

SPRINT #5

POWER BI: INTRODUCTION AND INDICATORS

student:
Leo Kalugin



Date: 07/11/2024

SUMMARY

In this sprint, I am exploring key concepts such as data transformation and loading, creating key performance indicators, Data Analysis Expressions (DAX), developing a comprehensive understanding and practical application of specific Power BI concepts, acquiring the skills and knowledge necessary for maximally effective utilization of the platform in various analytical scenarios.

RESULT

In this folder on the GitHub repository, you will find:

/db_data: csv-files with data and equivalent database dump

S5_01.pbix: dashboard in pbix

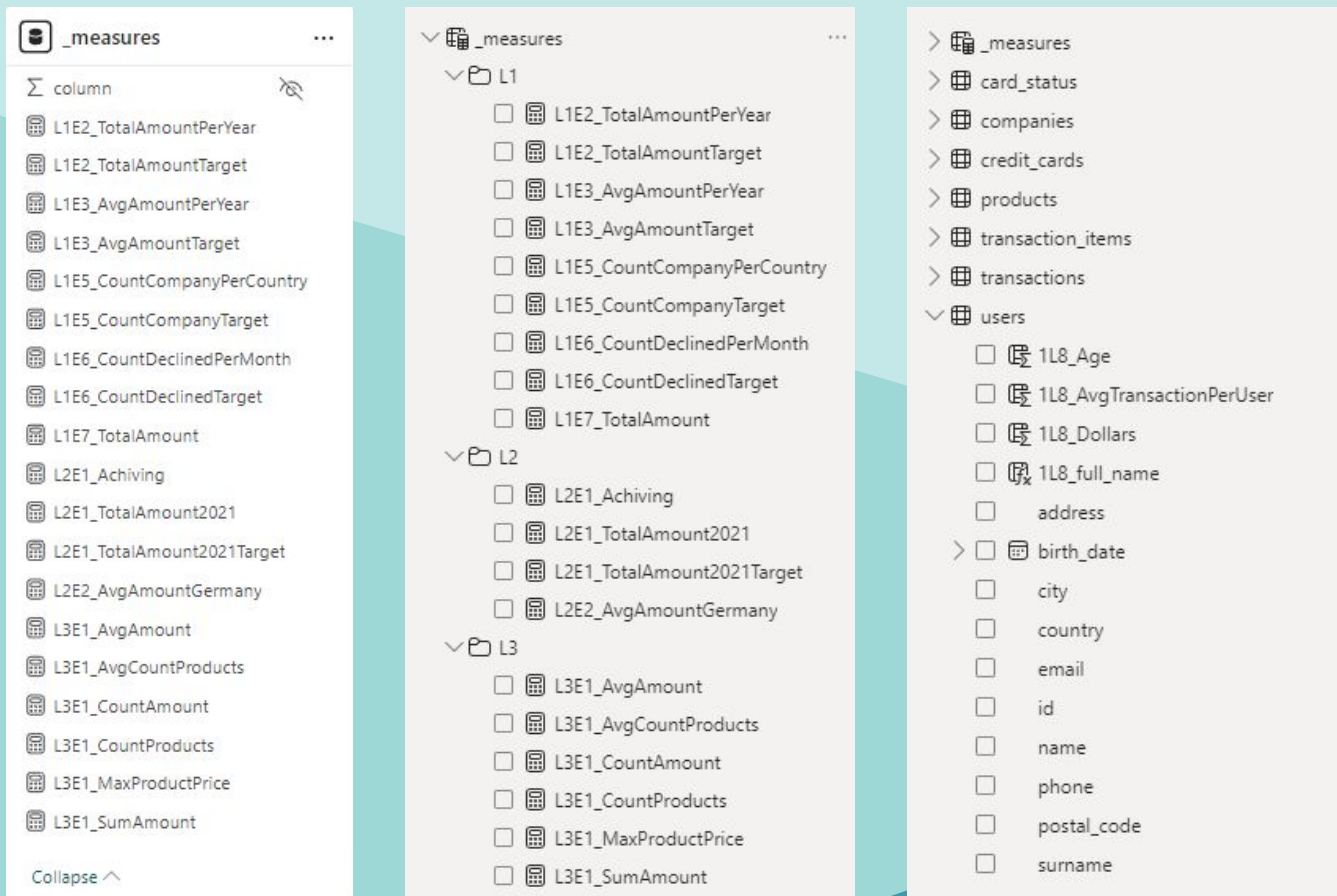
S5_01.pdf: dashboard in pdf

Sprint_5.pdf: interpretation of exercises and DAX formulas and descriptions with screenshots of all visuals and their results with comments

https://github.com/leocareer/DA_specialization/tree/main/Sprint_05

I analyzed all the tasks and modeled a dashboard from them. For each level, I made a separate page, three pages in total. For the first and second levels, the visuals of each task are organized into a separate block. I added visuals and dependencies between the blocks that I considered optimal in each case. Before each new level, I will show the corresponding dashboard page with a description.

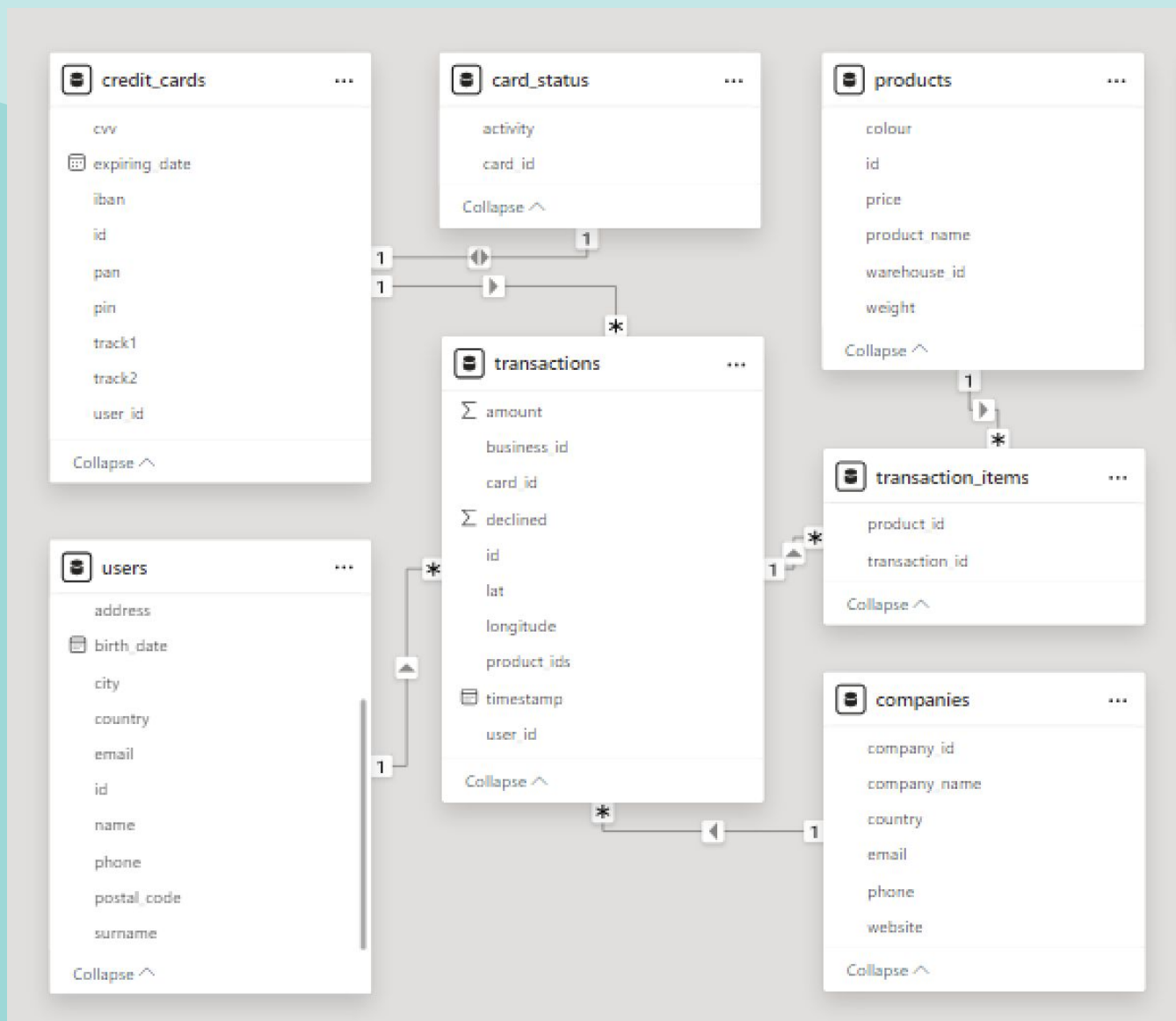
All the measures I created are in a separate table called 'measures' and are grouped into folders that correspond to levels (dashboard pages), the name of each measure consists of the level number and the task number. In addition, the 'users' table contains several calculated columns for level 1 exercise 8:



LEVEL 1 EXERCISE 1

It is necessary to import data from the previously used data base. After loading data, the database model is displayed in Power BI.

I used a connection between MySQL Workbench and Power BI to load the data. In the sprint folder on Git I also posted the equivalent csv-files and a dump of the database used (the '/db_data' folder). The base model in Power BI:

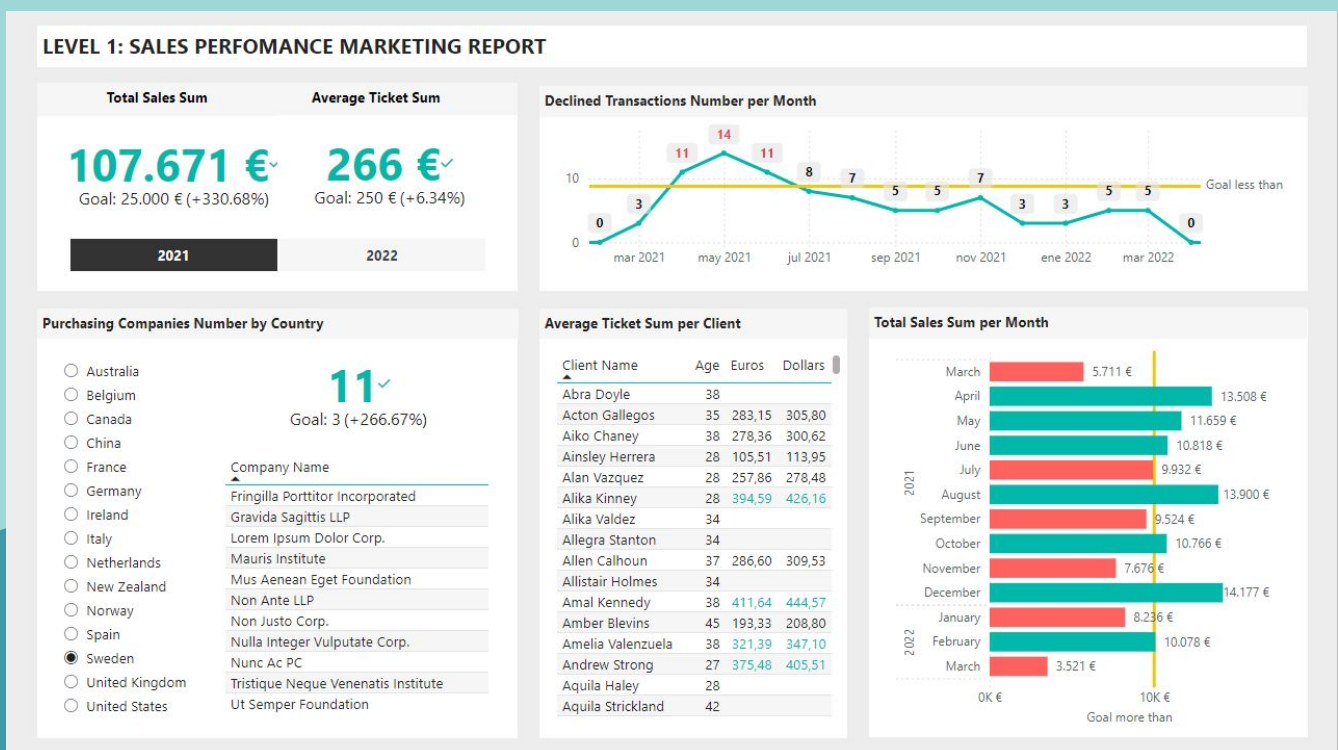


The database diagram is exactly the same as in the previous sprint, here I will skip the description of fields and relationships, it can be seen here:

github.com/leocareer/DA_specialization/blob/main/Sprint_04/Sprint_4.pdf

As for data types, I converted all the columns I needed to decimal or whole number, taking into account the need for rounding. I checked the date format, I decided not to create a separate table for the calendar, since I analyzed that the 'timestamp' field in the 'transactions' table in the 'Data Hierarchy' format would satisfy me. I also needed to convert the data for 'lat' and 'longitude', since some commas were lost when loading the data.

The first level and the first page of the dashboard. There are no common dependencies here and each block works separately. I considered the option of making a common filter by year and by country, and decided that it would be unnecessary and defocusing. Firstly, it was not requested. Secondly, the tasks are not related to each other, in fact, these are separate reports for different stakeholders, and common filters would create the illusion of connectivity. Thirdly, a filter by period would not bring good profit since all months and years are already presented on the charts and we only have 13 months. The third level will be presented as a single report and I practice with common filters there.



I analyzed the following three tasks, they ask for metrics that stakeholders often want to see at the same time, and also they are asked for the same period, and also the same 'card' visual suits them, so I decided to combine them into one block. Now I will give the three tasks for which this block was created, then I will give a description of the formulas and an analysis of the result.

LEVEL 1 EXERCISE 2

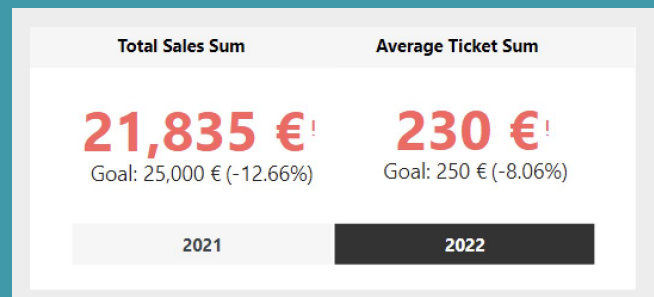
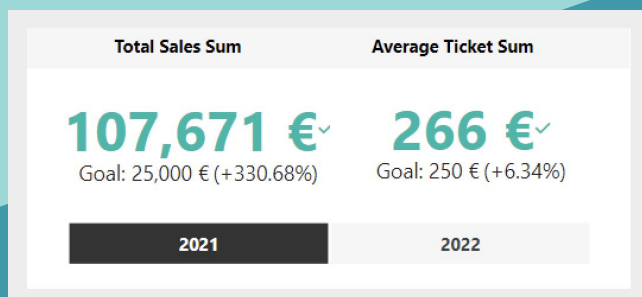
Your company is interested in evaluating the total sum of transactions carried out over the years. To achieve this, the creation of a key performance indicator (KPI) has been requested. The KPI must provide a clear visualization of the business objective of achieving a total sum of €25,000 for each year.

LEVEL 1 EXERCISE 3

Marketing asks you to create a new DAX measure that calculates the average sum of the transactions made during the year 2021. Visualize this average in a meter that reflects the sales made, remember that the company has a target of 250.

LEVEL 1 EXERCISE 4

Follow the same procedure as in Exercise 3 for 2022.



Over the past 2021, the goal for total sales was exceeded by 330%, and will obviously be achieved in the current 2022 in the first or second quarter, our company is doing an excellent job with this indicator, but I would also like to ask questions whether our goal is underestimated, probably It's time for the company to reconsider it, given that before this I was convinced that all departments had the correct data. The results of the average ticket look more realistic, but at the moment the company is not managing to achieve the figure of 250 euros, I would like to conduct an analysis to compare the current readings with the first quarter of last year – this is a period of low demand in many niches and perhaps we are reaching the annual figure in other quarters, I would also like to check for outliers in the data, perhaps we recently had a sale that could have affected the average check amount.

To create these visuals, I added measures that count the sum and average, and measures to achieve goals. I added a year selection in DAX. I set transactions[declined] = 0 to only count sales for sales that are not rejected, because obviously we only care about actual sales for which the company has received money.

```
1 L1E2_TotalAmountPerYear = CALCULATE(  
2 |   SUM(transactions[amount]),  
3 |   YEAR(transactions[timestamp]) = SELECTEDVALUE(transactions[timestamp].[Year]),  
4 |   transactions[declined] = 0  
5 | )
```

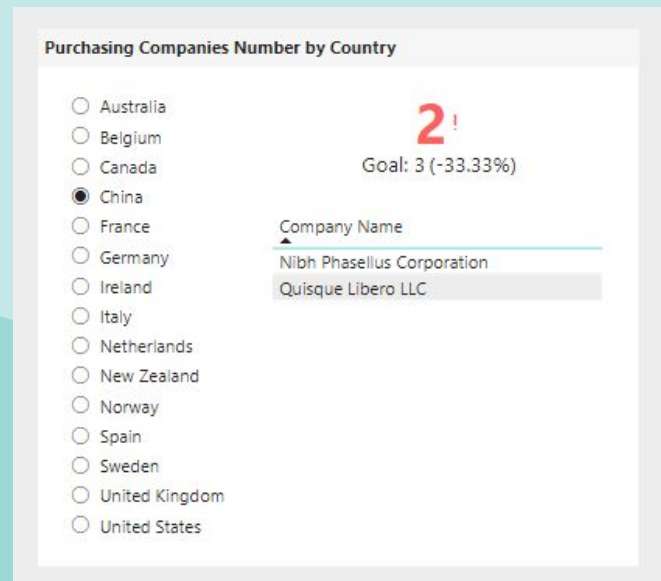
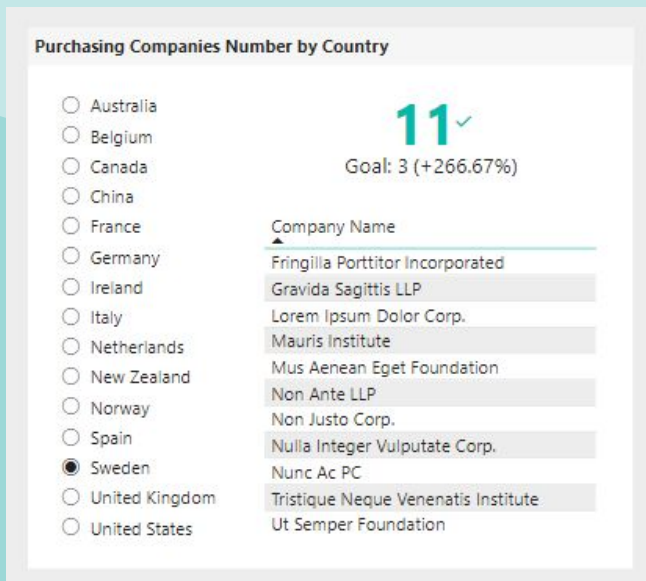
```
1 L1E2_TotalAmountTarget = 25000
```

```
1 L1E3_AvgAmountPerYear = CALCULATE(  
2 |   AVERAGE(transactions[amount]),  
3 |   YEAR(transactions[timestamp]) = SELECTEDVALUE(transactions[timestamp].[Year]),  
4 |   transactions[declined] = 0  
5 | )
```

```
1 L1E3_AvgAmountTarget = 250
```

LEVEL 1 EXERCISE 5

The objective of this exercise is to create a KPI that visualizes the number of companies per country that participate in transactions. The business goal is to ensure that there are at least 3 participating companies per country. To achieve this, it will be necessary to use DAX to calculate and represent this information in a clear and concise way.



To achieve the target of having at least 3 participating companies from each country, I recommend focusing efforts on attracting partners from China and Spain, where the current numbers fall short of the minimum target of 3 companies. Meanwhile, the success in Sweden demonstrates the country's potential for further collaboration and offers an example of successful participant engagement. The list of country names is added to the visual to clearly show which countries have met the participation target and where additional company engagement is needed.

To create these visuals, I added measure that counts the number of companies[company_id] with a country selection in DAX, and measure to achieve goals equal to 3 companies:

```
1 L1E5_CountCompanyPerCountry = CALCULATE(
2     COUNT(companies[company_id]),
3     companies[country] = SELECTEDVALUE(companies[country])
4 )
```

```
1 L1E5_CountCompanyTarget = 3
```

LEVEL 1 EXERCISE 6

Create a new KPI that allows you to visualize the number of declined transactions over time. The company has set a goal of having fewer than 10 declined transactions per month.

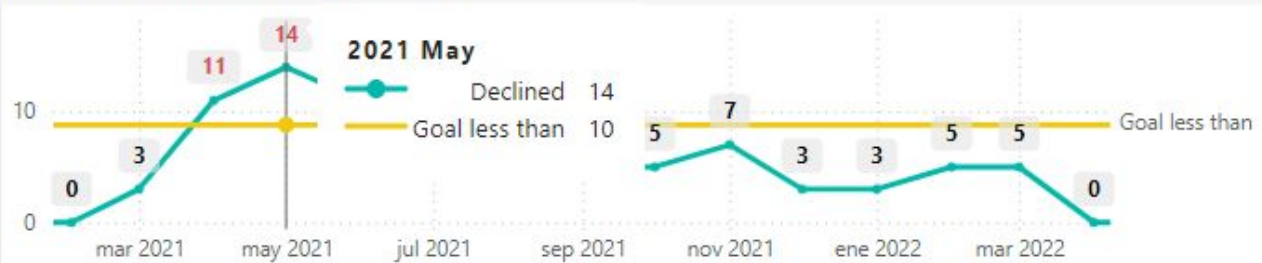
Declined Transactions Number per Month



Here are the data for all months in which there were sales. We see three periods when the company exceeded the limit, but this was in the second quarter of the previous year and the trend is decreasing, so in the current year the company does not exceed even half of the limit, which means that measures to reduce rejected transactions were effective.

To create this view I added a measure that counts the number of transaction[id] when declined = 1 with a month selection in DAX, it was important to specify [MonthNo] and not [Month]. In this measure I use the COALESCE function, as I wanted to display the zero boundary periods on the chart, so that it would be clear that this is the entire existing period. The second measure I added is the target of 10 transactions. When you hover over the chart, a hint with the current value is shown:

Declined Transactions Number per Month



```

1 L1E6_CountDeclinedPerMonth = COALESCE(
2   CALCULATE(
3     COUNT(transactions[id]),
4     transactions[declined] = 1,
5     MONTH(transactions[timestamp]) = SELECTEDVALUE(transactions[timestamp].[MonthNo])
6   ),
7   0
8 )

```

```

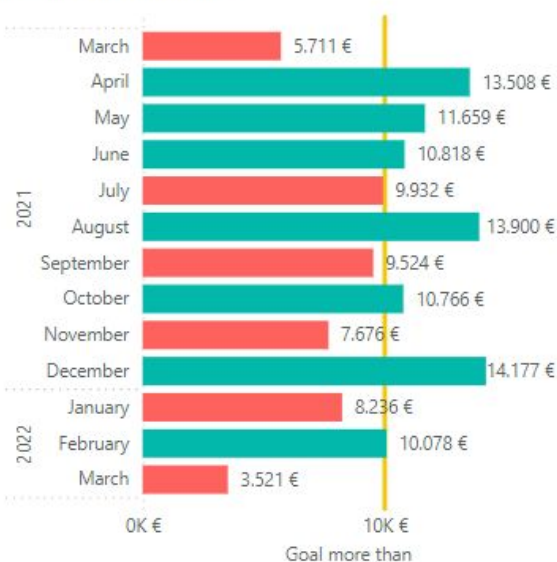
1 L1E6_CountDeclinedTarget = 10

```

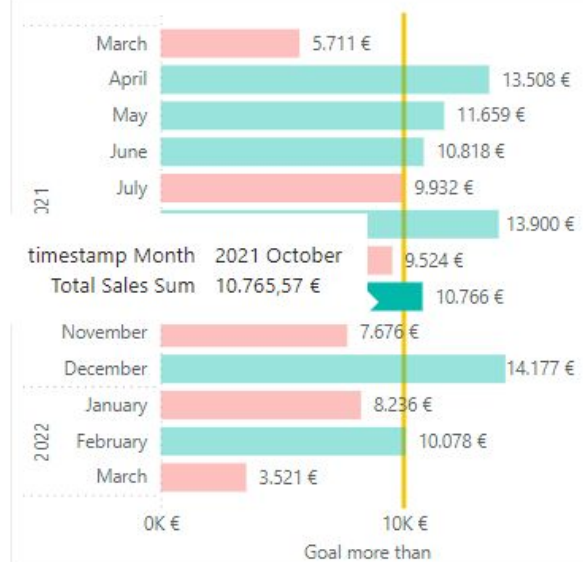
LEVEL 1 EXERCISE 7

Create a grouped bar chart that summarizes sales by month. The company's goal is to complete at least 10,000 transactions per month.

Total Sales Sum per Month



Total Sales Sum per Month



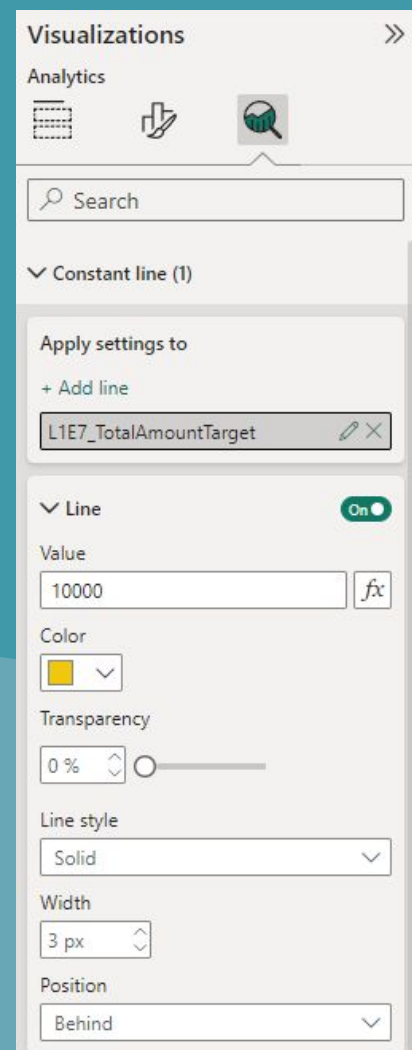
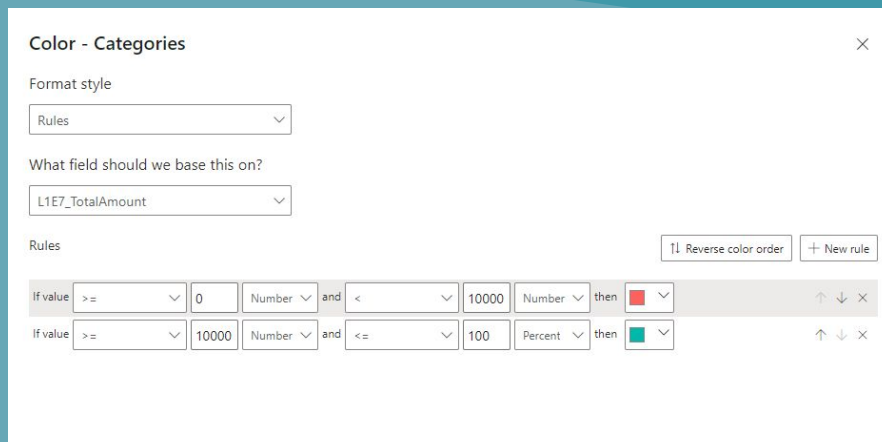
The trend is not very good because one month with the achieved goal is followed by months where the company does not reach it, it seems that last year the situation was more stable. It is necessary to study this indicator in more depth in order to plan measures to increase the indicators to achieve stable goal fulfillment.

To create these visuals, I added measures that count the sum of amount with declined = 0 to only count sales that were not rejected, because we only care about sales for which the company has received money.

```
1 L1E7_TotalAmount = CALCULATE(  
2     SUM(transactions[amount]),  
3     transactions[declined] = 0  
4 )
```

For this visual, I set up a goal via 'Constant Line' in the 'Analytics' block. →

I highlighted the columns of months that did not reach the target using the customize column color by rules feature. I got the impression that this functionality of Power BI is not perfect, for example, it does not work if you set the Max value in the final rule:



LEVEL 1 EXERCISE 8

In this exercise, we want to dig deeper into the transactions performed by each user and present the information in a clear and understandable form. In the table, present the following information:

- first and last name of users (a new column must be created to combine this information);
- age of users;
- average transactions in EUR;
- average transaction value in USD (conversion: 1 EUR equals 1.08 USD);
- necessary changes must be made to identify users who had an average of 300 or more EUR and 320 or more USD in transactions.

Average Ticket Sum per Client

Client Name	Age	Euros	Dollars
Keane McKinney	31	308,12	332,77
Dawn Murray	35	306,94	331,50
Lucas Dawson	41	304,43	328,78
Sasha Emerson	43	301,25	325,35
Sheila Dickerson	26	298,62	322,51
Olga Case	33	298,08	321,92
Gisela Johnston	31	295,55	319,19
Lynn Riddle	40	293,63	317,12
Theodore Barry	41	293,53	317,01
Lane Paul	41	292,86	316,28
Zoe Morrow	37	292,76	316,18
Yvonne Hatfield	43	292,06	315,42
Allen Calhoun	37	286,60	309,53
Acton Gallegos	35	283,15	305,80
Rhea Harvey	27	282,64	305,25
Neil Powers	44	281,90	304,45

Average Ticket Sum per Client

Client Name	Age	Euros	Dollars
Zahir Swanson	33	172,74	186,56
Blaze Duke	32	399,62	431,59
Cassandra Ferguson	32	274,28	296,22
Francis Bryant	32		
Garrett Mcconnell	32		
Linda Gaines	32	399,33	431,28
Linus Willis	32	270,78	292,44
Melissa Cameron	32	359,14	387,87
Philip Carey	32	234,52	253,28
Wyatt Morris	32	358,36	387,03
Burke Graham	31	414,58	447,74
Camilla Roach	31	445,48	481,12
Gisela Johnston	31	295,55	319,19
Keane McKinney	31	308,12	332,77
Keiko Guerra	31	234,49	253,24

Average Ticket Sum per Client

Client Name	Age	Euros	Dollars
Abra Doyle	38		
Alika Valdez	34		
Allegra Stanton	34		
Allistair Holmes	34		
Aquila Haley	28		
Aquila Strickland	42		
Aretha Chang	26		
Astra Alexander	41		
Barrett Andrews	29		
Benedict Wheeler	25		
Bruce Gill	34		
Chase Yang	25		
Ciaran Harrison	26		
Daquan Kirk	30		
Deacon Sharpe	45		
Diana Williamson	33		

With a superficial analysis, we can say that customers aged 40–42 and 30–32 buy more, here a detailed analysis is needed to verify this hypothesis (screenshot in the middle). We see the achievement and non-achievement of the goal for the same check in euros and dollars due to the disproportionate conversion of goals, it is necessary to clarify this point for an error (screenshot in the left). I left users with empty data in the table because it is important to see customers who do not make purchases at all (screenshot in the right).

To create this visual, I added several calculated columns to the users table. For the full name of the customers, I used the CONCATENATE function to glue the first and last names together. Age is obtained using DATADIFF, which returns the number of years between the birthday date and today's date. Average ticket sum is calculated with a filter by user. And the ticket sum in dollars takes this value and multiplies it by the exchange rate.

```
1 1L8_full_name = CONCATENATE(users[name], " " & users[surname])
```

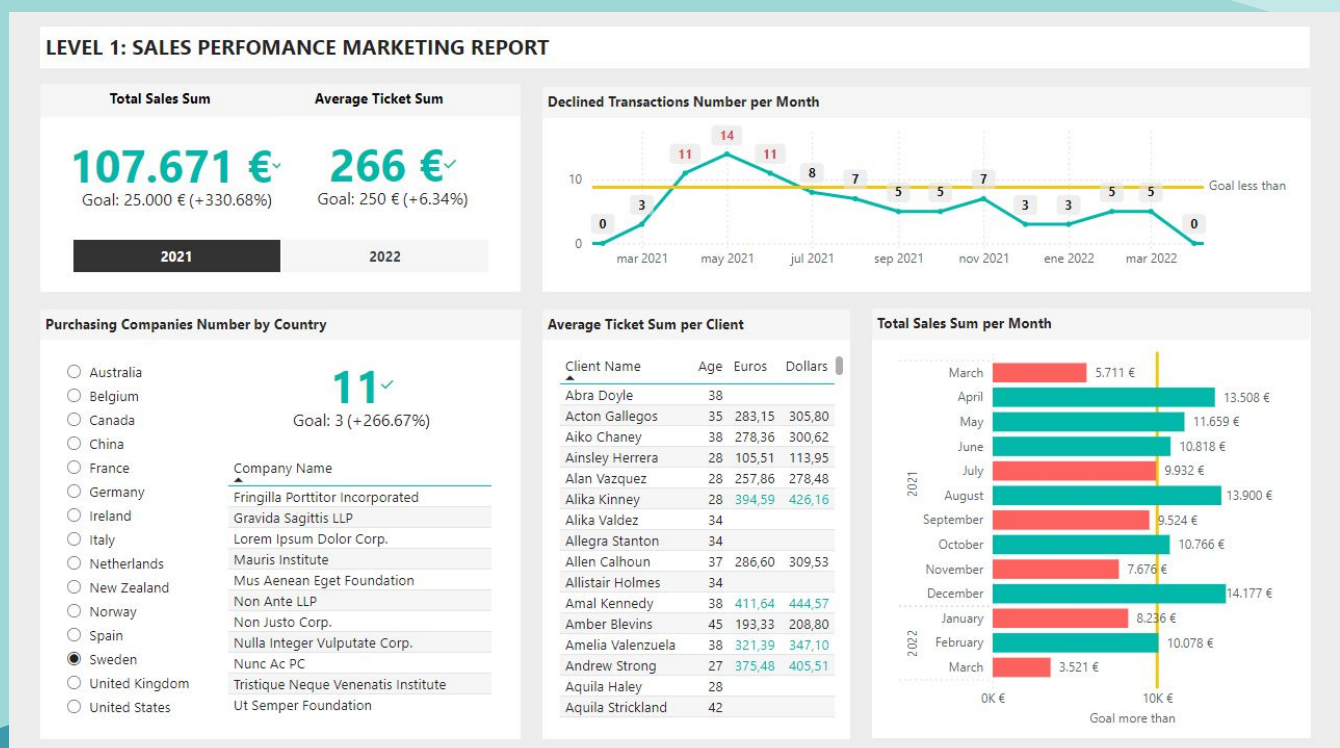
```
1 1L8_Age = DATEDIFF(users[birth_date].[Date], TODAY(), YEAR)
```

```
1 1L8_AvgTransactionPerUser =  
2   CALCULATE(  
3     AVERAGE(transactions[amount]),  
4     FILTER(transactions, transactions[user_id] = users[id])  
5   )
```

```
1 1L8_Dollars = users[1L8_AvgTransactionPerUser] * 1.08
```

LEVEL 1 EXERCISE 9

Write a short paragraph (up to 50 words) explaining the meaning of the numbers presented in Power BI visualizations. You can interpret the data as a whole or focus on a specific country. Accompany your interpretations with screenshots of the visualizations you will analyze.

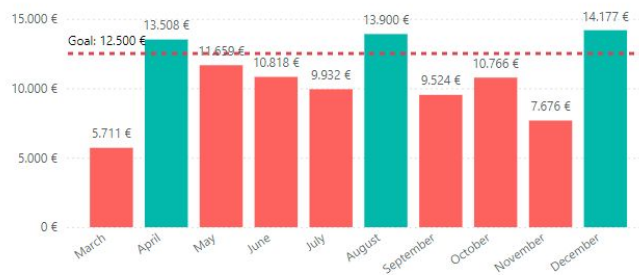


The annual plan for total sales and average check amounts was successfully fulfilled in 2021, and is on track for 2022. The company has secured a successful trend in the number of rejected transactions. We have successfully expanded our geography and, in general, fulfilled the plan for the number of companies in each country.

The second level and the second page of the dashboard. There are also no common dependencies here, and the two blocks work separately.

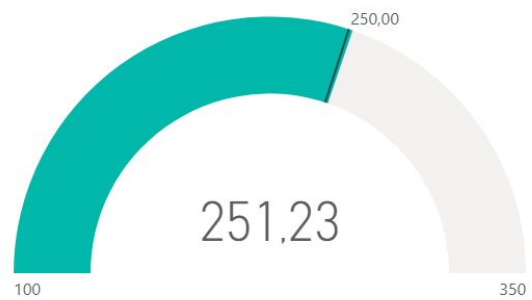
LEVEL 2: SALES METRICS ANALYSIS REPORT

Total Sales Sum in 2021 per Month



Month	Total Amount 2021	Achiving
December	14.177 €	achieved
August	13.900 €	achieved
April	13.508 €	achieved
May	11.659 €	not achieved
June	10.818 €	not achieved
October	10.766 €	not achieved
July	9.932 €	not achieved
September	9.524 €	not achieved
November	7.676 €	not achieved
March	5.711 €	not achieved

Average Ticket Sum by Germany



2021		2022
Company Name	Sales Number	Email
Nunc Interdum Incorporated	86	non@outlook.com
Ac Fermentum Incorporated	2	donec.porttitor.tellus@yahoo.net
Ac Industries	2	ipsum@yahoo.com
Aliquam PC	2	scelerisque.mollis@icloud.org
Auctor Mauris Corp.	2	eget.metus@protonmail.ca
Convallis In Incorporated	2	mauris.ut@aol.co.uk
Augue Foundation	1	mauris@yahoo.com

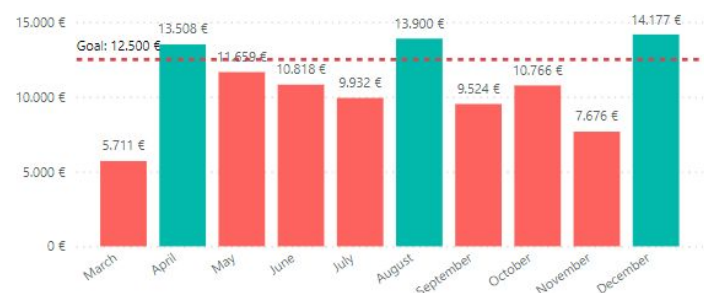
LEVEL 2 EXERCISE 1

From a marketing perspective, they need to study the monthly trend of transactions made in 2021, in particular, they want to know the change in transactions depending on the month. Remember to visualize the business goal of achieving a transaction amount of at least €12,500 per month. In this exercise, you will need to be able to identify the months in which the set goal was not achieved. If necessary, two views can be made.

Given that the company met the monthly transaction target of 12,500 euros only 30% of the time, I recommend investigating the factors that contributed to the high-performing months. For example, last month's earnings were nearly double those of the previous month, indicating potential underlying drivers. Identifying these events could help replicate successful strategies in lower-performing months and improve overall consistency in meeting monthly targets.

To create the graph, I added a measure as in the previous tasks with SUM and declined = 0, only specifically for 2021. I set up a goal via 'Constant Line' in the 'Analytics' block and highlighted the columns using the customize column color by rules feature.

Total Sales Sum in 2021 per Month



Month	Total Amount 2021	Achieving
December	14.177 €	achieved
August	13.900 €	achieved
April	13.508 €	achieved
May	11.659 €	not achieved
June	10.818 €	not achieved
October	10.766 €	not achieved
July	9.932 €	not achieved
September	9.524 €	not achieved
November	7.676 €	not achieved
March	5.711 €	not achieved

```

1 L2E1_TotalAmount2021 = CALCULATE(
2     SUM(transactions[amount]),
3     YEAR(transactions[timestamp]) = 2021,
4     transactions[declined] = 0
5 )

```

In this block I added a second visual for table lovers. For the table I added a measure of goal achievement by practicing with the IF operator in DAX:

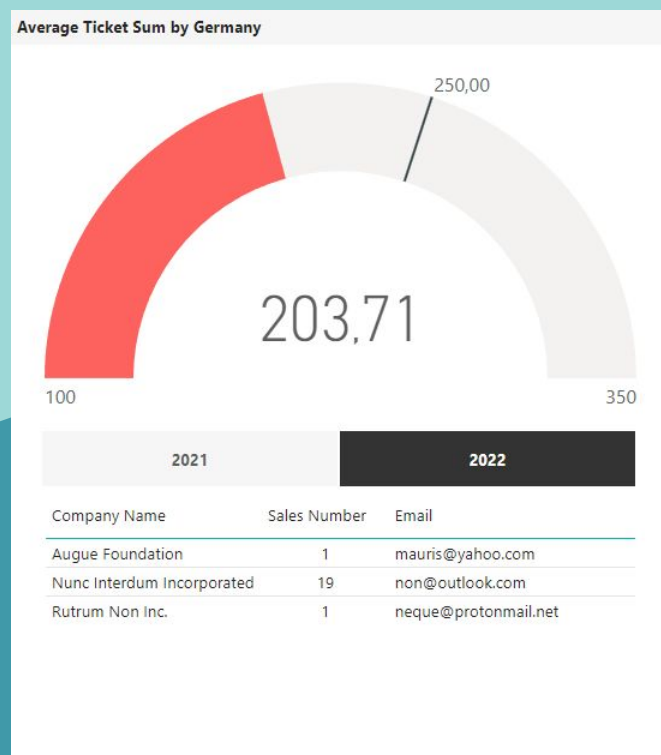
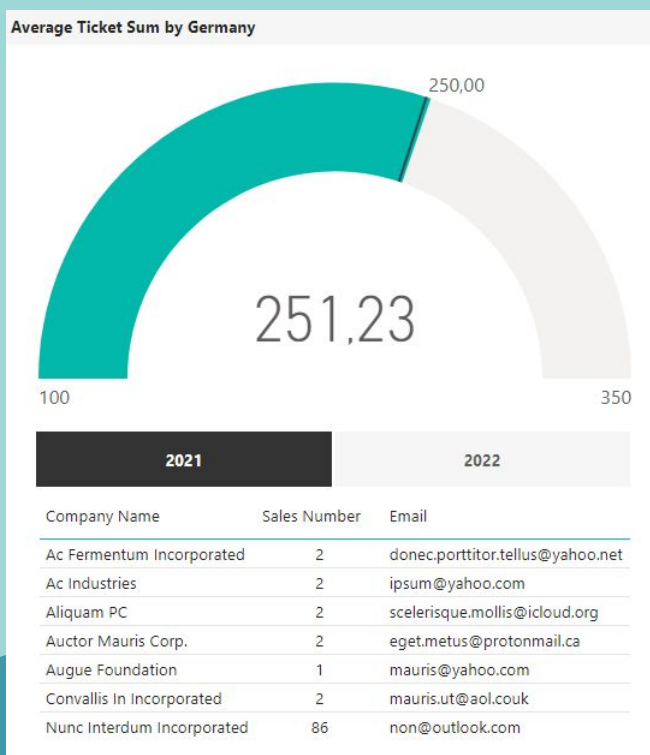
```

1 L2E1_Achieving = IF(
2     [L2E1_TotalAmount2021] >= [L2E1_TotalAmount2021Target],
3     "achieved",
4     IF(
5         [L2E1_TotalAmount2021] <> 0,
6         "not achieved"
7     )
8 )

```

LEVEL 2 EXERCISE 2

In your work, you want to gain a deeper understanding of transactions conducted in Germany. Therefore, you are asked to develop DAX metrics to create visualizations that reflect average sales in Germany. Keep in mind that the company is aiming to achieve a figure of 250 euros per year. Set the display so that the minimum value is 100 and the maximum is 350, thus providing a more effective presentation of the information.



The company fulfilled the plan in 2021, and has a tendency to exceed the plan in 2022. I have added to the visual a list of companies in Germany that contributed to sales in each year, with the number of transactions and email for contact. I recommend conducting a study of the sales of the company Nunk, as their number of transactions is fantastic compared to the rest.

For this visual I needed one measure with a condition companies[country] = "Germany", and the goal is configured in the Gauge visual itself.

```
1 L2E2_AvgAmountGermany = CALCULATE(  
2     AVERAGE(transactions[amount]),  
3     companies[country] = "Germany",  
4     transactions[declined] = 0,  
5     YEAR(transactions[timestamp]) = SELECTEDVALUE(transactions[timestamp].[Year])  
6 )
```

LEVEL 2 EXERCISE 3

Write a short paragraph, maximum 25 words, indicating in which month you did not achieve the suggested goal of Exercise 1.

The company failed to achieve its goal in 7 months out of 10. March was the most disastrous month – we did not even achieve half of the goal.

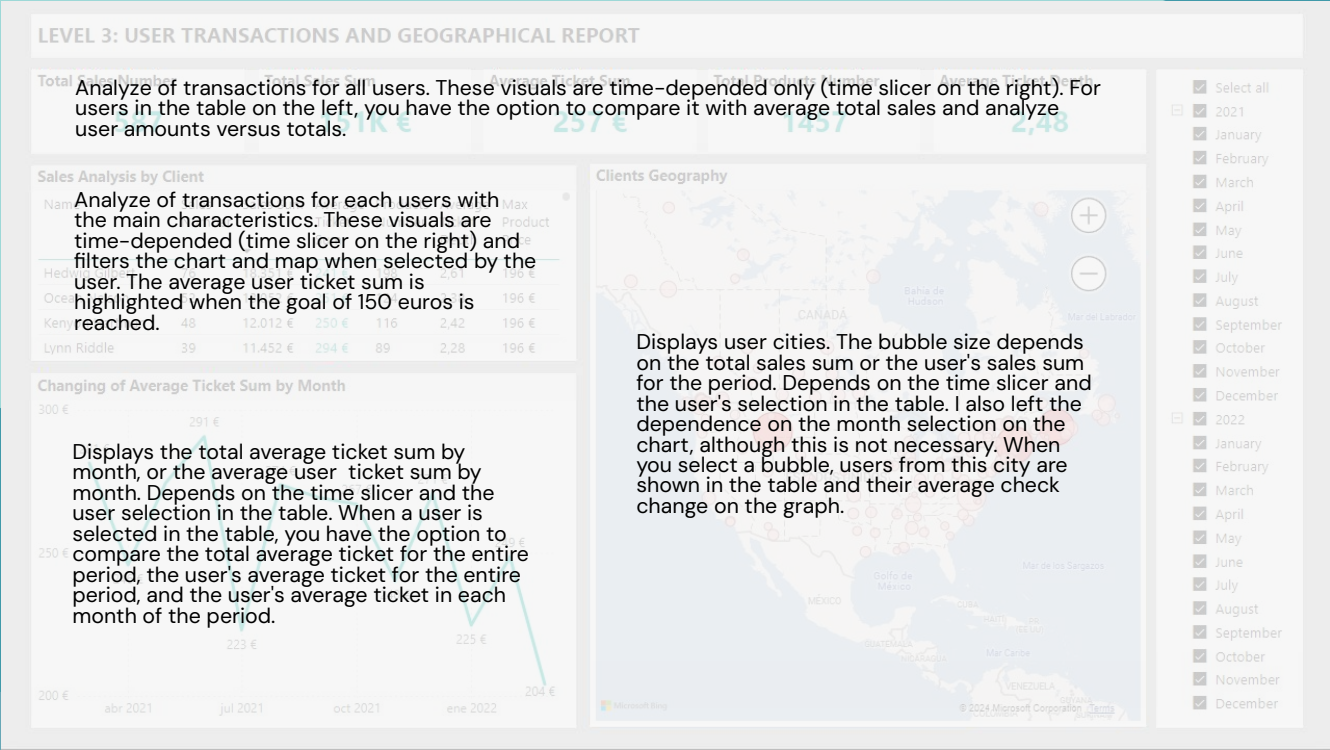
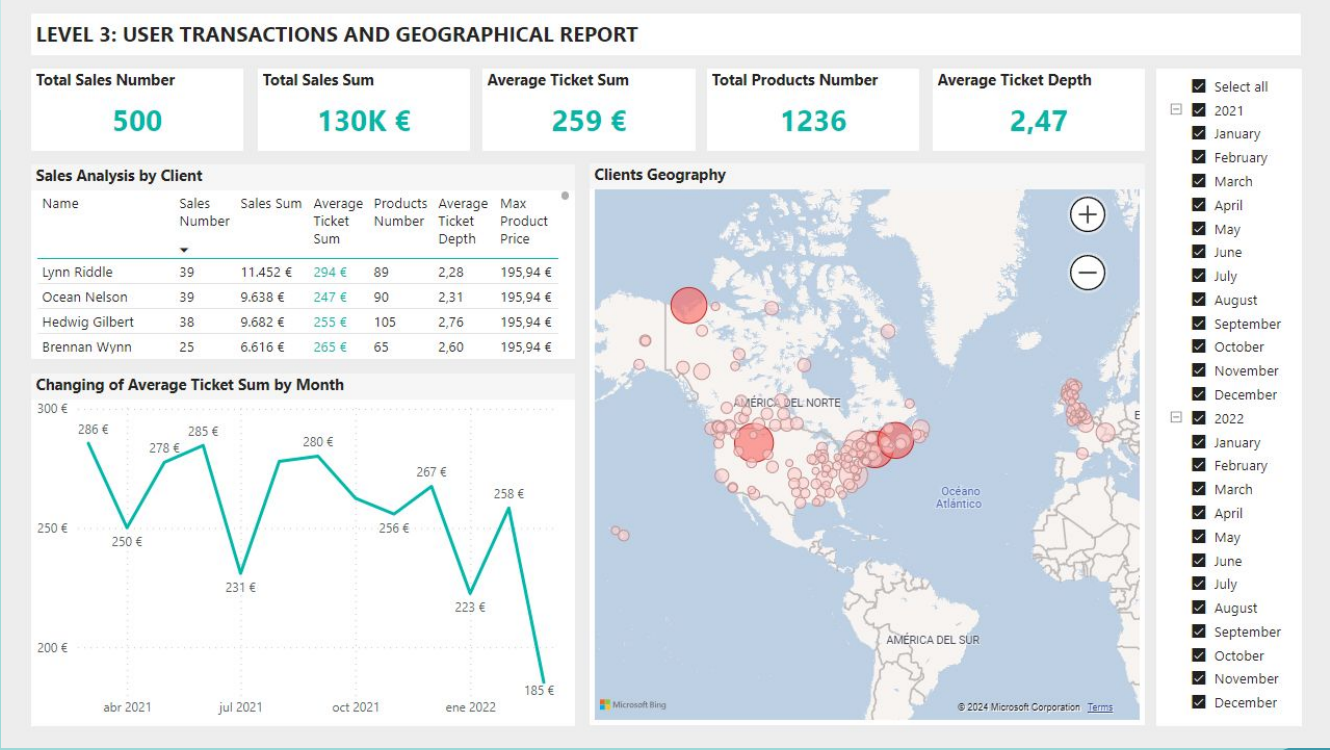
LEVEL 3 EXERCISE 1

The Marketing department wants to take a deeper look at the transactions made by users. Therefore, you are asked to prepare several visualizations that include:

- 1) key statistics of the variables that you consider important to understand the transactions made by users;
- 2) number of products purchased by each user;
- 3) average number of purchases made per user. See which users have an average number of purchases greater than 150 and which do not;
- 4) shows the price of the most expensive product purchased by each user;
- 5) visualize the geographical distribution of users.

In this exercise, you will need to make the necessary changes to each graph to improve its readability and comprehension. In this task, you are expected to carefully evaluate which variables are relevant to effectively convey the required information.

The approach I took to accomplish this task was to create a single dashboard with adding some views, dependencies, and slices, where you can also see the requested tasks. First, I'll show you a screenshot of the finished dashboard. Next, I will show a screenshot with a description of the graphs and dependencies, and after a screenshot with comments on which graph which requested task is performed. After that, I will provide the measures and the description of the DAX code that were created for this dashboard. I will finish the report with an analysis of two slices and conclusions for the company.

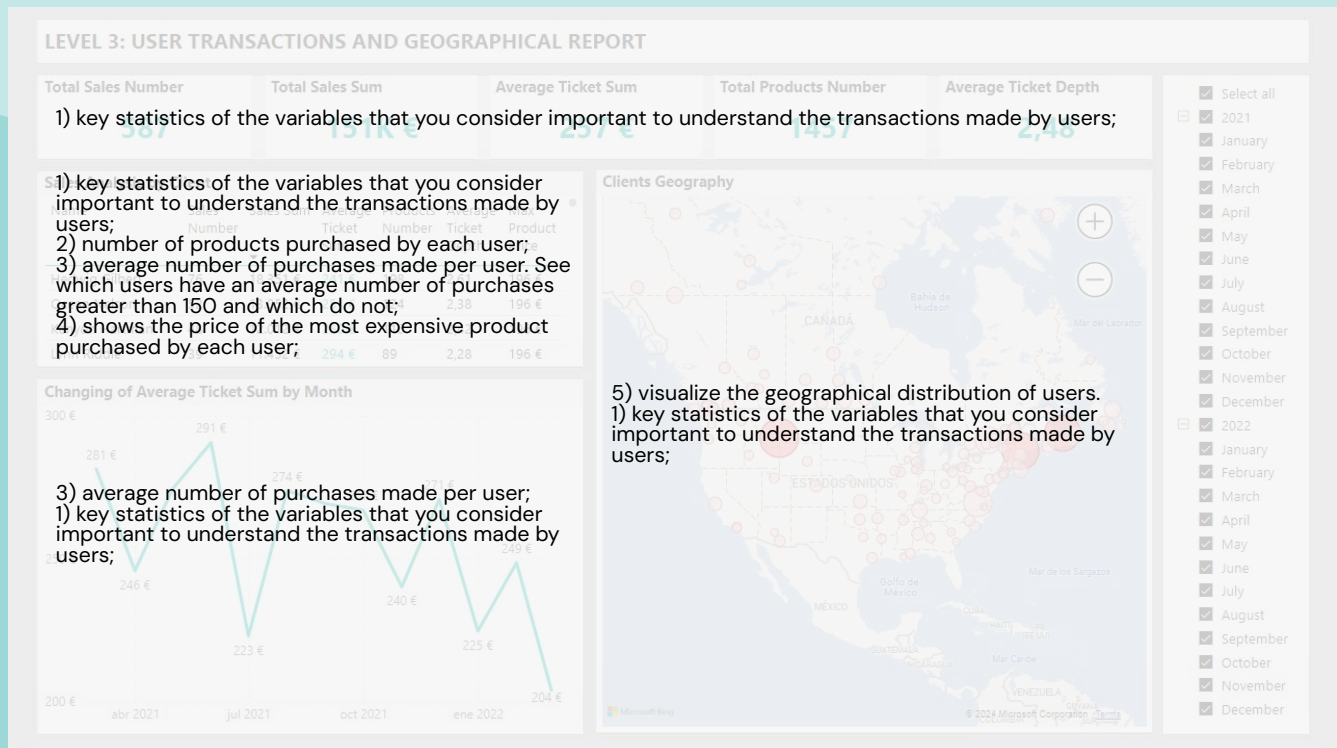


Analyze of transactions for all users. These visuals are time-depended only (time slicer on the right). For users in the table on the left, you have the option to compare it with average total sales and analyze user amounts versus totals.

Analyze of transactions for each users with the main characteristics. These visuals are time-depended (time slicer on the right) and filters the chart and map when selected by the user. The average user ticket sum is highlighted when the goal of 150 euros is reached.

Displays the total average ticket sum by month, or the average user ticket sum by month. Depends on the time slicer and the user selection in the table. When a user is selected in the table, you have the option to compare the total average ticket for the entire period, the user's average ticket for the entire period, and the user's average ticket in each month of the period.

Displays user cities. The bubble size depends on the total sales sum or the user's sales sum for the period. Depends on the time slicer and the user's selection in the table. I also left the dependence on the month selection on the chart, although this is not necessary. When you select a bubble, users from this city are shown in the table and their average check change on the graph.



For this dashboard, I chose the following key user-analyzed characteristics: number of transactions, total amount of all transactions, average transactions amount, number of purchased products, average transactions depth, and I also added the most important slice – their change over time. I also added cards for total sales for all these characteristics, so that they could be analyzed in comparison from any slice.

Now I will demonstrate what measures were created for this dashboard. There is a measure that duplicates the measure from the previous tasks, and theoretically it is possible to optimize the measures for all exercises and reduce their number, but since there is not much data and calculations, the dashboard is not loaded, the exercises are not related to each other, I chose as a higher priority in this training example for greater demonstrativeness to organize a full set of measures for each page and organize them by folders by pages. The measures created for this dashboard in the order of their arrangement in the cards and table:


```
1 L3E1_CountAmount = CALCULATE(  
2 |     COUNT(transactions[amount]),  
3 |     transactions[declined] = 0  
4 )
```

```
1 L3E1_SumAmount = CALCULATE(  
2 |     SUM(transactions[amount]),  
3 |     transactions[declined] = 0  
4 )
```

```
1 L3E1_AvgAmount = CALCULATE(  
2 |     AVERAGE(transactions[amount]),  
3 |     transactions[declined] = 0  
4 )
```

```
1 L3E1_CountProducts = CALCULATE(  
2 |     COUNT(transaction_items[product_id]),  
3 |     transactions[declined] = 0  
4 )
```

These are simple measures that are considered to be relevant functions and take into account that transactions must be accepted.

```
1 L3E1_AvgCountProducts = AVERAGEX(  
2 |     VALUES(transaction_items[transaction_id]),  
3 |     CALCULATE(  
4 |         COUNT(transaction_items[product_id]),  
5 |         transactions[declined] = 0  
6 |     )  
7 )
```

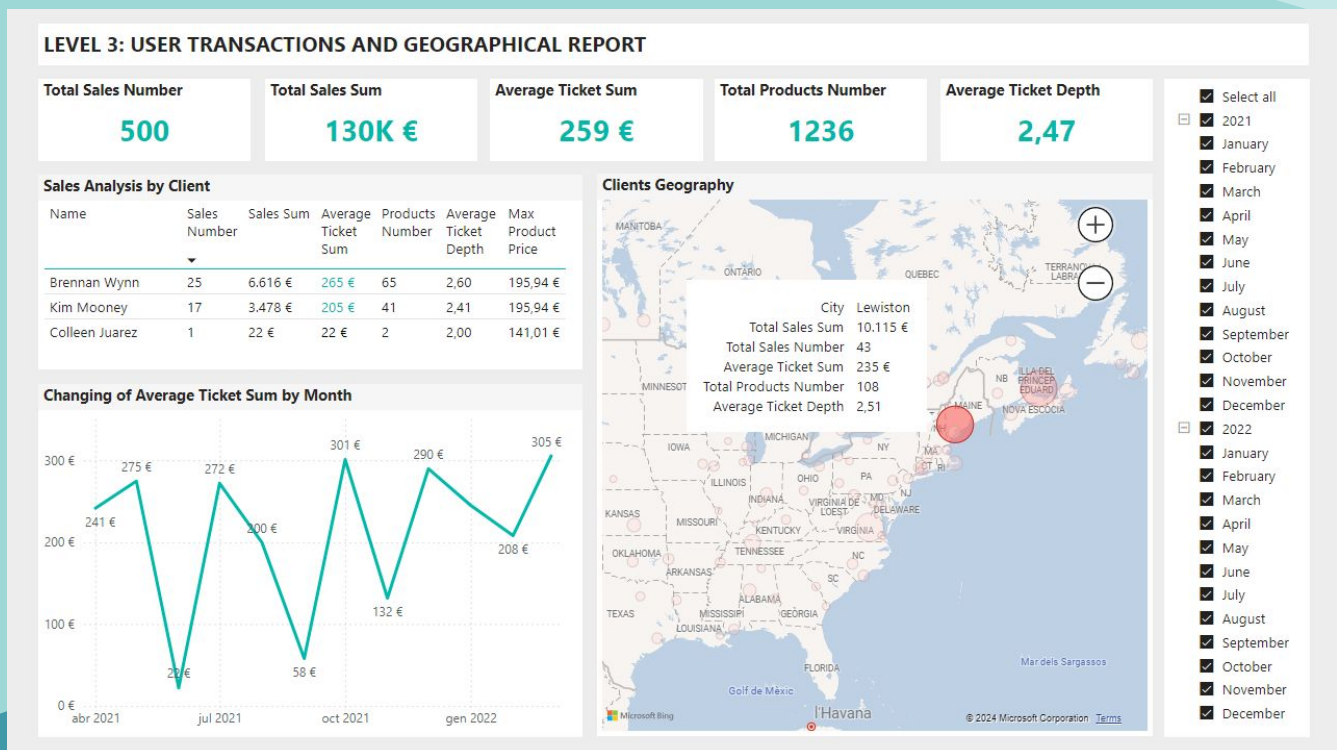
This measure calculates the average number of products per transaction where the transaction was not declined. `VALUES(transaction_items[transaction_id])` creates a set of unique transactions, and `AVERAGEX` iterates over each of them, counting the product counts `COUNT(transaction_items[product_id])` for transactions where `transactions[declined] = 0` (not declined). `AVERAGEX` then takes the average of all calculated quantities to get the final average product count for completed transactions.

```
1 L3E1_MaxProductPrice = MAXX(  
2   SUMMARIZE(  
3     transaction_items,  
4     transactions[user_id],  
5     products[id],  
6     "MaxPrice", MAX(products[price])  
7   ),  
8   [MaxPrice]  
9 )
```

This measure calculates the highest price of a product purchased by each user. SUMMARIZE creates a temporary table that groups rows by each user (transactions[user_id]) and product (products[id]), and for each combination, stores the maximum price of that product in the 'MaxPrice' column in the temporary table. MAXX then takes the maximum value from 'MaxPrice' across all rows in the temporary table, returning the highest price for the product for the user.

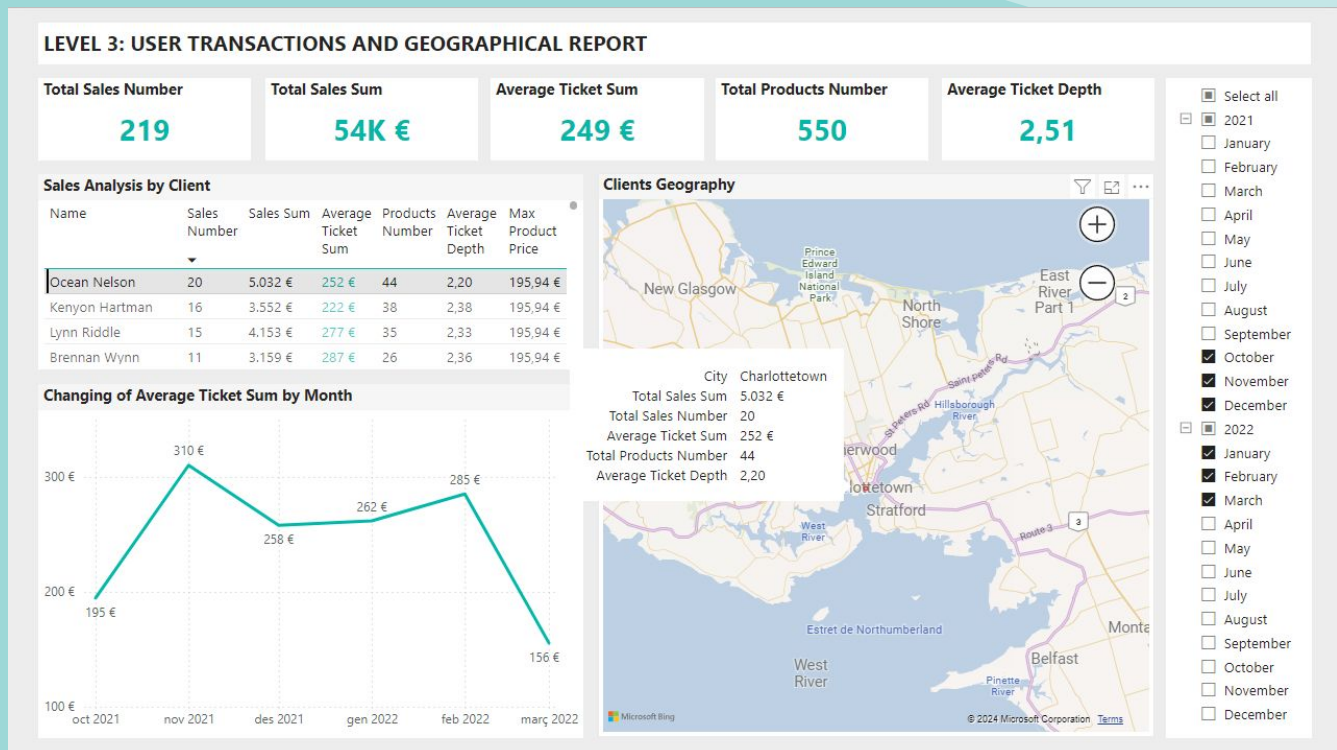
In addition to these measures, I also used the calculated column fullname from the users table for the table visual and city from the users table for the map.

Now I will analyze the results and draw conclusions on the visuals for the company. For this dashboard, I decided to work with the conclusions by selecting user scenarios and analyzing specific slices. The first slice is for the scenario when the stakeholder clicks on the big bubble on the map or purposefully clicks on a specific city to analyze sales in this city and does not interact with the time slicer, thus selecting the entire available period:



The city of Lewiston brought the company almost 8% of the money, while the average ticket sum has potential for growth compared to the general. Here we have three clients, two of which are especially large: Brennan Wynn made 5% of purchases and its average amount and depth are above average, we could work with this client to strengthen cooperation; Kim Mooney made more than 3% of purchases, but the average amount and depth, on the contrary, are significantly less than average, we should work with a valuable client to learn his needs and not lose him. The general trend of average sales growth in the city in recent months has been more stable.

And the second slice is for a scenario where the stakeholder is interested in the period from the last quarter of 2021 to the end of the first quarter of 2022 and he selects the client with the maximum number of transactions for analysis:



Ocean Nelson, the only client in his city Charlottetown, Canada, brought the company about 10% of the number of transactions and money. The client buys consistently every month and chooses premium products. In the last month, the client's ticket has fallen almost twice, we should process the reasons before it becomes a cause for concern.