

IQR y Z(score)

January 19, 2026

Análisis de patrones y outliers mediante IQR

```
[1]: import pandas as pd
df = pd.read_csv("/datasets/everpeak_clean.csv")

# calcular Q1, Q3 e IQR
Q1 = df['order_value'].quantile(0.25)
Q3 = df['order_value'].quantile(0.75)
IQR = Q3 - Q1

# calcular límite inferior y superior
limite_inferior = Q1 - 1.5 * IQR
limite_superior = Q3 + 1.5 * IQR

# Mostrar resultados
print('Primer cuartil: ', Q1)
print('Tercer cuartil: ', Q3)
print('IQR: ', IQR)

print("\nRegistros abajo del límite inferior")
print(df[df['order_value'] < limite_inferior])

print("\nRegistros arriba del límite superior")
print(df[df['order_value'] > limite_superior])
```

```
Primer cuartil: 3094.0
Tercer cuartil: 13160.5
IQR: 10066.5
```

Registros abajo del límite inferior

Empty DataFrame

Columns: [order_id, order_date, customer_id, product_category, price, quantity, order_value, payment_method, city, state, customer_age, quantity_invalid_flag, age_invalid_flag, state_missing_flag]

Index: []

Registros arriba del límite superior

```
order_id order_date customer_id product_category price quantity \
```

8	9	2024-03-14	2925	Beauty	55	833
123	124	2024-06-21	1389	Electronics	1947	16
155	156	2024-08-30	2901	Grocery	36	1278
199	200	2024-12-18	2919	Fashion	5174	7
265	266	2024-07-31	1905	Electronics	2897	10
...
4934	4935	2024-05-19	1502	Electronics	4282	11
4936	4937	2024-12-17	1907	Home	1522	19
4942	4943	2024-11-02	1979	Electronics	6478	11
4947	4948	2024-07-09	1447	Beauty	45	1015
4952	4953	2024-10-13	2993	Sports	2679	12

	order_value	payment_method	city	state	customer_age	\
8	45819	credit_card	Boston	MA	40	
123	31152	debit_card	Boston	MA	45	
155	45998	credit_card	Miami	FL	32	
199	36218	credit_card	Los Angeles	CA	24	
265	28970	paypal	Miami	FL	31	
...
4934	47102	credit_card	Seattle	WA	47	
4936	28918	credit_card	New York	NY	48	
4942	71258	debit_card	Seattle	WA	49	
4947	45697	credit_card	Seattle	WA	37	
4952	32148	credit_card	Los Angeles	CA	34	

	quantity_invalid_flag	age_invalid_flag	state_missing_flag
8	0	0	0
123	1	0	0
155	0	0	0
199	1	0	0
265	1	0	0
...
4934	1	0	0
4936	1	0	0
4942	1	1	0
4947	0	0	0
4952	1	0	0

[165 rows x 14 columns]

Se concluye que los valores extremos representan el 7.1% de la base de datos, por lo tanto, estos outliers representan casos reales del negocio que aportan información valiosa.

Analisis de outliers mediante Z(score)

```
[ ]: import pandas as pd
df = pd.read_csv("/datasets/everpeak_clean.csv")
```

```
#cálculo de la media
mean = df['order_value'].mean()

#cálculo de la desviación estándar
std = df['order_value'].std()

#Crea el z score
df['z'] = (df['order_value'] - mean) / std
#Calcula los valores extremos
print(df[df['z'].abs() > 3])
```

Se puede observar que ambos metodos indican la presencia de outliers, sin embargo, el IQR es mas eficiente si se requiere analizar sensibilidades al cambio en las retención del consumidor. Y el Z score es menos sensible ante esos cambios se puede analizar metricas típicas como el nivel de ventas, la estacionalidad, etc, pero no puede ayudar a predecir los cambios repentinos del mercado como lo haria IQR.