

Exploración y visualización de datos

Origen de los datos

Desarrollo la práctica y el análisis de los datos sobre un Dataset que recoge las transacciones de un E-Commerce.

<https://www.kaggle.com/datasets/carrie1/ecommerce-data/data>

Los datos iniciales de la tabla eran los siguientes:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
1	536365	85123A	WHITE HANGING HEART T-LIGHT HOLD...	6	12/01/2010 8:26:00	255	17850	United Kingdom
2	536365	71053	WHITE METAL LANTERN	6	12/01/2010 8:26:00	339	17850	United Kingdom
3	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/01/2010 8:26:00	275	17850	United Kingdom
4	536365	84029G	KNITTED UNION FLAG HOT WATER BOT...	6	12/01/2010 8:26:00	339	17850	United Kingdom
5	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	12/01/2010 8:26:00	339	17850	United Kingdom

En base a estos datos y utilizando distintas funciones de DAX, he procedió con la obtención de más columnas y medidas calculadas que nos permitirán, más adelante plantear distintos KPI a lo cuales daré respuesta en los distintos Dashboards.

Transformación manual de los datos

Las columnas InvoiceNO, StockCODE, Description los dejamos tal y como nos vienen de origen.

Para la columna Quantity le indicamos que se trata de un entero.

La columna InvoiceData le cambiamos el tipo y seleccionamos la configuración regional de Origen.

Para la columna de UnitPrice, sustituimos el punto que define el decimal y lo sustituimos por coma para poder hacer la conversión a número decimal.

El resto de columnas las dejamos tal y como vienen, salvo la columna de Country sobre la cual aplicamos un filtro para excluir al Reino Unido por ciertos problemas que daban las filas que contenían este país, tal y como se indicaba en las especificaciones de la data set.

Obtención de más columnas

Fecha

Las primeras columnas adicionales que he obtenido ha sido la de las siguientes:

Hour

```
Hour = HOUR('Invoice data'[InvoiceDate])
```

Hour Group

```
Hour Group = IF ('Invoice data'[Hour] < 12, "AM", "PM")
```

Day

```
Day = WEEKDAY('Invoice data'[InvoiceDate], 2)
```

Nota: el 2 responde al segundo parámetro de la función WEEKDAY en el cual hay que especificar cual es el primer día de la semana.

Day of Week

```
Day of Week = FORMAT('Invoice data'[InvoiceDate], "DDDD")
```

Nota: el segundo parámetro "DDDD" hace referencia a la forma en la que queremos visualizar el día de la semana. En este caso, esta es la sintaxis para el nombre completo del día, Lunes por ejemplo.

Categoría del producto

Con esta categoría establecemos en función del código del producto, si se trata de un producto nacional o de origen importado.

Para este ejercicio se define como producto importado todo aquel que tenga un StockCode con un LEN superior a 5.

Para mayor clasificación se definen otro criterio que indicara el origen de aquellos productos que han sido importados. "Imported from Europe" o "Imported Worldwide" en función de la última letra del StockCode

Con estos dos criterios calculamos y obtenemos las siguientes dos columnas:

Stock Category

```
Stock Category = IF(

RIGHT('Invoice data'[StockCode], 1) = "A" ||
RIGHT('Invoice data'[StockCode], 1) = "B" ||
RIGHT('Invoice data'[StockCode], 1) = "C",

"Imported from Europe",

IF(

RIGHT('Invoice data'[StockCode], 1) = "D" ||
RIGHT('Invoice data'[StockCode], 1) = "E" ||
RIGHT('Invoice data'[StockCode], 1) = "F" ||
RIGHT('Invoice data'[StockCode], 1) = "G",

"Imported Worldwide", "National Product"
))
```

Nota: Con la función RIGHT aplicada sobre una columna e indicando la posición del elemento obtenemos el primer elemento de cada StockCode y comprobamos si cumple la condición booleana.

Stock Category 2

```
Stock Category 2 = IF (LEN('Invoice data'[StockCode]) > 5, "Imported Product",
"National Product")
```

Categoría por género

Para este ejercicio establecemos la condición que determina el género del cliente en función de que el último elemento del CustomerID sea par o impar

Customer Gener

```
Customer Gener = IF(MOD((CONVERT(RIGHT('Invoice data'[InvoiceNo], 1), INTEGER)), 2) =
0, "Female", "Male")
```

Agrupación de países

Agrupamos a los países en dos categorías. Aquellos que tiene mas de un 10% de las ventas totales los incluimos en la categoría “Countries High amount %” y el resto de países con porcentajes inferiores en la categoría “Countries Low amount %”. Para ello utilizamos la función de agrupación de propio Power BI

Grupos

✕

Nombre *
Country (grupos)

Campo
Country

Tipo de grupo
Lista

Valores no agrupados

Grupos y miembros

- ▾ Countries High amount %
 - EIRE
 - France
 - Germany
 - Netherlands
- ▾ Countries Low amount %
 - Australia
 - Austria
 - Bahrain
 - Belgium
 - Brazil

AgruparDesagrupar

☐ Incluir otro grupo ⓘ

AceptarCancelar

Creación de tabla para medidas adicionales

Para llevar a cabo medidas adicionales necesarias para contestar a las nuevas columnas creo una nueva tabla DAX medidas para poder recogerlas.

Algunas de estas medidas, se han obtenido y aplicado sobre gráficos que finalmente muestro y otras finalmente no se han utilizado, pero aun así las mantengo por que pueden ser requeridas en cualquier momento si se amplía el alcance de los Dashboards.

Customer Count = `DISTINCTCOUNT('Invoice data'[CustomerID])`

Customer Count % = `[Customer Count] / CALCULATE([Customer Count], ALL('Invoice data'))`

Invoice Count = `DISTINCTCOUNT('Invoice data'[InvoiceNo])`

Invoice Count = `DISTINCTCOUNT('Invoice data'[InvoiceNo])`

National Product \$ = `CALCULATE([Total Amount], 'Invoice data'[Stock Category] = "National Product")`

National Producto % = `[National Product $] / [Total Amount]`

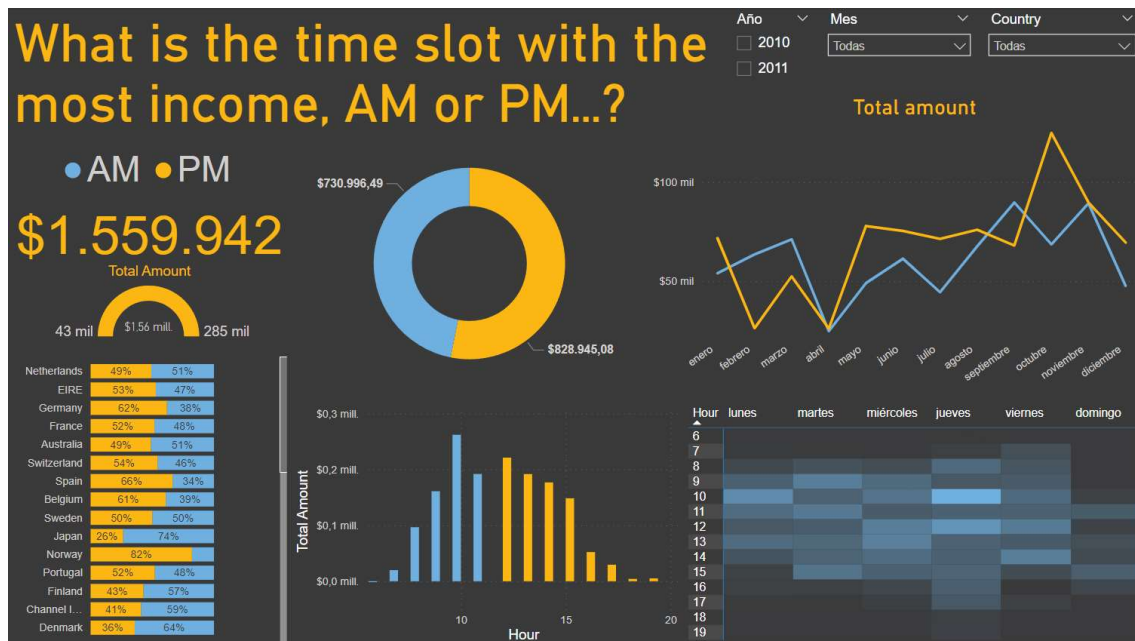
Total Amount = `SUMX('Invoice data', [Quantity] * [UnitPrice])`

Total amount % = `[Total Amount] / CALCULATE([Total Amount], ALL('Invoice data'))`

Unit Price Average = `AVERAGE('Invoice data'[UnitPrice])`

Dashboards

What is the time slot with the most income, AM or PM...?

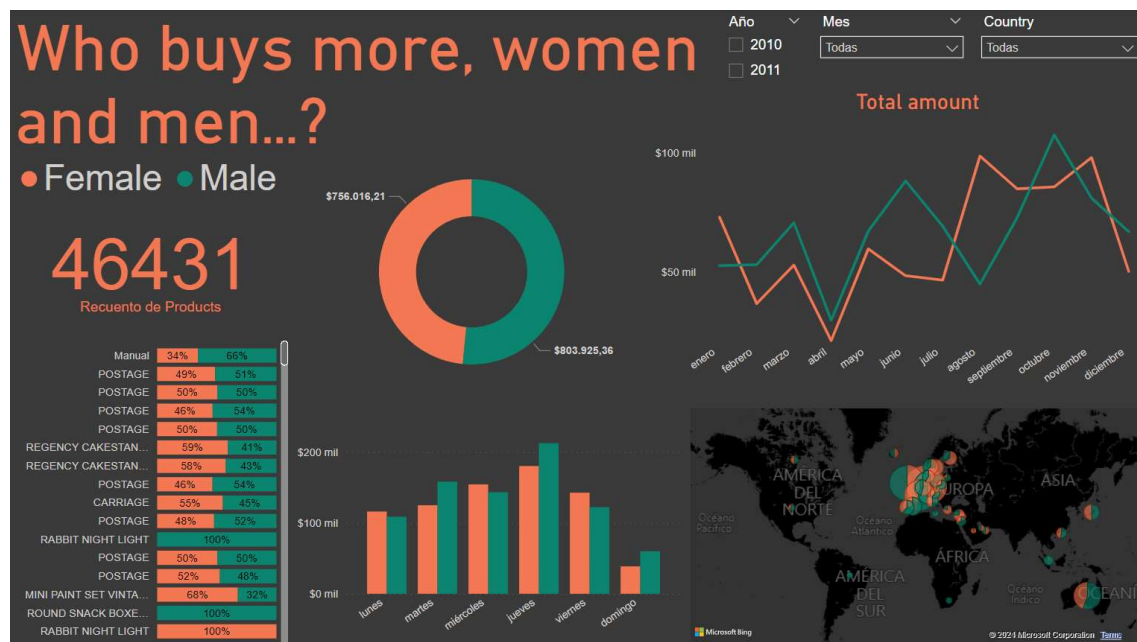


En el primer Dashboard doy respuesta a la pregunta de, “¿En franja horaria se factura más?

Los gráficos planteados en este dashboard responden a dicha pregunta de distintas maneras, permitiendo la segmentación y el análisis de los datos en mayor profundidad dependiendo de los parámetros de consulta que queramos trasladarle.

El panel incluye una gráfica que nos traslada como valor mínimo la media de facturación de todos los países y como valor máximo, la facturación total del país que más ha facturado.

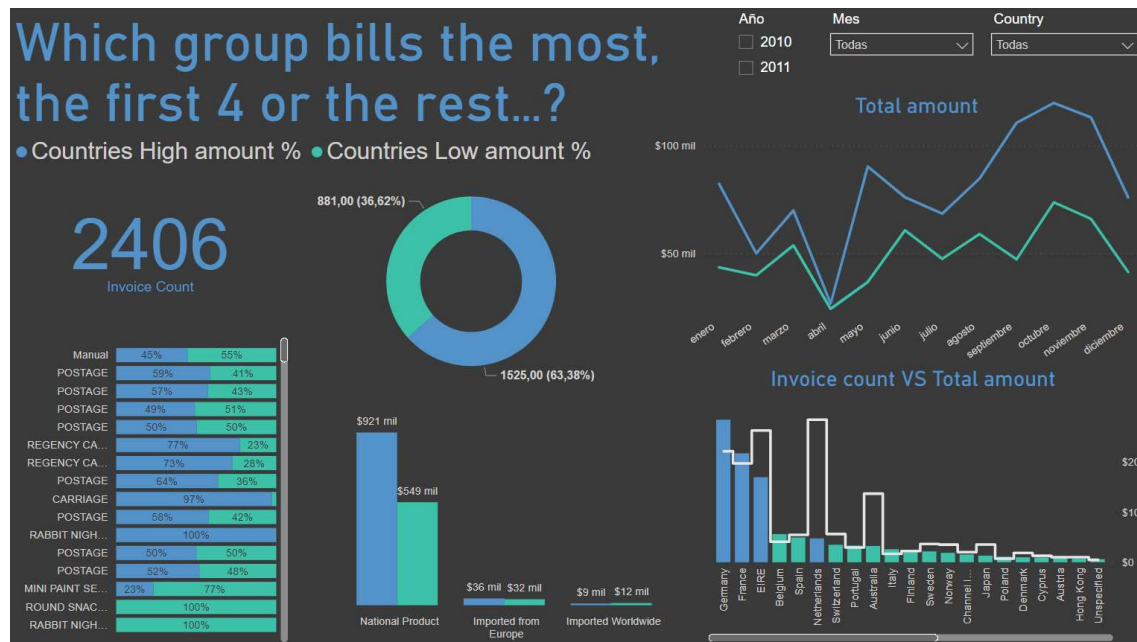
Who buys more, women and men...?



En el segundo dashboard resolvemos la pregunta que encabeza el panel pudiendo ver las diferencias entre las compras realizadas por mujeres respecto de las compras realizadas por los hombres.

Podemos analizar los productos que consumen con mayor frecuencia cualquier de los dos géneros. También incluye un mapa de forma que podamos trasladar la consulta sobre los hábitos de consumo a la geolocalización de los sujetos de estudio pudiendo de esta forma concretar por países nuestra consulta.

Which group bills the most, the first 4 or the rest...?



Por último, en este dashboard hemos agrupado los países de forma que tenemos en el primer grupo aquellos 4 países los cuales son los únicos que han facturado por si solos mas del 10% de la facturación total de todos los países juntos.

Para poder llevar a cabo este dashboard, me apoye previamente en las medidas obtenidas de la tabla DAX medidas. Estas medidas, pudieron permitirme observar datos relevantes los cuales considero importante tener en cuenta.

Un dato que me parece remarcable es que el país con mayor facturación es Netherlands, el cual alcanza, el puesto mas alto en el rankin, a pesar de tener menos muchas menos facturas que los otros 3 países que forman el grupo "Countries Hight amount %".

Netherlands, factura el 18,25% del total da la facturación y lo hace únicamente con un total de 101 facturas y 9 clientes. Comparando estos datos con los datos, por ejemplo de Germany, nos damos cuenta que este ultimo se queda \$60.000 por debajo de la facturación de Netherlands, a pesar de tener mas de 6 veces más facturas emitidas y mas de 10 veces más clientes.

\$1.559.942		2406		423	
Total Amount		Invoice Count		Customer Count	
Country	Total Amount	Total amount %	Invoice Count	Customer Count	Customer Count %
Netherlands	\$284.661,54	18,25 %	101	9	2,13 %
EIRE	\$263.276,82	16,88 %	300	4	0,95 %
Germany	\$221.680,21	14,21 %	603	95	22,46 %
France	\$197.403,90	12,65 %	461	68	20,80 %
Australia	\$137.077,27	8,79 %	69	9	2,13 %
Switzerland	\$56.385,35	3,61 %	74	22	5,20 %
Spain	\$54.774,58	3,51 %	105	31	7,33 %
Belgium	\$40.910,96	2,62 %	119	25	5,91 %
Sweden	\$36.595,91	2,35 %	46	8	1,89 %
Japan	\$35.340,62	2,27 %	28	8	1,89 %
Norway	\$33.163,46	2,25 %	40	10	2,38 %
Portugal	\$29.367,02	1,88 %	71	20	4,73 %
Finland	\$22.326,74	1,43 %	12	12	2,84 %
Channel Islands	\$20.086,29	1,29 %	33	9	2,13 %
Denmark	\$18.768,14	1,20 %	21	9	2,13 %
Italy	\$16.890,51	1,08 %	55	15	3,55 %
Cyprus	\$12.946,29	0,83 %	20	9	1,89 %
Austria	\$10.154,32	0,65 %	19	11	2,60 %
Hong Kong	\$10.117,04	0,65 %	15	1	0,24 %
Singapore	\$9.120,39	0,58 %	10	1	0,24 %
Israel	\$7.907,82	0,51 %	9	5	1,18 %
Poland	\$7.213,14	0,46 %	24	6	1,42 %
Unspecified	\$4.749,79	0,30 %	13	5	1,18 %
Greece	\$4.310,00	0,28 %	6	4	0,95 %
Iceland	\$3.666,38	0,24 %	7	1	0,24 %
Canada	\$2.505,47	0,16 %	10	2	0,47 %
Malta	\$2.505,47	0,16 %	10	2	0,47 %
Total	\$1.559.941,57	100,00 %	2406	423	100,00 %

InvoiceNo	CustomerID	Country	InvoiceDate	Total Amount
556917	12415	Australia	15/06/2011 13:37:00	\$22.775,93
563814	12415	Australia	19/08/2011 8:51:00	\$21.880,44
572035	14646	Netherlands	20/10/2011 12:03:00	\$20.277,92
563076	14646	Netherlands	11/08/2011 16:12:00	\$19.150,66
541220	14156	EIRE	14/01/2011 14:11:00	\$16.774,72
545475	12415	Australia	03/03/2011 10:59:00	\$16.726,84
Total				\$1.559.941,57

CustomerID	Total Amount	Total amount %	Invoice Count	Invoice Count %
14646	\$279.489,02	17,92 %	77	3,20 %
14911	\$132.572,62	8,50 %	248	10,31 %
12415	\$123.725,45	7,93 %	26	1,08 %
14156	\$113.384,14	7,27 %	66	2,74 %
17404	\$30.300,82	1,94 %	18	0,75 %
12678	\$27.749,15	1,78 %	73	3,03 %
12753	\$21.024,01	1,35 %	11	0,46 %
12731	\$18.793,41	1,20 %	14	0,58 %
12471	\$18.740,92	1,20 %	49	2,04 %
12678	\$17.588,26	1,13 %	13	0,54 %
12681	\$13.677,59	0,88 %	27	1,12 %
12621	\$13.612,07	0,87 %	23	0,96 %
12433	\$13.375,87	0,86 %	7	0,29 %
12540	\$13.309,04	0,85 %	19	0,79 %
12477	\$13.117,01	0,84 %	9	0,37 %
12682	\$12.288,22	0,79 %	31	1,29 %
12557	\$11.990,96	0,77 %	5	0,21 %
12409	\$11.056,93	0,71 %	7	0,29 %
12550	\$9.861,39	0,63 %	3	0,12 %
12709	\$9.294,10	0,60 %	26	1,08 %
12744	\$9.120,39	0,58 %	10	0,42 %
12567	\$9.114,94	0,58 %	11	0,46 %
12451	\$8.933,52	0,57 %	6	0,25 %
12683	\$8.221,09	0,53 %	20	0,83 %
14936	\$8.117,42	0,52 %	9	0,37 %
12428	\$7.877,20	0,50 %	12	0,50 %
12435	\$7.629,89	0,49 %	2	0,08 %
12583	\$7.187,14	0,46 %	18	0,75 %
Total	\$1.559.941,57	100,00 %	2406	100,00 %

He considerado muy vital la información aportada por las tablas de la pestaña “Tablas de estudio” pues con la ayuda de la visualización directa de los datos de facturación y valores porcentuales, así como la asignación de formatos condicionales a las distintas columnas, he podido comprender la relevación de los datos.

Sin estas tablas, hubiese sido muy complicado darme cuenta de esta relación de los distintos países respecto a la facturación total.

Otro dato a señalar, es que uno de los 9 clientes de Netherlands, superar por si solo, la facturación total del resto de países, manejando un volumen total de ventas del 17,92%.

\$284.662

Total Amount

101

Invoice Count

9

Customer Count

InvoiceNo	CustomerID	Country	InvoiceDate	Total Amount	
572035	14646	Netherlands	20/10/2011 12:03:00	\$20,277.92	
563076	14646	Netherlands	11/08/2011 16:12:00	\$19,150.66	
548011	14646	Netherlands	29/03/2011 11:14:00	\$15,719.56	
552883	14646	Netherlands	12/05/2011 10:13:00	\$14,415.74	
562789	14646	Netherlands	09/08/2011 13:50:00	\$12,341.21	
544479	14646	Netherlands	21/02/2011 10:28:00	\$11,956.10	
Total				\$284,661.54	

Country	Total Amount	Total amount %	Invoice Count	Customer Count	Customer Count %
Netherlands	\$284,661.54	18.25 %	101	9	2.13 %
LTIC	\$263,276.82	16.88 %	360	4	0.95 %
Germany	\$221,698.21	14.21 %	603	95	22.46 %
France	\$197,403.90	12.65 %	461	88	20.80 %
Australia	\$137,077.27	8.79 %	69	9	2.13 %
Switzerland	\$56,385.35	3.61 %	74	22	5.20 %
Spain	\$54,774.58	3.51 %	105	31	7.33 %
Belgium	\$40,910.96	2.62 %	119	25	5.91 %
Sweden	\$36,595.91	2.35 %	46	8	1.89 %
Japan	\$35,340.62	2.27 %	28	8	1.89 %
Norway	\$35,163.46	2.25 %	40	10	2.36 %
Portugal	\$29,367.02	1.88 %	71	20	4.73 %
Finland	\$22,326.74	1.43 %	48	12	2.84 %
Channel Islands	\$20,086.29	1.29 %	33	9	2.13 %
Denmark	\$18,768.14	1.20 %	21	9	2.13 %
Italy	\$16,890.51	1.08 %	55	15	3.55 %
Cyprus	\$12,946.29	0.83 %	20	8	1.89 %
Austria	\$10,154.32	0.65 %	19	11	2.60 %
Hong Kong	\$10,117.04	0.65 %	15	1	0.24 %
Singapore	\$9,120.39	0.58 %	10	1	0.24 %
Israel	\$7,907.82	0.51 %	9	5	1.18 %
Poland	\$7,213.14	0.46 %	24	6	1.42 %
Unspecified	\$4,749.79	0.30 %	13	5	1.18 %
Greece	\$4,710.52	0.30 %	6	4	0.95 %
Iceland	\$4,310.00	0.28 %	7	1	0.24 %
Canada	\$3,666.38	0.24 %	6	4	0.95 %
Malta	\$2,505.47	0.16 %	10	2	0.47 %
Total	\$1,559,941.57	100.00 %	2406	423	100.00 %

CustomerID	Total Amount	Total amount %	Invoice Count	Invoice Count %
14646	\$279,489.02	17.92 %	77	3.20 %
12759	\$1,531.44	0.10 %	3	0.12 %
12775	\$1,351.24	0.09 %	6	0.25 %
12778	\$804.23	0.05 %	7	0.29 %
12787	\$463.66	0.03 %	3	0.12 %
12802	\$412.58	0.03 %	2	0.08 %
12790	\$324.92	0.02 %	1	0.04 %
12791	\$192.60	0.01 %	1	0.04 %
12789	\$91.85	0.01 %	1	0.04 %
Total	\$284,661.54	18.25 %	101	4.20 %

La información que han aportado las tablas de estudio ha sido crucial para poder desarrollar los dashboards y poder obtener información de ellos.