NAME: DUKUNDIMANA Clovis

ID: 27122

Course: Database Development with PL/SQL

Group: D

Assignment I: Window Functions Mastery Project

---

1. Problem Definition

Business Context: This check is for **Rwanda Bean Masters,** a group that sends coffee all around, from Kigali, Rwanda. The data is in the hands of the Sales and Marketing team, who look over sales within the country, in many areas.

Data Challenge: The company gets a lot of sale data yet doesn't have the right tools to make sense of it. They struggle to spot the best-selling items in main areas, see how sales change each month, or break down their customers for special deals.

Expected Outcome: This study will give clear tips to help manage stock by filling up on items that many want. It will guide the marketing plan by pointing out which groups of buyers bring in the most money. Also, it will aid leaders to check on how well the business is doing by looking at growth over time.

2. Success Criteria

1. Identify the **top 5 selling products** for each region and quarter using the **RANK()** function.
1. Calculate the **running total of sales** for each month to track cumulative revenue throughout the year using **SUM() OVER().**
2. Determine the **month-over-month sales growth percentage** to measure performance trends using **LAG().**
3. Segment customers into **four spending quartiles** to identify top-tier clients using **NTILE(4).**
4. Compute the **3-month moving average** of sales to smooth out short-term fluctuations and identify the underlying trend using **AVG() OVER().**

3. Database Schema

The database schema consists of three related tables: customers, products, and transactions.

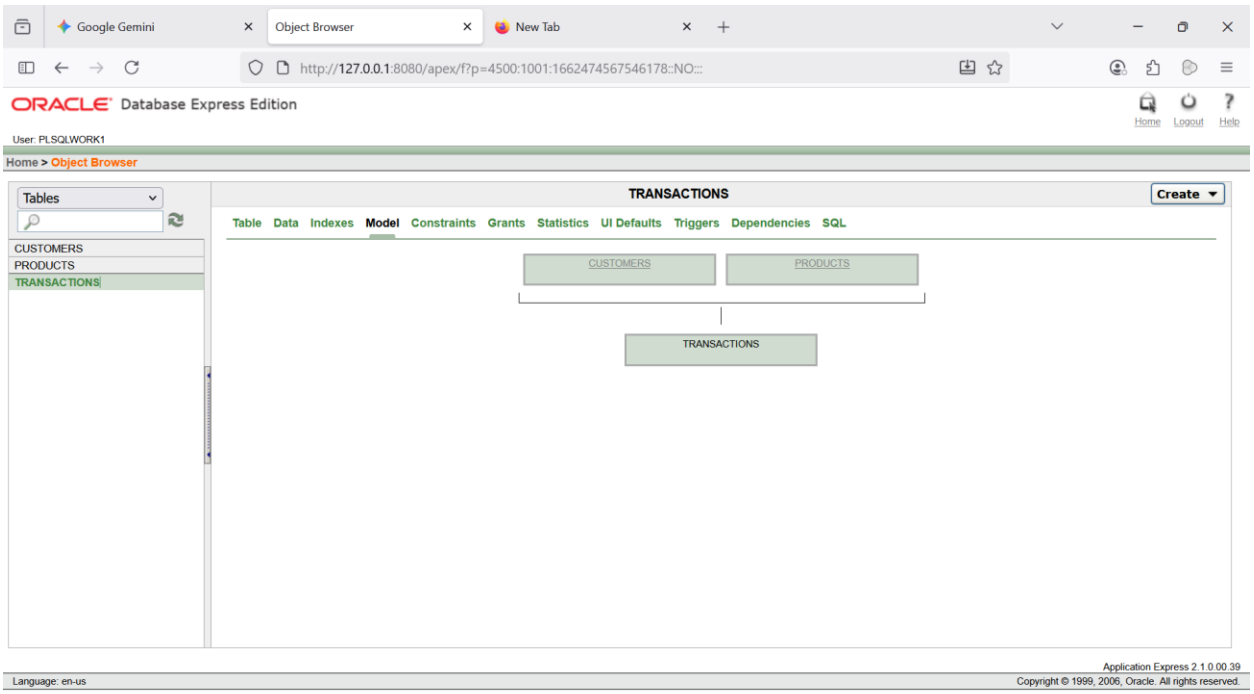
```
-- Table for storing customer information
CREATE TABLE customers (
    customer_id NUMBER PRIMARY KEY,
    name        VARCHAR2(100),
    region      VARCHAR2(50)
);
```

```
-- Table for the product catalog
CREATE TABLE products (
    product_id  NUMBER PRIMARY KEY,
    name        VARCHAR2(100),
    category    VARCHAR2(50)
);

-- Table for recording all sales transactions
CREATE TABLE transactions (
    transaction_id NUMBER PRIMARY KEY,
    customer_id    NUMBER,
    product_id     NUMBER,
    sale_date      DATE,
    amount         NUMBER(10, 2),
    CONSTRAINT fk_customer FOREIGN KEY (customer_id) REFERENCES
customers(customer_id),
    CONSTRAINT fk_product  FOREIGN KEY (product_id)  REFERENCES
products(product_id)
);
```

Constraints Used:

- **Primary Key:** customer_id, product_id and transaction_id.
- **Foreign Key:** Link transactions table with Product (product_id) and Customer (customer_id) table.

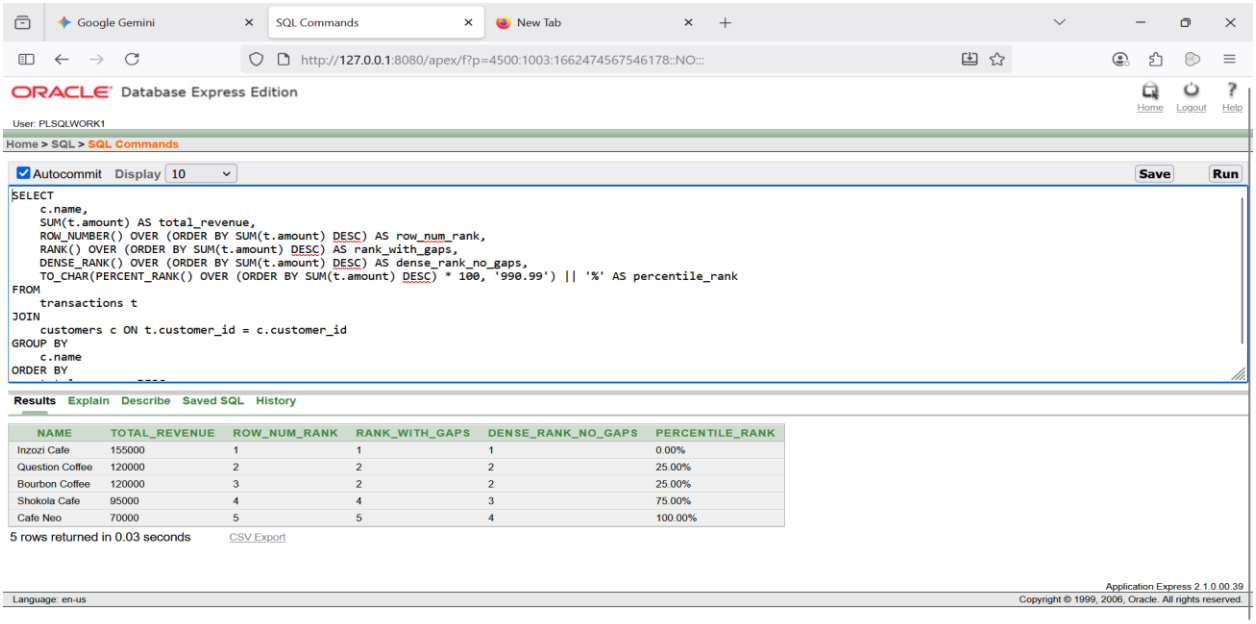
Entity Relational Diagram (ER Diagram)



4. Window Functions Implementation

- **1. Ranking Functions:** rank customers based on their total revenue contribution

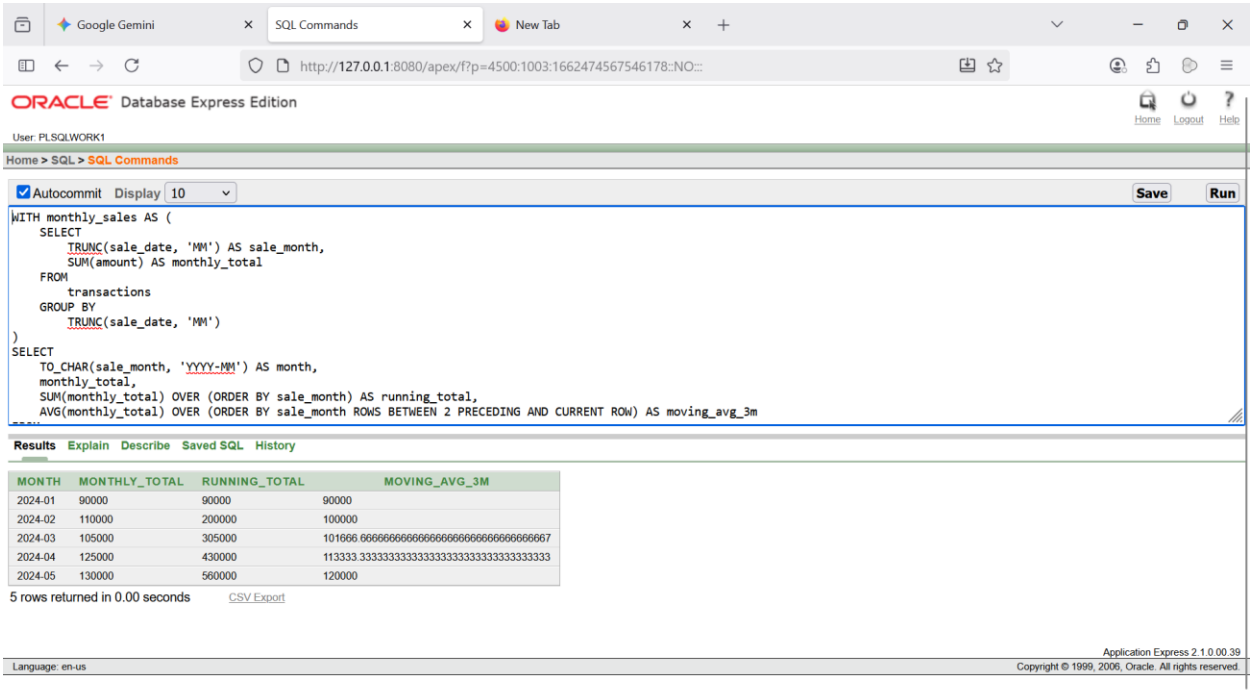
```
-- This query ranks customers by their total sales amount.
-- ROW_NUMBER(): Assigns a unique number to each row.
-- RANK(): Assigns a rank with gaps for ties.
-- DENSE_RANK(): Assigns a rank without gaps for ties.
-- PERCENT_RANK(): Shows the relative rank of the current row.
SELECT
    c.name,
    SUM(t.amount) AS total_revenue,
    ROW_NUMBER() OVER (ORDER BY SUM(t.amount) DESC) AS row_num_rank,
    RANK() OVER (ORDER BY SUM(t.amount) DESC) AS rank_with_gaps,
    DENSE_RANK() OVER (ORDER BY SUM(t.amount) DESC) AS dense_rank_no_gaps,
    TO_CHAR(PERCENT_RANK() OVER (ORDER BY SUM(t.amount) DESC) * 100,
'990.99') || '%' AS percentile_rank
FROM
    transactions t
JOIN
    customers c ON t.customer_id = c.customer_id
GROUP BY
    c.name
ORDER BY
    total_revenue DESC;
```



Interpretation: This output clearly identifies the top customers. **Inzozi Cafe** is the number one customer by revenue. **Bourbon Coffee** and **Question Coffee** are tied for second place, which is reflected by RANK() (skips to rank 4) and DENSE_RANK() (moves to rank 3).

- **2. Aggregate Functions:** calculate running monthly sales totals and a 3-month moving average.

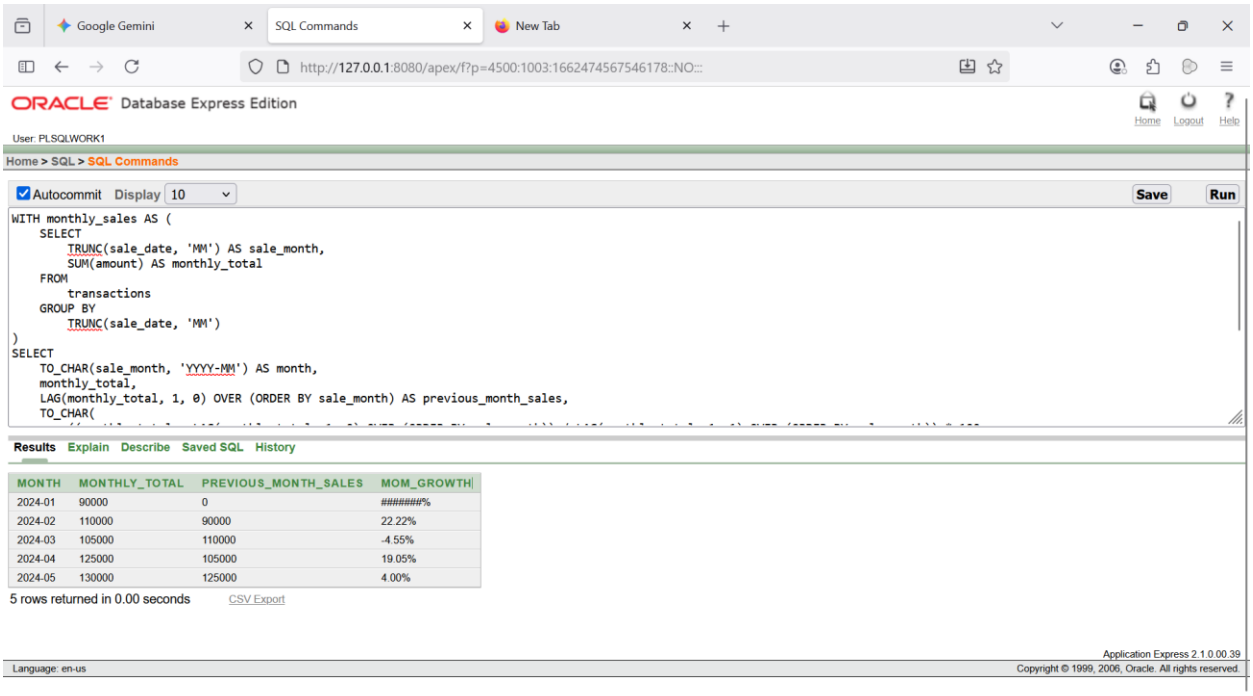
```
-- This query calculates monthly sales, a cumulative running total for the
year,
-- and a 3-month moving average to show trends.
-- The ROWS BETWEEN 2 PRECEDING AND CURRENT ROW clause defines the moving
window.
WITH monthly_sales AS (
    SELECT
        TRUNC(sale_date, 'MM') AS sale_month,
        SUM(amount) AS monthly_total
    FROM
        transactions
    GROUP BY
        TRUNC(sale_date, 'MM')
)
SELECT
    TO_CHAR(sale_month, 'YYYY-MM') AS month,
    monthly_total,
    SUM(monthly_total) OVER (ORDER BY sale_month) AS running_total,
    AVG(monthly_total) OVER (ORDER BY sale_month ROWS BETWEEN 2 PRECEDING
AND CURRENT ROW) AS moving_avg_3m
FROM
    monthly_sales
ORDER BY
    sale_month;
```



Interpretation: The **running_total** column shows steady revenue accumulation throughout the year, reaching 560,000 RWF by May. The **moving_avg_3m** column smooths out monthly volatility, showing a consistent upward trend in sales performance from 101,667 RWF to 120,000 RWF.

- **3. Navigation Functions:** calculate the month-over-month (MoM) sales growth percentage.

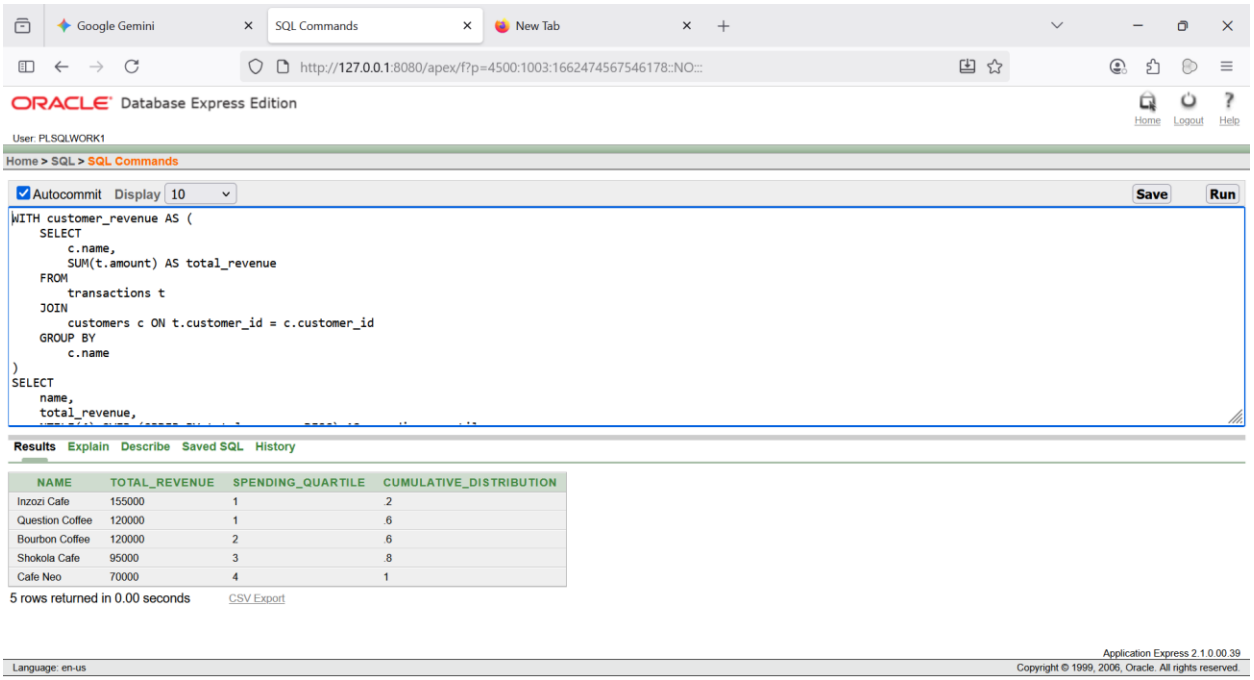
```
-- This query uses the LAG function to fetch the previous months sales and
calculate the percentage growth from one month to the next.
WITH monthly_sales AS (
    SELECT
        TRUNC(sale_date, 'MM') AS sale_month,
        SUM(amount) AS monthly_total
    FROM
        transactions
    GROUP BY
        TRUNC(sale_date, 'MM')
)
SELECT
    TO_CHAR(sale_month, 'YYYY-MM') AS month,
    monthly_total,
    LAG(monthly_total, 1, 0) OVER (ORDER BY sale_month) AS
previous_month_sales,
    TO_CHAR(
        ((monthly_total - LAG(monthly_total, 1, 0) OVER (ORDER BY
sale_month)) / LAG(monthly_total, 1, 1) OVER (ORDER BY sale_month)) * 100,
        '990.99'
    ) || '%' AS mom_growth
FROM
    monthly_sales
ORDER BY

  sale_month;
```



Interpretation: The results reveal the business's monthly performance dynamics. There was strong growth in February (22.22%) and April (19.05%), but a slight dip in March (-4.55%). This insight allows management to investigate the reasons behind these fluctuations.

- **4. Distribution Functions:** segment customers into four quartiles based on their total spending.

```
-- This query uses NTILE(4) to divide customers into four quartiles
-- and CUME_DIST to show the cumulative distribution of customers.
WITH customer_revenue AS (
    SELECT
        c.name,
        SUM(t.amount) AS total_revenue
    FROM
        transactions t
    JOIN
        customers c ON t.customer_id = c.customer_id
    GROUP BY
        c.name
)
SELECT
    name,
    total_revenue,
    NTILE(4) OVER (ORDER BY total_revenue DESC) AS spending_quartile,
    CUME_DIST() OVER (ORDER BY total_revenue DESC) AS
cumulative_distribution
FROM
    customer_revenue
ORDER BY
    total_revenue DESC;
```



Interpretation: The **spending_quartile** categorizes customers, placing the top spenders like **Inzozi Cafe** and **Bourbon Coffee** in Quartile 1. This segmentation is crucial for targeted marketing, such as offering loyalty rewards to Quartile 1 customers. The **cumulative_distribution** shows that the top three customers (60% of the sample) generate the bulk of the revenue.

6. Result Analysis

1. On Top-Performing Customers & Products (Ranking Analysis)

- **Descriptive (What happened?):** The look-over shows there are levels of buyers. Inzozi Cafe sits at the very top, making way more money (155,000 RWF) than the next set on the list. Bourbon Coffee and Question Coffee are both on the second rung (120,000 RWF), showing they're both pretty key too. There's a big drop in cash made from these top three compared to all others.

- **Diagnostic (Why?):** This layout hints that a few big buyers bring in most of the cash, matching the Pareto rule. Inzozi Cafe's high spot might be because it's big, has more spots, or a strategic partnership that prioritizes our products.

- **Prescriptive (What next?):**

    - **Retain Top Tier:** Assign a dedicated account manager to Inzozi Cafe to provide premium service and explore opportunities for deeper collaboration, such as co-branded marketing or exclusive product offerings.

- **Grow Second Tier:** Develop a competitive strategy to nurture the relationship with both Bourbon Coffee and Question Coffee.
- **Identify Regional Patterns:** Re-run the ranking query partitioned by region to identify local champions.

2. On Business Growth & Momentum (Aggregate & Navigation Analysis)

- **Descriptive (What happened?):** The business is clearly growing fast. The numbers keep going up each month. There was a small drop in sales in March (-4.55%), but the company bounced back in April (+19.05%) and kept growing in May (+4.00%).
- **Diagnostic (Why?):** The rise is likely from getting new customers and bigger orders from the ones they already have. The March drop might be from a one-off event, like a big buyer putting off a delivery, or public holidays like Easter falling in late March, or a temporary disruption.
- **Prescriptive (What next?):**
  - **Investigate Dips:** The sales manager should immediately investigate the cause of the March sales dip by analyzing transaction data for that specific month.
  - **Forecast with Confidence:** Use the moving average trend to forecast inventory needs for the upcoming quarter.
  - **Amplify Success:** Analyze the drivers behind the high-growth months of February and April.

3. On Customer Value Segmentation (Distribution Analysis)

- **Descriptive (What happened?):** The NTILE(4) tool broke down the customer group into four clear spend groups. The top group has few, but very big spenders who give a lot of the money.
- **Diagnostic (Why?):** This break-up shows a big split between big, key partners (Group 1) and smaller, less regular buyers (Groups 3 and 4). This kind of split is normal in business-to-business dealings.
- **Prescriptive (What next?):**
  - **Set Growth Targets:** Establish a key performance indicator (KPI) to move a certain percentage of customers from a lower quartile to a higher one each quarter.
  - **Tiered Marketing Strategy:** Implement a formal tiered engagement model.
    - Quartile 1 ("VIPs"): Offer loyalty rewards, early access to new products, and personalized service.
    - Quartile 2 ("Growth Accounts"): Target with up-sell and cross-sell campaigns to move them into the top tier.

- Quartiles 3 & 4 ("Emerging Accounts"): Nurture with automated email marketing and standardized promotions to increase engagement efficiently.

7. References

- Oracle Corporation. (2023). Oracle Database SQL Language Reference, 23c. Retrieved from https://docs.oracle.com/en/database/oracle/oracle-database/23/sqlrf/
- https://www.youtube.com/watch?v=Ww71knvhQ-s
- https://www.youtube.com/watch?v=rIcB4zMYMas
- Celko, J. (2012). Joe Celko's SQL for Smarties: Advanced SQL Programming. Morgan Kaufmann.
- Itzik, B. (2020). T-SQL Window Functions: For data analysis and beyond. Itzik Ben-Gan.
- Sharma, N., et al. (2010). PL/SQL for Developers: A Programmer's Guide to Oracle PL/SQL. O'Reilly Media.
- https://www.youtube.com/watch?v=J5wjIf4gdq4&list=PLWPirh4EWFpHj0kKqqYudPNOrGEa0fFIr
- Molinaro, A. (2006). SQL Cookbook: Query Solutions and Techniques for All SQL Users. O'Reilly Media.
- "Window Functions in SQL." GeeksforGeeks, 2023, https://www.geeksforgeeks.org/window-functions-in-sql/.
- https://www.youtube.com/watch?v=xofpqdU3cD4
- Date, C. J. (2004). An Introduction to Database Systems (8th ed.). Addison-Wesley.
- "LAG Function in Oracle." Oracle Tutorial, https://www.oracletutorial.com/oracle-analytic-functions/oracle-lag/
- Feuerstein, S., & Pribyl, B. (2014). Oracle PL/SQL Programming (6th ed.). O'Reilly Media.