

Mejorar operaciones y fidelización de clientes?

Un análisis con ciencia de datos.

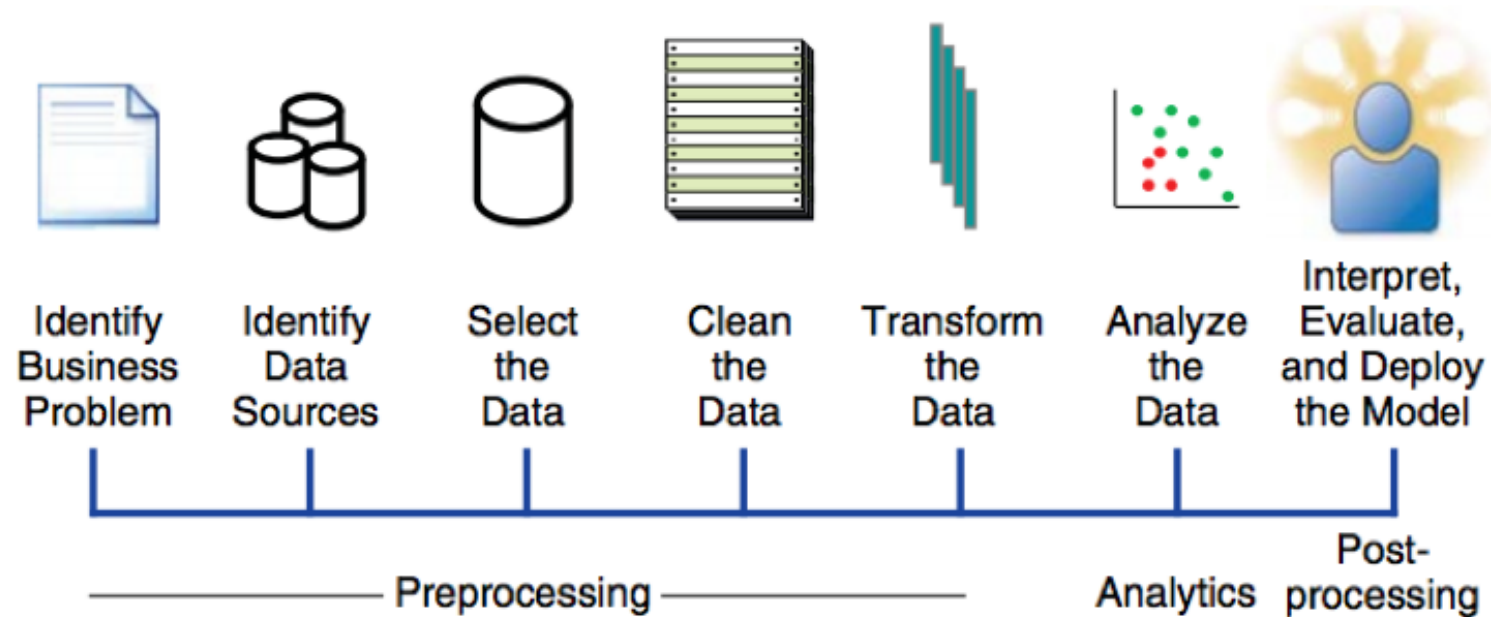


Objetivo:

- Responder a **cinco preguntas claves**, sugiriendo estrategias sobre como las operaciones de Tottus Chile se pueden mejorar con enfoque en la fidelización de clientes.
- Proveer en conjunto de **visualizaciones** que expongan los hallazgos.



Modelo de proceso analítico



[Fuente: SAS](#)

Datos:

- Ventas en Tottus.
- 2010/12/01 – 2011/12/09
- Datos recopilados de 38 países.
- 541.909 registros
- 8 variables



Variables:

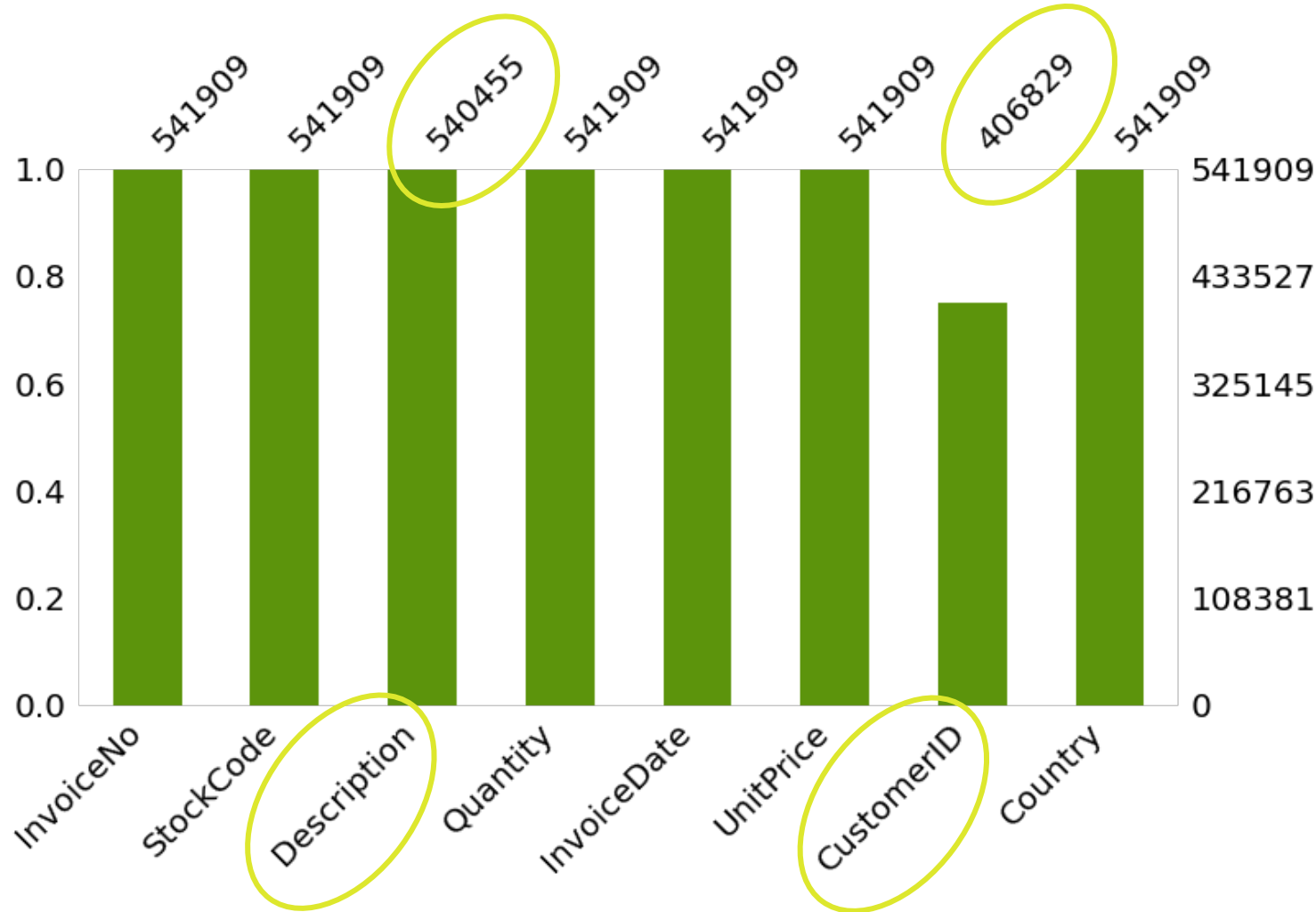
Variable	Descripción	Tipo	Valores Faltantes
InvoiceNo	Número de factura	Object	No
StockCode	Número de inventario	Object	No
Description	Descripción del Producto	Object	Si (1.454)
Quantity	Cantidad vendida	Integer	No
InvoiceDate	Fecha de factura	Object	No
UnitPrice	Precio de unidad	Float	No
CustomerID	Identificación del cliente	Float	Si (135.080)
Country	País	Object	No

Preprocesamiento

Problemas para tratar

- Valores faltantes
- Transformaciones de data type
- Registros de devoluciones
- Registros para ajustar deuda incobrable

Valores Faltantes



- Remover registros?
- Reemplazar con modo?
- Analizar StockCode?
- Analizar InvoiceNo?

Solución: Se remueven los registros con valores faltantes ya que aun así hay mas de 400.000 registros para incluir en el análisis.

UnitPrice Valores Negativo

- Dos registros que se refieren a ajustar deuda incobrable.
- Al remover los valores faltantes estos registros ya se removieron.

InvoiceDate → Date Time

- Se convirtió InvoiceDate de Object a DateTime.
- Se Crearon cuatro nuevas variables para:
 - Invoice_DayofWeek
 - Invoice_Hour
 - Invoice_month
 - Invoice_Time
 - Invoice_Day

Quantity < 0

- Hay 10.624 con cantidad menor que 0.
- Representan devoluciones.
- Se remueven de la base de datos.

CustomerID → Object

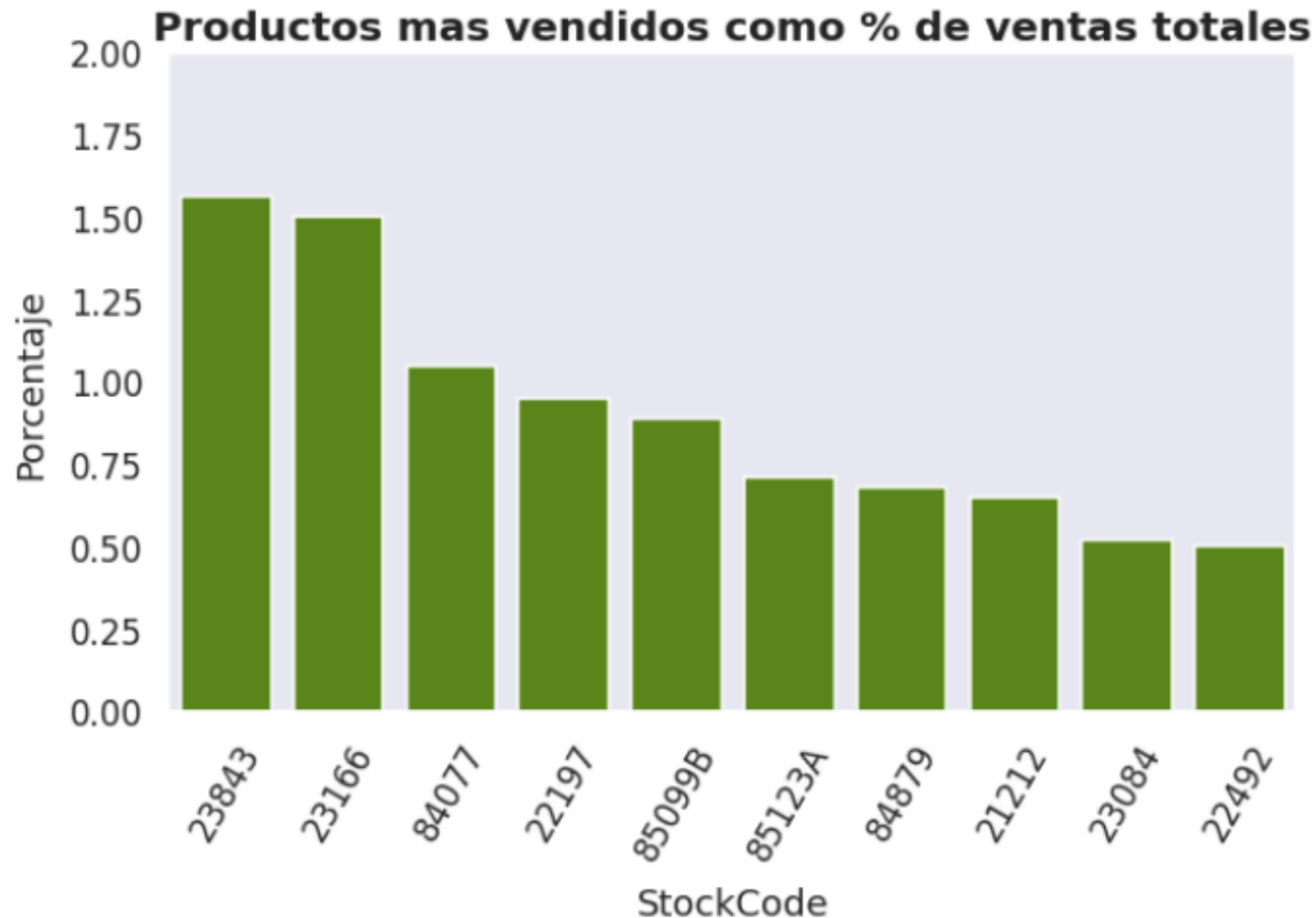
- Guardado como float.
- Se convierte a un integer y después a un object.

Resumen del preprocesamiento

- Se partió con 541.909 registros y 8 variables.
- Al terminar el preprocesamiento hay 397.924 registros y 14 variables.

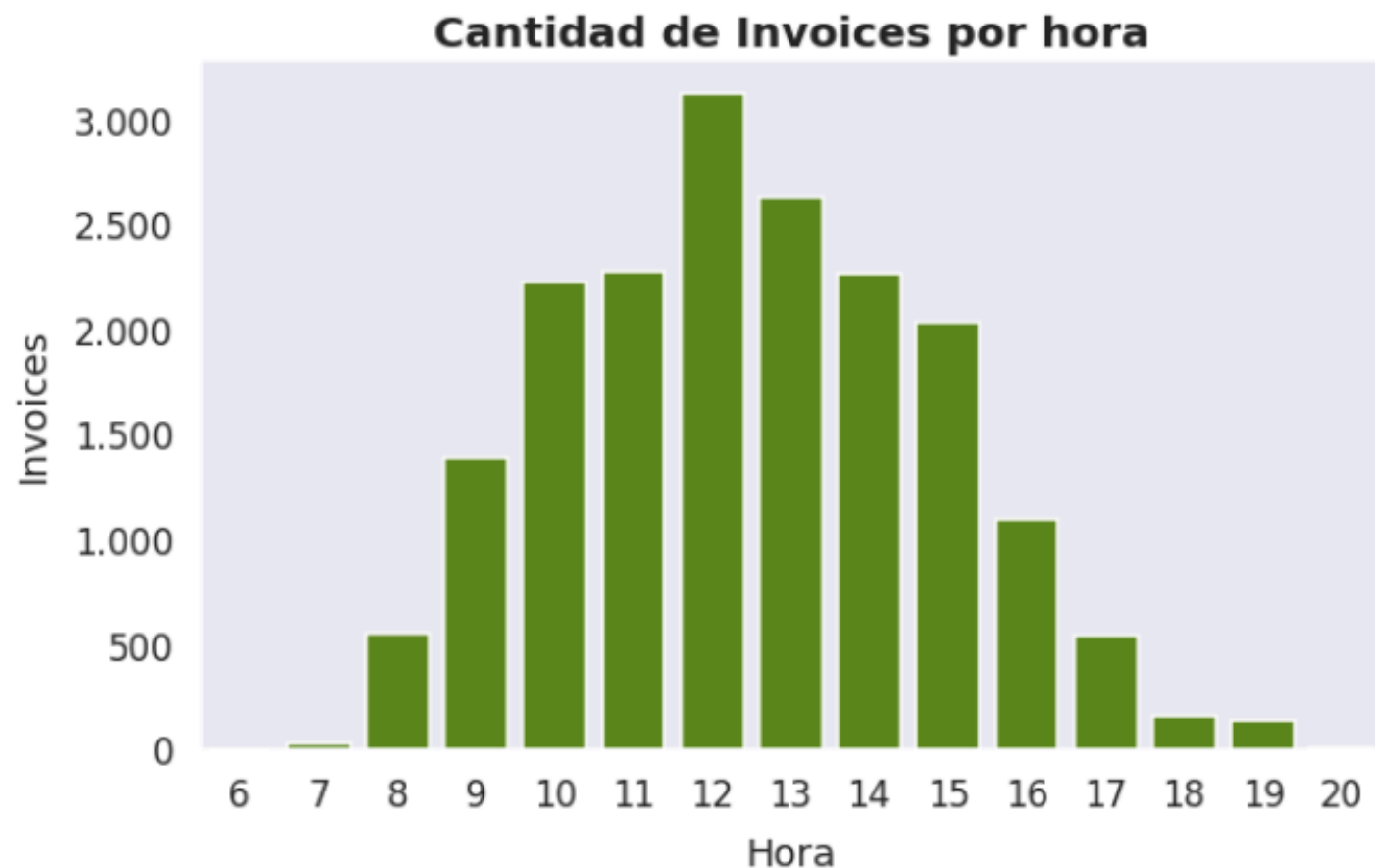
Preguntas Claves

Existen productos favoritos?



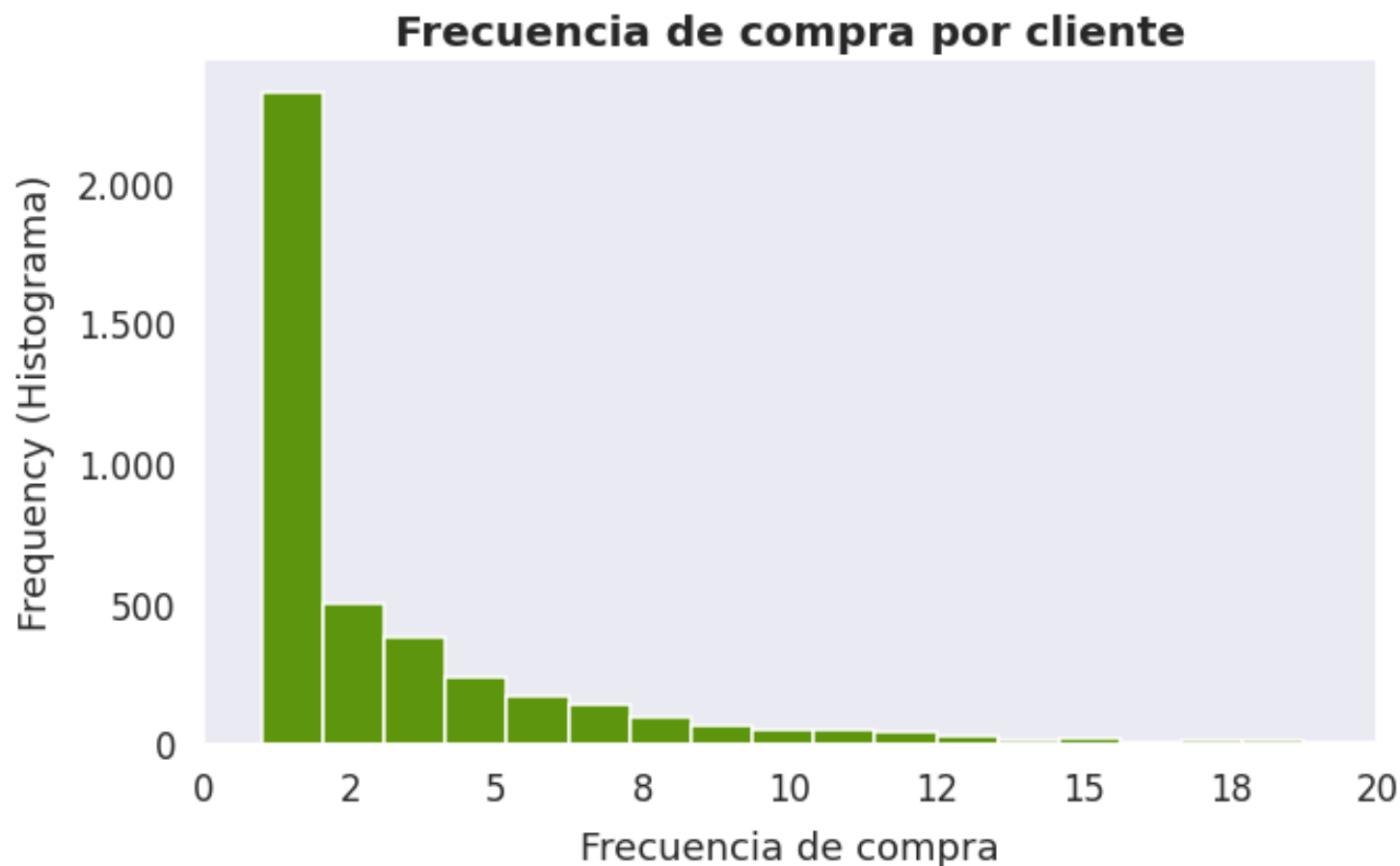
- **3.665** tipos de productos vendidos.
- Tres productos representan mas de 1% de todas las ventas (Stock code: **23843**, **23166**, **84077**).

En que horario se compra mas?



- 18.537 Invoices.
- **10am – 3pm** es el horario mas popular. Se aconseja poner **refuerzas en las tiendas** en este horario.
- Pocas compras antes de 9am y después de 5pm.

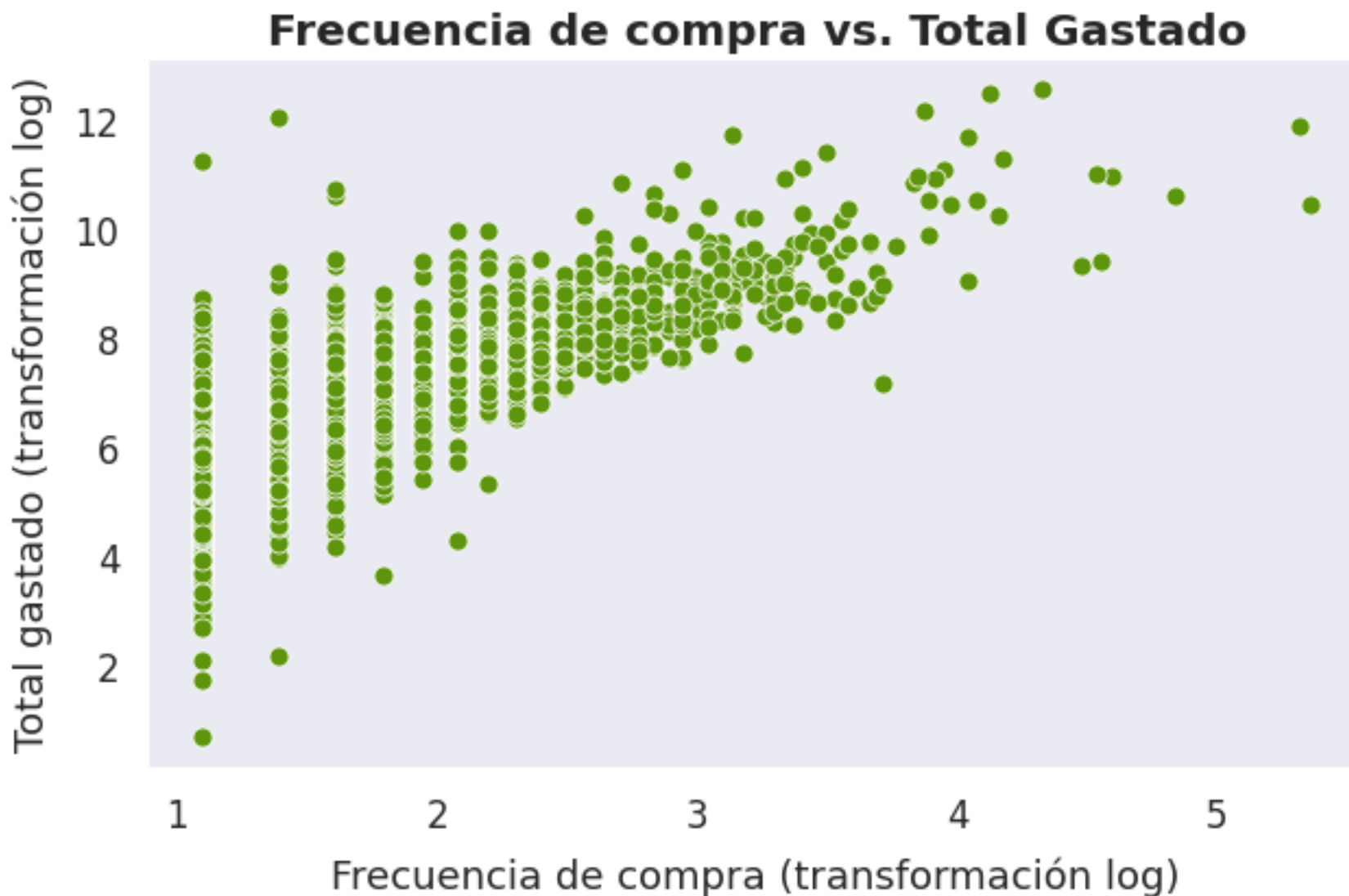
Frecuencia de compra



- 4.339 clients
- **34%** de los clientes tienen una frecuencia de compra de uno.
- **65%** de los clientes tienen una frecuencia de compra de uno, dos, o tres veces.
- El **10% de clientes** con mas frecuencia de compra tienen un rango de 9 a 210 veces, un **promedio de 19 veces**, y una mediana de 13 veces.



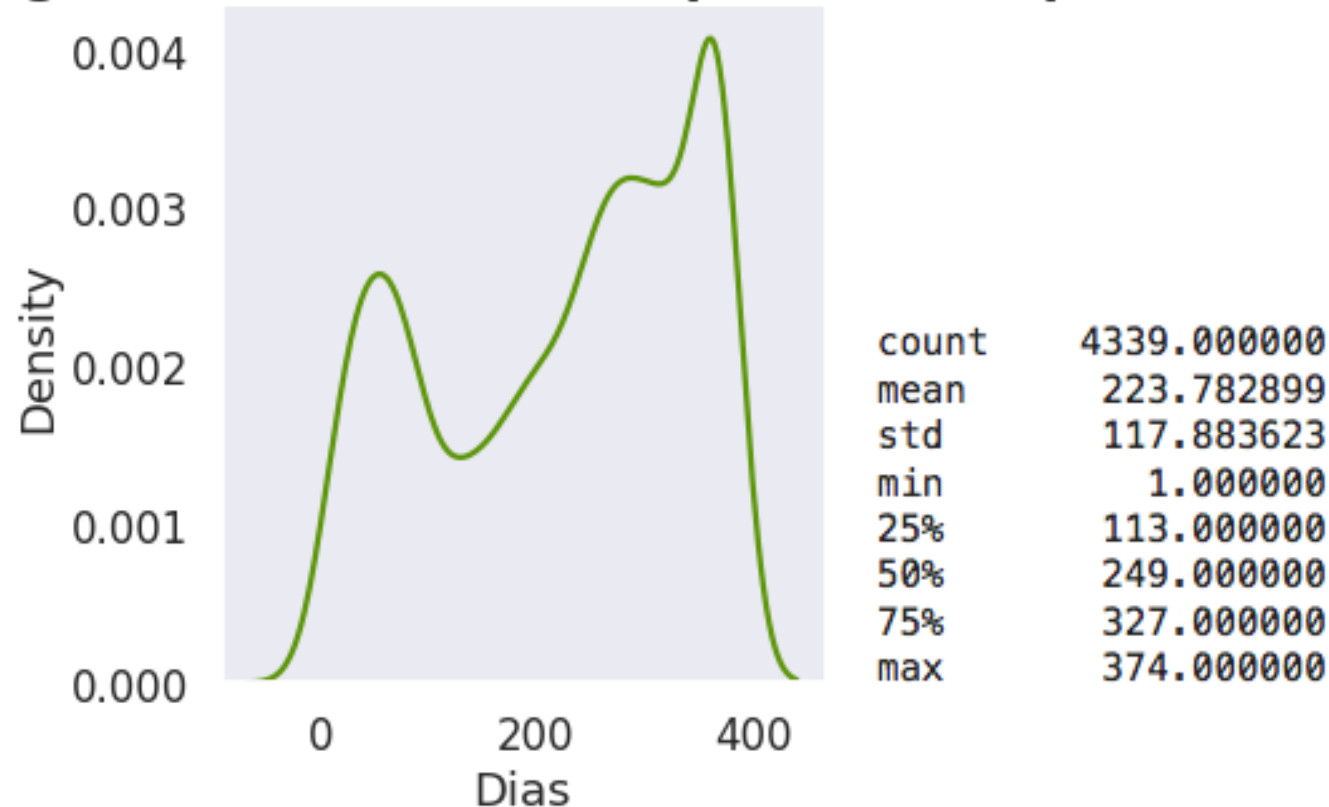
Cuanto gastan los clientes?



- Transformación log para visualizar la relación entre frecuencia de compra y total gastado de una manera mejor.
- La media gastada para cada cliente es **\$2.054**.
- La media para cada invoice es **\$419**.

Antigüedades – Días desde primera compra.

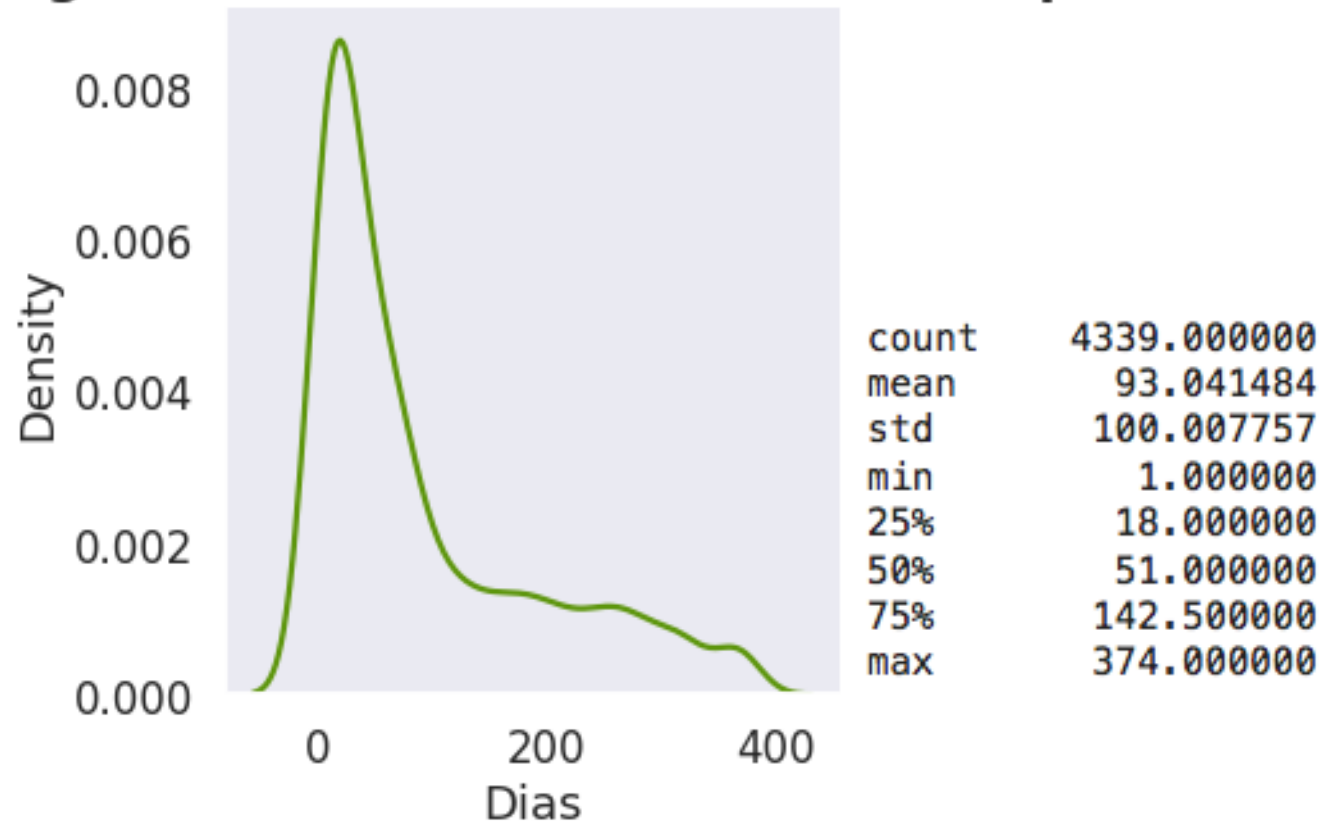
Antigüedades - Dias desde primera compra



- Promedio de 224 días desde primera compra.

Antigüedades – Dias desde ultima compra.

Antigüedades - Dias desde ultima compra



- Promedio de 93 días desde ultima compra.
- Mediana de 51 días.

Clustering clientes según frecuencia de compra

- Algoritmo Kmeans.
- Configuraciones.
- Tres grupos.
- Grupo 0 es mas poblado.
- Grupo 0 menor frecuencia de compra
- Grupo 1 mas frecuencia de compra.

```
kmeans = KMeans(  
    init="random",  
    n_clusters=3,  
    n_init=10,  
    max_iter=500,  
    random_state=42)
```

```
grupo  
0    4002.0  
1         8.0  
2    329.0  
Name: count, dtype: float64
```

```
grupo  
0    -0.184095  
1    15.280219  
2     1.867800  
Name: mean, dtype: float64
```



Características de los clusters

- Registros de **tres países** componen grupo 1.

1	United Kingdom	19645
	EIRE	5677
	Netherlands	2080

- En promedio grupo 2 compra en mayor **cantidad.**

Quantity	
grupo	
0	10.534921
1	15.020874
2	18.159463

- La media para **UnitPrice no varia** tanto entre los tres grupos.

UnitPrice	
grupo	
0	3.078103
1	3.016404
2	3.226255

- Grupo 2 **gasta mas** en promedio.

Total_gastado	
grupo	
0	17.599766
1	23.417366
2	32.983246



Resumen de insights

- Productos con Stock code: **23843**, **23166**, **84077** son los mas populares.
- **10am – 3pm** es el horario mas popular. Se aconseja poner **mas trabajadores** en las tiendas en este horario.
- Por otro lado hay pocas compras antes de 9am y después de 5pm.
- **65%** de los clientes tienen una frecuencia de compra de uno, dos, o tres veces.
- El **10% de clientes** con mas frecuencia de compra tienen un **promedio de 19 veces.**



Resumen de insights...

- La media gastado para cada cliente es **\$2.054**, mientras la media para cada invoice es **\$419**.
- Hay un promedio de **224 días** desde la primera compra de un cliente.
- Promedio de **93 días** desde ultima compra.
- Tres grupos creados con clustering. **Grupo 1** tiene un frecuencia de compra **mas alta**. Sugiere que los clientes en este grupo son mas fieles.