Art Beyond Frames

A Data Analysis of Museums and Masterpieces

Discovering Patterns in Paintings, Artists, and Museums with MySQL

For more info:
GitHub Link for complete Project

Table of contents

For more info:
GitHub Link for complete Project

Project Objectives and Overview :

❖ **Analyze Museum and Painting Data:**
- Understand the relationship between museums, paintings, and their associated data, such as artists, styles, and locations.

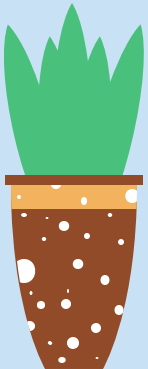
❖ **Identify Key Insights:**
- Extract meaningful insights on popular paintings, artists, and museums using MySQL queries.

❖ **Detect and Address Data Irregularities:**
- Identify invalid or duplicate data entries and clean the dataset for accurate analysis.

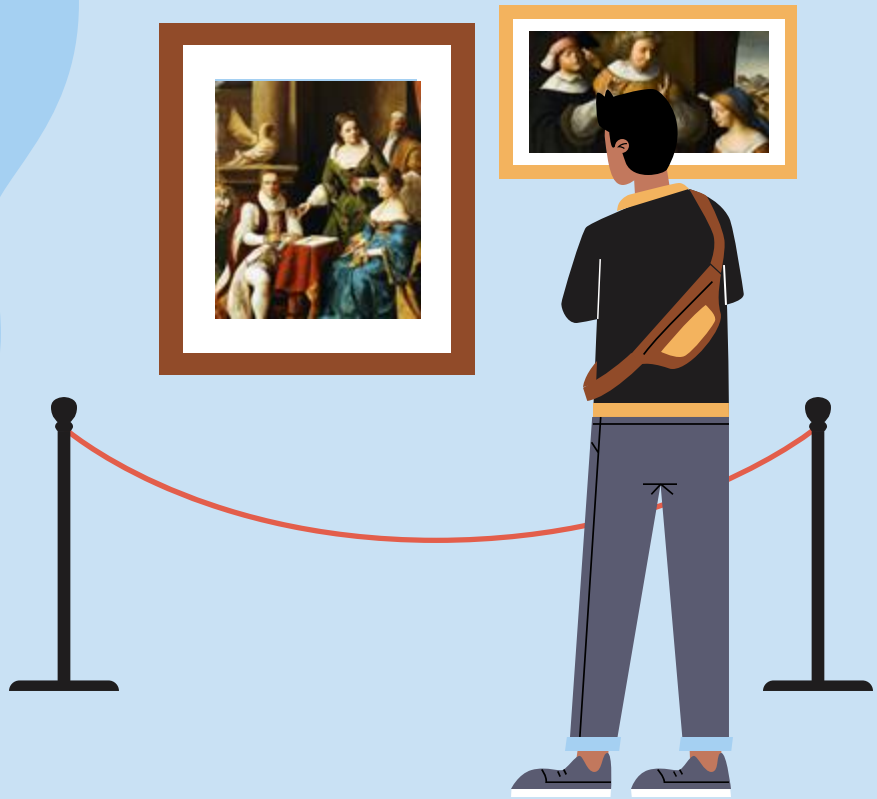
❖ **Discover Trends and Patterns:**
- Explore pricing trends, popular canvas sizes, and painting styles to uncover what drives demand and recognition in the art world.

❖ **Optimize Museum Operations:**
- Examine museum operations, including opening hours and city-wise distribution, to propose strategies for improving visitor engagement and efficiency.

The ultimate goal is to derive actionable insights for museum management and art strategies through data-driven decisions.



Dataset Exploration and Preparation :



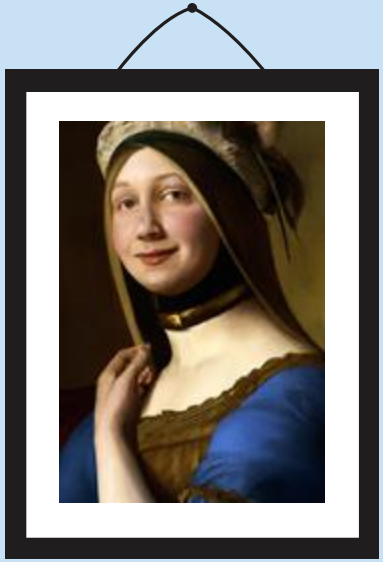
The dataset provides structured information on famous paintings, artists, museums, and related attributes. It is valuable for analyzing art history, artistic styles, museum collections, and pricing trends.

Key Tables & Columns

- Artist – Contains details about artists, including their full name, nationality, style, and lifespan.
- Work – Stores information on artworks, linking them to artists and museums.
- Subject – Lists subjects/themes depicted in each artwork.
- Image Link – Provides URLs for artwork images and thumbnails.
- Product Size – Includes pricing details for different artwork sizes.
- Canvas Size – Defines artwork dimensions.
- Museum – Contains museum details such as location, contact information, and website links.
- Museum Hours – Specifies opening and closing hours for museums.
- Dataset Source:
  This dataset is available on Kaggle at:

  Famous Paintings & Museums Dataset

Dataset Preparation & Ingestion Using Python:

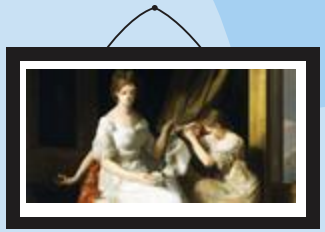


All the tables from the dataset have been initially ingested into a MySQL Workbench database using Python script below. The script automates the process of creating tables and inserting data efficiently.

```python
import pandas as pd
from sqlalchemy import create_engine
from urllib.parse import quote_plus

password = quote_plus('Hritik!7020603977')

# MySQL connection string
conn_string = 'mysql+pymysql://root:' + password + '@localhost/paintings'
db = create_engine(conn_string)
conn = db.connect()

# List of CSV filenames (without file extensions) that you want to insert into MySQL
files = ['artist', 'canvas_size', 'image_link', 'museum_hours', 'museum', 'product_size', 'subject', 'work']

# Path where the CSV files are stored
file_path_base = r'C:\Users\Hritik Kadam\Desktop\SQL\Famous Painting Case Study\\'

for file in files:
    # Construct the full file path
    file_path = file_path_base + file + '.csv'

    # Read the CSV file into a pandas DataFrame
    try:
        df = pd.read_csv(file_path)
        print(f"Loaded data from {file}.csv successfully.")

        # Insert data into MySQL (replace table if it exists)
        df.to_sql(file, con=conn, if_exists='replace', index=False)
        print(f"Data from {file}.csv inserted into MySQL table {file}.")

    except FileNotFoundError:
        print(f"Error: The file {file}.csv was not found.")
    except Exception as e:
        print(f"Error processing {file}.csv: {e}")

# Close the database connection
conn.close()
print("Database connection closed.")
```

Data Analysis and Insights :

Museums Open on Both Sunday and Monday



1. **Identify the museums which are open on both Sunday and Monday.**

Query:

```
SELECT
    DISTINCT m.name, m.city, mh1.day
FROM
    museum_hours mh1
        JOIN
    museum_hours mh2 ON mh1.museum_id = mh2.museum_id
        JOIN
    museum m ON m.museum_id = mh1.museum_id
WHERE
    mh1.day = 'Sunday' AND mh2.day = 'Monday';
```

Output:

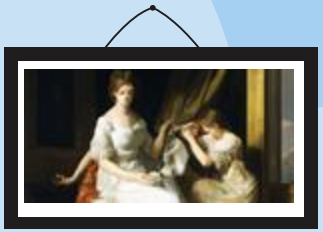
Result Grid | Filter Rows: | Export:

name	city	day
The Museum of Modern Art	New York	Sunday
Pushkin State Museum of Fine Arts	Moscow	Sunday
National Gallery of Victoria	Melbourne	Sunday
The Metropolitan Museum of Art	New York	Sunday
Museum of Grenoble	38000	Sunday
Nelson-Atkins Museum of Art	Kansas City	Sunday
Musée du Louvre	75001	Sunday
National Maritime Museum	London	Sunday
Museum of Fine Arts Boston	Boston	Sunday
Rijksmuseum	Amsterdam	Sunday

Result 2

Data Analysis and Insights :

Delete Duplicates While Keeping One Record



2. Delete duplicate records from work table and keep distinct record for each ID.

```sql
WITH ranked_rows AS (
    SELECT *,
           ROW_NUMBER() OVER (PARTITION BY work_id) AS row_num
    FROM work
)
DELETE FROM work
WHERE work_id IN (
    SELECT work_id
    FROM ranked_rows
    WHERE row_num > 1
);
```

: Query

Output:

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch

work_id	name	artist_id	style	museum_id
160228	Still Life with Flowers and a Watch	615	Baroque	43
160236	Still Life with Fruit and a Beaker on a Cock's Foot	615	Baroque	43
160244	Still Life with Fruit and a Goldfinch	615	Baroque	43
160252	Still Life with Fruit and Oysters	615	Baroque	43
160260	Still Life with Fruit, Oysters, and a Porcelain Bowl	615	Baroque	43
160268	The Overturned Bouquet	615	Baroque	43
125752	Arabian Horses at Pasture	757	Baroque	NULL
125818	Count Halm on His Basedow Estate	757	Baroque	NULL
23448	Horses at the Porch	757	Baroque	34
125763	Napoleon Before the Burning City of Smolensk	757	Baroque	NULL

Data Analysis and Insights :

Top 10 Most Famous Painting Subjects

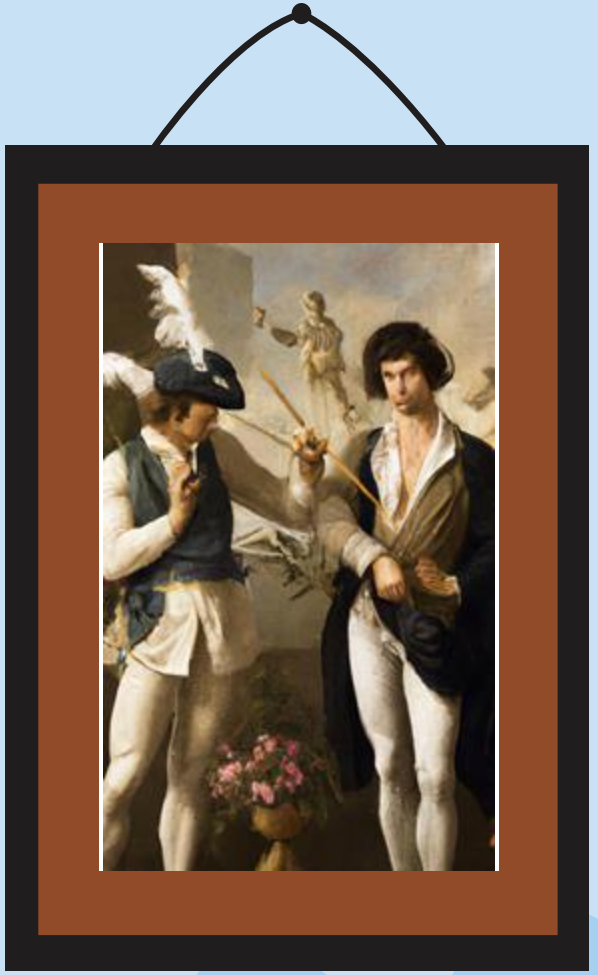
3. Fetch the top 10 most famous painting subject

Query:

```
with cte as (
    select s.subject, count(1) as no_of_subjects
    , rank() over(order by count(1) desc) as rn
    FROM work w
    join subject s on w.work_id=s.work_id
    group by 1
    order by 2 desc
)
select subject, no_of_subjects
from cte
where rn<=10;
```

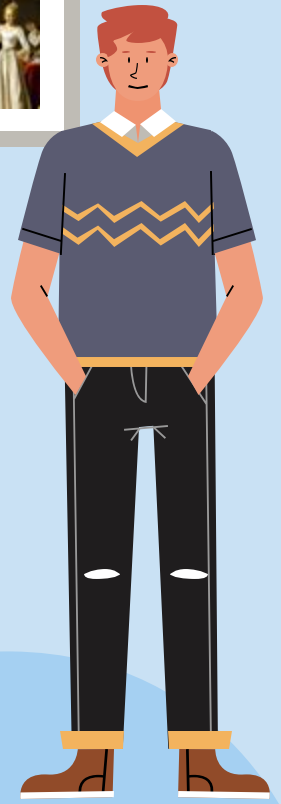
Output:

subject	no_of_subjects
Portraits	986
Nude	499
Landscape Art	392
Flowers	364
Abstract/Modern Art	294
Still-Life	290
Rivers/Lakes	290
Horses	250
Seascapes	250
Marine Art/Maritime	205
Gardens	205



Data Analysis and Insights :

Top 5 Most Prolific Artists



4. Who are the top 5 most popular artist? (Popularity is defined based on most no of paintings done by an artist)

Query:

```
SELECT
    full_name,
    count_of_paintings
FROM (
    SELECT
        a.full_name,
        COUNT(w.name) AS count_of_paintings,
        RANK() OVER (ORDER BY COUNT(w.name) DESC) AS rn
    FROM
        artist a
    JOIN
        work w ON a.artist_id = w.artist_id
    GROUP BY
        a.full_name
) x
WHERE
    rn <= 5;
```

Output:

full_name	count_of_paintings
Pierre-Auguste Renoir	458
Claude Monet	340
Vincent Van Gogh	292
Maurice Utrillo	253
Albert Marquet	226

Data Analysis and Insights :

Artists with Global Presence

5. Identify the artists whose paintings are displayed in multiple countries

Query:

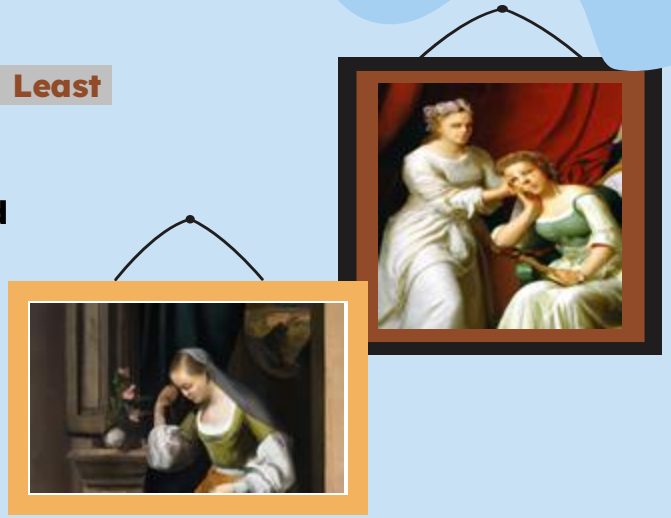
```
WITH cte AS (
    SELECT DISTINCT
        a.full_name AS artist,
        m.country
    FROM
        work w
    JOIN
        artist a
        ON a.artist_id = w.artist_id
    JOIN
        museum m
        ON m.museum_id = w.museum_id
)
SELECT
    artist,
    COUNT(DISTINCT country) AS no_of_countries
FROM
    cte
GROUP BY
    artist
Having
    COUNT(DISTINCT country) >1
ORDER BY
    2 desc;
```

Output:

artist	no_of_countries
Vincent Van Gogh	8
Claude Monet	7
Paul Gauguin	7
Francois Boucher	6
Pierre-Auguste Renoir	6
Rembrandt Van Rijn	6
Alfred Sisley	5
André Derain	5
Camille Pissarro	5
Edgar Degas	5
Édouard Vuillard	5
El Greco	5

Data Analysis and Insights :

Artists & Museums Housing the Most and Least Expensive Paintings



6. Identify the artist and the museum where the most expensive and least expensive painting is placed.

Query:

```
WITH cte AS (
    SELECT
        a.full_name AS artist,
        ps.sale_price,
        w.name AS painting,
        m.name AS museum,
        m.city AS city,
        RANK() OVER (ORDER BY sale_price DESC) AS rnk,
        RANK() OVER (ORDER BY sale_price ASC) AS rnk_asc
    FROM
        product_size ps
    JOIN
        work w ON w.work_id = ps.work_id
    JOIN
        museum m ON m.museum_id = w.museum_id
    JOIN
        artist a ON a.artist_id = w.artist_id
)
SELECT
    artist,
    sale_price,
    painting,
    museum,
    city,
    CASE
        WHEN cte.rnk = 1 THEN 'most_expensive'
        WHEN cte.rnk_asc = 1 THEN 'least_expensive'
    END AS expensive_level
FROM
    cte
WHERE
    cte.rnk = 1
    OR cte.rnk_asc = 1;
```

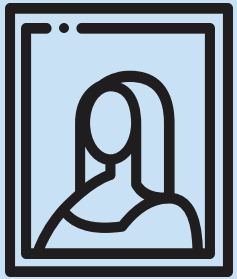
Output:

artist	sale_price	painting	museum	city	expensive_level
John Singleton Copley	20	Portrait of Mr. and Mrs. Thomas Mifflin (Sarah M...	Philadelphia Museum of Art	Philadelphia	least_expensive
John Singleton Copley	20	Portrait of Mr. and Mrs. Thomas Mifflin (Sarah M...	Philadelphia Museum of Art	Philadelphia	least_expensive
Peter Paul Rubens	1115	Fortuna	The Prado Museum	Madrid	most_expensive

Data Analysis and Insights :

Top 3 Most Popular & Least Popular Painting Styles

7. Which are the 3 most popular and 3 least popular painting styles?



Query:

```
WITH cte AS (
    SELECT
        style,
        COUNT(1) AS cnt_of_paintings,
        RANK() OVER (ORDER BY COUNT(1) ASC) AS asc_rnk,
        RANK() OVER (ORDER BY COUNT(1) DESC) AS desc_rnk
    FROM
        work
    GROUP BY
        style
)
SELECT
    style,
    cnt_of_paintings,
    CASE
        WHEN cte.desc_rnk <= 3 THEN 'most_popular'
        WHEN cte.asc_rnk <= 3 THEN 'least_popular'
    END AS popularity
FROM
    cte
WHERE
    asc_rnk <= 3
    OR desc_rnk <= 3;
```

Output:

style	cnt_of_paintings	popularity
Impressionism	2946	most_popular
Post-Impressionism	1616	most_popular
Realism	1140	most_popular
Avant-Garde	141	least_popular
Art Nouveau	106	least_popular
Japanese Art	69	least_popular

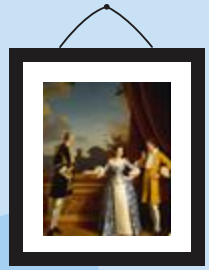
Data Analysis and Insights :

Museum with Longest Daily Hours

8. Which museum is open for the longest during a day?

: Query

```sql
SELECT
    museum_name,
    state,
    day,
    duration
FROM (
    SELECT
        m.name AS museum_name,
        m.state,
        day,
        STR_TO_DATE(open, '%h:%i %p') AS open_time,
        STR_TO_DATE(close, '%h:%i %p') AS close_time,
        TIMEDIFF(
            STR_TO_DATE(close, '%h:%i %p'),
            STR_TO_DATE(open, '%h:%i %p')
        ) AS duration,
        RANK() OVER (
            ORDER BY TIMEDIFF(
                STR_TO_DATE(close, '%h:%i %p'),
                STR_TO_DATE(open, '%h:%i %p')
            ) DESC
        ) AS rnk
    FROM
        museum_hours AS mh
    JOIN
        museum m
    ON
        m.museum_id = mh.museum_id
) x
WHERE
    x.rnk = 1;
```

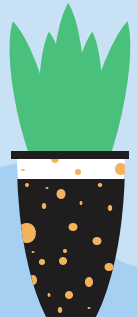


Output:

	museum_name	state	day	duration
▶	Musée du Louvre	Paris	Friday	12:45:00

Conclusions and Recommendations :

- Some paintings are not showcased in any museums, presenting an opportunity for new exhibitions or partnerships.
- A few museums lack paintings in their collections. These institutions should be reviewed for more diverse acquisitions or collaborations.
- Several paintings are listed with an asking price higher than the regular price, which could be part of a strategic pricing model. Conversely, some paintings are priced below 50% of their regular price, suggesting potential undervaluation or discount strategies.
- The most expensive canvas sizes should be highlighted in premium exhibitions, while less popular sizes could be used to create unique or niche collections.
- Some museums have extended hours, especially those open every day, providing enhanced visitor access. Promoting these museums can attract more visitors, especially those with non-traditional schedules.
- The most popular museums, based on the number of paintings, and artists, based on their total works, should be emphasized in global exhibitions to capitalize on their recognition and audience appeal.
- Museums with missing or invalid city information, as well as incorrect museum hours, need updating to maintain data accuracy and proper operations.
- Identifying the most and least popular painting styles can guide curators to focus on trending styles for future exhibitions while exploring niche markets for unique collections.
- Artists with paintings displayed in multiple countries should be promoted for international exhibits to expand their global presence.
- Identifying the most and least expensive paintings and comparing them across museums allows institutions to optimize their pricing strategies for maximum revenue.



Thanks!

If you have any questions?

hritikkadam74@gmail.com

GitHub Link for complete Project