

## Limpieza de datos (creación de la tabla *noise*)

El fichero *marketing\_campaign.csv* con el que se ha hecho el análisis de los datos, dispone 38 registros que pueden ser considerados como “ruido” debido a que están incompletos o reflejan una distorsión con respecto al volumen global de los mismos.

**Los 38 registros considerados no son eliminados del modelo pero si separados del análisis para poder informar al proveedor de los datos** con tal de que nos confirmen si realmente son datos eliminables de forma definitiva o podrían ser complementados con nueva información. Los criterios para separar los datos del modelo de estrella son:

- El acumulado de sus ventas agrupadas son de 42€ y no se especifican los canales por los cuáles se han hecho esas ventas.
- En concreto, se trata de los clientes con los códigos de cliente 3955, 5555, 6862, 9931, 11110 y 11181.
- Existen clientes con edades que van más allá de los 100 años o de añadido, con unos ingresos de 666.666 € que están muy alejados de la media de los datos.
- Existe también una columna de datos en la que todos los registros relevantes a los ingresos son 0 €, y por tanto, no permitiría al modelo el poder valorar la repercusión de las campañas sobre su capacidad de compra.

A nivel más general, se elimina la columna *Z\_Revenue* que muestra el valor 11 en todos los registros. Todo indica que se trata de “ruido” proveniente de la base de datos inicial. También se elimina la columna *Z\_Contact* que muestra el valor 3 en todos los registros. Todo indica que se trata de “ruido” proveniente de la base de datos inicial.

Se ha cambiado el formato de los datos de algunas de las columnas (por ejemplo de texto a entero largo) para poder facilitar el posterior cálculo y análisis de la información.

## Creación de la tabla de hechos *facts\_clean*

Por defecto todos los registros no que no están en la tabla *facts\_noise* están incluidos en la tabla *facts\_clean* que es con la que se va a crear todo el modelo de análisis de estrella.

## Creación de nuevas tablas (dimensiones y calendario)

A partir de la tabla de hechos, se crean las tablas de dimensiones que nos permiten desarrollar al máximo la granularidad del modelo de estrella. Se crean tablas de duplicados, se eliminan las columnas sobrantes, así como, la duplicidad de registros para que la tabla solo refleje datos únicos.

### Tablas de dimensiones

- *Estado Civil:*
- *Educación:*
- *Fidelización:*

Gracias al impacto que puedan tener las campañas sobre la decisión de compra de los clientes, estos se pueden agrupar en función de las veces que las campañas les han influido. Gracias a la información de la BB.DD. podemos crear varias categorías.

**Campañas    Clasificación**

0	Indiferente
1	Ocasional
2	Captado
3	Habitual
4	Fiel
5	Embajador

*Generación:*

Gracias a las franjas de edad que nos da el modelo podemos categorizar a los clientes por generación. Las categorías creadas son las siguientes:

**Generación            Nacimiento**

Generación X	1980..1995
Boomer I	1964..1979
Boomer II	1954..1963
Post War	1946..1953
WWII	1927..1945

Nivel de renta (clase social).

Se categorizan los clientes en base a los niveles de ingresos que se muestran en la BB.DD. por los siguientes tramos.

<b>Clase</b>	<b>Renta</b>
Baja	De 0 a 25.000
Media-Baja	De 25.000 a 50.000
Media	De 50.000 a 75.000
Media Alta	De 75.000 a 100.000
Alta	De 100.000 a 150.000
Rico	Más de 150.000

Creación de la tabla calendario

Desde cada uno de los registros de la tabla de hechos, se ha creado una tabla calendario como base para la creación de las medidas del modelo de datos en estrella, en la que se ha desglosado la información en columnas aplicando criterios de segmentación basados en años, semanas, días, ...

## Creación de la tabla de medidas DAX

Tal como se estipula en las instrucciones del ejercicio he creado una serie de medidas que están todas incluidas en el tabla medidas DAX. En concreto:

1. <u>%complain</u>	Calcula el porcentaje de reclamaciones
2. <u>%deals</u>	Calcula el porcentaje de compras
3. <u>acumulado_clientes</u>	Suma el total de clientes
4. complaintot	Suma el total de quejas
5. Deals_catalog	Suma el total de compras por catálogo
6. DealsStore	Suma el total de compras por tienda
7. DealsTot	Suma el total de compras acumuladas
8. DealsWeb	Suma el total de compras por web
9. <u>MntCursotmer2years</u>	Suma el total de compras de todos los clientes sin desglosar el tipo de producto que se haya adquirido
10. pdtFishTot	Suma el total de compras de pescado
11. pdtFruiTot	Suma el total de compras de fruta
12. pdtGoldTot	Suma el total de compras de oro (productos de lujo)
13. pdtMeatTot	Suma el total de compras de carne
14. PdtSweeTot	Suma el total de compras de caramelos (o productos dulces)
15. pdtWineTot	Suma el total de compras de vino
16. Ticket_medio	Da el importe del ticket medio de compra.

## Conclusiones fruto del análisis de los datos

Desde un punto de vista general se puede observar que existen una serie de conclusiones que si bien, pueden no ser del todo concluyentes, sí que nos pueden ayudar a entender como se están comportando los clientes y el impacto que están teniendo las campañas sobre ellos.

1. La segunda campaña ha tenido un impacto poco concluyente sobre los clientes dado que la captación sobre los mismos ha sido muy baja.
2. El público objetivo sobre el que la empresa ha podido incidir es sobre todo de un poder adquisitivo alto, con un nivel de educación elevado y que además, percibe el producto como de alto poder adquisitivo. Las clases más bajas no tienen demasiado interés en el producto que se ofrece.

Sería muy interesante el poder adaptar la estrategia de comunicación de la compañía, paquetizando la oferta de productos de tal modo que, los clientes que se muestran indiferentes pasarán a percibir el producto como asequible.

Los productos podrían ser ofrecidos en cantidades más pequeñas por un precio de venta final más bajo. Este tipo de estrategia, de ser efectiva tendría unos efectos positivos para la empresa que se traducirían en:

1. Más ventas con un margen comercial mejor.
2. La posibilidad de crear economías de escala al aprovechar mejor los canales de distribución.
3. La posibilidad de ir progresivamente fidelizando a clientes que todavía no han podido disfrutar de las características de los productos que ofrece la empresa.

## **Pestaña Recencia / Frecuencia**

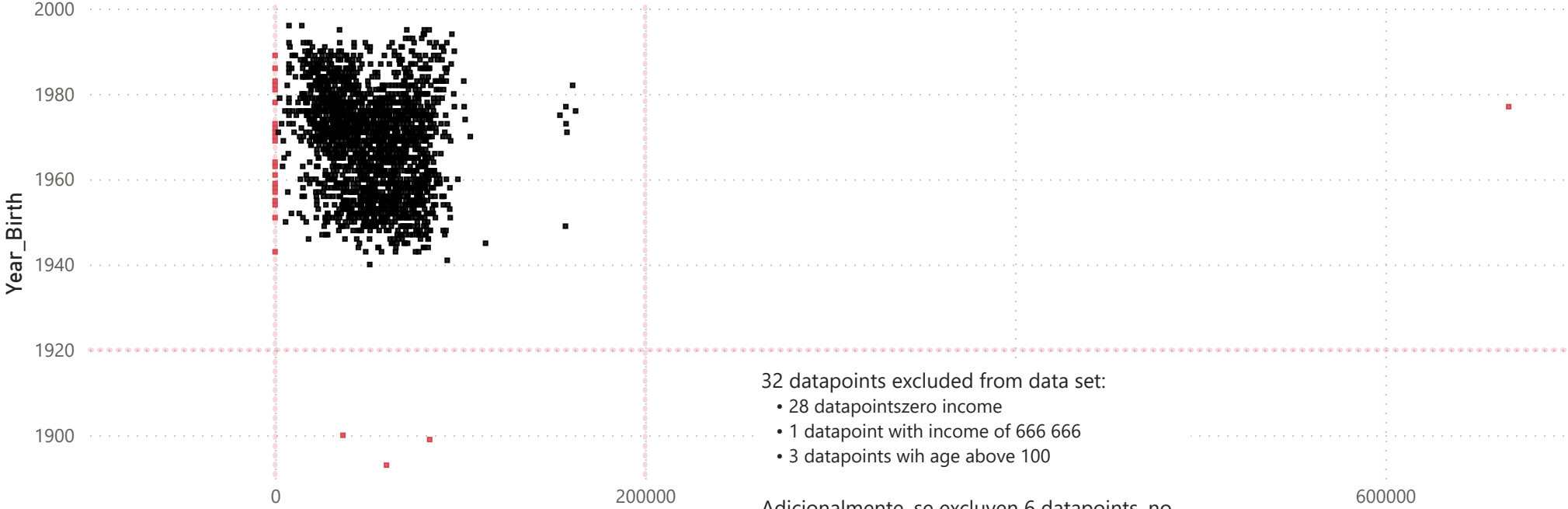
Se ha creado una hoja en la que se han puesto en relación la recencia (días desde la última compra) y la frecuencia de compra en dos ejes.

En esta hoja del modelo se han incorporado toda una serie de slicers para poder llegar a nivel de granularidad máxima y poder observar el comportamiento de esos dos ejes en función de la clase social, la educación, el estado civil, la generación a la que pertenecen los clientes y el grado de fidelización.

Como analista de datos considero que este tipo de información, más conductual y demográfica ayudará no solo al departamento de marketing a mejorar su estrategia de comunicación a los clientes actuales, sino también a los futuros.

Asimismo, ayudará a mejorar la oferta específica de productos del departamento comercial para que las operaciones que se cierren incrementen.

Income y Year\_Birth





Clase\_social

Todas

Education

Todas

Marital\_Status

Todas

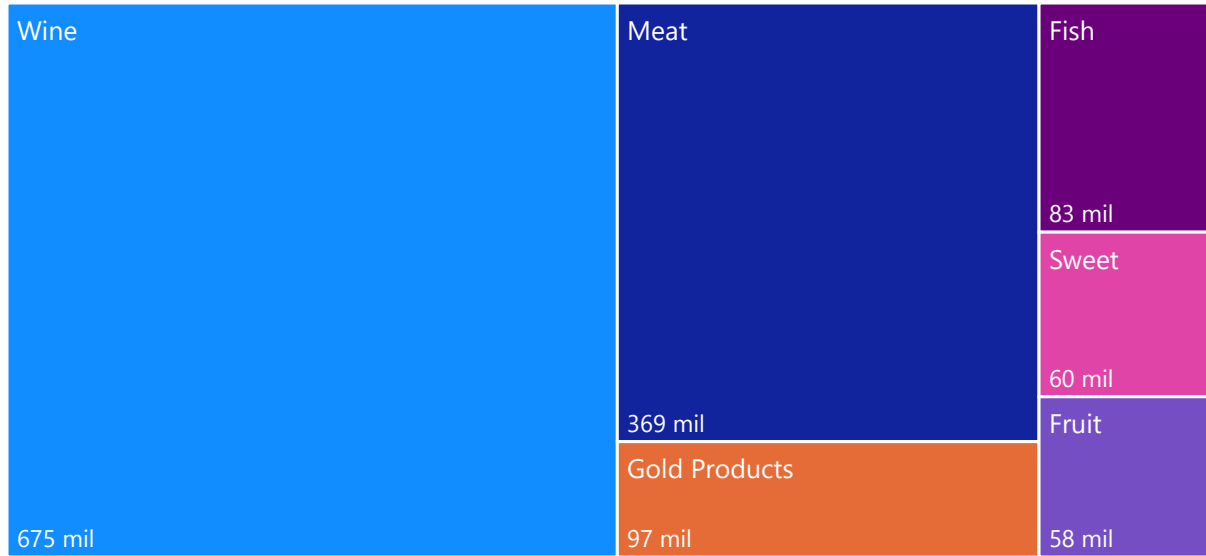
Generation

Todas

Fidelizacion

Todas

Ventas por categoría



Número quejas

20

Ventas Totales

1343235

clientes descontentos

0,000719

Tickets

28 mil

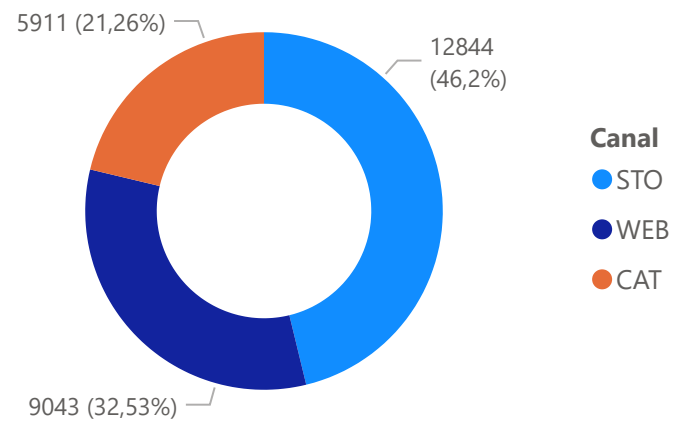
Ticket Medio

48,32

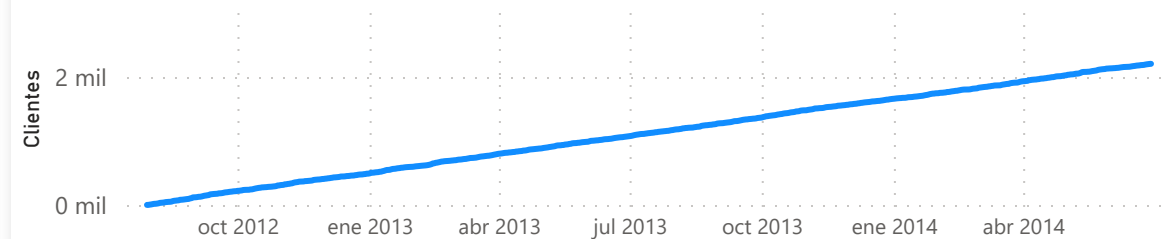
Tickets con descuento

0,18

Deals por canal



Evolución de clientes



Número de clientes

2206

Volumen de negocio

1343235

Recencia Media

49

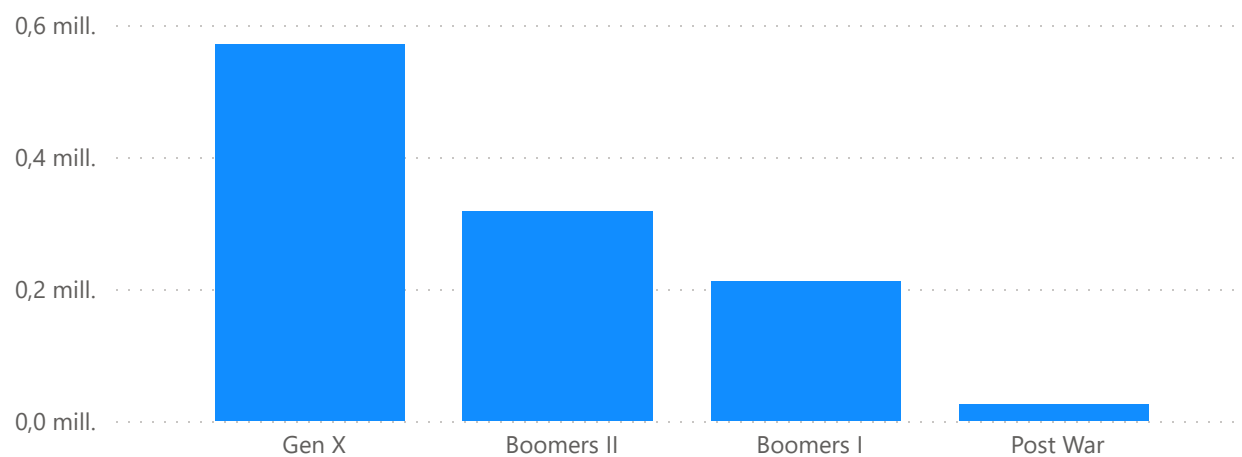
Tickets con descuento

0,18

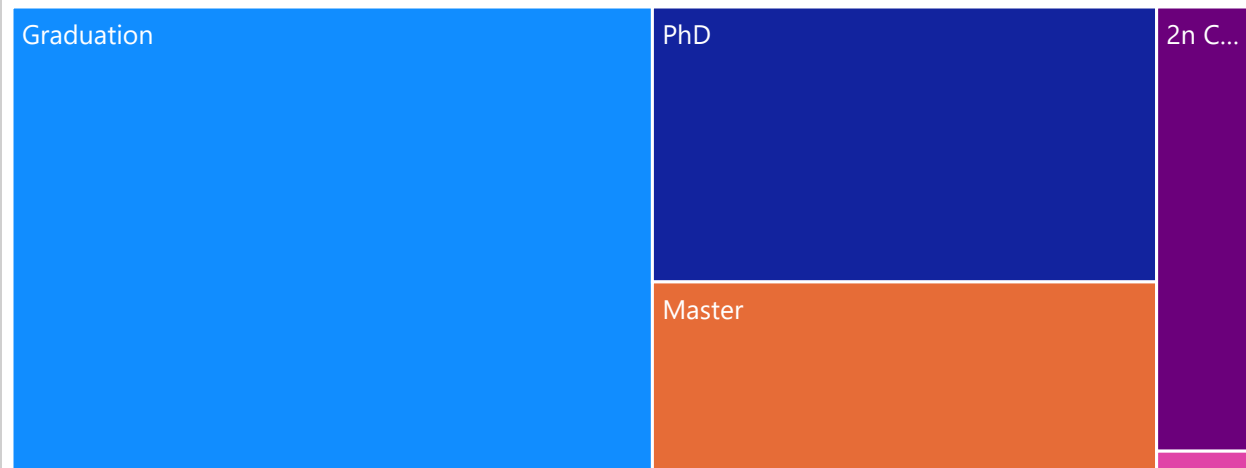
Ticket Medio

48,32

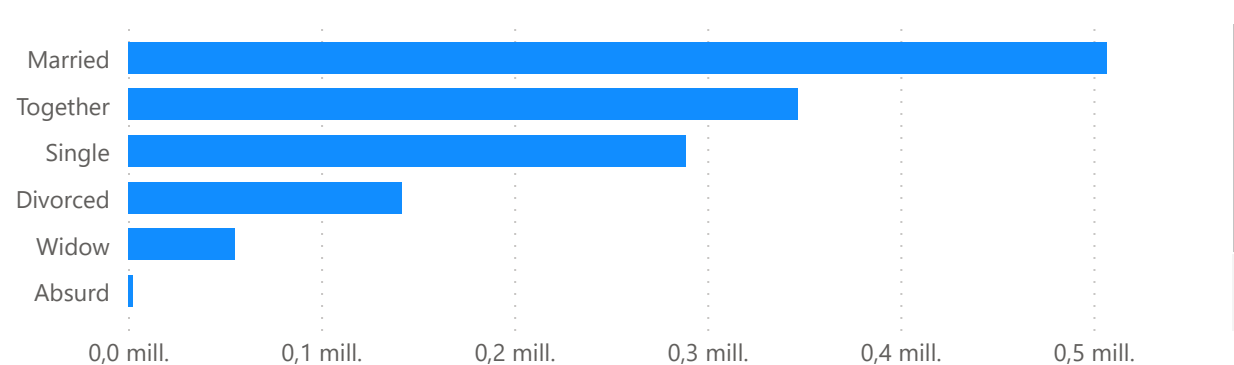
Ventas bianuales por generación



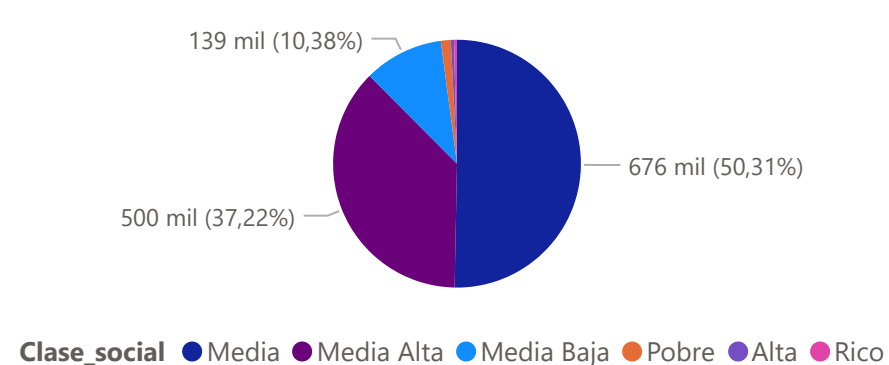
Ventas bianuales por educación



Ventas bianuales por estado civil



Ventas bianuales por clase social



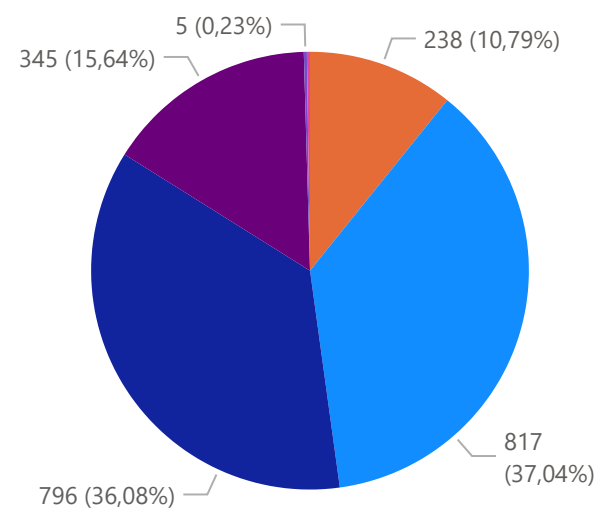
## Cientes y aceptación de campañas

142	30	163	164	161
Campaña 1	Campaña 2	Campaña 3	Campaña 4	Campaña 5

## Total Clientes Captados

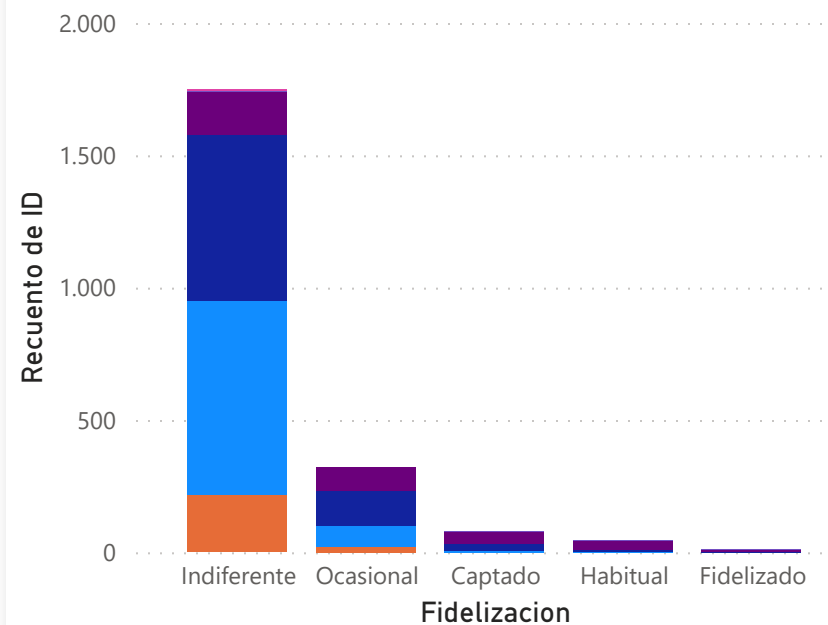
667

## Aceptadores de campañas por clase social



Clase\_social ● Pobre ● Media Baja ● Media ● Media Alta ● Alta

## Fidelización y clase social de los aceptadores



Clase\_social ● Pobre ● Media Baja ● Media ● Media Alta ● Alta

## Campañas aceptadas por un cliente

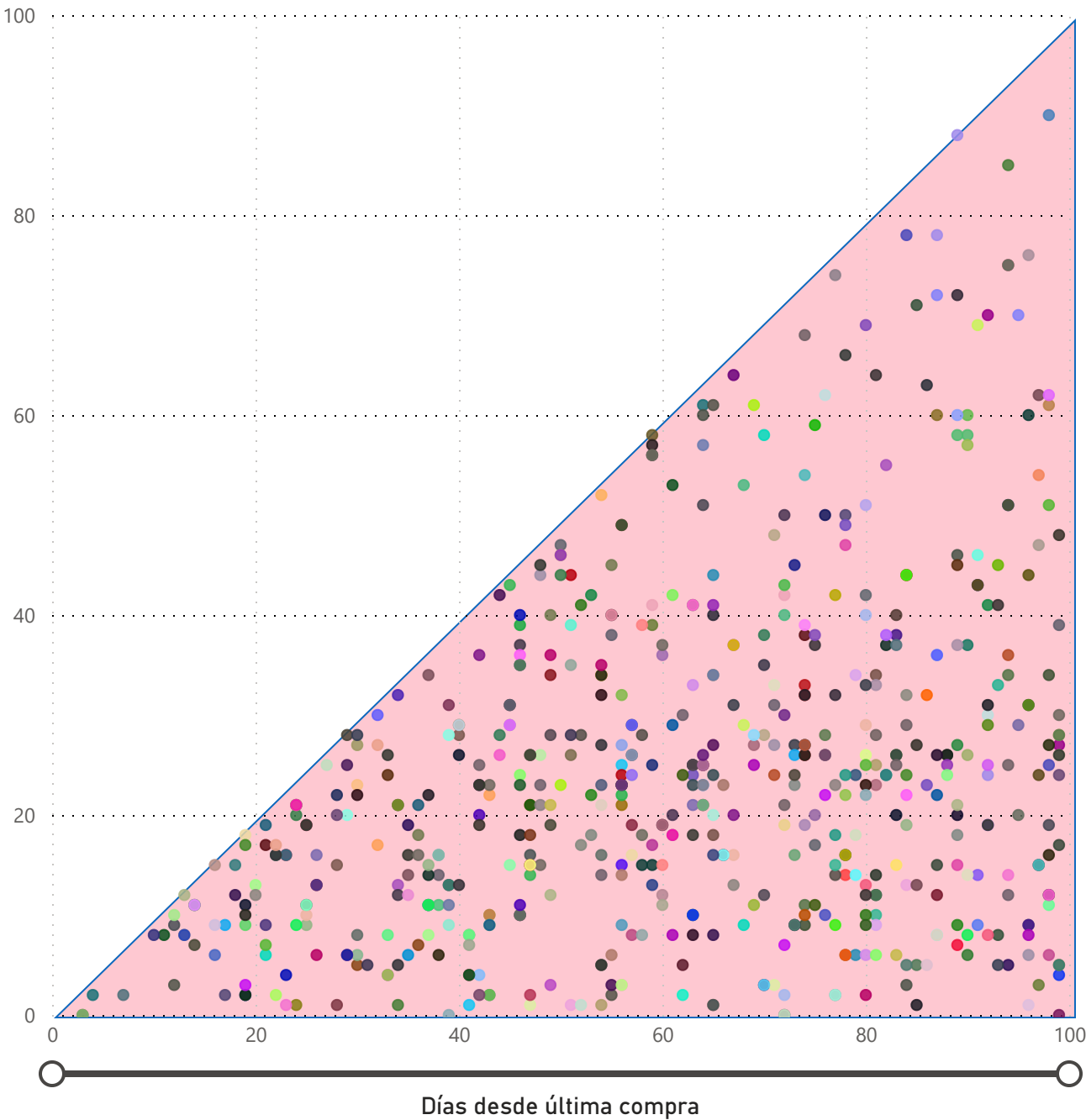
- ☐ 0
- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4





# Frecuencia vs Recencia (50 días) 1/Jul/2014

Días entre compra y compra



## Clase Social

- ☐ Rico
- ☐ Alta
- ☐ Media Alta
- ☐ Media
- ☐ Media Baja
- ☐ Pobre

## Generación

- ☐ WWII
- ☐ Post War
- ☒ Gen X
- ☐ Boomers II
- ☐ Boomers I

## Educación

- ☐ Basic
- ☐ 2n Cycle
- ☐ Graduation
- ☐ Master
- ☐ PhD

## Estado Civil

- ☐ Alone
- ☐ Single
- ☐ Together
- ☐ Married
- ☐ Divorced
- ☐ Widow
- ☐ YOLO
- ☐ Absurd

## Fidelización

- ☐ Indiferente
- ☐ Ocasional
- ☐ Captado
- ☐ Habitual
- ☐ Fidelizado

## Relacion códigos cliente con Recencia > Frecuencia

ID	Frecuencia	Recency
1371	0	72
2471	0	39
2831	0	99
4399	0	3
8093	0	72
9888	0	3
2225	1	54
2495	1	24

## Recencia > frecuencia

- ☐ Normal
- ☒ Reportar a ventas