Programming for data science with python

Project 1

Nuha

Question 1:

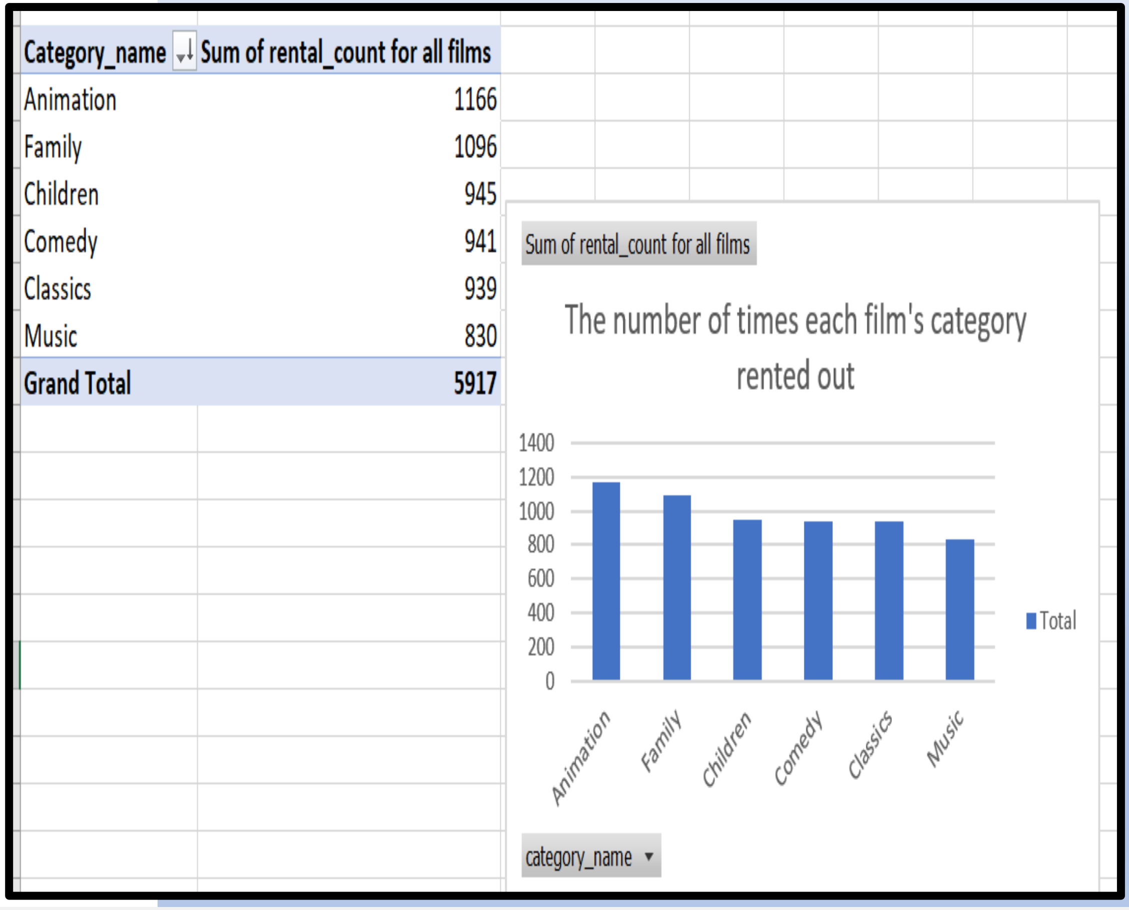
We want to understand more about the movies that families are watching. The following categories are considered family movies: Animation, Children, Classics, Comedy, Family and Music.

Create a query that lists each movie, the film category it is classified in, and the number of times it has been rented out.

```
WITH T1 AS(
   SELECT f.film_id, title film_title, name category_name
   FROM category c
   JOIN film_category fc
   ON c.category_id= fc.category_id
   JOIN film f
   ON fc.film_id = f.film_id
   WHERE name IN ('Animation', 'Children', 'Classics', 'Comedy', 'Family', 'Music'))

SELECT DISTINCT(T1.film_title), T1.category_name, count(r.rental_id) OVER (PARTITION BY
T1.film_title)  AS  rental_count
  FROM T1
  JOIN inventory i
  ON T1.film_id = i.film_id
  JOIN rental r
  ON r.inventory_id = i.inventory_id

ORDER BY category_name, film_title;
```



Category_name	Sum of rental_count for all films
Animation	1166
Family	1096
Children	945
Comedy	941
Classics	939
Music	830
Grand Total	5917

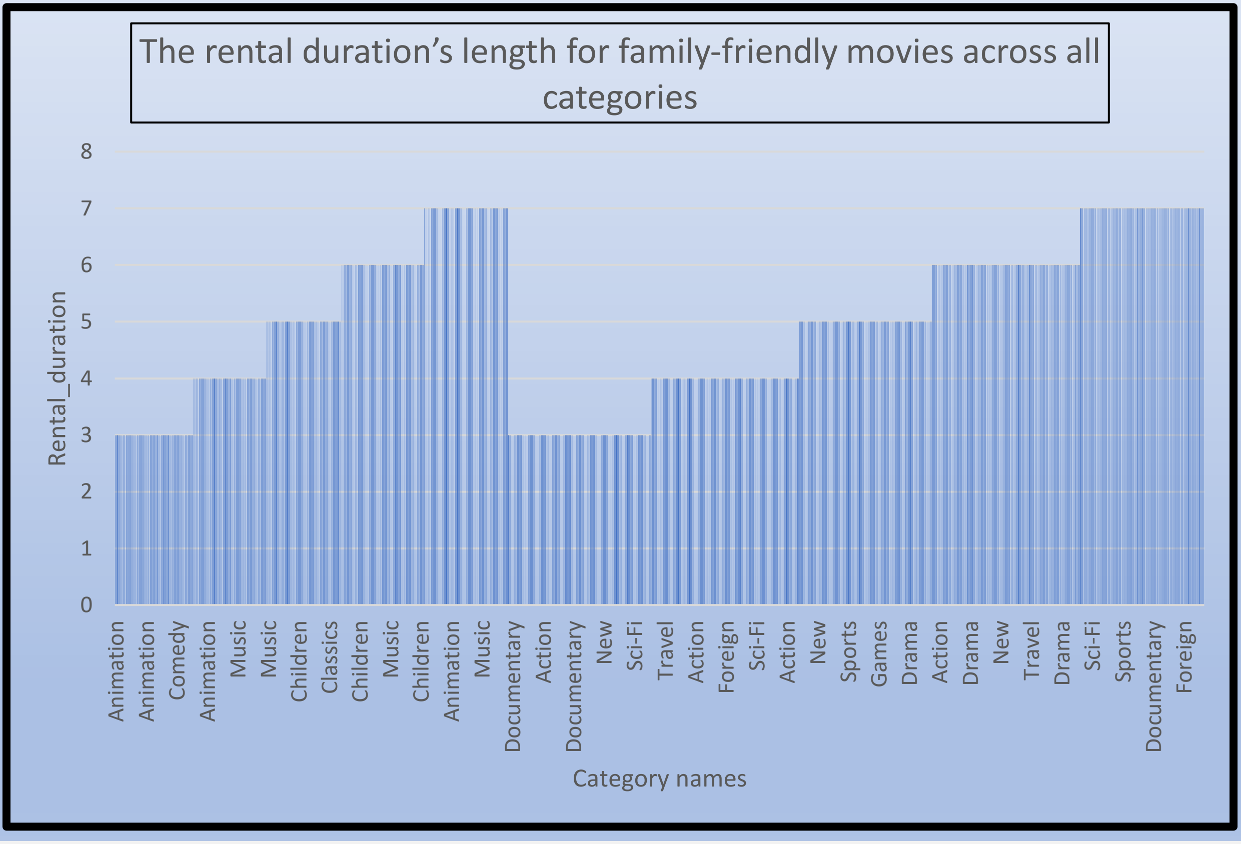
We can see here the five categorizations for the family-friendly movies with the sum of their rental count. Movies categorized as 'Animation' satisfy the highest rental out. However, music movies are the least.

# Question 2:

Now we need to know how the length of rental duration of these family- friendly movies compares to the duration that all movies are rented for. Can you provide a table with the movie titles and divide them into 4 levels (first_quarter, second_quarter, third_quarter, and final_quarter) based on the quartiles (25%, 50%, 75%) of the rental duration for movies across all categories? Make sure to also indicate the category that these family-friendly movies fall into.

```
WITH T1 AS(
SELECT f.title, c.name, f.rental_duration,
NTILE(4) OVER (ORDER BY f.rental_duration) AS standard_quartile
FROM category c
JOIN film_category fc
ON c.category_id = fc.category_id
JOIN film f
ON fc.film_id = f.film_id
WHERE c.name IN ('Animation', 'Children', 'Classics', 'Comedy',
'Family', 'Music')
ORDER BY 3),
T2 AS(
SELECT f.title, c.name, f.rental_duration,
NTILE(4) OVER (ORDER BY f.rental_duration) AS standard_quartile
FROM category c
JOIN film_category fc
ON c.category_id = fc.category_id
JOIN film f
ON fc.film_id = f.film_id
WHERE c.name NOT IN ('Animation', 'Children', 'Classics', 'Comedy',
'Family', 'Music')
ORDER BY 3)

Select *
From T1
UNION ALL
SELECT *
FROM T2
```



The figure shows the length of the rental duration for family-friendly movies across all categories. From here we can discover that family-friendly movies and other categories have the same rental duration length with a maximum of 7 times rental duration and a minimum of 3 times.

# Question 3:

Finally, provide a table with the family-friendly film category, each of the quartiles, and the corresponding count of movies within each combination of film category for each corresponding rental duration category. The resulting table should have three columns:

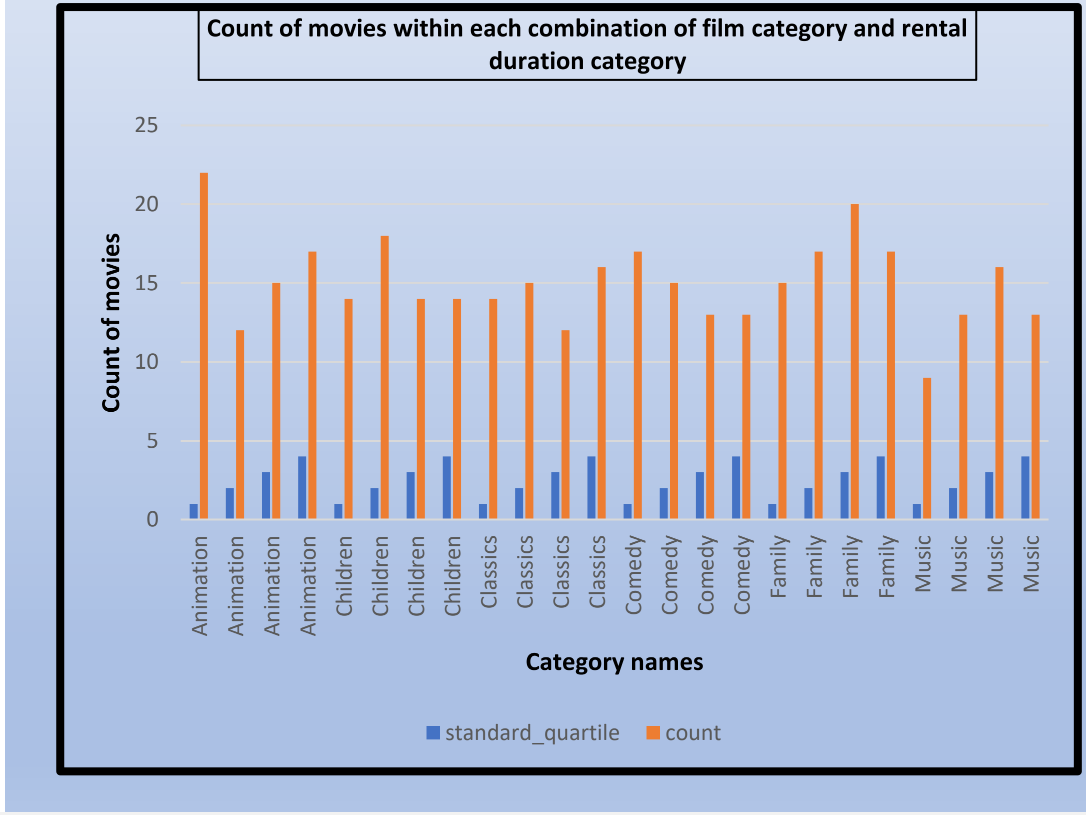
Category
Rental length category
Count

```
SELECT T1.name, T1.standard_quartile, COUNT(*)

FROM
(
SELECT f.title, c.name, f.rental_duration, NTILE(4) OVER
(ORDER BY f.rental_duration) AS standard_quartile
FROM category c
JOIN film_category fc
ON c.category_id = fc.category_id
JOIN film f
ON fc.film_id = f.film_id
WHERE c.name IN ('Animation', 'Children', 'Classics', 'Comedy',
'Family', 'Music'))T1

 GROUP BY 1,2

 ORDER BY 1,2
```



Count of movies within each combination of film category and rental duration category

Here we can see the count of movies within each combination of film category and rental duration category. The data shows that the maximum count is for the movies in the ‘Animation’ category, which have one-time rental duration, followed by ‘family’ movies with three-times rental duration.

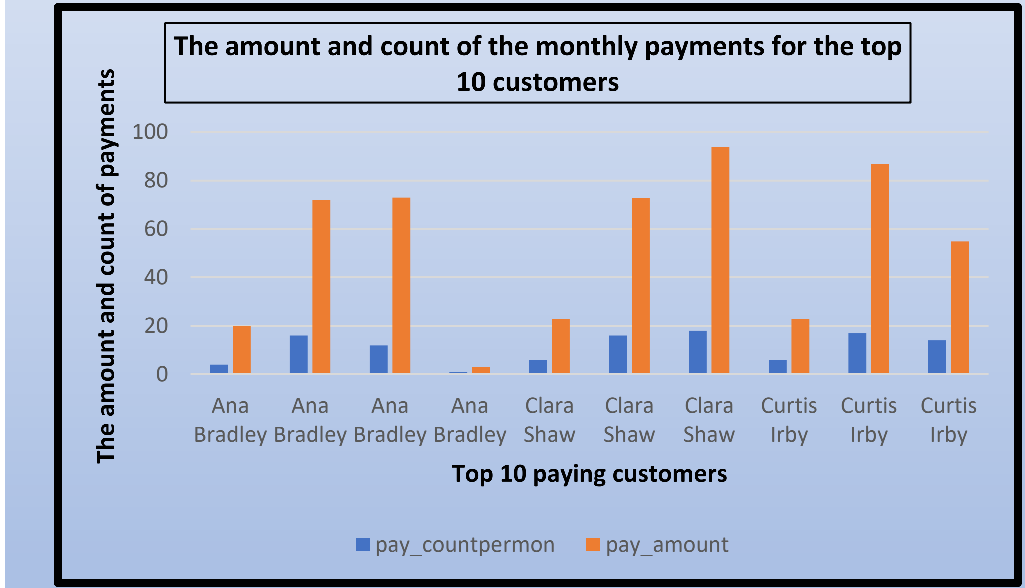
# Question 4:

We would like to know who were our top 10 paying customers, how many payments they made on a monthly basis during 2007, and what was the amount of the monthly payments. Can you write a query to capture the customer name, month and year of payment, and total payment amount for each month by these top 10 paying customers?

```
SELECT DATE_TRUNC('month', p.payment_date) pay_month,
T1.full_name,
COUNT(p.amount) AS pay_countpermon,
SUM(p.amount) AS pay_amount

FROM
(SELECT c.first_name || ' ' || c.last_name AS full_name,
SUM(p.amount) AS amount_total, c.customer_id
FROM customer c
JOIN payment p
ON p.customer_id = c.customer_id
GROUP BY 1,3
ORDER BY 2 DESC
LIMIT 10) T1
JOIN customer c
ON T1.customer_id = c.customer_id
JOIN payment p
ON p.customer_id = c.customer_id

WHERE p.payment_date BETWEEN '2007-01-01' AND '2008-01-01'
GROUP BY 2, 1
ORDER BY 2, 1, 3
LIMIT 10;
```



The figure shows the names for the top 10 paying customers with the number of times and the amount of their monthly payments. The data displays that the top 10 customers have roughly the same number of payments per month, but with different amount of payments.