# famous_painting Project

## table of contents



## Project Overview

This project analyses a museum's collection of paintings, focusing on details like display status, museum operations, and painting popularity. It involves querying data to find missing or incorrect information, top artworks, and museum activity patterns. The goal is to gain insights into the museum’s collection, operations, and visitor engagement.

### Data Sources

famous_painting Data: The primary dataset used in the analysis is a "artist.csv","canavs_size", "image_link.csv", "museum.csv"," museum_hours.csv", "product_size.csv", "subject.csv", and "work.csv", all files contain detailed information about all Museums, paintings, and hours of work.

### Tools

- MySQL: Cleansing process

-[download here](http://micrsoft.com)

- MySQL: Make the analysis project

### Data Cleaning Preparation

We performed the following tasks:

1- data loading and inception

2-Check the quality of the data for each table.

## Data Analysis

includes requirements and SQL syntax for each question.

## Questions:

```
use famous_painting;
1.Assign all the paintings that are not displayed in any museum ?
```sql

select work_id, museum_id   from work
where museum_id is null;
```

## 2. Are there museums without any paintings?

```sql

select work_id, m.museum_id
from work w
join museum m on w.museum_id = m.museum_id
where w.work_id is null;
----2
select  count(distinct m.museum_id)museun_no_paintings
from work w
join museum m on w.museum_id = m.museum_id

where w.work_id is null;

```

## 3. How many paintings have an asking price of more than their regular price?

```sql

select count(distinct work_id) Num_paintings
from product_size
where sale_price > regular_price;

```

## 4. Identify the paintings whose asking price is less than 50% of its regular price?

```sql

select *
from product_size
where sale_price <(regular_price*0.5)
```
```

| work_id | size_id | sale_price | regular_price |
|---|---|---|---|
| 31780 | 36 | 10 | 125 |
| 31780 | 30 | 10 | 95 |
| 31780 | 36 | 10 | 125 |
| 31780 | 30 | 10 | 95 |
| 198417 | 36 | 30 | 125 |
| 198417 | 30 | 30 | 95 |
| 31974 | 24 | 30 | 85 |
| 17351 | 24 | 10 | 85 |
| 17351 | 30 | 10 | 95 |
| 17351 | 36 | 10 | 125 |
| 30947 | 3024 | 285 | 575 |

5. Delete duplicate records from work, product_size, subject and image_link tables.

```sql
--Delete from work
with workcte as(
select *,ROW_NUMBER() over (partition by
name,artist_id,style,museum_id order by name)rnk
from work)
delete from workcte where rnk>1;
select * from work

----Delete from product_size
with product_sizeCte as(
select * ,ROW_NUMBER() over (partition by
size_id,sale_price,regular_price order by size_id)rnk
from product_size)
delete from product_sizeCte where rnk > 1;
select * from product_size

----delete from subject
with subjectCte as(
select *,ROW_NUMBER() over (partition by work_id,subject order by
work_id)rnk
from subject)
delete from subjectCte where rnk>1
select * from subject

-----delet Duplicate from image_link
with image_linkCte as(
select * , ROW_NUMBER() over (partition by
url,thumbnail_small_url,thumbnail_large_url order by url) as rnk
 from image_link)
 delete from image_linkCte where rnk > 1 ;
 select * from image_link
```

```

6. Identify the museums with invalid city information in the given dataset

```sql
select *
from museum
where city not like '%[^0-9]%'
```

7. Fetch the top 10 most famous painting subject?

```sql
with t1 as(
     select s.subject ,COUNT(*)num_repited_subject
     ,RANK() over( order by COUNT(*)desc ) rnk
     from work w
     join subject s
     on w.work_id = s.work_id
     group by s.subject)
     select * from t1
     where rnk<=10
```

8. Identify the museums which are open on both Sunday and Monday. Display

museum name, city.

```sql
select m.name,mh.museum_id,mh.day,
     from museum m
     join museum_hours mh on m.museum_id = mh.museum_id
     where mh.day in('Sunday' ,'Monday')
```

9. How many museums are open every single day?

```sql

```sql
with t1 as (
      select m.name,mh.museum_id,
          COUNT(day) over (partition by m.name,mh.museum_id order by
m.name )num_days
           from museum m
      join museum_hours mh on m.museum_id = mh.museum_id
      group by m.name,mh.museum_id,mh.day)
       select * from t1
        where num_days = 7
```

10. Which are the top 5 most popular museums? (Popularity is defined based on most number of paintings in a museum).

```sql
with t1 as(
         select m.museum_id,m.name,m.city,m.country,
         COUNT(1) no_of_painting
         from work w
         join museum m on m.museum_id = w.museum_id
         group by m.museum_id,m.name,m.city,m.country ),
        t2 as( select * , RANK() over( order by no_of_painting desc) rnk
         from t1)
        select t2.*
        from t2
        where rnk <5
        ==========
      with t1 as(
      select m.museum_id, count(1) as no_of_painintgs
                  , rank() over(order by count(1) desc) as rnk
                  from work w
                  join museum m on m.museum_id=w.museum_id
                  group by m.museum_id)
      select t1.* ,m.city,m.country
      from t1
      join museum m  on m.museum_id = t1.museum_id
      where t1.rnk <=5
```

11. Who are the top 5 most popular artists? (Popularity is defined based on most number of paintings done by an artist).

```sql
with t1 as(
      select a.artist_id ,count(1) no_painting_byartist,
      RANK() over(order by count(1)  desc )rnk
```

```
      from work w
      join artist a on a.artist_id = w.artist_id
      group by a.artist_id)
      select t1.* ,a.full_name,a.nationality
      from t1
      join artist a  on t1.artist_id = a.artist_id
      where t1.rnk < 5
```

## 12. Display the 3 least popular canvas sizes.

```sql
with t1 as(
        select p.size_id,count(1) no_popular_size,
        rank() over ( order by count(1) )ranking
        from product_size p
        join canvas_size c on c.size_id = p.size_id
        group by p.size_id)
        select t1.*, c.label
        from t1 join
        canvas_size c on t1.size_id = c.size_id
        where t1.ranking <=3
```

## 13. Which Museum has the greatest number of the most popular painting style?

```sql
with popular_style as
                  (select style,count(*) no_famous_style
                  ,rank() over(order by count(*) desc) as style_rnk
                  from work
                  group by style),
            t1 as
                  (select m.name as museum_name,w.museum_id,ps.style,
count(*) as no_paintings
                  ,rank() over(order by count(*) desc) as painting_rnk
                  from work w
                  join museum m on m.museum_id=w.museum_id
                  join popular_style ps on ps.style = w.style
                  where ps.style_rnk=1
                  and
                  w.museum_id is not null
                  group by w.museum_id, m.name,ps.style)
      select museum_name,style,no_paintings
      from t1
```

```

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

```sql
with t1 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(*) No_country
from t1
group by artist
order by count(*) desc
```

15. Display the country and the city with the most museums.

Output 2 separate

columns to mention the city and country. If there are multiple values, separate them with a comma.

```sql
with cte_country as(
          select country,count(*) No_museum,
           RANK() over (order by count(*) desc)country_rnk
          from museum
          group by country),
cte_city as (
           select city ,count(*) No_musum,RANK() over(order by
count(*) desc) city_rnk
         from museum
          group by city)
  select  string_agg(country,' , ')as country,string_agg(city,' ,
')as city
  from cte_country
  cross join cte_city
  where cte_country.country_rnk = 1
  and cte_city.city_rnk=1;
```

16. Identify the artist and the museum where the most expensive and the least expensive painting is placed. Display the artist's name,
```

sale_price, painting name, museum name, museum city and canvas label.

```sql
with cte as(
  select *,
  RANK() over (  order by sale_price desc) max_rnk,
  RANK() over (  order by sale_price ) least_rnk
  from product_size  )
  select w.name as painting,
  a.full_name as artist,
  cte.sale_price,
  m.name as museum,
  m.city as city,cs.label as canav_size ,
  cte.max_rnk,cte.least_rnk
  from cte
  join work w on w.work_id= cte.work_id
  join artist a on a.artist_id = w.artist_id
  join museum m on m.museum_id = w.museum_id
  join canvas_size cs on cs.size_id = cte.size_id
  where cte.max_rnk =1 or cte.least_rnk=1
```

17. Which country has the 5th highest no of paintings?

```sql
with cte as(
 select m.country,count(*)no_of_paintings,
 RANK() over (order by count(*) desc) country_rnk
 from museum m
 join work w on w.museum_id = m.museum_id
 group by country)
 select country,no_of_paintings
 from cte
 where country_rnk =5
```

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

```sql
with cte as(
 select style,COUNT(*) as no_painting,
 count(*) over() as  no_of_style,
 Rank() over (order by count(*) desc) rnk
 from work
 where style is not null
```

```
 group by style )
 select style,
 case when rnk <=3 then 'most_popular' else 'leaset_popular' end as
popularity
 from cte
 where rnk <=3
 or
rnk > no_of_style -3
```

19. Which artist has the most no of Portraits paintings outside USA? Display artist name, no of paintings and the artist nationality

```sql
with cte as(
select a.full_name as artist,a.nationality, count(*) no_of_painting,
Rank() over (order by count(*) desc) as  rnk
from artist a
join work w on w.artist_id = a.artist_id
join subject s on s.work_id=w.work_id
join museum m on m.museum_id = w.museum_id
where s.subject ='Portraits'
and m.country <>'USA'
group by full_name,nationality)
select  artist,no_of_painting,nationality
from cte
where rnk =1
```