**SQL Data Exploration: Famous Paintings**

I selected the famous paintings dataset available on Kaggle, comprising seven tables: Artist, Canvas\_Size, Museum, Museum\_Hours, Price\_Size, Subject, and Work. My choice of this dataset stems from a personal interest in art and paintings. Leveraging SQL queries, I delved into the dataset to gain insights by answering specific questions related to the world of art.

**Data Source**: https://www.kaggle.com/datasets/mexwell/famous-paintings

* **List top 10 artists who created highest number of famous paintings?**

1. Tables involved: The query involves two tables: **work** and **artist**
2. Join Condition: The LEFT JOIN connects the work table to the artist table using the artist\_id as the linking key.
3. Selecting and Grouping: The SELECT statement retrieves the full name of the artists and the count of paintings. The result set is grouped by the full name of the artists using the GROUP BY clause.
4. Ordering: The result set is ordered in descending order based on the count of paintings using ORDER BY clause.
5. Limiting results: The LIMIT 10 clause ensures that only the top 10 artists with the highest count of paintings are included in the result.

* **What was the most used style and subject by the top 10 artists?**

1. Tables Involved: The query involves two tables, work and artist
2. Join Condition: A LEFT JOIN is used to connect the work table to the artist table using the artist\_id as the linking key. Subject table is also joined to the work table using work\_id as the linking key.
3. Selection and Grouping: The SELECT statement retrieves the full name of the artists (a.full\_name) and the count of paintings (COUNT(w.work\_id)). The result set is then grouped by the full name of the artists using the GROUP BY clause.
4. Ordering: The result set is ordered in descending order based on the count of paintings using the ORDER BY clause.
5. Limiting Results: The LIMIT 10 clause ensures that only the top 10 artists with the highest count of paintings are included in the result.

* **What % of the artists' paintings were of their favourite subject?**

1. Tables Involved: The query involves three tables: artist, work, and subject
2. Join Conditions: The JOIN clauses connect the artist table to the work table using the artist\_id as the linking key and the work table to the subject table using the work\_id as the linking key.
3. Selection and Grouping: The SELECT statement retrieves the artist's full name, the predominant painting style, the predominant subject, and the count of paintings. The result set is grouped by the artist's full name, style, and subject.
4. Ordering: The result set is ordered in descending order based on the count of paintings using the ORDER BY clause.
5. Limiting Results: The LIMIT 10 clause ensures that only the top 10 artists with the highest count of paintings are included in the result.

* **What is the subject of most expensive painting (regular price)?**

The query aims to find subjects and their corresponding regular prices for paintings where the regular price matches the maximum regular price in the price\_size table.

1. Tables Involved: The query involves two tables: subject and price\_size.
2. Join Condition: A LEFT JOIN is used to connect the subject table to the price\_size table using the work\_id as the linking key.
3. Selection: The SELECT statement retrieves the subject and the regular price from the price\_size table.
4. Filtering: The WHERE clause filters the results to include only those rows where the regular price is equal to the maximum regular price in the entire price\_size table.

* **Which artist created the most expensive collection of paintings?**

The query provides a list of the top 10 artists, their IDs, full names, and the total cost of their collections, sorted in descending order of the total cost.

1. Tables Involved: The query involves three tables: artist, work, and price\_size.
2. Join Conditions: The JOIN clauses connect the artist table to the work table using the artist\_id as the linking key and the work table to the price\_size table using the work\_id as the linking key.
3. Selection and Aggregation: The SELECT statement retrieves the artist ID, full name, and the sum of regular prices for each artist. The results are aggregated using the GROUP BY clause, grouping by both the full name and artist ID.
4. Ordering: The result set is ordered in descending order based on the sum of regular prices using the ORDER BY clause.
5. Limiting Results: The LIMIT 10 clause ensures that only the top 10 artists with the highest total cost of collections are included in the result.

* Which artist had the highest average cost of their paintings?

The query provides a ranked list of the top 10 artists, along with their artist IDs, full names, and the average cost of their paintings, sorted in descending order of the average cost.

1. Tables Involved: The query involves three tables – artist, work, and price\_size
2. Join Conditions: The JOIN clauses connect the artist table to the work table using the artist\_id as the linking key and the work table to the price\_size table using the work\_id as the linking key.
3. Selection and Aggregation: The SELECT statement retrieves the artist ID, full name, and the average regular price for each artist. The results are aggregated using the GROUP BY clause, grouping by both the full name and artist ID.
4. Ordering: The result set is ordered in descending order based on the average regular price using the ORDER BY clause.
5. Limiting Results: The LIMIT 10 clause ensures that only the top 10 artists with the highest average cost of paintings are included in the result.

* **Which style of paintings is the most expensive?**

The query provides the painting style with the highest average regular price.

1. Tables Involved: The query involves two tables - work and price\_size.
2. Join Condition: The JOIN clause connects the work table to the price\_size table using the work\_id as the linking key.
3. Selection and Aggregation: The SELECT statement retrieves the painting style and the average regular price (avg(regular\_price\_$k)) for each style. The results are aggregated using the GROUP BY clause, grouping by the painting style.
4. Ordering: The result set is ordered in descending order based on the average regular price using the ORDER BY clause.
5. Limiting Results: The LIMIT 1 clause ensures that only the style with the highest average regular price is included in the result.

* **Is there a favoured style of paintings displayed by each museum?**

The query retrieves information about museums, painting styles, and the number of paintings, considering only the top 5 styles (by the number of paintings) for each museum where the number of paintings for that style is greater than 5.

1. Tables Involved: The query involves two tables - work and museum
2. Join Condition: The LEFT JOIN clause connects the work table to the museum table using the museum\_id as the linking key.
3. Subquery and Aggregation: The subquery (aliased as tbl1) selects the museum name (m.name), painting style, and the count of paintings for each style. The ROW\_NUMBER() window function is used to assign a ranking to each style within each museum based on the count of paintings, ordered in descending order.
4. Filtering: The outer query filters the results to include only rows where the row\_num is less than or equal to 5 (considering the top 5 styles) and where the num\_of\_paintings is greater than 5**.**

* **Which museum has the greatest number of Claude Monet paintings?**

The query provides the name of the museum with the highest count of paintings by the artist Claude Monet.

1. Tables Involved: The query involves two tables - work and museum.
2. Join Condition: The LEFT JOIN clause connects the work table to the museum table using the museum\_id as the linking key.
3. Filtering: The WHERE clause filters the results to include only rows where the artist\_id matches the artist\_id associated with the artist Claude Monet (retrieved from the artist table).
4. Aggregation: The SELECT statement retrieves the museum name and the count of paintings for each museum. The results are aggregated using the GROUP BY clause, grouping by the museum.
5. Ordering and Limiting Results: The result set is ordered in descending order based on the count of paintings, and the LIMIT 1 clause ensures that only the museum with the highest count of Claude Monet's paintings is included in the result.

* **What time will 'The Metropolitan Museum of Art' open on 2024-02-05?**

1. Table Involved: The query involves the museum\_hours table.
2. Filtering: The WHERE clause filters the results to include only rows where the day matches the day name of the specified date ('2024-02-05') and where the museum\_id matches the museum\_id associated with the museum named 'The Metropolitan Museum of Art' (retrieved from the museum table).
3. Selection: The SELECT statement retrieves the day and open columns for the specified conditions.