Muhammadjon Merzaqulov

MY PORTFOLIO PROJECT IN SQL
EXPLORATORY DATA ANALYSIS

INTRODUCTION

The database used for this analysis contains transactions related to car sales, their specifications, and dealership details. I will use only three tables in this report to analyze the research data:

- **FactCarSales**
- **DimCar**
- **DimLocation**

- **Sales Analysis** (focusing on sales trends and profitability).
- **Location Analysis** (Focused on geographic places)**.**
- **Product Analysis** (focusing on car models and specifications).

Objective

Tables are used to learn about data and answer different questions using different SQL queries to help businesses make more informed decisions.

Overview

Azure Data Studio was used to query this SQL project. The areas of SQL covered in this case study:

- **Basic aggregations**
- **CASE WHEN statements**
- **Window Functions**
- **Joins**
- **Date time functions**
- **CTSs**
- **Subquery**

Content

SALES ANALYSIS

1. What is the total revenue of the company?

```
SELECT
       FORMAT(SUM(Sales_Price),'C2') AS Total_Revenue,
       FORMAT(SUM(Cost_Price),'C2') AS Total_Cost,
       FORMAT(SUM((Sales_Price - Cost_Price) * Production_Quantity),'C2')
                                                          AS Total_Profit
FROM Fact
```

Results Messages

	Total_Revenue	Total_Cost	Total_Profit
1	$223,095,233.00	$170,700,854.00	$25,853,277,734.00

Important Note

I always use a special query to see the big picture of the entire data period before I dive in. By doing the analysis and basing it on this query, I can focus more on the details and have more confidence in the outcome of my analysis.

```
SELECT
   Year([Date]) AS Years,
   Min(Month([Date])) AS StartOfMonth,
   Max(Month([Date])) AS EndOfMonth,
   Count(DISTINCT Month([Date])) AS TotalNumOfMonths,
   Concat_ws('-', Datename(M, Min([Date])), Datename(DAY, Min([Date]))) AS
StrMonthAndDay,
   Concat_ws('-', Datename(M, Max([Date])), Datename(DAY, Max([Date]))) AS
EndMonthAndDay,
   Count(DISTINCT Datepart(DAYOFYEAR, [Date])) AS DayOfyear
FROM Fact
GROUP BY Year([Date])
ORDER BY 1, 2, 3, 4
```

Results Messages

	Years	StartOfMonth	EndOfMonth	TotalNumOfMonths	StrMonthAndDay	EndMonthAndDay	DayOfyear
1	2010	1	12	12	January-1	December-31	365
2	2011	1	12	12	January-1	December-31	365
3	2012	1	12	12	January-1	December-31	366
4	2013	1	12	12	January-1	December-31	365
5	2014	1	12	12	January-1	December-31	365
6	2015	1	12	12	January-1	December-31	365
7	2016	1	12	12	January-1	December-31	366
8	2017	1	12	12	January-1	December-31	365
9	2018	1	12	12	January-1	December-31	365
10	2019	1	12	12	January-1	December-31	365
11	2020	1	12	12	January-1	December-31	366
12	2021	1	12	12	January-1	December-31	365
13	2022	1	12	12	January-1	December-31	365

2. How many years has the company recorded sales?

```
SELECT DISTINCT YEAR([Date]) AS Years
FROM Fact
ORDER BY 1
```

Results Messages

	Years
1	2010
2	2011
3	2012
4	2013
5	2014
6	2015
7	2016
8	2017
9	2018
10	2019
11	2020
12	2021
13	2022

3. How do sales compare across weekdays and weekends for each year?

```
SELECT
   YEAR([Date]) AS Years,
   CASE
       WHEN DATEPART(WEEKDAY, [Date]) IN (1, 7) THEN 'Weekend'
       ELSE 'Weekday'
   END AS Day_Type,
   SUM(Production_Quantity) AS Total_Quantity
over
FROM Fact
GROUP BY YEAR([Date]),
   CASE
       WHEN DATEPART(WEEKDAY, [Date]) IN (1, 7) THEN 'Weekend'
       ELSE 'Weekday'
   END
ORDER BY 1,2
```

Results Messages

	Years	Day_Type	Total_Quantity
1	2010	Weekday	133068
2	2010	Weekend	52926
3	2011	Weekday	131485
4	2011	Weekend	52637
5	2012	Weekday	128876
6	2012	Weekend	55657
7	2013	Weekday	129141
8	2013	Weekend	49278
9	2014	Weekday	137777
10	2014	Weekend	50426
11	2015	Weekday	127104
12	2015	Weekend	47108
13	2016	Weekday	135716
14	2016	Weekend	49150
15	2017	Weekday	127445
16	2017	Weekend	50132
17	2018	Weekday	134736
18	2018	Weekend	53258
19	2019	Weekday	123589
20	2019	Weekend	51369
21	2020	Weekday	131300
22	2020	Weekend	57666
23	2021	Weekday	129847
24	2021	Weekend	50119
25	2022	Weekday	132720
26	2022	Weekend	53135

4. What is the profit of the company by year?

```
SELECT
   YEAR([Date]) AS Years,
   FORMAT(SUM((Sales_Price - Cost_Price) * Production_Quantity), 'C2') AS Profit
FROM Fact
GROUP BY YEAR([Date])
ORDER BY 2
```

Results Messages

	Years	Profit
1	2010	$1,451,325,052.00
2	2012	$1,778,366,932.00
3	2019	$1,786,194,775.00
4	2022	$1,892,014,631.00
5	2017	$1,937,238,862.00
6	2018	$1,994,910,800.00
7	2016	$2,073,215,182.00
8	2021	$2,080,398,424.00
9	2013	$2,120,982,945.00
10	2020	$2,126,859,665.00
11	2011	$2,151,946,528.00
12	2015	$2,203,329,560.00
13	2014	$2,256,494,378.00

5. What was the best month to sell for each year?

```
WITH best_month AS (
   SELECT
       YEAR([Date]) AS Years,
       DATENAME(MONTH, [Date]) AS MonthName,
       FORMAT(SUM((Sales_Price - Cost_Price) * Production_Quantity), 'C2') AS Profit,
       DENSE_RANK() OVER(
           PARTITION BY YEAR([Date])
           ORDER BY SUM((Sales_Price - Cost_Price) * Production_Quantity) DESC
       ) AS Rank
   FROM Fact
   GROUP BY YEAR([Date]), DATENAME(MONTH, [Date]))
SELECT Years, MonthName, Profit
FROM best_month
WHERE Rank = 1
ORDER BY 1;
```

	Years	MonthName	Profit
1	2010	July	$361,109,198.00
2	2011	March	$301,629,965.00
3	2012	June	$244,653,576.00
4	2013	March	$426,619,476.00
5	2014	October	$309,860,370.00
6	2015	May	$265,430,400.00
7	2016	April	$242,272,123.00
8	2017	May	$317,293,687.00
9	2018	October	$291,127,660.00
10	2019	April	$277,204,579.00
11	2020	November	$303,666,561.00
12	2021	October	$265,558,980.00
13	2022	November	$269,612,841.00

6. What was the best weekdays to sell for each year?

```
WITH best_weekday as (
   SELECT
       YEAR([Date]) AS Years,
       DATENAME(WEEKDAY, [Date]) AS WeekName,
       FORMAT(SUM((Sales_Price - Cost_Price) * Production_Quantity), 'C2') AS
Profit,
       DENSE_RANK() OVER(
           PARTITION BY YEAR([Date])
           ORDER BY SUM((Sales_Price - Cost_Price) * Production_Quantity) DESC
       ) AS Rank
   FROM Fact
   GROUP BY YEAR([Date]), DATENAME(WEEKDAY, [Date]))
SELECT Years, WeekName, Profit
FROM best_weekday
WHERE Rank = 1
ORDER BY 1
```

Results Messages

	Years	WeekName	Profit
1	2010	Tuesday	$519,991,043.00
2	2011	Saturday	$347,513,140.00
3	2012	Monday	$412,459,391.00
4	2013	Wednesday	$398,770,741.00
5	2014	Friday	$405,195,979.00
6	2015	Sunday	$374,260,306.00
7	2016	Thursday	$408,684,151.00
8	2017	Tuesday	$509,999,337.00
9	2018	Sunday	$407,359,234.00
10	2019	Sunday	$413,206,124.00
11	2020	Monday	$398,865,630.00
12	2021	Wednesday	$386,862,594.00
13	2022	Thursday	$386,337,625.00

7. What is the Year-over-Year (YoY) revenue change between the current year and the previous year?

```
SELECT
   FORMAT(Profit_CY,'C2') AS TotalProfit_CY,
   FORMAT(Profit_CY,'C2') AS TotalProfit_LY,
   FORMAT(Profit_CY - Profit_LY, 'C2') AS Dif_Profit,
   FORMAT(Profit_CY / Profit_LY - 1, 'P') AS Dif_Profit_P
FROM
(
   SELECT SUM((Sales_Price - Cost_Price) * Production_Quantity) AS Profit_CY
   FROM Fact
   WHERE YEAR([Date]) = (SELECT MAX(YEAR([Date])) FROM Fact)
)CurrentYear,
(
   SELECT SUM((Sales_Price - Cost_Price) * Production_Quantity) AS Profit_LY
   FROM Fact
   WHERE YEAR([Date]) = (SELECT MAX(YEAR([Date])) FROM Fact)-1
)LastYear
```

(another option) → both return the same response

```
WITH CurrentYear AS (
   SELECT SUM((Sales_Price - Cost_Price) * Production_Quantity) AS Profit_CY
   FROM Fact
   WHERE YEAR([Date]) = (SELECT MAX(YEAR([Date])) FROM Fact)
), LastYear AS (
   SELECT SUM((Sales_Price - Cost_Price) * Production_Quantity) AS Profit_LY
   FROM Fact
   WHERE YEAR([Date]) = (SELECT MAX(YEAR([Date])) FROM Fact)-1
)

SELECT
   FORMAT(Profit_CY,'C2') AS TotalProfit_CY,
   FORMAT(Profit_LY,'C2') AS TotalProfit_LY,
   FORMAT(Profit_CY - Profit_LY, 'C2') AS Dif_Profit,
   FORMAT(Profit_CY / Profit_LY - 1, 'P') AS Dif_Profit_P
FROM CurrentYear, LastYear
```

Results Messages

	TotalProfit_CY	TotalProfit_LY	Dif_Profit	Dif_Profit_P
1	$1,892,014,631.00	$2,080,398,424.00	($188,383,793.00)	-9.06%

8. What is the Month-over-Month (MoM) revenue change between the current month and the previous month?

```
SELECT
   FORMAT(TotalProfit_CM,'C2') AS TotalProfit_CM,
   FORMAT(TotalProfit_LM,'C2') AS TotalProfit_LM,
   FORMAT(TotalProfit_CM - TotalProfit_LM, 'C2') AS Dif_Profit,
   FORMAT(TotalProfit_CM / TotalProfit_LM - 1, 'P') AS Dif_Profit_P
FROM
(
   SELECT SUM((Sales_Price - Cost_Price) * Production_Quantity) AS TotalProfit_CM
   FROM Fact
   WHERE YEAR([Date]) = (SELECT MAX(YEAR([Date])) FROM Fact)
       AND MONTH([Date]) = (SELECT MAX(MONTH([Date])) FROM Fact)
)Current_Month,
(
   SELECT SUM((Sales_Price - Cost_Price) * Production_Quantity) AS TotalProfit_LM
   FROM Fact
   WHERE YEAR([Date]) = (SELECT MAX(YEAR([Date])) FROM Fact)
       AND MONTH([Date]) = (SELECT MAX(MONTH([Date])) FROM Fact) - 1
)Last_Month
```

(another option) → both return the same response

```
WITH Current_Month AS (
   SELECT SUM((Sales_Price - Cost_Price) * Production_Quantity) AS TotalProfit_CM
   FROM Fact
   WHERE YEAR([Date]) = (SELECT MAX(YEAR([Date])) FROM Fact)
       AND MONTH([Date]) = (SELECT MAX(MONTH([Date])) FROM Fact)
), Last_Month AS (
   SELECT SUM((Sales_Price - Cost_Price) * Production_Quantity) AS TotalProfit_LM
   FROM Fact
   WHERE YEAR([Date]) = (SELECT MAX(YEAR([Date])) FROM Fact)
       AND MONTH([Date]) = (SELECT MAX(MONTH([Date])) FROM Fact) - 1
)
SELECT
   FORMAT(TotalProfit_CM,'C2') AS TotalProfit_CM,
   FORMAT(TotalProfit_LM,'C2') AS TotalProfit_LM,
   FORMAT(TotalProfit_CM - TotalProfit_LM, 'C2') AS Dif_Profit,
   FORMAT(TotalProfit_CM / TotalProfit_LM - 1, 'P') AS Dif_Profit_P
FROM Current_Month, Last_Month;
```

Results Messages

	TotalProfit_CM	TotalProfit_LM	Dif_Profit	Dif_Profit_P
1	$119,073,067.00	$269,612,841.00	($150,539,774.00)	-55.84%

9. Sales by territories ranked by highest revenue.

```
SELECT DL.Dealership_Location, FORMAT(SUM((Sales_Price - Cost_Price) *
Production_Quantity), 'C2') AS Profit
FROM Fact F
JOIN DimLocation DL
      ON f.DealershipLocationID = DL.DealershipLocationID
GROUP BY DL.Dealership_Location
ORDER BY 2 DESC
```

Results Messages

	Dealership_Location	Profit
1	Suburban	$9,190,438,699.00
2	Urban	$8,441,308,104.00
3	Rural	$8,221,530,931.00

LOCATION ANALYSIS

1.How many manufacturing locations exist?

```
SELECT DISTINCT Manufacturing_Location
FROM DimLocation
```

Results Messages

	Manufacturing_Location
1	Germany
2	Japan
3	South Korea
4	USA

2. How many sales originate from all manufacturing locations?

```
SELECT DL.Manufacturing_Location, count(*) AS sales_quantity
FROM Fact F
LEFT JOIN DimLocation DL
   ON f.DealershipLocationID = DL.DealershipLocationID
GROUP BY DL.Manufacturing_Location
ORDER BY 2 DESC
```

Results Messages

	Manufacturing_Location	sales_quantity
1	USA	1246
2	South Korea	1200
3	Germany	1185
4	Japan	1117

3. Which manufacturing location generates the maximum profit?

```
SELECT TOP(1)
   DL.Manufacturing_Location,
   FORMAT(SUM((F.Sales_Price - F.Cost_Price) * F.Production_Quantity), 'C2') AS
Profit
FROM DimLocation DL
LEFT JOIN Fact F
   ON DL.DealershipLocationID = F.DealershipLocationID
GROUP BY DL.Manufacturing_Location
ORDER BY 2
```

Results Messages

	Manufacturing_Location	Profit
1	Japan	$5,762,168,035.00

4. How many unique countries of origin are there?

```
SELECT DISTINCT Country_of_Origin
FROM DimLocation
```

Results Messages

	Country_of_Origin
1	Germany
2	Italy
3	Japan
4	South Korea
5	USA

5. How many sales are attributed to each country of origin?

```
SELECT dl.Country_of_Origin, count(F.Color) AS sales_quantity
FROM DimLocation DL
LEFT JOIN Fact F
   ON DL.DealershipLocationID = F.DealershipLocationID
GROUP BY dl.Country_of_Origin
ORDER BY 2
```

Results Messages

	Country_of_Origin	sales_quantity
1	South Korea	909
2	Germany	936
3	USA	955
4	Italy	969
5	Japan	979

6. What is sales by each country of origin?

```
SELECT
   DL.Country_of_Origin,
   FORMAT(SUM(Sales_Price),'C2') AS Total_Revenue,
   FORMAT(SUM(Cost_Price),'C2') AS Total_Cost,
   FORMAT(SUM((Sales_Price - Cost_Price) * Production_Quantity),'C2') AS
Total_Profit
FROM DimLocation DL
LEFT JOIN Fact F
   ON DL.DealershipLocationID = F.DealershipLocationID
GROUP BY Country_of_Origin
ORDER BY 4 DESC, 2, 3
```

Results Messages

	Country_of_Origin	Total_Revenue	Total_Cost	Total_Profit
1	USA	$45,358,118.00	$34,198,513.00	$5,462,237,172.00
2	Italy	$45,818,619.00	$34,761,572.00	$5,432,349,651.00
3	Japan	$45,816,803.00	$35,161,238.00	$5,046,195,317.00
4	South Korea	$42,388,852.00	$32,761,767.00	$5,046,057,352.00
5	Germany	$43,712,841.00	$33,817,764.00	$4,866,438,242.00

7. Number of unique Dealership Location:

```
SELECT DISTINCT Dealership_Location
FROM DimLocation
```

Results Messages

	Dealership_Location
1	Rural
2	Suburban
3	Urban

8. Sales by Dealership Location:

```
SELECT DL.Dealership_Location, count(F.Color) AS sales_quantity
FROM DimLocation DL
LEFT JOIN Fact F
   ON DL.DealershipLocationID = F.DealershipLocationID
GROUP BY dl.Dealership_Location
ORDER BY 2
```

Results Messages

	Dealership_Location	sales_quantity
1	Rural	1548
2	Urban	1572
3	Suburban	1628

9. What is sales by each Dealership Location?

```
SELECT
   DL.Dealership_Location,
   FORMAT(SUM(Sales_Price),'C2') AS Total_Revenue,
   FORMAT(SUM(Cost_Price),'C2') AS Total_Cost,
   FORMAT(SUM((Sales_Price - Cost_Price) * Production_Quantity),'C2') AS
Total_Profit
FROM DimLocation DL
LEFT JOIN Fact F
   ON DL.DealershipLocationID = F.DealershipLocationID
GROUP BY Dealership_Location
ORDER BY 4 DESC, 2, 3
```

Results Messages

	Dealership_Location	Total_Revenue	Total_Cost	Total_Profit
1	Suburban	$76,725,248.00	$57,728,121.00	$9,190,438,699.00
2	Urban	$73,761,937.00	$56,383,303.00	$8,441,308,104.00
3	Rural	$72,608,048.00	$56,589,430.00	$8,221,530,931.00

10. Top Dealership Location by Profit:

```
SELECT TOP(1)
   DL.Dealership_Location,
   FORMAT(SUM((Sales_Price - Cost_Price) * Production_Quantity),'C2') AS
Total_Profit
FROM DimLocation DL
LEFT JOIN Fact F
   ON DL.DealershipLocationID = F.DealershipLocationID
GROUP BY Dealership_Location
ORDER BY 2
```

Results Messages

	Dealership_Location	Total_Profit
1	Rural	$8,221,530,931.00

11. Top Counrty_of_Origin by Profit:

```
SELECT TOP(1)
   DL.Country_of_Origin,
   FORMAT(SUM((Sales_Price - Cost_Price) * Production_Quantity),'C2') AS
Total_Profit
FROM DimLocation DL
LEFT JOIN Fact F
   ON DL.DealershipLocationID = F.DealershipLocationID
GROUP BY Country_of_Origin
ORDER BY 2
```

Results Messages

	Country_of_Origin	Total_Profit
1	Germany	$4,866,438,242.00

PRODUCT ANALYSIS

1. How many car models are available?

```
SELECT DISTINCT Car_Model
FROM DimCar
```

Results Messages

	Car_Model
1	Sports Car
2	Truck
3	Hatchback
4	Sedan
5	SUV

2. Which car model generates the most sales?

```
SELECT TOP(1) DC.Car_Model, COUNT(F.Color) AS sales_quantity
FROM DimCar DC
LEFT JOIN Fact F
   ON DC.DealershipLocationID = F.DealershipLocationID
GROUP BY DC.Car_Model
ORDER BY 2 DESC
```

Results Messages

	Car_Model	sales_quantity
1	SUV	1180

3. Which fuel types are most commonly used across car models?

```
SELECT Fuel_Type, COUNT(DISTINCT Car_Model) AS Model_Count
FROM DimCar
GROUP BY Fuel_Type
ORDER BY 2 DESC
```

Results Messages

	Fuel_Type	Model_Count
1	Diesel	5
2	Electric	5
3	Gasoline	5

4. Which car models have above average safety ratings?

```
SELECT DISTINCT DC.Car_Model
FROM DimCar DC
JOIN Fact F
   ON DC.DealershipLocationID = F.DealershipLocationID
WHERE F.Safety_Rating > (SELECT AVG(Safety_Rating) FROM Fact)
ORDER BY 1
```

Results Messages

	Car_Model
1	Hatchback
2	Sedan
3	Sports Car
4	SUV
5	Truck

5. What is the average warranty period for each car model?

```
SELECT DC.Car_Model, AVG(F.Warranty_Period_months) AS Warranty_Period_months
FROM DimCar DC
LEFT JOIN Fact F
   ON DC.DealershipLocationID = F.DealershipLocationID
GROUP BY DC.Car_Model
ORDER BY 2 DESC
```

Results Messages

	Car_Model	Warranty_Period_months
1	Hatchback	36
2	Sedan	35
3	SUV	35
4	Sports Car	35
5	Truck	35

6. How does the average cost price vary by car model and fuel type?

```
SELECT DC.Car_Model, DC.Fuel_Type,
   FORMAT(AVG(F.Cost_Price), 'C2') AS Avg_Cost_Price
FROM DimCar DC
LEFT JOIN Fact F
   ON DC.DealershipLocationID = F.DealershipLocationID
GROUP BY DC.Car_Model, DC.Fuel_Type
ORDER BY 1, 2, 3 DESC
```

Results Messages

	Car_Model	Fuel_Type	Avg_Cost_Price
1	Hatchback	Diesel	$35,980.64
2	Hatchback	Electric	$37,714.00
3	Hatchback	Gasoline	$36,788.25
4	Sedan	Diesel	$36,384.58
5	Sedan	Electric	$35,244.95
6	Sedan	Gasoline	$33,752.46
7	Sports Car	Diesel	$36,224.73
8	Sports Car	Electric	$35,386.28
9	Sports Car	Gasoline	$36,459.69
10	SUV	Diesel	$34,906.31
11	SUV	Electric	$35,444.58
12	SUV	Gasoline	$34,449.71
13	Truck	Diesel	$36,531.13
14	Truck	Electric	$37,093.87
15	Truck	Gasoline	$37,274.36

7. What is the total sales generated by each car model ?

```
SELECT
   DC.Car_Model,
   FORMAT(SUM(Sales_Price),'C2') AS Total_Revenue,
   FORMAT(SUM(Cost_Price),'C2') AS Total_Cost,
   FORMAT(SUM((Sales_Price - Cost_Price) * Production_Quantity),'C2') AS
Total_Profit
FROM DimCar DC
LEFT JOIN Fact F
   ON DC.DealershipLocationID = F.DealershipLocationID
GROUP BY DC.Car_Model
ORDER BY 4 DESC, 2, 3
```

Results Messages

	Car_Model	Total_Revenue	Total_Cost	Total_Profit
1	SUV	$55,478,586.00	$41,379,480.00	$7,160,143,622.00
2	Hatchback	$54,691,704.00	$41,494,786.00	$6,493,349,808.00
3	Sedan	$52,857,098.00	$40,538,761.00	$5,997,167,674.00
4	Sports Car	$32,525,579.00	$25,486,022.00	$3,476,279,983.00
5	Truck	$27,542,266.00	$21,801,805.00	$2,726,336,647.00

8. What is the most common car type in sales?

```
SELECT DC.Car_Model, COUNT(F.Color) AS sales_quantity
FROM DimCar DC
LEFT JOIN Fact F
   ON DC.DealershipLocationID = F.DealershipLocationID
GROUP BY DC.Car_Model
ORDER BY 2 DESC
```

Results Messages

	Car_Model	sales_quantity
1	SUV	1180
2	Hatchback	1137
3	Sedan	1135
4	Sports Car	708
5	Truck	588

9. Which car type generates the highest profit?

```
SELECT TOP(1)
   DC.Car_Model,
   FORMAT(SUM((Sales_Price - Cost_Price) * Production_Quantity),'C2') AS
Total_Profit
FROM DimCar DC
LEFT JOIN Fact F
   ON DC.DealershipLocationID = F.DealershipLocationID
GROUP BY DC.Car_Model
ORDER BY 2 DESC
```

Results Messages

	Car_Model	Total_Profit
1	SUV	$7,160,143,622.00

10. Sales volume for each color

```
SELECT DC.Color, COUNT(F.Color) AS Sales_quantity
FROM DimCar DC
LEFT JOIN Fact F
       ON DC.DealershipLocationID = F.DealershipLocationID
GROUP BY DC.Color
ORDER BY 2
```

Results Messages

	Color	Sales_quantity
1	Black	787
2	White	898
3	Blue	926
4	Red	1039
5	Silver	1098