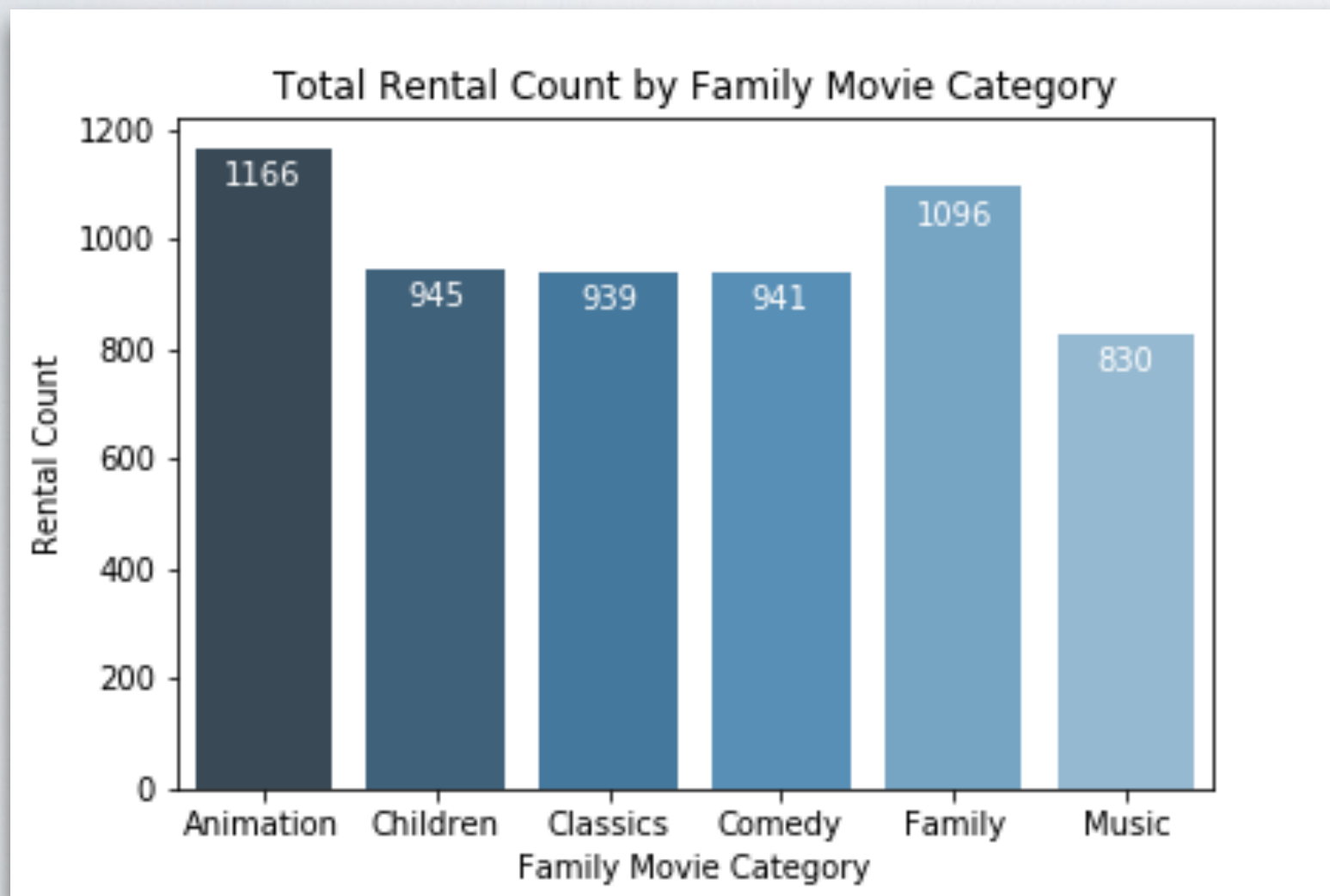


# INVESTIGATE THE SAKILA DVD RENTAL DATABASE

Morgan Ivey

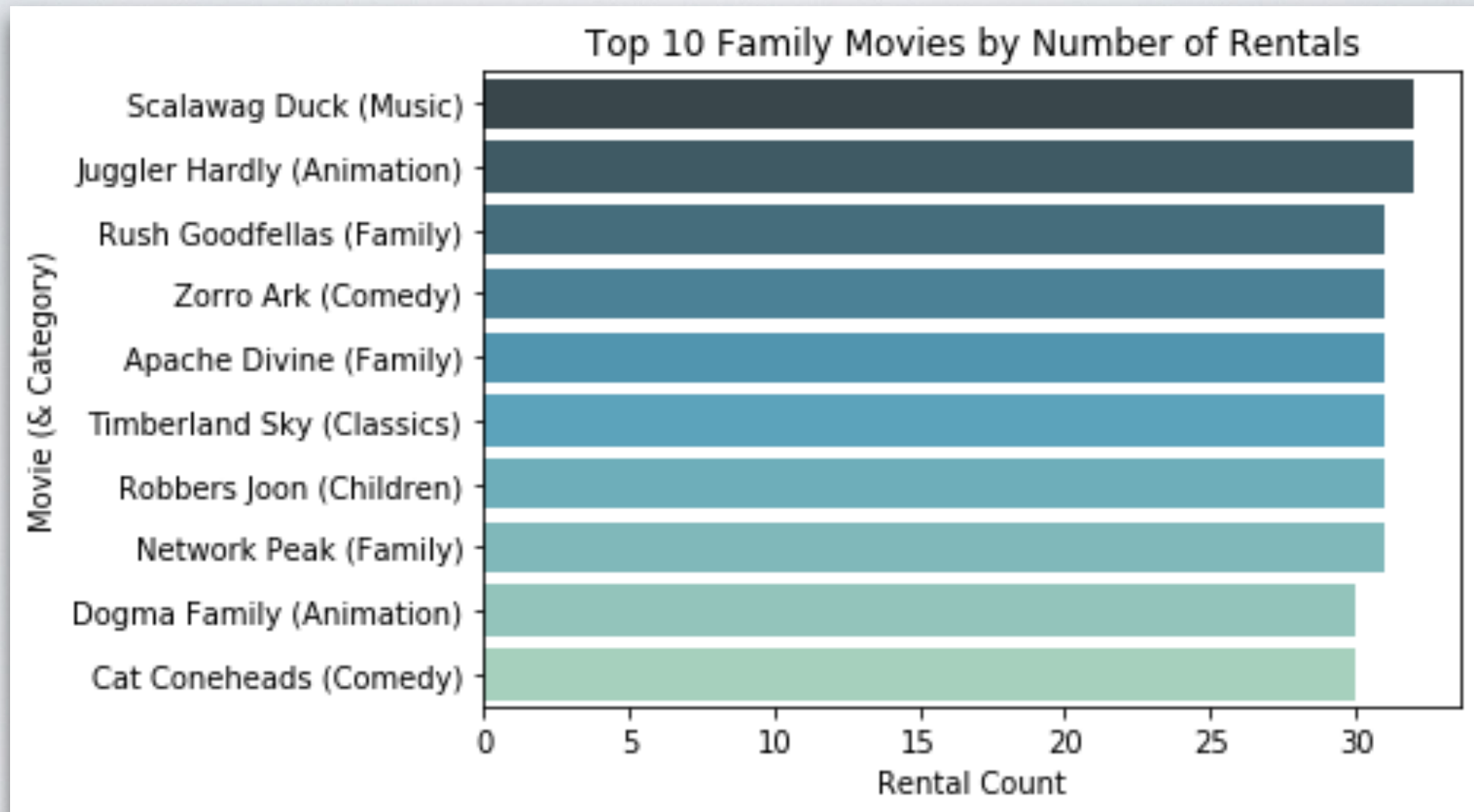
# QUESTION 1

Out of all of the family movies, what are the categories of those movies and how many times has it been rented out?



Out of all of the the family movie categories, *animation* has the highest rentals with 1,166 rentals. The second highest rental count is for *family* movies with 1,096 rentals. *Music* has the lowest rental count with 830 rentals. *Children*, *classics*, and *comedy* have similar rental counts in the range of 939 to 945.

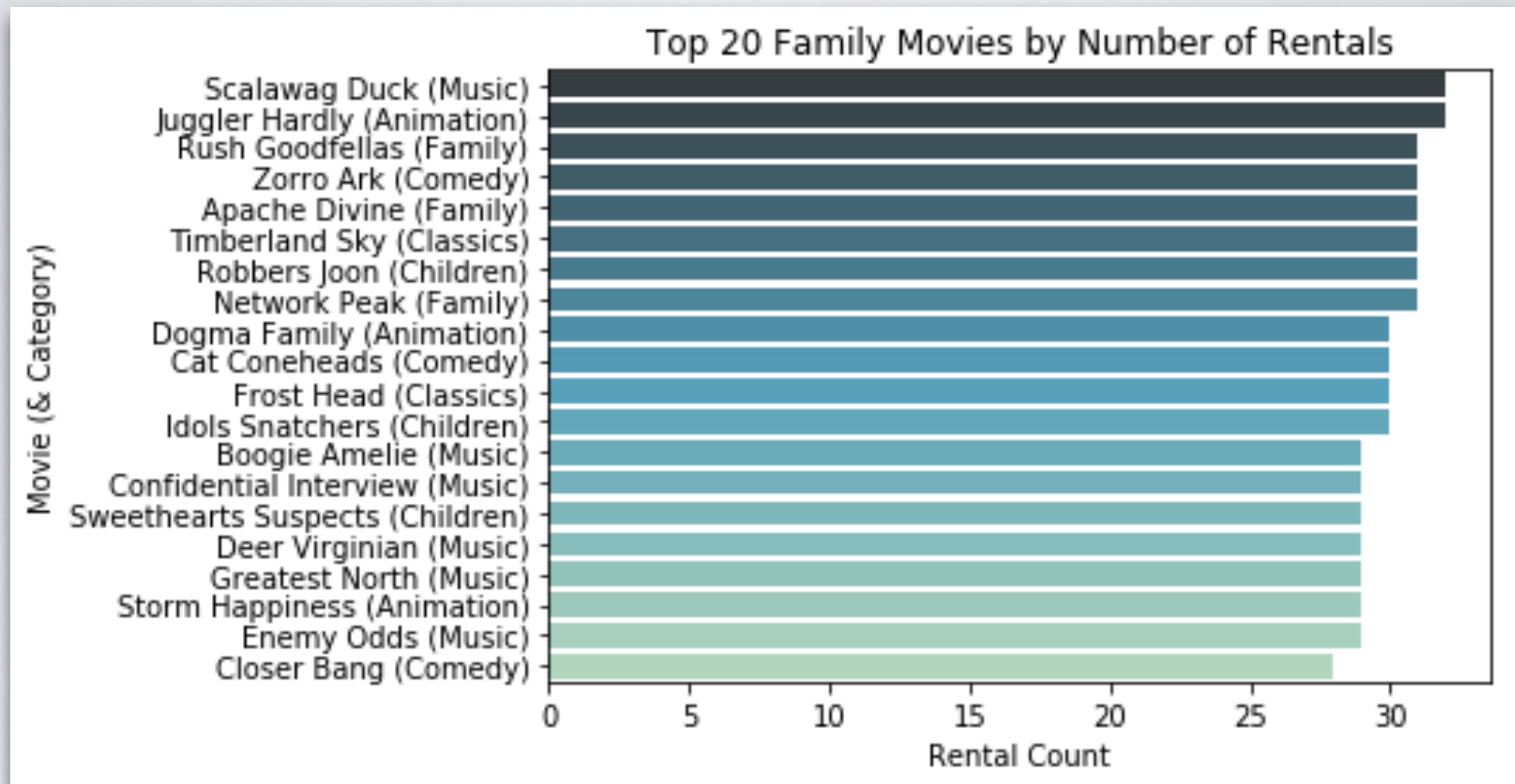
# QUESTION 1 (CONT)



The top 10 family movie rentals are *Scalawag Duck*, *Juggler Hardly*, *Rush Goodfellas*, *Zorro Ark*, *Apache Divine*, *Timberland Sky*, *Robbers Joon*, *Network Peak*, *Dogma Family*, and *Cat Coneheads*. Out of these top 10 movies, the top rental is *Scalawag Duck* and it the only movie in the music category. There are 2 animation movies, 3 family movies, 2 comedy movies, 1 classics movie, and 1 children movie in the top 10 family movie rentals.



# QUESTION 1 (CONT)

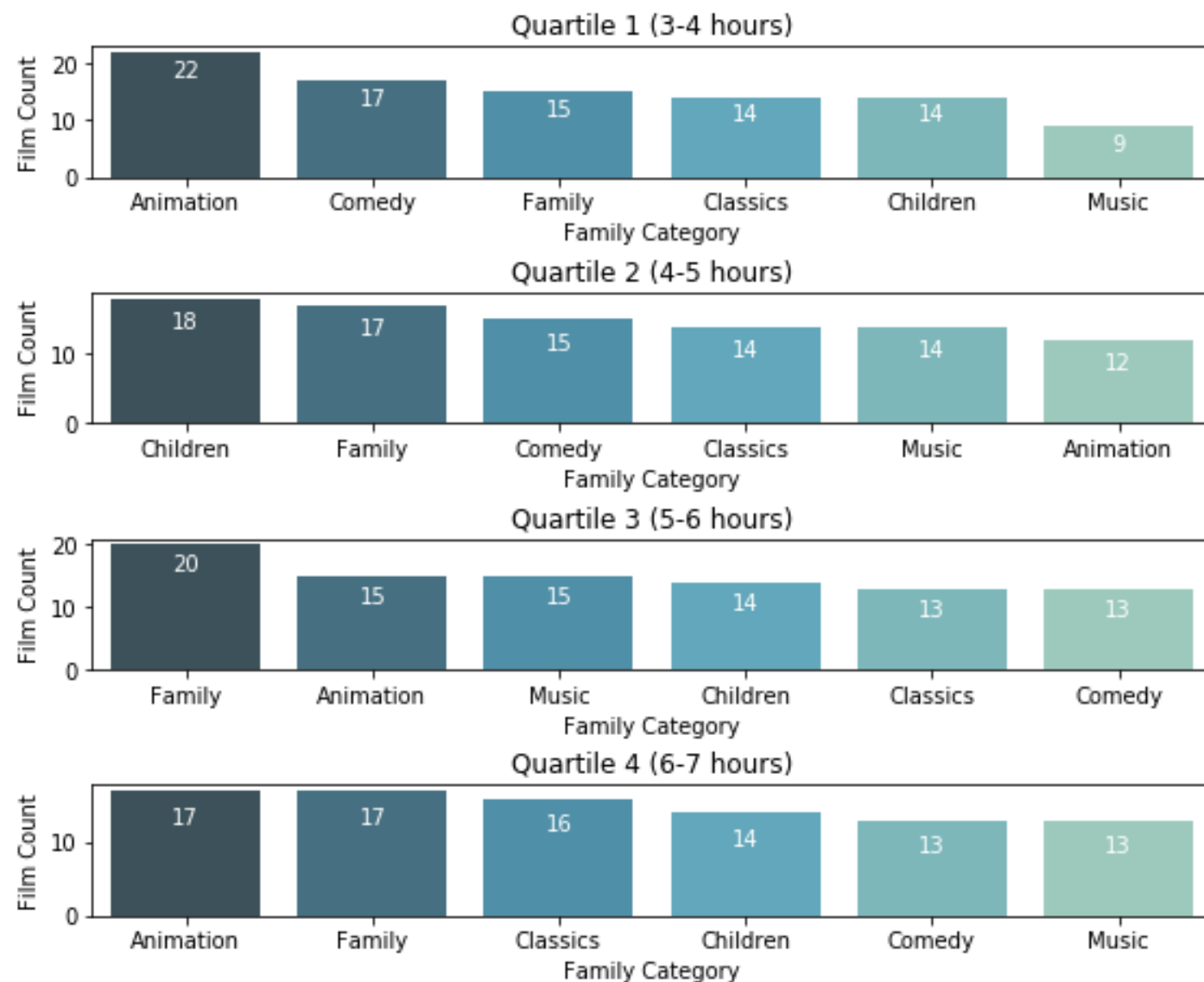


The top 20 family movie rentals include the movies from the previous slide, and also include *Frost Head*, *Idols Snatchers*, *Boogie Amelie*, *Confidential Interview*, *Sweethearts Suspects*, *Deer Virginian*, *Greatest North*, *Storm Happiness*, *Enemy Odds*, and *Closer Bang*. Out of all of these top 20 family movies, 60% of them are in the music category, 30% of them are in the children category, 30% of them are in the animation category, 30% of them are in the comedy category, 30% of them are in the family category, and 20% are in the classics category. Thus, the majority of the top 20 family movies are music movies.

# QUESTION 2

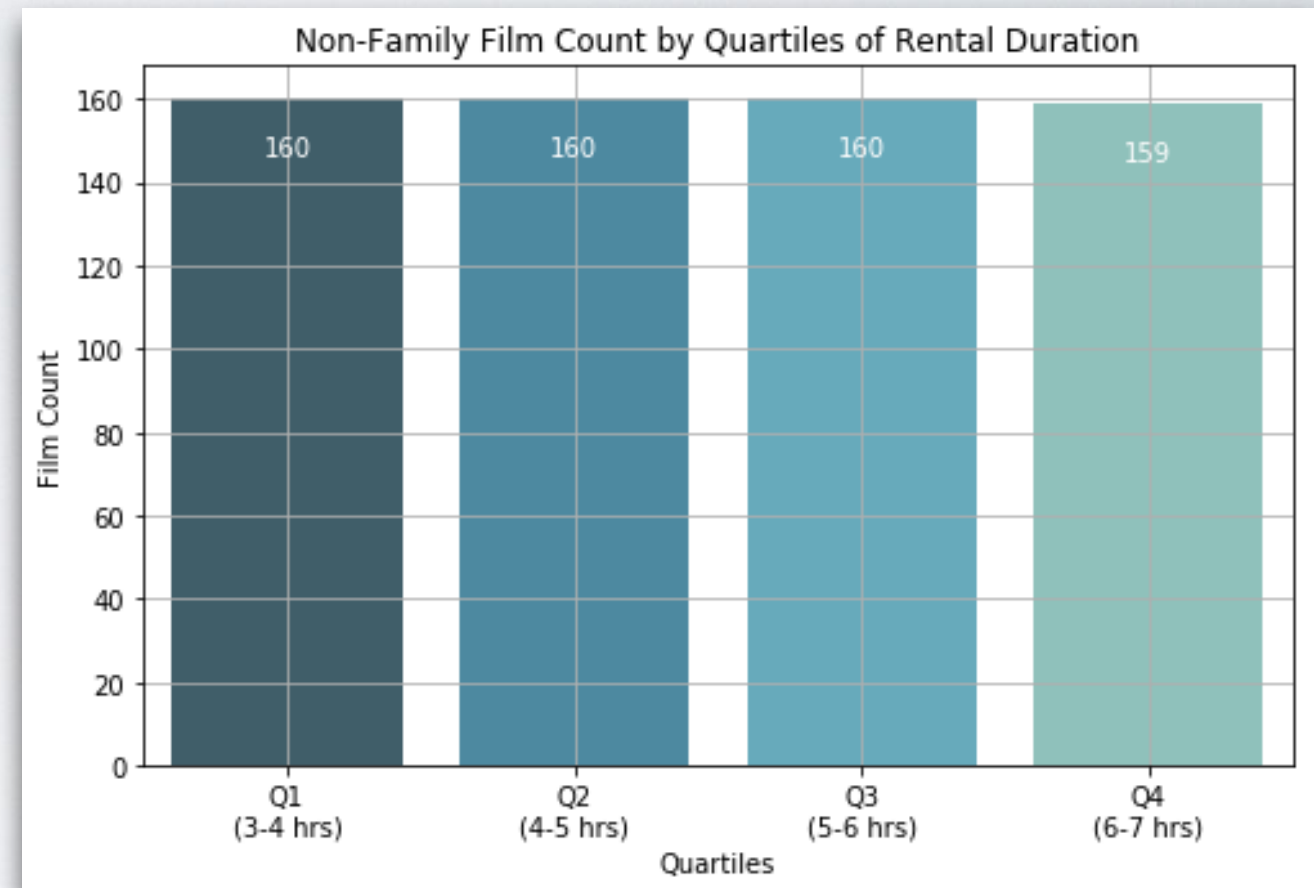
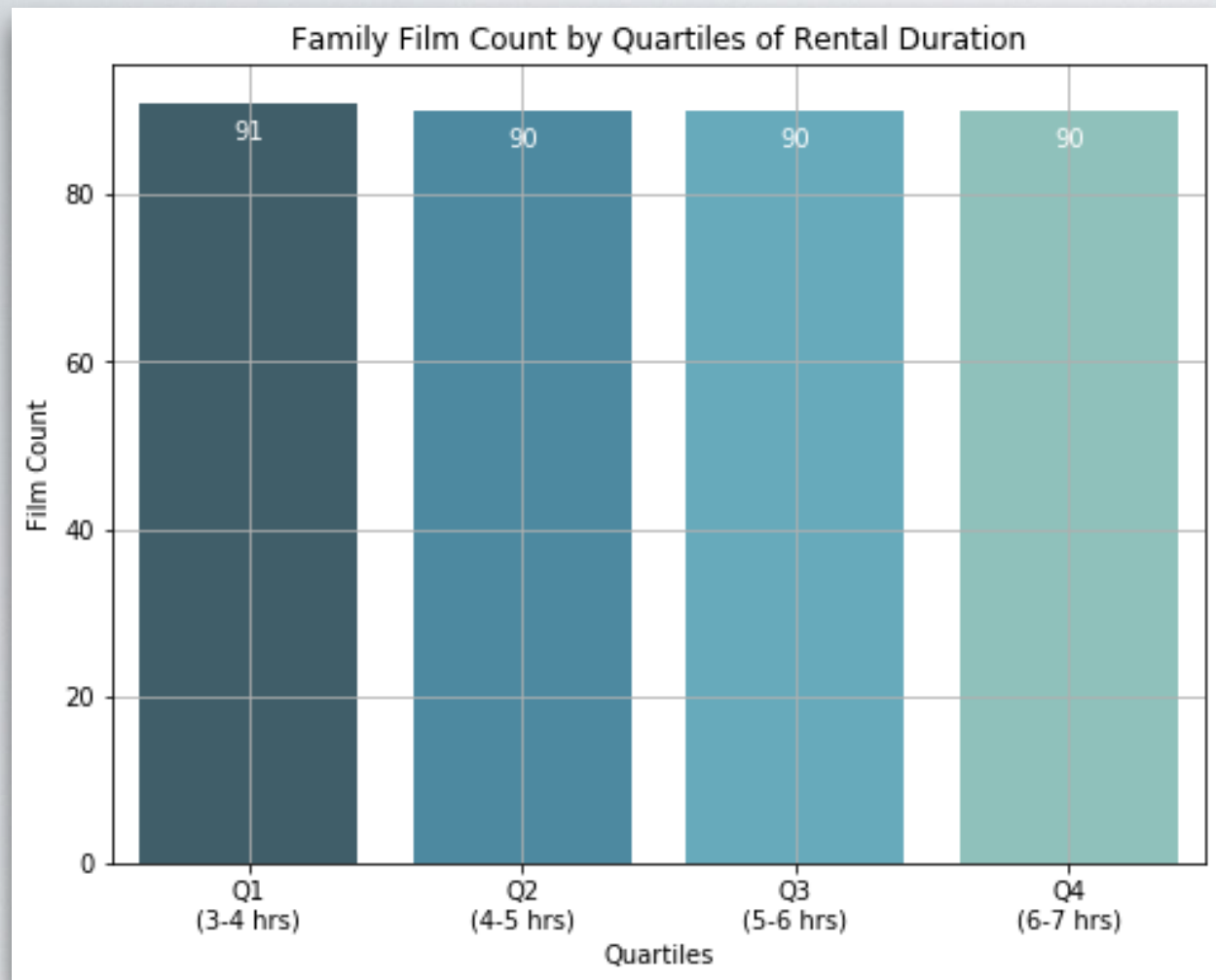
How does the length of rental duration of these family-friendly movies compare to the duration that all movies are rented for?

Movie Count for each Rental Duration Quartile & each Family Category



For Quartile 1, the *animation* category has the highest movie count (22 movies) and the *music* category has the lowest movie count (9 movies). For Quartile 2, the *children* category has the highest movie count (18 movies) and the *animation* category has the lowest movie count (12 movies). For Quartile 3, the *family* category has the highest movie count (20 movies), and the *classics* and *comedy* categories have the lowest movie counts (13 movies). For Quartile 4, the *animation* and *family* categories have the highest movie count (17 movies), and the *comedy* and *music* categories have the lowest movie counts (13 movies).

# QUESTION 2 (CONT)



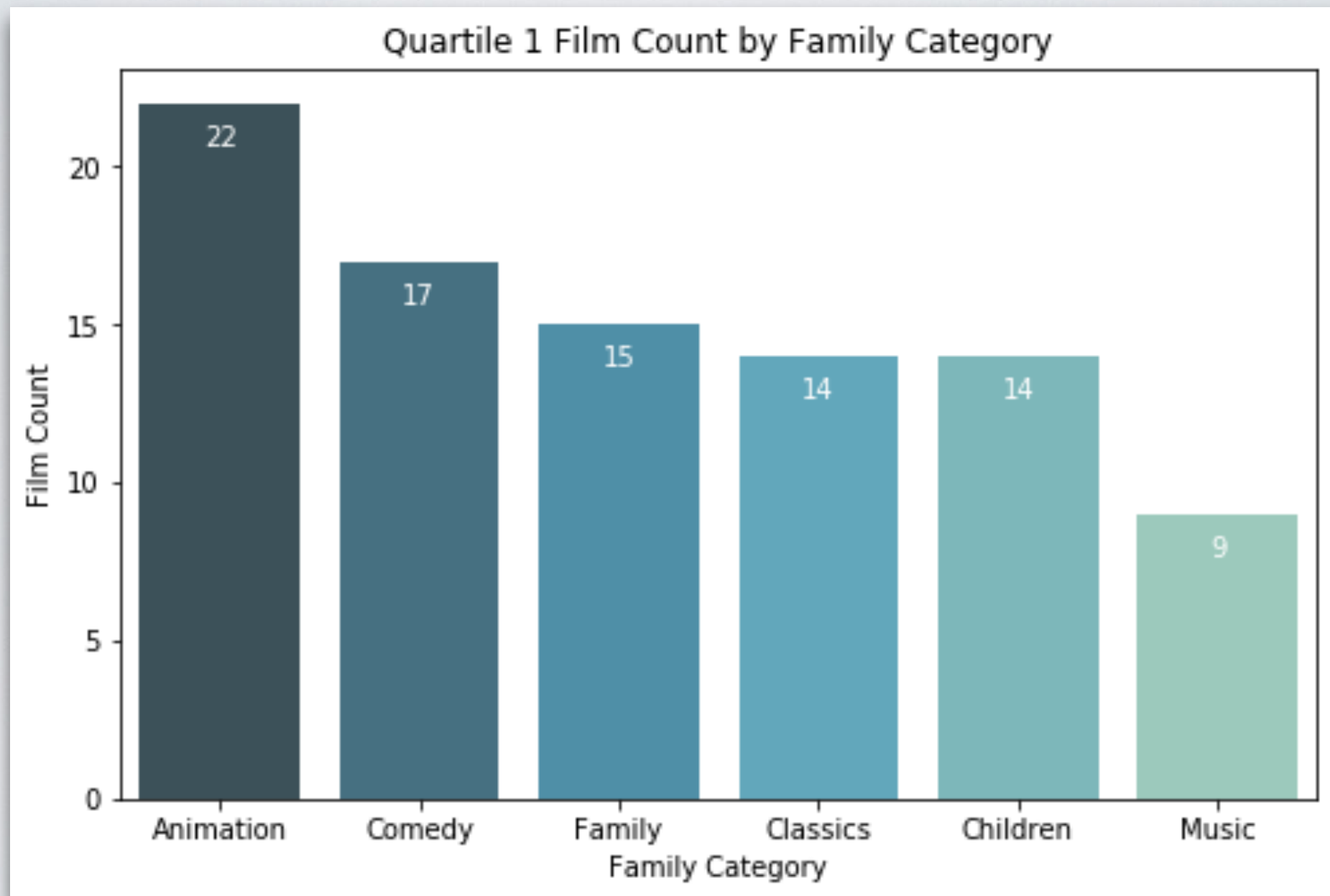
There are 91 family movies in the first quartile, and 90 family movies in the second, third, and fourth quartile.

There are 160 non-family movies in the first, second and third quartiles, and 159 non-family movies in the fourth quartile.

Thus, family movies are usually rented for 3–4 hours, compared to non-family movies, which are rented either the same or longer (for 3–6 hours).

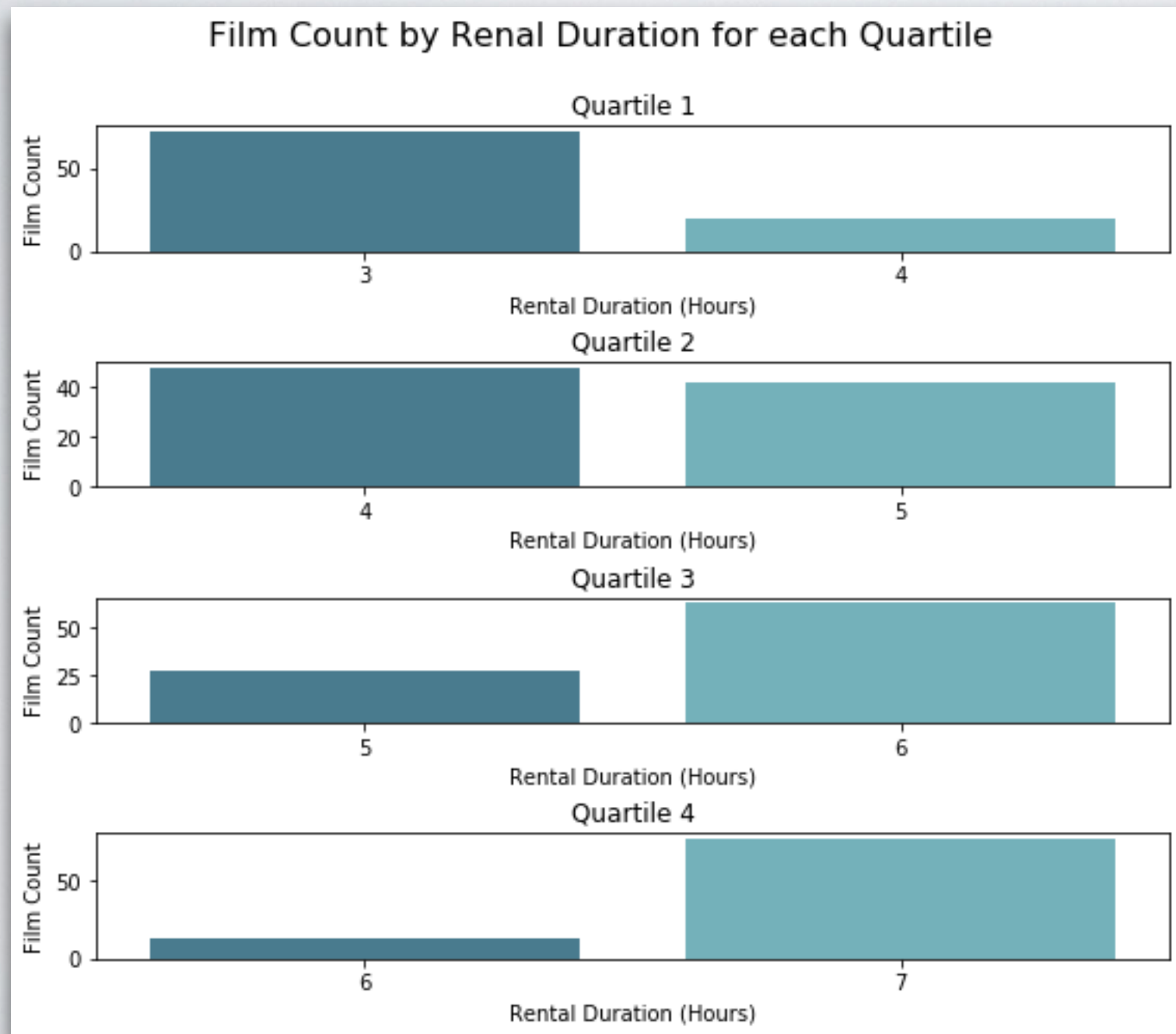


# QUESTION 2 (CONT)



For Quartile 1, the *animation* category has the highest movie count (22 movies) and the *music* category has the lowest movie count (9 movies)

# QUESTION 2 (CONT)



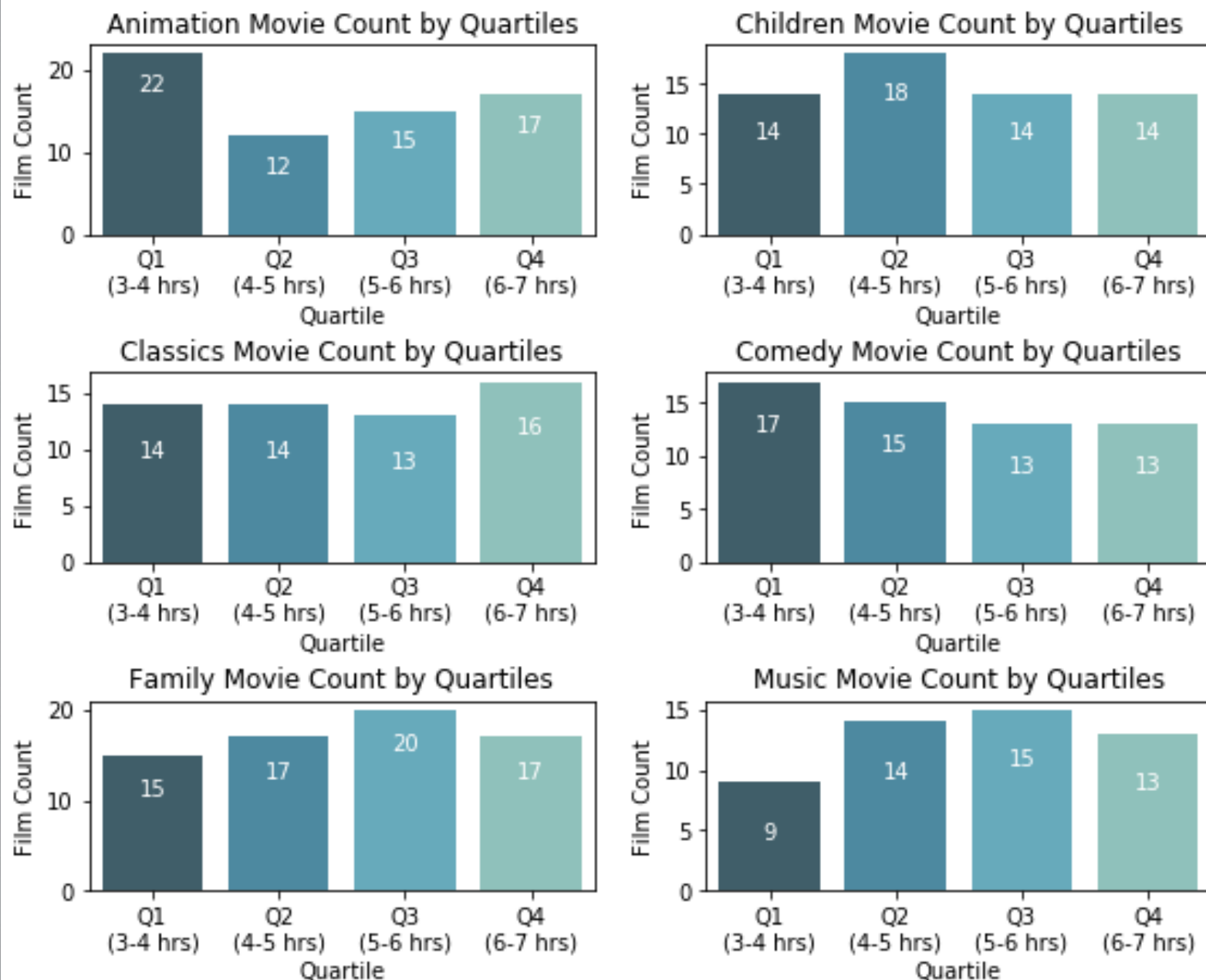
The majority of movies rented in the 1st Quartile are rented for 3 hours, in the 2nd Quartile are rented for 4 hours, in the 3rd Quartile are rented for 6 hours, and in the 4th Quartile are rented for 7 hours.



# QUESTION 3

How many movies are within each combination of film category for each corresponding rental duration category?

Movie Count for each Family Category & each Quartile



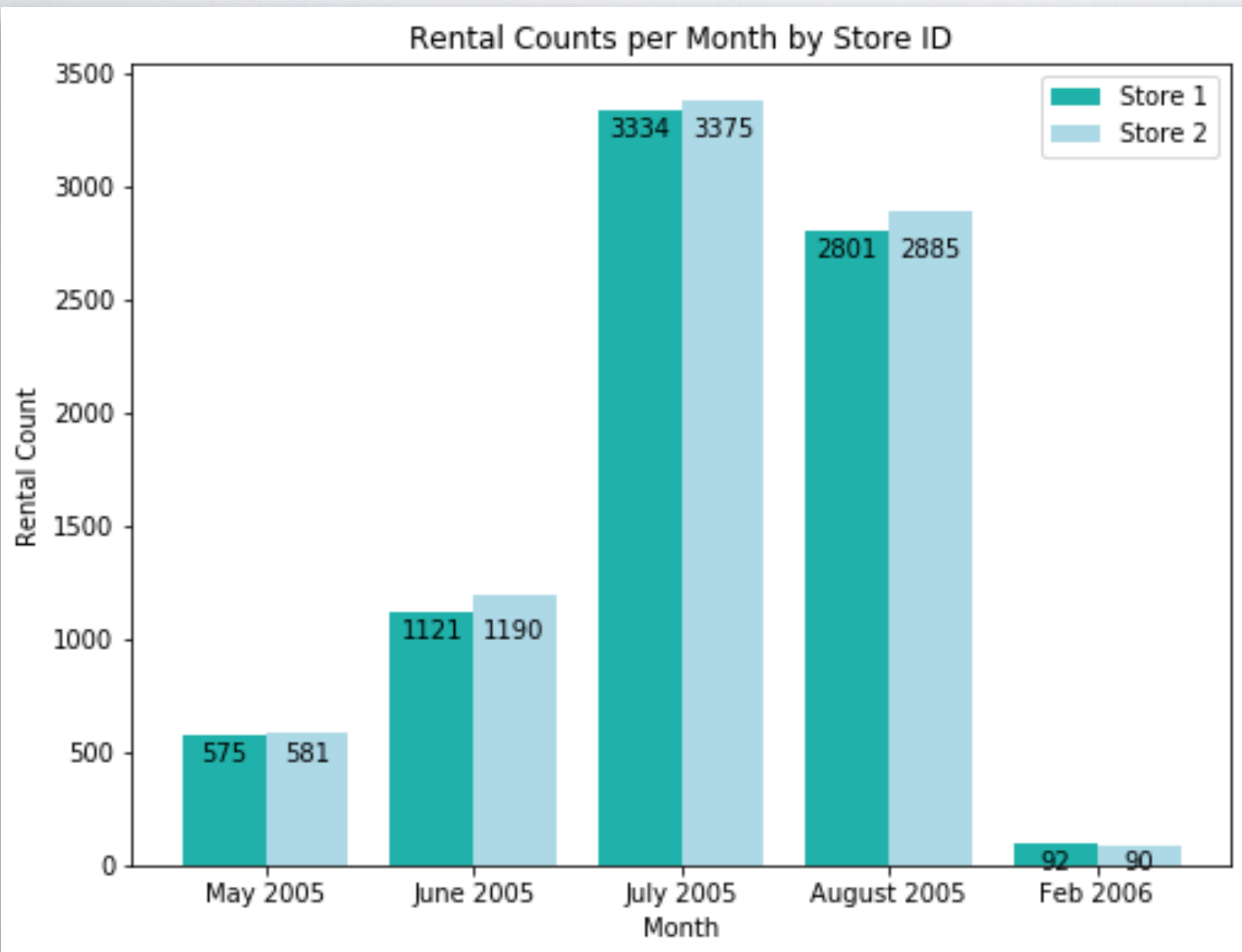
Quartile 1 (3-4 hour rental duration) has the lowest count for *music* and *family* films. Quartile 2 (4-5 hours rental duration) has the lowest count for *animation* films, which is a rental duration of 4-5 hours. Quartile 3 (5-6 hour rental duration) has the lowest count for *comedy* and *classics* films.

Moreover, Quartile 1 has the highest count for *animation* and *comedy* films, Quartile 2 has the highest count for *children* films, Quartile 3 has the highest count for *music* and *family* films, and Quartile 4 (6-7 hour rental duration) has the highest count for *classics* films.

In other words, most of the animation and comedy movies are rented out for 3-4 hours, most of the children movies are rented out for 4-5 hours, most of the music and family movies are rented out for 5-6 hours, and most of the classics movies are rented out for 6-7 hours. Alternatively, not many animation movies are rented out for 4-5 hours, not many music or family movies are rented out for 3-4 hours.

# QUESTION 4

How do the two stores compare in their count of rental orders during every month for all the years we have data for?



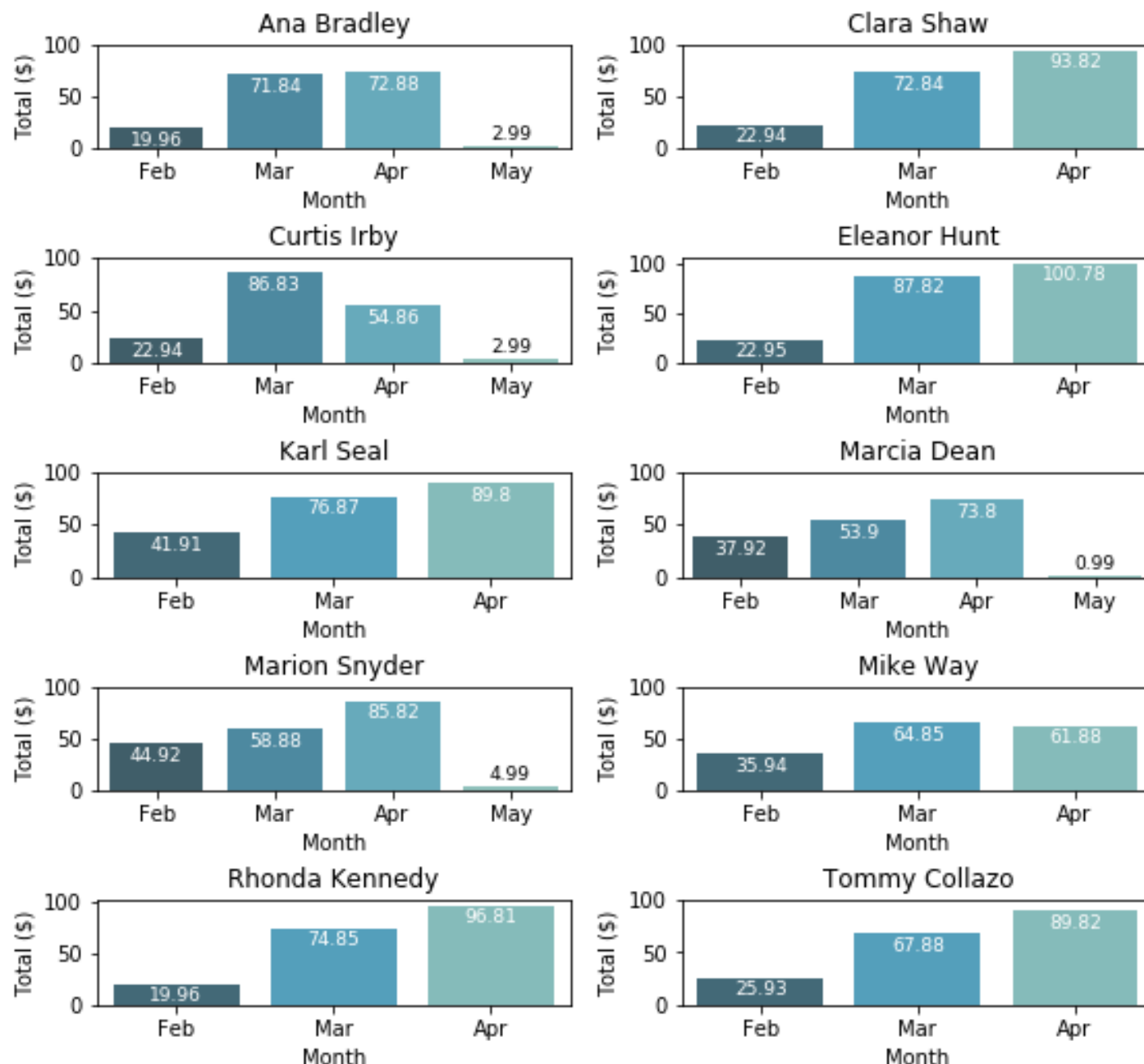
For each month in 2005, Store 2 had more rentals than Store 1. In May, Store 2 had 6 more rentals than Store 1. Store 2 had 69 more rentals in June, 41 more rentals in July, and 84 more rentals in August. In February 2006, Store 1 had 2 more rentals compared to Store 1.



# QUESTION 5

Who were the top 10 paying customers, how many payments did they make on a monthly basis during 2007, and what was the amount of the monthly payments?

Top 10 Customers' Purchase Totals Per Month in 2007



The top 10 paying customers in 2007 are Ana Bradley, Clara Shaw, Curtis Irby, Eleanor Hunt, Karl Seal, Marcia Dean, Marion Snyder, Mike Way, Rhonda Kennedy, and Tommy Collazo.

Curtis Irby and Mike Way paid the most in March compared to the other months, whereas Ana Bradley, Clara Shaw, Eleanor Hunt, Karl Seal, Marcia Dean, Marion Snyder, Rhonda Kennedy, and Tommy Collazo all paid the most in April compared to the other months.

Moreover, Ana Bradley, Curtis Irby, Marcia Dean, and Marion Snyder all spent the least amount in May compared to the other months. Clara Shaw, Eleanor Hunt, Karl Seal, Mike Way, Rhonda Kennedy, and Tommy Collazo all spent the least amount in February compared to the other months.

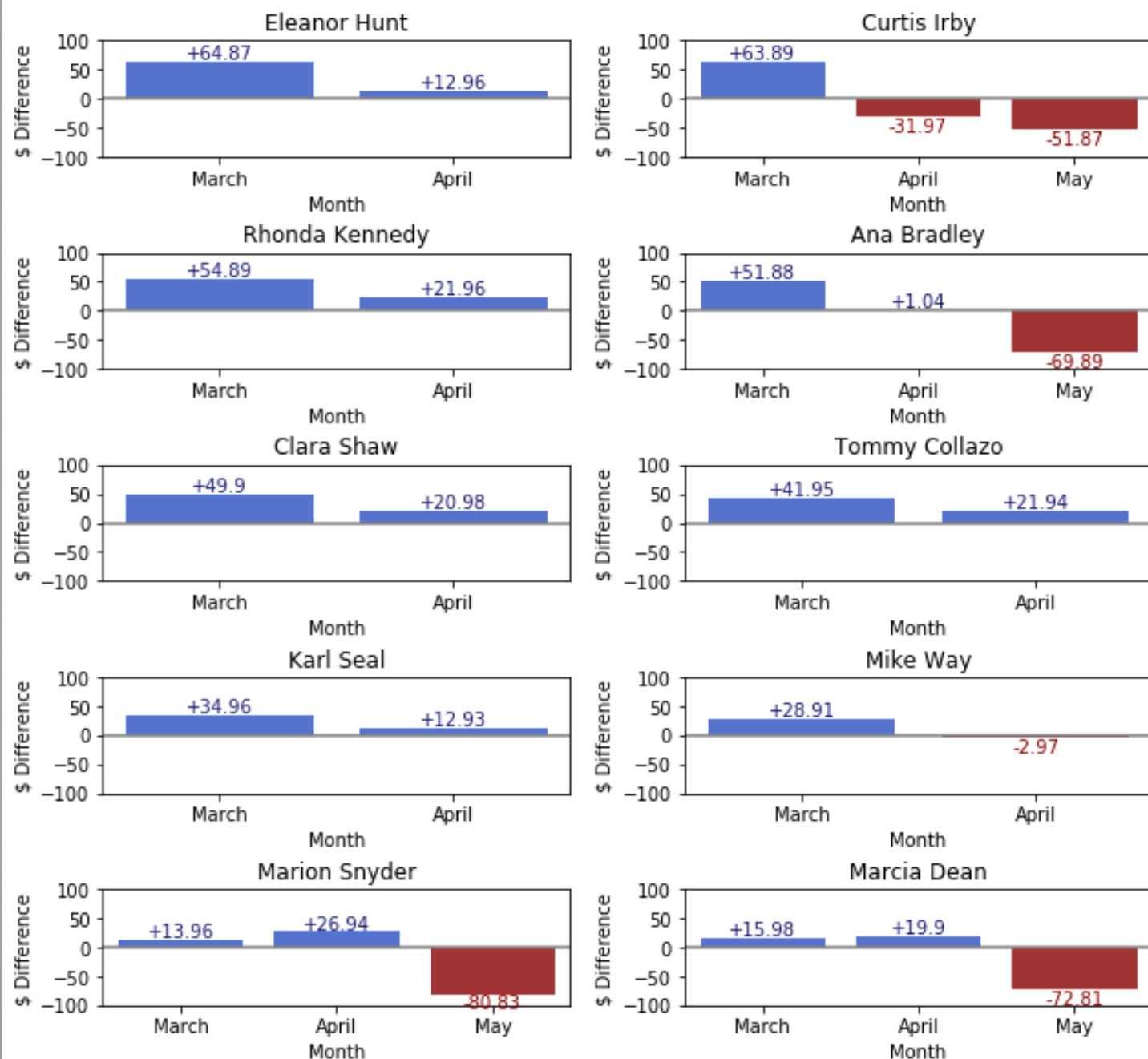


# QUESTION 6

What is the difference across their monthly payments during 2007 for each of the top 10 paying customers?

Who is the customer who paid the most difference in terms of payment?

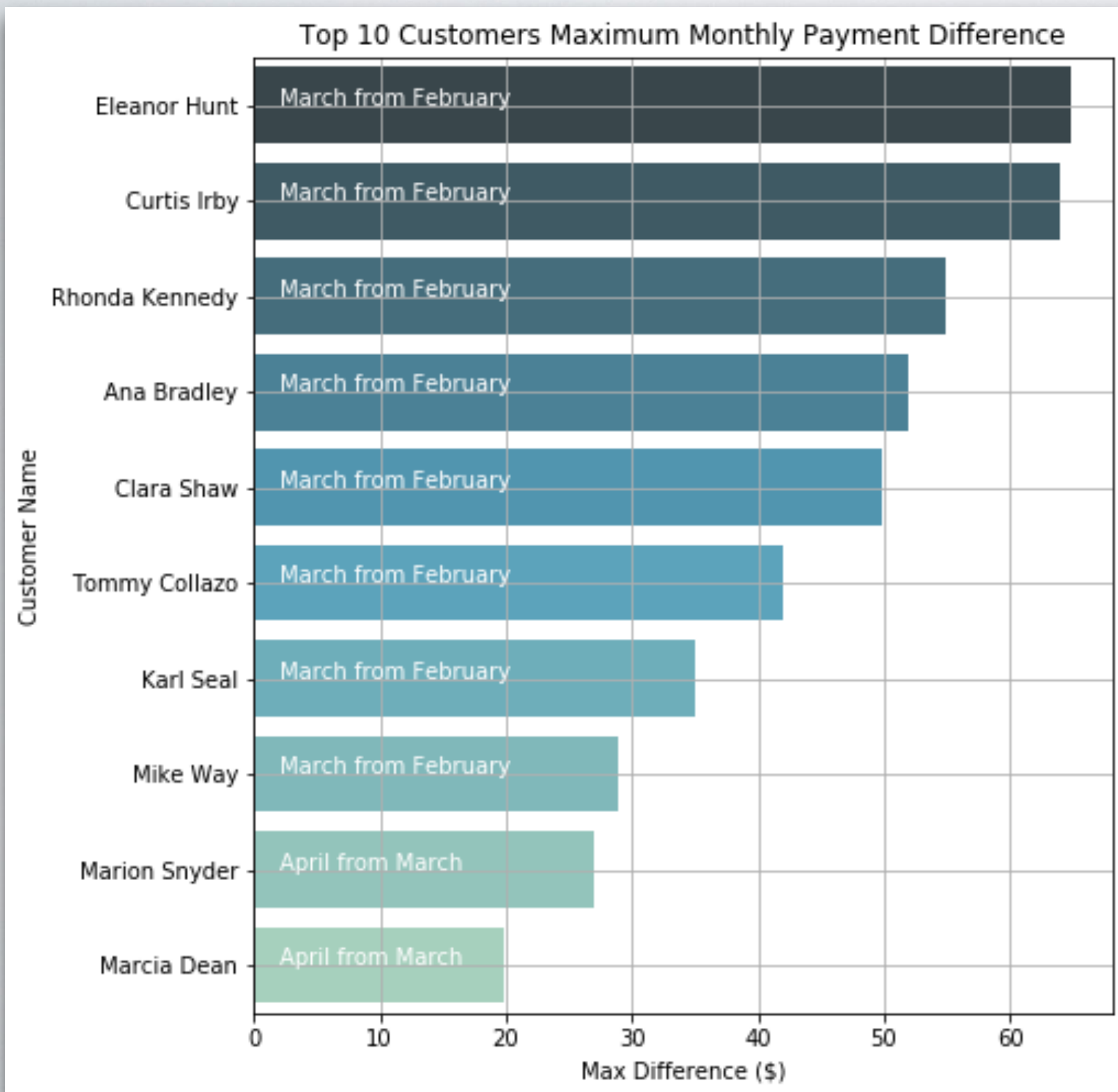
Top 10 Customers' Monthly Purchase Differences in 2007



The customer Eleanor Hunt paid the maximum difference of \$64.87 during March 2007 from February of 2007.

Curtis Irby paid the second highest difference during March 2007 from February of 2007, but then paid less in April and May from March and April, respectively. Marion Snyder, Marcia Dean, and Ana Bradley all had negative monthly payment differences in May 2007 from April 2007.

# QUESTION 6 (CONT)



The customer **Eleanor Hunt** paid the maximum difference of \$64.87 during March 2007 from February of 2007.

Curtis Irby, Rhonda Kennedy, Ana Bradley, Clara Shaw, Tommy Collazo, Karl Seal, and Mike Way all paid their respective maximum differences during March 2007 from February of 2007, whereas Marion Snyder and Marcia Dean paid their respective maximum differences during April 2007 from March of 2007.



# QUERIES

## Question 1

```
select f.title as film_title,
       c.name as category_name,
       count(r.rental_date) as rental_count
from film_category as fc
join category as c
on c.category_id = fc.category_id
join film as f
on f.film_id = fc.film_id
join inventory as i
on fc.film_id = i.film_id
join rental as r
on r.inventory_id = i.inventory_id
where c.name = 'Animation' or
       c.name = 'Children' or
       c.name = 'Classics' or
       c.name = 'Comedy' or
       c.name = 'Family' or
       c.name = 'Music'
group by film_title, category_name
order by 2,1;
```

## Question 2

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

## Question 3

```
SELECT category_name, standard_quartile, count(standard_quartile) as movie_count
FROM
  (SELECT
     c.name as category_name,
     f.title,
     NTILE(4) OVER(ORDER BY f.rental_duration) AS standard_quartile
   FROM film_category AS fc
   JOIN category AS c
   ON c.category_id = fc.category_id
   JOIN film AS f
   ON f.film_id = fc.film_id
   WHERE c.name IN ('Animation', 'Children', 'Classics', 'Comedy', 'Family', 'Music')) f1
GROUP BY category_name, standard_quartile
ORDER BY category_name, standard_quartile
```



## Question 4

```
SELECT
    rental_month, rental_year, store_id,
    sum(count_rentals_one) as count_rentals
FROM
    (SELECT
        EXTRACT(MONTH FROM r.rental_date) AS rental_month,
        EXTRACT(YEAR FROM r.rental_date) AS rental_year,
        i.store_id as store_id,
        count(r.rental_id) as count_rentals_one
    FROM rental r
    JOIN inventory i
    ON r.inventory_id = i.inventory_id
    GROUP BY r.rental_date, store_id) f1
GROUP BY rental_month, rental_year, store_id
ORDER BY count_rentals DESC;
```

## Question 5

```
SELECT f2.trunc_month, full_name, f2.pay_countpermon, f2.customer_amount
FROM
    (SELECT f1.full_name AS full_name,
        sum(customer_amount) OVER (PARTITION BY f1.full_name) AS sum_customer,
        f1.trunc_month, f1.pay_countpermon, f1.customer_amount
    FROM
        (SELECT
            date_trunc('month', p.payment_date) AS trunc_month,
            concat(c.first_name, ' ', c.last_name) AS full_name,
            count(c.first_name) AS pay_countpermon,
            sum(p.amount) AS customer_amount
        FROM payment p
        JOIN customer c
        ON c.customer_id = p.customer_id
        WHERE EXTRACT(YEAR FROM date_trunc('month', p.payment_date))=2007
        GROUP BY trunc_month, full_name) f1
        ORDER BY sum_customer DESC) f2
WHERE full_name IN
    (
        SELECT full_name
        FROM
            (SELECT
                date_trunc('month', p.payment_date) AS trunc_month,
                concat(c.first_name, ' ', c.last_name) AS full_name,
                count(c.first_name) AS pay_countpermon,
                sum(p.amount) AS customer_amount
            FROM payment p
            JOIN customer c
            ON c.customer_id = p.customer_id
            WHERE EXTRACT(YEAR FROM date_trunc('month', p.payment_date))=2007
            GROUP BY trunc_month, full_name) f3
            GROUP BY full_name
            ORDER BY sum(customer_amount) DESC
            LIMIT 10)
    ORDER BY full_name, trunc_month;
```

## Question 6

```
WITH top_10_customers (order_month, full_name, pay_count, customer_amount)
AS
(
    SELECT f2.trunc_month,
           full_name,
           f2.pay_countpermon,
           f2.customer_amount
    FROM
        (SELECT f1.full_name AS full_name,
                 sum(customer_amount) OVER (PARTITION BY f1.full_name) AS sum_customer,
                 f1.trunc_month, f1.pay_countpermon, f1.customer_amount
        FROM
            (SELECT
                 date_trunc('month', p.payment_date) AS trunc_month,
                 concat(c.first_name, ' ', c.last_name) AS full_name,
                 COUNT(c.first_name) AS pay_countpermon,
                 SUM(p.amount) AS customer_amount
            FROM payment p
            JOIN customer c
            ON c.customer_id = p.customer_id
            WHERE EXTRACT(YEAR FROM date_trunc('month', p.payment_date))=2007
            GROUP BY trunc_month, full_name) f1
        ORDER BY sum_customer DESC) f2
    WHERE full_name IN
        (SELECT
             full_name
        FROM
            (SELECT
                 date_trunc('month', p.payment_date) AS trunc_month,
                 concat(c.first_name, ' ', c.last_name) AS full_name,
                 COUNT(c.first_name) AS pay_countpermon,
                 SUM(p.amount) AS customer_amount
            FROM payment p
            JOIN customer c
            ON c.customer_id = p.customer_id
            WHERE EXTRACT(YEAR FROM date_trunc('month', p.payment_date))=2007
            GROUP BY trunc_month, full_name) f3
        GROUP BY full_name
        ORDER BY SUM(customer_amount) DESC
        LIMIT 10)
    ORDER BY full_name, trunc_month
)
```

```
SELECT full_name,
       order_month,
       MAX(monthly_payment_difference)
FROM
    (SELECT order_month,
            full_name,
            pay_count,
            customer_amount,
            (customer_amount - lag(customer_amount) over (partition by full_name)) AS monthly_payment_difference
    FROM top_10_customers) f4
WHERE monthly_payment_difference IS NOT NULL
GROUP BY full_name, order_month
ORDER BY 3 DESC;
```