

Tarea S5.01. Iniciación al análisis de datos con Power BI e indicadores

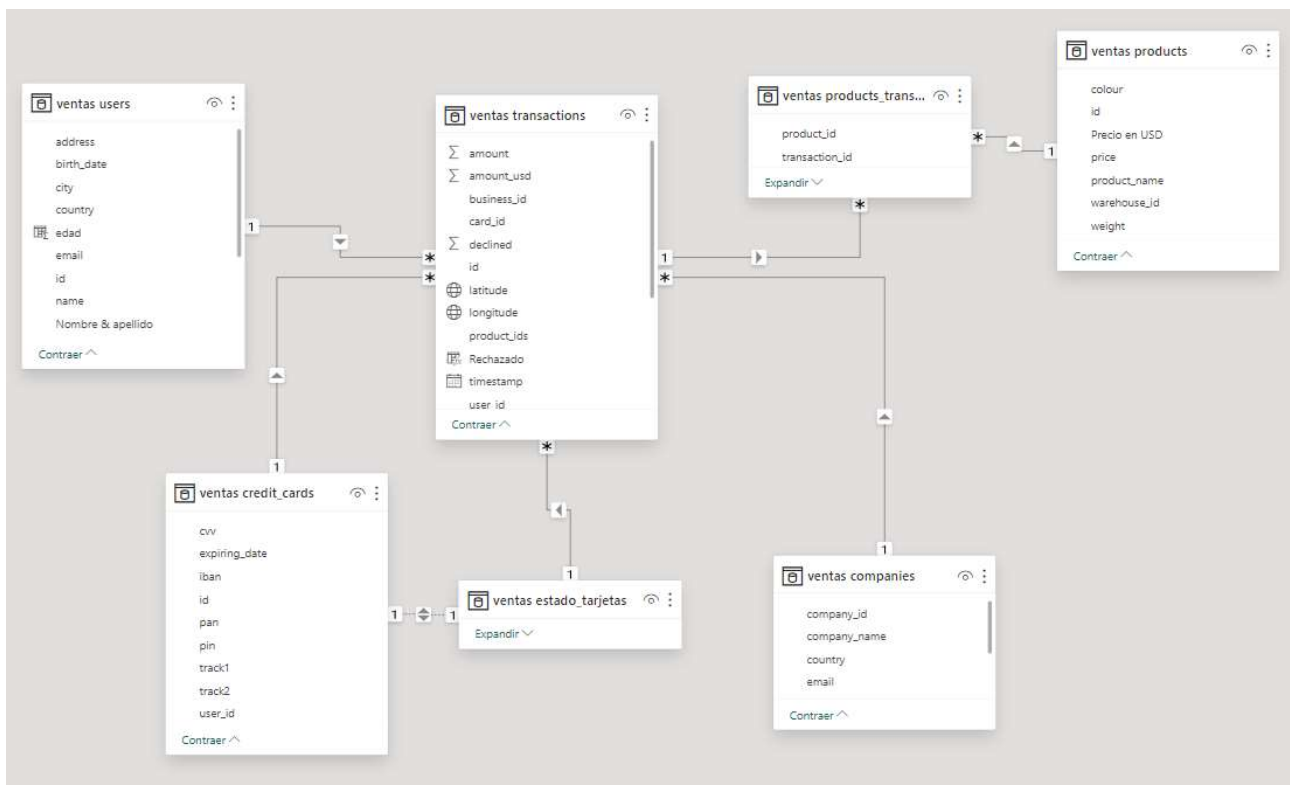
Radostin Pavlov

Nivel 1

- Ejercicio 1

Importa los datos de la base de datos utilizada previamente. Después de cargar los datos, muestra el modelo de la base de datos en Power BI.

Solución



- Ejercicio 2

Tu empresa está interesada en evaluar el rendimiento de las transacciones realizadas a lo largo de los años. Para ello, se ha solicitado la creación de un indicador clave de rendimiento (KPI). El KPI debe proporcionar una clara visualización del objetivo empresarial de conseguir una suma total de 25.000 transacciones por cada año.

Solución

Para crear el KPI primero se añade suma de 'amount' como valor de una visualización de tipo KPI y en eje de tendencia se pone año (que es parte de la jerarquía derivada del campo timestamp).

Después se crea una nueva medida con el valor objetivo de ventas anuales = 25000 y se pone en 'Destino'.

A continuación, se añade un filtro para el año y otro para el campo 'declined'. Para poder seleccionar sí o no en vez de 0 y 1, se crea una nueva columna con la siguiente fórmula de DAX: Rechazado = IF('ventas transactions'[declined] = 0, "No", "Sí"). Finalmente, se cambia el título del KPI por defecto a Valor Ventas Anuales.



- Ejercicio 3

Desde marketing te solicitan crear una nueva medida DAX que calcule la media de transacciones realizadas durante el año 2021. Posteriormente visualiza esta media en un medidor que refleje las ventas realizadas. Es importante recordar que la empresa tiene un objetivo de ventas establecido en 250 transacciones.

Solución

Se crean las siguientes medidas nuevas:

Valor medio de ventas anuales = `average('ventas transactions'[amount])`

Valor medio de ventas anuales objetivo = 250

El KPI se genera de la misma manera que el anterior, utilizando la primera medida como valor, y la segunda como objetivo pero representándolas en un objeto tipo medidor. Se mantiene el filtrado global de la página por año y por validez de la transacción.



- Ejercicio 4

Realiza el mismo procedimiento que realizaste en el ejercicio 3 para el año 2022.

Solución

Al tener el filtro por años configurado, solo hay que seleccionar el año 2022:



- Ejercicio 5

El objetivo de este ejercicio es crear una KPI que visualice la cantidad de empresas por país que participan en las transacciones. La meta empresarial es garantizar que haya al menos 3 empresas participantes por país. Para ello, será necesario utilizar DAX para calcular y representar esta información de manera clara y concisa.

Solución

Empezamos explorando la suma de ventas por país y por empresa. Se observa que todas las compañías han hecho transacciones tanto en 2021 como en 2022, independientemente de si se consideran las transacciones rechazadas o no.

Año			Rechazado	
<input type="checkbox"/> 2021			<input type="checkbox"/> Seleccionar todo	
<input checked="" type="checkbox"/> 2022			<input checked="" type="checkbox"/> No	
			<input type="checkbox"/> Sí	

country	company_name	Suma de amount
Germany	Nunc Interdum Incorporated	3,995.97
Ireland	Malesuada PC	3,423.85
United Kingdom	Enim Condimentum Ltd	3,405.42
Norway	Arcu LLP	2,718.74
United Kingdom	Non Institute	2,212.85
Canada	Lorem Eu Incorporated	1,737.49
Sweden	Ut Semper Foundation	1,218.47
Belgium	Sed LLC	416.66
Italy	Quam A Felis Industries	376.41
Sweden	Fringilla Porttitor Incorporated	354.28
Australia	Magna A Neque Industries	312.50
Germany	Rutrum Non Inc.	266.90
United States	Fusce Corp.	239.87
United States	Auctor Mauris Vel LLP	179.40
Australia	Mattis Foundation	155.57
China	Quisque Libero LLC	155.44
France	Donec Fringilla PC	119.68
United Kingdom	Orci Adipiscing Limited	111.64
Italy	Amet Nulla Donec Corporation	92.61
Belgium	Aliquam laculis Lacus Corp.	92.18
Ireland	Neque Tellus Imperdiet Corp.	78.29
Australia	Metus Vitae Associates	61.37
Norway	Donec Ltd	42.82
United Kingdom	Ac Libero Inc.	30.76
United States	Nascetur Ridiculus Mus Inc.	20.35
Total		21,834.57

A continuación, se hace un recuento de compañías por país y se observa que efectivamente algunos países tienen menos de 3 compañías que al ser filtradas quedan solo los países que nos interesan. Estos se pueden representar en un gráfico, indicando el mínimo deseado:



Si se desea un indicador más conciso, se puede generar una tabla con recuento de empresas para los países que tienen más de 3 compañías que han hecho compras, y compararlo con el número total de países (recuento de country → 15). Esto se puede hacer en Power Query mediante los siguientes pasos:

- Duplicando la tabla transactions
- Eliminando todas las columnas menos company_id y country
- Agrupando por países simultaneo recuento de las compañías
- Filtrando para dejar los países con más de 3 compañías y finalmente contando las fila

A continuación se muestra el resultado de los últimos dos pasos:

= Table.SelectRows("#Filas ordenadas1", each [Recuento] > 3)		
	country	Recuento
1	Sweden	11
2	Netherlands	10
3	United States	9
4	United Kingdom	9
5	Italy	9
6	Germany	8
7	Belgium	8
8	Norway	7
9	Australia	6
10	New Zealand	6
11	Ireland	6
12	Canada	5

= #table(1, {{"Recuento calculado"}})		
	Column1	
1		13

La misma tabla se puede crear con DAX:

```

1 Países con más de 3 compañías activas =
2 COUNTROWS(
3     FILTER(
4         SUMMARIZE(
5             'ventas companies',
6             'ventas companies'[country],
7             "CompañíasConTransacciones", COUNTROWS('ventas transactions')
8         ),
9         [CompañíasConTransacciones] > 3)

```

En este código, **SUMMARIZE(...)** genera una tabla basada en el recuento del número de compañías con agrupamiento por 'ventas companies'[country] que sería el equivalente a la tabla mostrada antes y **FILTER(...)** filtra la tabla por número de compañías con al menos 3 transacciones.

1	Tabla =
2	FILTER (
3	SUMMARIZE (
4	'ventas companies',
5	'ventas companies'[country],
6	"NumCompañías", COUNTROWS ('ventas companies')
7),
8	[NumCompañías] >= 3
9)

NumCompañías	country
11	Sweden
10	Netherlands
9	Italy
9	United Kingdom
9	United States
8	Germany
8	Belgium
7	Norway
6	Ireland
6	New Zealand
6	Australia
5	Canada
3	France

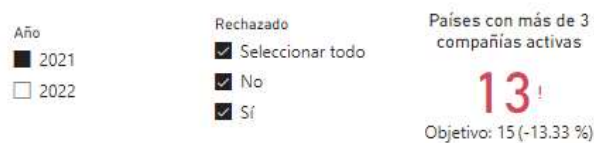
Finalmente se crea una nueva medida que cuenta las compañías de la tabla mostrada arriba:

```

1 Países con más de 3 compañías activas =
2 COUNTROWS(
3     FILTER(
4         SUMMARIZE(
5             'ventas companies',
6             'ventas companies'[country],
7             "NumCompañías", COUNTROWS('ventas companies')
8         ),
9         [NumCompañías] >= 3)
10 )

```

El resultado final es:

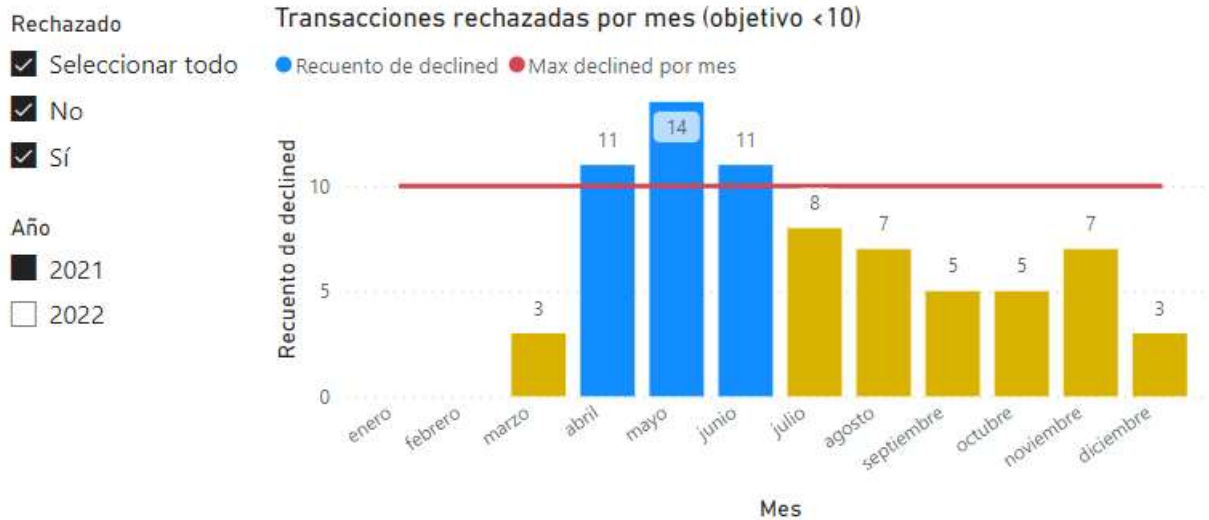


- Ejercicio 6

Crea una nueva KPI que permita visualizar la cantidad de transacciones declinadas a lo largo del tiempo. La empresa estableció un objetivo de tener menos de 10 transacciones declinadas por mes.

Solución

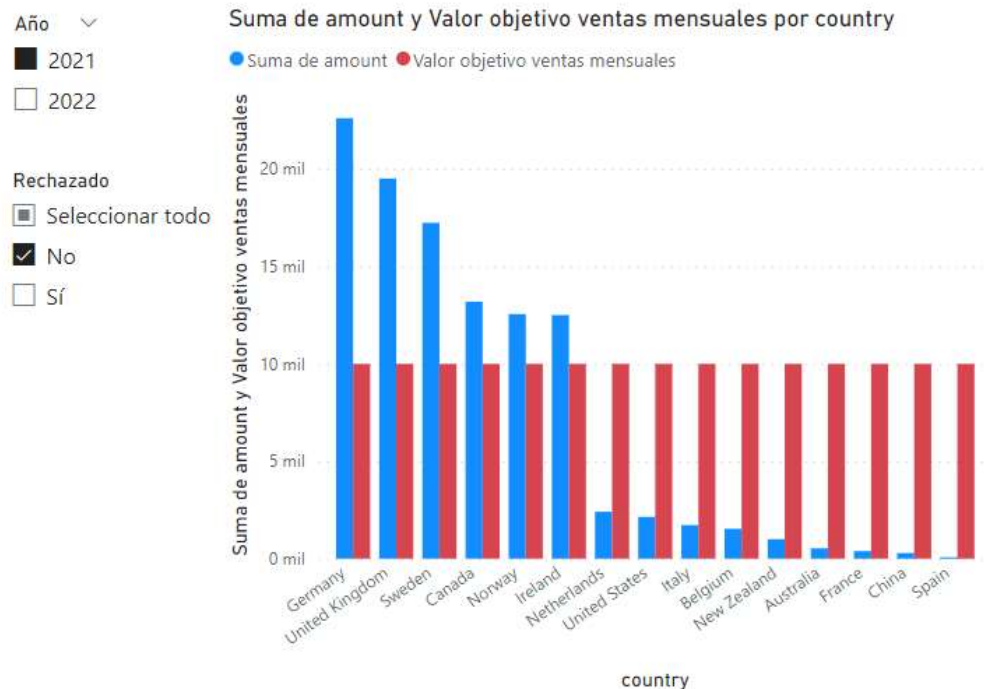
Se emplea un gráfico de columnas y líneas y formato condicional para diferenciar mejor los dos grupos:



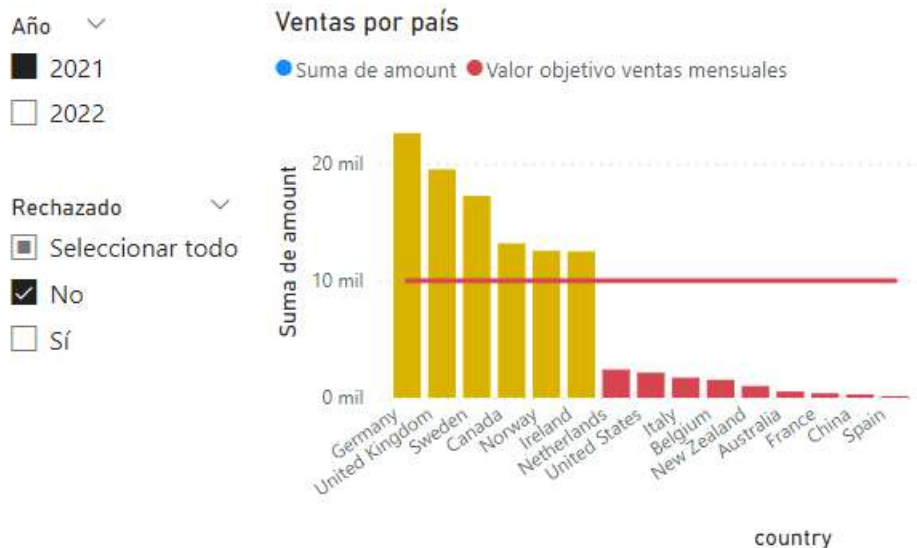
- Ejercicio 7

Crea un gráfico de columnas agrupadas que refleje la sumatoria de las ventas por país. El objetivo de la empresa es tener por lo menos 10.000 transacciones por mes.

Solución



Una representación alternativa y quizá más clara sería con otro gráfico de columnas y líneas y formato condicional:



-Ejercicio 8

En este ejercicio, se quiere profundizar en las transacciones realizadas por cada usuario y presentar la información de forma clara y comprensible. En una tabla, presenta la siguiente información: - Nombre y apellido de los usuarios/as (habrá que crear una nueva columna que combine esta información). - Edad de los usuarios/as. - Media de las transacciones en euros. - Media de las transacciones en dólares (conversión: 1 euro equivale a 1,08 dólares). Se deben realizar los cambios necesarios para identificar a los usuarios/as que tuvieron una media de 300 o más euros y 320 o más dólares en sus transacciones.

Solución

Hay al menos dos maneras de hacer este ejercicio. Una consiste en realizar los siguientes pasos **sin emplear DAX**:

- Combinar en Power Query las columnas Nombre y Apellido creando una columna nueva en la tabla users.
- Crear en Power Query con la función Columna Personalizada un campo nuevo en la tabla transactions convirtiendo todos los valores de 'amount' en dólares simplemente multiplicando el identificador del campo por 1.08.
- Combinar los campos creados y el de los valores en euros en una visualización de tipo tabla.
- Filtrar los valores para ocultar los que están por debajo de 300 EUR / 320 USD. Se puede emplear un filtro.

La segunda manera es generando toda la tabla **mediante código DAX**. También se pueden hacer combinaciones de las dos maneras.

1	'Promedios ventas por usuario' =
2	SELECT COLUMNS (
3	SUMMARIZE (
4	FILTER (
5	'ventas users',
6	CALCULATE(
7	AVERAGEX('ventas transactions', 'ventas transactions'[amount]) > 300
8)
9),
10	'ventas users'[name],
11	'ventas users'[surname],
12	"Nombre & apellido", 'ventas users'[name] & " " & 'ventas users'[surname],
13	"Promedio ventas en EUR", ROUND(AVERAGEX('ventas transactions', 'ventas transactions'[amount]), 2),
14	"Promedio ventas en USD", ROUND(1.08 * ROUND(AVERAGEX('ventas transactions', 'ventas transactions'[amount]), 2), 2)
15),
16	"Nombre & apellido", [Nombre & apellido],
17	"Promedio ventas en EUR", [Promedio ventas en EUR],
18	"Promedio ventas en USD", [Promedio ventas en USD]
19)

Nombre & apellido	Promedio ventas en EUR	Promedio ventas en USD
Sasha Emerson	301,25	325,35
Lucas Dawson	304,43	328,78
Dawn Murray	306,94	331,5
Keane Mckinney	308,12	332,77
Iona Soto	309,45	334,21
Warren Christian	317,04	342,4
Kimberley Avila	317,41	342,8
Chase Ellis	320,29	345,91
Keegan Watson	320,38	346,01
Amelia Valenzuela	321,39	347,1
Slade Downs	323,22	349,08
David Vance	324,43	350,38

En este código, summarize() recopila los campos y hace las transformaciones necesarias:

- Concatenación de name y surname con & “ “ &
- Calculo del promedio de amount con averagex() y redondeo con round()
- Conversión a dólares multiplicando el promedio de amount por 1.08

También dentro de summarize() se incluye la función filter() que filtra en base al promedio de amount (>300).

Finalmente se seleccionan las tres columnas deseadas (Nombre & apellido, Promedio en EUR y Promedio en USD ya que summarize() inevitablemente devuelve también las columnas nombre y apellido originales.

A continuación, se muestra una comparación de las tablas obtenidas de las dos maneras (están ordenadas por orden ascendente de la cantidad para apreciar el corte en 300 EUR). Como se puede ver, las dos tablas son prácticamente iguales:

Nombre & apellido	Promedio de amount	Promedio de amount_usd
Sasha Emerson	301,25	325,35
Lucas Dawson	304,43	328,78
Dawn Murray	306,94	331,50
Keane Mckinney	308,12	332,77
Iona Soto	309,45	334,21
Warren Christian	317,04	342,40
Kimberley Avila	317,41	342,80
Chase Ellis	320,29	345,91
Keegan Watson	320,38	346,01
Amelia Valenzuela	321,39	347,10
Slade Downs	323,22	349,08
David Vance	324,43	350,38
Brett Kirby	325,18	351,19
Total	387,19	418,17

Nombre & apellido	Promedio ventas en EUR	Promedio ventas en USD
Sasha Emerson	301,25	325,35
Lucas Dawson	304,43	328,78
Dawn Murray	306,94	331,50
Keane Mckinney	308,12	332,77
Iona Soto	309,45	334,21
Warren Christian	317,04	342,40
Kimberley Avila	317,41	342,80
Chase Ellis	320,29	345,91
Keegan Watson	320,38	346,01
Amelia Valenzuela	321,39	347,10
Slade Downs	323,22	349,08
David Vance	324,43	350,38
Brett Kirby	325,18	351,19

- Ejercicio 9

Redacta un párrafo breve, máximo 50 palabras, explicando el significado de las cifras presentadas en las visualizaciones de Power BI. Puedes interpretar los datos por lo general o centrarte en algún país específico. Acompaña las interpretaciones realizadas con la captura de pantalla de las visualizaciones que vas a analizar.

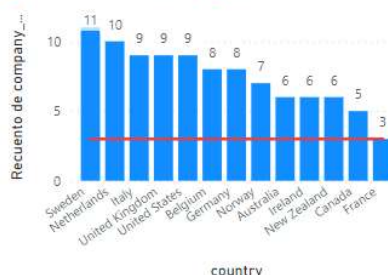
Solución

En el 2021 se ha superado con creces el objetivo de ventas y de valor medio de ventas anuales a pesar de la ausencia de datos de enero y febrero. 13 de los 15 países registrados superan el objetivo de mínimo 3 empresas por país. El objetivo de no superar 10 transacciones rechazadas por mes también se ha cumplido, salvo en abril, mayo y junio. 6 países han superado el objetivo de mínimo 10 mil euros en ventas mensuales.



Compañías por país (mínimo deseado 3)

● Recuento de company_name ● Numero mínimo de países



Países con más de 3 compañías activas

13!

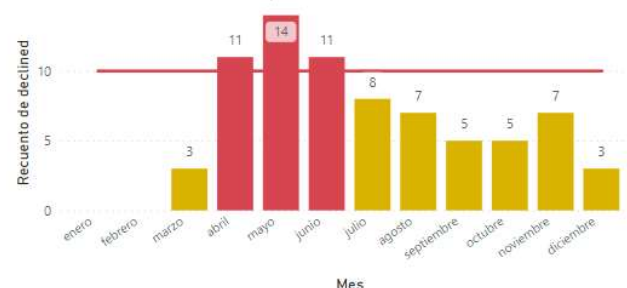
Objetivo: 15 (-13.33 %)

Año
■ 2021
□ 2022

Rechazado
☒ Seleccionar todo
☒ No
☒ Sí

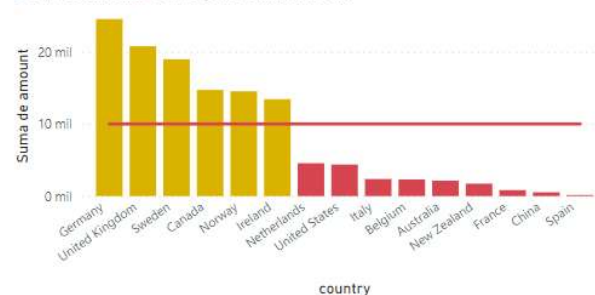
Transacciones rechazadas por mes (objetivo <10)

● Recuento de declined ● Max declined por mes



Ventas por país (objetivo > 10 mil)

● Suma de amount ● Valor objetivo ventas mensuales



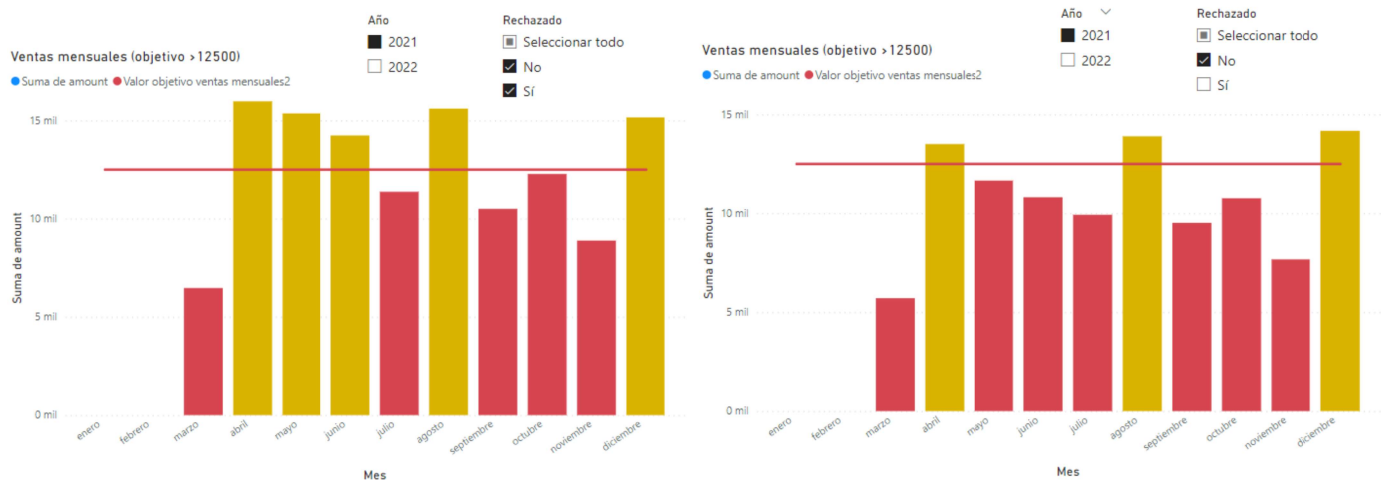
Nivel 2

Ejercicio 1

Desde el área de marketing necesitan examinar la tendencia mensual de las transacciones realizadas en 2021, específicamente, quieren conocer la variación de las transacciones en función del mes. Por tanto, te solicitan que visualices una KPI que incorpore la meta empresarial de conseguir al menos 12.500 transacciones por mes. En este ejercicio, será necesario que consigas identificar los meses en los que no se logró la meta establecida. Si es necesario puedes realizar dos visualizaciones.

Solución

A continuación, se muestran las ventas por mes para el 2021, con y sin incluir las rechazadas, diferenciando las que están por encima y por debajo del objetivo mediante formato condicional:



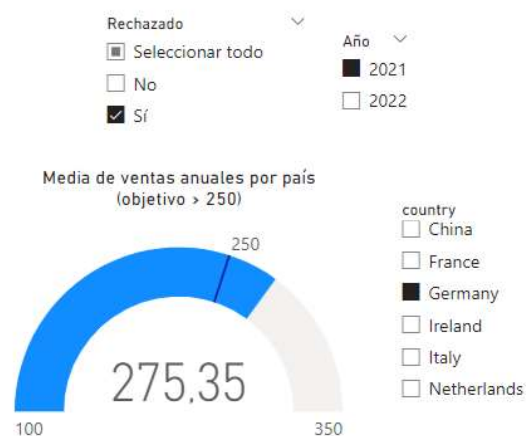
Ejercicio 2

En tu trabajo, se quiere profundizar en la comprensión de las transacciones realizadas en Alemania.

Por tanto, te solicitan que desarrolles medidas DAX para crear visualizaciones que destaquen la media de ventas en Alemania. Ten presente que la empresa tiene como objetivo conseguir una cifra de 250 euros anuales. Configura la visualización de forma que el valor mínimo sea 100 y el máximo 350, brindando así una representación más efectiva de la información.

Solución

Empleando DAX se generan los valores límite y objetivo.

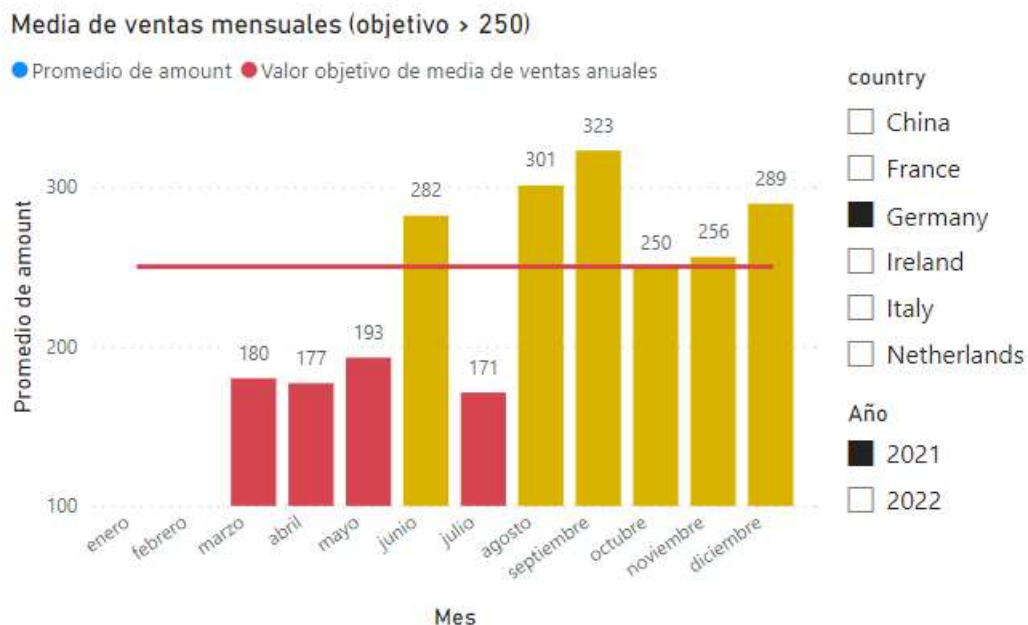


Ejercicio 3

Escribe un breve párrafo, máximo de 25 palabras, indica en qué mes no llegó a cumplirse con el objetivo propuesto del ejercicio 3 (2 ?).

Solución

Tras filtrar las transacciones rechazadas, se observa que marzo, abril, mayo y julio están muy alejados del objetivo de ventas medias anuales de 250 euros. Octubre y noviembre están en el límite.



Nivel 3

Ejercicio 1

La sección de marketing quiere profundizar en las transacciones realizadas por los usuarios y usuarias. En consecuencia, se te solicita la elaboración de diversas visualizaciones que incluyan:

- Las medidas estadísticas clave de las variables que consideres relevantes para comprender las transacciones realizadas por los usuarios/as.
- Cantidad de productos comprados por cada usuario/a.
- Media de ventas realizadas por usuario/a, visualiza qué usuarios/as tienen una media de ventas superior a 150 y cuáles no.
- Contabilizar el precio del producto más caro consumido por cada usuario/a.
- Visualiza la distribución geográfica de los usuarios/as.

En esta actividad, será necesario que realices los ajustes necesarios en cada gráfico para mejorar la legibilidad y comprensión. En el cumplimiento de esta tarea, se espera que evalúes cuidadosamente qué variables son relevantes para transmitir la información requerida de forma efectiva.

Solución

- Las medidas estadísticas clave de las variables que consideres relevantes para comprender las transacciones realizadas por los usuarios/as.

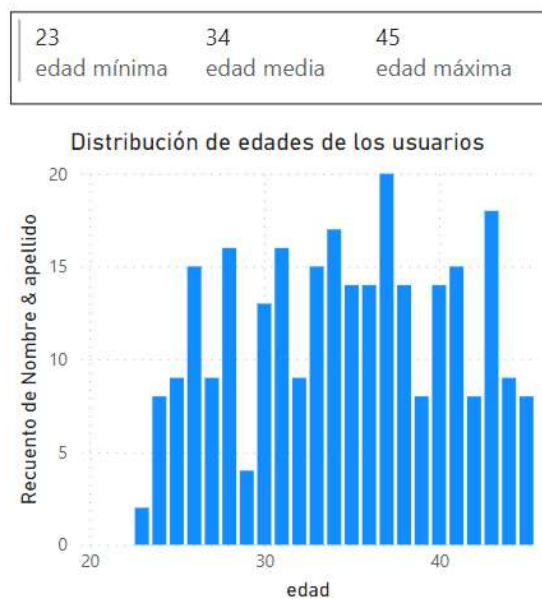
Un dato clave, aparte de los solicitados es la edad. Esta se obtiene como un campo calculado usando la siguiente fórmula:

```
edad = YEAR(TODAY()) - RIGHT([birth_date], 4)
```

También se generan valores para la edad mínima, media y máxima:

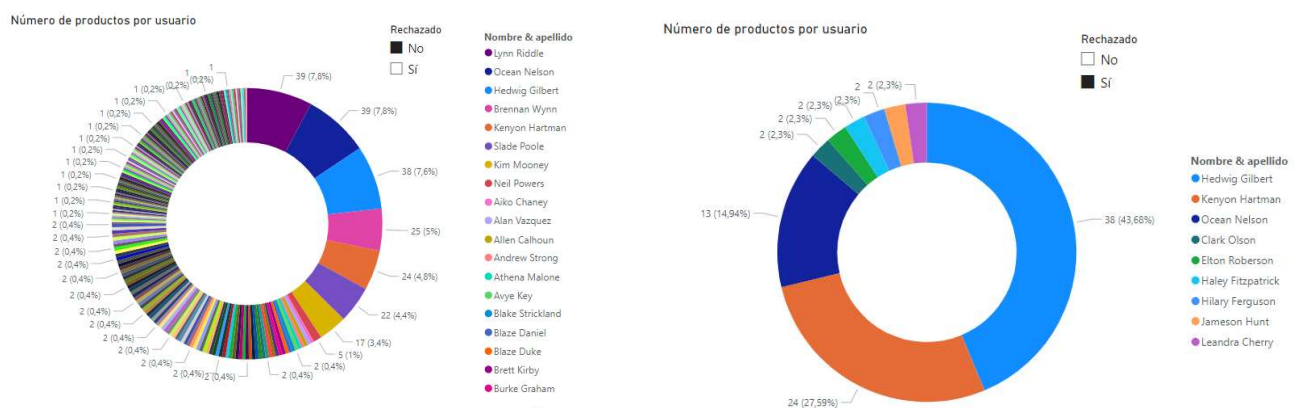
```
edad mínima = int(min('ventas users'[edad]))
edad media = int(average('ventas users'[edad]))
edad máxima = int(max('ventas users'[edad]))
```

A continuación, se muestran estos campos en una tarjeta de varias filas y también se crea una especie de histograma mediante un gráfico de barras con recuento de usuarios por valor de edad para ver la distribución por edades. No se observa una distribución uniforme debido a que los datos son ficticios.



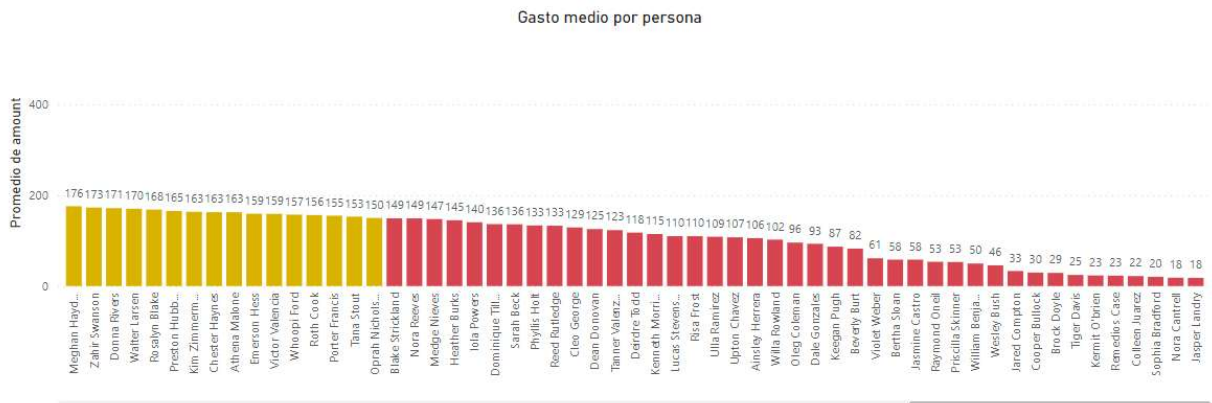
○ Cantidad de productos comprados por cada usuario/a.

Los gráficos de tipo donut representan con bastante claridad el número de compras aceptadas y rechazadas por cada usuario. La mayoría de las compras aceptadas son de 1 o 2 artículos. Hay 7 usuarios que al parecer son clientes fieles. Los usuarios que tienen productos rechazados son 9 y los tres de ellos que más compras rechazadas tienen son clientes fieles.



- Media de ventas realizadas por usuario/a, visualiza qué usuarios/as tienen una media de ventas superior a 150 y cuáles no.

Ordenando el gráfico de barras se observa una distribución de medias bastante uniforme sin grandes diferencias entre usuarios. El empleo del formato condicional permite diferenciar claramente los que tienen una media mayor de 150 USD de los que no:




- Contabilizar el precio del producto más caro consumido por cada usuario/a.

Para contabilizar el precio del producto más caro comprado por cada usuario/a conviene crear una nueva tabla, lo cual se puede hacer mediante Power Query, DAX o SQL

Procedimiento con Power Query:

Se duplica la tabla products

Se expande el vínculo a products_transactions y  luego a transactions

Se selecciona el campo Precio en USD creado anteriormente y user_id y se eliminan los demás campos.

Se agrega por id o nombre de usuario, buscando el máximo del precio.

El resultado de la última operación se muestra a continuación:

= Table.Sort(#"Filas agrupadas",{{"ventas.products_transactions.ventas.transactions.user_id", Order.Ascending}})		
1.2 Recuento	ventas.products_transactions.ventas.transactions.user_id	
1	null	196,65
2	60	195,94
3	61	114,09
4	62	167,2
5	63	26,51
6	64	167,2
7	65	63,33
8	66	167,2
9	67	91,89
10	68	167,2
11	69	141,01
12	70	49,7
13	71	141,01
14	72	195,94
15	73	139,59
16	74	59,8
17	75	141,01
18	76	139,59
19	77	172,78
20	78	171,22
21	79	195,94
22	80	132,86
23	81	161,11
24	82	171,22
25	83	169,96
26	84	132,86
27	85	91,89
28	86	172,78

Código DAX:

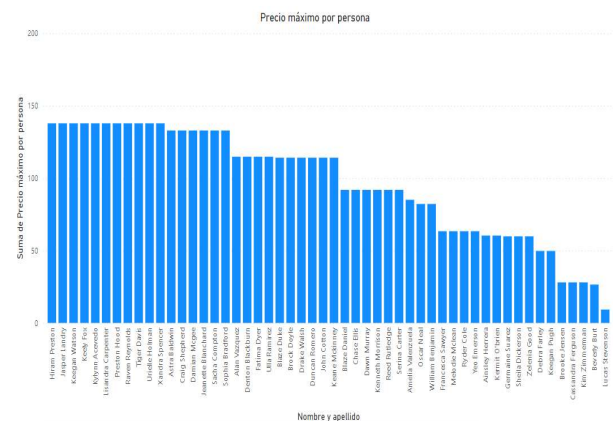
<pre> 1 ventas precio max persona = 2 3 SUMMARIZE(4 NATURALINNERJOIN(5 NATURALINNERJOIN(6 'ventas transactions', 7 'ventas products_transactions' 8), 9 'ventas products' 10), 11 'ventas users'[id], 'ventas users'[Nombre & apellido], 12 "Precio max", MAX('ventas products'[Precio en USD]) 13) 14 </pre>		
id	Precio max	Nombre & apellido
60	195,94	Roth Cook
61	114,09	Duncan Romero
62	167,2	Phyllis Holt
63	26,51	Beverly Burt
64	167,2	Irma Whitehead
65	63,33	Yeo Emerson
66	167,2	Bert Juarez
67	91,89	Kenneth Morrison
68	167,2	Whoopi Ford
69	141,01	Stone Atkinson
70	49,7	Debra Farley
71	141,01	Emerson Hess
72	195,94	Jael Robles
73	139,59	Warren Christian
74	59,8	Zelenia Good
75	141,01	Colleen Juarez

Código SQL:

```
62 • SELECT
63     user_id AS `ID usuario`,
64     MAX(CAST(SUBSTRING(price, 2, LENGTH(price) - 1) AS DECIMAL (10 , 2 ))) AS `Precio max`
65 FROM
66     transactions
67 JOIN
68     products_transactions ON products_transactions.transaction_id = transactions.id
69 JOIN
70     products ON products.id = products_transactions.product_id
71 GROUP BY user_id
72 ;
```

ID usuario	Precio max
60	195.94
61	114.09
62	167.20
63	26.51
64	167.20
65	63.33
66	167.20
67	91.89
68	167.20
69	141.01
70	49.70
71	141.01
72	195.94

A continuación se muestra un gráfico de barras con el precio máximo por id, así como un gráfico similar, pero con el nombre de persona en vez del id y precios en orden descendente:

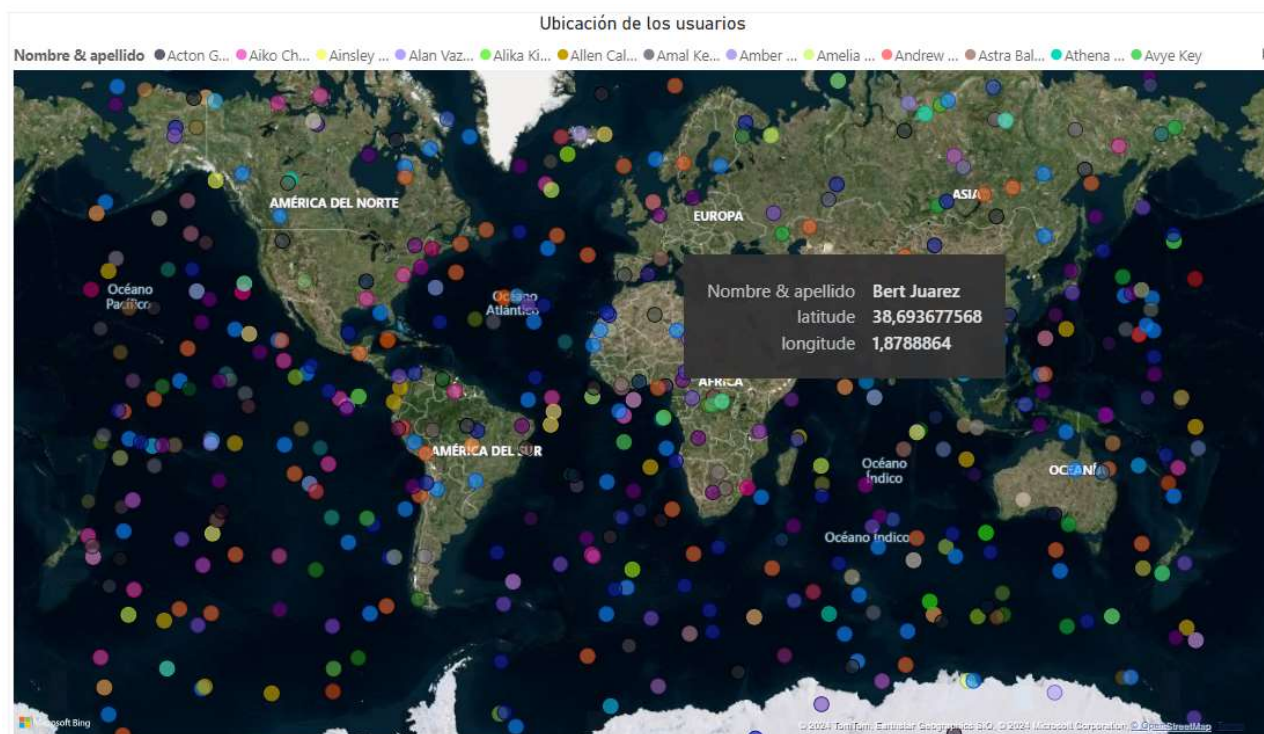


Se observa que hay unos pocos niveles de precios, de modo que se puede hacer un histograma mostrando cuantas personas se han permitido cada uno de esos precios. Curiosamente, el número de compradores aumenta con el precio:



○ Visualiza la distribución geográfica de los usuarios/as.

Para visualizar la distribución geográfica de los usuarios se puede usar el objeto mapa y las coordenadas de latitude y longitude de la tabla transactions. El hecho de que gran parte de los usuarios están repartidos por el océano indica que las coordenadas son ficticias.



Empleando códigos postales de la tabla users en lugar de coordenadas se logra un resultado más verosímil, si bien, tampoco cuadra con el hecho que los usuarios importados en la tabla users en teoría pertenecen solo a tres países (UK, USA, CAN), lo cual también se puede atribuir al hecho de que los datos son inventados:

Distribución geográfica de los usuarios

