# BI Analyst Capstone Project

Peter Löffelmann - March 2025

# Table of Contents

# 1 Introduction

**Name**: Peter Löffelmann

**Desired certificate**: IBM Business Intelligence (BI) Analyst Professional Certificate

**Time frame working on the certificate**: February – March 2025

**Target field of activity in the job**: Data-based optimization of production processes

**Note**: Some content (e.g. screenshots) may be on German since most of the Software was used in German language settings

Feedback is welcome! Thank you very much :)

# 2 Methodology

Modul 1: Data Cleaning and Analysis:

- Cleaned and analyzed sales data using Excel.
- Conducted Sales Data Analysis, Sales by City Analysis, and Product Performance Analysis.

Modul 2: Data Querying and Analysis:

- Used PostgreSQL to create data cubes with ROLLUP.
- Summarized data along hierarchies.
- Combined time and geographical data to identify trends.

Modul 3: Chart Creation and Regression Analysis:

- Created charts using Excel.
- Conducted regression analysis to find correlations between sales figures and dates.

Modul 4: Data Visualization:

- Used Tableau to create dynamic dashboards and visualizations.
- Presented key findings on sales performance, regional sales analysis, product analysis, and store performance.

# 3 Results – Module 1.1 Data Cleaning



**②**

Cleaned the data set.

- e.g. deleting missing values
- As seen on picture

**①** Uploaded every data set.

# 3 Results – Module 1.2 Pivot Tables

## ① Sales by store (pivot table)

| Zeilenbeschriftungen | Summe von sales | Summe von stock | Summe von revenue | Mittelwert von sales2 |
|---|---|---|---|---|
| ⊟ S0001 | | | | |
| Electro World (National Chain) | 11828 | 19955 | 78721 | 14,3369697 |
| ⊟ S0002 | | | | |
| DIGI | 16708 | 32237 | 84346 | 23,9713056 |
| ⊟ S0003 | | | | |
| Darty | 2503 | 20377 | 28631 | 9,932539683 |
| ⊟ S0004 | | | | |
| Elettrodomestici Rossi | 111 | 5831 | 34736 | 0,269417476 |
| ⊟ S0006 | | | | |
| Euronics Lisboa (National Chain) | 43 | 2599 | 5658 | 0,323308271 |
| ⊟ S0008 | | | | |
| Currys (National Chain) | 83 | 7415 | 18774 | 0,144347826 |
| ⊟ S0009 | | | | |
| Appliance Centre | 64 | 3400 | 11841 | 0,297674419 |
| ⊟ S0010 | | | | |

## ② Sales by cityid (pivot table)

| Zeilenbeschriftungen | Summe von sales | Summe von revenue | Mittelwert von sales |
|---|---|---|---|
| ⊟ S0001 | | | |
| C031 | 11828 | 78721 | 14,3369697 |
| ⊟ S0002 | | | |
| C007 | 16708 | 84346 | 23,9713056 |
| ⊟ S0003 | | | |
| C014 | 2503 | 28631 | 9,932539683 |
| ⊟ S0004 | | | |
| C022 | 111 | 34736 | 0,269417476 |
| ⊟ S0006 | | | |
| C024 | 43 | 5658 | 0,323308271 |
| ⊟ S0008 | | | |
| C024 | 83 | 18774 | 0,144347826 |
| ⊟ S0009 | | | |
| C020 | 64 | 11841 | 0,297674419 |
| ⊟ S0010 | | | |
| C014 | 28673 | 114124 | 64,14541387 |
| ⊟ S0011 | | | |
| C031 | 104 | 15695 | 0,495238095 |
| ⊟ S0012 | | | |
| C005 | 165 | 39516 | 0,294117647 |
| ⊟ S0013 | | | |

## ③ Sales by product (pivot table)

| Zeilenbeschriftungen | Summe von sales | Summe von revenue | Mittelwert von sales2 |
|---|---|---|---|
| ⊟ P0001 | | | |
| Side-by-Side Refrigerator | 10 | 1483 | 0,07518797 |
| ⊟ P0004 | | | |
| Beverage Center | 3 | 1143 | 0,15 |
| ⊟ P0005 | | | |
| Wine Cooler | 1 | 2873 | 0,00862069 |
| ⊟ P0011 | | | |
| Induction Cooktop | 3 | 12687 | 0,053571429 |
| ⊟ P0014 | | | |
| Microwave Oven | 0 | 0 | 0 |
| ⊟ P0015 | | | |
| Convection Oven | 62 | 13041 | 0,196202532 |

# 3 Results – Module 2.1 Data Querying

**1** BICapstone database tables and data from the BI-dump.sql file was restored using the command

**2** Data population of six tables was checked! ✔
Example table (city_names):

```
CREATE DATABASE
postgres=# \connect bicapstone
psql (14.17 (Ubuntu 14.17-0ubuntu0.22.04.1) server 13.2)
You are now connected to database "bicapstone" as user "postgres".
bicapstone=# \include BI-dump.sql
SET
SET
SET
SET
SET
 set_config
------------

(1 row)

SET
SET
SET
SET
SET
SET
CREATE TABLE
ALTER TABLE
CREATE TABLE
ALTER TABLE
CREATE TABLE
ALTER TABLE
CREATE TABLE
ALTER TABLE
CREATE TABLE
ALTER TABLE
CREATE TABLE
ALTER TABLE
CREATE TABLE
ALTER TABLE
COPY 37
COPY 699
COPY 699
COPY 19999
```
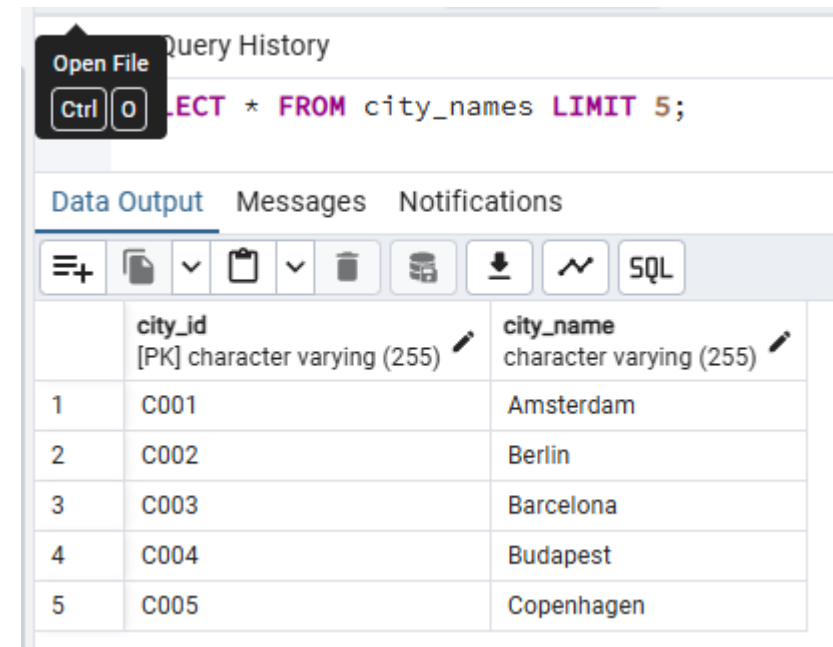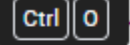
∨ ⊞ Tables (6)
  > ⊞ city_names
  > ⊞ product_hierarchy
  > ⊞ product_names
  > ⊞ sales
  > ⊞ store_cities
  > ⊞ store_names

Open File [Ctrl] [O] — Query History

`LECT * FROM city_names LIMIT 5;`

Data Output    Messages    Notifications

| | city_id<br>[PK] character varying (255) | city_name<br>character varying (255) |
|---|---|---|
| 1 | C001 | Amsterdam |
| 2 | C002 | Berlin |
| 3 | C003 | Barcelona |
| 4 | C004 | Budapest |
| 5 | C005 | Copenhagen |

# 3 Results – Module 2.1 Data Querying

**3** Sales performance was checked by joining tables and aggregating the data

```
1 ∨  SELECT
2       p.product_id,
3       c.store_id,
4       c.city_id,
5       SUM (sales) AS total_sales,
6       SUM (revenue) AS total_revenue,
7       AVG (price) AS average_price
8    FROM
9       sales sl
10   INNER JOIN product_hierarchy p ON sl.product_id = p.product_id
11   INNER JOIN store_cities c ON sl.store_id = c.store_id
12   GROUP BY p.product_id, c.store_id, c.city_id;
```

Data Output   Messages   Notifications

| | product_id character varying | store_id character varying | city_id character varying | total_sales double precision | total_revenue double precision | average_price double precision |
|---|---|---|---|---|---|---|
| 1 | P0001 | S0002 | C007 | 0 | 0 | 6.25 |
| 2 | P0001 | S0012 | C005 | 1 | 5.3 | 6.25 |
| 3 | P0001 | S0013 | C026 | 2 | 10.59 | 6.25 |
| 4 | P0001 | S0023 | C008 | 0 | 0 | 6.25 |
| 5 | P0001 | S0025 | C024 | 0 | 0 | 6.25 |
| 6 | P0001 | S0027 | C022 | 0 | 0 | 6.25 |
| 7 | P0001 | S0040 | C017 | 0 | 0 | 6.25 |
| 8 | P0001 | S0049 | C031 | 0 | 0 | 6.25 |
| 9 | P0001 | S0050 | C014 | 0 | 0 | 6.25 |
| 10 | P0001 | S0051 | C027 | 0 | 0 | 6.25 |
| 11 | P0001 | S0055 | C014 | 0 | 0 | 6.25 |
| 12 | P0001 | S0056 | C015 | 1 | 5.3 | 6.25 |
| 13 | P0001 | S0062 | C014 | 0 | 0 | 6.25 |

# 3 Results – Module 2.2 Data Analysis

**①** Sales trends over time were analyzed with ROLLUP

```
Query    Query History

1 ∨  SELECT
2        store_id,
3        TO_CHAR(date, 'YYYY-MM') AS date_formatted,
4        SUM(sales) AS total_sales
5    FROM
6        sales
7    GROUP BY
8        ROLLUP (store_id, TO_CHAR(date, 'YYYY-MM'));
```

Data Output    Messages    Notifications

| | store_id<br>character varying | date_formatted<br>text | total_sales<br>double precision |
|---|---|---|---|
| 1 | [null] | [null] | 9856.880999999998 |
| 2 | S0013 | 2017-03 | 27 |
| 3 | S0052 | 2017-02 | 23.675 |
| 4 | S0120 | 2017-02 | 19 |
| 5 | S0096 | 2017-03 | 36 |
| 6 | S0135 | 2017-02 | 32.305 |
| 7 | S0018 | 2017-02 | 38 |
| 8 | S0024 | 2017-02 | 85 |
| 9 | S0044 | 2017-02 | 22 |
| 10 | S0127 | 2017-03 | 9 |
| 11 | S0107 | 2017-03 | 27 |
| 12 | S0072 | 2017-02 | 80 |
| 13 | S0102 | 2017-03 | 23 |

# 3 Results – Module 2.2 Data Analysis

**2** Rolled-up summary of sales data across product hierarchy levels was performed

```
1  ∨  SELECT
2         hierarchy1_id, hierarchy2_id,
3         SUM(sales) AS total_sales
4     FROM
5       sales
6     JOIN
7       product_hierarchy p
8     ON
9         sales.product_id = p.product_id
10    GROUP BY
11      ROLLUP (hierarchy1_id, hierarchy2_id);
```

Data Output   Messages   Notifications

| | hierarchy1_id character varying | hierarchy2_id character varying | total_sales double precision |
|---|---|---|---|
| 1 | [null] | [null] | 9856.880999999998 |
| 2 | H03 | H0317 | 2 |
| 3 | H00 | H0001 | 617 |
| 4 | H03 | H0316 | 0 |
| 5 | H03 | H0314 | 190 |
| 6 | H03 | H0315 | 5 |
| 7 | H01 | H0107 | 374 |
| 8 | H03 | H0313 | 581 |
| 9 | H00 | H0000 | 1380 |
| 10 | H03 | H0311 | 17 |
| 11 | H00 | H0004 | 613.366 |
| 12 | H01 | H0108 | 172 |
| 13 | H01 | H0105 | 175 |

# 3 Results – Module 2.2 Data Analysis

**③** Data cube was created (combining sales data with geographical information)
→ analyze sales trend over time / regions



Query | Query History

```sql
1  SELECT
2      c.city_id,
3      TO_CHAR(s.date, 'YYYY-MM') AS date_formatted,
4      SUM(s.sales) AS total_sales
5  FROM
6      sales s
7  JOIN
8      store_cities c
9  ON
10     s.store_id = c.store_id
11 GROUP BY
12     ROLLUP (c.city_id, TO_CHAR(s.date, 'YYYY-MM'));
```

Data Output   Messages   Notifications

| | city_id<br>character varying | date_formatted<br>text | total_sales<br>double precision |
|---|---|---|---|
| 1 | [null] | [null] | 9856.880999999998 |
| 2 | C012 | 2017-02 | 39 |
| 3 | C006 | 2017-02 | 124 |
| 4 | C029 | 2017-03 | 92 |
| 5 | C033 | 2017-03 | 33 |
| 6 | C017 | 2017-03 | 44 |
| 7 | C021 | 2017-03 | 22.555 |
| 8 | C032 | 2017-02 | 49 |
| 9 | C017 | 2017-02 | 83 |
| 10 | C004 | 2017-03 | 110.208 |
| 11 | C018 | 2017-03 | 23 |
| 12 | C007 | 2017-03 | 28 |
| 13 | C011 | 2017-02 | 54.97 |

# 3 Results – Module 3.1 Data Visualization



**1** Excels Power Query was used to combine all files into one table (based on related columns x_id)

**2** Bar chart to dispaly the total sales per store

| Summe von sales | Spaltenbeschriftungen | | | | | | |
|---|---|---|---|---|---|---|---|
| Zeilenbeschriftungen | Edinburgh | Helsinki | London | Saint Petersburg | Vienna | (Leer) | Gesamtergebnis |
| Currys (National Chain) | | | | | 389 | | 389 |
| Darty | | | 1900,615 | | | | 1900,615 |
| DIGI | 5537,989 | | | | | | 5537,989 |
| Electro World (National Chain) | | 3674,995 | | | | | 3674,995 |
| Elettrodomestici Rossi | | | | 1247 | | | 1247 |
| Euronics Lisboa (National Chain) | | | | | 398 | | 398 |
| (Leer) | | | | | | | |
| Gesamtergebnis | 5537,989 | 3674,995 | 1900,615 | 1247 | 787 | | 13147,599 |



Store is outperforming!

Both stores in Vienna are underperforming!

# 3 Results – Module 3.1 Data Visualization

**(3)** Comparison of sales against the stock level
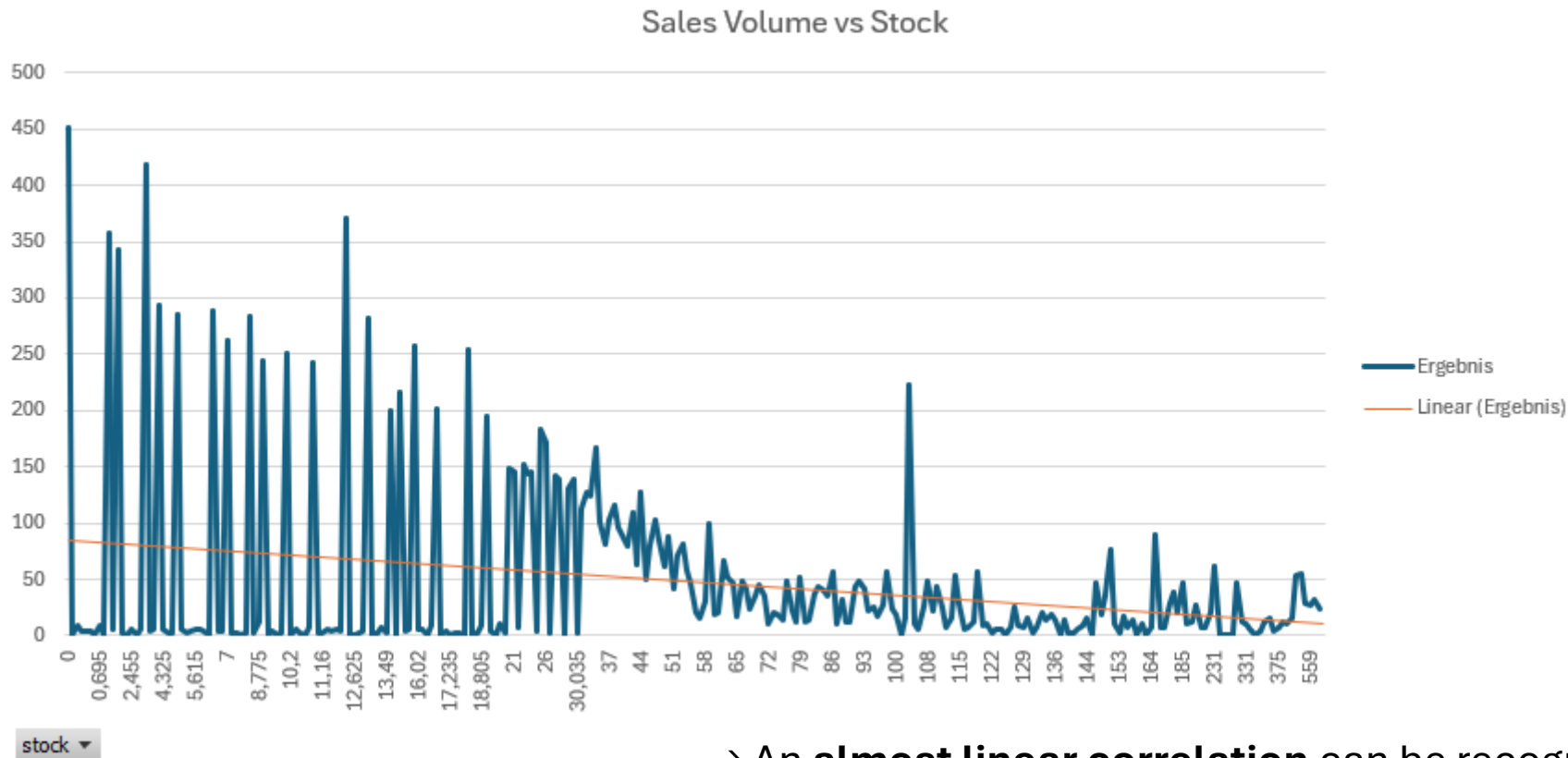


There is a simple correlation between the stock level and sales.

High sales lead to a reduction in the stock level. If the stock is low/empty, it is topped up.

**→ Accordingly, sales are high when the stock is low**

→ An **almost linear correlation** can be recognized by the orange trend line

**1** Raw results:

| Regressions-Statistik | |
|---|---|
| Multipler Korr | 0,00087942 |
| Bestimmthei | 7,7339E-07 |
| Adjustiertes E | -5,703E-05 |
| Standardfehle | 2,84048888 |
| Beobachtung | 17303 |

ANOVA

| | Freiheitsgrade | (dratsummen | Quadratsumr | Prüfgröße (F) | F krit |
|---|---|---|---|---|---|
| Regression | 1 | 0,10795808 | 0,10795808 | 0,0133804 | 0,90791253 |
| Residue | 17301 | 139590,992 | 8,06837708 | | |
| Gesamt | 17302 | 139591,1 | | | |

| | Koeffizienten | Standardfehle | t-Statistik | P-Wert | Untere 95% | Obere 95% | Untere 95,0% | Obere 95,0% |
|---|---|---|---|---|---|---|---|---|
| Schnittpunkt | 9,01293042 | 73,5838859 | 0,12248511 | 0,90251626 | -135,21893 | 153,244787 | -135,21893 | 153,244787 |
| X Variable 1 | -0,0001991 | 0,00172093 | -0,1156737 | 0,90791253 | -0,0035723 | 0,00317414 | -0,0035723 | 0,00317414 |

Date Coefficient p-value (0.907912531):

p-value for the date coefficient is much greater than 0.05
→ indicating that the date variable is not statistically significant
→ **this means changes in the date do not significantly affect sales**

R-squared (7.73388E-07):

The R-squared value is extremely low, almost zero
→ indicates that the date variable explains virtually none of the variance in sales
→ **model does not effectively predict sales based on the date**
(e.g. this could mean that it is not a business related to seasons)

Regression Equation

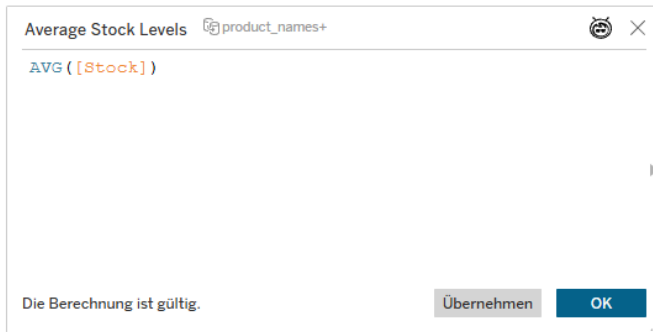**Equation: Sales = 9.012930425 + (-0.000199067 * Date)**

Given the non-significant coefficients and the low R-squared value (explanation above), this equation is not useful for predicting future sales based on the date.

# 3 Results – Module 4 Visualization

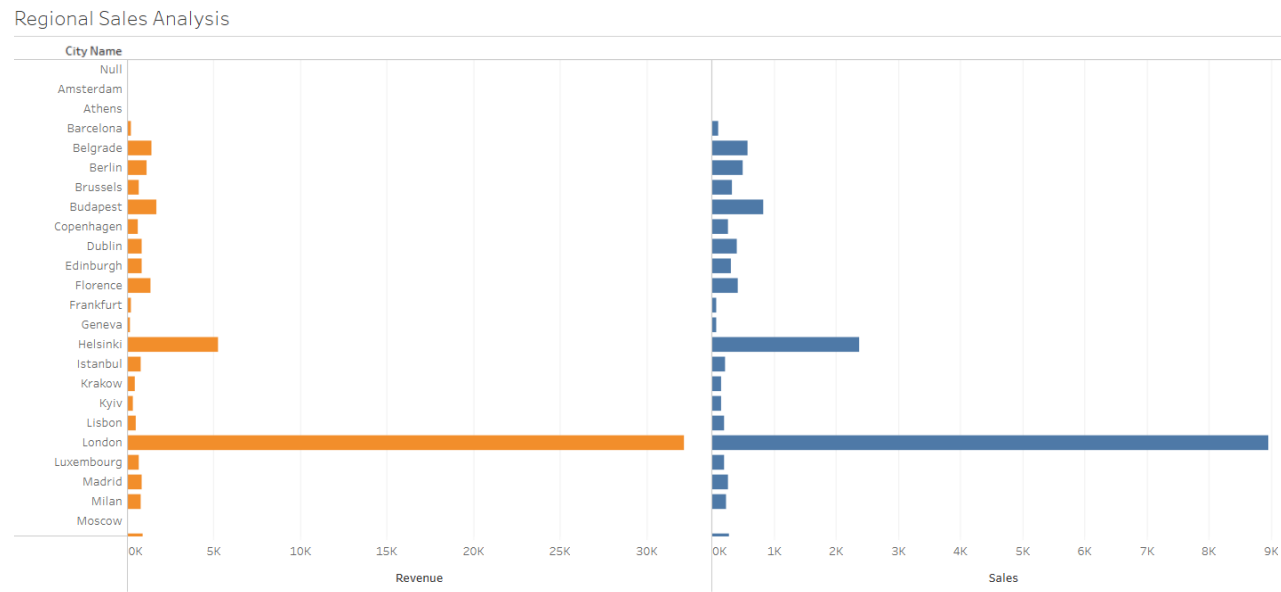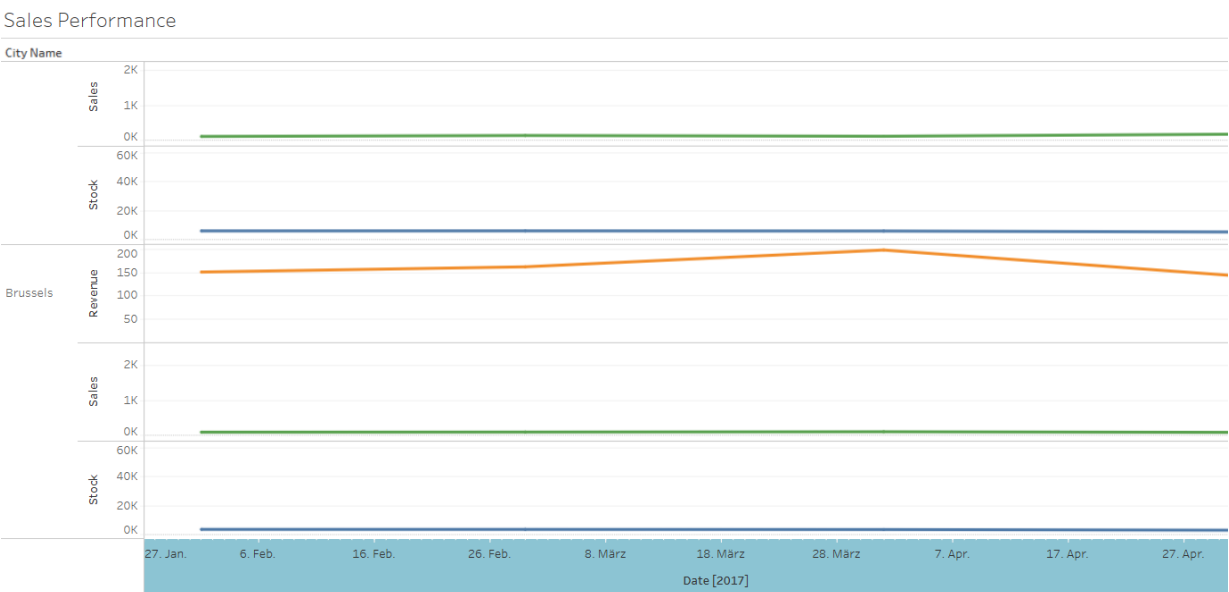**1** Data was imported & relationships established
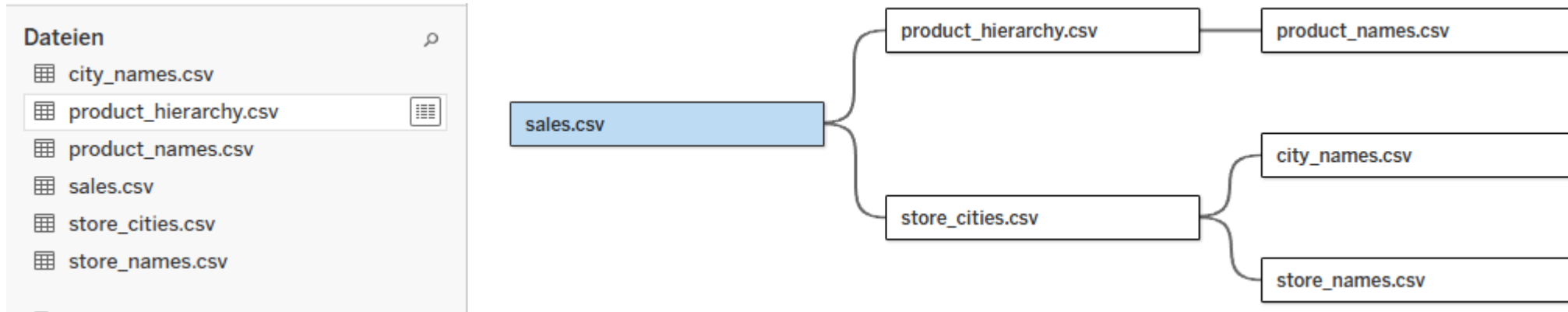


**2** Calculated fields were created

# 3 Results – Module 4 Visualization
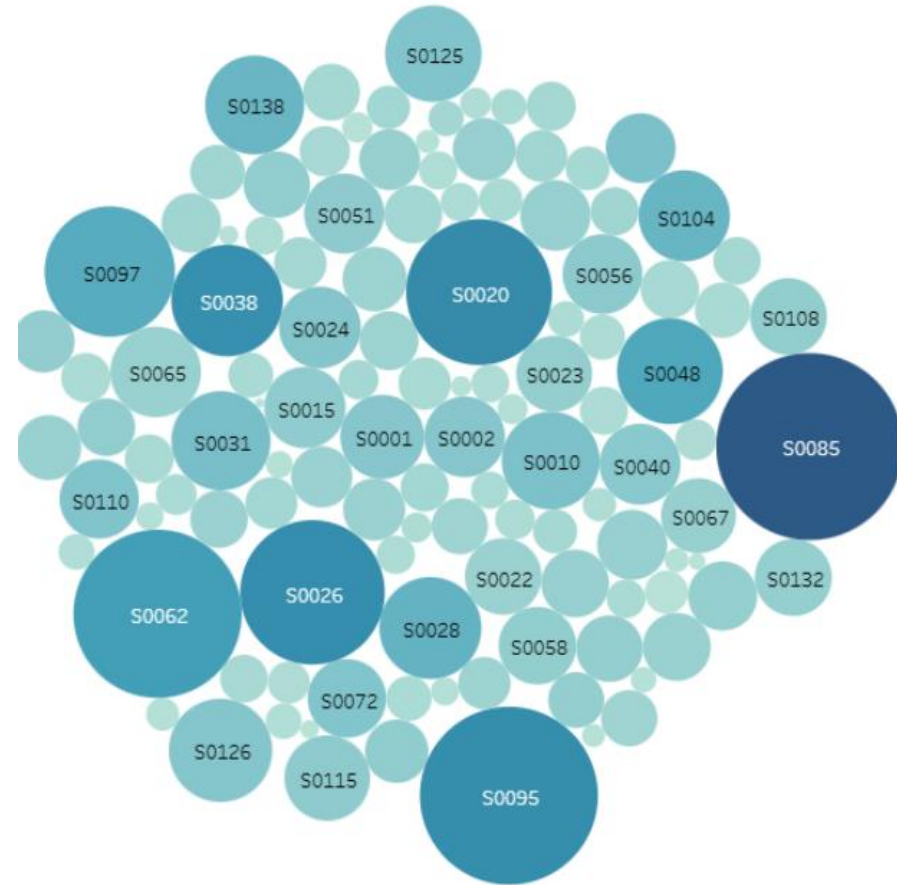
**③** Several visualizations were created

# 3 Results – Module 4 Visualization

**3** Several visualizations were created



Sales by Product



Sales by Store

# 3 Results – Module 4 Visualization

**4** Interactive Dashboard was created

# 4 Discussion – Insights 1 Geographics

- With the DIGI store doing so well in Edinborough, I would consider opening more stores following the same philosophy. The obvious location for the next store would be Glasgow. There, you can expect similar purchasing behavior based on geographical considerations.

- You could also consider turning this store into an (international) chain. I would start by trying to penetrate the UK market. Depending on market saturation, the store could very likely be well received by customers.

- Judging by the poor sales figures, I would drop or sell the two stores Currys and Euronics Lisboa in Vienna. I don't think they have a long-term profitable prospect.

# 4 Discussion – Insights 2 Products

- I would make sure that blockbusters like "NanoGrillflow" or the "Handheld Milk Frother" are always in stock in all stores. These products sell very well and are the cash cows.

- Products like the "NeoGrillPulse" don't sell at all. However, they have over 25,000 of them in stock. For such hard-to-sell products, I would start a discount campaign to boost sales and eventually remove them from the range. This can make room for products that bring in more money.

# 5 Conclusion

- Specific action were proposed. These can be proven by analysis and visualization (as seen in chapter 3 & 4)

- Of course, you can go much deeper into the analysis. However, this is beyond the scope of this Capstone project.

Personally, I was able to learn many skills throughout the certificate.

- Advanced Excel data analysis
- The application of SQL
- Tableau visualizations

I hope that I can apply theses skills in futures jobs & will benefit my career.

# Sources

[1]        Stock ID: 55646421

[2]        AI-generated on M365 Copilot

[3]        Stock ID: 47857043

[4]        Best Data Science Courses in 2022 | ServerWatch

[5]        Microsoft Excel – Wikipedia tiếng Việt

[6]        PostgreSQL | Reviews, Pricing & Demos - SoftwareAdvice GB

[7]        Tableau full logo transparent PNG – StickPNG

[8]        Stock ID: 494677327

[9]        Stock ID: 299479085

[10]       AI-generated on M365 Copilot