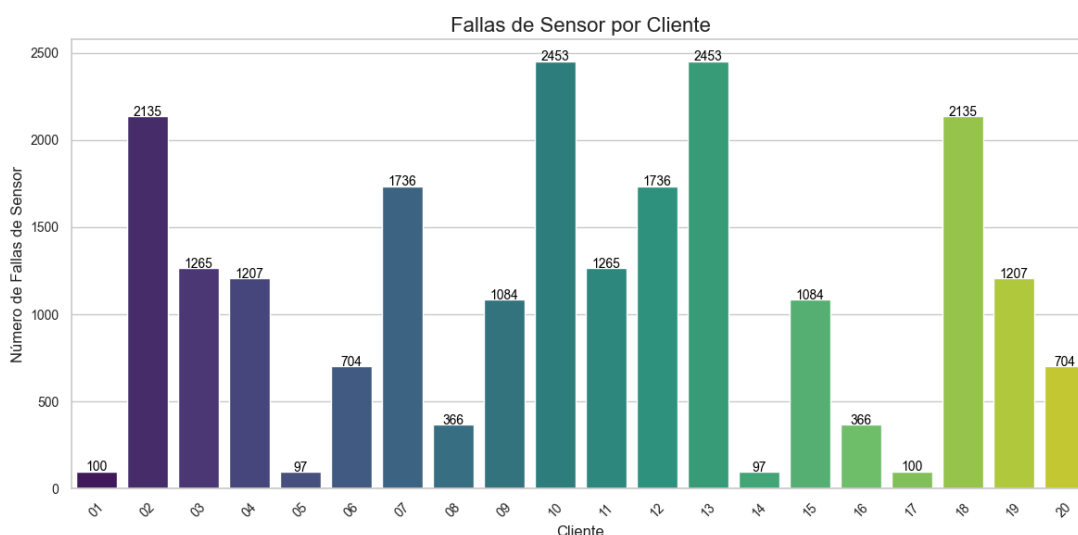
Técnicas de Limpieza de los Datos

1. Imputación de Valores Faltantes

Creamos una serie temporal completa para detectar valores faltantes. El objetivo de este paso es asegurarnos de que los datos de cada cliente tengan un rango temporal continuo desde el 2019-01-14 00:00:00 hasta el 2023-12-31 23:00:00 (registros cada hora). Esto nos permitirá identificar valores faltantes para cada cliente, interpolarlos y agregar una columna que identifique fallas en el sensor.

Al garantizar que todas las marcas de tiempo estén presentes, podemos identificar valores faltantes que, según lo explicado por los stakeholders, se debe a fallas en los sensores. Estas lagunas en los datos serán tratadas mediante interpolación, lo que permitirá preservar la continuidad de los datos sin introducir ruido adicional.

Para rellenar los valores faltantes, nos basamos en los valores adyacentes. La interpolación lineal es adecuada en este caso porque preserva la tendencia general de los datos sin introducir ruido. Una vez completado el proceso para cada cliente, unimos todos los resultados en un único DataFrame (`df_limpieza1`), organizando los datos por cliente y fecha y agregando una columna adicional `Sensor_Error` que identifica las fallas en los sensores.

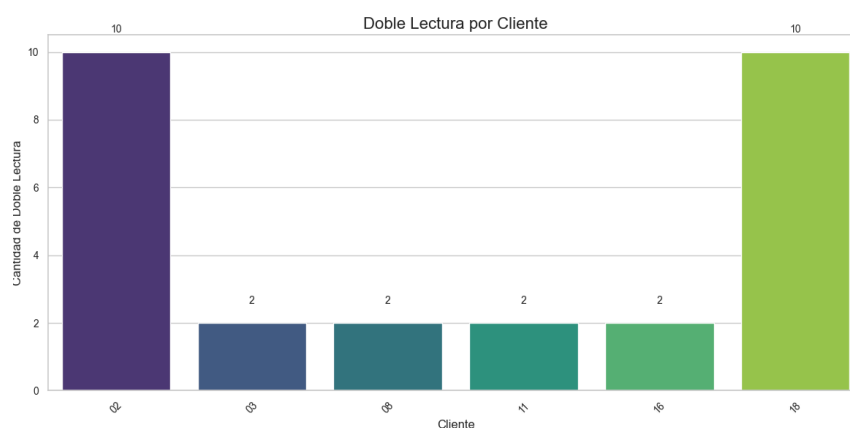


2. Eliminación de duplicados

La detección de dobles lecturas es clave para identificar posibles fallas en los sensores o errores de captura de datos, los cuales podrían influir en la calidad del monitoreo del gas.

Primero, verificamos duplicados exactos en cada cliente, definidos por coincidencias en las columnas Cliente y Fecha, y eliminamos uno de los registros duplicados, quedándonos con el primero. Posteriormente, detectamos y etiquetamos registros con valores duplicados pero diferentes en las variables de medición (Presión, Temperatura y Volumen). Estos casos fueron marcados con una nueva columna, Doble_Lectura, sin eliminar los registros, ya que representan información valiosa para identificar posibles fallas en sensores o errores de captura.

Esta columna servirá como una característica adicional en el entrenamiento del modelo, permitiendo que aprenda patrones asociados a fallas en los sensores. Además, su implementación práctica generará alertas cuando se detecten múltiples lecturas en un cliente, lo que facilitará la inspección técnica de los equipos de medición.



3. Verificación de Valores Nulos por Variable

La presencia de valores nulos en las variables de medición (Presión, Temperatura y Volumen) puede ser indicativa de otra falla en los sensores o problemas en la captura de datos, afectando directamente la calidad del monitoreo del gas y el desempeño de los modelos analíticos. Identificar y tratar estos valores es crucial para garantizar la continuidad y precisión de los datos utilizados en el proyecto.

Como parte de nuestro proceso de limpieza, se verificaron los valores nulos para cada cliente y variable de forma individual. Cada registro con al menos un valor nulo fue etiquetado con una nueva columna, Error_Lectura, la cual marca estos registros con un valor de 1 para facilitar su identificación. Posteriormente, los valores nulos fueron tratados mediante interpolación lineal, un método adecuado para series temporales que asegura la continuidad sin introducir ruido significativo.

Como resultado de este proceso, nuestra base de datos actual no contiene valores nulos en las variables de medición, garantizando que todos los registros estén completos.

Procesos de entendimiento de los datos

1. Generación de variables temporales

La extracción de características temporales a partir de la columna Fecha constituye un paso en el análisis de detección de anomalías. Este proceso permite transformar la información para que sea interpretable tanto en análisis exploratorio como en el entrenamiento del modelo.

Las variables temporales nos ayudarán a capturar relaciones entre el tiempo y las variables dependientes (Presión, Temperatura, Volumen). Por ejemplo, ciertos patrones de consumo de gas pueden ser diferentes entre días laborales y fines de semana. Además, al incluir información temporal, el modelo puede detectar anomalías relacionadas con horarios inusuales de consumo o variaciones atípicas en determinados días o meses.

	Fecha	Presion	Temperatura	Volumen	Cliente	Año	Sensor_Error	Double_Reading	Mes	Día	Hora	Día_Semana	Fin_de_Semana
0	2019-01-14 00:00:00	17.732563	28.209354	20.969751	01	2019	0	0	1	14	0	0	0
1	2019-01-14 01:00:00	17.747776	28.518614	17.845739	01	2019	0	0	1	14	1	0	0
2	2019-01-14 02:00:00	17.758916	28.230191	20.975914	01	2019	0	0	1	14	2	0	0
3	2019-01-14 03:00:00	17.727940	27.811509	20.592299	01	2019	0	0	1	14	3	0	0
4	2019-01-14 04:00:00	17.746484	27.795293	21.690626	01	2019	0	0	1	14	4	0	0
...
870249	2023-12-31 19:00:00	15.751139	27.460652	204.457549	20	2023	0	0	12	31	19	6	1
870250	2023-12-31 20:00:00	15.614858	27.010382	186.512096	20	2023	0	0	12	31	20	6	1
870251	2023-12-31 21:00:00	15.598944	26.709100	204.456461	20	2023	0	0	12	31	21	6	1
870252	2023-12-31 22:00:00	15.730040	27.266090	203.695596	20	2023	0	0	12	31	22	6	1
870253	2023-12-31 23:00:00	15.624457	27.481288	201.534548	20	2023	0	0	12	31	23	6	1

870254 rows × 13 columns

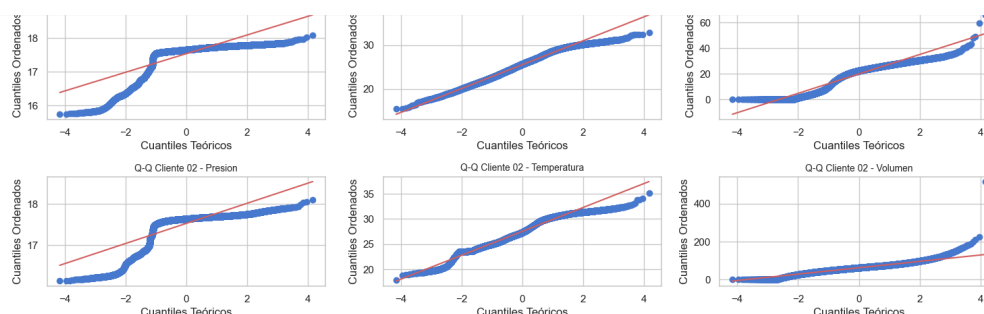
2. Escalado y estandarización de variables:

Según los gráficos Q-Q y los histogramas, las variables de presión y temperatura parecen seguir distribuciones relativamente normales o con ligeras desviaciones. Estas variables son independientes del uso directo y no presentan comportamientos drásticamente sesgados.

Por otro lado, Volumen tiene algunos clientes con distribución extremadamente sesgada con valores atípicos que reflejan el comportamiento de uso. Esto es esperable, ya que el volumen depende de patrones de consumo, lo que puede generar datos altamente no normales y sesgados hacia valores bajos.

Para el proceso de escalado tenemos que garantizar que todas las variables (Presión, Temperatura, Volumen) estén en la misma escala por cliente, permitiendo al modelo considerar sus efectos de manera proporcional. Es por esto que elegimos Estandarización (Z-score) como técnica de escalado debido a su capacidad para preservar valores atípicos para el caso de la variable volumen, lo que es esencial para la detección de anomalías, y en cuanto a las variables de Presión y Temperatura las ajustara correctamente sin perder sus patrones normales.

Aunque este proceso no es necesario en este punto, nos permite observar en detalle si nuestras variables siguen una distribución normal y poder determinar la mejor técnica para escalar nuestras variables cuando sea necesario introducir los datos a los modelo, asegurando que todas las variables contribuyan de manera equitativa, lo que evita que variables con magnitudes mayores (por ejemplo, Volumen) dominen el análisis.



3. Análisis del comportamiento de los ceros en Volumen

El análisis de registros con valores de volumen igual a cero nos permite comprender los patrones de consumo de gas por cliente y detectar otra posible anomalía. En este contexto, la presencia de ceros puede reflejar comportamientos normales asociados con horarios sin consumo, días no laborables, o horarios no operativos. Sin embargo, un comportamiento irregular o una frecuencia inusualmente alta de ceros o de consumos podría indicar problemas, como fallas en los sensores o anomalías en el suministro de gas.

Mes	1	2	3	4	5	6	7	8	9	10	11	12
Cliente												
01	2.714286	1.142857	4.545455	8.583333	6.666667	2.500000	5.384615	4.384615	1.200000	1.000000	1.000000	1.555556
02	1.000000	1.000000	1.058824	1.071429	1.000000	1.000000	1.000000	1.125000	1.000000	1.000000	1.000000	1.214286
03	1.166667	1.125000	1.000000	1.000000	1.000000	1.000000	1.000000	0.000000	1.000000	0.000000	0.000000	1.000000
04	17.737226	17.776978	17.353333	18.000000	18.080537	17.150685	17.765517	17.370629	17.537415	17.691781	17.903448	17.340000
05	4.718750	5.039370	5.148148	5.114504	5.753968	5.492308	5.143939	4.854839	5.043860	4.794643	5.009259	4.250000
06	3.369231	4.523810	5.049383	5.012658	1.303030	1.612903	1.586957	2.142857	1.984375	1.738462	1.818182	1.352941
07	0.000000	0.000000	11.311111	22.033613	21.958333	23.240000	23.193548	23.329032	23.273333	21.569231	1.000000	0.000000
08	1.100000	1.000000	1.187500	2.666667	1.457143	1.136364	1.458333	2.724138	4.265306	3.204082	1.615385	2.608696
09	21.643939	21.074627	21.758170	22.006803	21.310811	22.550336	22.098039	22.340136	22.333333	22.045752	21.978873	21.219858
10	21.237410	22.029851	21.543046	21.885714	22.607843	21.489796	21.267123	21.527027	22.162963	22.260274	21.462069	21.350993
11	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000
12	1.000000	1.000000	20.276923	23.783333	23.253425	23.773333	23.800000	23.819355	23.590604	22.309735	1.000000	0.000000
13	20.608696	21.942857	20.935484	21.546667	21.059211	21.130137	21.717241	21.413043	21.492754	21.474026	21.091549	21.046053
14	3.281250	2.970874	3.971963	3.904762	4.233645	4.592233	4.081633	3.505263	3.659341	2.967391	3.126437	2.744444
15	21.485714	21.124088	21.934211	21.598639	21.701299	21.048611	21.859060	21.731544	21.391608	22.331126	21.156028	21.328767
16	1.000000	1.000000	0.000000	2.333333	2.250000	1.000000	1.000000	3.521739	5.266667	2.916667	1.000000	2.111111
17	0.000000	0.000000	1.000000	1.000000	0.000000	1.000000	9.000000	8.500000	0.000000	0.000000	0.000000	0.000000
18	1.055556	1.083333	1.062500	1.166667	1.050000	1.090909	1.000000	1.000000	1.111111	1.066667	1.111111	1.047619
19	19.706767	18.926471	19.163265	19.972414	19.851351	19.602740	18.797297	19.097222	19.165517	20.120805	19.761905	19.720779
20	1.390244	3.019608	1.380000	2.929825	1.266667	1.294118	1.966667	2.236842	2.159420	1.966102	1.791667	1.816327

Clientes como el 11 y 16 prácticamente no tienen registros en cero. Esto indica un consumo constante o sensores que nunca registran valores en cero. Para este tipo de clientes, cualquier aparición de ceros sería una señal de anomalía o un posible error en la medición.

El cliente 9 se destaca por tener un porcentaje de ceros que alcanza el 87% de sus registros. Este comportamiento parece normal en su caso, ya que presenta un promedio de consumo de 2 horas al día. Sin embargo, cualquier aumento significativo en los ceros (como un incremento a 4 horas o más en ciertos días) podría considerarse como una anomalía.

El análisis por días de la semana muestra que ciertos clientes tienen tendencias regulares de ceros en días específicos, el cliente 1 tiene picos de ceros en algunos meses y días (por ejemplo, los miércoles y los meses de marzo a mayo), y nunca ha dejado de consumir jueves y viernes.

Los promedios calculados para meses, años, y días de la semana permiten establecer un umbral de comportamiento esperado para cada cliente. Mientras algunos presentan comportamientos consistentes, otros tienen fluctuaciones regulares que podrían parecer atípicas sin este contexto. Estos hallazgos son esenciales para diseñar un sistema de detección de anomalías que minimice falsos positivos y mejore la precisión del monitoreo.

4. Correlacion entre variables

La correlación entre las variables de presión, temperatura y volumen fue analizada de forma individual para cada cliente, con el objetivo de identificar relaciones significativas entre estas.

El análisis no reveló correlaciones positivas altas (mayores al 80%) que pudieran representar un problema para la inclusión de las variables en el modelo. Sin embargo, las correlaciones negativas altas observadas entre presión y

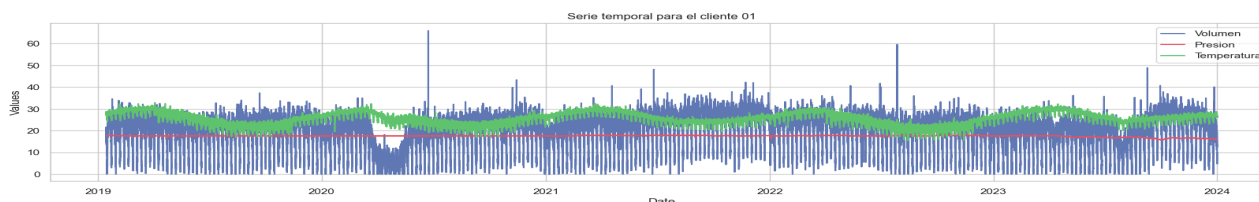
volumen en dos clientes destacan la necesidad de analizar estas relaciones más a fondo para descartar problemas en los sensores, aunque estas correlaciones no necesariamente representan un problema para el modelo si se entienden y justifican en el contexto del sistema medido.

5. Estadísticas Descriptivas por cliente

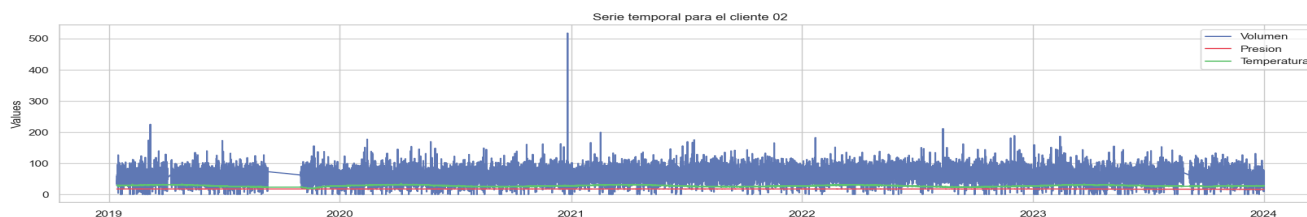
El análisis de métricas básicas por cliente permite entender en detalle las características de los datos y cómo se comportan las variables Presión, Temperatura y Volumen para cada cliente. Este conocimiento nos permite identificar comportamientos normales y posibles desviaciones que puedan indicar anomalías en los datos.

Cliente	Media_Presion	Mediana_Presion	Min_Presion	Max_Presion	Std_Presion	Media_Temperatura	Mediana_Temperatura	Min_Temperatura	Max_Temperatura	Std_Temperatura	Media_Volumen	Mediana_Volumen	Min_Volumen	Max_Volumen	Std_Volumen
1	17.53625213	17.65165391	15.7423367	18.07427385	0.3581138228	25.57531447	25.70483523	15.40180299	32.86911236	2.753574965	19.98746154	22.47499162	0	65.93664367	7.941182018
2	17.52698551	17.6389632	16.12901534	18.10640165	0.317948669	27.55745237	27.36175508	17.88405905	35.20834588	2.411789917	62.01965233	62.41729122	0	517.5648676	17.15896678
3	3.568596265	3.568166502	3.057170922	4.037030329	0.08055458042	26.37374785	26.36959043	14.93303167	34.0923348	2.635774304	117.65269	125.7107764	0	356.7240076	39.74806952
4	17.64248104	17.77983361	15.21499578	20.11293101	0.3876830284	23.25762481	23.6991558	12.71181264	36.89704413	3.023691605	17.46274105	0	0	363.0097759	61.05530001
5	17.48960099	17.61476238	14.48989628	19.0093516	0.3701592997	23.66453051	23.49573648	10.50063757	41.76223215	5.050807379	7.81885123	7.11855607	0	89.24505063	6.409062311
6	17.57060025	17.7074914	13.8100258	19.72086983	0.3955908457	26.37734961	26.34598658	14.15736626	34.18633759	2.721491817	152.8864087	197.8061898	0	366.6563824	85.33262725
7	17.4963032	17.49926935	14.41826092	20.30785175	0.4469026988	23.47364352	23.81295303	0.4334355568	39.9822393	4.949659346	26.47758337	0	0	175.7638578	33.68950794
8	16.70524172	16.56740921	14.99649003	18.95095539	0.5976634836	26.81886683	26.89673523	14.83870285	37.36895856	2.971478273	178.7658053	221.2128445	0	522.7808913	90.76039264
9	17.46679857	17.49667904	16.48647744	18.43046854	0.1568153645	22.065828	22.24626975	11.82010432	31.78396368	2.92792477	12.55790487	0	0	366.01612	49.16531134
10	17.47354564	17.49307259	16.46986398	18.57307889	0.1334072615	23.43801243	23.47012272	14.29103213	35.06972131	2.652476579	12.57862378	0	0	247.0729895	45.48629268
11	3.545151658	3.535421804	2.934872944	3.954039051	0.08242906812	26.33604697	26.22621627	13.73102519	34.35527609	2.71318852	131.7079678	136.2644622	0	298.2595728	34.22126292
12	17.83661354	17.83919038	13.74092216	20.23211047	0.418673319	26.51916288	26.76140178	-5.257899119	50.01985255	5.051278082	32.71658466	0	0	284.4750866	41.69664729
13	17.4936315	17.51000224	16.50507704	18.47104679	0.1177161026	21.53435767	21.49253935	12.24051121	29.75336441	2.861443496	9.909209743	0	0	253.8679162	40.38810436
14	17.51733643	17.63306811	13.61687666	20.02839495	0.3688608445	26.66132599	26.54237859	5.67693284	45.37817643	4.96641776	7.301760533	6.748793379	0	36.79315006	5.25644634
15	17.49543649	17.51432544	16.51449568	18.49323363	0.1388790194	24.0172316	24.10875841	14.71287364	35.66134854	3.117431027	12.46292809	0	0	398.0420267	47.40982952
16	16.70529469	16.61962095	14.73452103	19.44077978	0.4918821844	27.48175338	27.50048629	19.14610127	32.65845563	2.449027805	178.1860625	210.8837973	0	409.8722119	74.59863557
17	17.53389527	17.6535681	15.26570349	18.44537774	0.3797899558	25.84854279	25.76024922	18.44509636	31.9175602	2.361762178	20.56900489	21.75539679	0	48.50483292	5.562752539
18	17.55530945	17.64606968	16.2151046	17.82135868	0.2494967407	27.5577884	27.42364673	16.38078009	33.78936219	2.512794206	61.48135427	61.5697559	0	577.4134248	18.09880785
19	17.6694781	17.786668	16.24748361	18.25861885	0.3223937848	23.23930589	23.76087224	16.18198213	30.64224273	3.024355407	15.74127058	0	0	378.2678031	59.20646127
20	17.58113704	17.68932866	14.84973693	18.57820294	0.4069268782	25.40269963	25.36955456	14.76708132	33.80800577	2.683028454	160.4448343	201.7133671	0	315.8841531	82.91591749

Cliente 01: Muestra un comportamiento estable en términos de presión y temperatura, con promedios de 17.53 y 25.57, respectivamente. Su desviación estándar en ambas variables es baja, indicando un sistema operativo controlado. Sin embargo, su promedio de volumen (19.98) y desviación estándar (7.94) sugieren cierta variabilidad en el consumo, aunque no significativa. Este cliente tiene un porcentaje de ceros bajo (1.41%), lo que refuerza la estabilidad en su operación. Es poco probable que presente anomalías, pero cualquier aumento drástico en ceros podría ser un indicador de problemas.

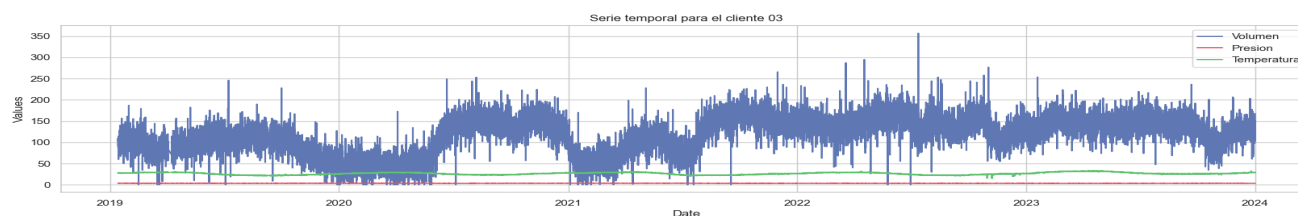


Cliente 02: Tiene un comportamiento regular en presión (promedio de 17.52) y temperatura (27.55), con bajas desviaciones estándar. Su consumo promedio de volumen es 62.01, siendo moderado en comparación con otros clientes, pero su desviación estándar es relativamente alta (17.15), lo que indica fluctuaciones en el consumo. Con solo un 0.32% de ceros, este cliente tiene patrones operativos estables y controlados con uno que otro outlier.

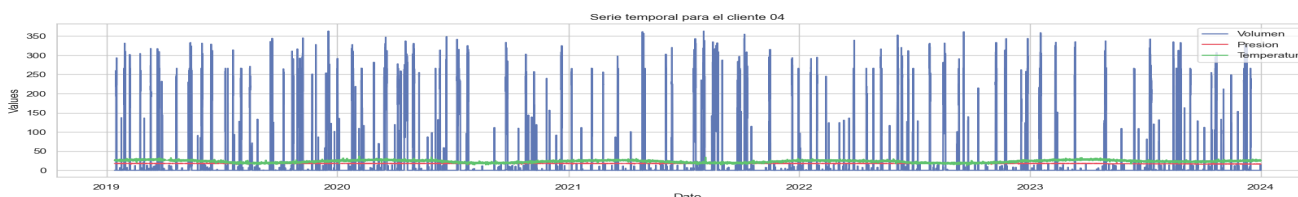


Cliente 03: Presenta un comportamiento único con la presión más baja entre todos los clientes (promedio de 3.56) y una desviación estándar muy baja (0.08), indicando un sistema de baja presión altamente estable. Sin embargo, su consumo de volumen promedio es relativamente alto (117.65) con una desviación significativa (39.74), que oscila entre los meses

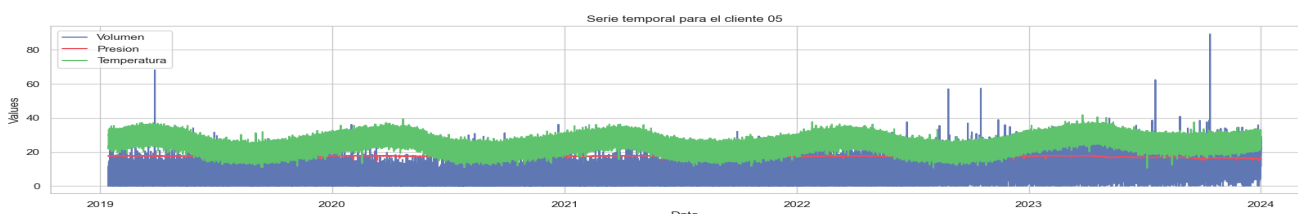
del año como se ve en la gráfica.



Cliente 04: Este cliente presenta alto porcentaje de ceros (70.50%). Su promedio de consumo descartando los periodos de inactividad es de 59.19 y con una desviación sd alta de 100.82, como se representa en la gráfica, el cliente presenta consumo intermitente pero alto.



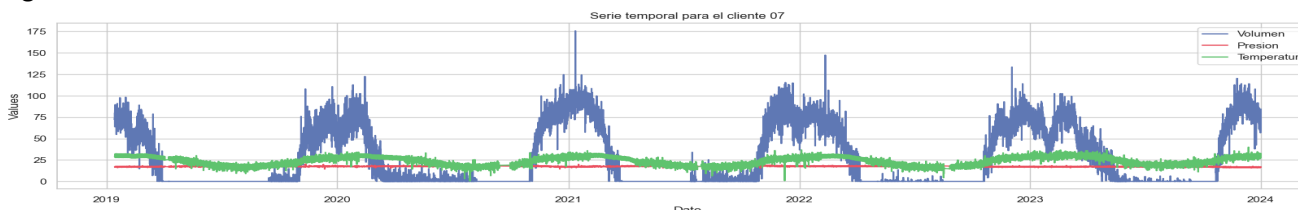
Cliente 05: Posee un promedio de volumen bajo (7.81) con una desviación estándar de 6.40. Su presión y temperatura son moderadas (17.48 y 23.66), con valores extremos notables en temperatura que llegan hasta 41.76, lo que sugiere un sistema que puede operar en condiciones extremas.



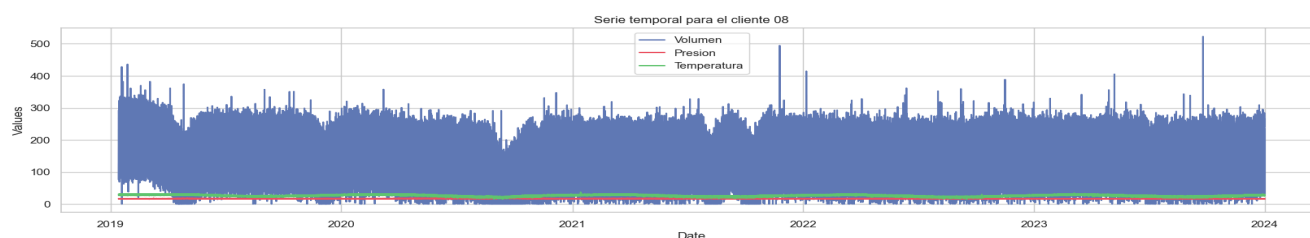
Cliente 06: Uno de los clientes con mayor consumo (152.88 en promedio), pero también con alta variabilidad. La presión (promedio de 17.57) y temperatura (26.37) son estables, aunque también muestran cierta variabilidad. Su porcentaje de ceros (4.76%) no es elevado.



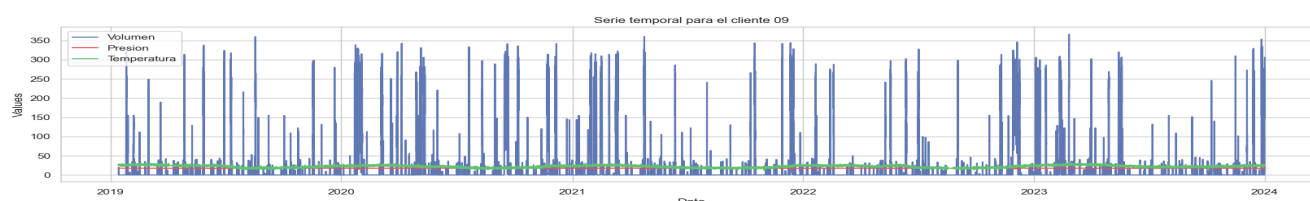
Cliente 07: Tiene un comportamiento distintivo ya que parece operar solo por la temporada de finales y principios de año y con un promedio de volumen bajo (56.96) cuando está en actividad. Su presión (17.49) y temperatura (23.47) son regulares.



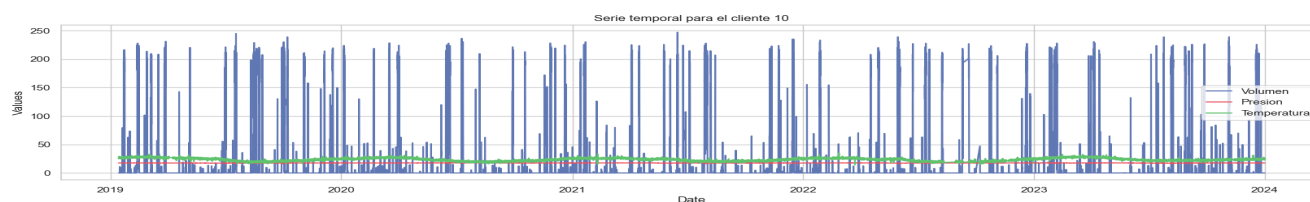
Cliente 08: Otro cliente de alto consumo con un promedio de volumen de 178.76, pero también con alta variabilidad (desviación estándar de 90.76). Su presión promedio (16.70) y temperatura (26.81) son estables. Con un porcentaje de ceros bajo (1.93%), este cliente es altamente operativo, y cualquier incremento en ceros podría ser una señal de alerta.



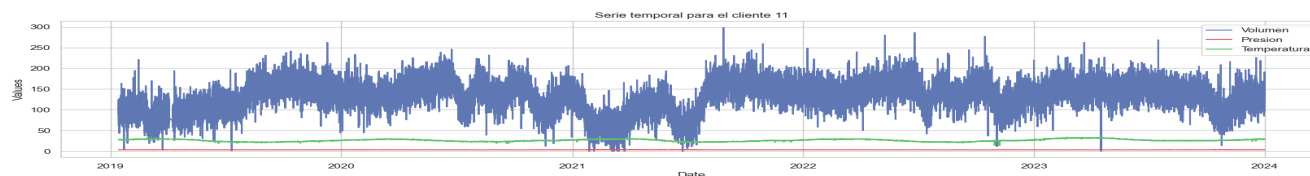
Cliente 09: Presenta un porcentaje de ceros extremadamente alto (87.62%), lo que refleja un comportamiento normal para este cliente. Su promedio de volumen (101.45 en actividad) aunque con una desviación estándar (102.52) lo que representa un comportamiento en picos como se puede observar en el gráfico.



Cliente 10: Patron muy similar al cliente 09 con un alto porcentaje de ceros (86.66%) y un promedio de volumen (94.30). Su presión y temperatura son consistentes y estables, con poca variabilidad.



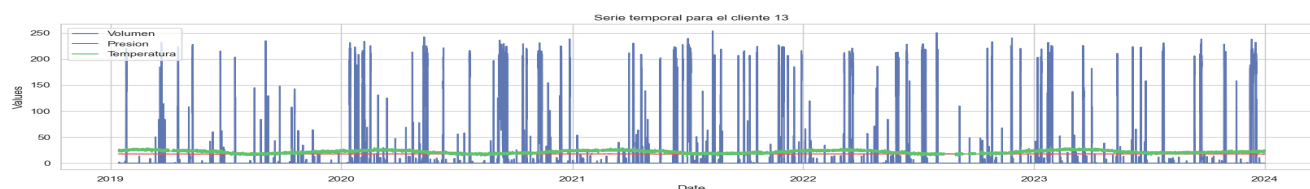
Cliente 11: Posee una presión promedio baja (3.54), con una desviación estándar mínima (0.08). Su volumen promedio (131.70) es moderado, con 0.01% de ceros.



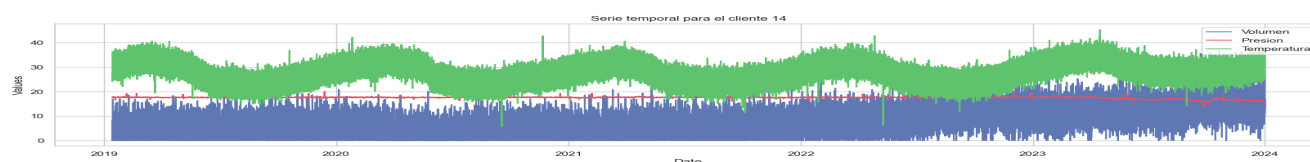
Cliente 12: Con un porcentaje de ceros (56.43%) indica periodos de inactividad que son normales, como se puede apreciar en la gráfica del comportamiento. Su promedio de volumen en actividad es de 75.09 con sd de 28.42.



Cliente 13: Este cliente tiene porcentaje de ceros alto (85.60%), sin embargo vemos picos de uso que llegan hasta un máximo de 253.86. Temperatura y presión estables.



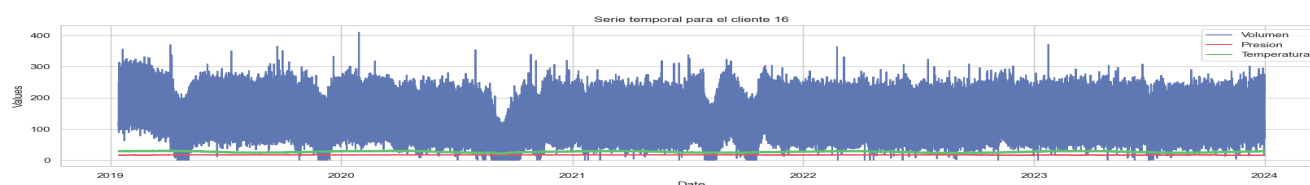
Cliente 14: Tiene promedios bajos en volumen (7.30) y cierta variabilidad (5.25), mientras su temperatura (26.66) tiene una variabilidad de 4.96, una de las más altas en el grupo. Este cliente tiene un porcentaje de ceros moderado (9.74%).



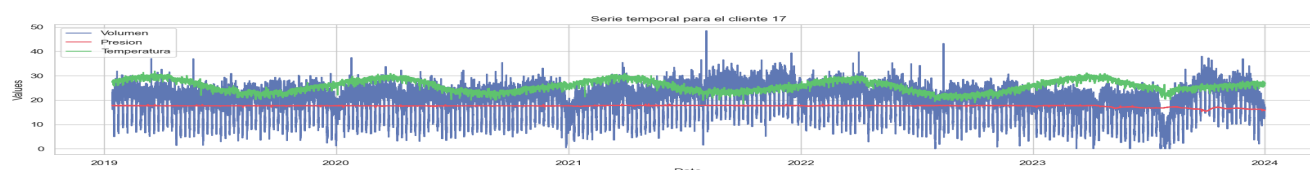
Cliente 15: Similar a los clientes 09, 10 y 13, alto porcentaje de ceros alto (86.88%), con consumos elevados cuando está en actividad con un promedio de 95.05.



Cliente 16: Tiene un alto promedio de volumen (178.18) y una desviación significativa (74.59), lo que indica operaciones intensivas con alta variabilidad. Su presión (16.70) y temperatura (27.48) son regulares.



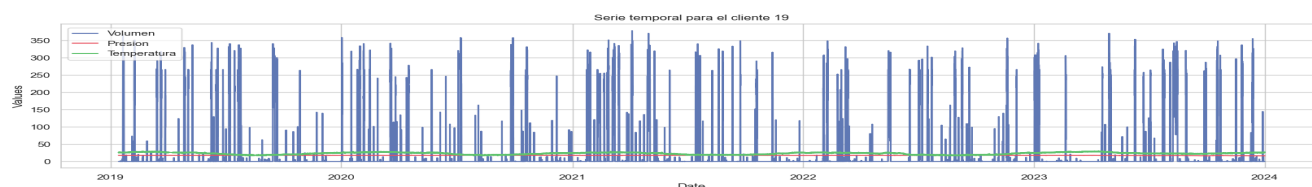
Cliente 17: Posee un comportamiento estable con consumos moderados en volumen (20.56) y una baja desviación estándar (5.56). Su porcentaje de ceros (0.14%) es mínimo, indicando que tiene un sistema altamente operativo.



Cliente 18: Con consumos moderados de volumen (61.48) y baja variabilidad. Su presión (17.55) y temperatura (27.55) son regulares, posee una larga pausa en 2019 pero que no se ha vuelto a repetir. Tiene unos picos de consumo que podrían ser algún tipo de anomalía. Es el cliente con el pico de consumo más alto de todo el cohort.



Cliente 19: Otro cliente con operación en pocas horas al día pero con alto consumo promedio en actividad (71.69). Tiene una presión y temperatura estable.



Cliente 20: Presenta un promedio alto de volumen (160.44) y una desviación significativa (82.91). Su presión y temperatura son estables, pero alta variabilidad en volumen.

