My SQL project

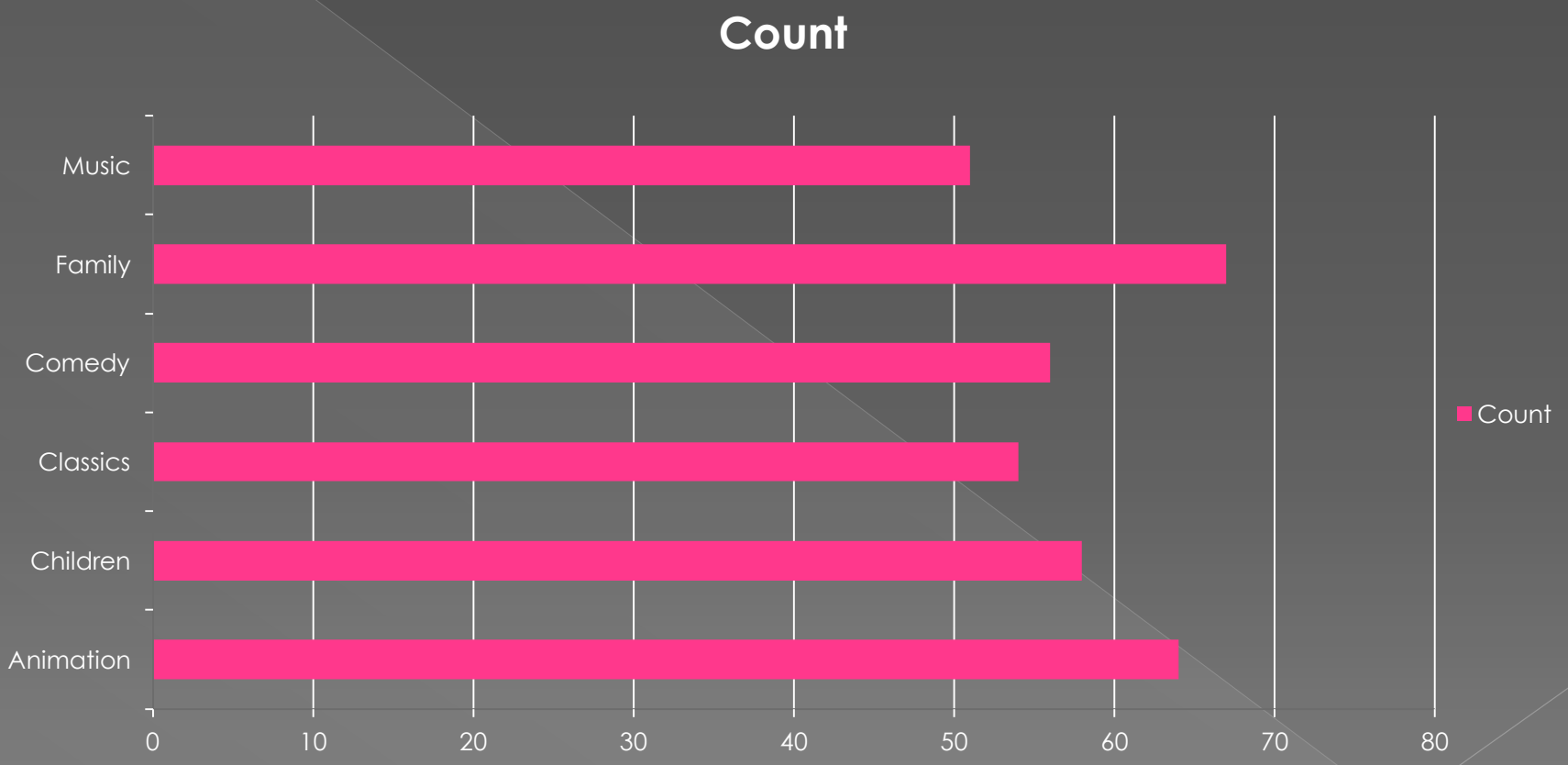
Prepared by Christopher Ozurumba

Workspace I question 1

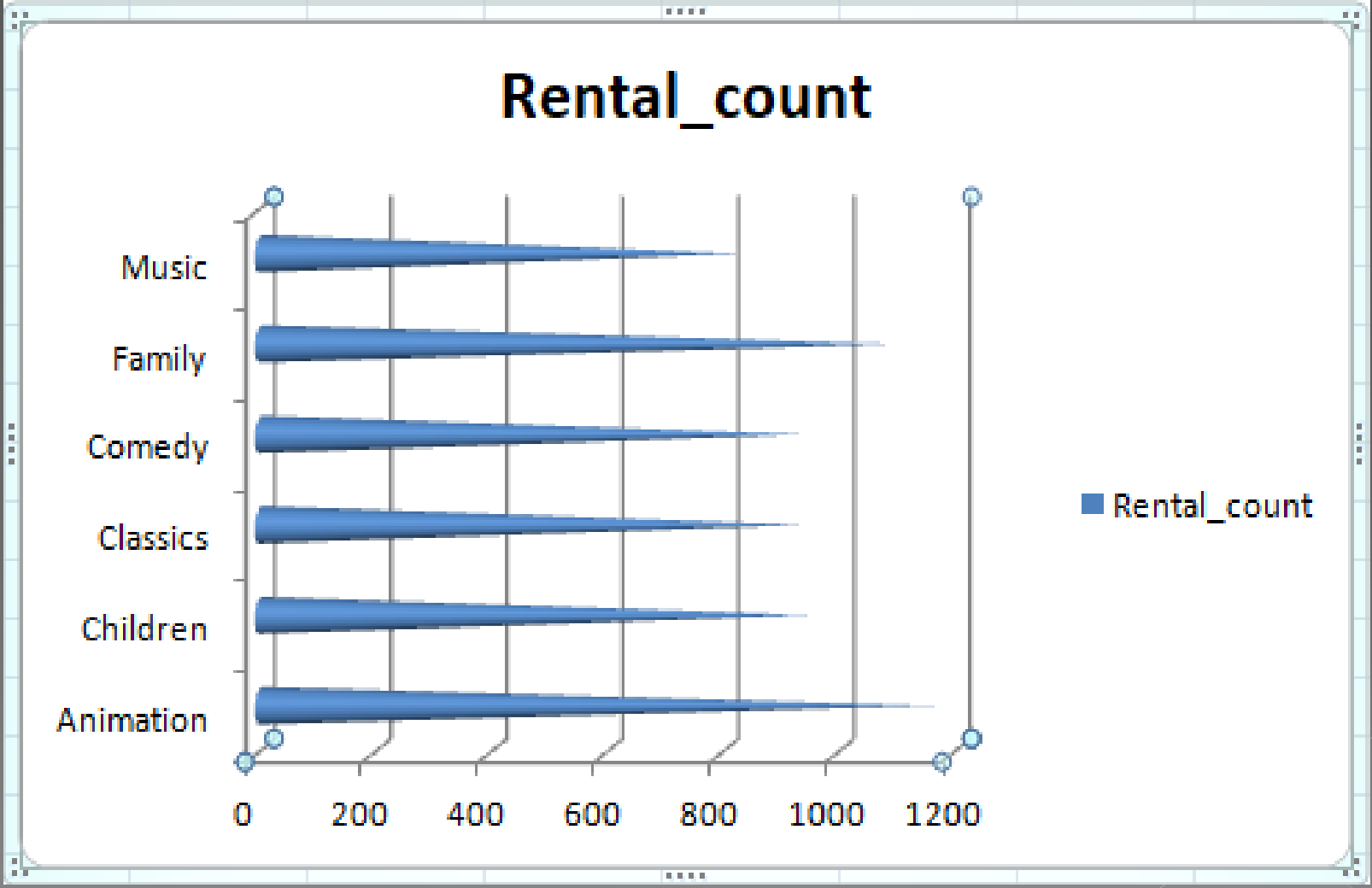
- We want to understand more about the movies that families are watching. The following categories are considered family movies: Animation, Children, Classics, Comedy, Family and Music.

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

Visualization I



Visualization II



Interpretation

- Highest rental count was observed with Animation (1166) which is 19.7% of the total rental count (5917).
- Lowest rental count was domiciled with music which accounted for 14% (830/5917)
- Highest category is Family (67) and lowest on the rank is music (51)

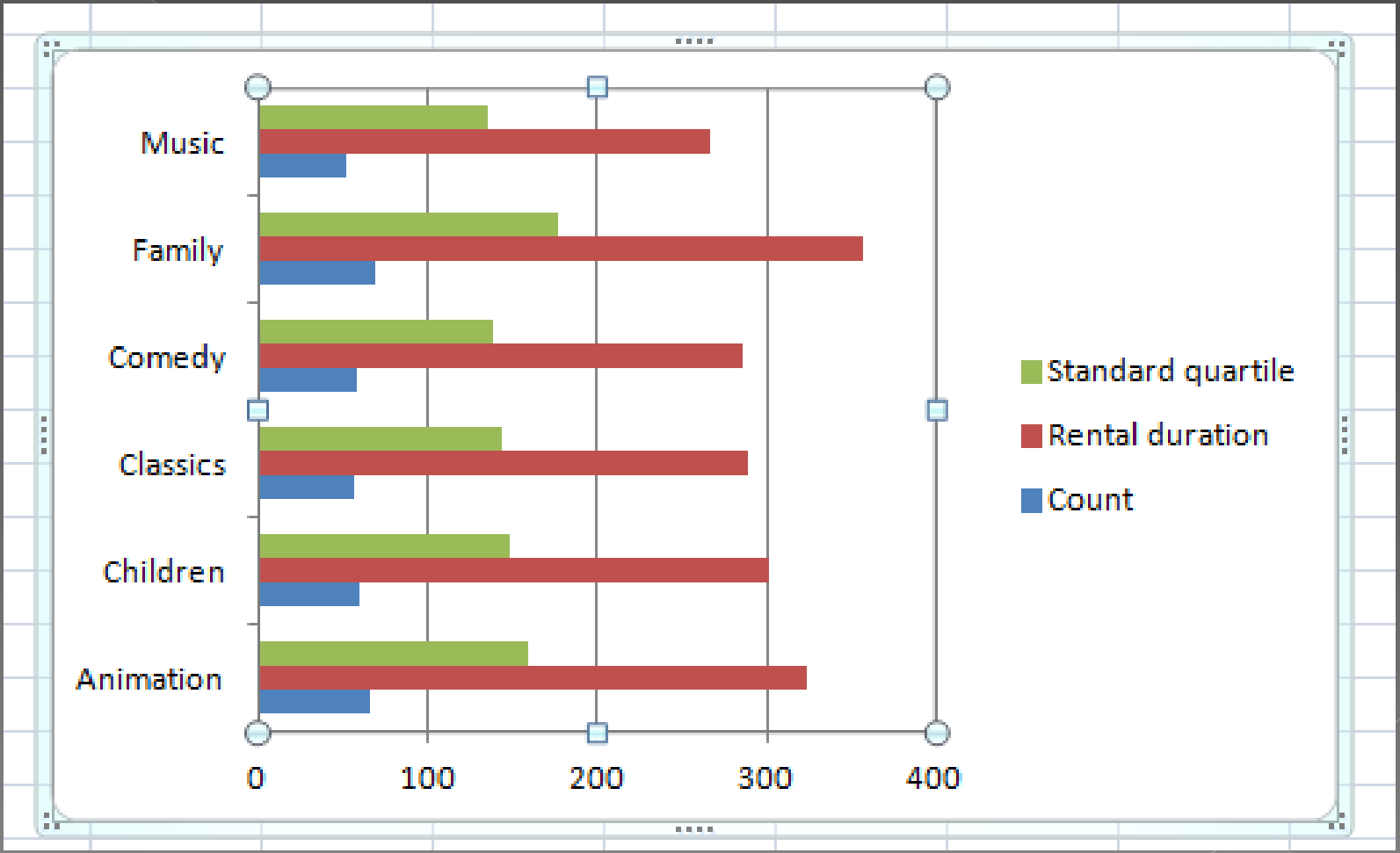
Workspace I question 2

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.

Data analysis

Fim category	Count	Rental duration	Standard quartile
Animation	66	323	159
Children	60	302	148
Classics	57	289	144
Comedy	58	286	138
Family	69	357	177
Music	51	267	135

Visualization



Interpretation

Family has the highest rental duration (357) followed by Animation (323) and once again, Music sits at the bottom (267)

Workspace 1 Question 3

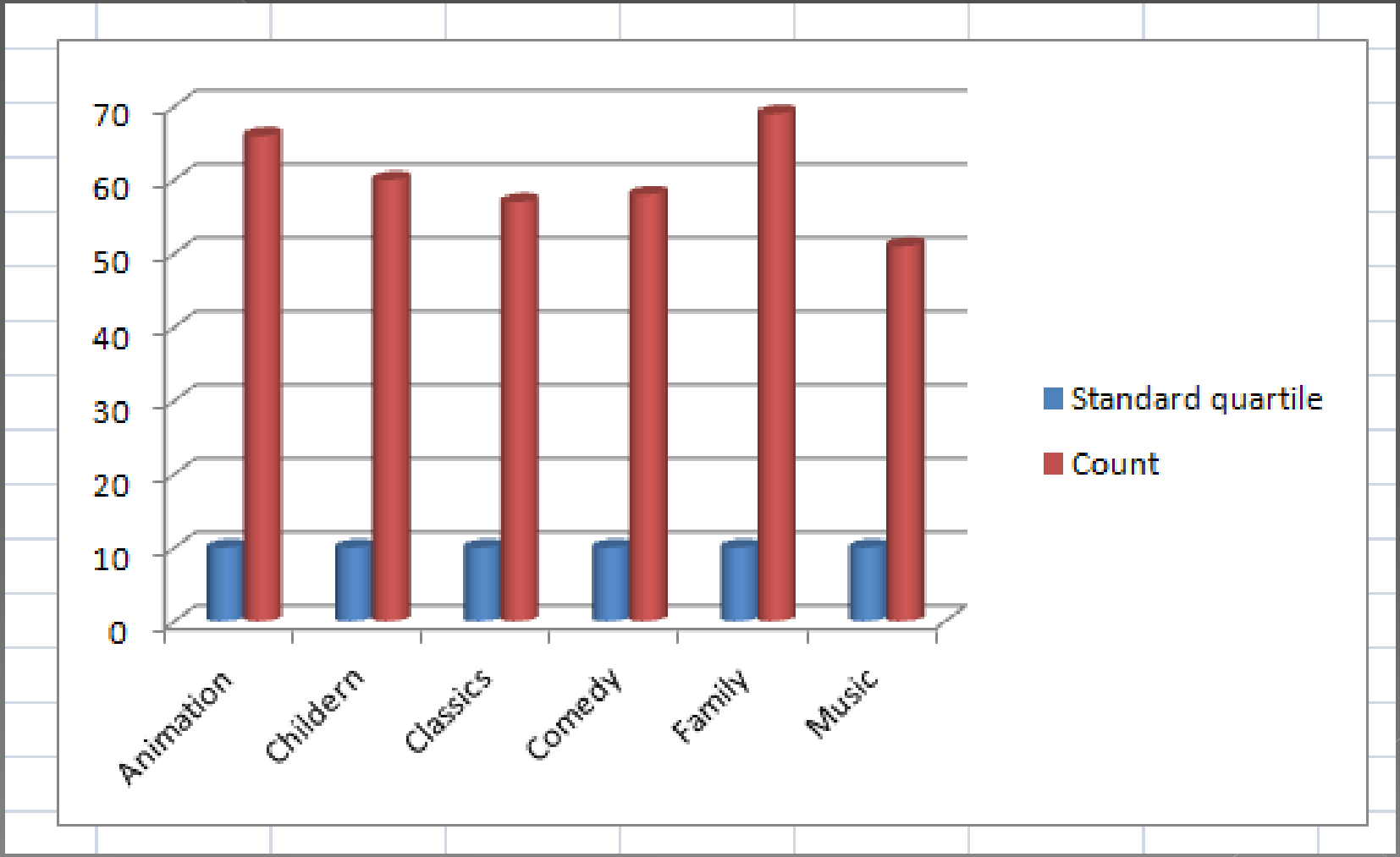
- Finally, provide a table with the family-friendly film category, each of the quartiles, and the corresponding count of movies within each combination of film category for each corresponding rental duration category. The resulting table should have three columns:

- a. Category
- b. Rental length category
- c. Count

Data analysis I

Category	Standard quartile	Count
Animation	10	66
Childern	10	60
Classics	10	57
Comedy	10	58
Family	10	69
Music	10	51

Visualization



Interpretation

- All categories has equal standard quartile (rental length) of 10.
- Family has the highest count of 69 followed by Animation.
- Music has the lowest count of 51.

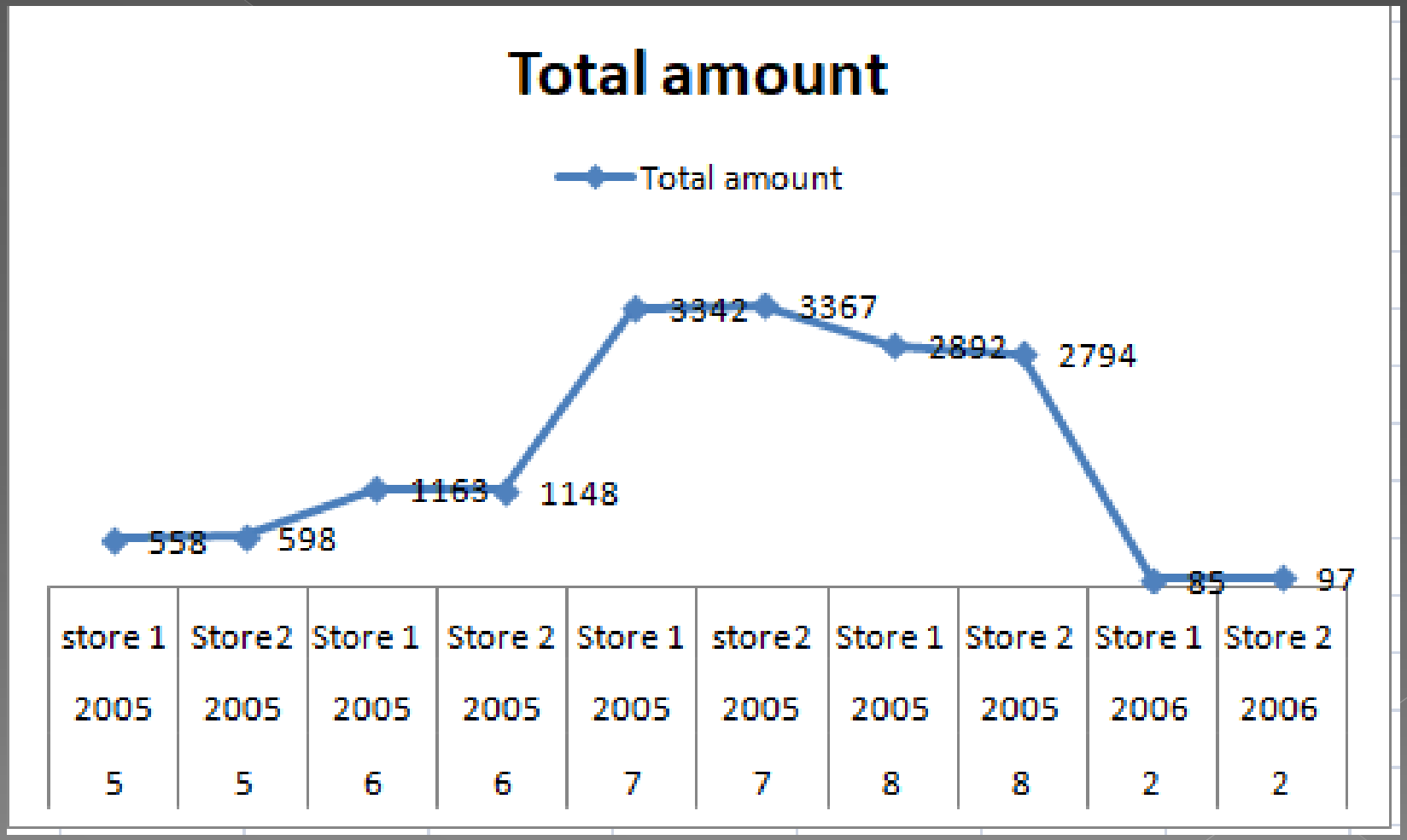
Workspace 2 question 1

We want to find out how the two stores compare in their count of rental orders during every month for all the years we have data for. Write a query that returns the store ID for the store, the year and month and the number of rental orders each store has fulfilled for that month. Your table should include a column for each of the following: year, month, store ID and count of rental orders fulfilled during that month

Data analysis

rental_month	rental_year	store	retail_count
5	2005	Store 1	558
5	2005	Store 2	598
6	2005	Store 1	1163
6	2005	Store 2	1148
7	2005	Store 1	3342
7	2005	Store 2	3367
8	2005	Store 1	2892
8	2005	Store 2	2794
2	2006	Store 1	85
2	2006	Store 2	97

Visualization



Interpretation

- More requests came in 2005 to 2006.
- Majority of request in 2005 was in Q2 and Q3 of the year.
- Requests in 2006 happened earlier (Q1).

Workspace 2 question 2

- 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

Data analysis I

Pay_month	Full_name	pay_countpermon	pay_amount
2007-02-01T00:00:00.000Z	Ana Bradley	4	19.96
2007-03-01T00:00:00.000Z	Ana Bradley	16	71.84
2007-04-01T00:00:00.000Z	Ana Bradley	12	72.88
2007-05-01T00:00:00.000Z	Ana Bradley	1	2.99
2007-02-01T00:00:00.000Z	Clara Shaw	6	22.94
2007-03-01T00:00:00.000Z	Clara Shaw	16	72.84
2007-04-01T00:00:00.000Z	Clara Shaw	18	93.82
2007-02-01T00:00:00.000Z	Curtis Irby	6	22.94
2007-03-01T00:00:00.000Z	Curtis Irby	17	86.83
2007-04-01T00:00:00.000Z	Curtis Irby	14	54.86
2007-05-01T00:00:00.000Z	Curtis Irby	1	2.99
2007-02-01T00:00:00.000Z	Eleanor Hunt	5	22.95
2007-03-01T00:00:00.000Z	Eleanor Hunt	18	87.82
2007-04-01T00:00:00.000Z	Eleanor Hunt	22	100.78
2007-02-01T00:00:00.000Z	Karl Seal	9	41.91
2007-03-01T00:00:00.000Z	Karl Seal	13	76.87
2007-04-01T00:00:00.000Z	Karl Seal	20	89.8
2007-02-01T00:00:00.000Z	Marcia Dean	8	37.92
2007-03-01T00:00:00.000Z	Marcia Dean	10	53.9
2007-04-01T00:00:00.000Z	Marcia Dean	20	73.8
2007-05-01T00:00:00.000Z	Marcia Dean	1	0.99
2007-02-01T00:00:00.000Z	Marion Snyder	8	44.92
2007-03-01T00:00:00.000Z	Marion Snyder	12	58.88
2007-04-01T00:00:00.000Z	Marion Snyder	18	85.82

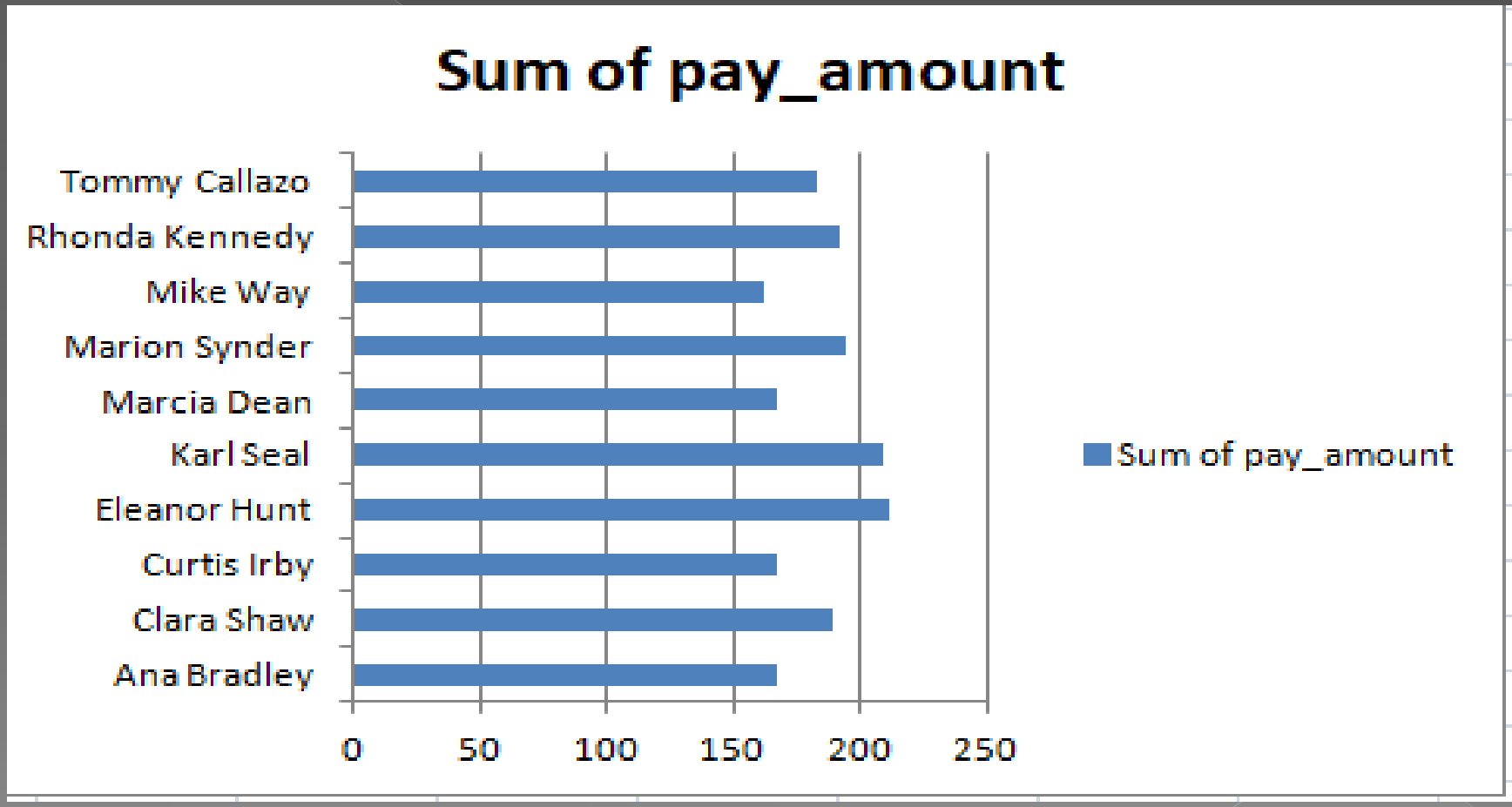
Data analysis II

26	2007-05-01T00:00:00.000Z	Marion Snyder	1	4.99
27	2007-02-01T00:00:00.000Z	Mike Way	6	35.94
28	2007-03-01T00:00:00.000Z	Mike Way	15	64.85
29	2007-04-01T00:00:00.000Z	Mike Way	12	61.88
30	2007-02-01T00:00:00.000Z	Rhonda Kennedy	4	19.96
31	2007-03-01T00:00:00.000Z	Rhonda Kennedy	15	74.85
32	2007-04-01T00:00:00.000Z	Rhonda Kennedy	19	96.81
33	2007-02-01T00:00:00.000Z	Tommy Collazo	7	25.93
34	2007-03-01T00:00:00.000Z	Tommy Collazo	12	67.88
35	2007-04-01T00:00:00.000Z	Tommy Collazo	18	89.82

Data analysis III

Row Labels	Sum of pay_amount
Ana Bradley	167.67
Clara Shaw	189.6
Curtis Irby	167.62
Eleanor Hunt	211.55
Karl Seal	208.58
Marcia Dean	166.61
Marion Snyder	194.61
Mike Way	162.67
Rhonda Kennedy	191.62
Tommy Collazo	183.63
Grand Total	**1844.16**

Visualization



Interpretation

- Eleanor Hunt has the highest record of sum_amount (211.55).
- Mike Way has the least record of sum_payment (162.67)

Workspace 2 Question 3

Finally, for each of these top 10 paying customers, I would like to find out the difference across their monthly payments during 2007. Please go ahead and write a query to compare the payment amounts in each successive month. Repeat this for each of these 10 paying customers. Also, it will be tremendously helpful if you can identify the customer name who paid the most difference in terms of payments.

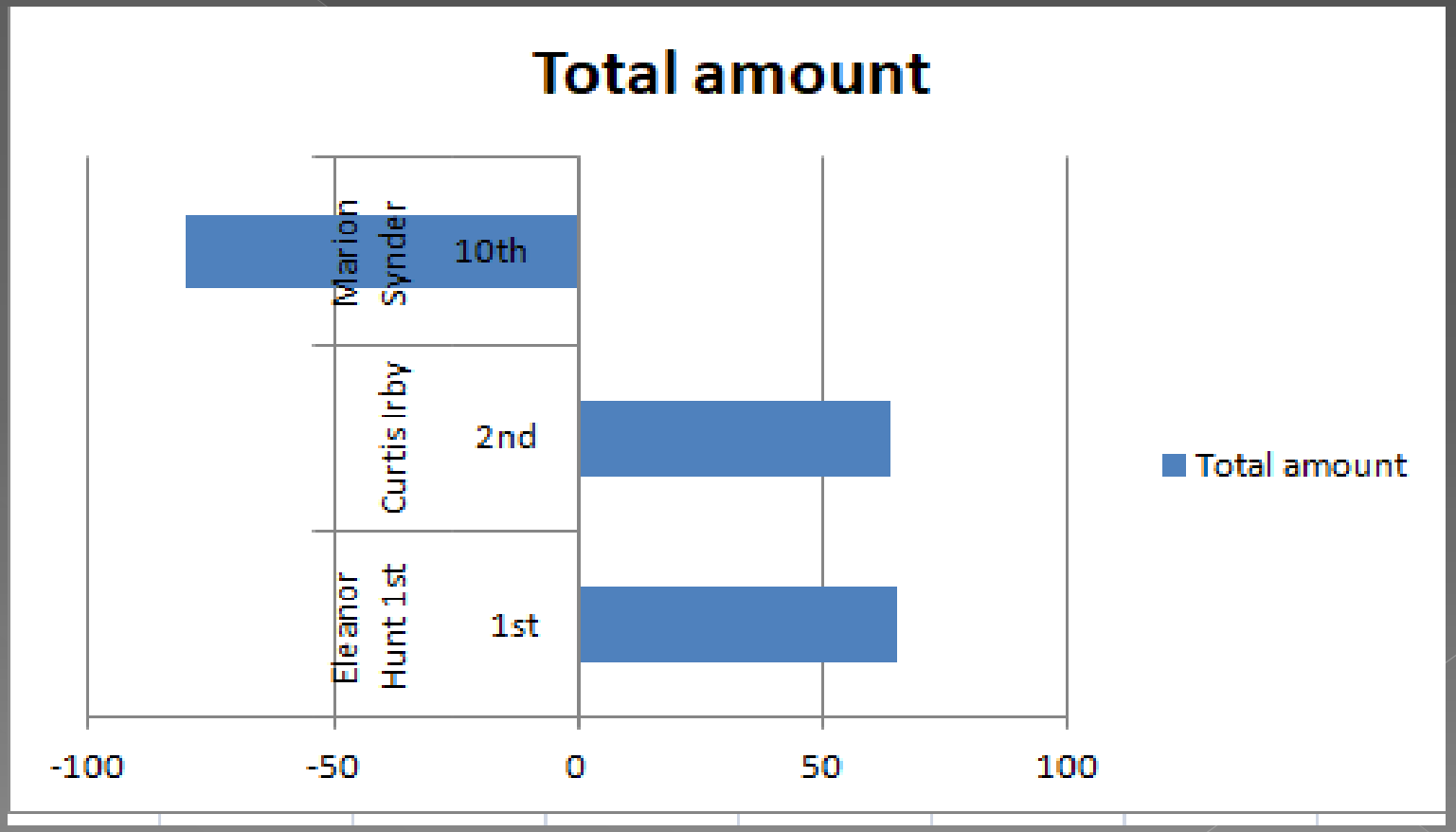
Data analysis I

Name	payment_month	payment_year	count	sum	total	lead	lead_diff
Ana Bradley	2	2007	4	19.96	19.96	71.84	51.88
Ana Bradley	3	2007	16	71.84	71.84	72.88	1.04
Ana Bradley	4	2007	12	72.88	72.88	2.99	-69.89
Ana Bradley	5	2007	1	2.99	2.99		
Clara Shaw	2	2007	6	22.94	22.94	72.84	49.9
Clara Shaw	3	2007	16	72.84	72.84	93.82	20.98
Clara Shaw	4	2007	18	93.82	93.82		
Curtis Irby	2	2007	6	22.94	22.94	86.83	63.89
Curtis Irby	3	2007	17	86.83	86.83	54.86	-31.97
Curtis Irby	4	2007	14	54.86	54.86	2.99	-51.87
Curtis Irby	5	2007	1	2.99	2.99		
Eleanor Hunt	2	2007	5	22.95	22.95	87.82	64.87
Eleanor Hunt	3	2007	18	87.82	87.82	100.78	12.96
Eleanor Hunt	4	2007	22	100.78	100.78		
Karl Seal	2	2007	9	41.91	41.91	76.87	34.96
Karl Seal	3	2007	13	76.87	76.87	89.8	12.93
Karl Seal	4	2007	20	89.8	89.8		
Marcia Dean	2	2007	8	37.92	37.92	53.9	15.98
Marcia Dean	3	2007	10	53.9	53.9	73.8	19.9
Marcia Dean	4	2007	20	73.8	73.8	0.99	-72.81
Marcia Dean	5	2007	1	0.99	0.99		
Marion Snyder	2	2007	8	44.92	44.92	58.88	13.96
Marion Snyder	3	2007	12	58.88	58.88	85.82	26.94

Data analysis II

Marion Snyder	4	2007	18	85.82	85.82	4.99	-80.83
Marion Snyder	5	2007	1	4.99	4.99		
Mike Way	2	2007	6	35.94	35.94	64.85	28.91
Mike Way	3	2007	15	64.85	64.85	61.88	-2.97
Mike Way	4	2007	12	61.88	61.88		
Rhonda Kennedy	2	2007	4	19.96	19.96	74.85	54.89
Rhonda Kennedy	3	2007	15	74.85	74.85	96.81	21.96
Rhonda Kennedy	4	2007	19	96.81	96.81		
Tommy Collazo	2	2007	7	25.93	25.93	67.88	41.95
Tommy Collazo	3	2007	12	67.88	67.88	89.82	21.94
Tommy Collazo	4	2007	18	89.82	89.82		

Visualization



Interpretation

- Eleanor Hunt has the highest cash difference within the period.
- Second on the list is Curtis Irby.
- Least on the list is Marion Snyder