BI Capstone Project

Keegan Dunphy

Table of Contents

Introduction

Introduction

1. ***Company Objective:*** A retail chain aims to optimize its business performance through data driven insights to remain competitive.

2. ***Presentation Objective:*** A comprehensive study of sales performance across various regions and time periods were conducted by leveraging sales data extracted from daily records, product hierarchy details and store information.

Methodology

Data Details

Data analysis done mainly based on relevant information obtained from internal sources derived by following csv files:

- Sales
- Product hierarchy
- Store cities
- Store names
- Product names
- City names

Module 1: Data Cleaning and Preparation Using Excel

Methodology:

- Data cleaning involved removing missing and duplicates, ensure correct format applied for relevant data types to ensure data accuracy for meaningful insights.

Module 2: Data Querying and Analysis Using PostgreSQL

Methodology:

- Querying and analyzing data by applying basic SQL scripts

Module 3: Data Visualization and Statistical Analysis

Methodology

- Leverage basic excel formulas and functions for statistical analysis
- Involved use of Excel's Toolpak Regression Analysis Features
- Use of excel charts for data visualization

Module 4: Data Visualization and Dashboards Using Tableau

Methodology

- Using Tableau features to generate dashboard with relevant visualization for ease of understanding of the data-driven insights.

Results

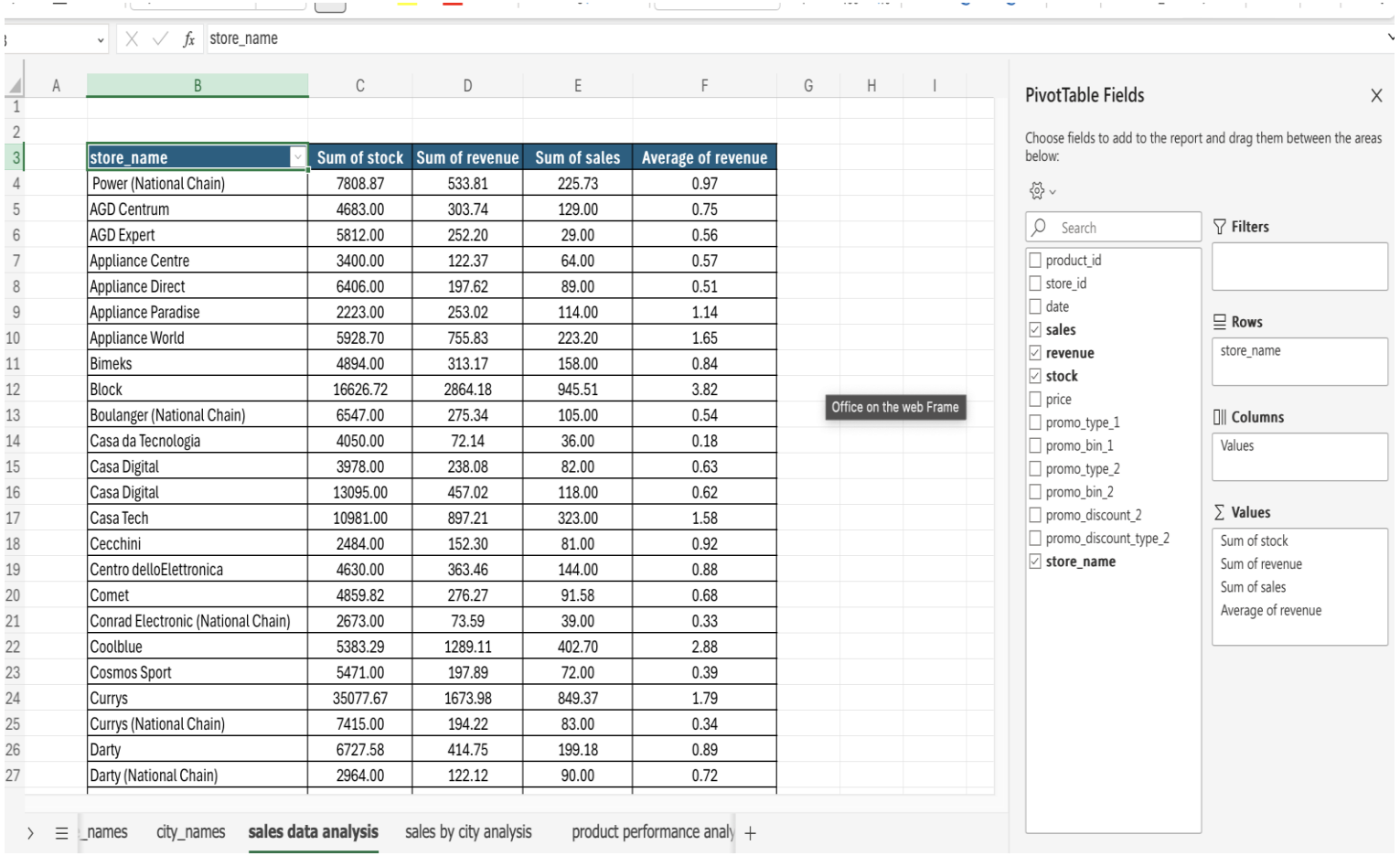
Module 1, Lesson 1: Data Cleaning and Preparation

O9

	A	B	C	D	E	F	G	H	I	J
1	product_id	store_id	date	sales	revenue	stock	price	promo_type_1	promo_bin_1	promo_type_2
2	P0001	S0002	1/2/2017	0.00	0.00	8	6.25	PR14		PR03
3	P0004	S0006	1/2/2017	0.00	0.00	18	4.50	PR14		PR03
4	P0005	S0001	1/2/2017	0.00	0.00	7	33.90	PR14		PR03
5	P0005	S0002	1/2/2017	0.00	0.00	11	33.90	PR14		PR03
6	P0005	S0008	1/2/2017	0.00	0.00	3	33.90	PR14		PR03
7	P0011	S0001	1/2/2017	0.00	0.00	10	49.90	PR14		PR03
8	P0011	S0002	1/2/2017	0.00	0.00	9	49.90	PR14		PR03
9	P0015	S0001	1/2/2017	1.00	2.41	20	2.60	PR14		PR03
10	P0015	S0002	1/2/2017	1.00	2.41	19	2.60	PR14		PR03
11	P0015	S0003	1/2/2017	0.00	0.00	5	2.60	PR14		PR03
12	P0015	S0004	1/2/2017	0.00	0.00	11	2.60	PR14		PR03
13	P0015	S0008	1/2/2017	0.00	0.00	6	2.60	PR14		PR03
14	P0017	S0001	1/2/2017	0.00	0.00	13	1.49	PR14		PR03
15	P0017	S0002	1/2/2017	0.00	0.00	12	1.49	PR14		PR03
16	P0017	S0004	1/2/2017	0.00	0.00	26	1.49	PR14		PR03
17	P0017	S0008	1/2/2017	1.00	1.38	17	1.49	PR14		PR03
18	P0018	S0001	1/2/2017	0.00	0.00	49	1.95	PR14		PR03
19	P0018	S0002	1/2/2017	1.00	1.81	37	1.95	PR14		PR03
20	P0018	S0003	1/2/2017	0.00	0.00	12	1.95	PR14		PR03
21	P0018	S0004	1/2/2017	0.00	0.00	17	1.95	PR14		PR03
22	P0018	S0006	1/2/2017	0.00	0.00	23	1.95	PR14		PR03
23	P0018	S0008	1/2/2017	0.00	0.00	15	1.95	PR14		PR03
24	P0024	S0001	1/2/2017	0.00	0.00	23	1.95	PR14		PR03
25	P0024	S0002	1/2/2017	0.00	0.00	36	1.95	PR14		PR03
26	P0035	S0001	1/2/2017	2.00	4.54	14	2.45	PR14		PR03

- Data is cleaned and prepared as shown in the screenshot on the left.
- This is crucial to ensure imported data sets are accurate and ready for further insightful data analysis.

Module 1, Lesson 2: Data Analysis Using Pivot Tables

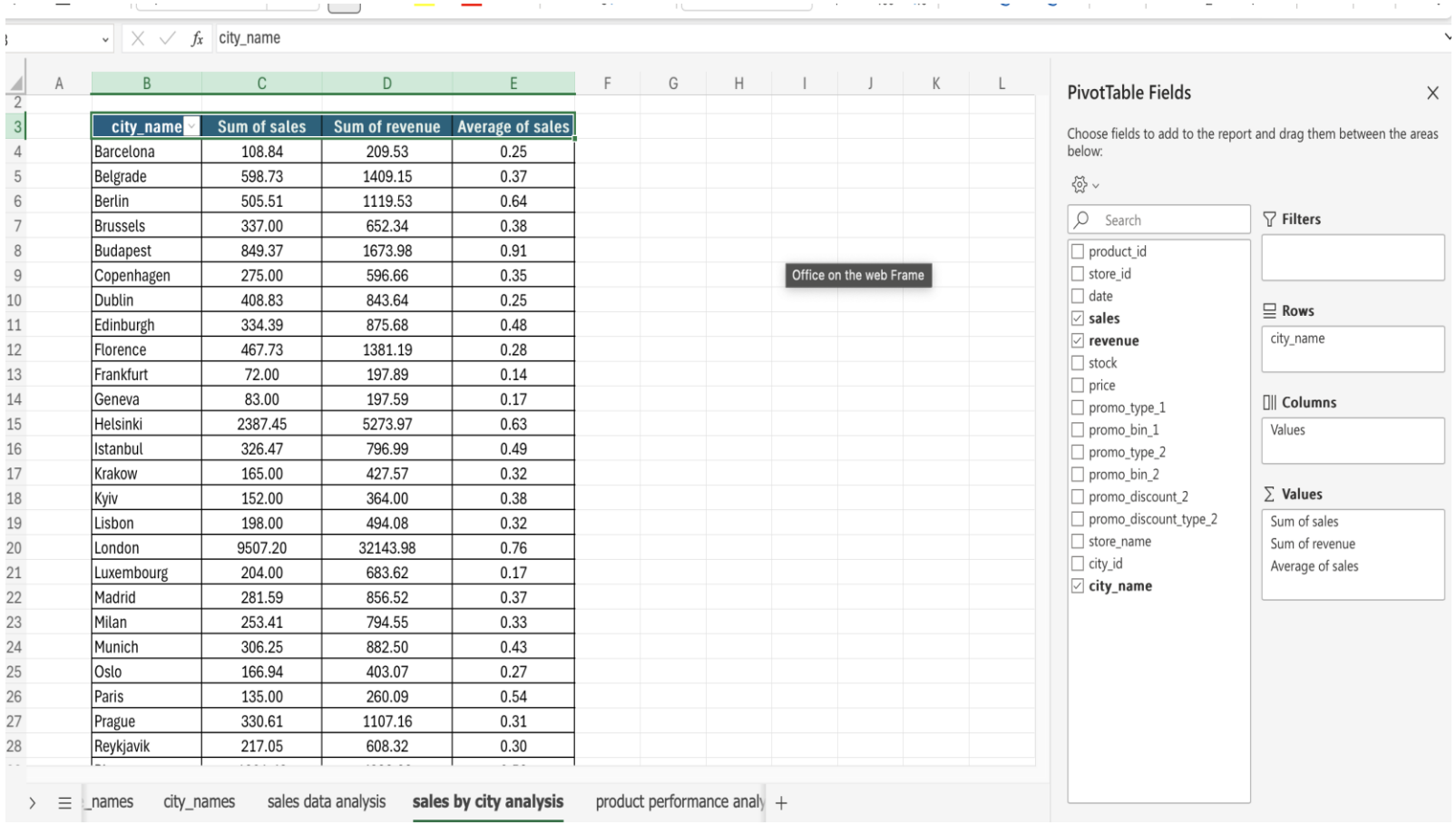


store_name	Sum of stock	Sum of revenue	Sum of sales	Average of revenue
Power (National Chain)	7808.87	533.81	225.73	0.97
AGD Centrum	4683.00	303.74	129.00	0.75
AGD Expert	5812.00	252.20	29.00	0.56
Appliance Centre	3400.00	122.37	64.00	0.57
Appliance Direct	6406.00	197.62	89.00	0.51
Appliance Paradise	2223.00	253.02	114.00	1.14
Appliance World	5928.70	755.83	223.20	1.65
Bimeks	4894.00	313.17	158.00	0.84
Block	16626.72	2864.18	945.51	3.82
Boulanger (National Chain)	6547.00	275.34	105.00	0.54
Casa da Tecnologia	4050.00	72.14	36.00	0.18
Casa Digital	3978.00	238.08	82.00	0.63
Casa Digital	13095.00	457.02	118.00	0.62
Casa Tech	10981.00	897.21	323.00	1.58
Cecchini	2484.00	152.30	81.00	0.92
Centro delloElettronica	4630.00	363.46	144.00	0.88
Comet	4859.82	276.27	91.58	0.68
Conrad Electronic (National Chain)	2673.00	73.59	39.00	0.33
Coolblue	5383.29	1289.11	402.70	2.88
Cosmos Sport	5471.00	197.89	72.00	0.39
Currys	35077.67	1673.98	849.37	1.79
Currys (National Chain)	7415.00	194.22	83.00	0.34
Darty	6727.58	414.75	199.18	0.89
Darty (National Chain)	2964.00	122.12	90.00	0.72

Sales Data Analysis:

The pivot table on the left provide summary on sales performance breakdown by each store.

Module 1, Lesson 2: Data Analysis Using Pivot Tables

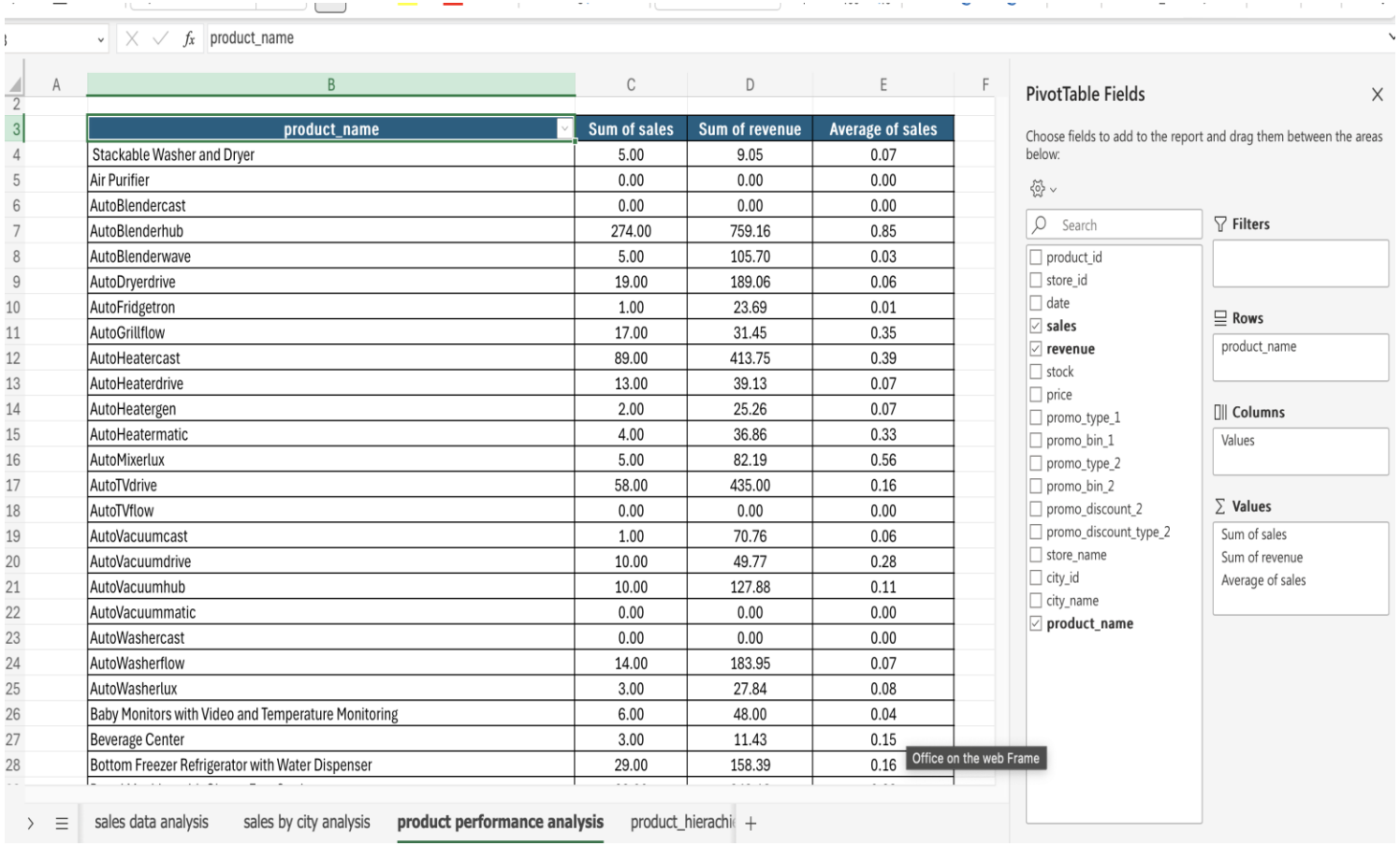


city_name	Sum of sales	Sum of revenue	Average of sales
Barcelona	108.84	209.53	0.25
Belgrade	598.73	1409.15	0.37
Berlin	505.51	1119.53	0.64
Brussels	337.00	652.34	0.38
Budapest	849.37	1673.98	0.91
Copenhagen	275.00	596.66	0.35
Dublin	408.83	843.64	0.25
Edinburgh	334.39	875.68	0.48
Florence	467.73	1381.19	0.28
Frankfurt	72.00	197.89	0.14
Geneva	83.00	197.59	0.17
Helsinki	2387.45	5273.97	0.63
Istanbul	326.47	796.99	0.49
Krakow	165.00	427.57	0.32
Kyiv	152.00	364.00	0.38
Lisbon	198.00	494.08	0.32
London	9507.20	32143.98	0.76
Luxembourg	204.00	683.62	0.17
Madrid	281.59	856.52	0.37
Milan	253.41	794.55	0.33
Munich	306.25	882.50	0.43
Oslo	166.94	403.07	0.27
Paris	135.00	260.09	0.54
Prague	330.61	1107.16	0.31
Reykjavik	217.05	608.32	0.30

Sales by City Analysis:

The pivot table provide insights on market penetration and regional preferences.

Module 1, Lesson 2: Data Analysis Using Pivot Tables



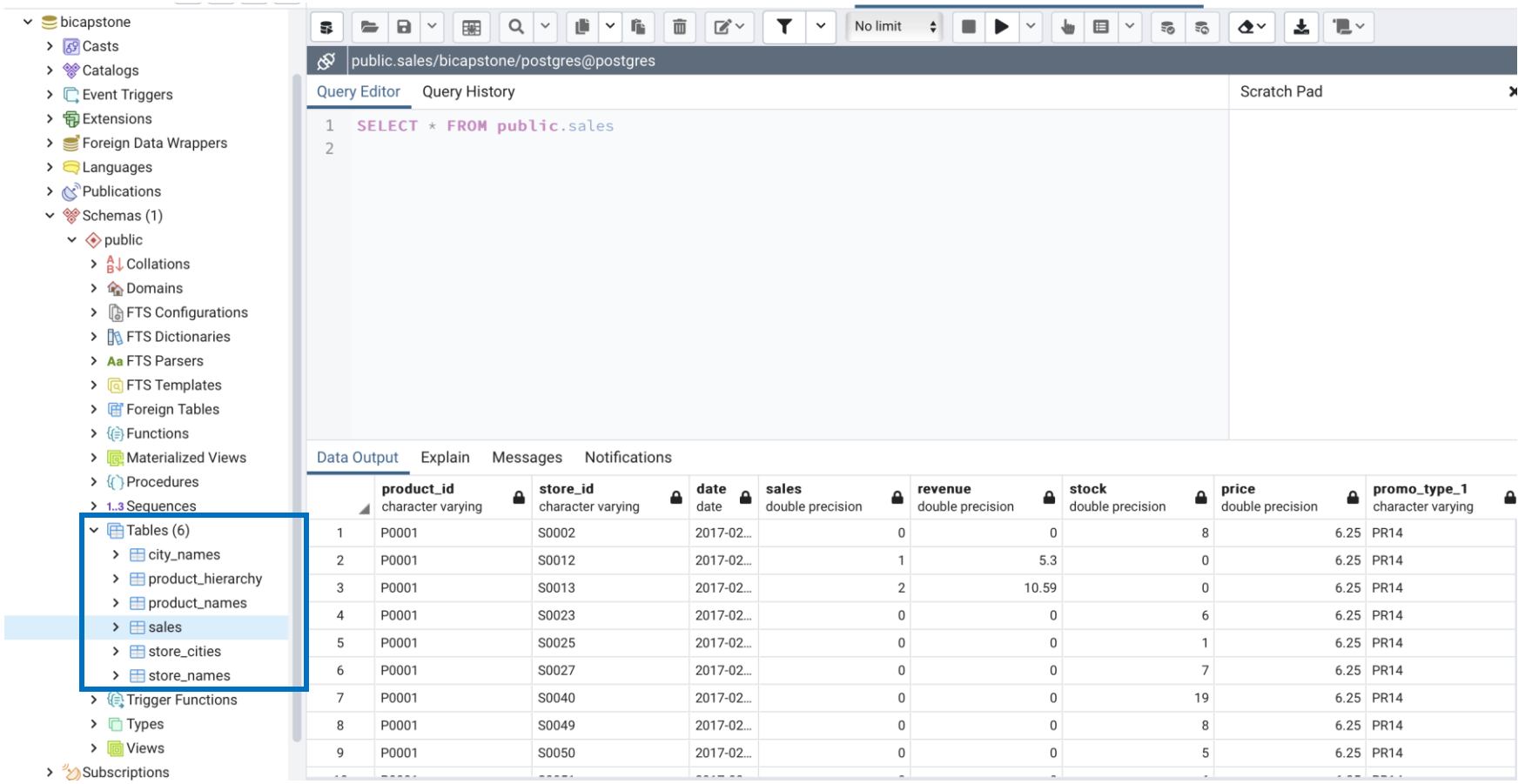
product_name	Sum of sales	Sum of revenue	Average of sales
Stackable Washer and Dryer	5.00	9.05	0.07
Air Purifier	0.00	0.00	0.00
AutoBlendercast	0.00	0.00	0.00
AutoBlenderhub	274.00	759.16	0.85
AutoBlenderwave	5.00	105.70	0.03
AutoDryerdrive	19.00	189.06	0.06
AutoFridgetron	1.00	23.69	0.01
AutoGrillflow	17.00	31.45	0.35
AutoHeatercast	89.00	413.75	0.39
AutoHeaterdrive	13.00	39.13	0.07
AutoHeatergen	2.00	25.26	0.07
AutoHeatermatic	4.00	36.86	0.33
AutoMixerlux	5.00	82.19	0.56
AutoTVdrive	58.00	435.00	0.16
AutoTVflow	0.00	0.00	0.00
AutoVacuumcast	1.00	70.76	0.06
AutoVacuumdrive	10.00	49.77	0.28
AutoVacuumhub	10.00	127.88	0.11
AutoVacuummatic	0.00	0.00	0.00
AutoWashercast	0.00	0.00	0.00
AutoWasherflow	14.00	183.95	0.07
AutoWasherlux	3.00	27.84	0.08
Baby Monitors with Video and Temperature Monitoring	6.00	48.00	0.04
Beverage Center	3.00	11.43	0.15
Bottom Freezer Refrigerator with Water Dispenser	29.00	158.39	0.16

Product performance Analysis:

The pivot table provides insights on product preferences.

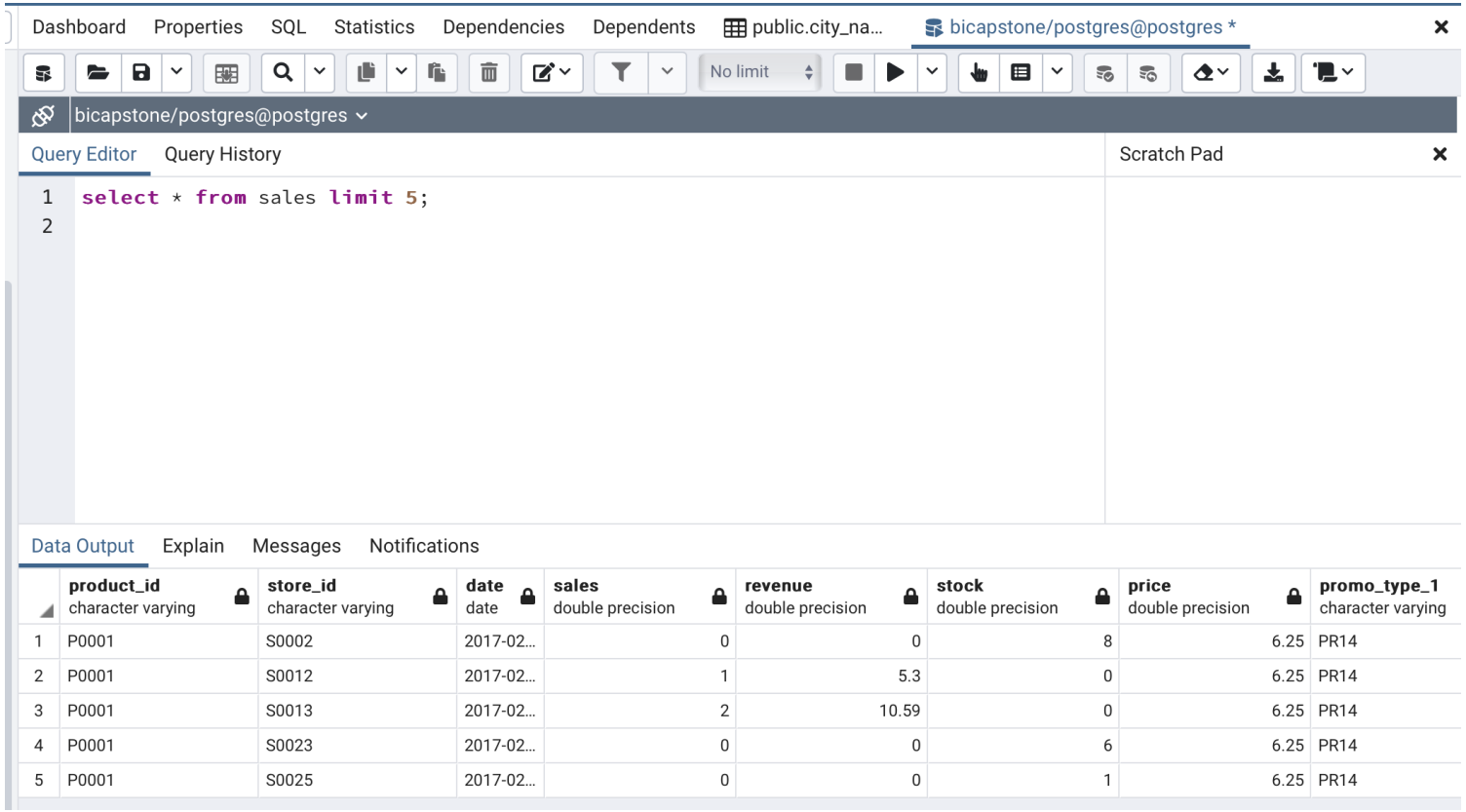
Module 2, Lesson 1: Data Querying Using PostgreSQL

- The pgAdmin interface shows the list of tables in the ‘BICapstone’ database scheme.



Module 2, Lesson 1: Data Querying Using PostgreSQL

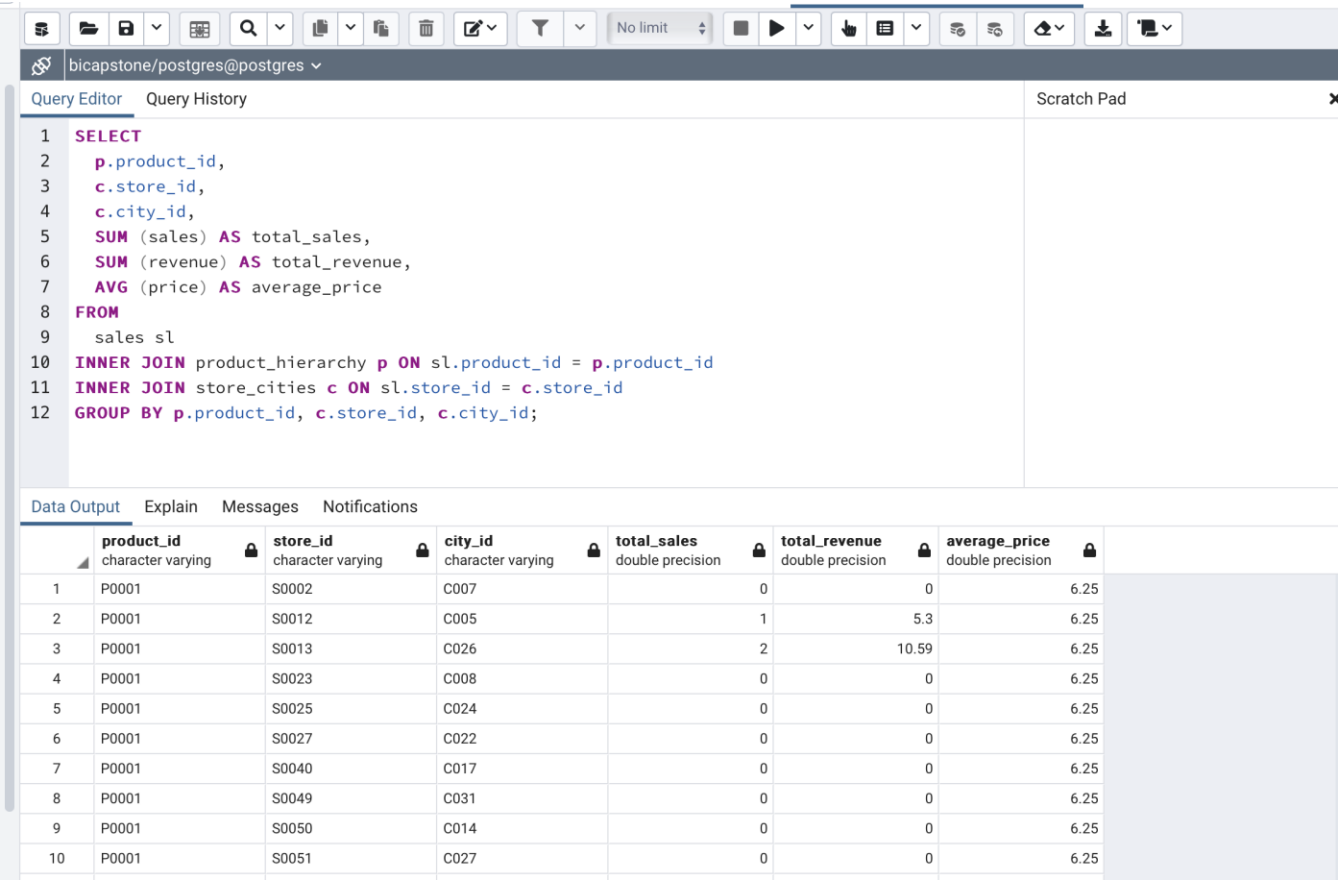
- A written SQL query to check whether data is populated in the table (choose any one table) and the results of running the query.



	product_id character varying	store_id character varying	date date	sales double precision	revenue double precision	stock double precision	price double precision	promo_type_1 character varying
1	P0001	S0002	2017-02...	0	0	8	6.25	PR14
2	P0001	S0012	2017-02...	1	5.3	0	6.25	PR14
3	P0001	S0013	2017-02...	2	10.59	0	6.25	PR14
4	P0001	S0023	2017-02...	0	0	6	6.25	PR14
5	P0001	S0025	2017-02...	0	0	1	6.25	PR14

Module 2, Lesson 1: Data Querying Using PostgreSQL

- A written SQL query that performs the sales performance analysis and the results of running the query.

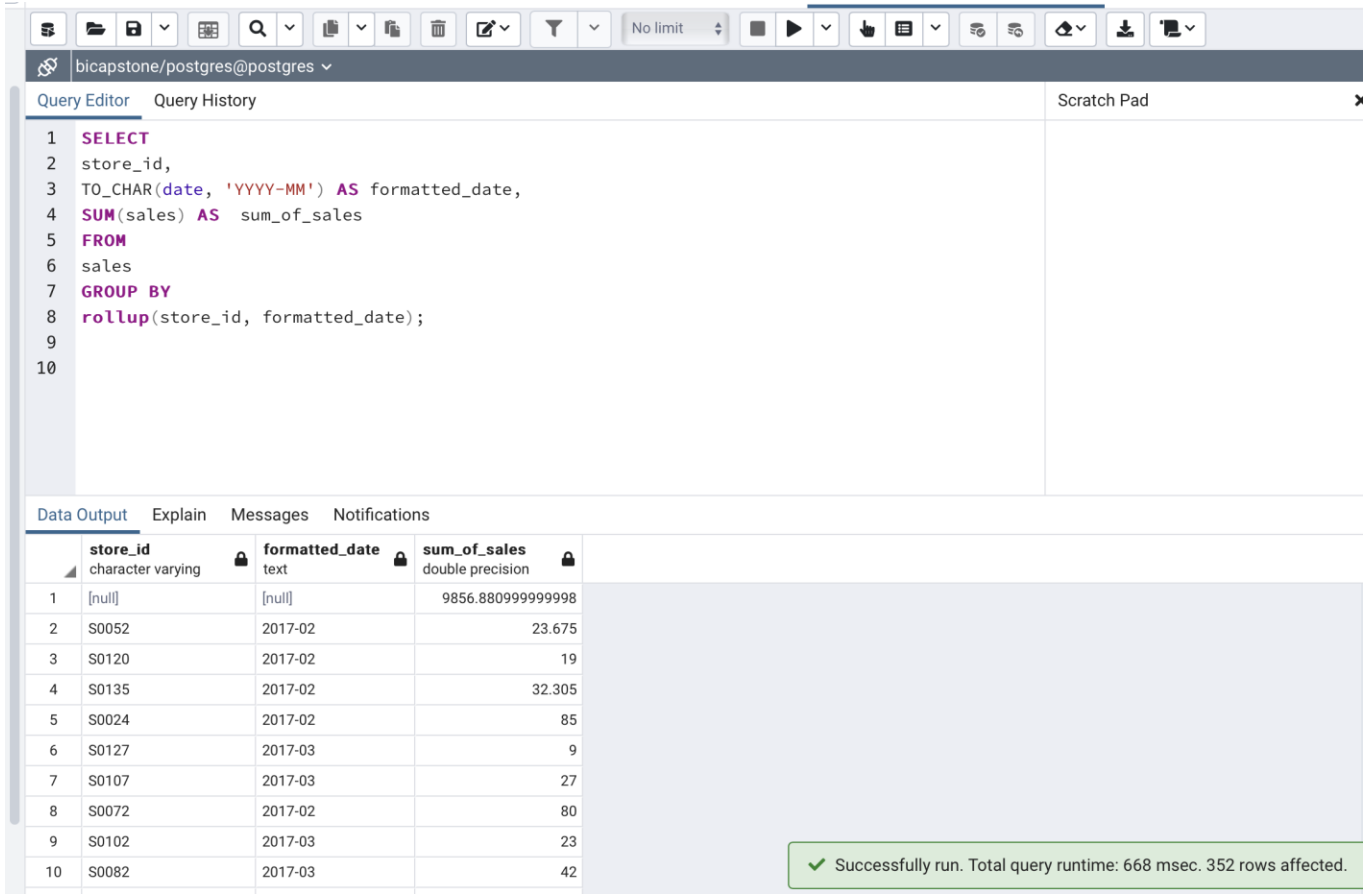


	product_id character varying	store_id character varying	city_id character varying	total_sales double precision	total_revenue double precision	average_price double precision
1	P0001	S0002	C007	0	0	6.25
2	P0001	S0012	C005	1	5.3	6.25
3	P0001	S0013	C026	2	10.59	6.25
4	P0001	S0023	C008	0	0	6.25
5	P0001	S0025	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

Setting up the database help to ensure relevant and useful data is available for subsequent analysis.

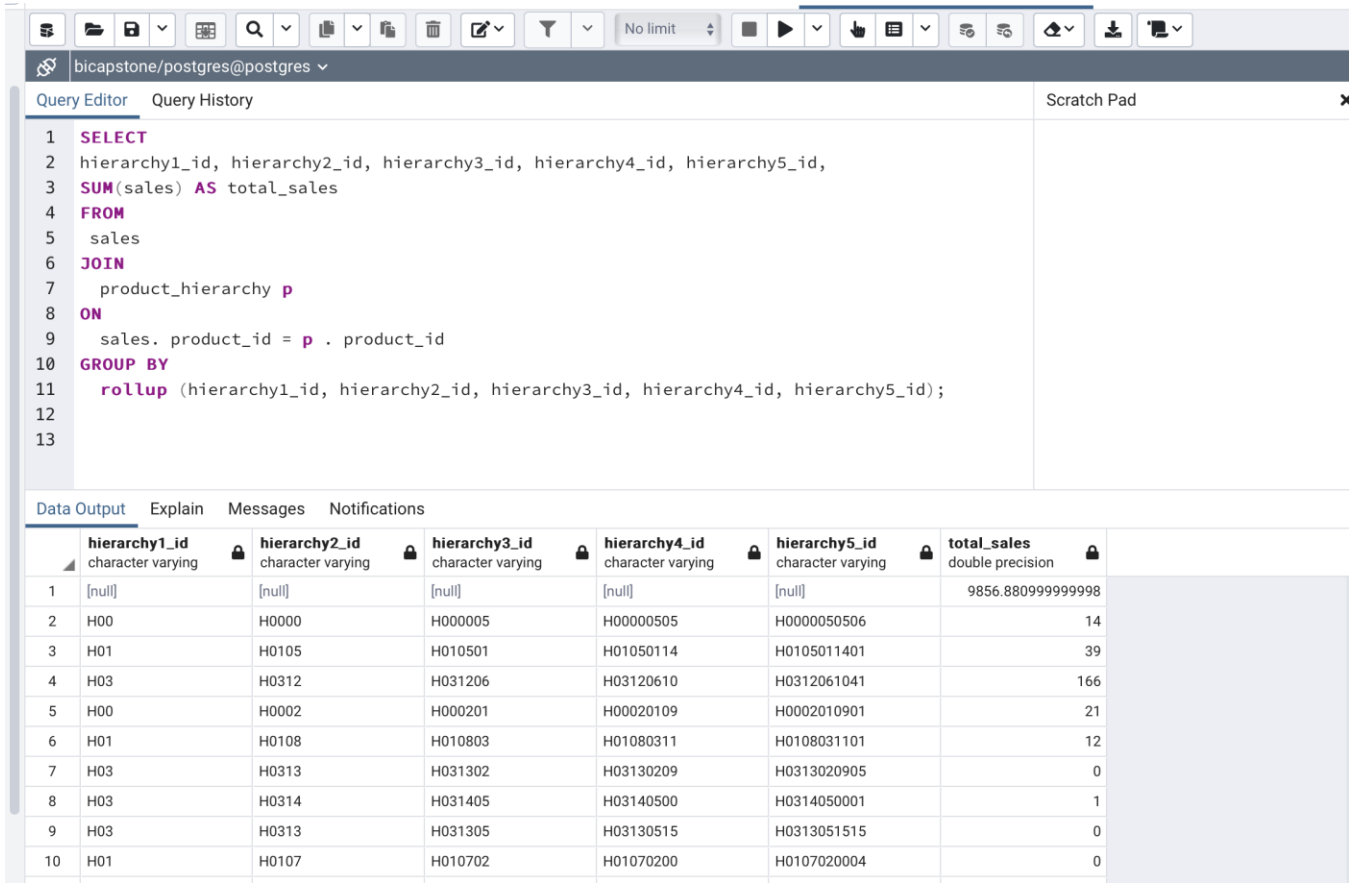
Module 2, Lesson 2: Data Analysis Using PostgreSQL

Creation of data cubes with ROLLUP



Module 2, Lesson 2: Data Analysis Using PostgreSQL

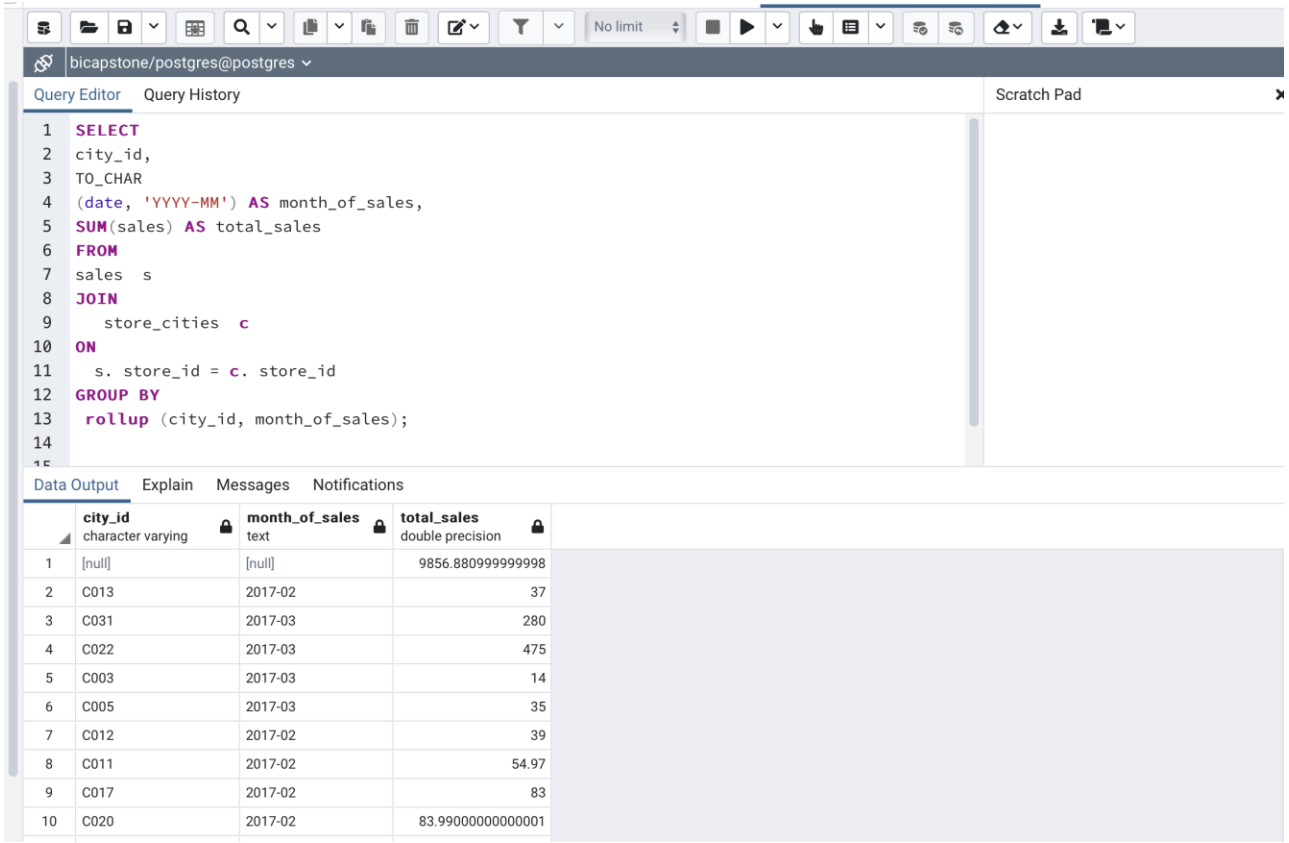
- Summarizing data along hierarchies



	hierarchy1_id character varying	hierarchy2_id character varying	hierarchy3_id character varying	hierarchy4_id character varying	hierarchy5_id character varying	total_sales double precision
1	[null]	[null]	[null]	[null]	[null]	9856.880999999998
2	H00	H0000	H000005	H00000505	H0000050506	14
3	H01	H0105	H010501	H01050114	H0105011401	39
4	H03	H0312	H031206	H03120610	H0312061041	166
5	H00	H0002	H000201	H00020109	H0002010901	21
6	H01	H0108	H010803	H01080311	H0108031101	12
7	H03	H0313	H031302	H03130209	H0313020905	0
8	H03	H0314	H031405	H03140500	H0314050001	1
9	H03	H0313	H031305	H03130515	H0313051515	0
10	H01	H0107	H010702	H01070200	H0107020004	0

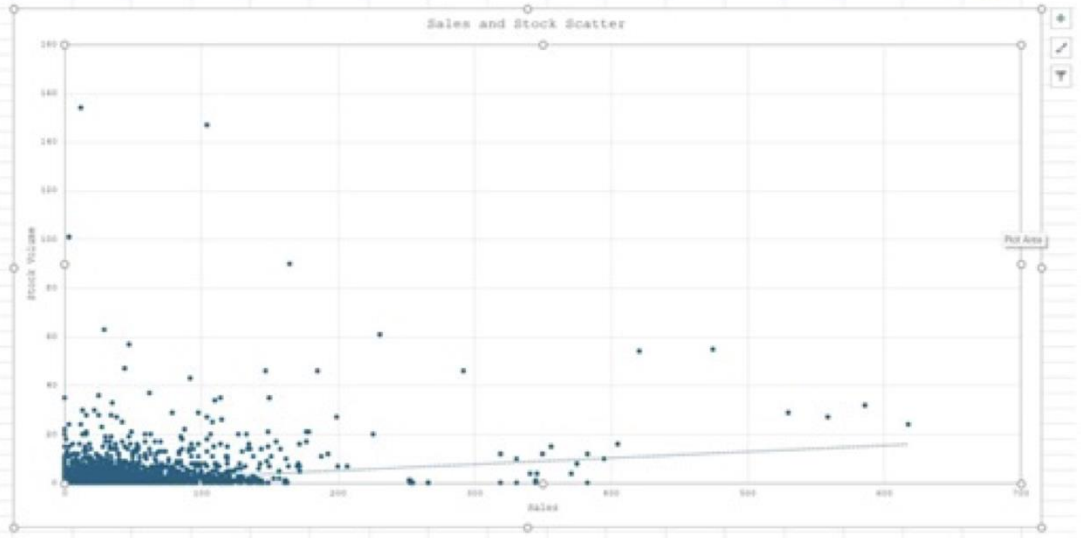
Module 2, Lesson 2: Data Analysis Using PostgreSQL

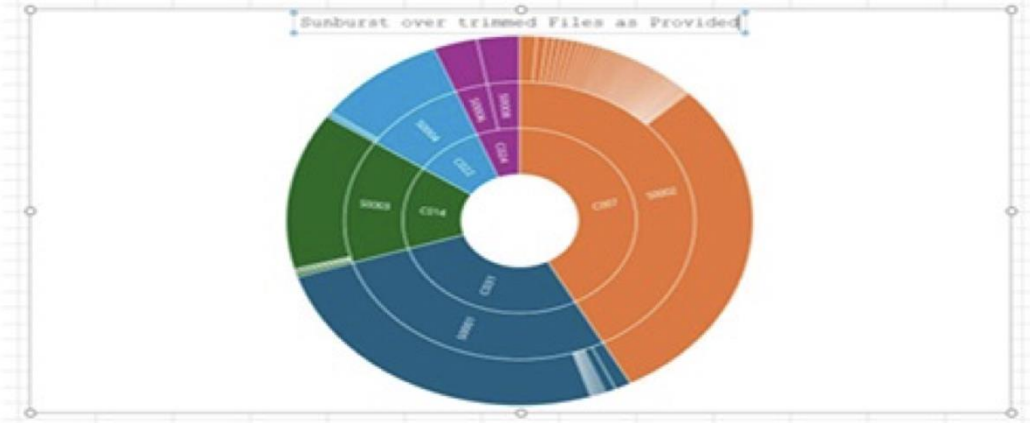
- Identifying sales trends over time and across various regions



These insights provide deeper analysis into data at different levels of granularity which could help with optimization sales strategies and stock management.

Module 3, Lesson 1: Data Visualization Using Excel





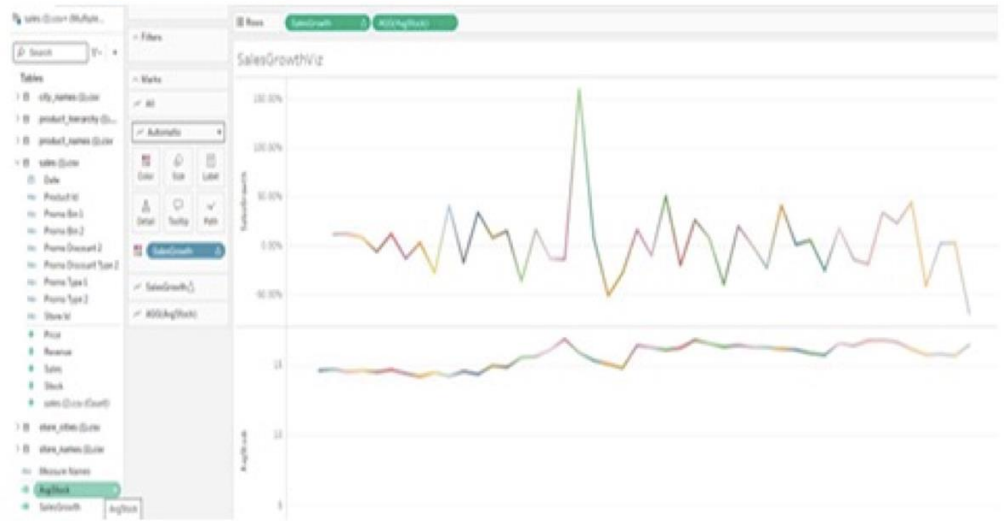
Through the use of the following excel visualizations, we can see that the Digi store clearly contributes to the bulk of the sales generated. It can be noted through the Sunburst chart that the most popular product is C007. It can be noted that the most popular sold item we struggle to regularly keep high inventory of. All quality assessments.

Module 3, Lesson 2: Statistical Analysis

O	P	Q	R	S	T	U	V	W	X
	SUMMARY OUTPUT								
	Regression Statistics								
	Multiple R	0.00254784							
	R Square	6.4915E-06							
	Adjusted R Square	-4.352E-05							
	Standard Error	2.80571308							
	Observations	19999							
	ANOVA								
		df	SS	MS	F	Significance F			
	Regression	1	1.02187736	1.021877363	0.12981123	0.71863165			
	Residual	19997	157416.902	7.872025898					
	Total	19998	157417.924						
		Coefficients	Standard Erro	t Stat	P-value	Lower 95%	Upper 95%	Lower 95%	Upper 95%
	Intercept	-23.188401	65.8376272	-0.352205905	0.724687494	-152.23559	105.858784	-152.23559	105.858784
	date	0.00055476	0.00153976	0.360293254	0.718631733	-0.0024633	0.00357282	-0.0024633	0.00357282

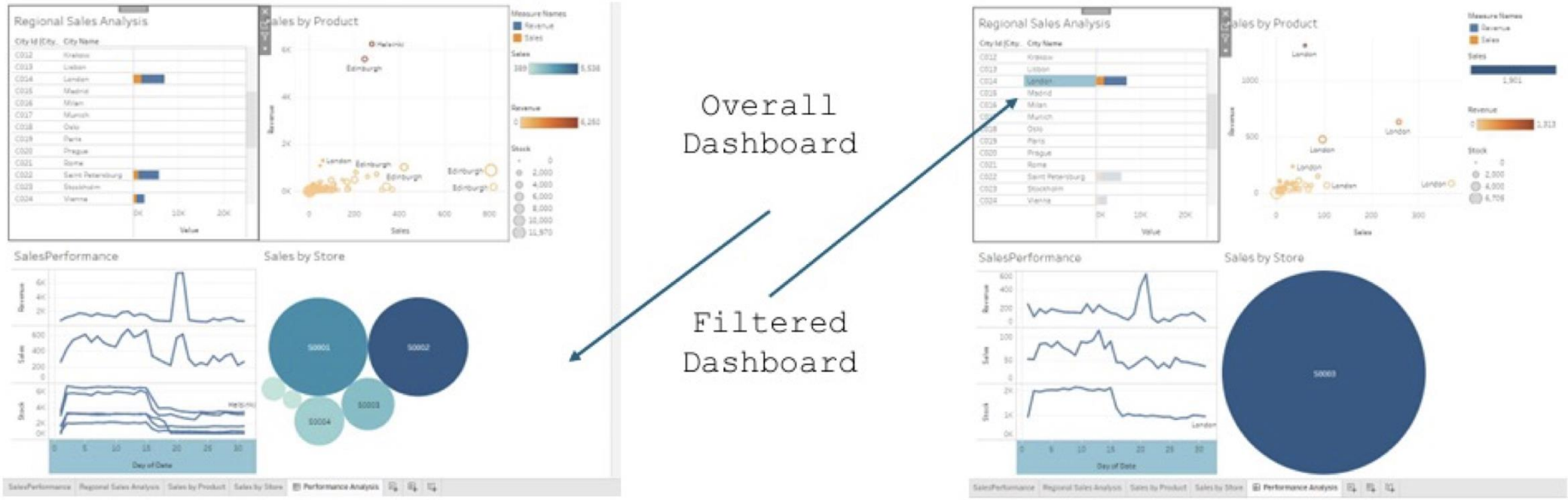
R squared is negligible, while P values is more than 0.05 which mean both factors has almost no impact to the sales level.

Module 4, Lesson 1: Basic Tableau Visualizations



Through the use of Tableau's sheet visualization generating features, clear images can be referred to when attempting to visually identify and isolate developing trends. For example, through the included visualizations herein, we can now determine fairly quickly what product sells best and in what city. We can see what overall sales and sales growth trends are. We can see the relative size of sales impact by city. Insights can be clearly communicated this way.

Module 4, Lesson 2: Advanced Visualizations Using Tableau



Through the use of Tableau's dashboard generating features, seamless interaction between data critical to the business decision-making processes is now available. The right image shows a high-level overview of what readily-available data is ready to refer to. The left image shows a drilled-down version of the same dashboard for one of the attributes selected within the Regional Sales Analysis container.

Discussion

Insights and Recommendations

- Key Products and Stock Levels:
  - By consistently following these analytical techniques, you've identified key products.
  - It's crucial to closely monitor stock levels for these significant revenue contributors.
- Best Revenue-Generating Cities and Products:
  - Your analysis revealed the cities and products driving the most revenue.
  - Use this information strategically to optimize sales and marketing efforts.
- Recommendations:
  - Improve internal data management for smoother analysis processes.
  - Explore connections with Tableau or other tools to enhance data handling.
  - Consider using a Google Sheets filing system for streamlined data scrubbing.
- Business Implications:
  - Understanding what constitutes a high-revenue product is essential.
  - Evaluate resource allocation to maximize success.

Conclusion

Summary

- Critical Insights and Further Analysis:
  - The existing dashboards offer valuable insights, but there's potential for even more critical insights through deeper analysis.
  - Consider exploring additional data dimensions or applying advanced techniques to uncover hidden patterns.
- Customization of Dashboards:
  - The generated dashboards are not set in stone; they can be further customized as needed.
  - If specific requirements arise, such as evaluating product-specific activity, you can adapt the dashboards accordingly.
- Product Strategy Monitoring:
  - Monitoring product strategy implementation is crucial.
  - By categorizing product hierarchy based on the newly implemented strategy, you can effectively track its impact.