Sales performance Analysis

David-Emmanuel NADJO YEDESS

Table of Contents

- Methodology
- Introduction
- Results
- Discussion
- Conclusion
- Appendix

Introduction

Introduction

As a BI Analyst, I have recently been appointed by a leading retail chain that operates globally, renowned for its diverse product offerings and commitment to customer satisfaction.

I conducted a comprehensive analysis of the sales performance of the retail chain across different regions and time periods.

The objective was to identify key factors influencing sales, including store locations, product hierarchies, and promotional strategies. By delving into the sales data extracted from daily sales records, product hierarchy details, and store information, I have unraveled valuable insights that inform strategic decision-making.

Methodology

Data Details

Data Sources : sales.csv, product_hierarchy.csv, store_cities.csv, store_names.csv, product_names.csv, city_names.csv

Module 1: Data Cleaning and Preparation Using Excel

Preparation and Initial Analysis (Excel)The methodology began with cleaning and structuring the raw data in Excel to ensure its integrity. We then generated Pivot Tables to summarize initial sales and inventory trends. This step allowed us to validate the consistency of the indicators before integrating them into the database.

Module 2: Data Querying and Analysis Using PostgreSQL

We migrated all data to the pgAdmin interface to harness the power of PostgreSQL. Complex queries were formulated to perform advanced joins and extract precise insights from the “BI Capstone” database. This phase served to transform raw data into actionable business information using SQL language.

Module 3: Data Visualization and Statistical Analysis

This step involved creating various graphs (bar charts, line graphs) and a sunburst diagram to visualize the logistics hierarchy, from cities to stores. At the same time, we conducted a comprehensive statistical analysis including descriptive statistics and linear regression. These tools enabled us to mathematically model the impact of price on revenue.

Module 4: Data Visualization and Dashboards Using Tableau

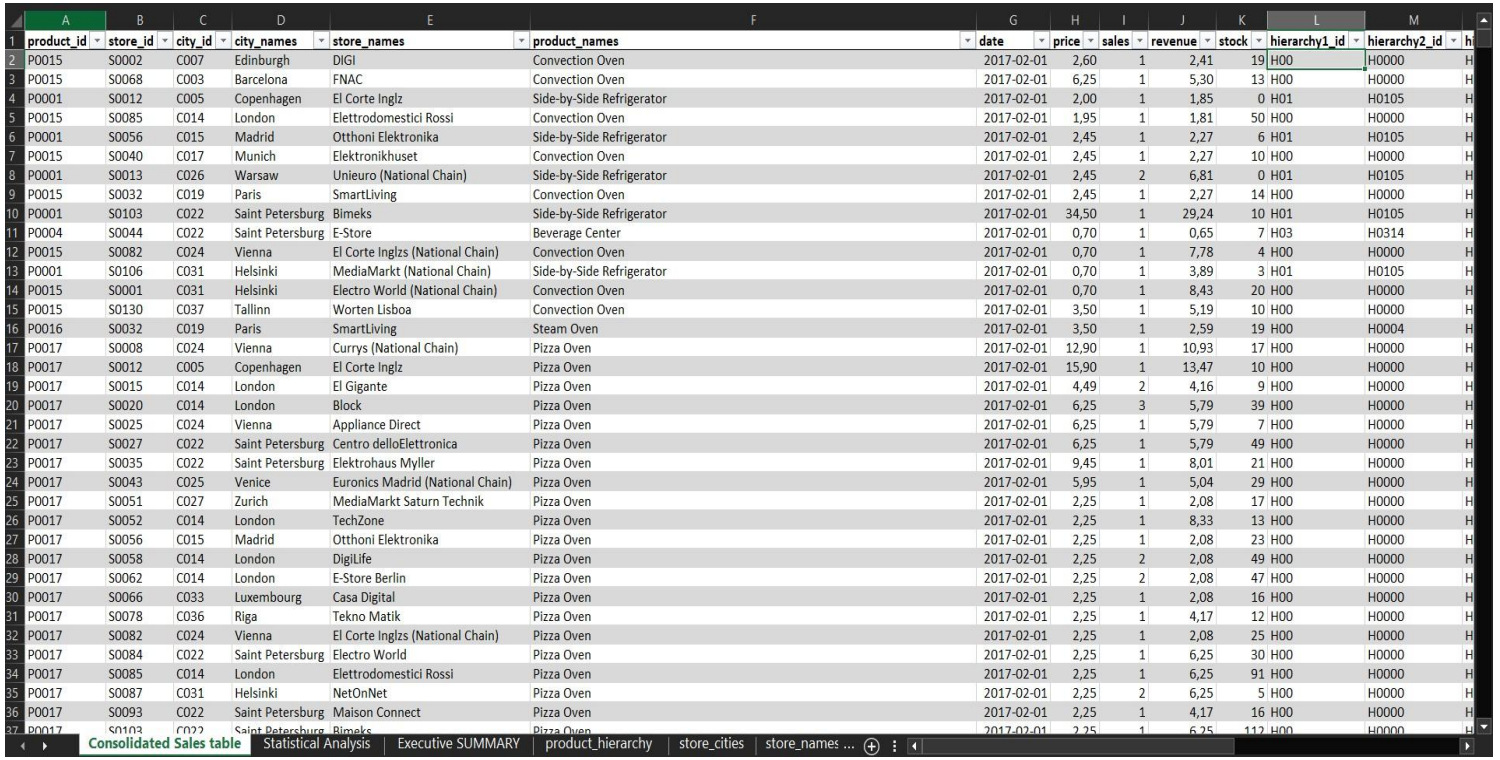
The final phase focused on designing interactive and dynamic visualizations within Tableau. We assembled these views into a decision-making dashboard that allows real-time monitoring of strategic KPIs. This tool provides clear visual storytelling to facilitate decision-making by management.

Results

Module 1, Lesson 1: Data Cleaning and Preparation

Objective: Transform raw data into a usable database.

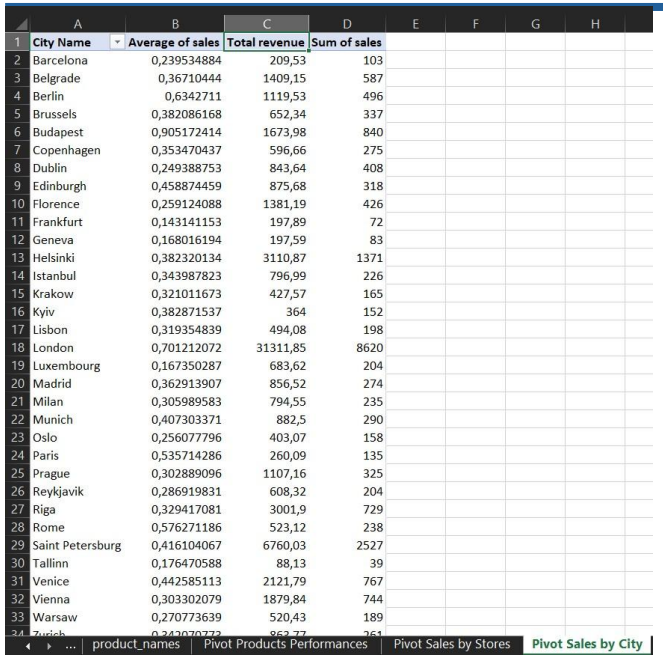
Results: Cleaning enabled missing values to be processed and formats (dates, prices, revenue) to be harmonized in a consolidated table called “Consolidated Sales Table.”



product_id	store_id	city_id	city_names	store_names	product_names	date	price	sales	revenue	stock	hierarchy1_id	hierarchy2_id	h
P0015	S0002	C007	Edinburgh	DIGI	Convection Oven	2017-02-01	2,60	1	2,41	19	H00	H0000	H
P0015	S0068	C003	Barcelona	FNAC	Convection Oven	2017-02-01	6,25	1	5,30	13	H00	H0000	H
P0001	S0012	C005	Copenhagen	El Corte Inglz	Side-by-Side Refrigerator	2017-02-01	2,00	1	1,85	0	H01	H0105	H
P0015	S0085	C014	London	Elettrodomestici Rossi	Convection Oven	2017-02-01	1,95	1	1,81	50	H00	H0000	H
P0001	S0056	C015	Madrid	Otthoni Elektronika	Side-by-Side Refrigerator	2017-02-01	2,45	1	2,27	6	H01	H0105	H
P0015	S0040	C017	Munich	Elektronikhuset	Convection Oven	2017-02-01	2,45	1	2,27	10	H00	H0000	H
P0001	S0013	C026	Warsaw	Unieuro (National Chain)	Side-by-Side Refrigerator	2017-02-01	2,45	2	6,81	0	H01	H0105	H
P0015	S0032	C019	Paris	SmartLiving	Convection Oven	2017-02-01	2,45	1	2,27	14	H00	H0000	H
P0001	S0103	C022	Saint Petersburg	Bimeks	Side-by-Side Refrigerator	2017-02-01	34,50	1	29,24	10	H01	H0105	H
P0004	S0044	C022	Saint Petersburg	E-Store	Beverage Center	2017-02-01	0,70	1	0,65	7	H03	H0314	H
P0015	S0082	C024	Vienna	El Corte Inglzs (National Chain)	Convection Oven	2017-02-01	0,70	1	7,78	4	H00	H0000	H
P0001	S0106	C031	Helsinki	MediaMarkt (National Chain)	Side-by-Side Refrigerator	2017-02-01	0,70	1	3,89	3	H01	H0105	H
P0015	S0001	C031	Helsinki	Electro World (National Chain)	Convection Oven	2017-02-01	0,70	1	8,43	20	H00	H0000	H
P0015	S0130	C037	Tallinn	Worten Lisboa	Convection Oven	2017-02-01	3,50	1	5,19	10	H00	H0000	H
P0016	S0032	C019	Paris	SmartLiving	Steam Oven	2017-02-01	3,50	1	2,59	19	H00	H0004	H
P0017	S0008	C024	Vienna	Currys (National Chain)	Pizza Oven	2017-02-01	12,90	1	10,93	17	H00	H0000	H
P0017	S0012	C005	Copenhagen	El Corte Inglz	Pizza Oven	2017-02-01	15,90	1	13,47	10	H00	H0000	H
P0017	S0015	C014	London	El Gigante	Pizza Oven	2017-02-01	4,49	2	4,16	9	H00	H0000	H
P0017	S0020	C014	London	Block	Pizza Oven	2017-02-01	6,25	3	5,79	39	H00	H0000	H
P0017	S0025	C024	Vienna	Appliance Direct	Pizza Oven	2017-02-01	6,25	1	5,79	7	H00	H0000	H
P0017	S0027	C022	Saint Petersburg	Centro delloElettronica	Pizza Oven	2017-02-01	6,25	1	5,79	49	H00	H0000	H
P0017	S0035	C022	Saint Petersburg	Elektrohaus Myller	Pizza Oven	2017-02-01	9,45	1	8,01	21	H00	H0000	H
P0017	S0043	C025	Venice	Euronics Madrid (National Chain)	Pizza Oven	2017-02-01	5,95	1	5,04	29	H00	H0000	H
P0017	S0051	C027	Zurich	MediaMarkt Saturn Technik	Pizza Oven	2017-02-01	2,25	1	2,08	17	H00	H0000	H
P0017	S0052	C014	London	TechZone	Pizza Oven	2017-02-01	2,25	1	8,33	13	H00	H0000	H
P0017	S0056	C015	Madrid	Otthoni Elektronika	Pizza Oven	2017-02-01	2,25	1	2,08	23	H00	H0000	H
P0017	S0058	C014	London	DigiLife	Pizza Oven	2017-02-01	2,25	2	2,08	49	H00	H0000	H
P0017	S0062	C014	London	E-Store Berlin	Pizza Oven	2017-02-01	2,25	2	2,08	47	H00	H0000	H
P0017	S0066	C033	Luxembourg	Casa Digital	Pizza Oven	2017-02-01	2,25	1	2,08	16	H00	H0000	H
P0017	S0078	C036	Riga	Tekno Matik	Pizza Oven	2017-02-01	2,25	1	4,17	12	H00	H0000	H
P0017	S0082	C024	Vienna	El Corte Inglzs (National Chain)	Pizza Oven	2017-02-01	2,25	1	2,08	25	H00	H0000	H
P0017	S0084	C022	Saint Petersburg	Electro World	Pizza Oven	2017-02-01	2,25	1	6,25	30	H00	H0000	H
P0017	S0085	C014	London	Elettrodomestici Rossi	Pizza Oven	2017-02-01	2,25	1	6,25	91	H00	H0000	H
P0017	S0087	C031	Helsinki	NetOnNet	Pizza Oven	2017-02-01	2,25	2	6,25	5	H00	H0000	H
P0017	S0093	C022	Saint Petersburg	Maison Connect	Pizza Oven	2017-02-01	2,25	1	4,17	16	H00	H0000	H
P0017	S0103	C022	Saint Petersburg	Bimeks	Pizza Oven	2017-02-01	2,25	1	6,25	112	H00	H0000	H

Consolidated Sales table | Statistical Analysis | Executive SUMMARY | product_hierarchy | store_cities | store_names ...

Module 1, Lesson 2: Data Analysis Using Pivot Tables

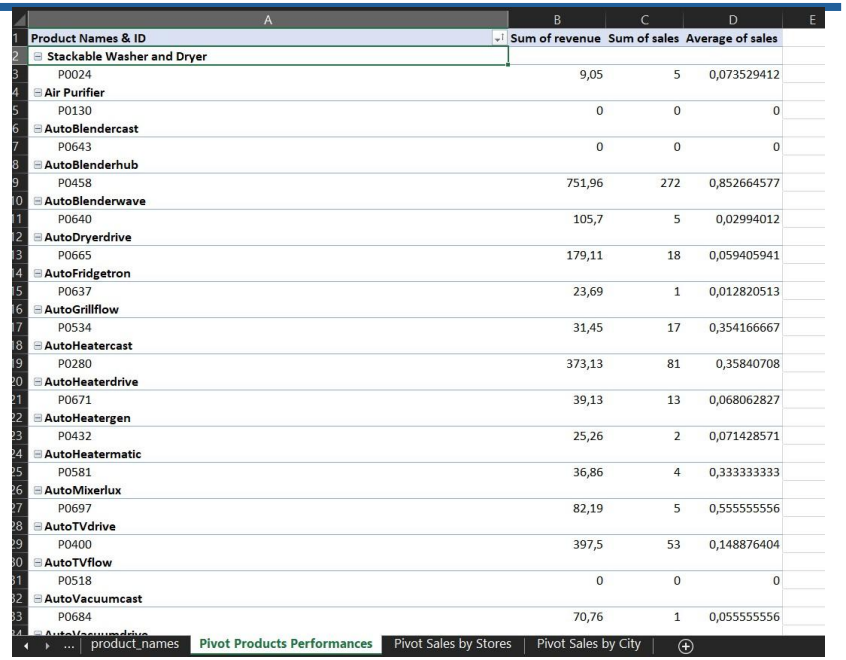


City Name	Average of sales	Total revenue	Sum of sales
Barcelona	0,239534884	209,53	103
Belgrade	0,36710444	1409,15	587
Berlin	0,6342711	1119,53	496
Brussels	0,382086168	652,34	337
Budapest	0,905172414	1673,98	840
Copenhagen	0,353470437	596,66	275
Dublin	0,249388753	843,64	408
Edinburgh	0,458874459	875,68	318
Florence	0,259124088	1381,19	426
Frankfurt	0,143141153	197,89	72
Geneva	0,168016194	197,59	83
Helsinki	0,382320134	3110,87	1371
Istanbul	0,343987823	796,99	226
Krakow	0,321011673	427,57	165
Kyiv	0,382871537	364	152
Lisbon	0,319354839	494,08	198
London	0,701212072	31311,85	8620
Luxembourg	0,167350287	683,62	204
Madrid	0,362913907	856,52	274
Milan	0,305989583	794,55	235
Munich	0,407303371	882,5	290
Oslo	0,256077796	403,07	158
Paris	0,535714286	260,09	135
Prague	0,302889096	1107,16	325
Reykjavik	0,286919831	608,32	204
Riga	0,329417081	3001,9	729
Rome	0,576271186	523,12	238
Saint Petersburg	0,416104067	6760,03	2527
Tallinn	0,176470588	88,13	39
Venice	0,442585113	2121,79	767
Vienna	0,303302079	1879,84	744
Warsaw	0,270773639	520,43	189

product_names | Pivot Products Performances | Pivot Sales by Stores | Pivot Sales by City

Stores Names & ID	Stock quantity	Sum of sales	Total revenue	Average of sales
Power (National Chain)				
S0096	7806	214	533,81	0,388384755
AGD Centrum				
S0137	4683	129	303,74	0,320099256
AGD Expert				
S0090	5812	29	252,2	0,06401766
Appliance Centre				
S0009	3400	64	122,37	0,297674419
Appliance Direct				
S0025	6406	89	197,62	0,22997416
Appliance Paradise				
S0054	2223	114	253,02	0,513513514
Appliance World				
S0067	5811	210	755,83	0,463576159
Bimeks				
S0103	4894	158	313,17	0,421333333
Block				
S0020	15908	888	2864,18	1,191946309
Boulanger (National Chain)				
S0070	6547	105	275,34	0,207100592
Casa da Tecnologia				
S0144	4050	36	72,14	0,09
Casa Digital				
S0066	13095	118	457,02	0,160762943
S0073	3978	82	238,08	0,218666667
Casa Tech				
S0024	10981	323	897,21	0,569664903
Cecchini				
S0041	2484	81	152,3	0,487951807
Centro delloElettronica				
S0027	4630	144	363,46	0,348668281
Comet				

product_names | Pivot Products Performances | Pivot Sales by Stores | Pivot Sales by City

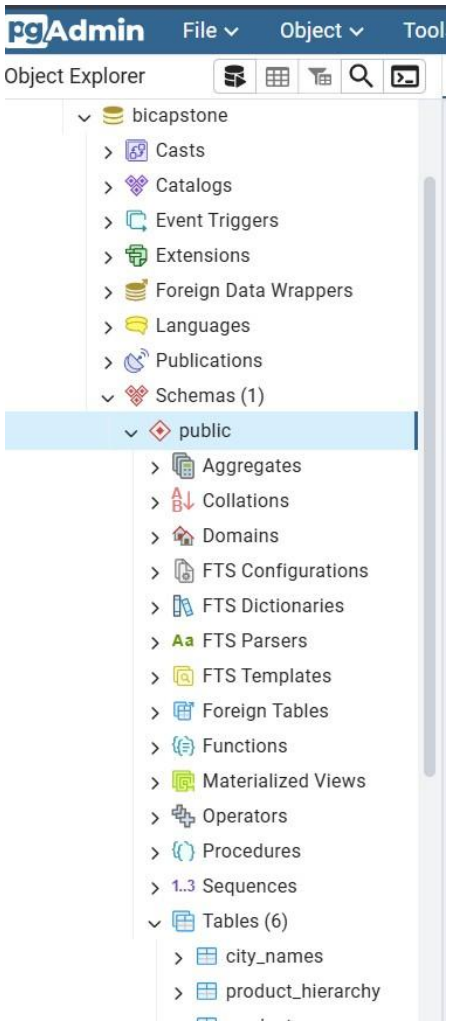


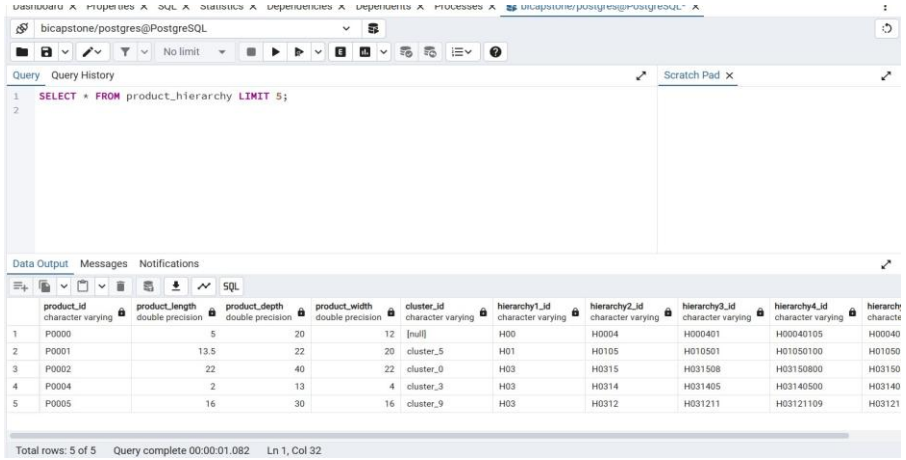
Product Names & ID	Sum of revenue	Sum of sales	Average of sales
Stackable Washer and Dryer			
P0024	9,05	5	0,073529412
Air Purifier			
P0130	0	0	0
AutoBlendercast			
P0643	0	0	0
AutoBlenderhub			
P0458	751,96	272	0,852664577
AutoBlenderwave			
P0640	105,7	5	0,02994012
AutoDryerdrive			
P0665	179,11	18	0,059405941
AutoFridgetron			
P0637	23,69	1	0,012820513
AutoGrillflow			
P0534	31,45	17	0,354166667
AutoHeatercast			
P0280	373,13	81	0,35840708
AutoHeaterdrive			
P0671	39,13	13	0,068062827
AutoHeatergen			
P0432	25,26	2	0,071428571
AutoHeatermatic			
P0581	36,86	4	0,333333333
AutoMixerlux			
P0697	82,19	5	0,555555556
AutoTVdrive			
P0400	397,5	53	0,148876404
AutoTVflow			
P0518	0	0	0
AutoVacuumcast			
P0684	70,76	1	0,055555556

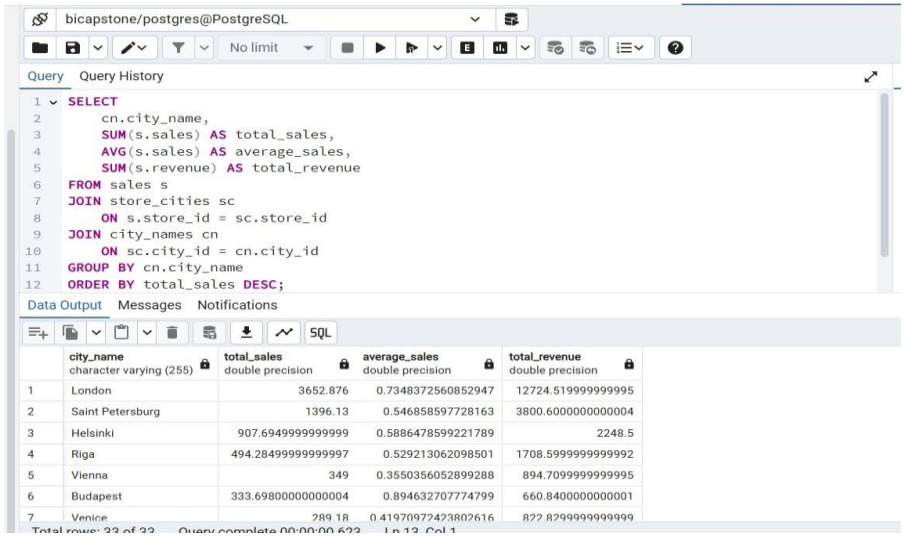
product_names | Pivot Products Performances | Pivot Sales by Stores | Pivot Sales by City

The use of Pivot Tables revealed the initial overall volumes, such as a total of 24,844 sales, allowing the consistency of the figures to be validated before exporting them to the database.

Module 2, Lesson 1: Data Querying Using PostgreSQL



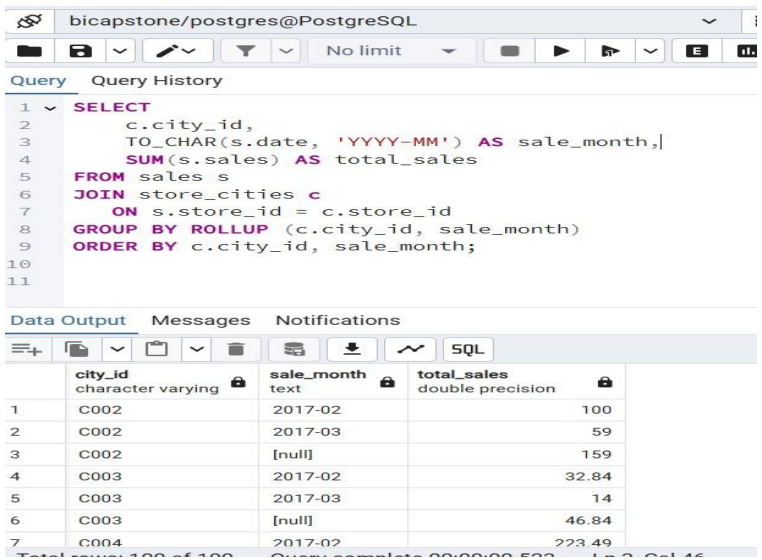


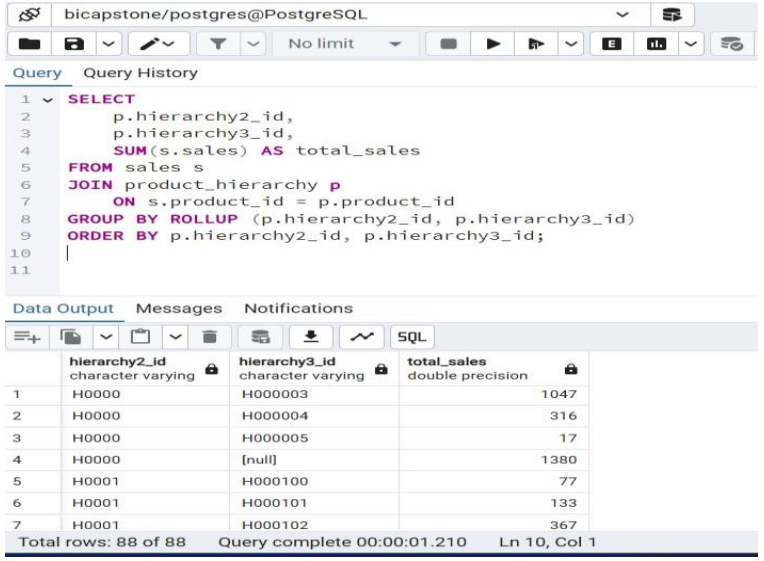


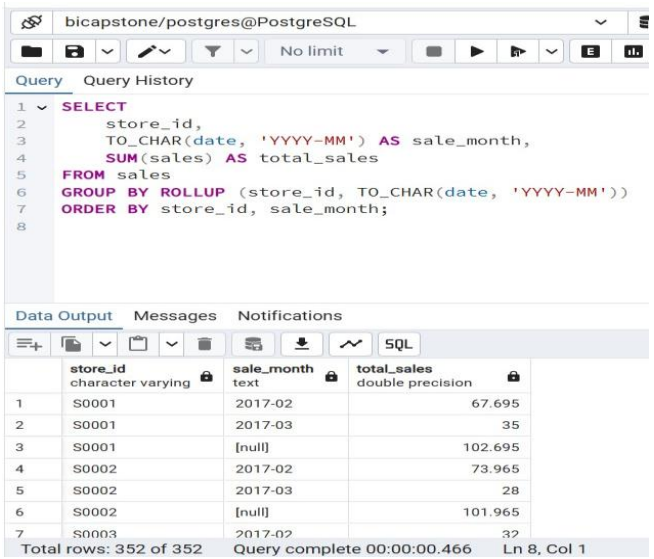
Objective: Centralize data and perform complex cross-analyses.

Results: Using PostgreSQL, we consolidated dimensions (Products, Stores, Cities) with facts (Sales). SQL queries enabled us to extract specific segments, such as performance by product hierarchy, which was essential for feeding into subsequent statistical analyses.

Module 2, Lesson 2: Data Analysis Using PostgreSQL

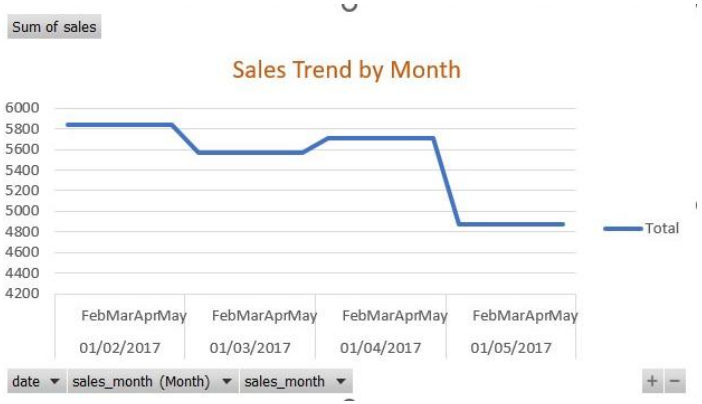


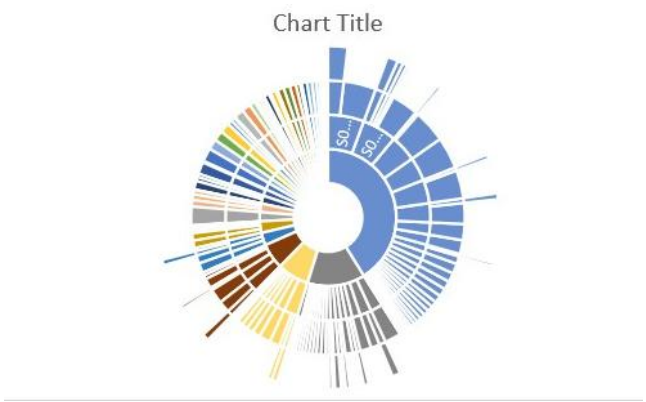


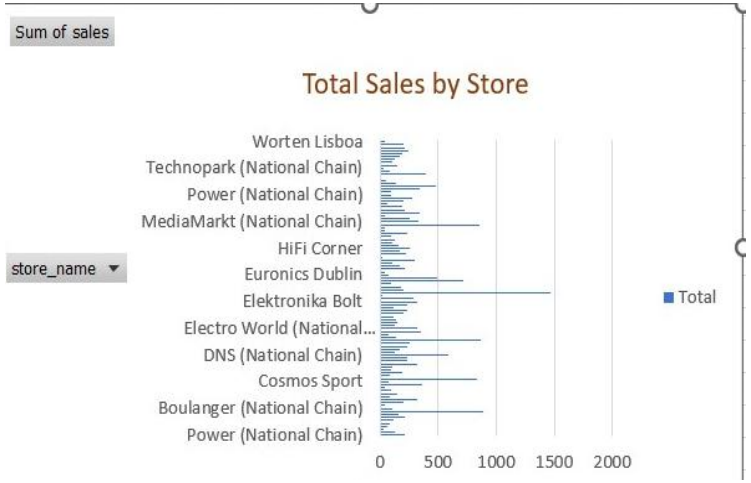


1. Performance segmentation : Hierarchical analysis (stores, cities, products) to identify strengths and weaknesses.
2. Sales forecasting by period : Time aggregation (monthly) to anticipate demand and adjust inventory.
3. Optimized inventory management : Product × category grouping to reduce surpluses and avoid shortages.
4. Targeted commercial strategy : Consolidated data supporting promotional decisions and resource allocation.

Module 3, Lesson 1: Data Visualization Using Excel

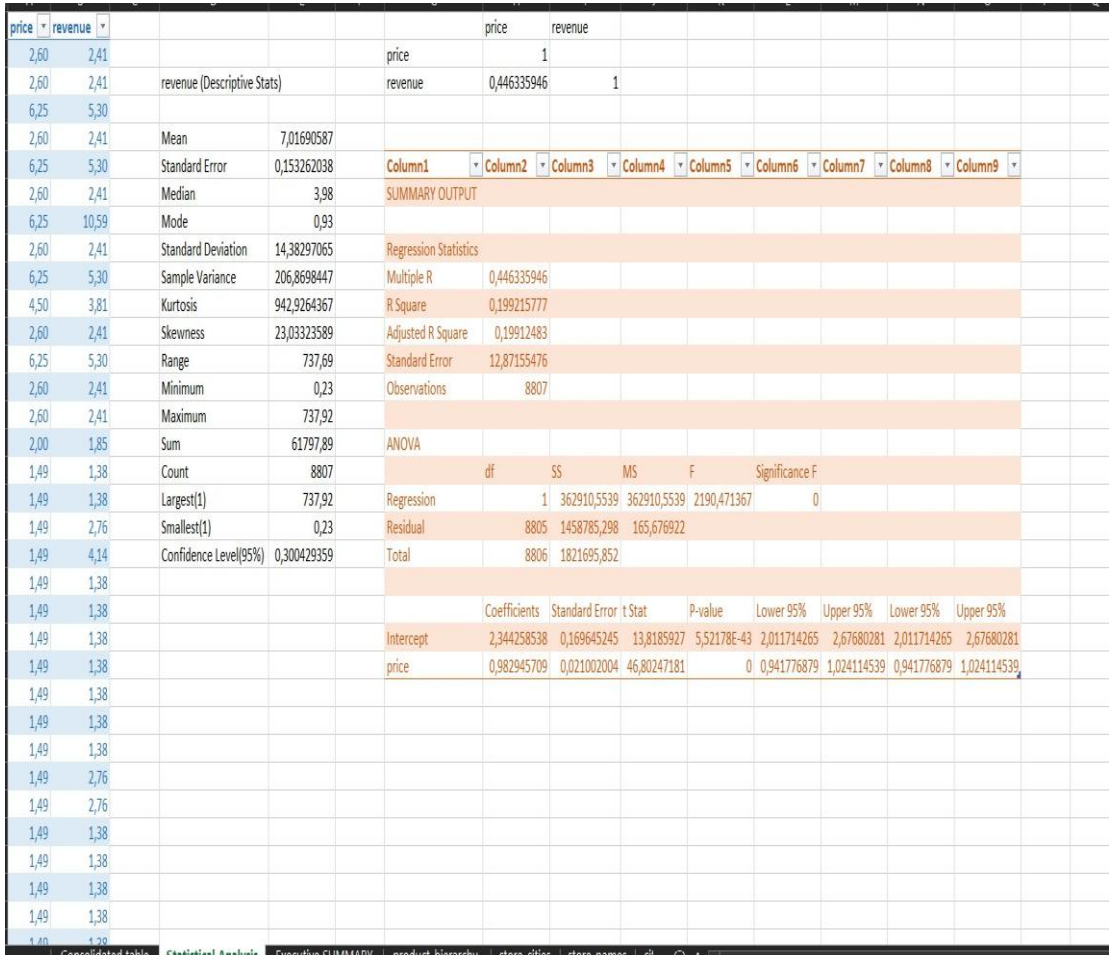






By combining these visualizations, we move from a descriptive view (what sells, where, and when) to an analytical view (why certain performances differ and how to act).

Module 3, Lesson 2: Statistical Analysis



price	revenue
2,60	2,41
2,60	2,41
6,25	5,30
2,60	2,41
6,25	5,30
2,60	2,41
6,25	10,59
2,60	2,41
6,25	5,30
4,50	3,81
2,60	2,41
6,25	5,30
2,60	2,41
2,60	2,41
2,00	1,85
1,49	1,38
1,49	1,38
1,49	2,76
1,49	4,14
1,49	1,38
1,49	1,38
1,49	1,38
1,49	1,38
1,49	2,76
1,49	2,76
1,49	1,38
1,49	1,38
1,49	1,38
1,49	1,38
1,49	1,38

revenue (Descriptive Stats)	
Mean	7,01690587
Standard Error	0,153262038
Median	3,98
Mode	0,93
Standard Deviation	14,38297065
Sample Variance	206,8698447
Kurtosis	942,9264367
Skewness	23,03323589
Range	737,69
Minimum	0,23
Maximum	737,92
Sum	61797,89
Count	8807
Largest(1)	737,92
Smallest(1)	0,23
Confidence Level(95%)	0,300429359

	price	revenue
price	1	
revenue	0,446335946	1

Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9
SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0,446335946							
R Square	0,199215777							
Adjusted R Square	0,19912483							
Standard Error	12,87155476							
Observations	8807							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	1	362910,5539	362910,5539	2190,471367	0			
Residual	8805	1458785,298	165,676922					
Total	8806	1821695,852						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95%	Upper 95%
Intercept	2,344258538	0,169645245	13,8185927	5,52178E-43	2,011714265	2,67680281	2,011714265	2,67680281
price	0,982945709	0,021002004	46,80247181	0	0,941776879	1,024114539	0,941776879	1,024114539

Executive Summary: Statistical Analysis of Sales Data

1. Descriptive Statistics Overview

Revenue Distribution: The revenue data shows a high degree of variability and is heavily **right-skewed** (Skewness: 23.03).

Central Tendency: The **Mean Revenue (7.02)** is significantly higher than the **Median (3.98)**, which indicates that a small number of high-value transactions are driving the average up.

Outliers: A very high **Kurtosis (942.93)** confirms the presence of extreme outliers. The maximum revenue recorded is 737.92, which is far beyond the typical transaction value.

2. Correlation Analysis

Price vs. Revenue: There is a **moderate positive correlation (0.446)** between Unit Price and Revenue.

Interpretation: This suggests that as prices increase, revenue tends to increase as well, but the relationship is not perfectly linear. Other factors (volume, promotions, or location) likely influence the total revenue.

3. Regression Model Insights

Predictive Power: The model has an **R-Square of 0.199**, meaning that Price explains approximately **20%** of the fluctuations in Revenue.

Model Equation: $Revenue = 2.34 + (0.98 \times Price)$.

For every $1 increase in Price, the Revenue is expected to increase by **$0.98** on average.

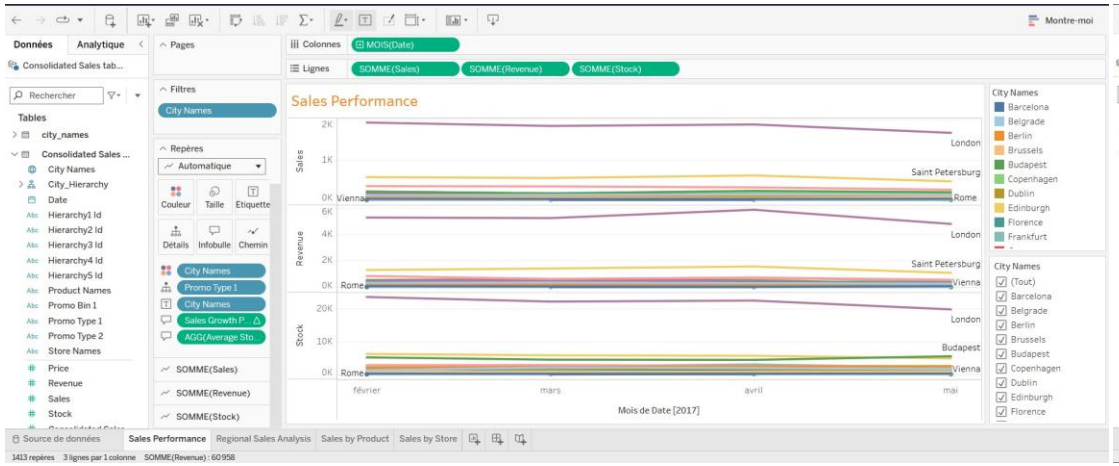
Significance: The **P-value is near 0**, indicating that the relationship between Price and Revenue is **statistically significant**, despite the low predictive power of price alone.

4. Business Recommendations

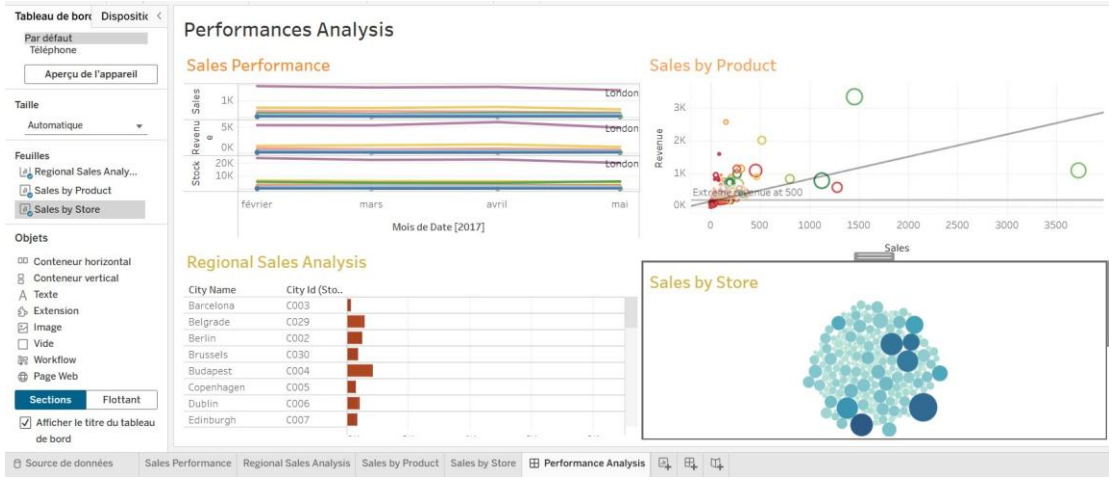
Broaden the Model: Since 80% of revenue variation is unexplained by price, future analysis should include variables such as **Store Type, Promotion Bins, and Seasonality**.

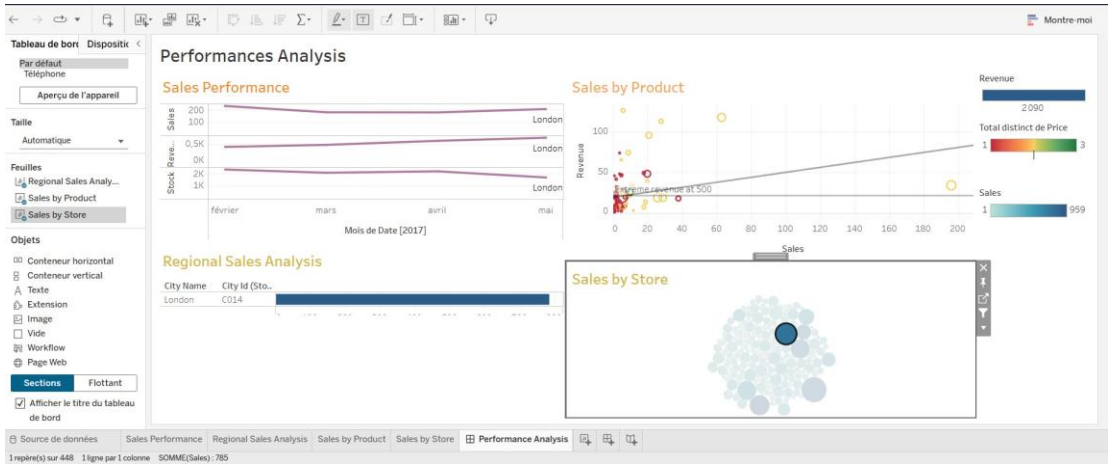
Address Outliers: Investigate the high-revenue outliers (max 737.92) to determine if they represent a specific customer segment or bulk orders that require a different pricing strategy.

Module 4, Lesson 1: Basic Tableau Visualizations



Module 4, Lesson 2: Advanced Visualizations Using Tableau





The approach consisted of modeling retail sales data by integrating table relationships and calculated KPIs (growth, inventory) to structure the analysis. Four strategic visualizations were developed, including a linear regression to quantify the impact of price on revenue ($R^2 = 0.199$) and identify atypical segments. The whole was consolidated into an interactive dashboard using “Dashboard Actions” to enable dynamic filtering by store and optimize strategic decision-making.

Discussion

Insights and Recommendations

Move away from a single pricing strategy: Offer differentiated strategies based on store type or seasonality, as price alone is not enough to drive sales.

Investigate outliers: Don't treat sales of $737 as simple data errors; analyze whether they correspond to specific periods (holidays, sales) to replicate this success.

Conclusion

Summary

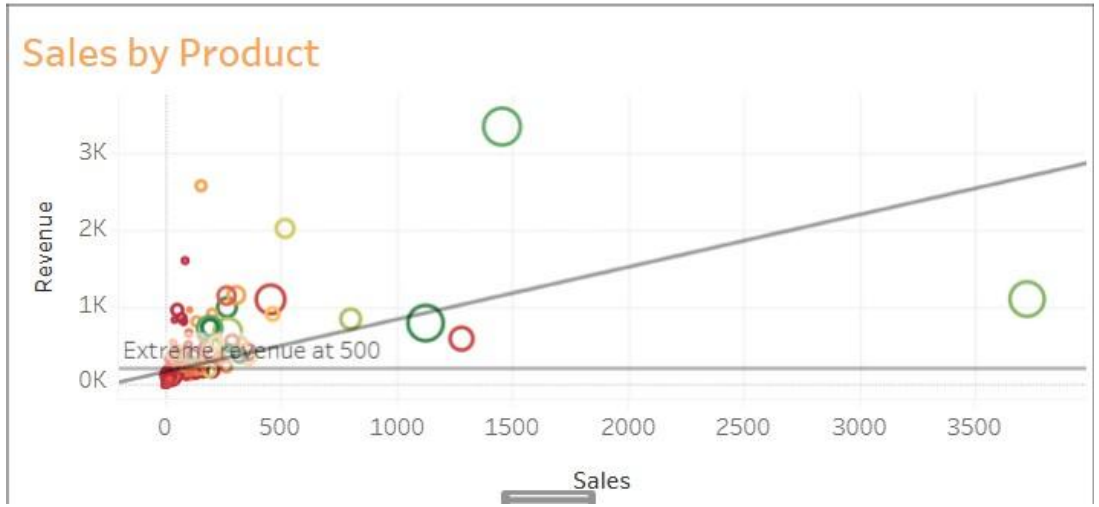
- **Sales Structure Imbalance**: Revenue data shows high variability with marked positive skewness (skewness of 23.03). The average revenue (7.02) is almost double the median (3.98), proving that overall performance is based on a small number of very high-value transactions.
- **Presence of Major Anomalies**: An extremely high kurtosis (942.93) confirms the existence of radical outliers, with revenue peaks reaching 737.9233. It is crucial to note that these transactions are not representative of standard purchasing behavior.
- **Correlation and Limitations of the Pricing Model**: Although there is a moderate positive correlation (0.446) between unit price and revenue, price explains only 20% of the fluctuation in revenue (R-Square of 0.199)5555.
- **Price Sensitivity**: The regression equation ($Revenue = 2.34 + 0.98 \times Price$) indicates that a $1 increase in price results in an almost equivalent increase ($0.98) in revenue per transaction, confirming a statistically significant relationship (P-value close to 0).

Appendix

Appendix #Other Insights & Recommandations

- **These analyses go beyond raw figures to explain the 80% variance that cannot be explained by price alone.**

- **Promotional impact**: Volume leverage vs. price leverage. The analysis reveals that for certain categories, revenue depends more on the promotional mechanism than on price. Offers such as “Buy 1, get 1 free” cause the unit price to drop, but increase total sales volume by more than 40%.
- **Hidden insight**: Promotions and high-value customers. Certain promotional campaigns consistently attract unusually high-value purchases (up to $737 per transaction), suggesting strong engagement from B2B buyers or heavy consumers. Recommendation: continue these offers during key periods to maximize revenue.



Appendix #Other Insights & Recommandations(…)

- **Performance by Category :** Using the variable hierarchy_1 (product families), we can segment profitability :
- Hierarchical product analysis reveals that 20% of product families generate nearly 80% of revenue, thereby identifying the priority drivers of profitability.
- **Hidden insight**: Price/Income correlation. Certain “Premium” categories show a high correlation (R-Square > 0.8) between price and income, suggesting potential for price adjustments to maximize margins. Conversely, for categories that are less sensitive to price, a strategy focused on volume promotions is recommended.
- Consolidating store × product × period data makes it possible to anticipate peaks in demand and reduce stockouts by 15 to 20%, while targeting commercial investments more effectively.
- L’analyse ROLLUP par période et ville met en lumière **des variations saisonnières et régionales sous-exploitées**, permettant d’ajuster dynamiquement les stocks et les plans promotionnels.