MOTION PICTURES DATA ANALYSIS

MySQL CASE STUDY

BY - TEJUSVI SARDANA

AGENDA

Introduction

Database used

Business Tasks

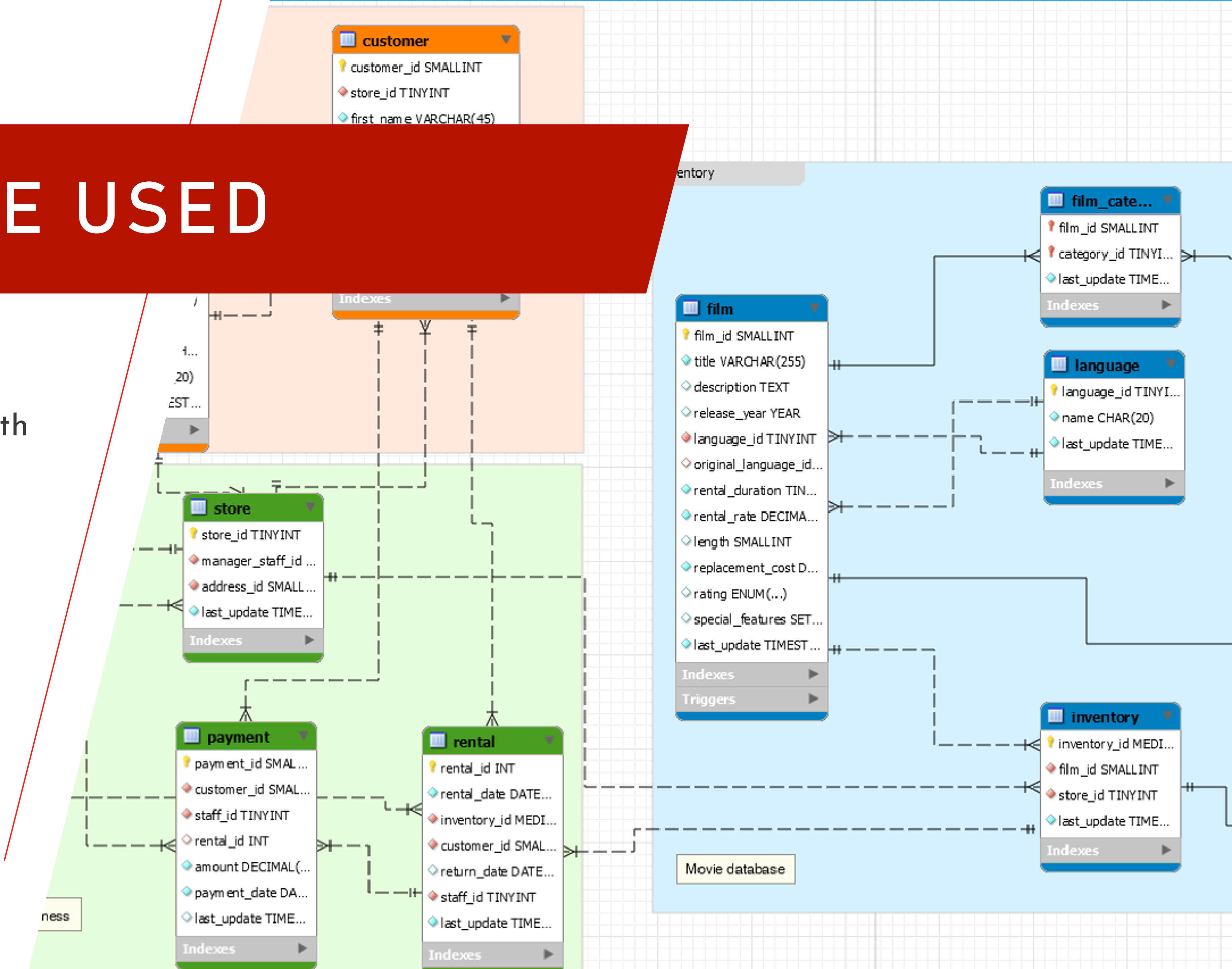
Summary

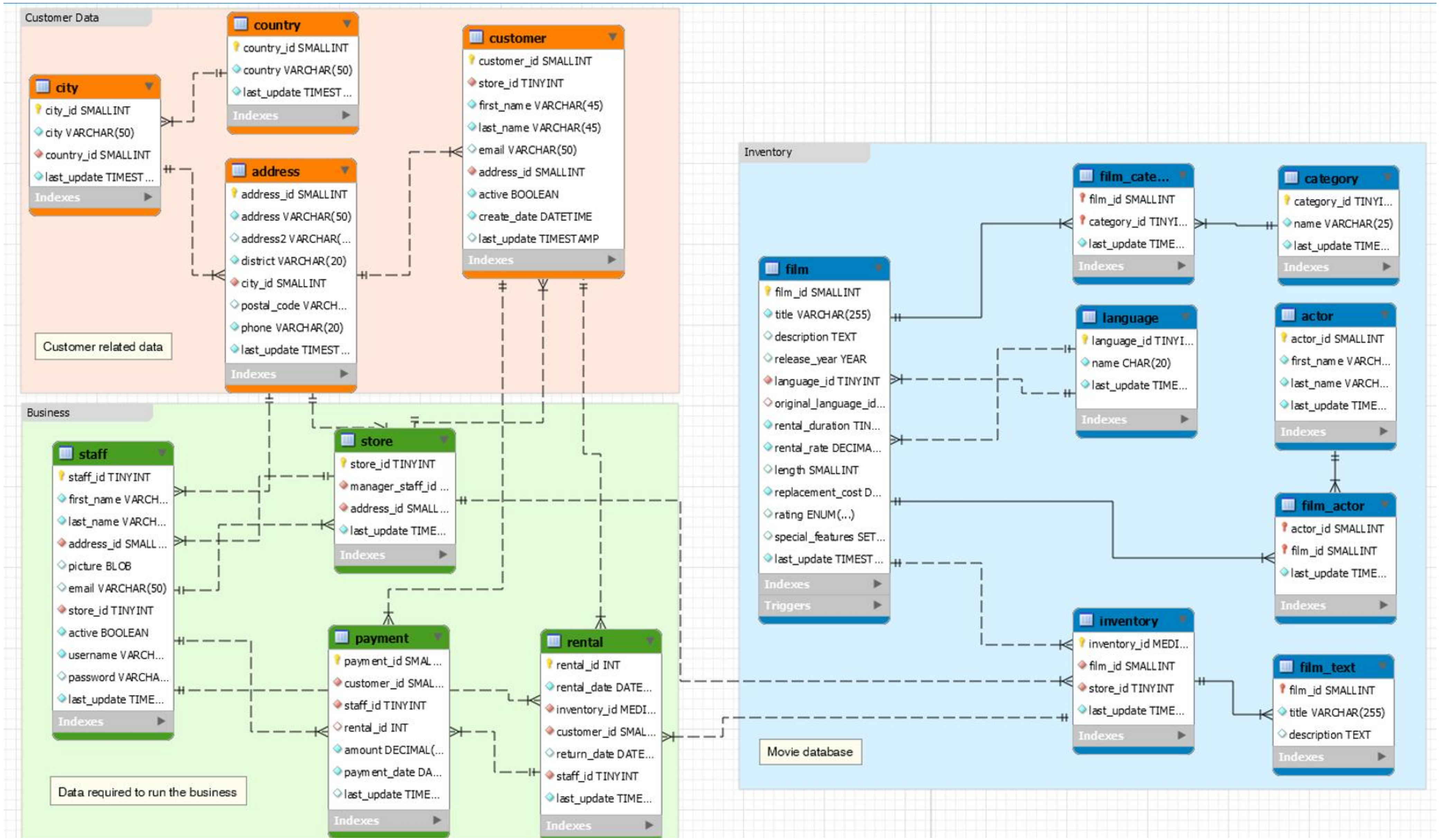
INTRODUCTION

- The Motion Picture Association (MPA) is a movie rating system.
- So, as an Analyst I've to analyze the data and help them stock up the inventory of movies to improve their business.

DATABASE USED

Sakila database i.e. movie rental store with vast collection of movies.





Customer Data

country
country_id SMALLINT
country VARCHAR(50)
last_update TIMEST...
Indexes

customer
customer_id SMALLINT
store_id TINYINT
first_name VARCHAR(45)
last_name VARCHAR(45)
email VARCHAR(50)
address_id SMALLINT
active BOOLEAN
create_date DATETIME
last_update TIMESTAMP
Indexes

city
city_id SMALLINT
city VARCHAR(50)
country_id SMALLINT
last_update TIMEST...
Indexes

address
address_id SMALLINT
address VARCHAR(50)
address2 VARCHAR(...
district VARCHAR(20)
city_id SMALLINT
postal_code VARCH...
phone VARCHAR(20)
last_update TIMEST...
Indexes

Customer related data

Business

staff
staff_id TINYINT
first_name VARCH...
last_name VARCH...
address_id SMALL...
picture BLOB
email VARCHAR(50)
store_id TINYINT
active BOOLEAN
username VARCH...
password VARCHA...
last_update TIME...
Indexes

store
store_id TINYINT
manager_staff_id ...
address_id SMALL...
last_update TIME...
Indexes

payment
payment_id SMAL...
customer_id SMAL...
staff_id TINYINT
rental_id INT
amount DECIMAL(...
payment_date DA...
last_update TIME...
Indexes

rental
rental_id INT
rental_date DATE...
inventory_id MEDI...
customer_id SMAL...
return_date DATE...
staff_id TINYINT
last_update TIME...
Indexes

Data required to run the business

Inventory

film_cate...
film_id SMALLINT
category_id TINYI...
last_update TIME...
Indexes

category
category_id TINYI...
name VARCHAR(25)
last_update TIME...
Indexes

film
film_id SMALLINT
title VARCHAR(255)
description TEXT
release_year YEAR
language_id TINYINT
original_language_id...
rental_duration TIN...
rental_rate DECIMA...
length SMALLINT
replacement_cost D...
rating ENUM(...)
special_features SET...
last_update TIMEST...
Indexes
Triggers

language
language_id TINYI...
name CHAR(20)
last_update TIME...
Indexes

actor
actor_id SMALLINT
first_name VARCH...
last_name VARCH...
last_update TIME...
Indexes

film_actor
actor_id SMALLINT
film_id SMALLINT
last_update TIME...
Indexes

inventory
inventory_id MEDI...
film_id SMALLINT
store_id TINYINT
last_update TIME...
Indexes

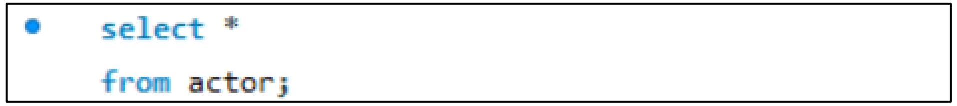
film_text
film_id SMALLINT
title VARCHAR(255)
description TEXT
Indexes

Movie database

TASK-1

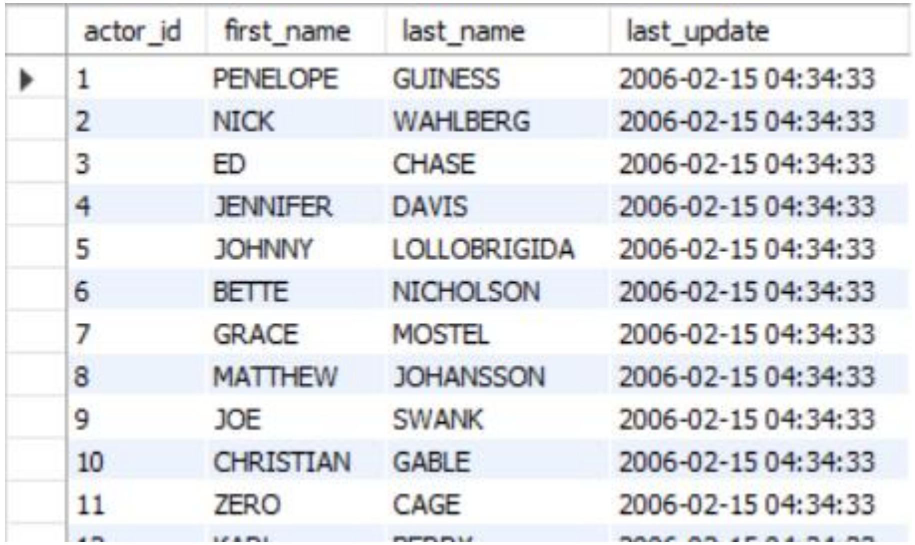
The Sakila rental store management wants to know the names of all the actors in their movie collection. **Display the first names, last names, actor IDs, and the details of the last updated column.**

Query:-



```
select *
from actor;
```

Output:-



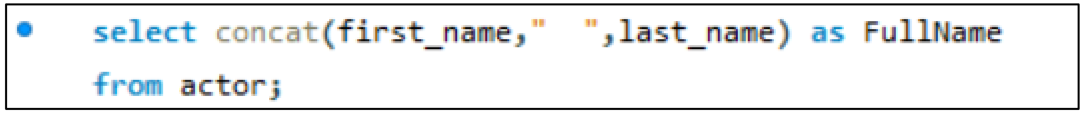
actor_id	first_name	last_name	last_update
1	PENELOPE	GUINESS	2006-02-15 04:34:33
2	NICK	WAHLBERG	2006-02-15 04:34:33
3	ED	CHASE	2006-02-15 04:34:33
4	JENNIFER	DAVIS	2006-02-15 04:34:33
5	JOHNNY	LOLLOBRIGIDA	2006-02-15 04:34:33
6	BETTE	NICHOLSON	2006-02-15 04:34:33
7	GRACE	MOSTEL	2006-02-15 04:34:33
8	MATTHEW	JOHANSSON	2006-02-15 04:34:33
9	JOE	SWANK	2006-02-15 04:34:33
10	CHRISTIAN	GABLE	2006-02-15 04:34:33
11	ZERO	CAGE	2006-02-15 04:34:33

TASK-2

Many actors have adopted attractive screen names mostly at the behest of producers and directors. The management wants to know the following:

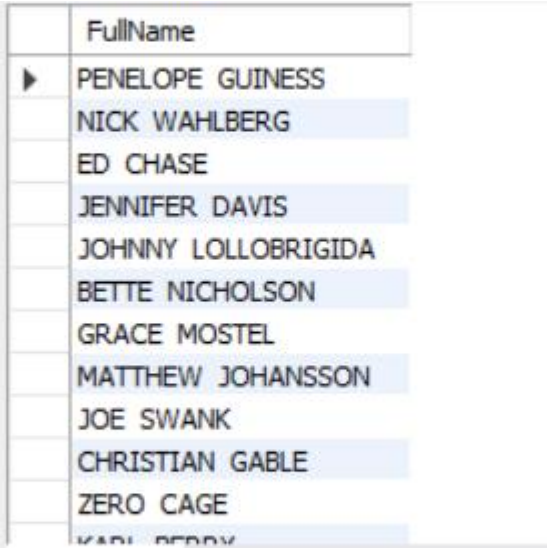
a. Display the full names of all actors.

Query:-



```
select concat(first_name,"  ",last_name) as FullName
from actor;
```

Output:-



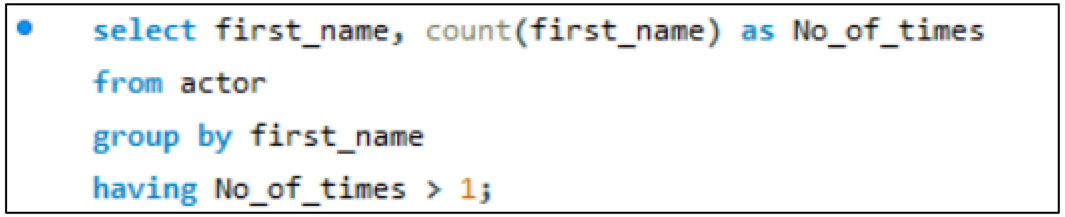
```
FullName
PENELOPE  GUINESS
NICK  WAHLBERG
ED  CHASE
JENNIFER  DAVIS
JOHNNY  LOLLOBRIGIDA
BETTE  NICHOLSON
GRACE  MOSTEL
MATTHEW  JOHANSSON
JOE  SWANK
CHRISTIAN  GABLE
ZERO  CAGE
```

TASK-2

Many actors have adopted attractive screen names mostly at the behest of producers and directors. The management wants to know the following:

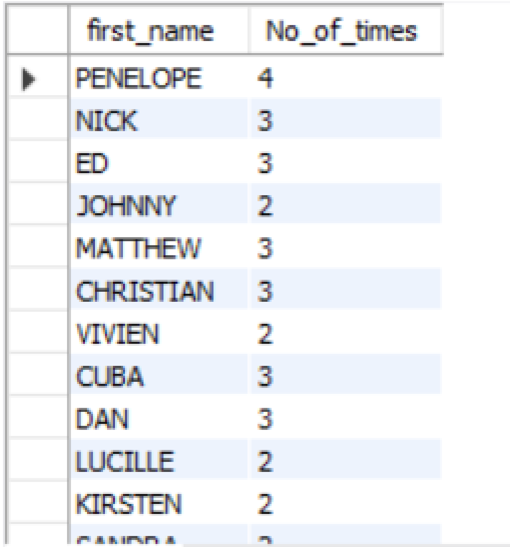
b. Display the first names of actors along with the count of repeated first names.

Query:-



```
select first_name, count(first_name) as No_of_times
from actor
group by first_name
having No_of_times > 1;
```

Output:-



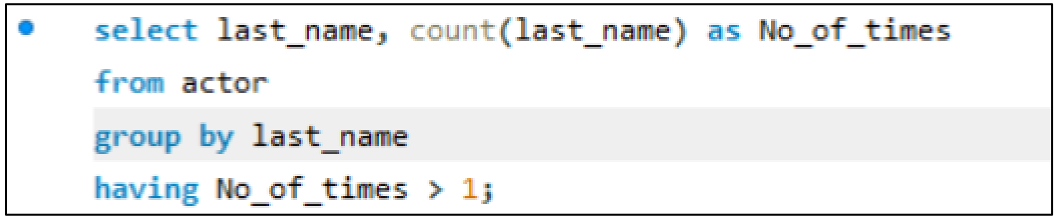
first_name	No_of_times
PENELOPE	4
NICK	3
ED	3
JOHNNY	2
MATTHEW	3
CHRISTIAN	3
VIVIEN	2
CUBA	3
DAN	3
LUCILLE	2
KIRSTEN	2
SANDRA	2

TASK-2

Many actors have adopted attractive screen names mostly at the behest of producers and directors. The management wants to know the following:

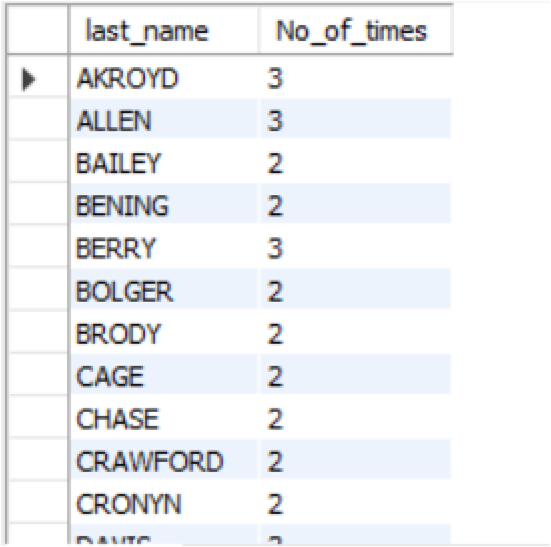
c. Display the first names of actors along with the count of repeated first names.

Query:-



```
select last_name, count(last_name) as No_of_times
from actor
group by last_name
having No_of_times > 1;
```

Output:-

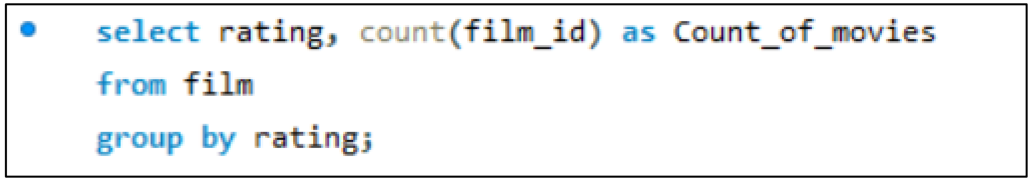


last_name	No_of_times
AKROYD	3
ALLEN	3
BAILEY	2
BENING	2
BERRY	3
BOLGER	2
BRODY	2
CAGE	2
CHASE	2
CRAWFORD	2
CRONYN	2
DAVIS	2

TASK-3

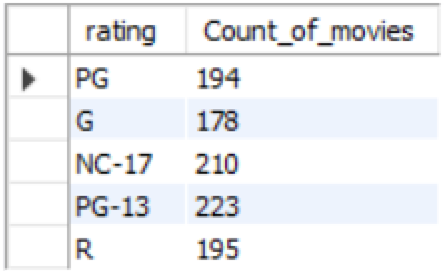
Display the count of movies grouped by the ratings.

Query:-



```
select rating, count(film_id) as Count_of_movies
from film
group by rating;
```

Output:-

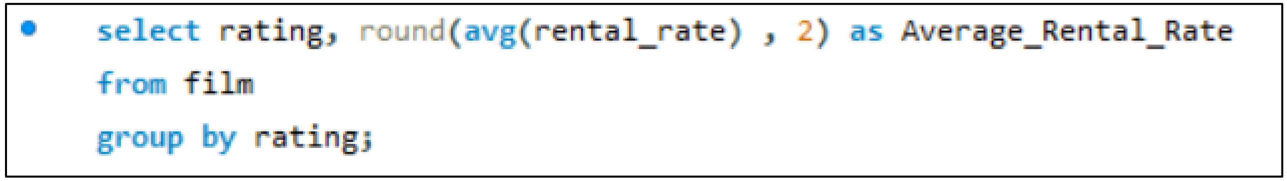


rating	Count_of_movies
PG	194
G	178
NC-17	210
PG-13	223
R	195

TASK-4

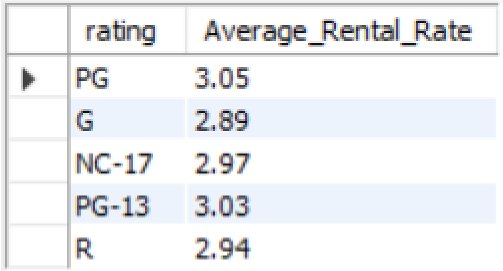
Calculate and Display the average rental rates based on the movie ratings.

Query:-



```
select rating, round(avg(rental_rate) , 2) as Average_Rental_Rate
from film
group by rating;
```

Output:-



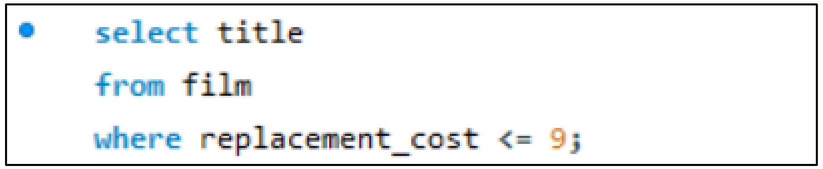
rating	Average_Rental_Rate
PG	3.05
G	2.89
NC-17	2.97
PG-13	3.03
R	2.94

TASK-5

The management wants the data about replacement cost of movies. Replacement cost is the amount of money required to replace an existing asset (DVD/blue ray disc) with an equally valued or similar asset at the current market price.

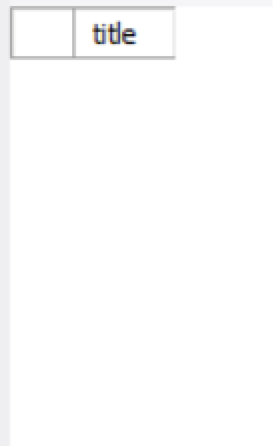
Part a) – Display the movie titles where the replacement cost is upto $9.

Query:-



```
select title
from film
where replacement_cost <= 9;
```

Output:-

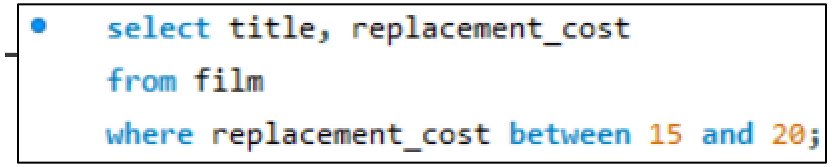


```
title
```

TASK-5

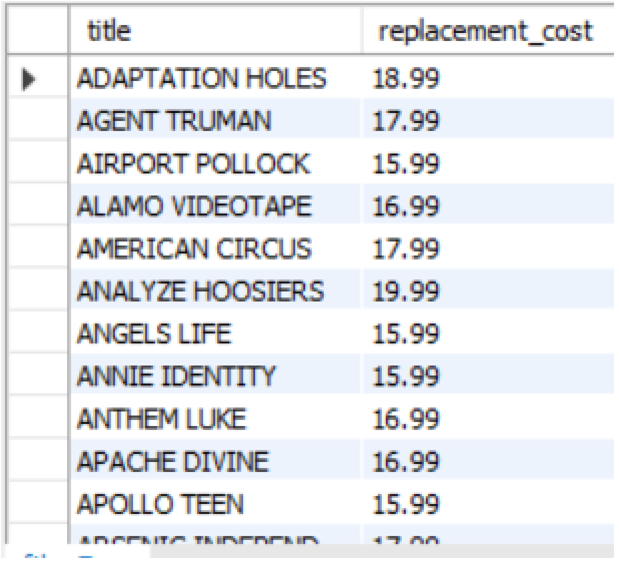
Part b) – Display the movie titles where the replacement cost is between 15$ and 20$.

Query:-



```
select title, replacement_cost
from film
where replacement_cost between 15 and 20;
```

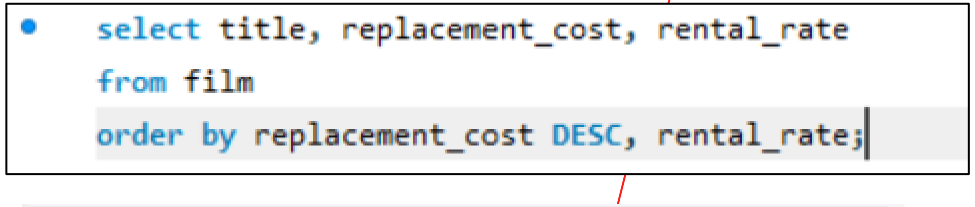
Output:-



title	replacement_cost
ADAPTATION HOLES	18.99
AGENT TRUMAN	17.99
AIRPORT POLLOCK	15.99
ALAMO VIDEOTAPE	16.99
AMERICAN CIRCUS	17.99
ANALYZE HOOSIERS	19.99
ANGELS LIFE	15.99
ANNIE IDENTITY	15.99
ANTHEM LUKE	16.99
APACHE DIVINE	16.99
APOLLO TEEN	15.99

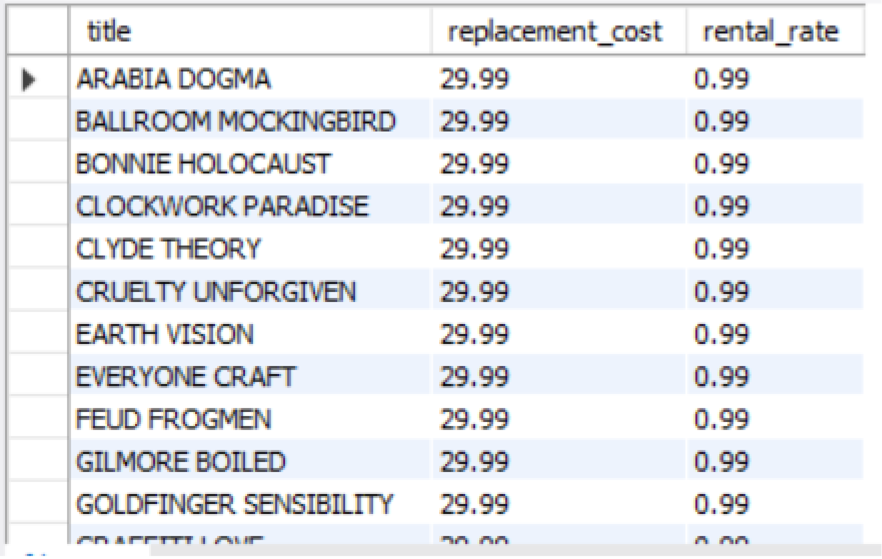
Part c) – Display the movie titles with the highest replacement cost and lowest rental cost.

Query:-



```
select title, replacement_cost, rental_rate
from film
order by replacement_cost DESC, rental_rate;
```

Output:-

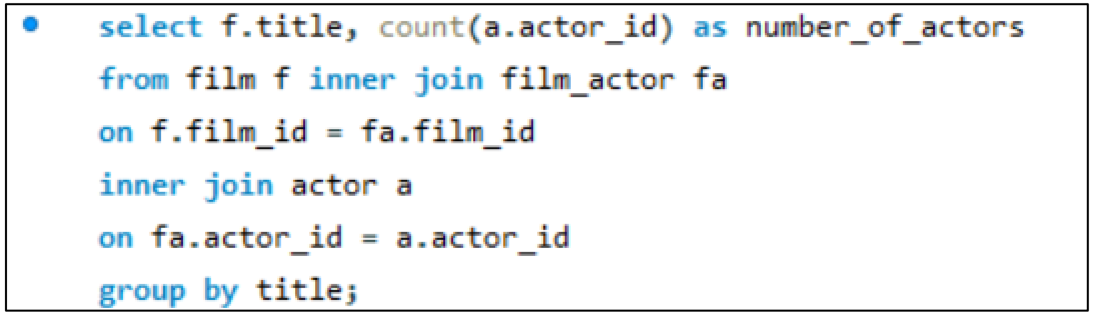


title	replacement_cost	rental_rate
ARABIA DOGMA	29.99	0.99
BALLROOM MOCKINGBIRD	29.99	0.99
BONNIE HOLOCAUST	29.99	0.99
CLOCKWORK PARADISE	29.99	0.99
CLYDE THEORY	29.99	0.99
CRUELTY UNFORGIVEN	29.99	0.99
EARTH VISION	29.99	0.99
EVERYONE CRAFT	29.99	0.99
FEUD FROGMEN	29.99	0.99
GILMORE BOILED	29.99	0.99
GOLDFINGER SENSIBILITY	29.99	0.99

TASK-6

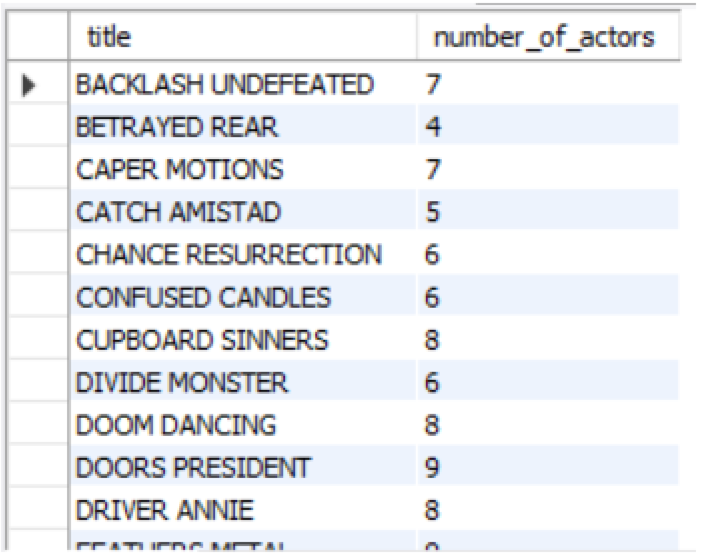
The management needs to know the **list all the movies along with the number of actors listed for each movie.**

Query:-



```
select f.title, count(a.actor_id) as number_of_actors
from film f inner join film_actor fa
on f.film_id = fa.film_id
inner join actor a
on fa.actor_id = a.actor_id
group by title;
```

Output:-

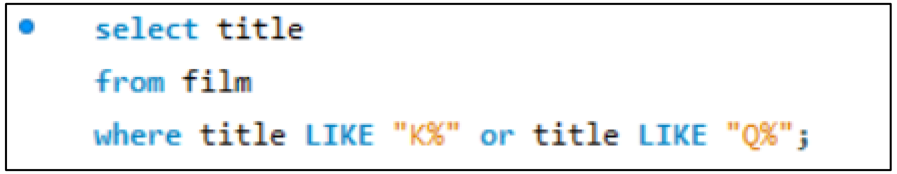


title	number_of_actors
BACKLASH UNDEFEATED	7
BETRAYED REAR	4
CAPER MOTIONS	7
CATCH AMISTAD	5
CHANCE RESURRECTION	6
CONFUSED CANDLES	6
CUPBOARD SINNERS	8
DIVIDE MONSTER	6
DOOM DANCING	8
DOORS PRESIDENT	9
DRIVER ANNIE	8
FEATHERS METAL	[illegible]

TASK-7

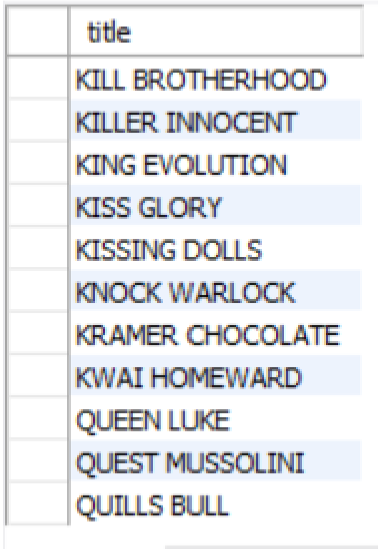
The Music of Queen and Kris Kristofferson has seen an unlikely resurgence. As an unintended consequence, movies starting with the letters 'K and 'Q" have also soared in popularity. **Display the movie titles starting with the letters "K' and 'Q'.**

Query:-



```
select title
from film
where title LIKE "K%" or title LIKE "Q%";
```

Output:-

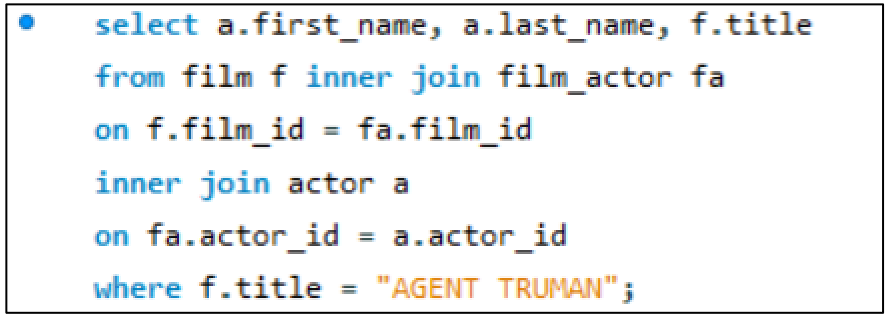


title
KILL BROTHERHOOD
KILLER INNOCENT
KING EVOLUTION
KISS GLORY
KISSING DOLLS
KNOCK WARLOCK
KRAMER CHOCOLATE
KWAI HOMEWARD
QUEEN LUKE
QUEST MUSSOLINI
QUILLS BULL

TASK-8

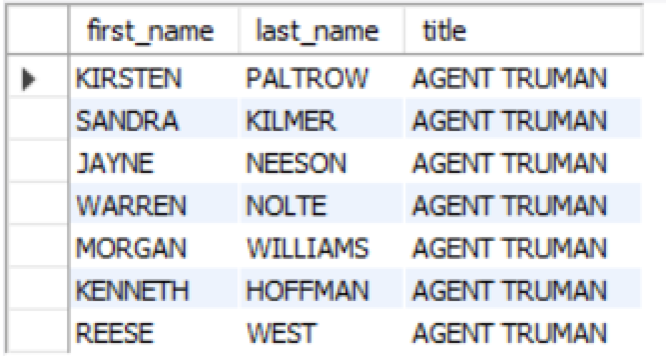
The movie ‘AGENT TRUMAN' has been a great success. **Display the first names and last names of all actors who are a part of this movie.**

Query:-



```
select a.first_name, a.last_name, f.title
from film f inner join film_actor fa
on f.film_id = fa.film_id
inner join actor a
on fa.actor_id = a.actor_id
where f.title = "AGENT TRUMAN";
```

Output:-

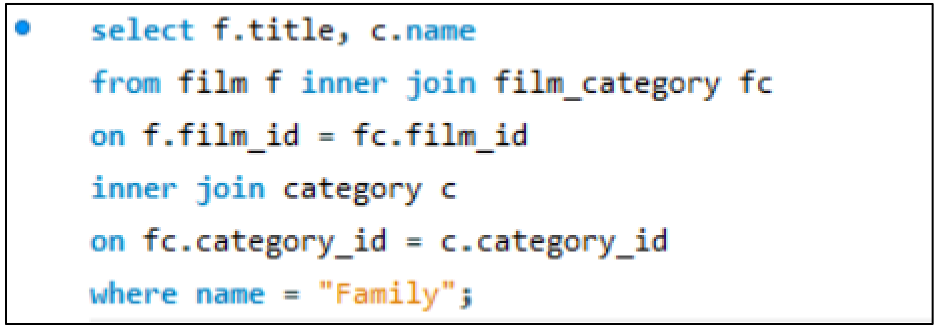


first_name	last_name	title
KIRSTEN	PALTROW	AGENT TRUMAN
SANDRA	KILMER	AGENT TRUMAN
JAYNE	NEESON	AGENT TRUMAN
WARREN	NOLTE	AGENT TRUMAN
MORGAN	WILLIAMS	AGENT TRUMAN
KENNETH	HOFFMAN	AGENT TRUMAN
REESE	WEST	AGENT TRUMAN

TASK-9

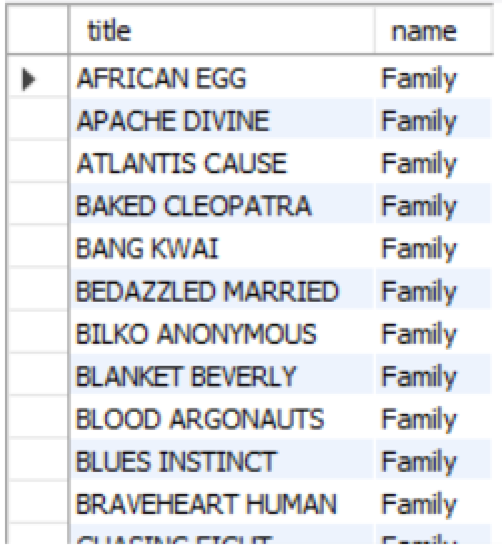
Sales has been down among the family audience with kids. The management wants to promote the movies that fall under the ‘children’ category. **Identify and display the names of the movies in the family category.**

Query:-



```
select f.title, c.name
from film f inner join film_category fc
on f.film_id = fc.film_id
inner join category c
on fc.category_id = c.category_id
where name = "Family";
```

Output:-

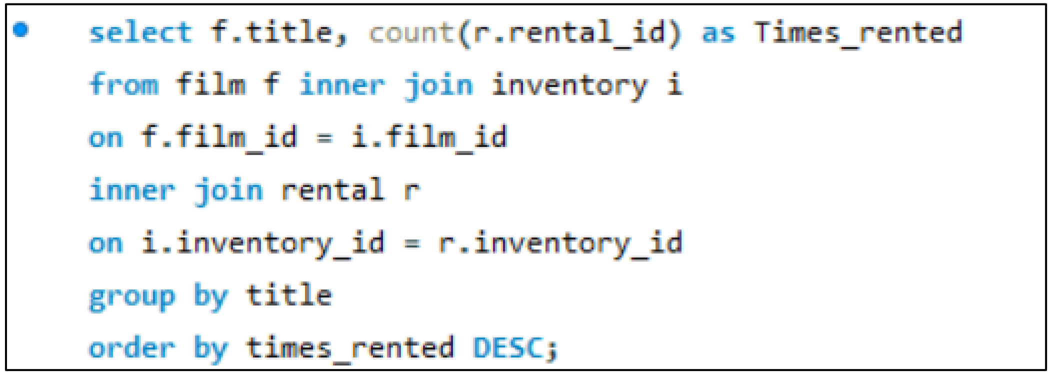


title	name
AFRICAN EGG	Family
APACHE DIVINE	Family
ATLANTIS CAUSE	Family
BAKED CLEOPATRA	Family
BANG KWAI	Family
BEDAZZLED MARRIED	Family
BILKO ANONYMOUS	Family
BLANKET BEVERLY	Family
BLOOD ARGONAUTS	Family
BLUES INSTINCT	Family
BRAVEHEART HUMAN	Family
CHASING FIGHT	Family

TASK-10

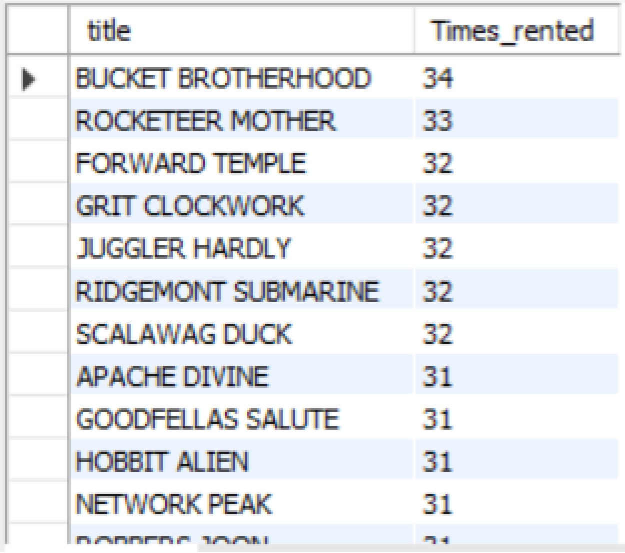
Display the names of the most frequently rented movies in descending order, so that the management can maintain more copies of such movies.

Query:-



```
select f.title, count(r.rental_id) as Times_rented
from film f inner join inventory i
on f.film_id = i.film_id
inner join rental r
on i.inventory_id = r.inventory_id
group by title
order by times_rented DESC;
```

Output:-

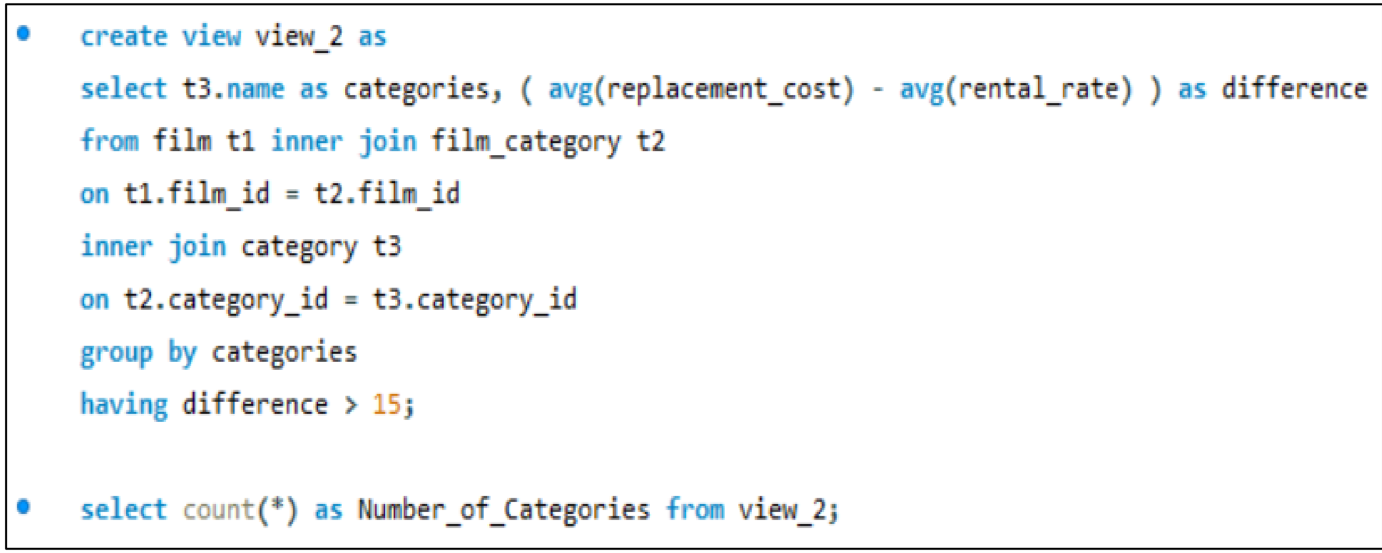


title	Times_rented
BUCKET BROTHERHOOD	34
ROCKETEER MOTHER	33
FORWARD TEMPLE	32
GRIT CLOCKWORK	32
JUGGLER HARDLY	32
RIDGEMONT SUBMARINE	32
SCALAWAG DUCK	32
APACHE DIVINE	31
GOODFELLAS SALUTE	31
HOBBIT ALIEN	31
NETWORK PEAK	31
ROBBERS JOON	31

TASK-11

Calculate and display the number of movie categories where the average difference between the movie replacement cost and the rental rate is greater than $15.

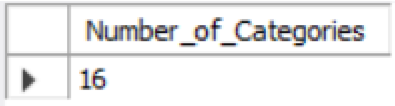
Query:-



```
create view view_2 as
select t3.name as categories, ( avg(replacement_cost) - avg(rental_rate) ) as difference
from film t1 inner join film_category t2
on t1.film_id = t2.film_id
inner join category t3
on t2.category_id = t3.category_id
group by categories
having difference > 15;

select count(*) as Number_of_Categories from view_2;
```

Output:-

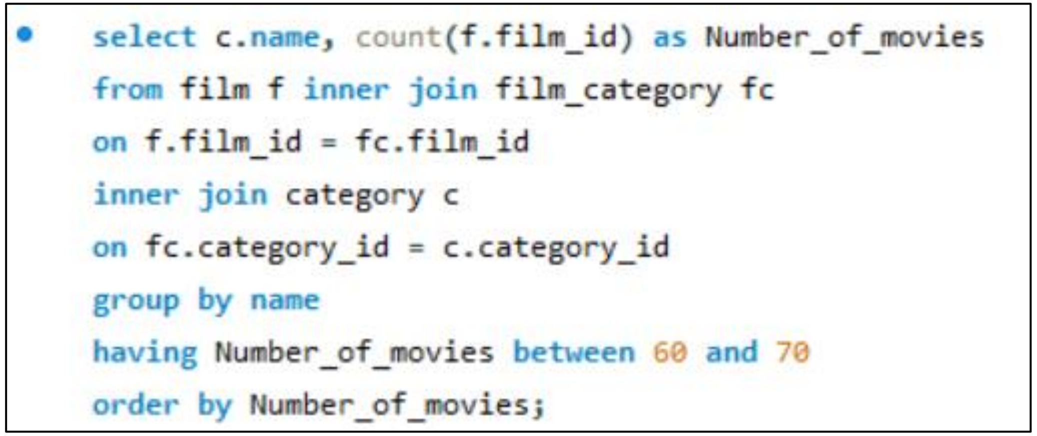


	Number_of_Categories
▶	16

TASK-12

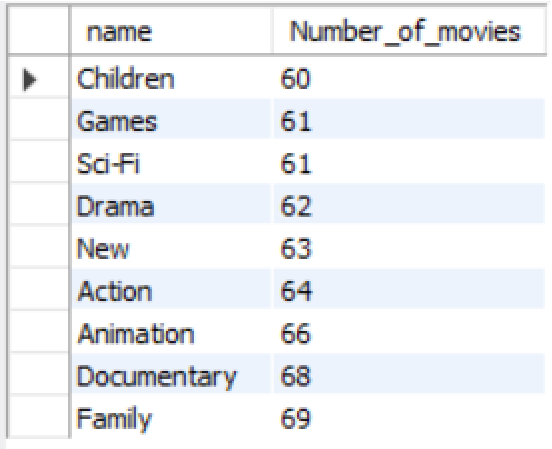
The management wants to identify all the genres that consist of 60-70 movies. The genre details are captured in the category column. **Display the names of these categories and the number of movies per category, sorted by the number of movies.**

Query:-



```
select c.name, count(f.film_id) as Number_of_movies
from film f inner join film_category fc
on f.film_id = fc.film_id
inner join category c
on fc.category_id = c.category_id
group by name
having Number_of_movies between 60 and 70
order by Number_of_movies;
```

Output:-



	name	Number_of_movies
▶	Children	60
	Games	61
	Sci-Fi	61
	Drama	62
	New	63
	Action	64
	Animation	66
	Documentary	68
	Family	69

SUMMARY

Solved the tasks related to business which could really help the Motion Picture Association in stocking up certain movies which will be good for their business growth.

THANK YOU

Tejusvi Sardana

sardanatejus8@gmail.com