(BI) Analyst Capstone Project

Introduction

Business Scenario

As a BI Analyst in a leading retail chain that operates globally, renowned for its diverse product offerings and commitment to customer satisfaction

Tasks

Conduct a comprehensive analysis of the sales performance of the retail chain across different regions and time periods

Methodology

Retail Sales Data

Consisted 3 files, daily sales data, product hierarchy, stores locations

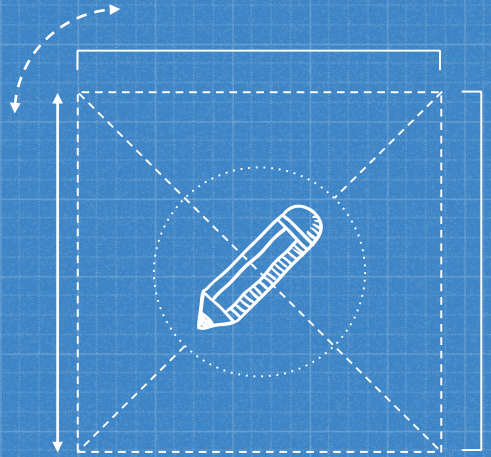
Steps Taken

Data Cleaning and Analysis

Data Querying and Analysis

Chart Creation and Regression Analysis

Data Visualization



RESULTS

Module 1 Lesson 1: Data Cleaning and Preparation

Screenshot of the cleaned and prepared dataset

The cleaned dataset enables analysis much easier and accurate

	product_id	store_id	date	sales	revenue	stock	price	promo_type_1	promo_bin_1	promo_type_2	promo_bin_2	promo_dis
0	P0005	S0001	2017-02-01	0.00	0.00	7.00	33.90	PR14	0	PR03	0.00	
1	P0011	S0001	2017-02-01	0.00	0.00	10.00	49.90	PR14	0	PR03	0.00	
2	P0015	S0001	2017-02-01	1.00	2.41	20.00	2.60	PR14	0	PR03	0.00	
3	P0017	S0001	2017-02-01	0.00	0.00	13.00	1.49	PR14	0	PR03	0.00	
4	P0018	S0001	2017-02-01	0.00	0.00	49.00	1.95	PR14	0	PR03	0.00	
...	...	...	...	...	...	...	...	...	...	...	...	
29994	P0514	S0008	2017-01-18	0.00	0.00	2.00	8.90	PR14	0	PR03	0.00	
29995	P0527	S0008	2017-01-18	0.00	0.00	5.00	1.95	PR14	0	PR03	0.00	
29996	P0536	S0008	2017-01-18	0.00	0.00	20.00	2.95	PR14	0	PR03	0.00	
29997	P0543	S0008	2017-01-18	0.00	0.00	28.00	2.50	PR14	0	PR03	0.00	
29998	P0551	S0008	2017-01-18	0.00	0.00	24.00	1.85	PR14	0	PR03	0.00	

29999 rows × 13 columns

Module 1 Lesson 2: Basic Analysis Using Pivot Tables

Sales Data Analysis

This information can guide decisions about product development, marketing strategies, and pricing.

Sales by City Analysis

It can reveal regional preferences and market penetration, helping businesses tailor their strategies to different locations

store_id	sales
S0001	0.38
S0002	0.68
S0003	0.64
S0004	0.25
S0006	0.25
S0008	0.15

store_id	revenue
S0001	1.86
S0002	2.21
S0003	1.72
S0004	0.90
S0006	0.41
S0008	0.41

Module 2 Lesson 1: Data Querying Using PostgreSQL

A written SQL query to check whether data is populated in the table

```
SELECT COUNT(*) FROM sales_table;
```

A written SQL query that performs the sales performance analysis

```
SELECT store_id, AVG(sales) as average_sales
FROM sales_table
GROUP BY store_id;
```

Module 2 Lesson 2: Data Analysis Using PostgreSQL

Creation of data cubes with ROLLUP

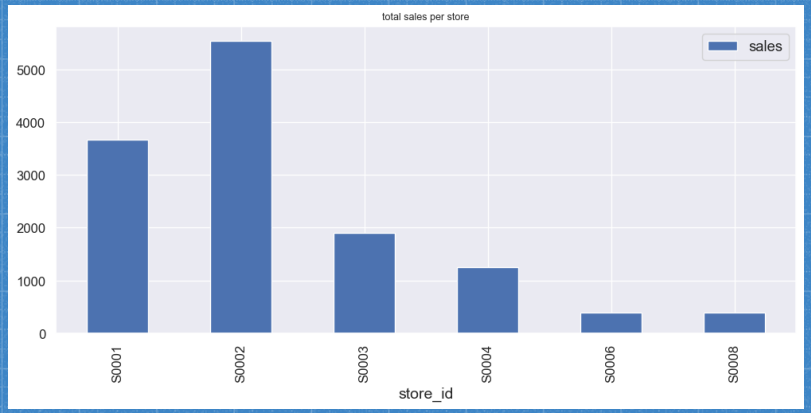
```
SELECT store_id, product_id, SUM(sales) as total_sales
FROM sales_table
GROUP BY ROLLUP(store_id, product_id);
```

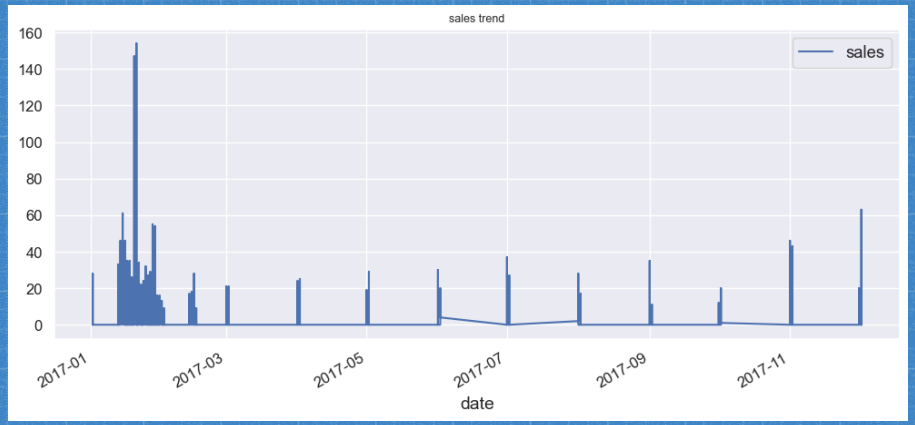
Summarizing data along hierarchies

```
SELECT city, store_name, SUM(sales) as total_sales
FROM sales_table
GROUP BY ROLLUP(city, store_name);
```

Module 3 Lesson 1: Data Visualization Using MS Excel

Line chart can be particularly useful for forecasting future values or understanding the impact of specific events on your data. Bar plot is used to compare different groups or to track changes over time.





Module 3 Lesson 2: Statistical Analysis

R squared at 0.51, it means that 51% of the variability in the outcome data can be explained by the model's inputs.

Independent variables of revenue, stock and price is significant in this prediction

OLS Regression Results

Dep. Variable:	sales	**R-squared:**	0.514
Model:	OLS	**Adj. R-squared:**	0.514
Method:	Least Squares	**F-statistic:**	1.058e+04
Date:	Thu, 02 May 2024	**Prob (F-statistic):**	0.00
Time:	18:35:31	**Log-Likelihood:**	-57647.
No. Observations:	29999	**AIC:**	1.153e+05
Df Residuals:	29995	**BIC:**	1.153e+05
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.1356	0.014	9.393	0.000	0.107	0.164
revenue	0.0464	0.000	165.110	0.000	0.046	0.047
stock	0.0222	0.000	56.128	0.000	0.021	0.023
price	-0.0122	0.001	-14.621	0.000	-0.014	-0.011

Module 4 Lesson 1: Basic Tableau Visualizations

Scatter plots are used to display the relationship between two numerical variables. They can help identify correlations, trends, and outliers in the data

Packed bubbles charts are useful for visualizing hierarchical data and part-to-whole relationships. They can help highlight the relative importance or proportion of different categories.

Module 4 Lesson 2: Advanced Visualizations Using Tableau

In a Tableau dashboard, these visualizations can be combined and interacted with, allowing decision-makers to explore the data from multiple angles and drill down into areas of interest. The ability to filter and drill down into the data makes the insights gained from these visualizations even more powerful.

Discussion

Summary

Thanks!

ANY QUESTIONS?

You can find me at:

@username
user@mail.me