**Case Study #7 - Balanced Tree**

**A. High Level Sales Analysis**

**1 - What was the total quantity sold for all products?**

SELECT

product\_name,

SUM(qty) AS total\_quantity

FROM sales AS s

JOIN product\_details AS pd ON pd.product\_id = s.prod\_id

GROUP BY product\_name

**2 - What is the total generated revenue for all products before discounts?**

SELECT

product\_name,

SUM(qty \* s.price) AS total\_revenue

FROM sales AS s

JOIN product\_details AS pd ON pd.product\_id = s.prod\_id

GROUP BY product\_name

**3 - What was the total discount amount for all products?**

SELECT

product\_name,

SUM(s.price \* qty \* discount/100) AS total\_discount

FROM sales AS s

JOIN product\_details AS pd ON pd.product\_id = s.prod\_id

GROUP BY product\_name

**B. Transaction Analysis**

**1 - How many unique transactions were there?**

SELECT

COUNT(DISTINCT txn\_id) AS trans\_count

FROM sales

**2 - What is the average unique products purchased in each transaction?**

SELECT

ROUND(SUM(qty) / (SELECT COUNT(DISTINCT txn\_id) FROM sales)) AS avg\_qty

FROM sales AS s

**3 - What are the 25th, 50th and 75th percentile values for the revenue per transaction?**

WITH revenue\_quartiles AS (

SELECT

SUM(price\*qty) AS revenue,

NTILE(4) OVER(ORDER BY SUM(price\*qty)) AS quartile

FROM sales

GROUP BY txn\_id

)

SELECT

MAX(CASE WHEN quartile = 1 THEN revenue END) AS percentile\_25th,

MAX(CASE WHEN quartile = 2 THEN revenue END) AS percentile\_50th,

MAX(CASE WHEN quartile = 3 THEN revenue END) AS percentile\_75th

FROM revenue\_quartiles;

**4 - What is the average discount value per transaction?**

SELECT

ROUND(SUM(price \* qty \* discount/100)/(SELECT COUNT(DISTINCT txn\_id) FROM sales),2) as avg\_discount

FROM sales

**5 - What is the percentage split of all transactions for members vs non-members?**

SELECT

ROUND(SUM(IF(member,1,0)) / (SELECT COUNT(txn\_id) FROM sales)\*100) AS member\_percentage,

100 - ROUND(SUM(IF(member,1,0)) / (SELECT COUNT(txn\_id) FROM sales)\*100) AS non\_member\_percentage

FROM sales

**6 - What is the average revenue for member transactions and non-member transactions?**

SELECT

ROUND(SUM(IF(member,qty\*price,0))/(SELECT COUNT(DISTINCT txn\_id) FROM sales WHERE member),2) as member\_avg\_revenue,

ROUND(SUM(IF(!member,qty\*price,0))/(SELECT COUNT(DISTINCT txn\_id) FROM sales WHERE !member),2) as non\_member\_avg\_revenue

FROM sales

**C. Product Analysis**

**1 - What are the top 3 products by total revenue before discount?**

SELECT

product\_name,

SUM(qty \* s.price) AS total\_revenue

FROM sales AS s

JOIN product\_details AS pd ON pd.product\_id = s.prod\_id

GROUP BY product\_name

ORDER BY total\_revenue DESC

LIMIT 3

**2 - What is the total quantity, revenue and discount for each segment?**

SELECT

segment\_name,

SUM(qty) as total\_quantity,

SUM(qty \* s.price) AS total\_revenue,

SUM(qty \* s.price \* discount/100) AS total\_discount

FROM sales AS s

JOIN product\_details AS pd ON pd.product\_id = s.prod\_id

GROUP BY segment\_name;

**3 - What is the top selling product for each segment?**

SELECT

segment\_name,

product\_name,

total\_quantity

FROM

(SELECT

product\_name,

segment\_name,

SUM(qty) AS total\_quantity,

DENSE\_RANK() OVER (PARTITION BY segment\_name ORDER BY SUM(qty) DESC) AS ranking

FROM sales AS s

JOIN product\_details AS pd ON pd.product\_id = s.prod\_id

GROUP BY product\_name,segment\_name) as x

WHERE ranking = 1

**4 - What is the total quantity, revenue and discount for each category?**

SELECT

category\_name,

SUM(qty) as total\_quantity,

SUM(qty \* s.price) AS total\_revenue,

SUM(qty \* s.price \* discount/100) AS total\_discount

FROM sales AS s

JOIN product\_details AS pd ON pd.product\_id = s.prod\_id

GROUP BY category\_name;

**5 - What is the top selling product for each category?**

SELECT

category\_name,

product\_name,

total\_quantity

FROM

(SELECT

product\_name,

category\_name,

SUM(qty) AS total\_quantity,

DENSE\_RANK() OVER (PARTITION BY category\_name ORDER BY SUM(qty) DESC) AS ranking

FROM sales AS s

JOIN product\_details AS pd ON pd.product\_id = s.prod\_id

GROUP BY product\_name,category\_name) as x

WHERE ranking = 1

**6 - What is the percentage split of revenue by product for each segment?**

WITH total\_revenue\_segment AS

(SELECT

segment\_name,

SUM(qty \* s.price) AS total\_revenue

FROM sales AS s

JOIN product\_details AS pd ON pd.product\_id = s.prod\_id

GROUP BY segment\_name)

SELECT

pd.segment\_name,

product\_name,

SUM(qty\*s.price) / total\_revenue \* 100 as revenue\_percentage

FROM sales AS s

JOIN product\_details AS pd ON pd.product\_id = s.prod\_id

JOIN total\_revenue\_segment AS trs ON trs.segment\_name = pd.segment\_name

GROUP BY pd.segment\_name,product\_name

ORDER BY pd.segment\_name

**7 - What is the percentage split of revenue by segment for each category?**

WITH total\_revenue\_category AS

(SELECT

category\_name,

SUM(qty \* s.price) AS total\_revenue

FROM sales AS s

JOIN product\_details AS pd ON pd.product\_id = s.prod\_id

GROUP BY category\_name)

SELECT

pd.category\_name,

product\_name,

SUM(qty\*s.price) / total\_revenue \* 100 as revenue\_percentage

FROM sales AS s

JOIN product\_details AS pd ON pd.product\_id = s.prod\_id

JOIN total\_revenue\_category AS trc ON trc.category\_name = pd.category\_name

GROUP BY pd.category\_name,product\_name

ORDER BY pd.category\_name

**8 - What is the percentage split of total revenue by category?**

SELECT

category\_name,

SUM(qty \* s.price)/(SELECT SUM(qty\*price) OVER() FROM sales LIMIT 1)\*100 AS total\_revenue

FROM sales AS s

JOIN product\_details AS pd ON pd.product\_id = s.prod\_id

GROUP BY category\_name

**9 - What is the total transaction “penetration” for each product? (hint: penetration = number of transactions where at least 1 quantity of a product was purchased divided by total number of transactions)**

WITH count\_prod\_trans AS

(SELECT

product\_name,

txn\_id,

ROW\_NUMBER() OVER(PARTITION BY product\_name) as ranking

FROM sales AS s

JOIN product\_details AS pd ON pd.product\_id = s.prod\_id

GROUP BY product\_name,txn\_id)

SELECT

DISTINCT product\_name,

LAST\_VALUE(ranking) OVER(PARTITION BY product\_name) / (SELECT COUNT(DISTINCT txn\_id) FROM sales) AS penetration

FROM count\_prod\_trans

**10 - What is the most common combination of at least 1 quantity of any 3 products in a 1 single transaction?**