Keiffer Tan
AD450 – Data Science

Data Analysis for Global Coffee Company

Identifying Top Exporter (1990-1994):

Scenario: The company is looking to understand historical trends in coffee exports.

Task: Determine which country exported the most coffee cumulatively from 1990 to 1994.

Expected Output: A report stating the top exporting country along with the total export volume for the specified period.

```
-- Identifying Top Exporter (1990-1994) --
WITH total_exports AS (
    SELECT country,
    SUM("1990"+"1991"+"1992"+"1993"+"1994") AS total_export
    FROM coffee_export
    GROUP BY country
)
SELECT country,
    total_export,
    DENSE_RANK() OVER(
        ORDER BY total_export DESC
    ) AS export_rank
FROM
    total_exports;
```

	country text	total_export numeric	export_rank bigint
1	Brazil	5521260000	1
2	Colombia	4106580000	2
3	Indonesia	1676700000	3
4	Côte d'Ivoire	1146660000	4
5	Mexico	995940000	5
6	Guatemala	985500000	6
7	El Salvador	709080000	7
8	Uganda	696420000	8

- From this query, we found that Brazil, Colombia, and Indonesia were the top exporters between the years of 1990-1994.

Keiffer Tan
AD450 – Data Science

Top Coffee Producers (1999-2004):

Scenario: To plan future coffee sourcing strategies, the company needs to identify key coffee-producing countries.

Task: Find the five countries with the most coffee production from 1999 to 2004.

Expected Output: A table listing the countries, their rank, and their combined production from 1999 to 2004.

```sql
WITH total_producers AS (
    SELECT country,
        coffee_type,
        SUM("1999_2000"+"2000_2001"+"2001_2002"+"2002_2003"+"2003_2004") AS coffee_produced
    FROM coffee_production
    GROUP BY country, coffee_type
)
SELECT country,
    coffee_type,
    DENSE_RANK() OVER(
        ORDER BY coffee_produced DESC
    ) AS producer_rank,
    coffee_produced
FROM total_producers
LIMIT 5;
```

	country text	coffee_type text	producer_rank bigint	coffee_produced numeric
1	Brazil	Arabica/Robusta	1	11248680000
2	Viet Nam	Robusta/Arabica	2	3988560000
3	Colombia	Arabica	3	3284760000
4	Indonesia	Robusta/Arabica	4	2010720000
5	Mexico	Arabica/Robusta	5	1441440000

- From this query, the countries with the most coffee produced is Brazil, Vietnam, Colombia, Indonesia, and Mexico in 1999-2004. It seems that Brazil out produced every other country by a large margin during that period.

Keiffer Tan
AD450 – Data Science

Second Highest Coffee Production by Type (1990-1994):

Scenario: Understanding the diversity in coffee production is crucial for market analysis.

Task: Identify countries with the second highest coffee production for each coffee type from 1990-1994. Note that there can be multiple countries tied for second place.

Expected Output: A list of coffee types along with the corresponding countries that ranked second in production for each type.

```sql
WITH total_producers AS (
    SELECT country,
        coffee_type,
        SUM("1990_1991"+"1991_1992"+"1992_1993"+"1993_1994") AS coffee_produced
    FROM coffee_production
    GROUP BY country, coffee_type
),
ranked_producers AS (
    SELECT country,
        coffee_type,
        DENSE_RANK() OVER(
            PARTITION BY coffee_type
            ORDER BY coffee_produced DESC
        ) AS producer_rank,
        coffee_produced
    FROM total_producers
)
SELECT country,
    coffee_type,
    producer_rank,
    coffee_produced
FROM ranked_producers
WHERE producer_rank = 2;
```

	country text	coffee_type text	producer_rank bigint	coffee_produced numeric
1	Costa Rica	Arabica	2	640740000
2	Mexico	Arabica/Robusta	2	1069500000
3	Madagascar	Robusta	2	208740000
4	India	Robusta/Arabica	2	731100000

- The second highest producers of each coffee type between 1990-1994 are Costa Rica, Mexico, Madagascar, and India

Keiffer Tan
AD450 – Data Science

Top Five Countries in Combined Exports and Imports (1995-2000):

Scenario: The company is assessing its global trade footprint.

Task: Name the top five countries with the most combined coffee export and import volume between 1995 to 2000.

Expected Output: A CSV file containing columns for country, export volume, import volume, and total combined volume for the specified period.

```sql
-- USING CTE METHOD --
WITH total_imp AS(
    SELECT country,
    SUM("1995"+"1996"+"1997"+"1998"+"1999"+"2000") AS total_import_1995_2000
    FROM coffee_import
    GROUP BY country
),
total_exp AS (
    SELECT country,
        SUM("1995"+"1996"+"1997"+"1998"+"1999"+"2000") AS total_export_1995_2000
    FROM coffee_export
    GROUP BY country
)
SELECT COALESCE(total_imp.country,total_exp.country) AS country,
    COALESCE(total_imp.total_import_1995_2000,0) AS total_import,
    COALESCE(total_exp.total_export_1995_2000,0) AS total_export,
    COALESCE(total_exp.total_export_1995_2000,0) + COALESCE(total_imp.total_import_1995_2000,0)AS total_combined
FROM total_imp
FULL JOIN total_exp
    ON total_imp.country = total_exp.country
ORDER BY total_combined DESC
LIMIT 5;
```

```sql
-- USING UNION METHOD --
SELECT country,
    SUM(total_import_1995_2000) AS total_import,
    SUM(total_export_1995_2000) AS total_export,
    SUM(total_import_1995_2000 + total_export_1995_2000) AS total_combined
FROM (
    SELECT country,
        SUM("1995"+"1996"+"1997"+"1998"+"1999"+"2000") AS total_import_1995_2000,
        0 AS total_export_1995_2000
    FROM coffee_import
    GROUP BY country
    UNION ALL
    SELECT country,
        '0' AS total_import_1995_2000,
        SUM("1995"+"1996"+"1997"+"1998"+"1999"+"2000") AS total_export_1995_2000
    FROM coffee_export
    GROUP BY country
) as combined_data
GROUP BY country
ORDER BY total_combined DESC;
```

	country text	total_import numeric	total_export numeric	total_combined numeric
1	United States of America	7466280000	0	7466280000
2	Brazil	0	6349740000	6349740000
3	Germany	4933140000	0	4933140000
4	Colombia	0	3705180000	3705180000
5	France	2361000000	0	2361000000

- It seems during this time, companies that export coffee do not import coffee, and it also is true the other way around as well.

Keiffer Tan
AD450 – Data Science

Import Analysis with Country of Origin:

Scenario: For supply chain optimization, the company seeks detailed import data.

Task: Attempt to show imports for every country, including the country of origin for those imports, ranked by import volume.

Expected Output: Explain why this data might not be retrievable from the current database structure. Suggest alternative methods or additional data that could be used to fulfill this request.

With the information provided, the only query I will be able to show is the imports for every country and ranking which country imports the most. The tables and data within the database only contain one-dimensional information such as how much is imported, exported, and produced. To be able to show the imports of each country along with the origin of the imports, there must be a table that breaks down the exports of each country to the export destination. With that information, I can retrieve queries that show the countries that export to a specific destination. From that information, I can find the amount of coffee a country imports from specific exporting countries.