EDA

Q1. How many rows are in each table?

Table Name	Rows	Code
Accounts	351	`SELECT COUNT(*) FROM Accounts;`
Orders	6912	`SELECT COUNT(*) FROM orders;`
Region	4	`SELECT COUNT(*) FROM Region;`
Sales_Reps	50	`SELECT COUNT(*) FROM sales_reps;`
Web_events	9073	`SELECT COUNT(*) FROM web_events;`

Q2. What is the date range of orders?

```
SELECT
      TO_CHAR(MIN(DATE_TRUNC('day',occurred_at)),'YYYY-MM-DD') AS
      Earliest_order_date,
       TO_CHAR(MAX(DATE_TRUNC('day',occurred_at)),'YYYY-MM-DD') AS Latest_order_date
FROM Orders;
```

earliest_order_date	latest_order_date
2013-12-04	2017-01-02

Q3. How many of each type of paper has been sold?

```
SELECT SUM(standard_qty) AS Total_Standard_Sales,
      SUM(gloss_qty) AS Total_Gloss_Sales,
      SUM(poster_qty) AS Total_Poster_Sales
FROM Orders;
```

total_standard_sales	total_gloss_sales	total_poster_sales
1938346	1013773	723646

Q4. How much, in dollars, have each of the paper types sold?

```
SELECT SUM(standard_amt_usd) AS Total_Standard_Sales_usd,
       SUM(gloss_amt_usd) AS Total_Gloss_Sales_usd,
       SUM(poster_amt_usd) AS Total_Poster_Sales_usd
FROM Orders;
```

total_standard_sales_usd	total_gloss_sales_usd	total_poster_sales_usd
9672346.54	7593159.77	5876005.52

Q5. What is the most profitable paper type?

From above, **Standard** is the most profitable paper type.

Q6. What are the top five accounts by average total amount?

```
SELECT a.name Account_Name,
       ROUND(AVG(o.total_amt_usd),2) AS Avg_Total_Amt_usd
FROM Accounts a
JOIN Orders o
ON a.id= o.account_id
GROUP BY 1
ORDER BY 2 DESC
LIMIT 5;
```

account_name	avg_total_amt_usd
Pacific Life	19639.94
Fidelity National Financial	13753.41
Kohl's	12872.17
State Farm Insurance Cos.	12423.39
AmerisourceBergen	9685.45

Q7. What channel do most of the online sales come from?

```
SELECT channel,
       COUNT(*) AS Count_events
FROM web_events
GROUP BY channel_name
ORDER BY Count_events DESC
```

```
LIMIT 1;
```

channel	count_events
direct	5298

Most Online Sales came from **Direct** channel.

Q8. Which region_id has the largest number of sales persons?

```
SELECT Region_id, COUNT(DISTINCT(name)) AS sales_reps_count
FROM sales_reps
GROUP BY 1
ORDER BY 2 DESC
LIMIT 1;
```

region_id	sales_reps_count
1	21

Region_id 1 has largest number of sales persons.

Joins

Q9: Which web events channel had the highest total quantity sold of all three types of paper?

```
SELECT w.channel channel_name,
      SUM(o.total) AS Total_sold,
      SUM(o.standard_qty) AS Standard_Sales,
      SUM(o.poster_qty) AS Poster_Sales,
      SUM(o.gloss_qty) AS Gloss_Sales
FROM Accounts a
JOIN Orders o
ON a.id= o.account_id
JOIN web_events w
ON a.id= w.account_id
GROUP BY 1
ORDER BY 2 DESC
```

```
LIMIT 1;
```

channel_name	total_sold	standard_sales	poster_sales	gloss_sales
direct	102408010	55810806	19350472	27246732

Direct had most sales of all 3 types of paper.

Q10: Which region, by name, has the highest amount of sales in USD?

```
SELECT r.name region_name,
       SUM(o.total_amt_usd) AS Total_sales_usd
FROM Accounts a
JOIN Orders o
ON a.id= o.account_id
JOIN sales_reps s
ON s.id= a.sales_rep_id
JOIN Region r
ON r.id= s.region_id
GROUP BY 1
ORDER BY 2 DESC
LIMIT 1;
```

region_name	total_sales_usd
Northeast	7744405.36

Northeast had highest amount of Sales in USD.

CTEs, Sub queries, and temp tables

Q11: Categorize each region's average sales as "Above Average" or "Below Average" based on the average for the company as a whole.

```
WITH T1 AS
             (SELECT ROUND(AVG(total_amt_usd),2) AS Company_average_sales
             FROM Orders o)
```

```
SELECT r.name region_name,
       ROUND(AVG(o.total_amt_usd),2) Region_Average_Sales,
       T1.Company_average_sales,
CASE
       WHEN AVG(o.total_amt_usd)> T1.Company_average_sales THEN 'Above Average'
       ELSE 'Below Average'
END AS Avg_Sales_Category
FROM Accounts a
JOIN Orders o
ON a.id= o.account_id
JOIN sales_reps s
ON s.id= a.sales_rep_id
JOIN Region r
ON r.id= s.region_id
CROSS JOIN T1
GROUP BY 1,3
ORDER BY 2 DESC ;
```

region_name	region_average_sales	company_average_sales	avg_sales_category
West	3626.15	3348.02	Above Average
Midwest	3359.52	3348.02	Above Average
Northeast	3285.7	3348.02	Below Average
Southeast	3190.96	3348.02	Below Average

Q12: What are the total quantities of each paper type sold for the top region?

```
WITH region_totals AS (
               SELECT
                 r.name AS region_name,
                 SUM(o.standard_qty) AS standard_qty,
                 SUM(o.gloss_qty)  AS gloss_qty,
                 SUM(o.poster_qty)  AS poster_qty,
                 SUM(o.total) AS total_qty
               FROM orders o
```

```
                JOIN accounts a  ON o.account_id = a.id
                JOIN sales_reps s  ON a.sales_rep_id = s.id
                JOIN region r  ON s.region_id = r.id
                GROUP BY r.name
),
top_region AS (
                SELECT *
                FROM region_totals
                ORDER BY total_qty DESC
                LIMIT 1
)
SELECT
        region_name,
        standard_qty,
        gloss_qty,
        poster_qty
FROM top_region;
```

region_name	standard_qty	gloss_qty	poster_qty
Northeast	646871	351679	231828

Windowing Functions

Q13: What are the average sales in USD by region and sales person? Include region name, sales person's name, and account name. Include first 20 rows

```
WITH T1 AS(
            SELECT r.name region_name,
            s.name sales_rep,
            a.name account_name,
            ROUND(AVG(o.total_amt_usd) OVER(PARTITION BY r.name, s.name, a.name),2)
            AS avg_sales,
            ROW_NUMBER() OVER(PARTITION BY r.name, s.name, a.name ORDER BY o.id) AS
            rn
            FROM Accounts a
```

```
            JOIN Orders o
            ON a.id= o.account_id
            JOIN Sales_reps s
            ON s.id= a.sales_rep_id
            JOIN Region r
            on r.id= s.region_id)
SELECT region_name,
       Sales_rep,
       Account_name,
       Avg_Sales
FROM T1
WHERE rn=1
ORDER BY avg_sales DESC
LIMIT 20;
```

region_name	sales_rep	account_name	avg_sales
West	Dawna Agnew	Pacific Life	19639.94
West	Soraya Fulton	Fidelity National Financial	13753.41
Midwest	Cordell Rieder	Kohl's	12872.17
Northeast	Julia Behrman	State Farm Insurance Cos.	12423.39
Northeast	Ayesha Monica	AmerisourceBergen	9685.45
Southeast	Babette Soukup	CBS	8648.07
Northeast	Renetta Carew	Berkshire Hathaway	7474.32
Midwest	Charles Bidwell	Starbucks	7437.6
Southeast	Vernita Plump	CenturyLink	7422.51
Southeast	Earlie Schleusner	Edison International	7402.19
Northeast	Tia Amato	CHS	7272.69
West	Brandie Riva	Precision Castparts	7082.65
West	Arica Stoltzfus	EOG Resources	6175.38
West	Brandie Riva	Public Service Enterprise Group	6152.75
Southeast	Saran Ram	Illinois Tool Works	6130.9
Southeast	Dorotha Seawell	Facebook	5880.7
West	Brandie Riva	Republic Services	5877.22
West	Micha Woodford	Eversource Energy	5705.98
Northeast	Necole Victory	IBM	5446.99
Midwest	Chau Rowles	Halliburton	5440.9

Q14: What is the running total of sales by month? Return the first twenty rows

```
WITH Monthly_sales AS(
                      SELECT EXTRACT(Year FROM occurred_at) AS Yr,
                     EXTRACT(Month FROM occurred_at) AS Month,
                     SUM(total_amt_usd) AS Total_Month_Sales
                     FROM Orders
                     GROUP BY 1,2
                     )
SELECT Yr,
       Month,
       Total_Month_Sales,
       SUM(total_month_sales) OVER( ORDER BY Yr, Month
                            ROWS BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW)
                            AS Running_Total_Month_Sales
FROM Monthly_Sales
ORDER BY Yr, Month
LIMIT 20;
```

yr	month	total_month_sales	running_total_month_sales
2013	12	377331.00	377331.00
2014	1	286140.27	663471.27
2014	2	349721.34	1013192.61
2014	3	341512.32	1354704.93
2014	4	344893.99	1699598.92
2014	5	319210.40	2018809.32
2014	6	297655.65	2316464.97
2014	7	289128.19	2605593.16
2014	8	366685.41	2972278.57
2014	9	299968.38	3272246.95
2014	10	495333.59	3767580.54
2014	11	311893.88	4079474.42
2014	12	366963.12	4446437.54
2015	1	347804.30	4794241.84
2015	2	333688.01	5127929.85
2015	3	519403.40	5647333.25
2015	4	451753.57	6099086.82
2015	5	390830.84	6489917.66
2015	6	420906.13	6910823.79

2015	7	461895.49	7372719.28

Q15: Create a seven-day moving average of orders (hint: CTE, temp table, or subquery).

```
WITH Daily_Sales AS (SELECT
                    TO_CHAR(DATE_TRUNC('Day', Occurred_at), 'YYYY-MM-DD') AS
                    Order_Date,
                     COUNT(*) AS Total_Orders,
                    SUM(Total_amt_usd) AS Day_Sales
                    FROM Orders
                    GROUP BY 1)

SELECT Order_Date,
ROUND(
      AVG(Total_Orders) OVER( ORDER BY Order_Date
      ROWS BETWEEN 6 PRECEDING AND CURRENT ROW),2)
AS Seven_day_moving_avg_order_total_orders,
ROUND(
      AVG(Day_Sales) OVER(ORDER BY Order_Date
            ROWS BETWEEN 6 PRECEDING AND CURRENT ROW),2)
AS Seven_Day_Moving_Avg_total_Sales
FROM Daily_Sales
ORDER BY Order_Date;
```

order_date	seven_day_moving_avg_order_total_orders	seven_day_moving_avg_total_sales
12/4/2013	3	5983.87
12/5/2013	2.5	4631.43
12/6/2013	4	11111.33
12/8/2013	5	22907.2
12/9/2013	4.6	19369.72
12/10/2013	4.5	17352.52
12/11/2013	4.71	18006.49
12/12/2013	5.14	19951.29

12/13/2013	5.14	20669.34
12/14/2013	4.86	19213.04
12/15/2013	3.86	11245.57
12/16/2013	3.86	11344.67
12/17/2013	3.86	12262.73
12/18/2013	3.43	11348.98
12/19/2013	3.14	10924.37
12/21/2013	3.29	11230.84
12/22/2013	3.43	12385.72
12/23/2013	3.57	12933.2
12/24/2013	3.43	13291.59
12/25/2013	3.29	12658.98