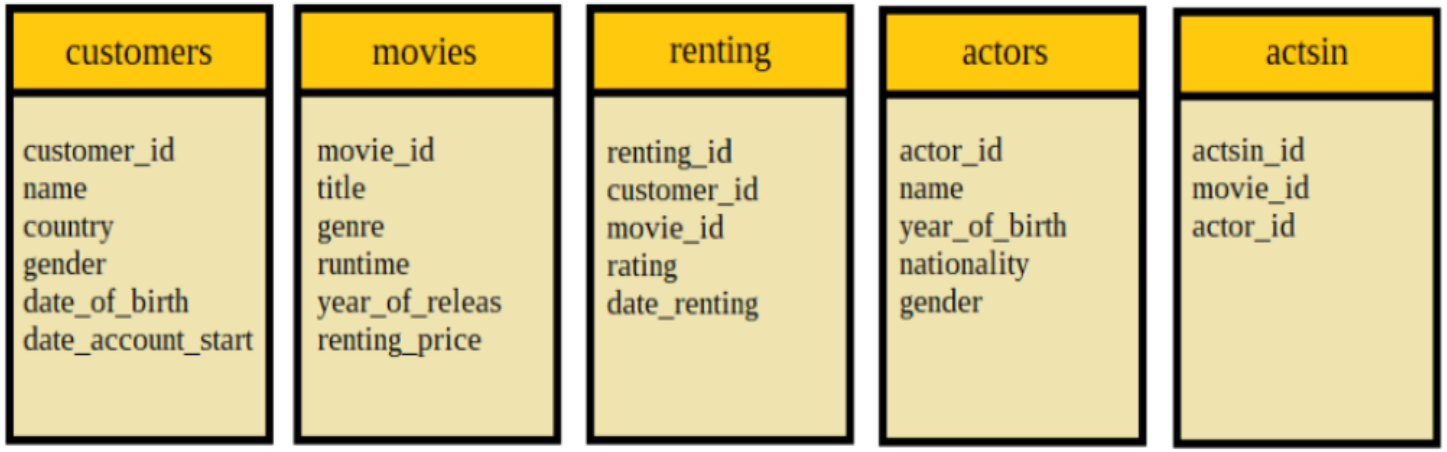
PROJECT: DATA-DRIVEN DECISION MAKING WITH SQL

BY: MAI HAN TRAN

PROJECT GOAL

- Gather information for operational decisions
- Gather information for strategic decisions
- About the company:
  MovieNow is a movie rental company that has a PostgreSQL database with 4 tables

MovieNow data structure



Source: DataCamp

EXPLORATORY ANALYSIS

- Conduct analysis to see when first customer accounts were created for each country:

```
SELECT country,
        MIN(date_account_start) AS first_account
FROM customers
GROUP BY country
ORDER BY country;
```

country	first_account
Austria	2017-11-22
Belgium	2017-01-28
Denmark	2017-04-30
France	2017-01-13

- Report avg. rating, number of ratings, and number of views by each movie

```
SELECT movie_id,
        ROUND (AVG (rating),4) AS avg_rating,
        COUNT (rating) AS number_ratings,
        COUNT (*) AS number_renting
FROM renting
GROUP BY movie_id
ORDER BY avg_rating DESC;
```

movie_id	avg_rating	number_ratings	number_renting
11	null	0	6
48	10.0000	2	5
5	9.7500	4	8

EXPLORATORY ANALYSIS

- Calculate the revenue coming from movie rentals, the # of movie rentals, and the # of customers who rented a movie:

```
SELECT
        SUM (m.renting_price) AS revenue,
        COUNT (*) AS n_movies,
        COUNT (DISTINCT (r.customer_id)) AS cust
FROM renting AS r
LEFT JOIN movies AS m
ON r.movie_id = m.movie_id;
```

revenue	n_movies	cust
658.02	298	93

- Give an overview of which actors play in which movie:

```
SELECT
        m.title,
        a.name
FROM actsin as ai
LEFT JOIN movies as m
ON m.movie_id = ai.movie_id
LEFT JOIN actors AS a
ON a.actor_id = ai.actor_id;
```

title	name
Candy	Abbie Cornish
Jack and Jill	Adam Sandler
Simone	Al Pacino
The Recruit	Al Pacino

EXPLORATORY ANALYSIS

- Identify favorite movies for a group of customers:

```
SELECT
        m.title,
        COUNT (*),
        ROUND (AVG (r.rating),2)
FROM renting AS r
LEFT JOIN customers AS c
ON c.customer_id = r.customer_id
LEFT JOIN movies AS m
ON m.movie_id = r.movie_id
WHERE c.date_of_birth
        BETWEEN '1970-01-01' AND '1979-12-31'
GROUP BY m.title
HAVING COUNT (*) > 1
ORDER BY AVG (rating) DESC;
```

title	count	round
Showtime	5	null
Harry Potter and the Deathly Hallows - Part 2	2	null
Waking Up in Reno	2	null
Ray	2	null
Django Unchained	4	10.00
The Fighter	4	10.00
One Night at McCool's	2	10.00
No Country for Old Men	3	10.00
The Fellowship of the Ring	2	10.00

EXPLORATORY ANALYSIS

- Report a list of frequent customers:

```
SELECT *
FROM customers
WHERE customer_id IN
        ( SELECT customer_id
          FROM renting
          GROUP BY customer_id
           HAVING COUNT (*) > 10);
```

customer_id	name	country	gender	date_of_birth	date_account_start
21	Avelaine Corbeil	France	female	1986-03-17	2017-06-11
28	Sidney Généreux	France	male	1980-12-01	2017-02-04
49	Havasy Kristof	Hungary	male	1998-06-13	2017-01-18

- Report movies with rating above average:

```
SELECT movie_id,
        avg (rating)
FROM renting
GROUP BY movie_id
HAVING AVG (rating) >
        ( SELECT AVG (rating)
          FROM renting);
```

movie_id	avg
54	8.1666666666666667
29	8.0000000000000000
71	8.0000000000000000
34	8.0000000000000000

EXPLORATORY ANALYSIS

- Create a table with the total number of male and female customers for each country

```
SELECT gender,
        country,
        COUNT (*)
FROM customers
GROUP BY CUBE (gender, country)
ORDER BY country;
```

gender	country	count
female	Austria	3
male	Austria	1
null	Austria	4
female	Belgium	3

- Create a table for national preference of the customers from MovieNow

```
SELECT country,
        genre,
        AVG (r.rating) AS avg_rating
FROM renting AS r
LEFT JOIN movies AS m
ON m.movie_id = r.movie_id
LEFT JOIN customers AS c
ON r.customer_id = c.customer_id
GROUP BY CUBE (country, genre);
```

country	genre	avg_rating
null	null	7.9390243902439024
France	Mystery & Suspense	6.0000000000000000
Slovenia	Action & Adventure	null
Spain	Animation	null

CUSTOMER BEHAVIOR INSIGHT

- A new advertising campaign is going to focus on customer who rented fewer than 5 movies.
- Extract all customer information within the target group.

```
SELECT *
FROM customers AS c
WHERE 5 >
            ( SELECT COUNT (*)
              FROM renting AS r
              WHERE r.customer_id =
                          c.customer_id);
```

customer_id	name	country	gender	date_of_birth	date_account_start
2	Wolfgang Ackermann	Austria	male	1971-11-17	2018-10-15
3	Daniela Herzog	Austria	female	1974-08-07	2019-02-14
4	Julia Jung	Austria	female	1991-01-04	2017-11-22
5	Juliane Kirsch	Austria	female	1977-03-01	2018-12-16
6	Rowanne Couperus	Belgium	female	1994-04-05	2018-08-26
9	Antal van Looij	Belgium	male	1982-06-18	2019-03-10

CUSTOMER BEHAVIOR INSIGHT

- Identify customers who were not satisfied with movies they watched on MovieNow.

```
SELECT *
FROM customers AS c
WHERE 4 >
            ( SELECT MIN (rating)
              FROM renting AS r
              WHERE r.customer_id =
                          c.customer_id);
```

customer_id	name	country	gender	date_of_birth	date_account_start
28	Sidney Généreux	France	male	1980-12-01	2017-02-04
41	Zara Mitchell	Great Britan	female	1994-07-08	2017-06-12
86	Albin Jaworski	Poland	male	1984-05-01	2017-12-15
120	Robin J. Himes	USA	male	1988-11-30	2018-08-06

- Identify a list of customers with at least one rating

```
SELECT *
FROM customers AS c
WHERE EXISTS
            ( SELECT *
              FROM renting AS r
              WHERE rating IS NOT NULL
              AND r.customer_id
                 = c.customer_id);
```

customer_id	name	country	gender	date_of_birth	date_account_start
2	Wolfgang Ackermann	Austria	male	1971-11-17	2018-10-15
4	Julia Jung	Austria	female	1991-01-04	2017-11-22
7	Annelous Sneep	Belgium	female	1993-11-14	2018-05-12
8	Jaëla van den Dolder	Belgium	female	1990-08-31	2018-02-08

BUSINESS CASE: *MovieNow considers investing in new movies.*

- Analysis 1 – Should the company invest in recently-produced or older movies?
  And is there difference across country?

➢ *Restrictions: pull records of movies with at least 4 ratings and rentals since 2018-04-01*

```
SELECT c.country,
        m.year_of_release,
        COUNT (*) AS n_rentals,
        COUNT (DISTINCT r.movie_id) AS n_movies,
        AVG (rating) AS avg_rating
FROM renting AS r
LEFT JOIN customers AS c
ON c.customer_id = r.customer_id
LEFT JOIN movies AS m
ON m.movie_id = r.movie_id
WHERE r.movie_id IN
        ( SELECT  movie_id
          FROM renting
          GROUP BY movie_id
          HAVING COUNT (rating) >= 4 )
AND r.date_renting >= '2018-04-01'
GROUP BY ROLLUP (m.year_of_release, country)
ORDER BY c.country, m.year_of_release;
```

country	year_of_release	n_rentals	n_movies	avg_rating
Austria	2002	1	1	7.0000000000000000
Austria	2004	1	1	null
Austria	2007	1	1	6.0000000000000000
Austria	2011	1	1	6.0000000000000000
Belgium	2001	5	3	10.0000000000000000
Belgium	2002	1	1	null
Belgium	2003	3	2	10.0000000000000000
Belgium	2005	1	1	null
Belgium	2006	2	2	7.0000000000000000

Insight:

➢ The customers have no clear preference for most recent movies over older ones.

BUSINESS CASE: *MovieNow considers investing in new movies.*

- Analysis 2 – Find out movies of the best rated genres for future investment

  - *Restrictions: pull records of movies with at least 4 ratings and rentals since 2018-04-01*

```
SELECT genre,
        AVG (rating) AS avg_rating,
        COUNT (rating) AS n_rating,
        COUNT (*) AS n_rentals,
        COUNT (DISTINCT m.movie_id) AS n_movies
FROM renting AS r
LEFT JOIN movies AS m
ON m.movie_id = r.movie_id
WHERE r.movie_id IN
        ( SELECT movie_id
          FROM renting
          GROUP BY movie_id
          HAVING COUNT (rating) > =4 )
AND r.date_renting >= '2018-04-01'
GROUP BY genre
ORDER BY AVG (r.rating) DESC;
```

genre	avg_rating	n_rating	n_rentals	n_movies
Action & Adventure	8.7142857142857143	7	9	2
Animation	7.8333333333333333	6	10	2
Art House & International	8.5000000000000000	4	5	1
Comedy	7.9500000000000000	20	31	5
Drama	7.7482517482517483	143	245	34
Mystery & Suspense	7.4285714285714286	7	19	3

Insight:

- Action & Adventure is the genre with highest average rating

BUSINESS CASE: *MovieNow considers investing in new movies.*

- Analysis 3 – Analyze customer preferences for actors to identify possible investment option

➢ *Restrictions: pull records of movies with at least 4 ratings and rentals since 2018-04-01*

```
SELECT a.nationality,
            a.gender,
            AVG (r.rating) AS avg_rating,
            COUNT (r.rating) AS n_rating,
            COUNT ( DISTINCT a.actor_id) AS n_actors
FROM renting AS r
LEFT JOIN actsin AS ai
ON ai.movie_id = r.movie_id
LEFT JOIN actors AS a
ON ai.actor_id = a.actor_id
WHERE r.movie_id IN
            ( SELECT movie_id
              FROM renting
              GROUP BY movie_id
               HAVING COUNT (rating) >=4)
AND r.date_renting >= '2018-04-01'
GROUP BY GROUPING SETS ((a.nationality, a.gender),
            (a.nationality), (a.gendger),  () );
```

nationality	gender	avg_rating	n_rating	n_rentals	n_actors
Argentina	male	8.5000000000000000	4	5	1
Argentina	null	8.5000000000000000	4	5	1
Australia	female	8.6666666666666667	3	5	1
Australia	male	7.4545454545454545	11	17	3
Australia	null	7.7142857142857143	14	22	4
Austria	male	8.5000000000000000	2	6	1
Austria	null	8.5000000000000000	2	6	1

Insight:

➢ Australian actress typically receives higher rating than the nation's average

THANK YOU