Home Assignment 3 - MØLBA3001
Data Engineering

Nutan Gupta, Student ID: 274037
Geraldine Soto, Student ID: 274032

Exercise 1

You are now going to explore the Sakila dataset. Can you find for which years and months we have data for payments?

```
SELECT DISTINCT
   YEAR(payment_date) AS payment_year,
   MONTH(payment_date) AS payment_month
FROM payment
ORDER BY payment_year,payment_month ASC;
```

Result:

	payment_year	payment_mon...
	2005	5
	2005	6
	2005	7
	2005	8
	2006	2

Exercise 2

a) Can you find how many customers there are in the database, by using the count function?

```
SELECT count(*) AS total_count
FROM customer;
```

Result:

	total_count
	599

b) Can you also find how many customers have last names containing 'OO' (the letter O)?

```
SELECT COUNT(*)  AS number_of_customer
FROM customer
WHERE last_name REGEXP 'OO';
```

Result:

number_of_custo...
8

Exercise 3

Can you find the first and last rental date for Patricia Johnson who has customer_id = 2? Present the first name, last name, first rental date, and last rental date.

```
SELECT
   c.first_name,
   c.last_name,
   MIN(r.rental_date) AS first_rental_date,
   MAX(r.rental_date) AS last_rental_date
FROM customer c
JOIN rental r ON c.customer_id = r.customer_id
WHERE c.customer_id = 2
GROUP BY c.first_name, c.last_name;
```

Result:

first_name	last_name	first_rental_date	last_rental_date
PATRICIA	JOHNSON	2005-05-27 00:09:24	2005-08-23 17:39:35

Exercise 4

How many rentals are there per week? In which week did we see the most rentals?Present the year, week, and count.

```
SELECT WEEK(rental_date) AS rental_week,
count(*) AS rental_count
FROM rental
GROUP BY rental_week;
```

Result:

rental_week	rental_count
21	681
22	475
24	1357
25	954
27	2017
28	1436
30	2577
31	1993
33	2489
34	1883
7	182

```
SELECT
   YEAR(rental_date) AS rental_year,
   WEEK(rental_date) AS rental_week,
   COUNT(*) AS rental_count
FROM rental
GROUP BY rental_year, rental_week
ORDER BY rental_count DESC
LIMIT 1;
```

Result:

	rental_year	rental_week	rental_count
	2005	30	2577

Exercise 5

Which customers rented the movie "TRAFFIC HOBBIT"? Provide the first name and last name of the customers. (Here, you must join four tables)

```
SELECT
   c.first_name,
   c.last_name,
   f.title
FROM customer c
JOIN rental r ON c.customer_id = r.customer_id
JOIN inventory i ON r.inventory_id = i.inventory_id
JOIN film f ON i.film_id = f.film_id
WHERE f.title = 'TRAFFIC HOBBIT';
```

Result:

	first_name	last_name	title
	MAXINE	SILVA	TRAFFIC HOBBIT
	LEE	HAWKS	TRAFFIC HOBBIT
	CHRISTOPHER	GRECO	TRAFFIC HOBBIT
	JULIE	SANCHEZ	TRAFFIC HOBBIT
	JO	FOWLER	TRAFFIC HOBBIT

Exercise 6

Which five movies have been rented more than 31 times? Provide the title and the number of times the movie has been rented. (Here, you must join three tables)

```
SELECT
   f.title,
   COUNT(*) AS rental_count
FROM rental r
JOIN inventory i ON r.inventory_id = i.inventory_id
JOIN film f ON i.film_id = f.film_id
GROUP BY f.title
HAVING rental_count > 31
ORDER BY rental_count DESC;
```

Result:

	title	rental_count
	BUCKET BROTHERHOOD	34
	ROCKETEER MOTHER	33
	FORWARD TEMPLE	32
	GRIT CLOCKWORK	32
	JUGGLER HARDLY	32
	RIDGEMONT SUBMARINE	32
	SCALAWAG DUCK	32

Exercise 7

The analyst in your company is in need of a table to fulfil an analyst task and has come to you to make a query. The table should have four columns:
1) ”customer_id”. This is the customer ID.
2) “customer_name”. This is the customer’s name on the format “WILLIAMS, LINDA”.
3) “rental_date”. This is the date and time when the customer rented a movie.
4) “previous_rental”. This is the previous date and time when the customer rented a movie.

```
USE sakila;
SHOW TABLES;
SELECT
   c.customer_id,
   CONCAT(UPPER(c.last_name), ', ', UPPER(c.first_name)) AS customer_name,
   r.rental_date,
   (
      SELECT MAX(r2.rental_date)
      FROM rental r2
      WHERE r2.customer_id = r.customer_id
      AND r2.rental_date < r.rental_date
   ) AS previous_rental
FROM
   customer c
JOIN
   rental r ON c.customer_id = r.customer_id
ORDER BY
   c.customer_id, r.rental_date;
```

Result:

customer_id	customer_name	rental_date	previous_rental
3	WILLIAMS, LINDA	2005-05-29 22:43:55	2005-05-27 17:17:09
3	WILLIAMS, LINDA	2005-06-16 01:34:05	2005-05-29 22:43:55
3	WILLIAMS, LINDA	2005-06-16 15:19:10	2005-06-16 01:34:05
3	WILLIAMS, LINDA	2005-06-17 05:15:15	2005-06-16 15:19:10
3	WILLIAMS, LINDA	2005-06-19 08:34:53	2005-06-17 05:15:15
3	WILLIAMS, LINDA	2005-07-07 10:23:25	2005-06-19 08:34:53
3	WILLIAMS, LINDA	2005-07-08 12:47:11	2005-07-07 10:23:25
3	WILLIAMS, LINDA	2005-07-27 04:54:42	2005-07-08 12:47:11
3	WILLIAMS, LINDA	2005-07-27 20:23:12	2005-07-27 04:54:42
3	WILLIAMS, LINDA	2005-07-28 03:59:21	2005-07-27 20:23:12
3	WILLIAMS, LINDA	2005-07-28 04:46:30	2005-07-28 03:59:21
3	WILLIAMS, LINDA	2005-07-28 11:46:45	2005-07-28 04:46:30
3	WILLIAMS, LINDA	2005-07-28 18:17:14	2005-07-28 11:46:45
3	WILLIAMS, LINDA	2005-07-29 11:07:04	2005-07-28 18:17:14
3	WILLIAMS, LINDA	2005-07-30 13:31:20	2005-07-29 11:07:04
3	WILLIAMS, LINDA	2005-07-30 21:45:46	2005-07-30 13:31:20
3	WILLIAMS, LINDA	2005-07-31 03:27:58	2005-07-30 21:45:46
3	WILLIAMS, LINDA	2005-07-31 11:32:58	2005-07-31 03:27:58
3	WILLIAMS, LINDA	2005-08-01 14:19:48	2005-07-31 11:32:58
3	WILLIAMS, LINDA	2005-08-18 14:49:55	2005-08-01 14:19:48
3	WILLIAMS, LINDA	2005-08-19 22:18:07	2005-08-18 14:49:55
3	WILLIAMS, LINDA	2005-08-20 06:14:12	2005-08-19 22:18:07
3	WILLIAMS, LINDA	2005-08-21 20:50:48	2005-08-20 06:14:12
3	WILLIAMS, LINDA	2005-08-22 09:37:27	2005-08-21 20:50:48
3	WILLIAMS, LINDA	2005-08-23 07:10:14	2005-08-22 09:37:27

Exercise 8 (optional)

Add a column to the query in Exercise 7. This column should contain the number of hours since the last rental. Call the column "hours_since_last_rental." (here, you can use the timestamp diff function)

```
SELECT
   c.customer_id,
   CONCAT(UPPER(c.last_name), ', ', UPPER(c.first_name)) AS customer_name,
   r.rental_date,
   (
      SELECT MAX(r2.rental_date)
      FROM rental r2
      WHERE r2.customer_id = r.customer_id
      AND r2.rental_date < r.rental_date
   ) AS previous_rental,
   TIMESTAMPDIFF(HOUR, (
      SELECT MAX(r2.rental_date)
      FROM rental r2
      WHERE r2.customer_id = r.customer_id
      AND r2.rental_date < r.rental_date
```

```
    ), r.rental_date) AS hours_since_last_rental
FROM
    customer c
JOIN
    rental r ON c.customer_id = r.customer_id
ORDER BY
    c.customer_id, r.rental_date;
```

Result:

customer_id	customer_name	rental_date	previous_rental	hours_since_last_rental
3	WILLIAMS, LINDA	2005-05-29 22:43:55	2005-05-27 17:17:09	53
3	WILLIAMS, LINDA	2005-06-16 01:34:05	2005-05-29 22:43:55	410
3	WILLIAMS, LINDA	2005-06-16 15:19:10	2005-06-16 01:34:05	13
3	WILLIAMS, LINDA	2005-06-17 05:15:15	2005-06-16 15:19:10	13
3	WILLIAMS, LINDA	2005-06-19 08:34:53	2005-06-17 05:15:15	51
3	WILLIAMS, LINDA	2005-07-07 10:23:25	2005-06-19 08:34:53	433
3	WILLIAMS, LINDA	2005-07-08 12:47:11	2005-07-07 10:23:25	26
3	WILLIAMS, LINDA	2005-07-27 04:54:42	2005-07-08 12:47:11	448
3	WILLIAMS, LINDA	2005-07-27 20:23:12	2005-07-27 04:54:42	15
3	WILLIAMS, LINDA	2005-07-28 03:59:21	2005-07-27 20:23:12	7
3	WILLIAMS, LINDA	2005-07-28 04:46:30	2005-07-28 03:59:21	0
3	WILLIAMS, LINDA	2005-07-28 11:46:45	2005-07-28 04:46:30	7
3	WILLIAMS, LINDA	2005-07-28 18:17:14	2005-07-28 11:46:45	6
3	WILLIAMS, LINDA	2005-07-29 11:07:04	2005-07-28 18:17:14	16
3	WILLIAMS, LINDA	2005-07-30 13:31:20	2005-07-29 11:07:04	26
3	WILLIAMS, LINDA	2005-07-30 21:45:46	2005-07-30 13:31:20	8
3	WILLIAMS, LINDA	2005-07-31 03:27:58	2005-07-30 21:45:46	5
3	WILLIAMS, LINDA	2005-07-31 11:32:58	2005-07-31 03:27:58	8
3	WILLIAMS, LINDA	2005-08-01 14:19:48	2005-07-31 11:32:58	26
3	WILLIAMS, LINDA	2005-08-18 14:49:55	2005-08-01 14:19:48	408
3	WILLIAMS, LINDA	2005-08-19 22:18:07	2005-08-18 14:49:55	31
3	WILLIAMS, LINDA	2005-08-20 06:14:12	2005-08-19 22:18:07	7
3	WILLIAMS, LINDA	2005-08-21 20:50:48	2005-08-20 06:14:12	38
3	WILLIAMS, LINDA	2005-08-22 09:37:27	2005-08-21 20:50:48	12
3	WILLIAMS, LINDA	2005-08-23 07:10:14	2005-08-22 09:37:27	21