



Caso 407b Online Retail, Analisis Cohortes

4 Avanzando en Dax. Proyectos para tu Portafolio de Analista de Datos

Jose Ignacio González Gómez

Departamento de Economía, Contabilidad y Finanzas - Universidad de La Laguna

www.jggomez.eu

V. 2.5

Ejercicio adaptado de: [Quick Cohort Analysis in PowerBI](#) - [Online Retail](#) - [UCI Machine Learning Repository](#)

Contenido

Resumen y objetivos	2
Presentacion – Información disponible	2
Se pide	3
FASE I: Proceso ETL e Informe I	3
Creación del fichero de trabajo y conexión al Dataframe. Ajustes y depuración básicos	3
Parámetro de conexión	3
Creacion de la consulta Clientes Unicos.....	3
Tabla calendario y de medidas. Modelo de datos v1.....	3
Informe I Análisis Cohorte matriz absoluto y porcentual	5
Medidas necesarias	7
FASE II: Informe II PDTE DE DESARROLLO	8
Orientación y notas	9
Proceso ETL	9
Creacion fichero de trabajo y parámetro de conexión	9
Creación de la consulta clientes unicos.....	9
Creacion de tabla calendario básica.....	10
Diseño del modelo de datos v1 – Modelados y relaciones.....	10
Fase I Informe Documentando el proyecto	10
Informe Analisis Cohorte Absoluto.	10
Medidas: INFO.VIEW.MEASURES().....	17
Bibliografía y conceptos técnicos relacionados	18
Análisis Cohorte, breve analisis.....	18
IA Crear tabla calendario.....	19

Resumen y objetivos

FASE I

Power Query. Creando parámetros de conexión. Creación de consulta de valores únicos. Modelo semántico. Crear tabla calendario. Crear tablas para organizar las medidas.

Funciones DAX aplicadas: CALCULATE, DATEDIFF, RELATED, DISTINCTCOUNT, DIVIDE, EPMONTH, FILTER, ISBLANK, NOT,

Presentacion – Información disponible

La empresa Online Retail domiciliada en el Reino Unido pertenece al sector comercial de venta directa al consumidor¹ a través de varios marketplace (Amazon, eBay, Etsy y otros) sin tener una tienda física tradicional contando con un amplio catálogo accesible en línea donde los clientes, en su mayoría mayoristas, pueden explorar productos, leer descripciones, ver imágenes y realizar los pedidos.

Cuenta con su propia logística de envío que implica la gestión del almacenamiento, empaquetado y envío de productos.

Disponemos de un DataFrame en formato Excel (Online Retail.xlsx) que contiene todas las transacciones ocurridas entre el 01/12/2010 y el 09/12/2011, tal y como se muestra en la siguiente ilustración.

	A	B	C	D	E	F	G	H
1	voiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
2	536365	85123A	WHITE HANGING HE	6	01/12/2010 8:26	2,55	17850	United Kingdom
3	536365	71053	WHITE METAL LANT	6	01/12/2010 8:26	3,39	17850	United Kingdom
4	536365	84406B	CREAM CUPID HEAF	8	01/12/2010 8:26	2,75	17850	United Kingdom
5	536365	84029G	KNITTED UNION FLA	6	01/12/2010 8:26	3,39	17850	United Kingdom
6	536365	84029E	RED WOOLLY HOTTI	6	01/12/2010 8:26	3,39	17850	United Kingdom

Ilustración 1

El significado de los campos o columnas es el siguiente.

Variable	Descripción	Observaciones
InvoiceNo	FacturaNo	Nominal, un número entero de 6 dígitos asignado de forma única a cada transacción. Si este código comienza con la letra 'c', indica una cancelación.
StockCode	Código Producto	Código artículo. Un número entero de 5 dígitos asignado de forma única a cada producto
Description	Descripción	Nombre del producto
Quantity	Cantidad	Las cantidades de cada producto (artículo) por transacción
InvoiceDate	Fecha de factura	Día y hora en que se generó cada transacción
UnitPrice	Precio unitario	Precio del producto por unidad en libras esterlinas.
CustomerID	ID de cliente	Código cliente, número de 5 dígitos asignado de forma única a cada cliente
Country	País	País donde reside cada cliente

¹ Adquiere productos (de fabricantes o mayoristas) y los ofrece al público

Se pide

FASE I: Proceso ETL e Informe I

Creación del fichero de trabajo y conexión al Dataframe. Ajustes y depuración básicos

Comenzamos por **analizar el Dataframe** con especial atención al contenido de las columnas y el tipo de información que contiene, detectando posibles relaciones y tipos entre las fuentes de datos. Si existieran varias tablas o fuentes de datos, categorizarlas entre tablas de Dimensión o de Hechos.

En esta primera fase es imprescindible llevar los ajustes necesarios en el editor de consultas (Power Query) como son el ajuste del tipo de columna, eliminar columna, verificar el tipo de datos y auditamos los datos importados asegurándonos que no hay errores, ni valores en blanco en las columnas clave.

Es importante confirmar y revisar la no existencias de errores una vez llevada a cabo el proceso de ajuste y depuración contando con las opciones de Vista en Power Query que nos permite un análisis detallado columna por columna.

Parámetro de conexión

Deberá disponer de un **parámetro de conexión al Dataset**.

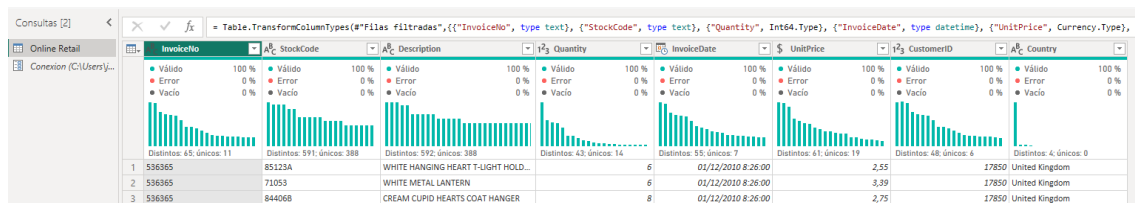


Ilustración 2

Creacion de la consulta Clientes Únicos

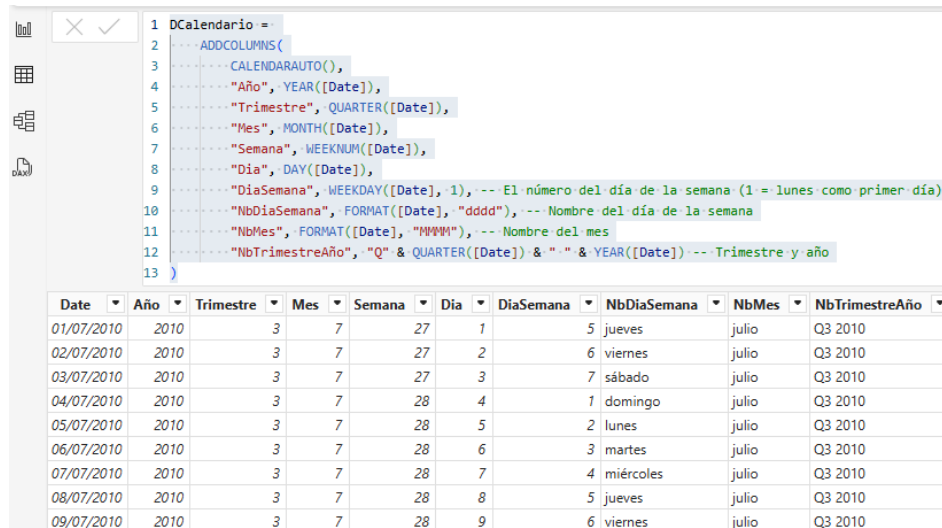
Necesitamos disponer de una nueva consulta que nos devuelva una tabla con los registros únicos de los clientes (CustomerID) y que evidentemente no puede tener vacíos. A esta consulta la llamaremos “Clientes Unicos”. Esta sera una tabla de dimensión

Tabla calendario y de medidas. Modelo de datos v1

Tabla DCalendario, para crear la tabla básica calendario lo adaptaremos de la propuesta disponible en el anexo.

Tabla de Medidas, también deberemos disponer (crear) de una tabla para almacenar las medidas y que llamaremos “**Medidas**”

Crear una tabla basica calendario con la función CALENDARAUTO y que contenga las siguientes columnas (ver Anexo, IA Crear tabla calendario.):



```

1 DCalendario =
2 ...ADDCOLUMNS(
3 ...CALENDAR(AUTO()),
4 ..."Año", YEAR([Date]),
5 ..."Trimestre", QUARTER([Date]),
6 ..."Mes", MONTH([Date]),
7 ..."Semana", WEEKNUM([Date]),
8 ..."Dia", DAY([Date]),
9 ..."DiaSemana", WEEKDAY([Date], 1), -- El número del día de la semana (1 = lunes como primer día)
10 ..."NbDiaSemana", FORMAT([Date], "dddd"), -- Nombre del día de la semana
11 ..."NbMes", FORMAT([Date], "MMM"), -- Nombre del mes
12 ..."NbTrimestreAño", "Q" & QUARTER([Date]) & " " & YEAR([Date]) -- Trimestre y año
13 )

```

Date	Año	Trimestre	Mes	Semana	Dia	DiaSemana	NbDiaSemana	NbMes	NbTrimestreAño
01/07/2010	2010	3	7	27	1	5	jueves	julio	Q3 2010
02/07/2010	2010	3	7	27	2	6	viernes	julio	Q3 2010
03/07/2010	2010	3	7	27	3	7	sábado	julio	Q3 2010
04/07/2010	2010	3	7	28	4	1	domingo	julio	Q3 2010
05/07/2010	2010	3	7	28	5	2	lunes	julio	Q3 2010
06/07/2010	2010	3	7	28	6	3	martes	julio	Q3 2010
07/07/2010	2010	3	7	28	7	4	miércoles	julio	Q3 2010
08/07/2010	2010	3	7	28	8	5	jueves	julio	Q3 2010
09/07/2010	2010	3	7	28	9	6	viernes	julio	Q3 2010

Ilustración 3

Para completar esta fase nos queda relacionar las tablas y diseñando el modelo de datos. Es posible que Power BI haya creado las relaciones automáticamente y por tanto es necesario revisarlas en cuanto puede contener algún error, en cualquier caso, es aconsejable tener desactivado esta opción de creación de relaciones automáticamente para lo cual exponemos unas notas relacionadas en el anexo sobre como desactivar esta opción.

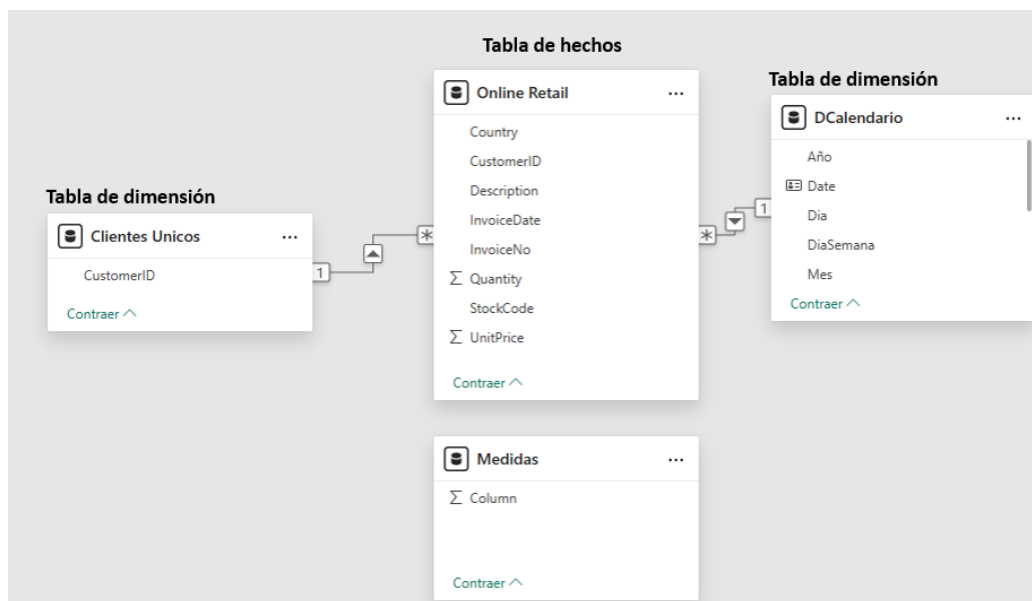


Ilustración 4

Nota. Señalar en este caso que para que la relación entre la tabla de hechos (Online Retail) y la tabla calendario (Date) opere correctamente, es necesario que los tipos de campos no solo sean iguales (tipo fecha) sino además el formato, en el sentido de que si el campo fecha InvoiceDate es FechaHora y el campo Date es solo Fecha, la relación podría no funcionar correctamente, por tanto, tienen que coincidir. Si quisiéramos disponer además de la hora para operar tendríamos que duplicar la columna solo para poder disponer de las horas.

Informe I Análisis Cohorte matriz absoluto y porcentual

Presentacion

El objetivo de este proyecto es el diseño de un informe en Power BI que permita llevar a cabo un analisis de Cohorte absoluto y porcentual tal y como se muestra en la siguiente ilustración.

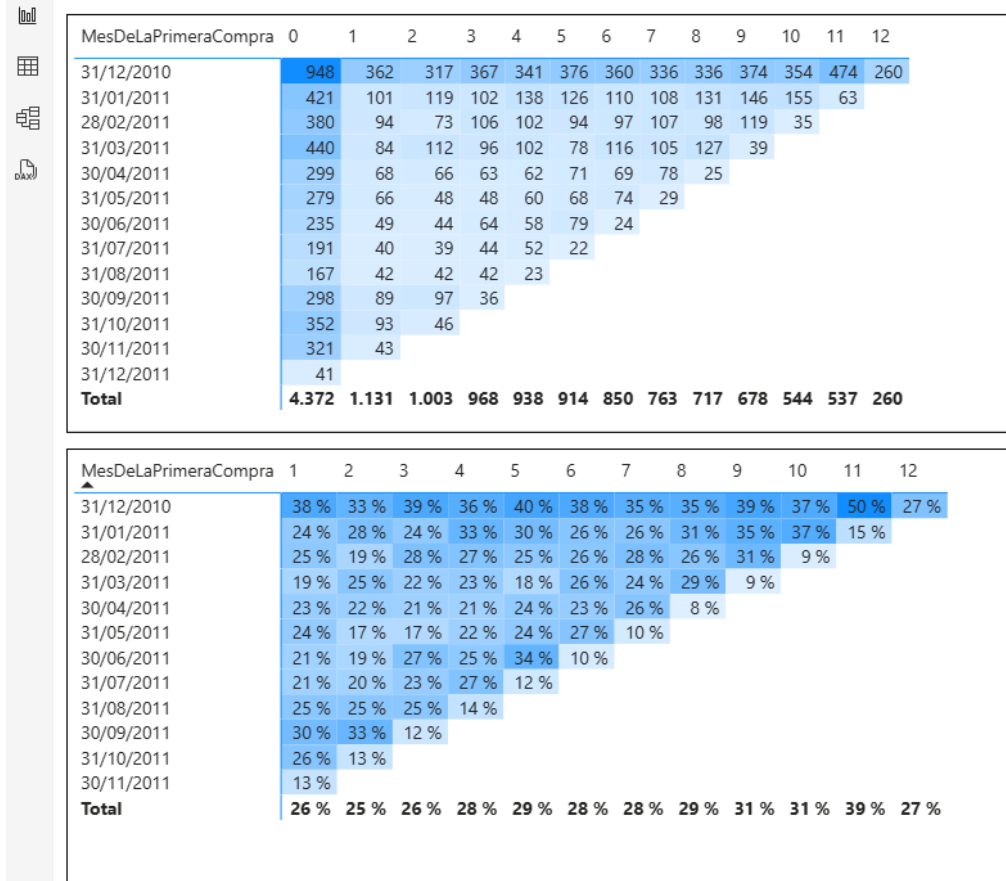


Ilustración 5

Para llevar a cabo este analisis es necesario crear una serie de medidas que a traves del objeto visual matriz nos permita configurar las dos visualizaciones. Estas muestras la evolución de los clientes que han permanecido a lo largo del tiempo desde que han realizado la primera compra. Asi para el caso de aquellos clientes que compraron por primera vez en diciembre de 2010 (948) han permanecido comprado durante 12 meses 260 clientes e igualmente han permanecido comprado durante 6 meses 360 clientes o lo que es proporcionalmente el 27% de los clientes que compraron en diciembre de 2010 seguían comprando o permanencia despues de 12 meses e igualmente a los seis meses permanencias el 38%

Como orientación presentamos a continuacion una notas y orientaciones necesarias.

Columnas calculadas y medidas básicas

Destacar que para este analisis es necesario crear una serie de columnas calculadas y medidas como son las siguientes:

Columnas calculadas:

- En la tabla calendario es necesario identificar el final de mes para lo cual crearemos una nueva columna que llamaremos **FinDeMes**.

- En la tabla Clientes Unicos es necesario determinar el mes de la primera compra de cada cliente **MesDeLaPrimeraCompra**.

Medidas necesarias

- #NºClientesQueHanComprado**, la primera medida necesaria será determinar el numero de clientes que nos han comprado (Nº de clientes que siguieron comprando) y esta medida se creara en la tabla de Medidas.

Matriz Cohorte absoluto v.1

Por tanto, el primer paso será obtener el siguiente objeto visual basado en una matriz, tal y como mostramos a continuación

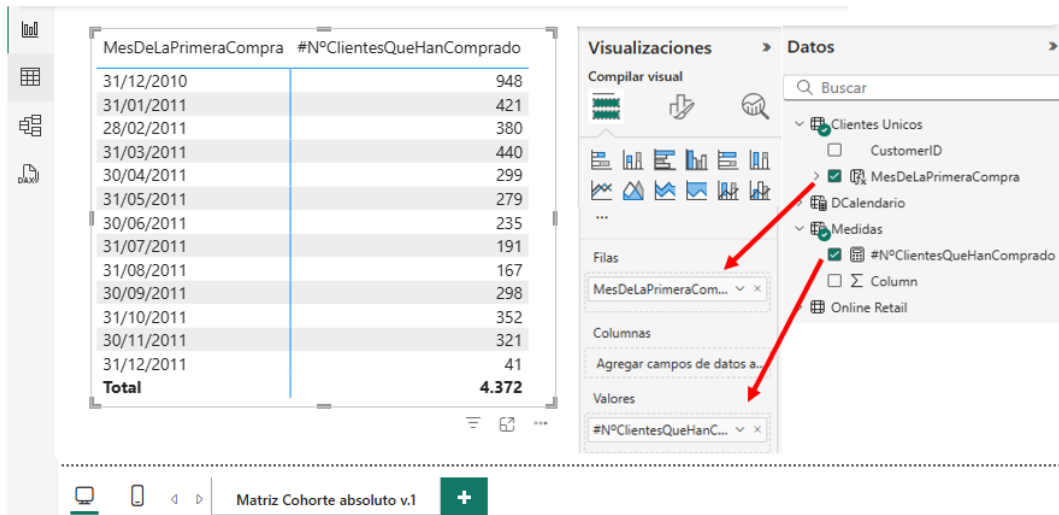


Ilustración 6

Para seguir avanzando necesitamos conocer el número de meses que han transcurrido desde el primer pedido y para ello crearemos una columna calculada en la tabla de hechos Online Retail.

Columnas calculadas:

- En la tabla de hechos, creamos la columna calculada con el nombre “NºdeMesesDespuesDelºCompra”.

Incorporando esto como columna a la ilustración anterior el resultado es el siguiente:

MesDeLaPrimeraCompra	0	1	2	3	4	5	6	7	8	9	10	11	12	Total
31/12/2010	948	362	317	367	341	376	360	336	336	374	354	474	260	948
31/01/2011	421	101	119	102	138	126	110	108	131	146	155	63		421
28/02/2011	380	94	73	106	102	94	97	107	98	119	35			380
31/03/2011	440	84	112	96	102	78	116	105	127	39				440
30/04/2011	299	68	66	63	62	71	69	78	25					299
31/05/2011	279	66	48	48	60	68	74	29						279
30/06/2011	235	49	44	64	58	79	24							235
31/07/2011	191	40	39	44	52	22								191
31/08/2011	167	42	42	42	23									167
30/09/2011	298	89	97	36										298
31/10/2011	352	93	46											352
30/11/2011	321	43												321
31/12/2011	41													41
Total	4.372	1.131	1.003	968	938	914	850	763	717	678	544	537	260	4.372

Ilustración 7

Matriz Cohorte absoluto v.1 con formato condicional

Solo nos quedaría quitar el total de la columna de la derecha y dar formato condicional, para que el resultado final sea el siguiente:

MesDeLaPrimeraCompra	0	1	2	3	4	5	6	7	8	9	10	11	12
31/12/2010	948	362	317	367	341	376	360	336	336	374	354	474	260
31/01/2011	421	101	119	102	138	126	110	108	131	146	155	63	
28/02/2011	380	94	73	106	102	94	97	107	98	119	35		
31/03/2011	440	84	112	96	102	78	116	105	127	39			
30/04/2011	299	68	66	63	62	71	69	78	25				
31/05/2011	279	66	48	48	60	68	74	29					
30/06/2011	235	49	44	64	58	79	24						
31/07/2011	191	40	39	44	52	22							
31/08/2011	167	42	42	42	23								
30/09/2011	298	89	97	36									
31/10/2011	352	93	46										
30/11/2011	321	43											
31/12/2011	41												
Total	4.372	1.131	1.003	968	938	914	850	763	717	678	544	537	260

Ilustración 8

Matriz Cohorte relativo, porcentual

La interpretación del análisis de cohortes mejora significativamente cuando los valores se representan en términos porcentuales, para lo cual es necesario crear una nueva medida.

Para ello duplicamos la matriz de la Ilustración 8 y sobre la misma aplicamos la nueva medida para lo cual obtendremos como resultado final el siguiente:

MesDeLaPrimeraCompra	1	2	3	4	5	6	7	8	9	10	11	12
31/12/2010	38,19 %	33,44 %	38,71 %	35,97 %	39,66 %	37,97 %	35,44 %	35,44 %	39,45 %	37,34 %	50,00 %	27,43 %
31/01/2011	23,99 %	28,27 %	24,23 %	32,78 %	29,93 %	26,13 %	25,65 %	31,12 %	34,68 %	36,82 %	14,96 %	
28/02/2011	24,74 %	19,21 %	27,89 %	26,84 %	24,74 %	25,53 %	28,16 %	25,79 %	31,32 %	9,21 %		
31/03/2011	19,09 %	25,45 %	21,82 %	23,18 %	17,73 %	26,36 %	23,86 %	28,86 %	8,86 %			
30/04/2011	22,74 %	22,07 %	21,07 %	20,74 %	23,75 %	23,08 %	26,09 %	8,36 %				
31/05/2011	23,66 %	17,20 %	17,20 %	21,51 %	24,37 %	26,52 %	10,39 %					
30/06/2011	20,85 %	18,72 %	27,23 %	24,68 %	33,62 %	10,21 %						
31/07/2011	20,94 %	20,42 %	23,04 %	27,23 %	11,52 %							
31/08/2011	25,15 %	25,15 %	25,15 %	13,77 %								
30/09/2011	29,87 %	32,55 %	12,08 %									
31/10/2011	26,42 %	13,07 %										
30/11/2011	13,40 %											
Total	25,87 %	22,94 %	22,14 %	21,45 %	20,91 %	19,44 %	17,45 %	16,40 %	15,51 %	12,44 %	12,28 %	5,95 %

Ilustración 9

Medidas y columnas necesarias

Las medidas y columnas calculadas necesarias a crear son las mostradas en la siguiente tabla.

Medidas	Tabla	Comentario
#NºClientesQueHanComprado	Medidas	Nº de clientes únicos que han comprado
#Cohorte	Medidas Fase I	Grupos específicos de usuarios (cohortes) a lo largo del tiempo, por ejemplo, que nos han comprado en el intervalo de un mes, dos meses, tres meses.....

Columnas calculadas

Columna	Tabla	
NºdeMesesDespuesDelºCompra	Online Retail	Determina en nº de meses transcurridos desde la primera compra del cliente.

MesDeLaPrimeraCompra	Cientes Unicos	Determina el mes de la primera compra.
FinDeMes	Dcalendario	Determina a que fecha corresponde el final de mes.

Funciones a aplicar: CALCULATE, DATEDIFF, RELATED, DISTINCTCOUNT, DIVIDE, EOMONTH, FILTER, ISBLANK, NOT,

FASE II: Informe II PDTE DE DESARROLLO

Pendiente de desarrollo, en este caso se pedirá realizar un analisis de compras según franja horaria, para lo cual debeermos duplicar en el PQ la columna de fecha y hora de venta, que no se ve por que cambiamos el formato a solo fecha y ouycltando hora para poder crear los conexiones en el modelo de daots, por eso debemos duplicarla para que con ella mostrar las franjas horarias con columna condicciona y para trabajar con fines de semansa..

Es decir vamos a trabajar con tiempos horarios, por ejemplo franja horaria

Tambien podríamos aplicar segmentadores y filñtrps---

Orientación y notas

Proceso ETL

Creacion fichero de trabajo y parámetro de conexión

Comenzamos creando nuestro fichero de trabajo y a través del editor de consultas creamos nuestro primeros ajustes.

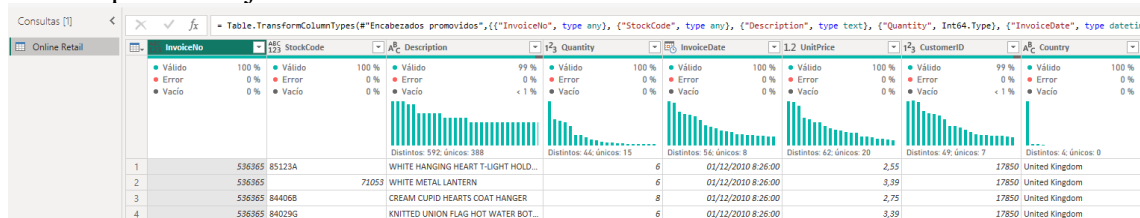


Ilustración 10

Una vez disponible el fichero de trabajo creamos el parámetro de conexión.

Administrar parámetros

Nuevo

A^B_C Conexion

X

Nombre

Conexion

Descripción

☒ Requerido

Tipo

Cualquiera

Valores sugeridos

Cualquier valor

Valor actual

C:\Users\joseignacio\cod\Desktop\407a Teoria Cohort Analysis

C:\Users\joseignacioicod\Desktop\407a Teoria Cohort Analysis\Caso Power BI\407b Online Retail\Mi adaptación\Online Retail.xlsx

Ilustración 11

Creación de la consulta clientes únicos

Para crear esta consulta, procedemos a duplicar la anterior y cambiarle el nombre por el de Clientes Únicos. De esta forma disponemos de todos los valores y el parámetro de conexión. Ahora seleccionamos la columna que nos interesa CustomerID que es el código del cliente y seleccionamos quitar duplicados, de esta forma hemos generado una nueva tabla con el código de los clientes únicos disponibles.

Debemos asegurarnos que no contiene vacíos pq entonces no podríamos relacionar las tablas correspondientes.

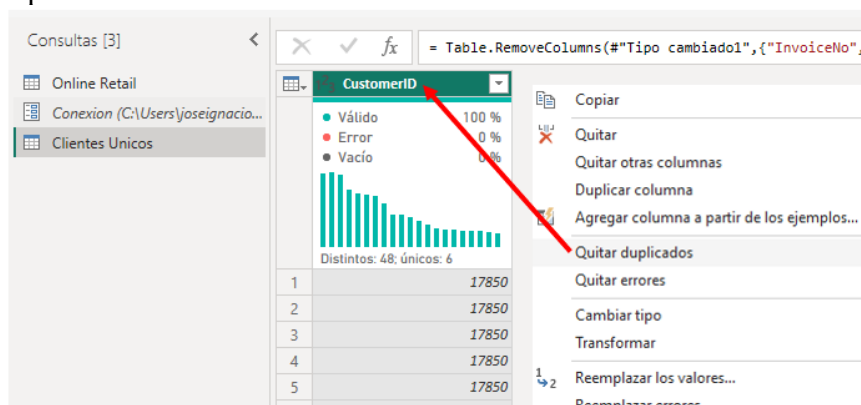


Ilustración 12

Creacion de tabla calendario básica

Esta tabla se llamara DCalendario y aplicamos el codigo disponible en el anexo. Debemos recordar marcada como tabla de fechas.

No debemos olvidarnos de esta tabla marcarla como tabla de fechas.

Diseño del modelo de datos v1 – Modelados y relaciones.

Una vez realizadas las conexiones al Dataframe y disponiendo del calendario asi como de la tabla Medidas, disponemos en el modelo de datos de todos estos elementos sin relacionar. Por tanto, para finalizar el proceso solo nos queda establecer las conexiones correspondientes. Ver la nota relacionada con la Ilustración 4 relacionada con el formato de fecha de las dos tablas.

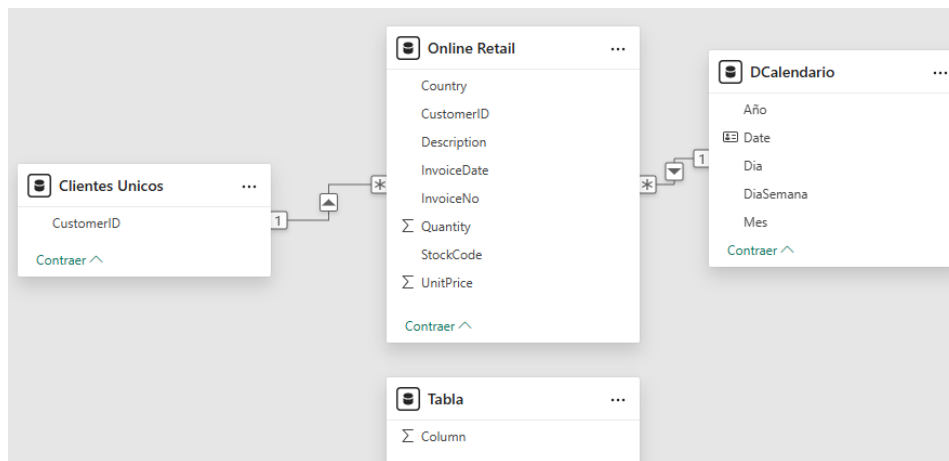


Ilustración 13

Modelado – Administrar relaciones INFO.VIEW.RELATIONSHIPS()

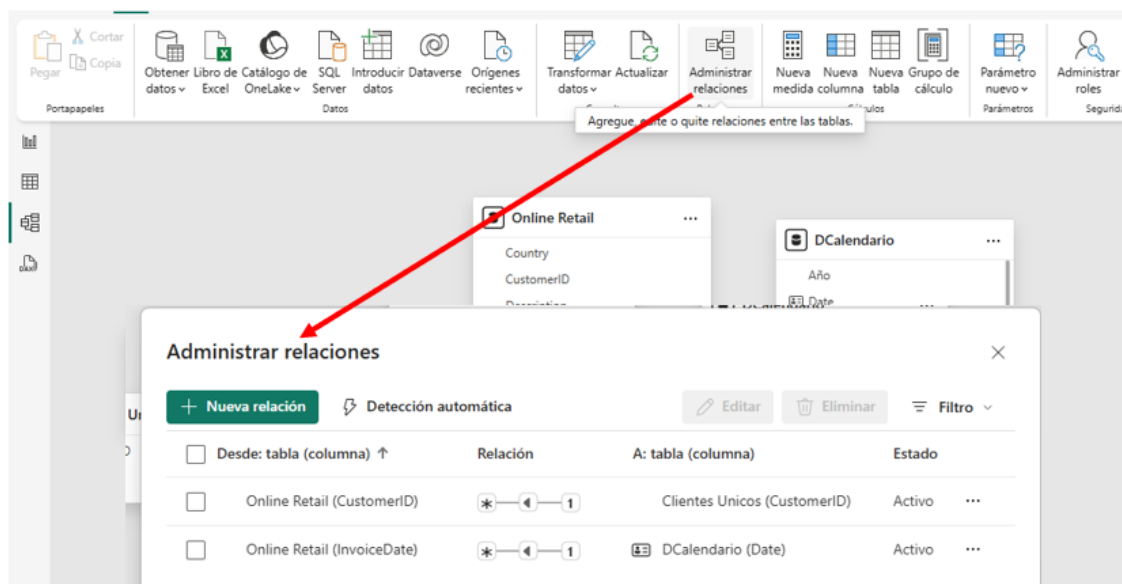


Ilustración 14

Fase I Informe Documentando el proyecto

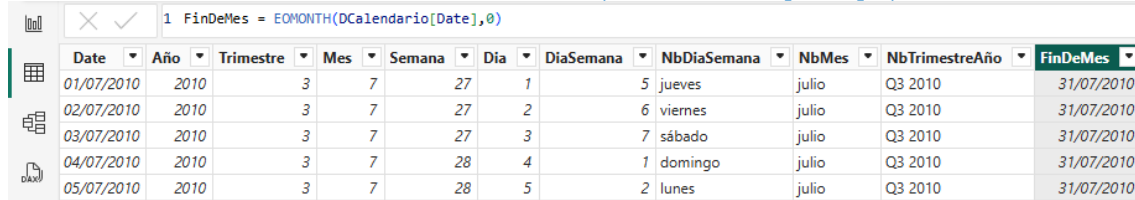
Informe Analisis Cohorte Absoluto.

Para elaborar el primer objeto visual es necesario llevar a cabo una serie de pasos previos relacionados con columnas calculadas y medidas.

Columna calculada FinDeMes, Tabla Calendario

Para elaborar el primer objeto visual es necesario llevar a cabo una serie de pasos previos, comenzamos creando en la tabla Dcalendario la columna que llamaremos FinDeMes y que nos devolverá el fin de mes correspondiente a cada fecha del calendario

$$\text{FinDeMes} = \text{EOMONTH}(\text{Dcalendario}[\text{Date}], 0)$$



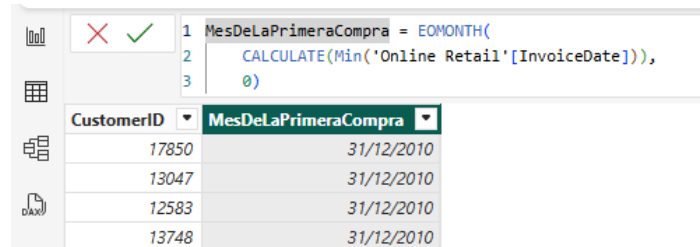
The screenshot shows the Power BI formula bar with the formula: `1 FinDeMes = EOMONTH(Dcalendario[Date],0)`. Below it, a table view of the 'Dcalendario' table is shown with columns: Date, Año, Trimestre, Mes, Semana, Día, DíaSemana, NbDíaSemana, NbMes, NbTrimestreAño, and FinDeMes. The data rows show dates from 01/07/2010 to 05/07/2010, all corresponding to July 2010, with the FinDeMes column consistently showing 31/07/2010.

Date	Año	Trimestre	Mes	Semana	Día	DíaSemana	NbDíaSemana	NbMes	NbTrimestreAño	FinDeMes
01/07/2010	2010	3	7	27	1	5	jueves	julio	Q3 2010	31/07/2010
02/07/2010	2010	3	7	27	2	6	viernes	julio	Q3 2010	31/07/2010
03/07/2010	2010	3	7	27	3	7	sábado	julio	Q3 2010	31/07/2010
04/07/2010	2010	3	7	28	4	1	domingo	julio	Q3 2010	31/07/2010
05/07/2010	2010	3	7	28	5	2	lunes	julio	Q3 2010	31/07/2010

Ilustración 15

Columna calculada MesDeLaPrimeraCompra, Tabla Clientes Unicos

En este caso queremos determinar el mes de la primera compra de cada cliente y para ello definimos en la tabla Clientes Unicos la Columna MesDeLaPrimeraCompra.



The screenshot shows the Power BI formula bar with the formula: `1 MesDeLaPrimeraCompra = EOMONTH(2 CALCULATE(Min('Online Retail'[InvoiceDate])),3 0)`. Below it, a table view of the 'Clientes Unicos' table is shown with columns: CustomerID and MesDeLaPrimeraCompra. The data rows show CustomerIDs 17850, 13047, 12583, and 13748, all with a MesDeLaPrimeraCompra of 31/12/2010.

CustomerID	MesDeLaPrimeraCompra
17850	31/12/2010
13047	31/12/2010
12583	31/12/2010
13748	31/12/2010

Ilustración 16

$$\text{MesDeLaPrimeraCompra} = \text{EOMONTH}(\text{CALCULATE}(\text{Min}(\text{'Online Retail'}[\text{InvoiceDate}]]), 0)$$

Medida Base #NºClientesQueHanComprado

En este caso determinamos el nº total de clientes que nos han comprado, es decir Nº de clientes que siguieron comprando y esta medida se creara en la tabla de Medidas.

$$\begin{aligned} \# \text{NºClientesQueHanComprado} = & \text{CALCULATE} \\ & \text{DISTINCTCOUNT}(\text{'Online Retail'}[\text{CustomerID}]), \\ & \text{NOT}(\text{ISBLANK}(\text{'Online Retail'}[\text{CustomerID}])) \end{aligned}$$

Esta fórmula calcula el nº de clientes únicos que han realizado una compra, ignorando los registros donde el campo CustomerID está en blanco. Explicación paso a paso:

- `CALCULATE(...)`: Esta función cambia el contexto de evaluación para aplicar filtros antes de hacer un cálculo. Es como decir: "Calcula esto, pero solo bajo ciertas condiciones".
- `DISTINCTCOUNT('Online Retail'[CustomerID])`: Cuenta cuántos clientes distintos hay en la columna CustomerID.
- `NOT(ISBLANK('Online Retail'[CustomerID]))`: Aplica un filtro para excluir los registros donde el CustomerID está en blanco. Es decir, solo cuenta los clientes que realmente están identificados.

Matriz Cohorte absoluto v.1

El primer paso en la configuración del objeto visual es la mostrada en la Ilustración 6, pero toca

MesDeLaPrimeraCompra	#NºClientesQueHanComprado
31/12/2010	948
31/01/2011	421
28/02/2011	380
31/03/2011	440
30/04/2011	299
31/05/2011	279
30/06/2011	235
31/07/2011	191
31/08/2011	167
30/09/2011	298
31/10/2011	352
30/11/2011	321
31/12/2011	41
Total	4372

Ilustración 17

Como hemos comentado para ampliar la Ilustración anterior necesitamos disponer de una nueva columna calculada “NºdeMesesDespuesDelºCompra” y al hacer columna calculada nos permitirá filtrar por su contenido.

Columna calculada “NºdeMesesDespuesDelºCompra”.

- En la tabla de hechos, creamos la columna calculada con el nombre “NºdeMesesDespuesDelºCompra”.

La formulación propuesta para esta columna calculada es:

NºdeMesesDespuesDelºCompra =

```
DATEDIFF(
    RELATED('Clientes Unicos'[MesDeLaPrimeraCompra]),
    Related(DCalendario[FinDeMes]),MONTH
)
```

Explicamos brevemente esta fórmula.

1 Objetivo del cálculo

Calcular **cuántos meses han pasado desde la primera compra de un cliente hasta la fecha de cada venta registrada en la tabla de hechos Online Retail.**

2 Desglose del cálculo

NºdeMesesDespuesDelºCompra =

```
DATEDIFF(
    RELATED('Clientes Unicos'[MesDeLaPrimeraCompra]),
    RELATED(DCalendario[FinDeMes]), MONTH
)
```

a) RELATED('Clientes Unicos'[MesDeLaPrimeraCompra])

- Esta función busca el campo [MesDeLaPrimeraCompra] en la tabla Clientes Unicos.
- Utiliza la relación entre Online Retail y Clientes Unicos (a través del código de cliente).
- Devuelve el **mes de la primera compra** del cliente para cada fila de la tabla Online Retail.

b) RELATED(DCalendario[FinDeMes])

- Busca el campo [FinDeMes] en la tabla de calendario DCalendario.
- Usa la relación entre Online Retail y DCalendario (a través del campo de fecha).
- Devuelve la **fecha de fin de mes** correspondiente a la fecha de cada venta.

c) DATEDIFF(..., ..., MONTH)

- Calcula la **diferencia en meses** entre los dos valores anteriores.
- La diferencia es entre el mes de la primera compra del cliente y la fecha de venta (con formato de fin de mes).
- El resultado es un número entero que representa **cuántos meses han pasado desde que el cliente hizo su primera compra**.

3 ¿Por qué usar FinDeMes?

Usar FinDeMes en vez de la fecha directamente puede estandarizar la medición, asegurando que todas las diferencias se calculen hasta el final del mes, y no entre fechas específicas (por ejemplo, del 3 al 27).

4 Resultado esperado

Para cada fila en Online Retail, la nueva columna N°deMesesDespuesDel°Compra mostrará un número como:

- 0 si fue la primera compra (en el mismo mes).
- 1, 2, 3... si fue después de 1, 2, 3... meses desde la primera compra.

De esta forma disponemos finalmente de la matriz de cohorte absoluto incorporando esta última columna calculada

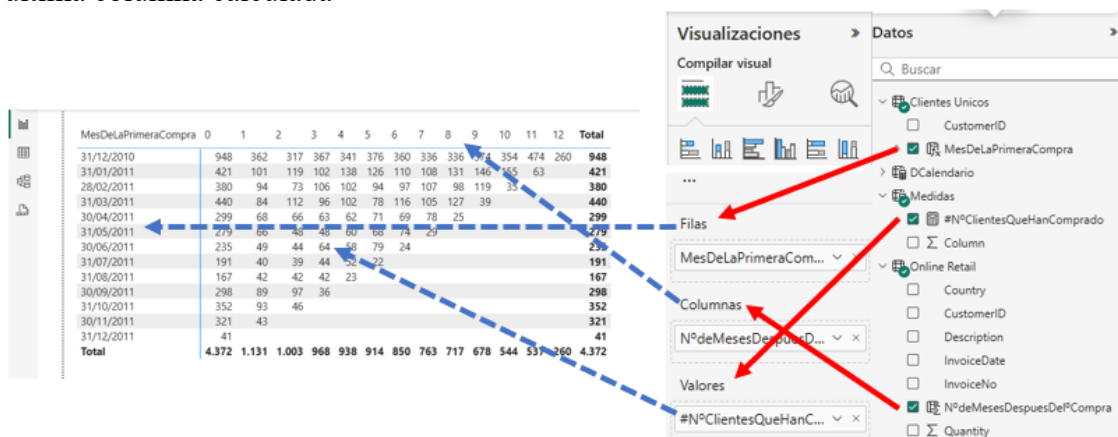


Ilustración 18

Matriz Cohorte absoluto v.1 con formato condicional

Vamos a continuación aplicar dos ajustes: eliminación del total de columna y aplicar formato condicionales para los valores.

MesDeLaPrimeraCompra	0	1	2	3	4	5	6	7	8	9	10	11	12	Total
31/12/2010	948	362	317	367	341	376	360	336	336	374	354	474	260	948
31/01/2011	421	101	119	102	138	126	110	108	131	146	155	63		421
28/02/2011	380	94	73	106	102	94	97	107	98	119	35			380
31/03/2011	440	84	112	96	102	78	116	105	127	39				440
30/04/2011	299	68	66	63	62	71	69	78	25					299
31/05/2011	279	66	48	48	60	68	74	29						279
30/06/2011	235	49	44	64	58	79	24							235
31/07/2011	191	40	39	44	52	22								191
31/08/2011	167	42	42	42	23									167
30/09/2011	298	89	97	36										298
31/10/2011	352	93	46											352
30/11/2011	321	43												321
31/12/2011	41													41
Total	4372	1131	1003	968	938	914	850	763	717	678	544	537	260	4372

Ilustración 19

1. Eliminación del total de columna

Como la columna total no interesa, la quitamos para lo cual accedemos a las propiedades y desactivamos Subtotales de columnas

2. Formato condicional para los valores

También nos interesa dar un formato condicional a los valores del tipo escala de color, esto permite aplicar estilos visuales a los datos basándose en sus valores. Esto ayuda a resaltar información importante y a mejorar la interpretación de los datos.

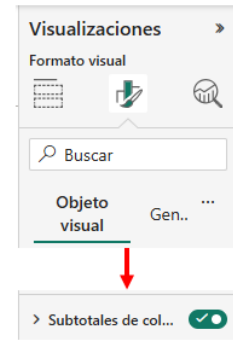


Ilustración 20

Para aplicar el formato condicional podremos hacerlo de varias formas, accedemos al panel de visualizaciones, haciendo clic derecho o seleccionando la flecha hacia abajo junto al campo que se desea formatear y eligiendo "Formato condicional" Color de fondo, tal y como se presenta en la Ilustración 21

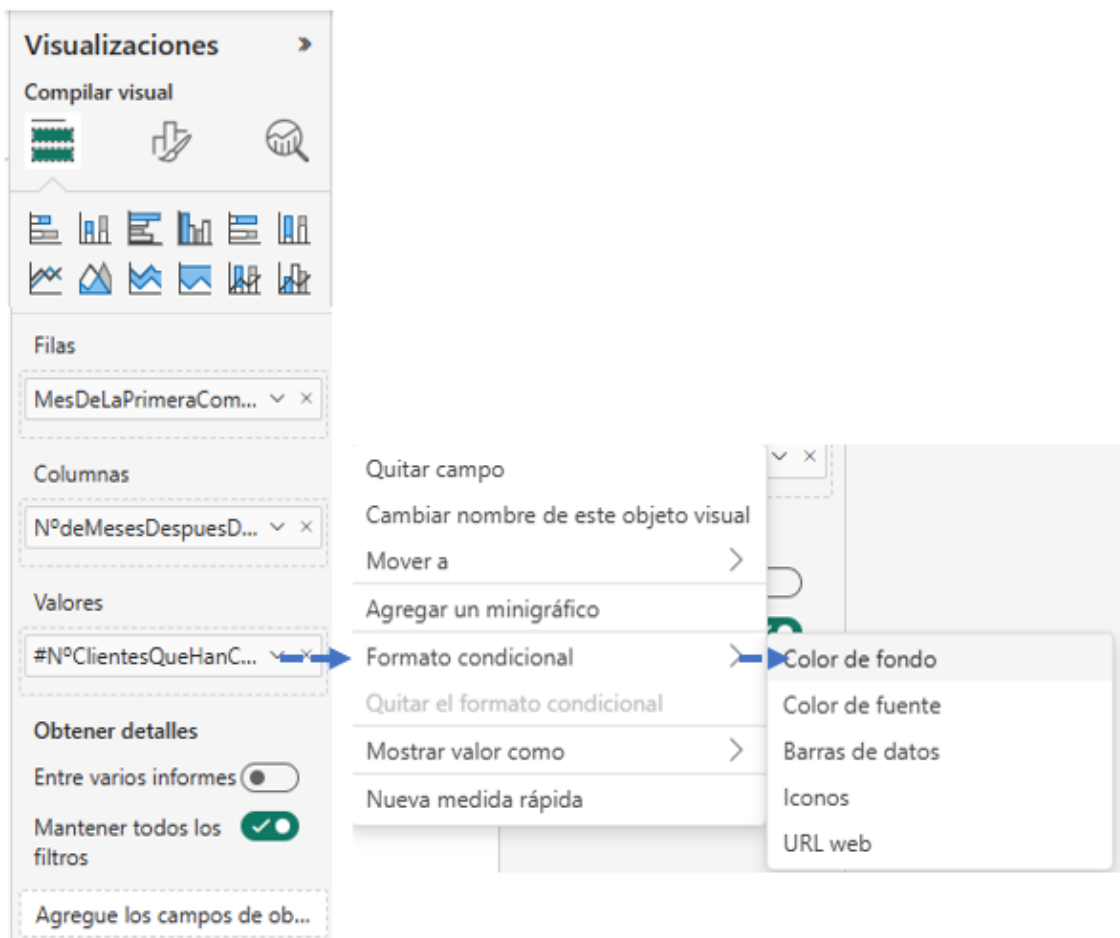


Ilustración 21

Ello no da acceso a un cuadro que permite configurar el color de fondo de las celdas según sus valores, utilizando escalas de color o reglas personalizadas.

Color de fondo - #NºClientesQueHanComprado ✕

Estilo de formato: Degradado Aplicar a: Solo valores

¿En qué campo debemos basar esto? #NºClientesQueHanComprado

¿Cómo se deben dar formato a los valores vacíos? Como cero

Mínimo: Valor más bajo □

Máximo: Valor más alto ■

Escribir un valor Escribir un valor

☐ Agregar un color medio

Aceptar Cancelar

Ilustración 22

El resultado final de aplicarlo serie:

MesDeLaPrimeraCompra	0	1	2	3	4	5	6	7	8	9	10	11	12
31/12/2010	948	362	317	367	341	376	360	336	336	374	354	474	260
31/01/2011	421	101	119	102	138	126	110	108	131	146	155	63	
28/02/2011	380	94	73	106	102	94	97	107	98	119	35		
31/03/2011	440	84	112	96	102	78	116	105	127	39			
30/04/2011	299	68	66	63	62	71	69	78	25				
31/05/2011	279	66	48	48	60	68	74	29					
30/06/2011	235	49	44	64	58	79	24						
31/07/2011	191	40	39	44	52	22							
31/08/2011	167	42	42	42	23								
30/09/2011	298	89	97	36									
31/10/2011	352	93	46										
30/11/2011	321	43											
31/12/2011	41												
Total	4.372	1.131	1.003	968	938	914	850	763	717	678	544	537	260

Ilustración 23

Matriz Cohorte relativo, porcentual

Para convertir los valores de la matriz anterior a porcentuales es necesario crear una nueva medida, pero comencemos duplicando el objeto visual anterior en una nueva página.

Medida #NºClientesDistintos

Comenzamos creando la medida cliente distintos que nos devolverá para una determinada fecha el número de clientes distintos

#NºClientesDistintos = DISTINCTCOUNT('Online Retail'[CustomerID])

Explicación paso a paso

- **DISTINCTCOUNT(...)**: Esta función cuenta el número de valores únicos (distintos) en una columna.
- **'Online Retail'[CustomerID]**: Es la columna que contiene los identificadores de los clientes en la tabla llamada 'Online Retail'.

¿Qué devuelve esta medida?

Devuelve el número total de clientes únicos que aparecen en la tabla 'Online Retail'. Es decir, si un cliente ha hecho varias compras, solo se cuenta una vez.

Medida #Cohortes

Creamos la segunda medida que necesitamos y la llamaremos #Cohorte

#Cohorte =

```
Var _PrimerMes=CALCULATE([#NºClientes],FILTER(ALL('Online
Retail'[NºdeMesesDespuesDelºCompra]),'Online
Retail'[NºdeMesesDespuesDelºCompra]=0))
Var _Cohort=DIVIDE([#NºClientes],_PrimerMes)
```

RETURN

_Cohort

Explicación y objetivo

Esta medida calcula el porcentaje de retención o permanencia de clientes en una cohorte, comparando el número de clientes en un mes determinado con el número de clientes en el mes 0 (el mes de la primera compra).

Paso a paso:

Variable _PrimerMes

```
VAR _PrimerMes = CALCULATE(
    [#NºClientes],
    FILTER(
        ALL('Online Retail'[NºdeMesesDespuesDelºCompra]),
        'Online Retail'[NºdeMesesDespuesDelºCompra] = 0
    )
)
```

Ilustración 24

- **¿Qué hace?**

Calcula el número de clientes en el mes 0, es decir, el mes en que los clientes hicieron su primera compra.

- **¿Cómo lo hace?**

- FILTER(ALL(...)): Elimina cualquier filtro aplicado sobre la columna [NºdeMesesDespuesDelºCompra] para asegurarse de que se evalúe **solo el mes 0**.
- CALCULATE([#NºClientes], ...): Aplica ese filtro para obtener el número total de clientes en ese mes inicial.

Variable _Cohort

- **¿Qué hace?**

Calcula el porcentaje de clientes actuales (en el mes actual del contexto de evaluación) respecto al número de clientes en el mes 0.

- **¿Por qué DIVIDE?**

DIVIDE es más seguro que usar / directamente, ya que evita errores de división por cero.

RETURN _Cohort

Devuelve el valor calculado: el porcentaje de retención de la cohorte en el mes actual.

Resultado final

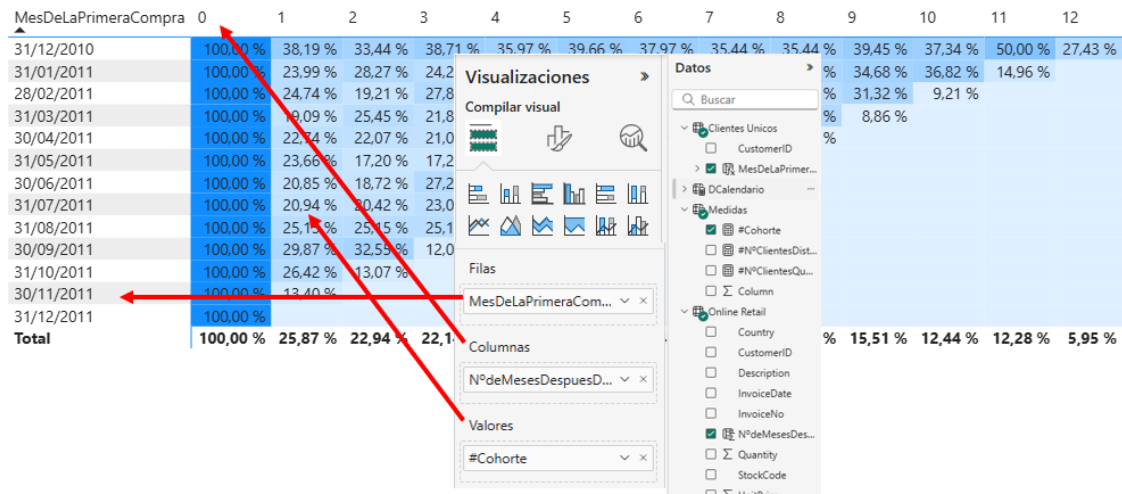


Ilustración 25

Medidas y columnas calculadas necesaria: INFO.VIEW.MEASURES()

Medidas	Tabla	Comentario	[Expression]
#NºClientesQueHanComprado	Medidas Fase I	Nº de clientes únicos que han comprado	#NºClientesQueHanComprado = CALCULATE(DISTINCTCOUNT('Online Retail'[CustomerID]), NOT(ISBLANK('Online Retail'[CustomerID])))
#Cohorte	Medidas Fase I	Segmento	#Cohorte = Var _PrimerMes= CALCULATE([#NºClientesQueHanComprado],FILTER(ALL('Online Retail'[NºdeMesesDespuesDelºCompra]),'Online Retail'[NºdeMesesDespuesDelºCompra]=0)) Var _Cohort=DIVIDE([#NºClientesQueHanComprado],_PrimerMes) RETURN _Cohort

Columnas calculadas, resumen

Columna	Tabla	[Expression]
NºdeMesesDespuesDelºCompra	Online Retail	NºdeMesesDespuesDelºCompra = DATEDIFF(RELATED('Clientes Unicos'[MesDeLaPrimeraCompra]), RELATED(DCalendario[FinDeMes]),MONTH)
MesDeLaPrimeraCompra	Clientes Unicos	MesDeLaPrimeraCompra = EOMONTH(CALCULATE(Min('Online Retail'[InvoiceDate])), 0)
FinDeMes	Dcalendario	FinDeMes = EOMONTH(DCalendario[Date],0)

Funciones empleadas CALCULATE, DATEDIFF, RELATED, DISTINCTCOUNT, DIVIDE, EOMONTH, FILTER, ISBLANK, NOT,

Bibliografía y conceptos técnicos relacionados

Análisis Cohorte, breve analisis

El **análisis de cohortes** es una técnica que estudia el comportamiento de grupos específicos de usuarios (cohortes) a lo largo del tiempo, es decir divide o clasifica a los usuarios, clientes o individuos en grupos (cohortes) que comparten una característica común durante un periodo de tiempo, como el mes de registro, la primera compra o la instalación de una aplicación. El objetivo es analizar cómo se comporta cada grupo a lo largo del tiempo, permitiendo identificar patrones, tendencias, diferencias y cambios en el comportamiento entre cohortes.

¿Qué es una cohorte y para qué se utiliza?

Una **cohorte** es un grupo de personas que comparten una característica común dentro de un período específico, como la fecha de registro, la primera compra, o el primer uso de un producto.

Ejemplo clásico: usuarios que se registraron en un sitio web durante el mes de enero 2025.

El análisis de cohortes permite entender patrones de retención, abandono y comportamiento de usuarios agrupándolos según cuándo comenzaron a interactuar con tu producto o servicio. La cohorte más común es la "cohorte de adquisición", donde se agrupa a los usuarios por la fecha de su primera interacción.

¿Para qué se utiliza?

Se utiliza para entender la retención, el comportamiento repetido, la lealtad, o el desgaste de usuarios a lo largo del tiempo, de manera más precisa que mirando promedios globales.

- **Medir la retención de usuarios**, ver cuántos usuarios de una cohorte siguen activos tras cierto tiempo.
- **Detectar cambios de comportamiento**. Identificar si una acción (como una campaña de marketing) afecta de forma diferente a cohortes distintas.
- **Optimizar estrategias**. Ajustar acciones según el comportamiento específico de cada cohorte.

El análisis se enfoca en comparar el comportamiento de diferentes cohortes a lo largo del tiempo. Por ejemplo, puedes analizar cómo los usuarios que se registraron en enero se comportan en los meses siguientes, y compararlos con los que se registraron en febrero, marzo, etc.

Normalmente se ve cómo un grupo de usuarios se comporta cada periodo posterior al evento inicial, dando respuestas, por ejemplo:

- ¿Cuánto tiempo permanecen activos los usuarios después de registrarse?
- ¿Cómo cambia el comportamiento de los usuarios según el mes en que se unieron?
- ¿Qué tan efectiva fue una campaña de marketing en términos de retención?
- Cuántos vuelven al sitio web en los siguientes meses.
- Cuántos hacen una segunda compra.
- Cuántos siguen usando el producto.
- Cuánto gastan con el tiempo.

Beneficios del análisis

Este análisis ayuda a identificar patrones estacionales, evaluar el impacto de cambios en el producto, optimizar estrategias de marketing, predecir comportamientos futuros y calcular métricas de negocio como el LTV de manera más precisa.

El análisis de cohortes es especialmente valioso porque elimina el sesgo de crecimiento general del negocio, permitiendo ver el comportamiento real de grupos específicos de usuarios y tomar decisiones basadas en datos más precisos.

En resumen, lo importante al interpretar es:

- **Comparar cohortes entre sí:** ver qué cohortes están funcionando mejor y por qué.
- **Analizar la pendiente:** una bajada muy rápida implica bajo engagement.
- **Medir el impacto de cambios:** si haces una mejora en onboarding en marzo, ¿mejora la retención respecto a febrero?

- **Permite analizar el comportamiento real de los usuarios a lo largo del tiempo**, no solo promedios generales.
- **Ayuda a identificar problemas específicos en etapas concretas del ciclo de vida del cliente.**
- **Facilita la toma de decisiones** basadas en datos segmentados y no en agregados.
- El análisis Cohort es fundamental para **entender la evolución de los usuarios y optimizar productos, servicios y estrategias de marketing** de manera precisa y efectiva.

IA Crear tabla calendario.

Fuente: Consulta <https://copilot.microsoft.com/shares/ir8ayY3E9PKnfguBUB68w>

En Power BI como crear una tabla calendario con calendarauto y que contenga las siguientes columnas Año, Trimestre, Mes, Semana, Dia, DiaSemana, NbDiaSemana, NbMes, NbTrimestreAño.

En Power BI, puedes crear una tabla de calendario utilizando la función CALENDARAUTO() y luego agregar las columnas necesarias con funciones de DAX. Aquí tienes un ejemplo de cómo hacerlo:

DCalendario =

```
ADDCOLUMNS(
    CALENDARAUTO(),
    "Año", YEAR([Date]),
    "Trimestre", QUARTER([Date]),
    "Mes", MONTH([Date]),
    "Semana", WEEKNUM([Date]),
    "Dia", DAY([Date]),
    "DiaSemana", WEEKDAY([Date], 1), -- El número del día de la semana (1 = lunes como primer día)
    "NbDiaSemana", FORMAT([Date], "dddd"), -- Nombre del día de la semana
    "NbMes", FORMAT([Date], "MMMM"), -- Nombre del mes
    "NbTrimestreAño", "Q" & QUARTER([Date]) & " " & YEAR([Date]) -- Trimestre y año
)
```

Explicación:

- *CALENDARAUTO() genera automáticamente un rango de fechas basado en los datos disponibles en el modelo.*
- *YEAR([Date]), QUARTER([Date]), MONTH([Date]), WEEKNUM([Date]), DAY([Date]) extraen los valores correspondientes.*
- *WEEKDAY([Date], 2) obtiene el número del día de la semana, empezando por lunes.*
- *FORMAT([Date], "dddd") y FORMAT([Date], "MMMM") transforman la fecha en el nombre del día y del mes.*
- *"Q" & QUARTER([Date]) & " " & Format(YEAR([Date]), "YY") crea una etiqueta que combina el trimestre y el año.*