

# **Proyecto Final - CRISP-DM (Etapas 4,5 y análisis)**

Alumno: Berenice Jazmín Antonio Leyva

Resumen de entregables:

- Evaluación del modelo y resultados      Dataset: RAMEN\_DATA\_SET.xlsx
- Análisis de confiabilidad (Alfa de Cronbach)
- Matriz de correlaciones y KPIs
- Gráficas: distribuciones y mapa de correlación
- Interpretaciones y propuestas para despliegue

Fecha de generación: 2025-11-30

## Vista previa del dataset (primeras filas)

Review #	Brand	Style	Country	Total	Unnamed: 5	K	4
2580	1.0	1.0	1.0	2583.0		SVi	252591.0
2579	2.0	2.0	2.0	2585.0		Vt	719.6549165120591
2578	3.0	1.0	3.0	2585.0			
2577	4.0	2.0	2.0	2585.0		SECCION1	1.333333333333333333
2576	5.0	2.0	4.0	2587.0		SECCION2	-349.9890562885738
2575	6.0	2.0	5.0	2588.0		ABSOLUTO S2	349.9890562885738
2574	7.0	1.0	1.0	2583.0			
2573	8.0	3.0	1.0	2585.0		Alpha	466.6520750514317
2572	9.0	2.0	1.0	2584.0			
2571	10.0	2.0	6.0	2589.0			

## Estadísticos descriptivos (columnas numéricas seleccionadas)

Variable	count	mean	std	min	25%	50%	75%	max
Review #	100.0	5012.0797	.184584781335	2481.0	2505.75	2530.5	2556.25	250502.0
Brand	100.0	24.586	2868977346024	1.0	3.0	9.5	20.25	1234.0
Style	100.0	2.380	5457117666718	1.0	2.0	2.0	2.0	219.0
Country	100.0	12.621	4342820319882	1.0	2.75	5.0	10.25	636.0
Total	95.51	424242424220.32	6384708194645	2493.0	2530.0	2554.0	2574.5	2590.0

## Análisis de Confiabilidad - Alfa de Cronbach

Columnas usadas para Alfa: Review #, Brand, Style, Country, Total

Alfa de Cronbach calculado: 0.5079

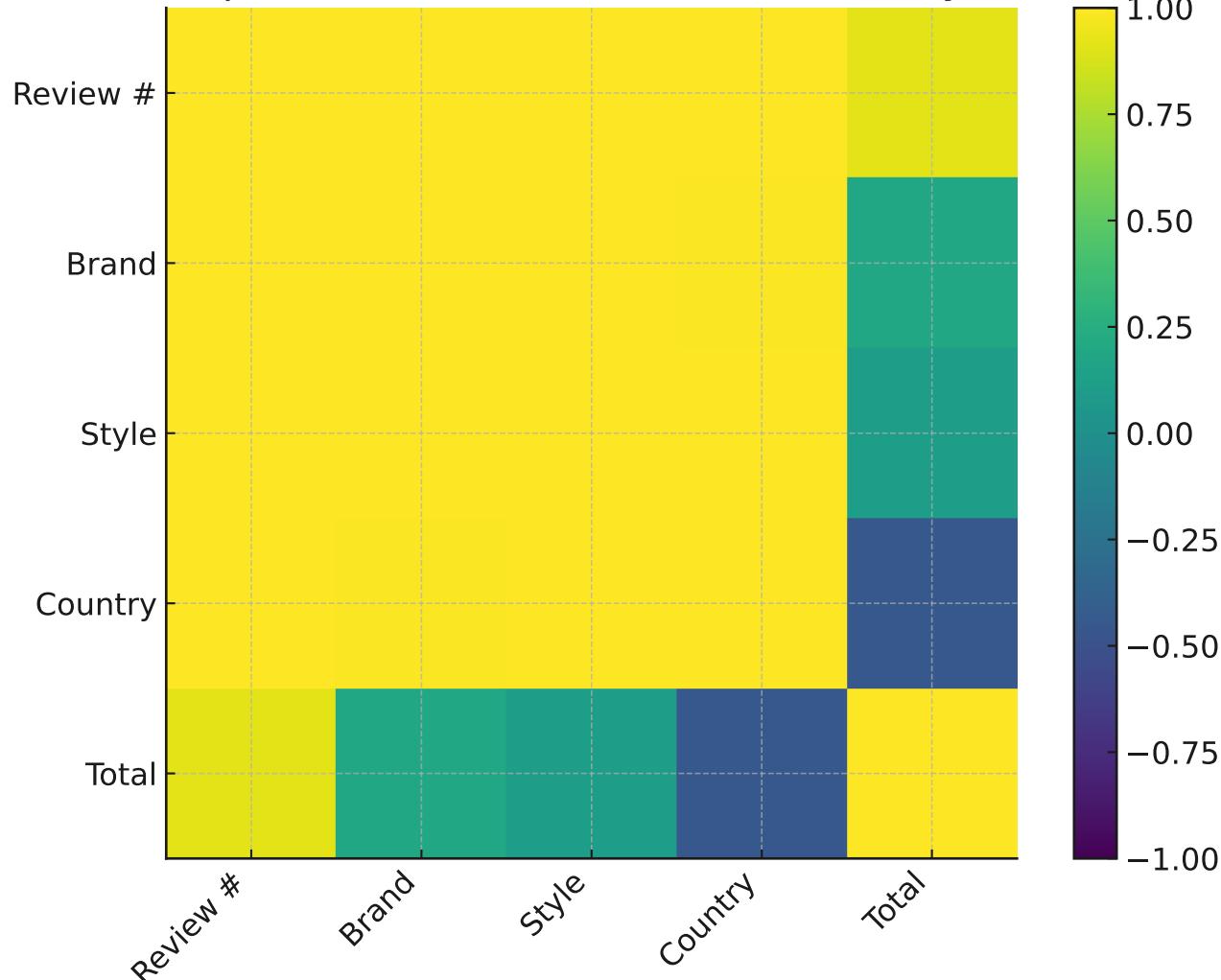
Interpretación:

Baja consistencia interna; los ítems no miden bien la misma construcción.

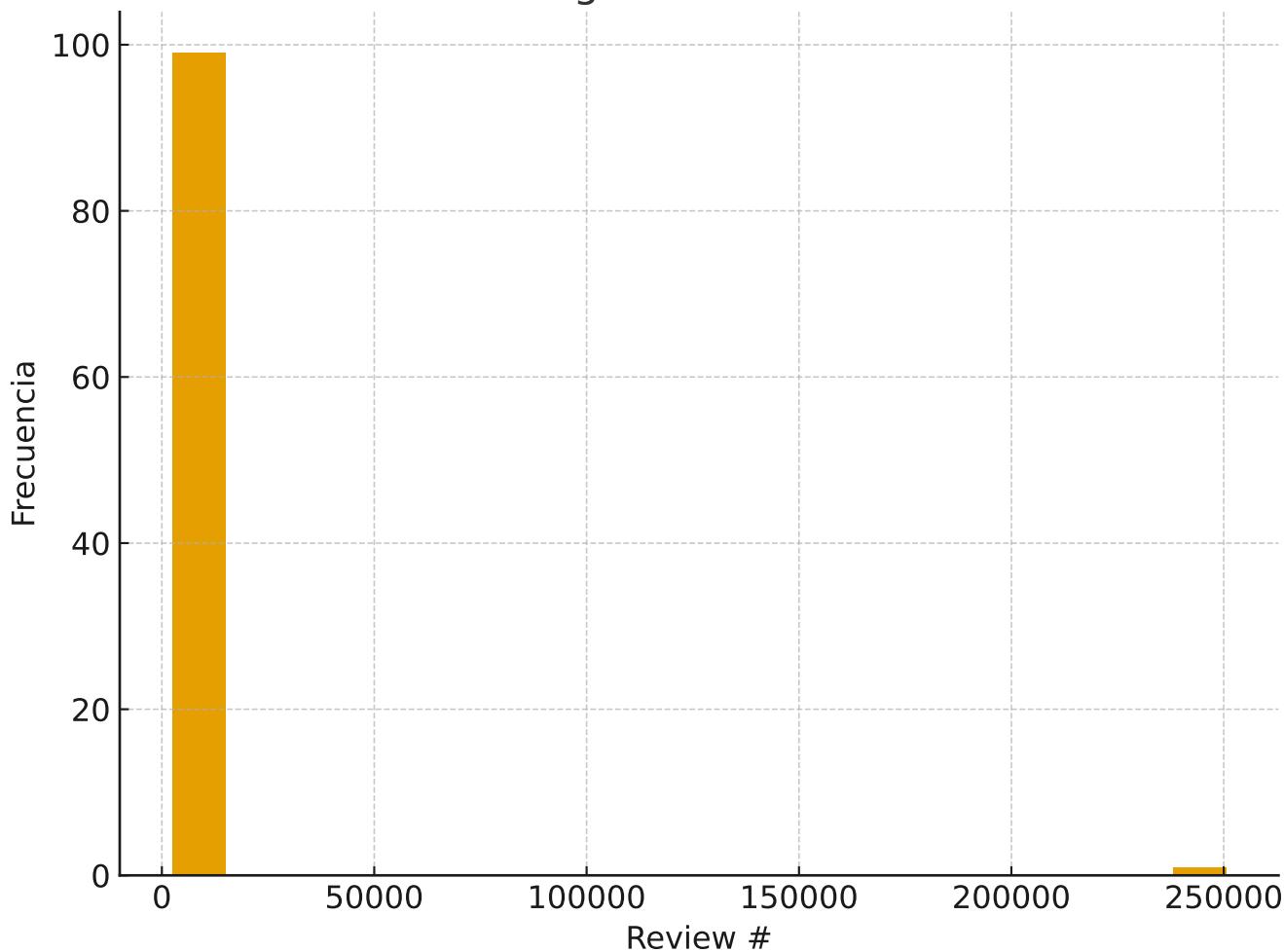
## Matriz de correlaciones (Pearson)

	Review #	Brand	Style	Country	Total
Review #	1.0	0.996	0.999	0.997	0.921
Brand	0.996	1.0	0.996	0.992	0.19
Style	0.999	0.996	1.0	0.995	0.11
Country	0.997	0.992	0.995	1.0	-0.446
Total	0.921	0.19	0.11	-0.446	1.0

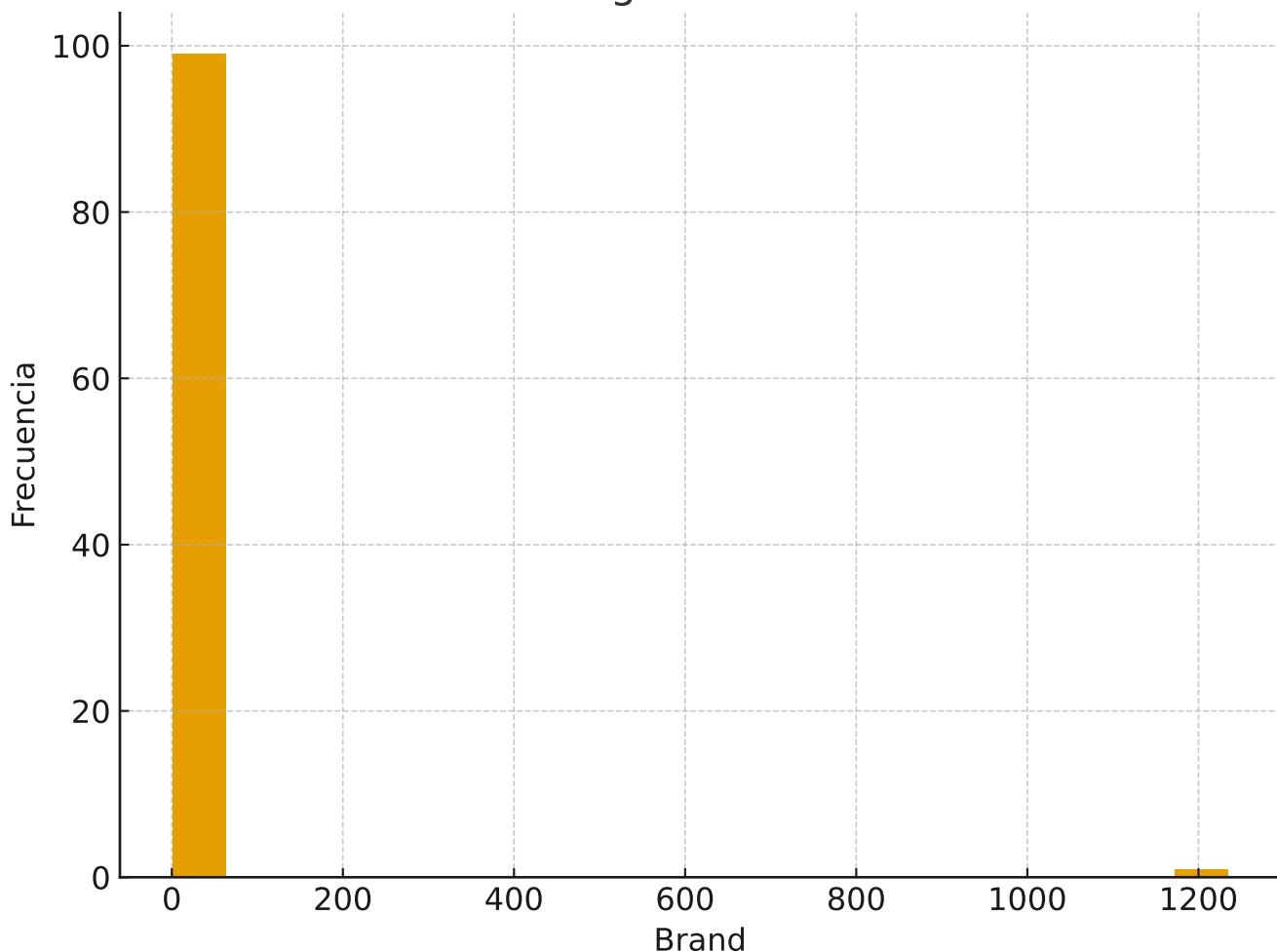
## Mapa de correlación (valores entre -1 y 1)



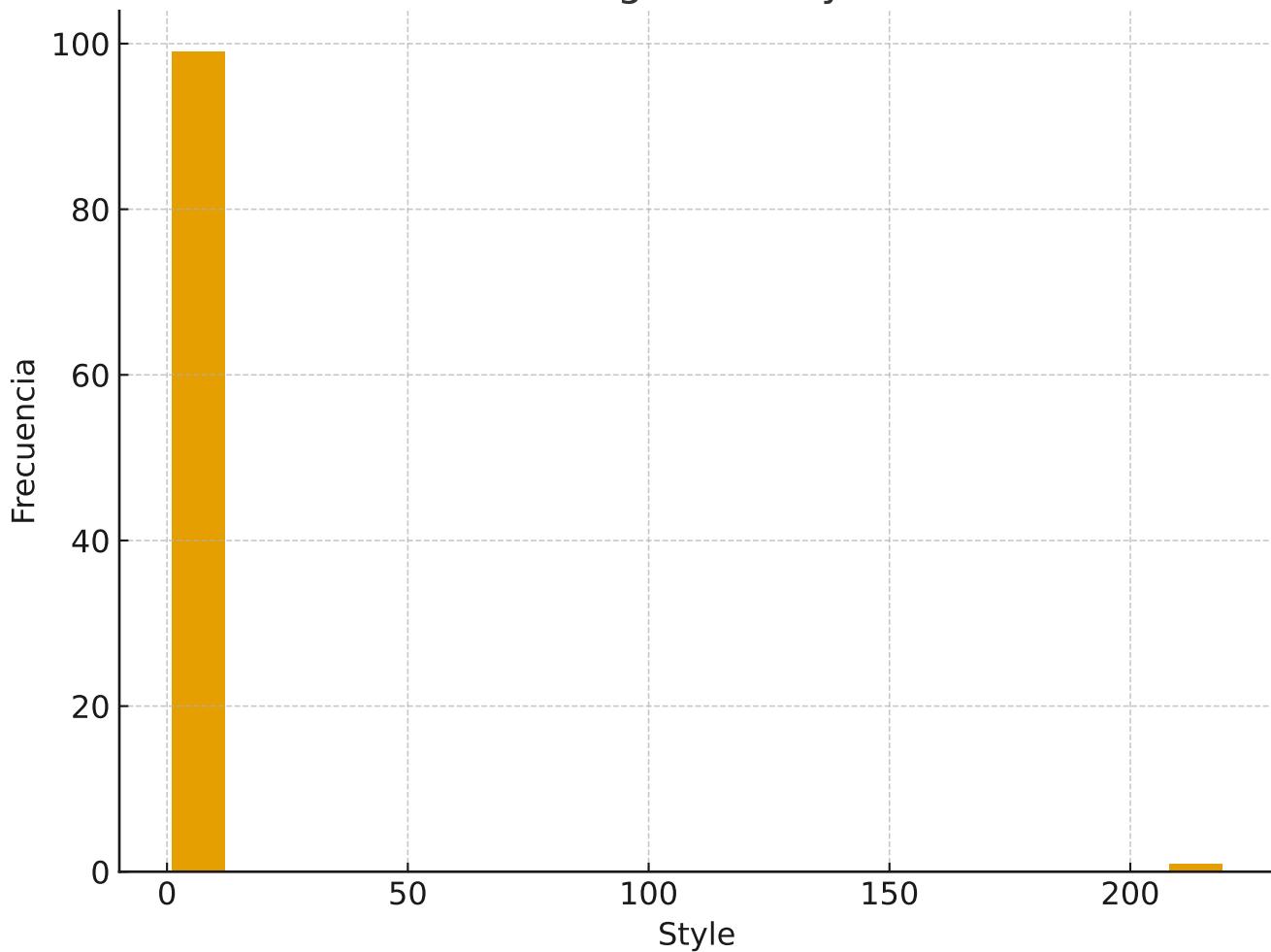
# Histograma: Review #



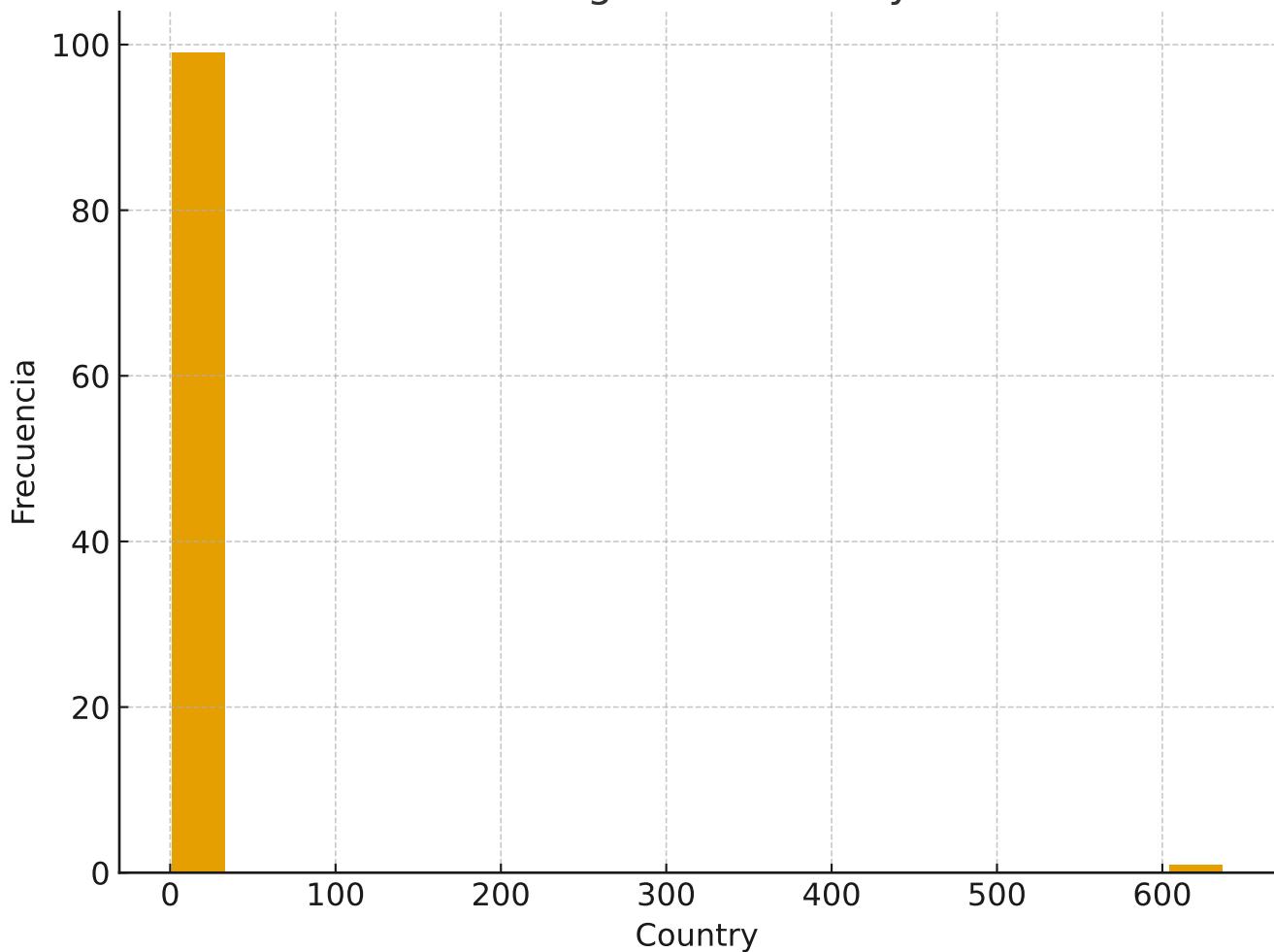
## Histograma: Brand



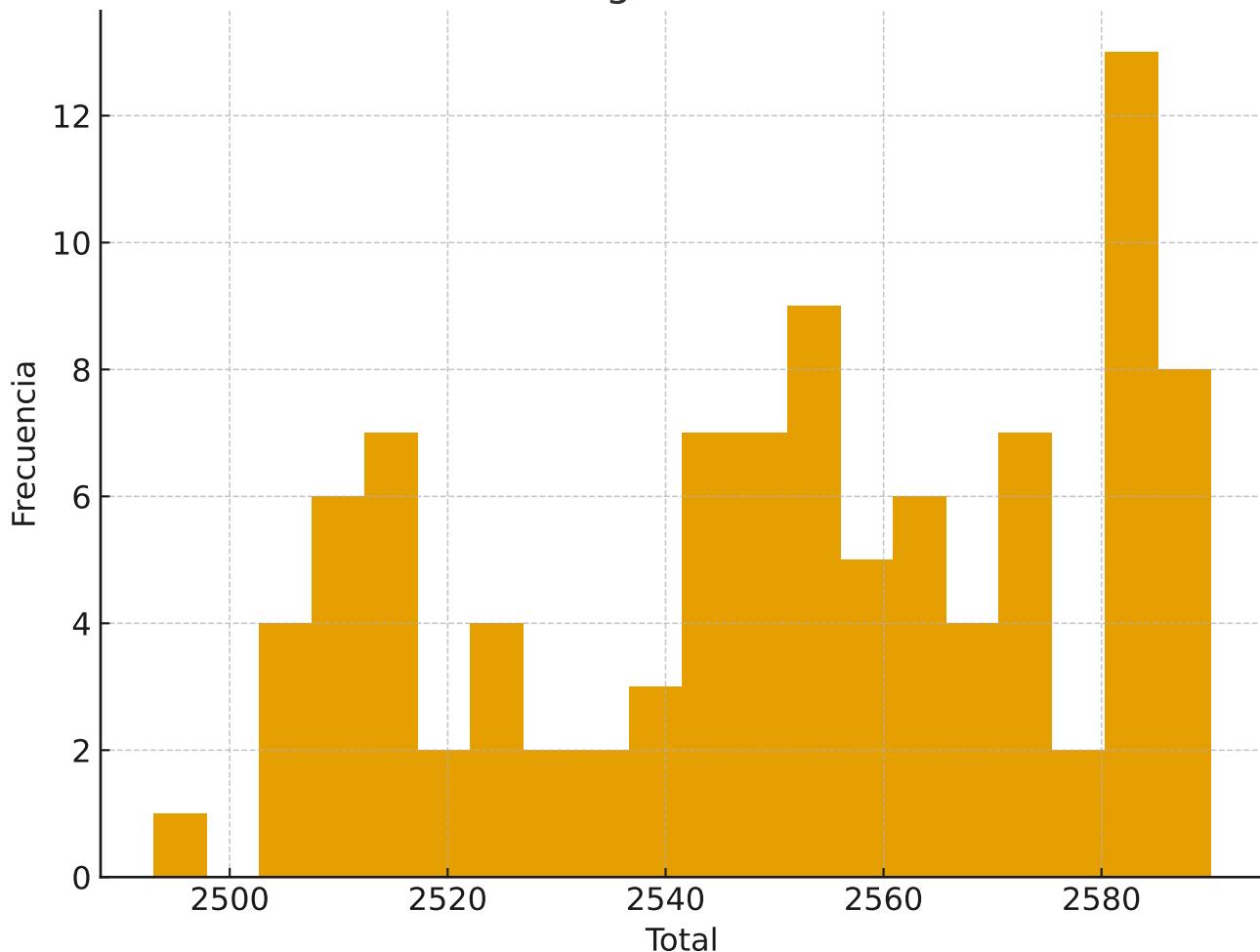
## Histograma: Style



## Histograma: Country



## Histograma: Total



Brand ↔ Style:  $r = 0.996 \rightarrow$  Nivel: Muy alta  
Review # ↔ Brand:  $r = 0.996 \rightarrow$  Nivel: Muy alta  
Style ↔ Country:  $r = 0.995 \rightarrow$  Nivel: Muy alta  
Brand ↔ Country:  $r = 0.992 \rightarrow$  Nivel: Muy alta  
Review # ↔ Total:  $r = 0.921 \rightarrow$  Nivel: Muy alta  
Country ↔ Total:  $r = -0.446 \rightarrow$  Nivel: Moderada  
Brand ↔ Total:  $r = 0.190 \rightarrow$  Nivel: Muy baja  
Style ↔ Total:  $r = 0.110 \rightarrow$  Nivel: Muy baja

## Pares de variables con correlaciones más relevantes (ordenadas por magnitud)

## Etapa 5 - Despliegue (sugerencias)

Sugerencias para despliegue del modelo / análisis:

- 1) Validación adicional en dataset ampliado: recolectar más muestras y validar estabilidad de correlaciones.
- 2) Preparar pipelines de preprocesamiento (limpieza, imputación, codificación de variables categóricas).
- 3) Implementar dashboard de KPIs (por ejemplo: r de correlación, medias por marca/país, tendencias).
- 4) Documentar supuestos y limitaciones (baja/muy baja consistencia interna en algunos ítems).
- 5) Si se pretende usar para predicción (p. ej. predecir Brand desde Style y Country), entrenar modelos supervisados y comparar m