Shakti prasad Nayak

Data Analytics Project
Bhubaneswar, Odisha

Comprehensive Data Analysis on Adventure Works Database Using SQL

September 10, 2024

Project Description

This project involves extracting and analyzing key business data using SQL queries. It covers a range of tasks, including sales analysis, customer segmentation, regional performance evaluation, and financial reporting. By solving various real-world business scenarios, the project demonstrates how SQL can be used to provide insights for sales, marketing, HR, and finance, supporting data-driven decision-making.

Approach

1. **Understanding Requirements:** Carefully analyzed each problem statement to identify the key data elements and conditions needed for query construction.

2. **Data Extraction:** Used SQL queries to retrieve data from tables like `FactInternetSales`, `DimCustomer`, `DimEmployee`, and `FactFinance`, focusing on filtering, grouping, and joining relevant data.

3. **Data Analysis:** Applied functions such as `SUM`, `GROUP BY`, `HAVING`, and `CASE` to aggregate, filter, and categorize data based on business needs (e.g., sales performance, customer segmentation, and expenditure summaries).

4. **Optimization:** Ensured efficient queries by utilizing indexing, appropriate filtering, and limiting data to specific criteria (e.g., region, currency).

5. **Presentation:** Organized the results into clear and concise summaries to enable stakeholders to make informed decisions based on the analysis.

Tech-Stack Used

Azure Data Studio: For writing, executing, and managing SQL queries, and visualizing data from the Adventure Works database.

SQL Server: The backend database management system hosting the Adventure Works dataset, enabling efficient data storage, retrieval, and processing.

Structured Query Language (SQL): For extracting, filtering, aggregating, and analyzing data to generate insights and reports.

Result

1. **Sales Insights:** Identified high-value invoices and detailed invoice line items, revealing top-selling products and sales trends, particularly in the U.S. market.

2. **Customer Analysis:** Segmented homeowner customers based on car ownership, enabling targeted marketing opportunities.

3. **Regional Performance:** Provided detailed summaries of sales activities by sales representatives in Europe, along with a focused sales report for specific territories during peak seasons.

4. **Financial Summary:** Generated expenditure reports and calculated percentage contributions to total costs, supporting financial planning and resource allocation.

5. **Data Organization:** Delivered clear, structured outputs that simplified complex data, aiding stakeholders in making data-driven business decisions.

REPORT

Problem statement 1a

Create a list of product costs, grouped by invoice numbers.

1. Write a query to return InvoiceNumber and TotalProductCost from the FactinternetSales table.

2. Return only invoices that HAVE a total product cost per Invoice Number > 2000.

SQL Query

```
SELECT [SalesOrderNumber] as InvoiceNumber,

SUM(TotalProductCost) as TotalProductCost

FROM FactInternetSales

group by SalesOrderNumber

HAVING SUM(TotalProductCost) > 2000

order by TotalProductCost DESC;
```

Top 10 Results

InvoiceNumber	TotalProductCost
SO43697	2171.294
SO43702	2171.294
SO43703	2171.294
SO43706	2171.294
SO43707	2171.294
SO43709	2171.294
SO43710	2171.294
SO43711	2171.294
SO43712	2171.294

Problem Statement 1b

We need a detailed list of invoices and invoice line numbers, but we're only interested in currency key 100.

1. Write a query to return InvoiceNumber, Invoice LineNumber and SalesAmount from the FactInternetSales table.

2. Return only lines WHERE the currency key is 100.

SQL Query

```
SELECT SalesOrderNumber as InvoiceNumber,

SalesOrderLineNumber as InvoiceLineNumber,

SalesAmount

from FactInternetSales

WHERE CurrencyKey = 100;
```

Top 10 Results

InvoiceNumber	InvoiceLineNumber	SalesAmount
SO43699	1	3399.99
SO43700	1	699.0982
SO43702	1	3578.27
SO43706	1	3578.27
SO43707	1	3578.27
SO43711	1	3578.27
SO43713	1	3578.27
SO43718	1	3578.27
SO43719	1	3578.27

Problem Statement 1c

We have a new data analyst in the team who wants to see a unique list of sales territory keys. This will help her to better understand the database.

1. Write a query to return the sales territory column from the FactInternetSales table.

2. Return a unique list of territories only.

3. Order the results alphabetically for ease.

SQL Query

```
SELECT distinct [SalesTerritoryKey]

FROM FactInternetSales

ORDER BY SalesTerritoryKey;
```

Top 10 Results

SalesTerritoryKey
1
2
3
4
5

6

7

8

9

10

Problem Statement 2a

Sales territory 1 need a summary of their sales for the lead up period to Christmas.

1. Write a query against the FactInternet Sales table that returns orders placed in December for the Sales Territory 1

The query should include SalesOrderNumber,SalesOrderLine Number, SalesAmount and TaxAmount.

SQL Query

```
SELECT salesordernumber,

[SalesOrderLineNumber],

[SalesAmount],

[TaxAmt]

from FactInternetSales

WHERE DATENAME(MONTH,orderdate) = N'December' AND SalesTerritoryKey
=1
```

Top 10 Results

salesordernumber	SalesOrderLineNumber	SalesAmount	TaxAmt
SO43699	1	3399.99	271.9992
SO46406	1	3578.27	286.2616
SO46431	1	3578.27	286.2616
SO46445	1	3578.27	286.2616
SO46446	1	3578.27	286.2616
SO46452	1	3374.99	269.9992
SO46466	1	3578.27	286.2616
SO46467	1	3578.27	286.2616
SO46470	1	699.0982	55.9279
SO46476	1	3374.99	269.9992

Problem Statement 2b

Marketing need a list of homeowner customers, along with the number of cars owned.

1. Write a query against the dimCustomer table that returns all customers that are homeowners and have more than 1 car.

2. The query should include full customer names, number of cars owned, and email.

3. The numbers of cars owned should categorize customers into groups: -2-3 4+

SQL Query

```
SELECT
CONCAT(FirstName, ' ', LastName) as CustomerName,
case
    when NumberCarsOwned in (2,3) THEN '2-3'
    when NumberCarsOwned >= 4 then '4+'
End as Numbercarsowned,
EmailAddress
from DimCustomer
WHERE HouseOwnerFlag =1 AND NumberCarsOwned > 1
```

Top 10 Results

CustomerName	Numbercarsowned	EmailAddress
Elizabeth Johnson	4+	elizabeth5@adventure-works.com
Marco Mehta	2-3	marco14@adventure-works.com
Rob Verhoff	2-3	rob4@adventure-works.com
Curtis Lu	4+	curtis9@adventure-works.com
Lauren Walker	2-3	lauren41@adventure-works.com
Ian Jenkins	2-3	ian47@adventure-works.com
Shannon Wang	2-3	shannon1@adventure-works.com
Clarence Rai	2-3	clarence32@adventure-works.com
Russell Xie	2-3	russell7@adventure-works.com
Alejandro Beck	2-3	alejandro45@adventure-works.com

Problem Statement 3a

Summarize the Internet Sales by Subcategory and return the top 5 subcategories.

1. Write a query that returns the top 5 best-selling subcategories by SalesAmount.

2. We're only interested in sales from our website (internet sales).

3. Finally, the data should only include sales where the country is United States and the currency is US Dollar.

4. You are avoid using the view we created.

SQL Query

```sql
SELECT TOP(5)
    dpsc.EnglishProductSubcategoryName AS SubCategory,
    SUM(fs.SalesAmount) AS SalesAmount
FROM FactInternetSales AS fs
    JOIN DimProduct AS dp
    ON fs.ProductKey=dp.ProductKey
    JOIN DimProductSubcategory AS dpsc
    ON dp.ProductSubcategoryKey=dpsc.ProductSubcategoryKey
    JOIN DimCurrency AS dcy
    ON fs.CurrencyKey=dcy.CurrencyKey
    JOIN DimSalesTerritory AS dst
    ON fs.SalesTerritoryKey=dst.SalesTerritoryKey
```

```
WHERE dst.SalesTerritoryCountry =N'United States'

AND dcy.CurrencyName=N'US Dollar'

GROUP BY dpsc.EnglishProductSubcategoryName

ORDER BY SalesAmount DESC
```

ALL Results

SubCategory	SalesAmount
Road Bikes	4289926
Mountain Bikes	3417458
Touring Bikes	1292476
Tires and Tubes	88762.86
Helmets	76663.09

Problem Statement 3b

It's performance review time. HR Europe need to see sales by sales representative, and by currency.

1. Write a query that will return a list of all current Sales Representatives or Sales Managers in the European territory.

2. For each person, HR need to see sales amounts grouped by currency.

3. Please include the following fields: Full employee name, Employee Title, Currency Name and total sales amount.

4. The query should be sorted by Employee Name and Sales Amount.

SQL Query

```sql
SELECT
    CONCAT(de.FirstName, ' ', de.LastName) AS EmployeeName,
    de.Title AS EmployeeTitle,
    dcy.CurrencyName AS Currency,
    SUM(fs.SalesAmount) AS TotalSalesAmount
FROM FactResellerSales AS fs
    JOIN DimEmployee AS de
    ON fs.EmployeeKey=de.EmployeeKey
    JOIN DimSalesTerritory AS dst
    ON fs.SalesTerritoryKey=dst.SalesTerritoryKey
    JOIN DimCurrency AS dcy
    ON fs.CurrencyKey=dcy.CurrencyKey
WHERE dst.SalesTerritoryGroup =N'Europe'
AND de.[Status] =N'Current'
GROUP BY de.FirstName, de.LastName, de.Title, dcy.CurrencyName
ORDER BY EmployeeName, TotalSalesAmount DESC
```

ALL Results

EmployeeName	EmployeeTitle	Currency	TotalSalesAmount
Amy Alberts	European Sales Manager	United Kingdom Pound	441081.6
Amy Alberts	European Sales Manager	EURO	200960.6
Amy Alberts	European Sales Manager	US Dollar	90036.24
José Saraiva	Sales Representative	United Kingdom Pound	3837927
Rachel Valdez	Sales Representative	EURO	1790640
Ranjit Varkey Chudukatil	Sales Representative	US Dollar	4026954
Ranjit Varkey Chudukatil	Sales Representative	EURO	482934.9

Problem Statement 4a

Create a summary of expenditure accounts.

1. Write a query that will return the sum of actuals from the FactFinance table.

2. Filter the data to meet the following conditions:

- January, 2011 only

- Southwest Division only expenditure accounts only

3. For each row, list the Organization, Account Type and Account Name.

4. Group the rows by organization and account.

SQL Query

```sql
SELECT
o.OrganizationName,
a.AccountType,
a.AccountDescription,
SUM(ff.amount) as Amount
FROM FactFinance as ff
INNER JOIN DimScenario as s
on ff.ScenarioKey = s.ScenarioKey
INNER JOIN DimDate as d
ON d.DateKey = ff.DateKey
INNER JOIN DimOrganization as o
on o.OrganizationKey = ff.OrganizationKey
INNER JOIN DimAccount as a
on a.AccountKey = ff.AccountKey
WHERE s.ScenarioName = N'Actual'
AND d.CalendarYear = 2011
AND d.EnglishMonthName = N'January'
AND o.OrganizationName = N'Southwest Division'

AND a.AccountType = N'Expenditures'

GROUP BY o.OrganizationName, a.AccountType, a.AccountDescription

ORDER BY Amount DESC
```

Top 10 Results

OrganizationName	AccountType	AccountDescription	Amount
Southwest Division	Expenditures	Standard Cost of Sales	122573
Southwest Division	Expenditures	Salaries	67520
Southwest Division	Expenditures	Taxes	36583
Southwest Division	Expenditures	Commissions	15640
Southwest Division	Expenditures	Variances	13034
Southwest Division	Expenditures	Payroll Taxes	6469
Southwest Division	Expenditures	Employee Benefits	5291
Southwest Division	Expenditures	Telephone	4162
Southwest Division	Expenditures	Returns and Adjustments	3910

Southwest Division	Expenditures	Building Leasehold	1593

Problem Statement 4b

Create a summary of expenditure account totals, and then calculate a Pct of Total.

1. First, write a query that will return Account Description, and amounts corresponding to actuals.

2. Filter the results to meet the following conditions:

- Actuals only

- Canadian division only

- Calendar year 2013 only

- Expenditure accounts only

3. Create a subquery to calculate the total sales that meet the same conditions.

SQL Query

```sql
SELECT

a.AccountDescription,

SUM(f.Amount) AS amount,

SUM(f.Amount) / (select SUM(f.Amount) AS amount

From FactFinance as f

INNER JOIN DimAccount as a

on f.AccountKey = a.AccountKey

INNER JOIN DimScenario as s
ON s.ScenarioKey = f.ScenarioKey

INNER JOIN DimOrganization as o

ON o.OrganizationKey = f.OrganizationKey

INNER JOIN DimDate as d

ON d.DateKey = f.DateKey

WHERE s.ScenarioName = N'Actual'

AND o.OrganizationName = N'Canadian Division'

AND d.CalendarYear = 2013

AND a.AccountType = N'Expenditures') AS [Percent]

From FactFinance as f

INNER JOIN DimAccount as a

on f.AccountKey = a.AccountKey

INNER JOIN DimScenario as s

ON s.ScenarioKey = f.ScenarioKey

INNER JOIN DimOrganization as o

ON o.OrganizationKey = f.OrganizationKey

INNER JOIN DimDate as d

ON d.DateKey = f.DateKey

WHERE s.ScenarioName = N'Actual'

AND o.OrganizationName = N'Canadian Division'

AND d.CalendarYear = 2013

AND a.AccountType = N'Expenditures'
```

```
GROUP BY a.AccountDescription

ORDER BY amount DESC
```

Top 10 Results

AccountDescription	amount	Percent
Standard Cost of Sales	2672904	0.357217
Salaries	2163556	0.289145
Taxes	665875.4	0.08899
Variances	441498.9	0.059004
Commissions	320161.8	0.042788
Returns and Adjustments	260098.6	0.034761
Payroll Taxes	223394.4	0.029855
Employee Benefits	164690.3	0.02201
Discounts	101335.4	0.013543
Telephone	99489.04	0.013296

END