

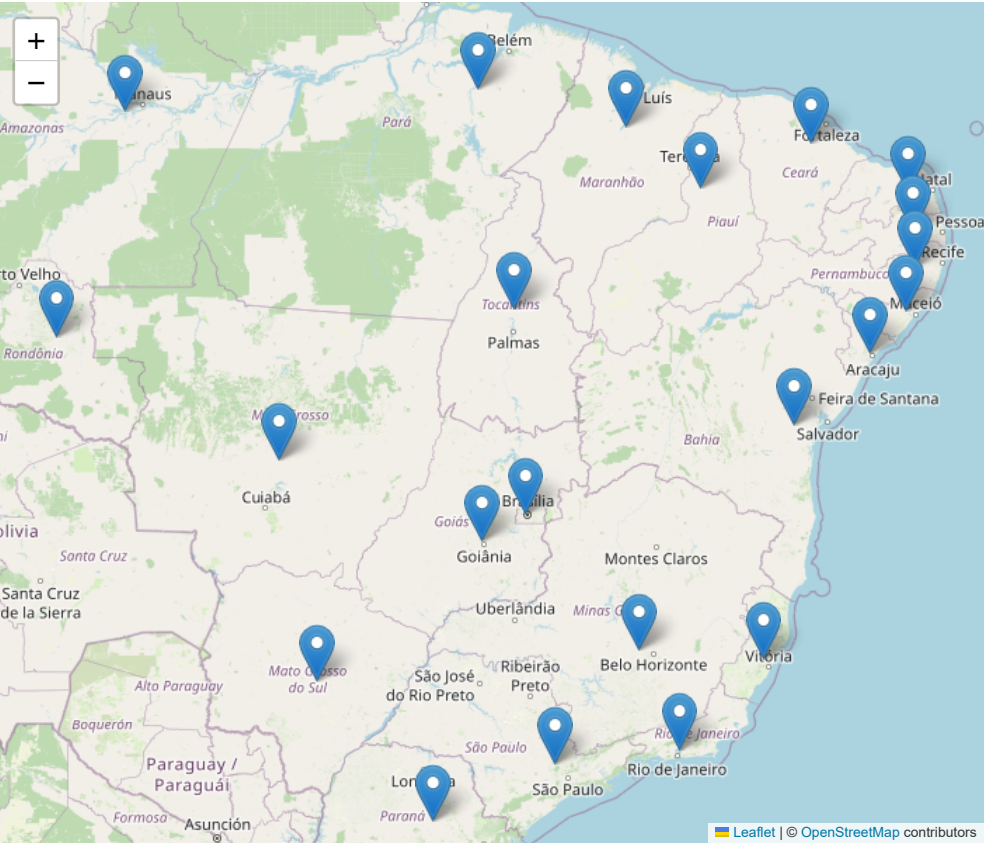
Resultados Analisis de Datos BD Olist E-commerce

Conjunto de datos públicos de comercio electrónico brasileño de pedidos realizados en Olist Store. El conjunto de datos tiene información de 100 000 pedidos de 2016 a 2018 realizados en varios mercados de Brasil. Sus características permiten ver un pedido desde múltiples dimensiones: desde el estado del pedido, el precio, el pago y el rendimiento del flete hasta la ubicación del cliente, los atributos del producto y, finalmente, las reseñas escritas por los clientes. También publicamos un conjunto de datos de geolocalización que relaciona los códigos postales brasileños con las coordenadas de latitud y longitud.

Realizare la identificación de las áreas geográficas que presentan mayores demoras en las entregas de pedidos, a través del análisis espacial de los datos de geolocalización de clientes y vendedores, junto con la información sobre tiempos de entrega y costos de envío

Estados Ubicación Clientes

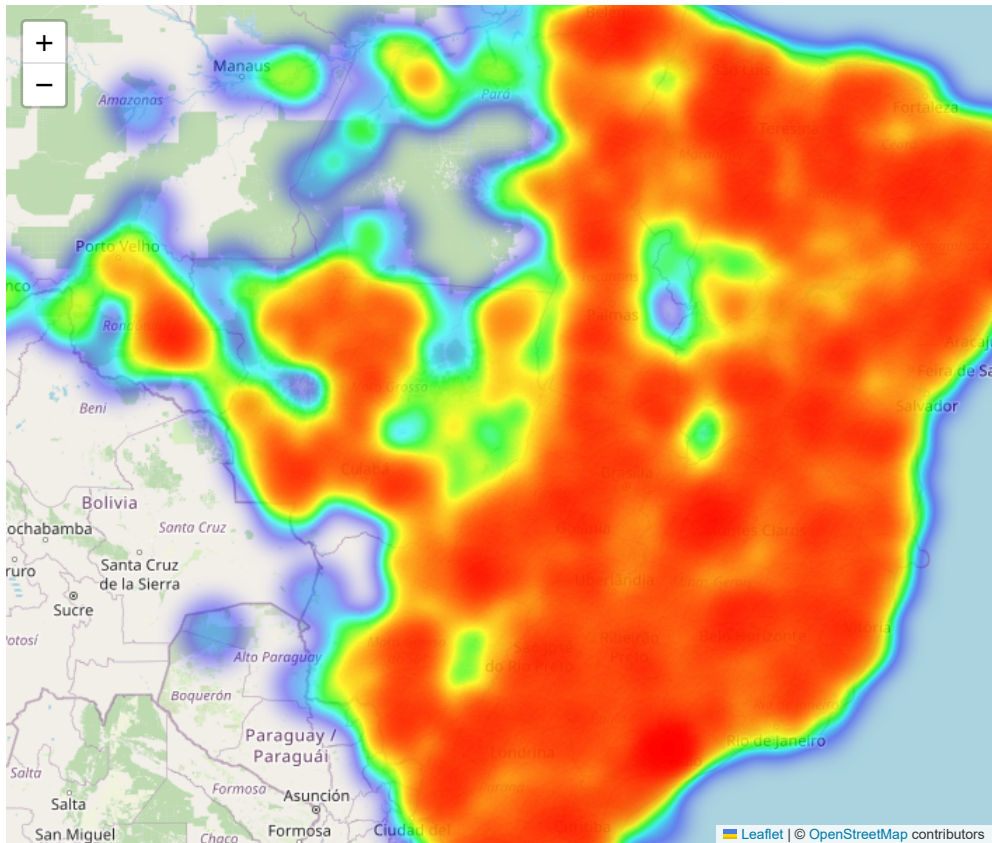
Visualización de estados donde existen clientes.



El mapa se contruye con una imagen del OpenStreetMap, representa la ubicación geográfica de los estados en Brasil donde existen clientes de la plataforma e-commerce. Los marcadores indican la posición aproximada de cada estado. La representación refleja mayor concentración de clientes en las regiones sudeste y noreste del país, con estados como São Paulo, Rio de Janeiro, y Minas Gerais, entre otros, destacándose como los principales mercados.

Mapa de calor - Ubicación Clientes

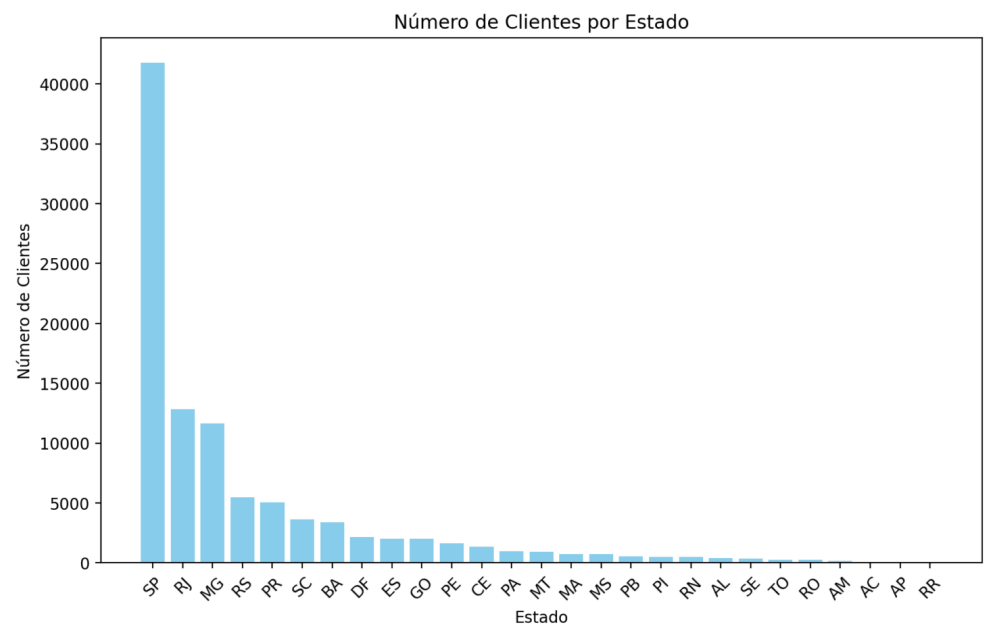
Ubicación de clientes en Brasil. Los colores varían desde tonos de azul, verde y amarillo, indicando las áreas con diferentes concentraciones de clientes.



El mapa se contruye con una imagen del OpenStreetMap, la cual incluyendo detalles geográficos como ciudades, carreteras y áreas protegidas. El color Verdes/Amarillo: Indican una alta concentración de clientes en esa ubicación específica. Estos son los puntos de mayor densidad, donde se observa la mayor actividad de clientes. El color Azul: Indican una menor concentración de clientes, pero aún así significativa, estas zonas tienen menos actividad en comparación con las áreas más cálidas.

Número de Clientes por Estado

Distribuyen los clientes de la plataforma e-commerce por estado, Sao Paulo como el principal mercado.



El gráfico representa la distribución del número de clientes por estado en la plataforma de e-commerce. El estado de Sao Paulo (SP) destaca como el principal mercado, con la mayor cantidad de clientes, seguido por Río de Janeiro (RJ) y Minas Gerais (MG). Los estados con menor población de clientes incluyen

Roraima (RR) y Amapá (AP), lo que sugiere una concentración de clientes en las áreas más pobladas y urbanizadas de Brasil.

Ciudades con Más de 500 Pedidos

Esta sección muestra las ciudades con más de 500 pedidos realizados en la plataforma de e-commerce.

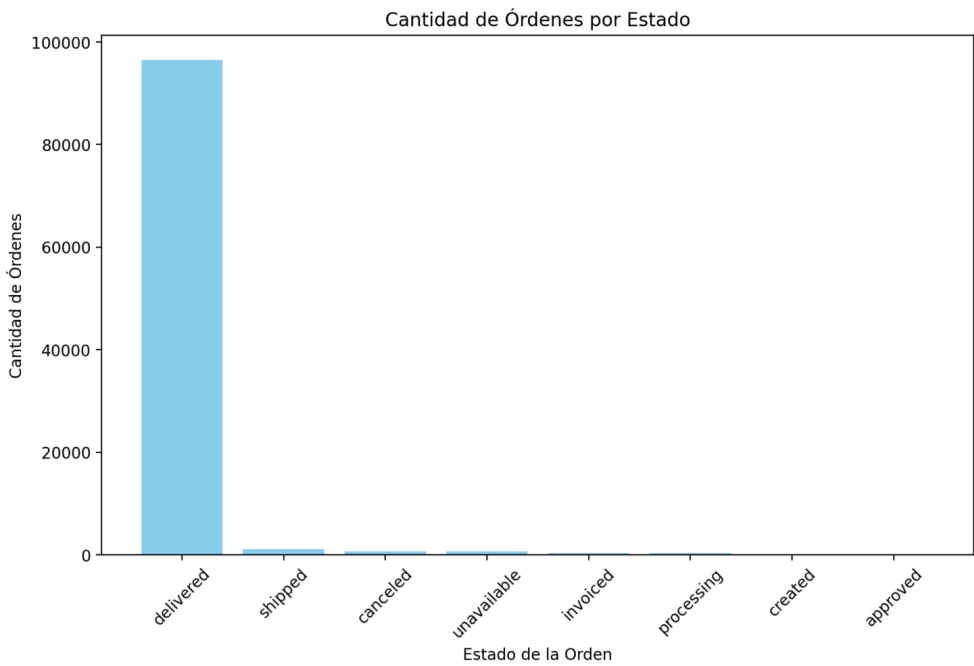
Top 10 Ciudades con Más de 500 Pedidos

A continuación se presenta una tabla que destaca las ciudades con más de 500 pedidos en la plataforma. Estas ciudades son los principales mercados que concentran un gran volumen de transacciones.

	ciudad	cantidadpedidos
0	sao paulo	15,540
1	rio de janeiro	6,882
2	belo horizonte	2,773
3	brasilia	2,131
4	curitiba	1,521
5	campinas	1,444
6	porto alegre	1,379
7	salvador	1,245
8	guarulhos	1,189
9	sao bernardo do campo	938

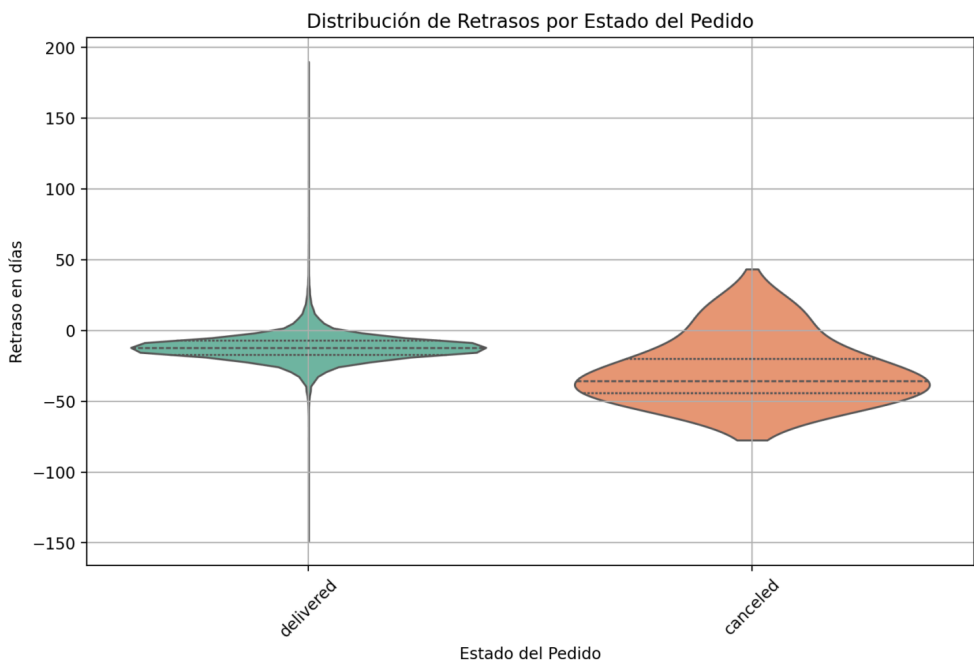
La tabla muestra las 10 principales ciudades que han generado más de 500 pedidos en la plataforma de e-commerce. Estas ciudades son puntos clave en la operación del comercio electrónico, concentrando la mayor actividad de compras. São Paulo es típicamente la ciudad con más actividad, seguida de otras grandes urbes como Rio de Janeiro y Belo Horizonte.

Estado de las ordenes y retrasos en la entrega



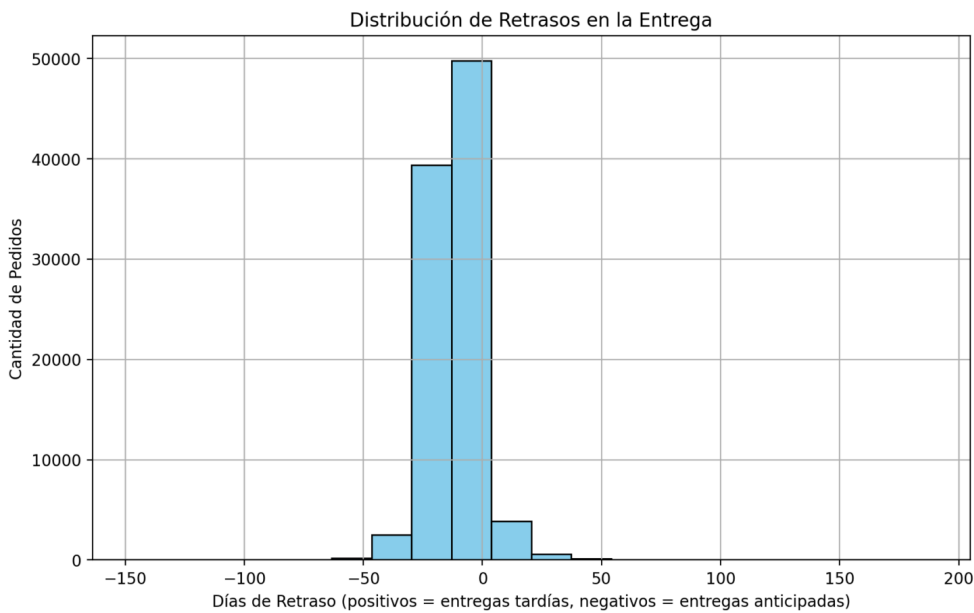
El gráfico muestra la distribución de órdenes según su estado en la plataforma de e-commerce. Se observa que la gran mayoría de las órdenes (alrededor de 100,000) han sido entregadas, representando el estado predominante. Los demás estados, como "shipped", "canceled", y otros, presentan una cantidad de órdenes significativamente menor, lo que indica que la mayoría de las transacciones alcanzan su finalización exitosa con la entrega del pedido..

Estado del Pedido y Retraso en la Entrega



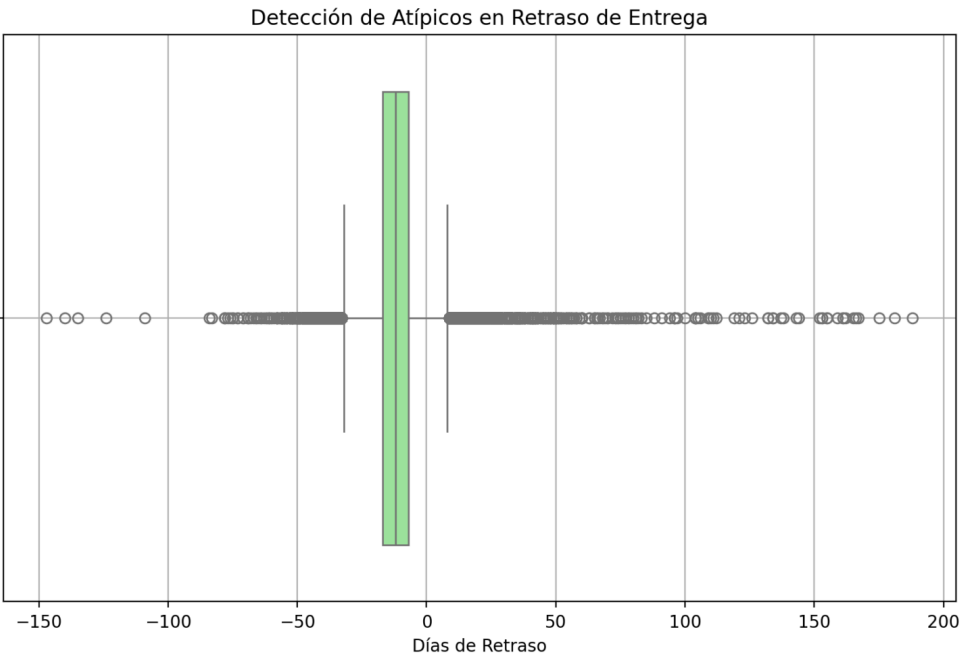
La gráfica muestra la distribución de los retrasos en la entrega según el estado del pedido. Se observa que, en general, los pedidos entregados tienen una distribución tanto de entregas anticipadas (valores negativos de retraso) como de pequeños retrasos (valores positivos de retraso). Por otro lado, los pedidos cancelados tienden a ser cancelados antes de la fecha estimada de entrega.

Distribución de Retrasos en la Entrega



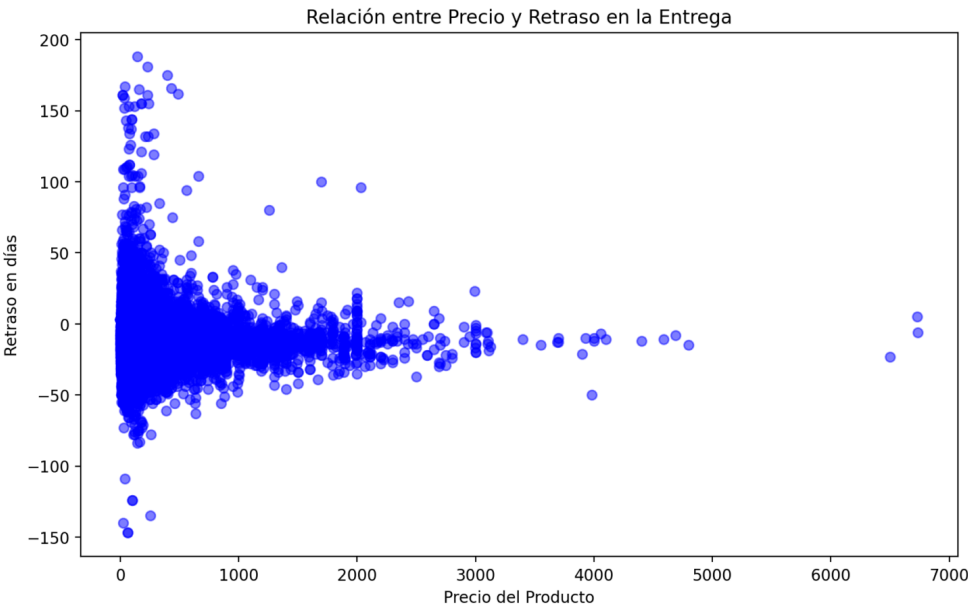
La mayoría de las entregas se realizan antes de la fecha estimada, con un pico alrededor de los 10 a 30 días antes de la entrega esperada..

Detección de Atípicos en Retraso de Entrega



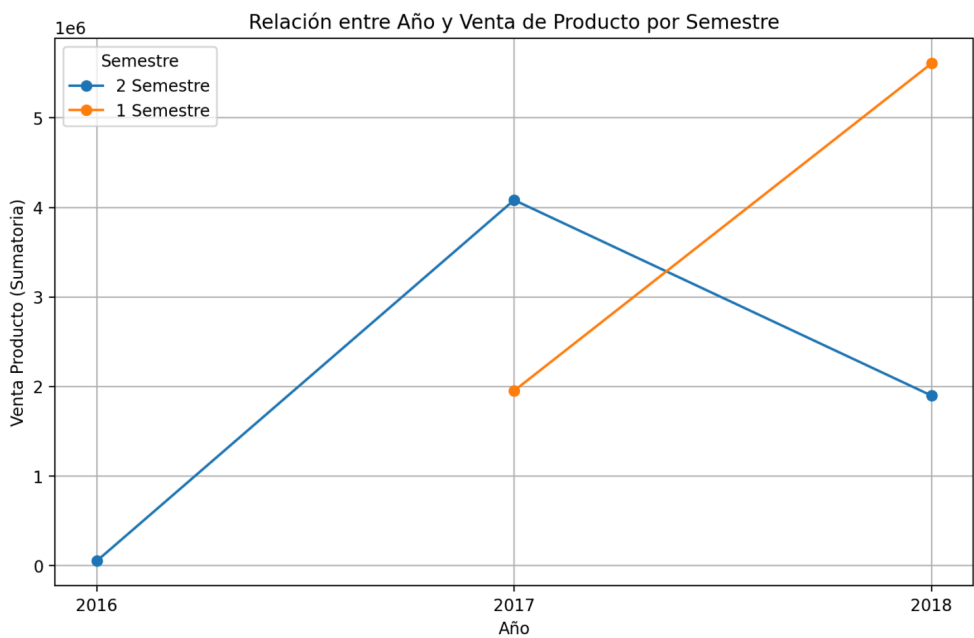
La gráfica evidencia que la mayoría de las entregas se realizan en torno a la fecha estimada, pero existen casos notables de entregas con retrasos o anticipaciones considerables, identificados como outliers en la distribución.

Relación entre Precio y Retraso en la Entrega



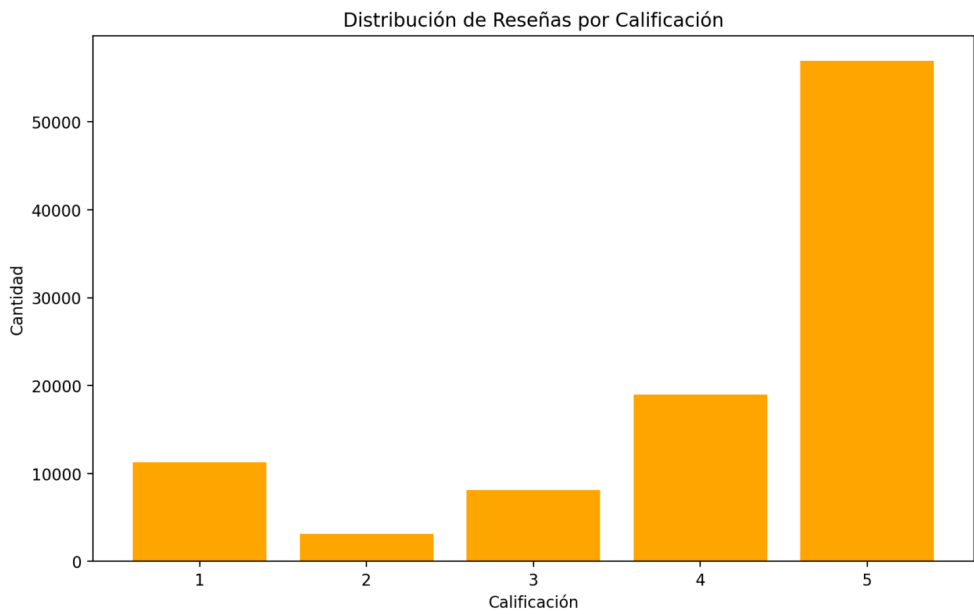
Este gráfico de dispersión muestra la relación entre el precio del producto y el retraso en la entrega. Se observa que la mayoría de los puntos se concentran en productos de bajo precio, generalmente por debajo de 1000 unidades monetarias. Para estos productos, los retrasos en la entrega varían considerablemente, abarcando desde entregas anticipadas hasta entregas tardías. A medida que el precio del producto aumenta, parece que la variabilidad en los retrasos disminuye, sugiriendo que los productos más caros tienden a cumplir mejor con las fechas de entrega estimadas.

Relación entre Año y Venta de Producto por Semestre



- a) Año 2016: Durante este año, se observa un crecimiento inicial, pero limitado, en las ventas del segundo semestre, mientras que no se registran ventas significativas en el primer semestre.
- b) Año 2017: Este año destaca por un fuerte incremento en las ventas durante el segundo semestre, alcanzando el pico más alto de todo el período analizado. En contraste, el primer semestre de 2017 muestra ventas significativamente más bajas.
- c) Año 2018: En 2018, se aprecia una inversión en el comportamiento observado en 2017. El primer semestre muestra un crecimiento notable, superando las ventas del segundo semestre del año anterior, mientras que el segundo semestre de 2018 muestra una caída considerable, llegando a niveles de ventas por debajo de los registrados en 2016.

Análisis de Reseñas y Puntuaciones



Visualización de Información de Productos con Caras Personalizadas



Esta visualización utiliza caras personalizadas para representar las características de los productos en diferentes categorías. El tamaño de los ojos representa el número de productos en esa categoría, y la forma de la boca refleja el peso promedio de los productos.

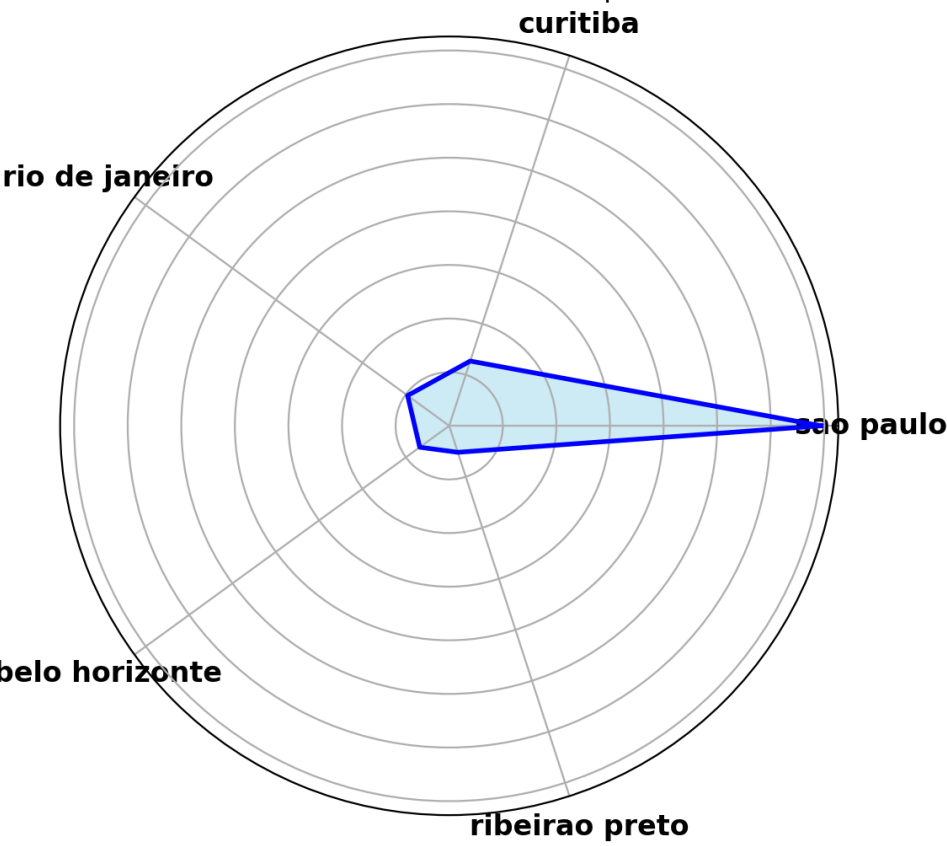
Información sobre los Vendedores

Distribución de vendedores en las principales ciudades donde opera la plataforma de e-commerce.

Columnas cargadas:

```
[
  0 : "ciudad"
  1 : "cnt_vendedores"
]
```

Cantidad de Vendedores en las Principales Ciudades

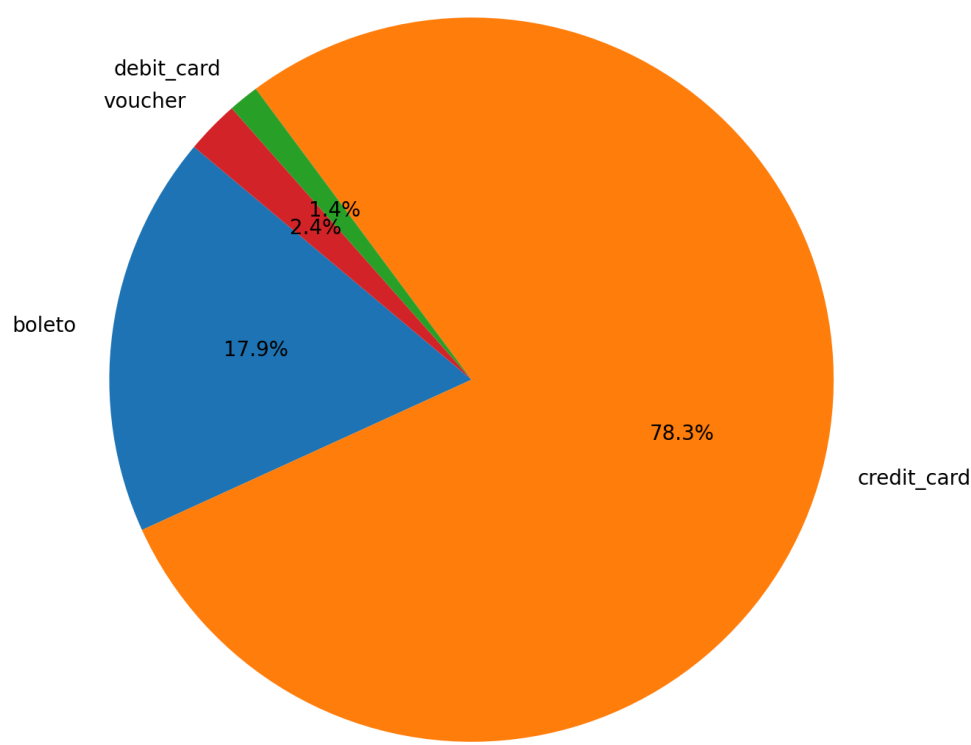


a) Sao Paulo: Representada en la esquina derecha del gráfico, es la ciudad con la mayor cantidad de vendedores, lo que se refleja en la longitud del eje correspondiente. b) Rio de Janeiro: Ubicada en la parte superior izquierda del gráfico, es la segunda ciudad con más vendedores. c) Belo Horizonte: También a la izquierda, muestra un número menor de vendedores en comparación con Sao Paulo y Rio de Janeiro. d) Curitiba: En la parte superior, tiene una cantidad de vendedores similar a la de Belo Horizonte. e) Ribeirao Preto: En la parte inferior del gráfico, muestra la menor cantidad de vendedores entre las cinco ciudades..

Métodos de Pago

El gráfico muestra la distribución de los métodos de pago utilizados por los clientes

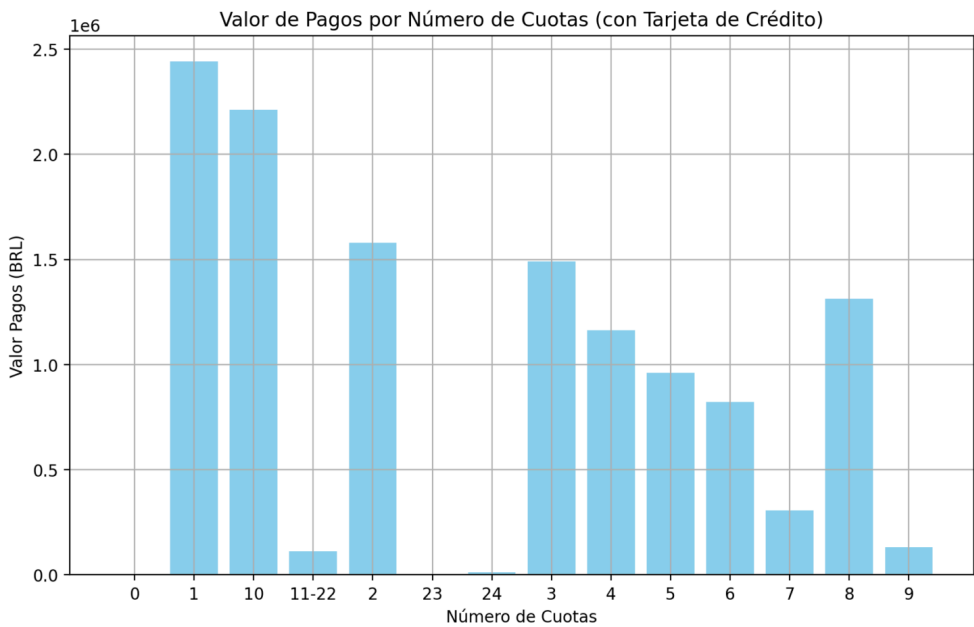
Métodos de Pago



La tarjeta de credito representa la mayor parte de los pagos, con un 78.3% del total. Boleto bancario: Es el segundo método más común, con un 17.9% de los pagos. Tarjeta de débito: Representa el 2.4% de los pagos, siendo un método menos frecuente. Voucher: Con 1.4% del total, este es el método de pago menos utilizado entre los presentados.

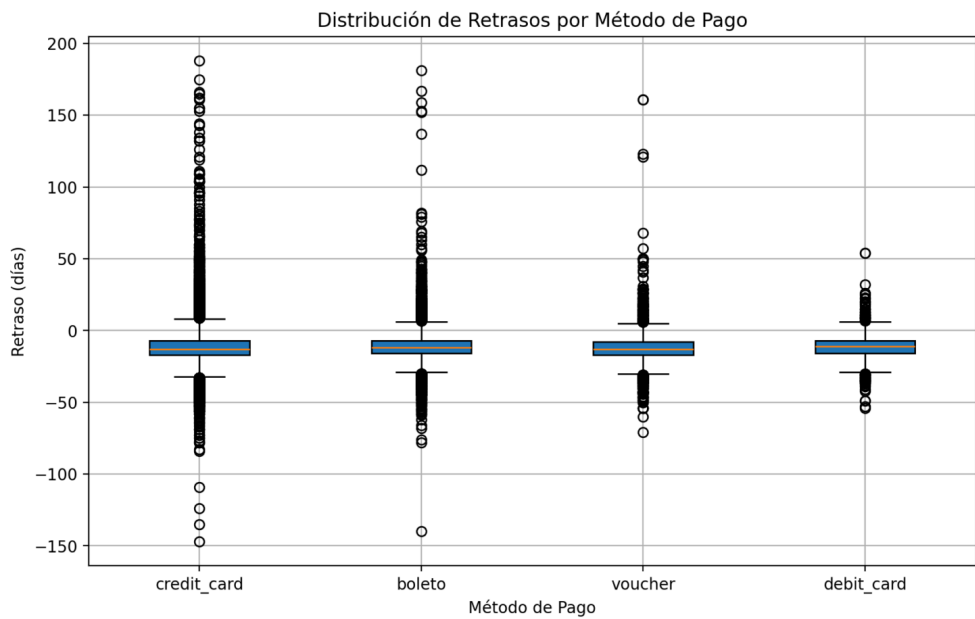
Pagos con Tarjeta de Crédito por Número de Cuotas

Gráfico de barras para visualizar el valor total de pagos en función del número de cuotas



El gráfico muestra la distribución del valor total de pagos realizados con tarjeta de crédito en función del número de cuotas. Observamos que los pagos a 1 y 2 cuotas son los más comunes, representando la mayor parte del valor total de pagos.

Métodos de Pago y Retrasos en la Entrega



Em el gráfico de caja y bigotes se observa la distribución de los retrasos en la entrega según el método de pago, en promedio todos los métodos tienen entregas anticipadas, no se observan diferencias significativas entre los métodos de pago. Existen valores atípicos notables, con algunas entregas extremadamente anticipadas o muy tardías, lo que sugiere variabilidad en los tiempos de entrega independientemente del método de pago.

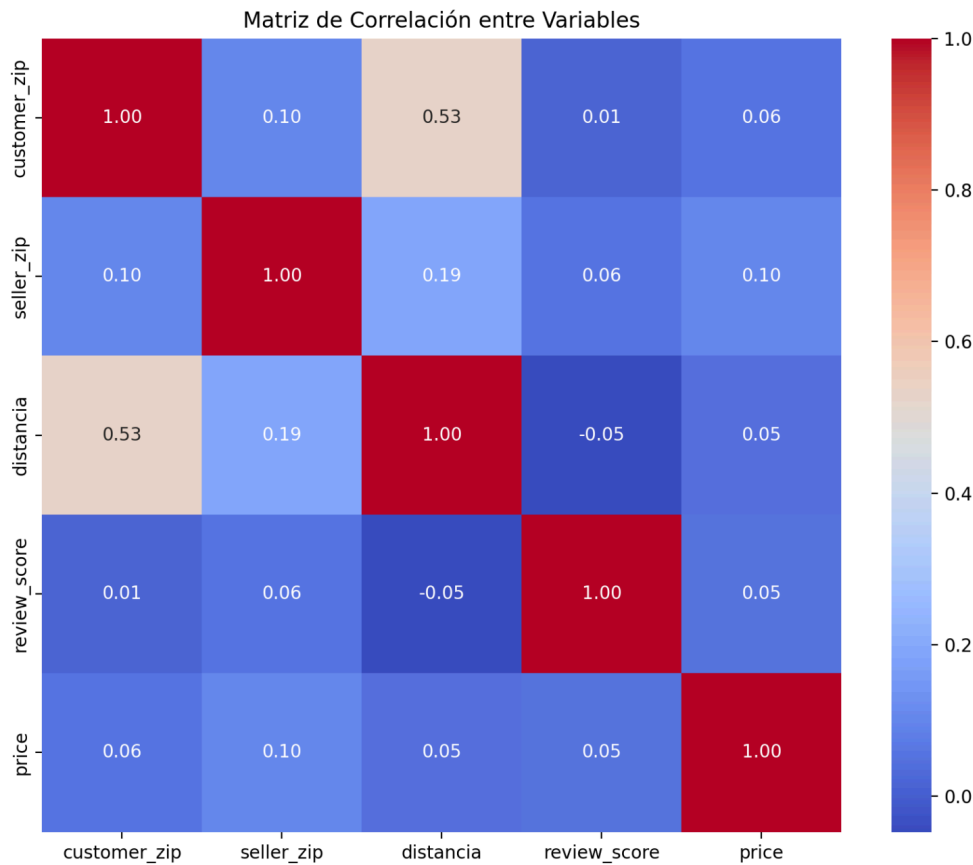
Análisis de Reseñas de Pedidos

Distribución de Calificaciones de Reseñas:

	calificacion	cantidad	porcentaje
0	1	11,282	11
1	2	3,114	3
2	3	8,097	8
3	4	19,007	19
4	5	56,910	58

Error al conectar a la base de datos: 'Cantidad'

Matriz de Correlación entre Variables

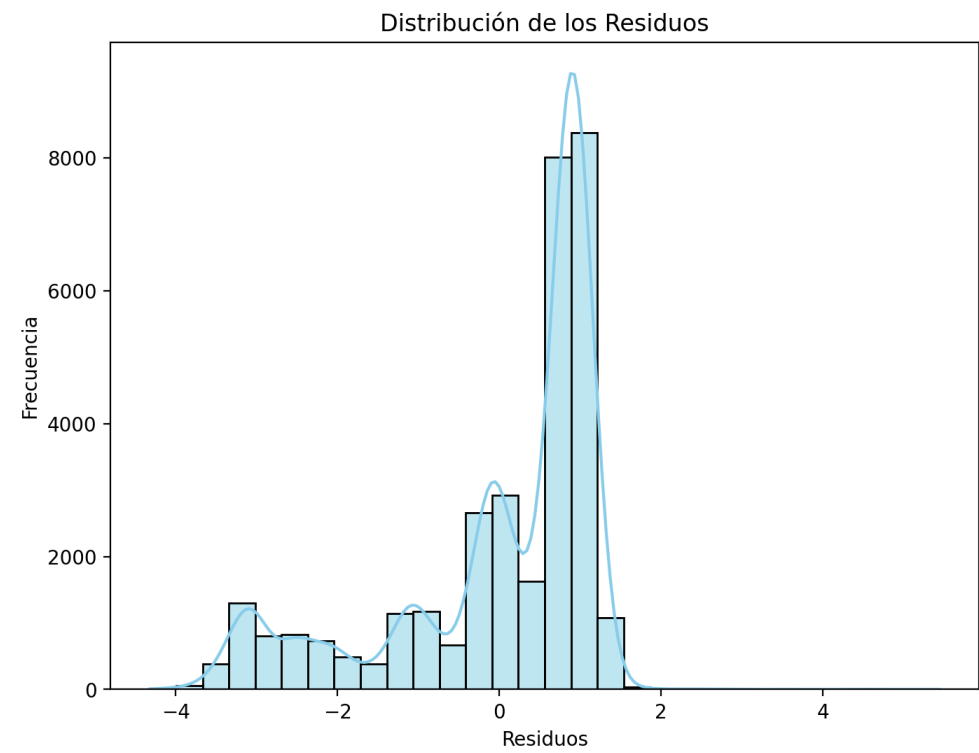


1. Hay una alta correlación positiva entre el código postal del cliente y el del vendedor, por ello se identifica que que los clientes y vendedores tienden a estar geográficamente cercanos
2. Hay una correlación negativa significativa entre la distancia y la puntuación de la reseña. Los clientes que están más lejos del vendedor tienden a dar peores puntuaciones, posiblemente debido a tiempos de entrega más largos o mayores costos de envío. price vs seller_zip (0.99):
3. El precio tiene una correlación casi perfecta con el código postal del vendedor, ciertos vendedores en áreas específicas tienden a vender productos a precios similares por ello se indica que los precios están asociados a ciertas regiones.

Modelo de Regresión Lineal: Predicción del Puntaje de Reseña

Error Cuadrático Medio (MSE): 1.71

Coefficiente de Determinación (R^2): 0.05



1. **Concentración de Residuos:** La mayor parte de los residuos está concentrada alrededor de 0, lo que indica que el modelo predice razonablemente bien la mayoría de las veces, pero hay cierta dispersión.
2. **Distribución Asimétrica:** Existe una asimetría en la distribución de los residuos, con algunos valores concentrándose más hacia residuos ligeramente negativos, lo que podría sugerir que el modelo tiende a subestimar ligeramente el puntaje de las reseñas en algunos casos.
3. **Picos en la Distribución:** Los picos en torno a 0 y 1 indican que el modelo realiza muchas predicciones con errores pequeños, pero también hay varias predicciones con errores moderados, visibles en los picos secundarios..

Ajuste del Modelo de Regresión Lineal y Obtención de Coeficientes

Coeficientes de la Regresión Lineal

	Coeficientes
seller_zip	0
price	0
delivery_delay	-0.0309

Intercepto del Modelo

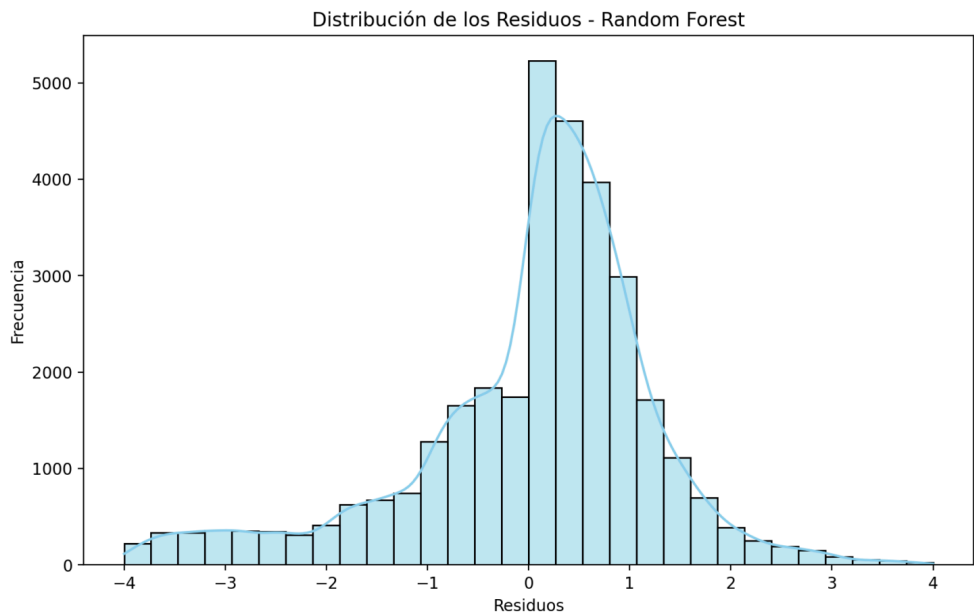
Intercepto: 3.7138252083899777

El modelo sugiere que el retraso en la entrega tiene un impacto negativo en la puntuación de las reseñas, mientras que el código postal del vendedor y el precio del producto no parecen influir significativamente en la puntuación de las reseñas. Sin embargo, la magnitud del impacto del retraso es pequeña, y otros factores no capturados por el modelo podrían estar influyendo en las calificaciones.

Evaluación del Modelo de Random Forest

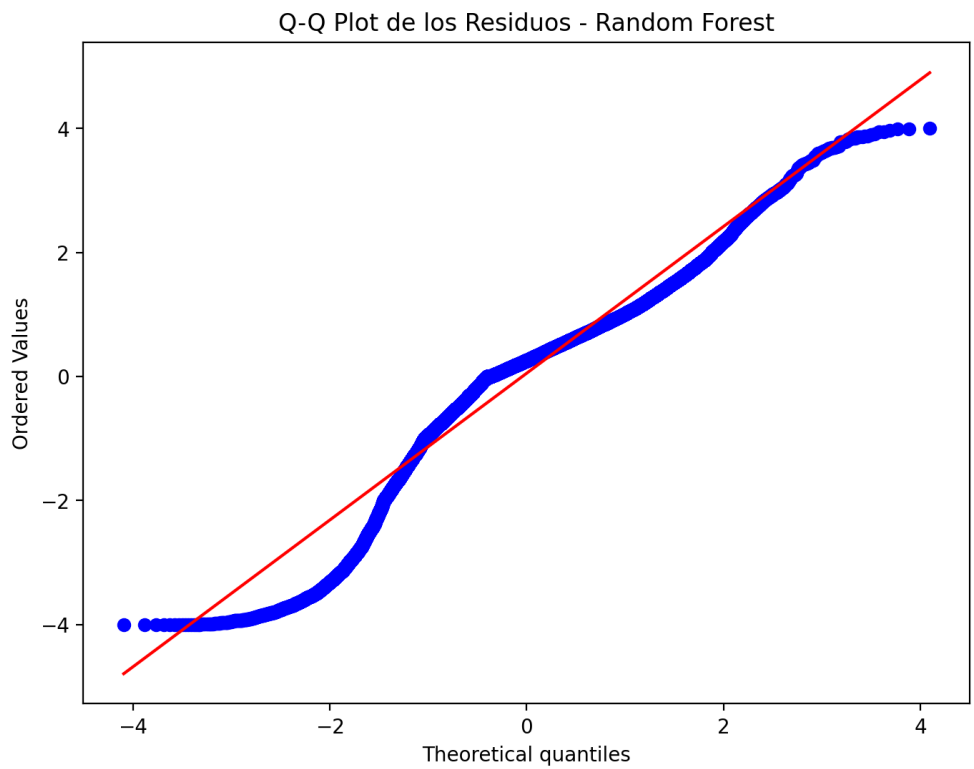
Random Forest - Error Cuadrático Medio (MSE): 1.501967536152944

Random Forest - Coeficiente de Determinación (R^2): 0.169163741321128



El modelo de Random Forest tiene un rendimiento intermedio según los valores de MSE y R^2 . La distribución de los residuos muestra una tendencia a un ajuste decente, el bajo valor de R^2 sugiere que el modelo no captura toda la complejidad del fenómeno que se está tratando de predecir. Es por ello el modelo no lineal o más profundo pueda ser más adecuado para mejorar la precisión.

Q-Q Plot de los Residuos - Random Forest



Los puntos en el Q-Q plot no siguen una línea recta, sino que se desvían significativamente de la línea roja diagonal (que representa la distribución normal). Esta curvatura indica que los residuos del modelo

Random Forest no siguen una distribución normal.

© 2024. Jorge Andres Melo 2024. Programación en Ciencia de Datos.

Análisis culminado.