# Tasca S8.02. Power BI amb Python

Aquesta tasca consisteix en l'elaboració d'un informe de Power BI, aprofitant les capacitats analítiques de Python. S'utilitzaran els scripts de Python creats prèviament en la Tasca I per a generar visualitzacions personalitzades amb les biblioteques Seaborn i Matplotlib. Aquestes visualitzacions seran integrades en l'informe de Power BI per a oferir una comprensió més profunda de la capacitat del llenguatge de programació en l'eina Power BI.

# Configuració Power BI i connexió amb script de Python

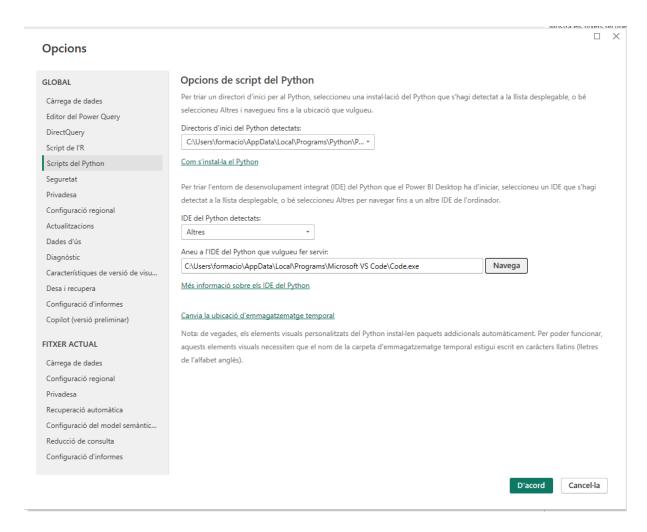
Nos aseguramos de tener las librerías pandas (para el manejo de dataframes), matplotlib y seaborn (para la creación de gráficos) en el equipo local.

En una instancia del terminal ejecuto el Python Launcher y le pido una lista de paquetes instalados mediante pip:

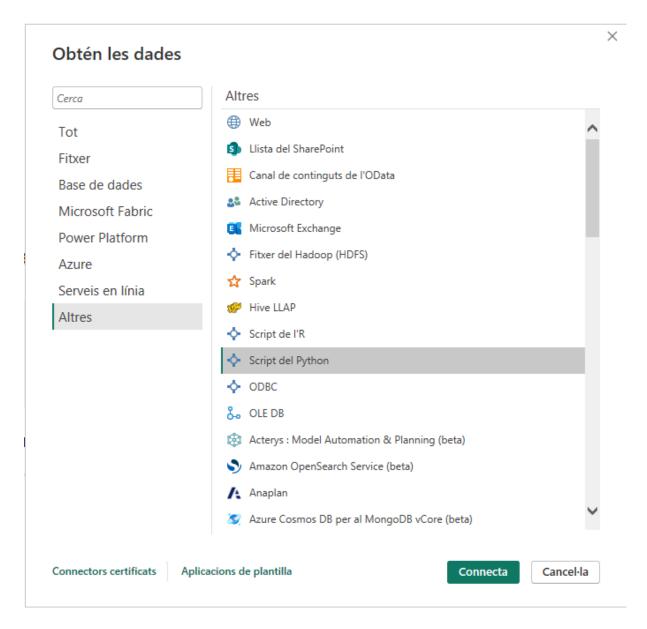
```
PS C:\Users\formacio\Documents\DataAnalyticsAlex> py --list
   -V:3.12 * Python 3.12 (64-bit)
PS C:\Users\formacio\Documents\DataAnalyticsAlex> py -m pip list
               Version
 asttokens 2.4.1 certifi 2024.12.14
 charset-normalizer 3.4.1 colorama 0.4.6 comm 0.2.2
contourpy 1.3.0
cycler 0.12.1
debugpy 1.8.5
decorator 5.1.1
et-xmlfile 1.1.0
executing 2.1.0
fontawesomefree 6.6.0
fonttools 4.53.1
greenlet 3.1.1
 highlight-text 0.2 idna 3.10 ipykernel
 ipykernel
ipython
                           6.29.5
                           8.27.0
 jedi 6.12
jupyter_client 8.6.2
jupyter_core 5.7.2
kiwisolver 1.4.7
3.10.0
 matplotlib-inline 0.1.7
 mysql-connector-python 9.1.0
 nest-asyncio 1.6.0
                              2.1.1
 numpy
 openpyx1
 packaging
                             24.1
 panda
                            0.3.1
 pandas
                             2.2.2
 parso
                             0.8.4
 pillow
                              10.4.0
```

pillow	10.4.0			
pip	25.0			
platformdirs	4.2.2			
prompt_toolkit	3.0.47			
psutil	6.0.0			
pure_eval	0.2.3			
Pygments	2.18.0			
pypalettes	0.1.5			
pyparsing	3.1.4			
python-dateutil	2.9.0.post0			
pytz	2024.1			
pywaffle	1.1.1			
pywin32	306			
pyzmq	26.2.0			
requests	2.32.3			
seaborn	0.13.2			
setuptools	75.8.0			
six	1.16.0			
SQLAlchemy	2.0.37			
squarify	0.4.4			
stack-data	0.6.3			
tornado	6.4.1			
traitlets	5.14.3			
typing_extensions	4.12.2			
tzdata	2024.1			
urllib3	2.3.0			
wcwidth	0.2.13			
PS C:\Users\formacio\Documents\DataAnalyticsAlex>				

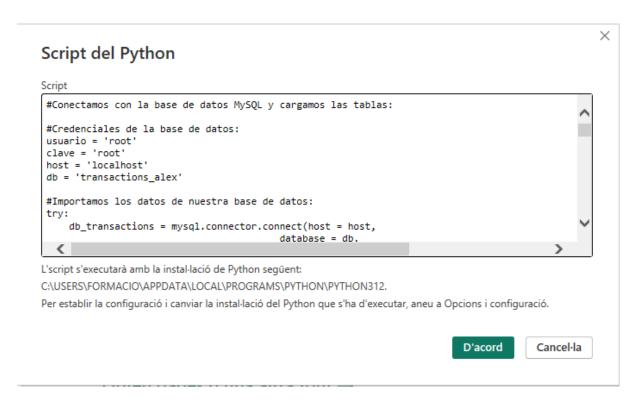
En Power BI, habilitamos la ejecución de scripts de Python:



Escogeremos la opción Obtener datos / Script de Python:

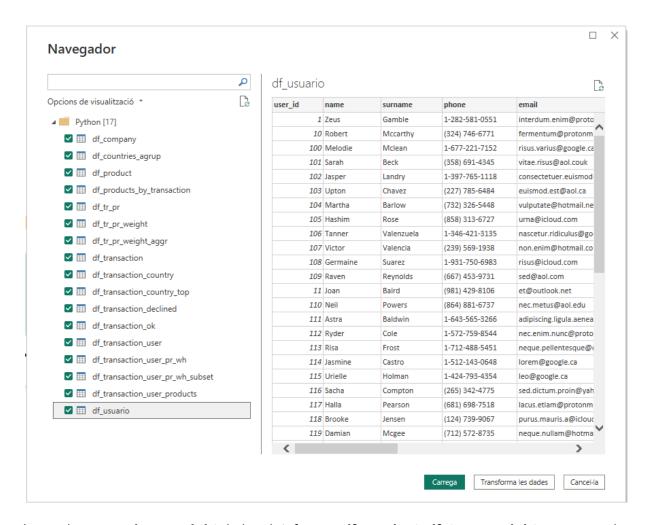


Y en el cuadro habilitado para el script cargaremos los dataframes que hemos creado en el script de la TascaS801, tanto los creados a partir de la base de datos **transaction\_alex** que importamos a través de MySQL Connector para Python como los dataframes creados a partir de la función merge() de la librería pandas con las transformaciones pertinentes.



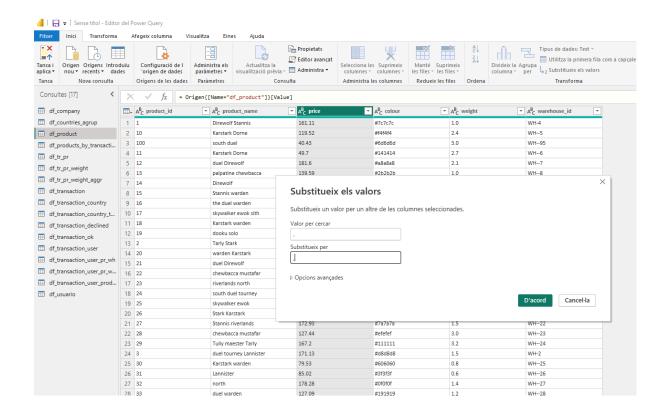
(Se puede ver el script completo en la primera celda del archivo TascaS802.ipynb.)

Una vez cargados los dataframes, pasamos a revisar los datos y realizar las transformaciones pertinentes:



Las columnas price y weight de los dataframes df\_product, df\_tr\_pr\_weight aparecen sin el separador de decimales, así como la columna amount en los dataframes df\_transaction, df\_transaction\_country, df\_transaction\_country\_top, df\_transaction\_ok, df\_transaction\_user, df\_transaction\_user\_pr\_wh y df\_transaction\_user\_pr\_wh\_subset. En estos dos últimos dataframes, fruto de la unión con la tabla product, vemos que el resultado de la agregación multiplica los resultados en el script, a causa de la desaparición del separador de decimales.

Procedemos a transformar estos datos antes de cargar los dataframes. Por ejemplo, en el caso de la columna **price**, eliminamos la transformación automática de texto a número decimal, sustituimos el signo '.' por ',' y transformamos a decimal fijo. También eliminaremos la conversión de algunos id que Power BI ha transformado de string a número entero.



# Nivell 1

# - Exercici 1

Els 7 exercicis del nivell 1 de la tasca 01

En las representaciones de las gráficas de los tres niveles seguiremos el mismo mecanismo: seleccionamos **Elemento visual de Python**, seleccionamos en la columna **Datos** el dataframe que contiene la(s) columna(s) con los datos que queremos graficar, los añadimos al campo **Datos** de la pestaña **Compilación** y, a continuación, copiamos y adaptamos el script correspondiente del archivo **Tasca S801.ipynb** en el recuadro inferior para generar el gráfico en Power BI.

En los siguientes apartados incluiré solo las acciones adicionales al mecanismo descrito aquí arriba. Los scripts adaptados se pueden consultar en el archivo **Tasca S802.ipynb** 

Nota: como aprovechaba la característica de Visual Studio Code de interpretar los objetos de matplotlib y seaborn, a los scripts les añado las instrucciones **plt.tight\_layout()** para evitar que la gráfica aparezca cortada en el contenedor de Power BI, y **plt.show()** para poder representarlos en Power BI.

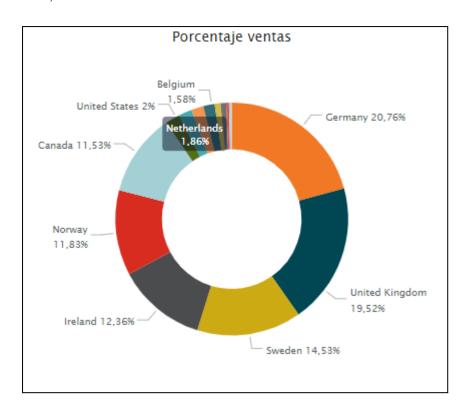
# Ejercicio 1. Gráfica con una variable numérica

He sobreimpresionado la leyenda sobre el gráfico para mejorar la escala.

L'empresa està interessada a obtenir una visió general de les transaccions realitzades per cada país. La teva tasca és crear una visualització que identifiqui el percentatge de les vendes per país.

Usamos la métrica **Ventas totales**, con la que trazaremos un gráfico de anillos con el porcentaje de ventas de cada país:

```
Ventas totales = CALCULATE(
    SUM('transaction'[amount]),
    'transaction'[declined]=false
)
```



Este gráfico nos confirma que nuestros mercados más importantes están en Europa, con Alemania, Reino Unido y Suecia, que abarcan más de la mitad del pastel, como los mercados más importantes; más Noruega e Irlanda, que, junto con Canadá, se sitúan alrededor del 12 % de la cuota total.

A la luz de este gráfico deducimos que los esfuerzos para aumentar el promedio de compras de Canadá se antoja más importante que el retorno que nos pueda proporcionar las acciones para mercados como el estadounidense y el chino que comentábamos en el apartado anterior.

Dissenya un indicador visual en Power BI per a analitzar la diferència de vendes entre els anys 2022 i 2021 en cada país. L'empresa està interessada a comprendre com han variat les vendes en diferents països durant aquest període i desitja identificar qualsevol disminució o augment significatiu en les vendes.

Como se indicaba anteriormente, solo tenemos datos de ventas hasta marzo del 2022. Lógicamente, los resultados por país van a salir negativos, aunque podemos identificar algunas tendencias.

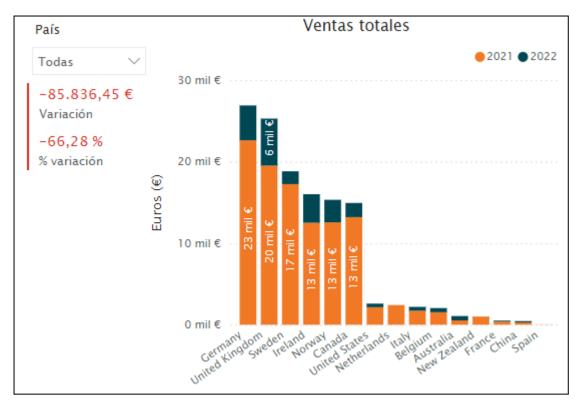
Primero, crearemos las siguientes métricas para calcular la diferencia entre años y el porcentaje de variación de las ventas:

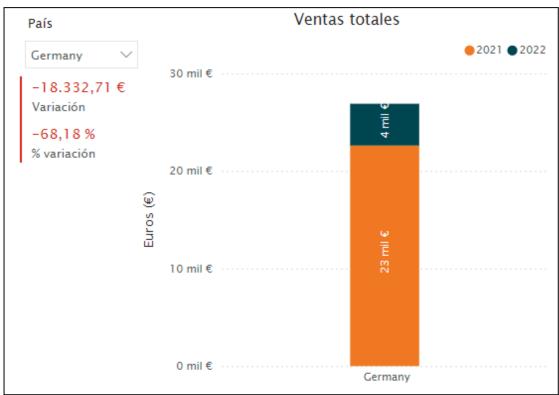
```
Ventas t 2021 = CALCULATE(
    SUM('transaction'[amount]),
    'transaction'[declined]=false,
    'calendar'[year] = 2021
)

Ventas t 2022 = CALCULATE(
    SUM('transaction'[amount]),
    'transaction'[declined]=false,
    'calendar'[year] = 2022
)

% Ventas 22-21 = DIVIDE([Ventas t 2022] - [Ventas t 2021],[Ventas totales])
```

Se incluye un segmentador por país, para poder analizar individualmente la evolución de las ventas.





Se puede observar un buen inicio de año de los mercados británico y alemán, también el irlandés, no así el canadiense, con una variación del -76 %, ni, sobre todo, el sueco, con un -83 %, en el que habría que incidir con acciones comerciales para impulsar las ventas, dada la importancia de estos mercados en el total de ingresos.

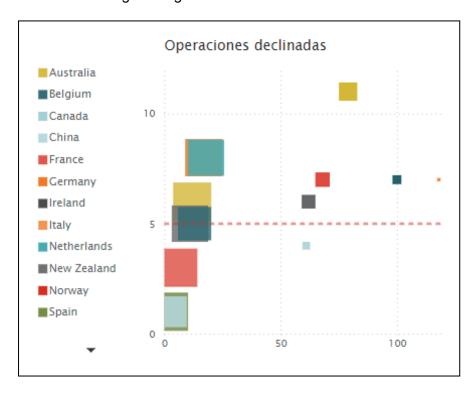
Crea una visualització en la qual es pugui comptabilitzar el nombre de transaccions rebutjades en cada país per a mesurar l'eficàcia de les operacions. Recorda que l'empresa espera tenir menys de 5 transaccions rebutjades per país.

Usamos un gráfico de dispersión para poder comparar, en un solo visual, la cantidad de transacciones declinadas de cada país en comparación con las transacciones totales y su incidencia en el total de transacciones. Usamos la métrica que introdujimos en el sprint anterior:

Y creamos otra para calcular el porcentaje de transacciones declinadas respecto al total:

```
% trans rech = DIVIDE([Declinadas], [Num transacciones totales],0)
```

#### Obtenemos el siguiente gráfico:

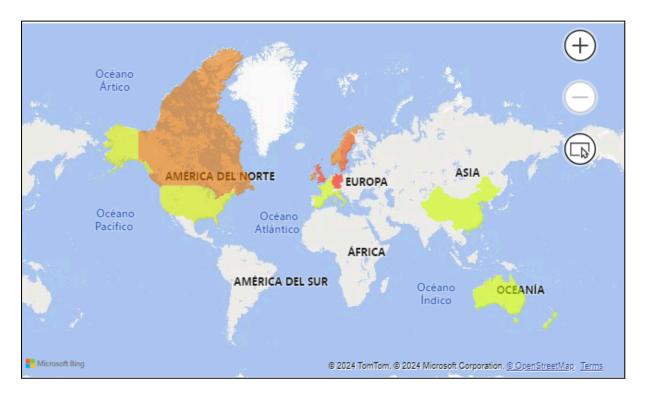


En el gráfico destaca Suecia, que, siendo uno de los mercados importantes por volumen de negocio y promedio de compras, tiene una alta tasa de transacciones rechazadas. Habría que estudiar quiénes son los usuarios que compran para empresas de Suecia y analizar las causas. Por otra parte, señalaría que la alta tasa de cancelaciones respecto al total de operaciones en países como Holanda, Italia, Estados Unidos y Australia afecta negativamente a la implantación en estos mercados, reduciendo dramáticamente su impacto en el total.

Aunque, por nivel de ingresos y cantidad de operaciones, el porcentaje de operaciones canceladas en nuestros mercados más importantes (Alemania, Reino Unido, Noruega e Irlanda) no afecte tanto a nivel global, nuestro objetivo es reducirlo por debajo del máximo establecido y acercarlos al nivel de Canadá. Sería interesante comprobar si estas cancelaciones se producen por algún agente en concreto.

L'empresa busca comprendre la distribució geogràfica de les vendes per a identificar patrons i oportunitats específiques en cada regió. Selecciona la millor visualització per a mostrar aquesta informació.

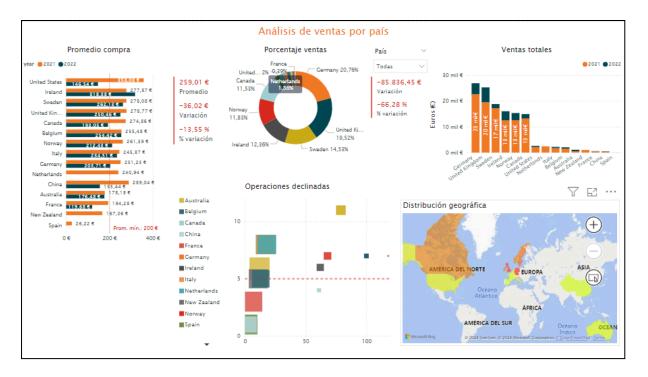
Tomaremos la medida **Ventas totales** y la visualizaremos en el objeto visual mapa de Power BI.



Comprobamos que los países con más peso son Alemania, Reino Unido, Suecia, Noruega, Irlanda (centro y norte de Europa e islas británicas) y Canadá.

El teu cap t'ha demanat preparar una presentació per al teu equip en la qual es detallin la informació de tots els gràfics visualitzats fins ara. Per a complir amb aquesta sol·licitud, has de proporcionar una interpretació de les visualitzacions obtingudes. La presentació pot realitzar-se amb la informació general o seleccionant un element en particular, com per exemple, els resultats d'Espanya.

#### Vamos a centrarnos en Alemania:



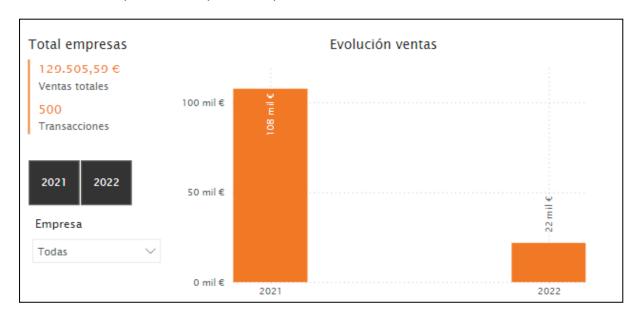
Con cerca del 21 % de cuota de mercado, Alemania lidera tanto en ingresos como en número de operaciones, y mantiene una excelente ratio de operaciones aprobadas y baja tasa de operaciones rechazadas. Aunque cumple con el promedio de compra mínimo, interesaría emprender acciones de márketing para aumentar ese promedio y acercarlo al nivel de Suecia o Irlanda.

# Nivell 2

# Exercici 1

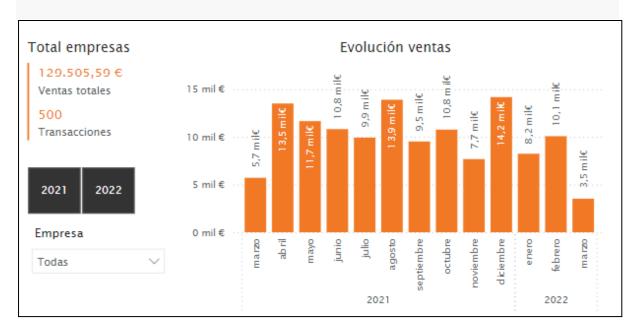
La teva tasca consisteix a implementar un filtre interactiu que permeti seleccionar les vendes per a cada any.

En la esquina superior fijaremos los datos de ventas y transacciones totales. Incluimos una segmentación por año, permitiendo también seleccionar todos los años. Creamos un visual con el total de ventas por año y por mes, y otra segmentación adicional para poder analizar los datos de cada empresa. Mediante la navegación por jerarquía podemos seleccionar qué nivel de detalle queremos, si por año o por mes.



### Exercici 2

La gerència està interessada a analitzar més a fons les vendes en relació amb el mes. Per tant, et demanen que facis els ajustos necessaris per a mostrar la informació d'aquesta manera.

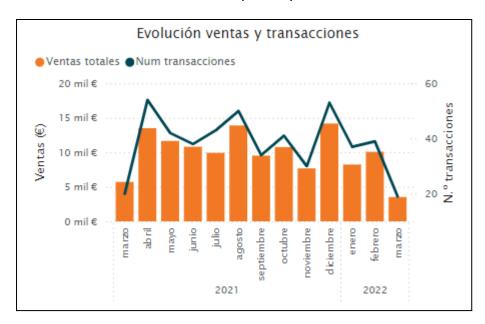


Si proyectamos los ingresos de los tres últimos trimestres del 2021 para el total de empresas, aquí ya podemos suponer que, si mantenemos la tendencia del año anterior, alcanzaremos sin problema los ingresos del 2022.

### Exercici 3

Visualitza el total de vendes i la quantitat de transaccions realitzades. Si és necessari, pots crear dues visualitzacions separades.

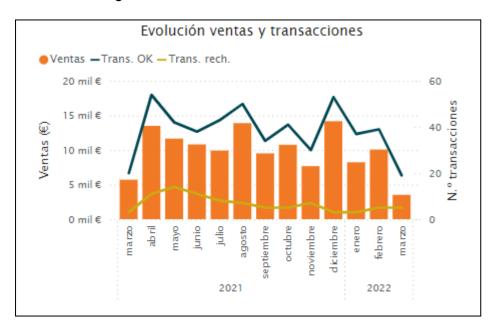
Modificamos el gráfico anterior para incorporar un gráfico de líneas que nos muestre la evolución de las transacciones aceptadas por mes.



### Exercici 4

Crea una visualització que permeti observar de manera efectiva i clara la quantitat de les vendes realitzades i la quantitat de transaccions rebutjades.

Incluimos en el gráfico anterior la métrica **Núm. transacciones rechazadas**:

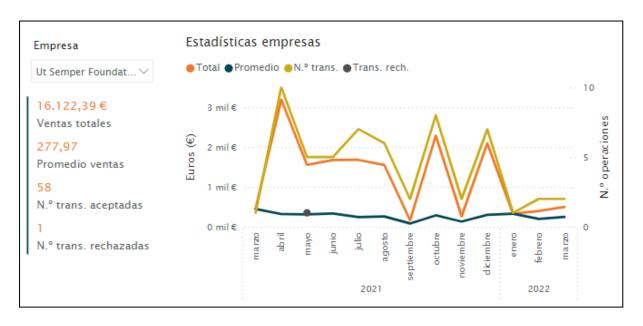


Vemos una relación entre la caída de ventas de mayo y junio y el incremento de transacciones rechazadas en esos dos meses, que subraya la importancia de combatir las posibles causas de estas incidencias para mejorar los resultados.

### Exercici 5

Selecciona una visualització en la qual es reflecteixi els estadístics descriptius de les empreses que van realitzar transaccions. Recorda mostrar el total de cada estadístic.

Representamos los estadísticos descriptivos en un gráfico de líneas para comparar las ventas, el promedio, el número de transacciones y de transacciones rechazadas por periodo temporal, que también mostraremos en sendas tarjetas que podremos comparar con los datos fijados en el ejercicio 1.



Para poder analizar con mayor profundidad los resultados por empresa, añadimos la siguiente tabla:

Empresa	País	Total €	Trans.	Rech.	% rech.	Compra media
Nunc Interdum Incorporated	Germany	25.266,56€	105	1	0,95 %	242,95€
Ut Semper Foundation	Sweden	16.122,39€	59	1	1,69 %	277,97€
Enim Condimentum Ltd	United Kingdom	14.578,03€	57	1	1,75 %	260,32€
Arcu LLP	Norway	13.762,79€	56	1	1,79 %	250,23€
Lorem Eu Incorporated	Canada	13.916,44€	54	1	1,85 %	262,57€
Malesuada PC	Ireland	14.885,80€	52	1	1,92 %	291,88€
Non Institute	United Kingdom	8.911,24€	30	0		297,04€
A Institute	Belgium	266,09€	2	1	50,00 %	266,09€
Ac Fermentum Incorporated	Germany	293,57€	2	1	50,00 %	293,57€
Ac Industries	Germany	396,15€	2	1	50,00 %	396,15€
Ac Libero Inc.	United Kingdom	30,76€	2	1	50,00 %	30,76€
Total		129.505,59	587	87	14,82 %	259,01 €

Identificamos a Nunc Interdum Incorporated como nuestro mejor cliente en global y que es quien realiza casi todas las compras en Alemania. Dentro de nuestro ránking, las siete empresas más importantes, tanto en volumen de ventas como de operaciones, representan el 83 % de nuestros ingresos.

# Nivell 3

### Exercici 1

En la teva empresa, volen aprofundir en l'anàlisi de les característiques dels usuaris que participen en les transaccions, així com en els productes venuts. T'han demanat que creïs visualitzacions rellevants per a millorar estratègicament les campanyes publicitàries i augmentar les vendes. Les visualitzacions que has d'incloure són les següents:

- o Informació personal dels usuaris/es.
- Quantitat de transaccions realitzades i rebutjades. L'empresa espera que cada usuari/ària tingui almenys 10 transaccions per any, i que tinguin menys de 2 transaccions rebutjades per any.
- o Identificació del producte més barat i més car comprat per cada usuari/ària, juntament amb el seu preu.
- o Distribució geogràfica dels usuaris/es.
- o Mitjana de compres realitzades.
- L'usuari/ària ha de tenir l'opció de seleccionar si desitja mirar la informació d'un any únicament.

Després de crear els gràfics, has de presentar la informació de l'usuari/ària amb l'ID 96 amb una breu descripció de les dades a través d'una presentació de diapositives. Assegura't d'optimitzar la llegibilitat i comprensió de les visualitzacions mitjançant ajustos adequats.

Para ayudarnos con las gráficas y tablas, crearemos las siguientes medidas:

Número de transacciones rechazadas:

```
Num transacciones rechazadas =
var rech = CALCULATE(
     DISTINCTCOUNT('transaction'[id]),
    'transaction'[declined] = true
    )
var total = CALCULATE(DISTINCTCOUNT('transaction'[id]))
RETURN
IF(ISBLANK(rech), IF(ISNUMBER(total), 0), rech)
```

Para sustituir los valores nulos en la columna **Transacciones rechazadas** por ceros, en la fórmula crearemos una variable **rech** donde almacenaremos el resultado de la consulta DISTINCTCOUNT. En el caso que nos devuelva un valor nulo, validaremos que el número de transacciones del usuario sea diferente de cero; en ese caso, sustituiremos el valor null por cero. De no comprobar que el usuario no ha realizado transacciones, nos aparecerían dichos usuarios con un valor nulo en **Transacciones realizadas** y cero en **Transacciones rechazadas**.

Cálculo del precio del producto más barato adquirido por el cliente:

```
Min precio unidad = CALCULATE(
    MIN('transaction_products'[product.price]),
    'transaction'[declined]=false
)
```

#### Obtención de los datos de los artículos correspondientes:

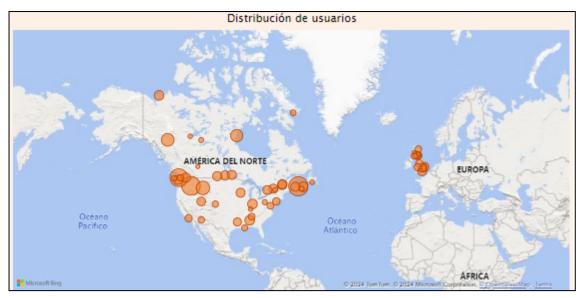
Crearemos una tabla donde listaremos los usuarios con transacciones, el número de transacciones (en verde los que superan el mínimo de transacciones por año, y en rojo los que no), el número de transacciones rechazadas (marcamos con iconos si tienen menos de dos, dos o más de dos), el número de unidades compradas en total, el promedio de la compra del usuario, y el valor y datos de los artículos más caros y más baratos adquiridos por cada usuario.





Mapa de distribución de usuarios correspondiente al 2021, donde comprobamos que Hedwig Gilbert, al norte de Canadá, es también uno de los usuarios que más ingresos aporta a nuestra empresa.





Mapa de distribución de usuarios del 2022. Quien más ingresos ha aportado hasta ahora es también el canadiense Ocean Nelson desde Charlottetown.

Hedwig Gilbert, del norte de Canadá, es el usuario más importante del 2021 y uno de los tres más importantes del 2022 en cuanto a número de transacciones; pero, a su vez, es uno de los que más incidencias sufre en las transacciones, con casi la mitad de transacciones del 2021 rechazadas y seis de nueve del 2022.

Creamos el panel para el usuario que responde con el id 96:

