SQL Project

Greencycles
Online Movie Rental Shop
Data Analysis

Using PostgreSQL

Presented by
Nicken Shidqia Nurahman

Nicken Shidqia Nurahman

About Me

Civil engineer graduate with some experience in administration and project management, who is interested in data science.
Detail oriented, and time management person, and familiar with Microsoft Office, Python, SQL and Jupyter.
Motivated to continue to learn and grow as a professional.

My Experience

- Data Science Bootcamp Student – RAKAMIN ACADEMY
  Oct 2023 - Now
- Project Management Masters Degree Student – UNIVERSITAS INDONESIA
  Sep 2021 - Sep 2023
- Engineering Administration and Project Control Staff - PT. ISTAKA KARYA
  Aug 2019 - Sep 2021
- Project Control Intern - PT. ISTAKA KARYA
  Feb 2019 - Jul 2019
- Surveying Laboratory Assistant – UNIVERSITAS TRISAKTI
  Jul 2017- Agust 2019

Case Study

Problem

Inventory Management

Ensuring the availability of popular movies and managing inventory effectively

Challenges

Help the company operate big query and gain insight from data.

Goal

Optimized Inventory

Optimize the movie catalog by identifying and promoting popular content, while efficiently managing licensing costs to maximize profitability

Tools Used

***) Solution Code :**

https://github.com/nickenshidqia/Online_Movie_Rental_Shop_Data_Analysis_SQL_Project/blob/9ab5a4d2b8f69a7341c924b5952b23026e0bfff8/SQL%20Project%20Greencycle%20Movie%20Rental%20Company.sql

Query Task

1. In the email system there was a problem with names where either the first name or the last name is more than 10 characters long. Find these customers and output the list of these first and last names in all lower case

first_names text	last_names text	email text
william	satterfield	william.satterfield@sakilacustomer.org
christopher	greco	christopher.greco@sakilacustomer.org
henry	billingsley	henry.billingsley@sakilacustomer.org
roger	quintanilla	roger.quintanilla@sakilacustomer.org
jonathan	scarborough	jonathan.scarborough@sakilacustomer.org

2. Extract the last 5 characters of the email address first. The email address always ends with '.org'. How can you extract just the dot '.' from the email address?

right text	left text
r.org	.
r.org	.
r.org	.
r.org	.
r.org	.

3. You need to create an anonymized version of the email addresses. It should be the first character followed by '***' and then the last part starting with '@'.

anonymized_email text
M***@sakilacustomer.org
P***@sakilacustomer.org
L***@sakilacustomer.org

4. In this challenge you have only the email address and the last name of the customers. You need to extract the first name from the email address and concatenate it with the last name. It should be in the form: "Last name, First name".

email text	position integer	left text	?column? text
MARY.SMITH@sakilacustomer.org	6	MARY	SMITH,MARY
PATRICIA.JOHNSON@sakilacustomer.org	10	PATRICIA	JOHNSON,PATRICIA

5. You need to create an anonymized form of the email addresses

email text	?column? text
MARY.SMITH@sakilacustomer.org	M***.S***@sakilacustomer.org
PATRICIA.JOHNSON@sakilacustomer.org	P***.J***@sakilacustomer.org

Query Task

6. What's the highest amount one customer has spent in a week?

week numeric	customer_id smallint	total_payment_amount numeric
18	459	73.88
12	21	72.86
18	2	65.88

7. You need to sum payments and group in the following formats

total_amount numeric	day text
39.91	Thu,03:45
16.96	Thu,10:05

8. You need to create a list for the suppcity team of all rental durations of customer with customer_id 35. Also you need to find out for the suppcity team which customer has the longest average rental duration?

customer_id smallint	avg interval
315	6 days 14:13:22.5

9.Your manager is thinking about increasing the prices for films that are more expensive to replace. Create a list of the films including the relation of rental rate where the rental rate is less than 4% of the replacement cost.

film_id [PK] integer	percentage numeric
417	3.30
663	3.30
52	3.30

10. a. Rating is 'PG' or 'PG-13' or length is more then 210 min: 'Great rating or long (tier 1)
b. Description contains 'Drama' and length is more than 90min: 'Long drama (tier 2)'
c. Description contains 'Drama' and length is not more than 90min: 'Shcity drama (tier 3)'
d. Rental_rate less than $1: 'Very cheap (tier 4)'
How can you filter to only those movies that appear in one of these 4 tiers?

title text	case text
ACADEMY DINOSAUR	Great rating or long (tier 1)
AGENT TRUMAN	Great rating or long (tier 1)
AIRPLANE SIERRA	Great rating or long (tier 1)

Query Task

11. sum rating by category but in pivot table

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

12. Handling missing value in return_date from rental table with 'Not returned'

rental_date timestamp with time zone	coalesce character varying
2020-02-14 22:16:03+07	Not returned
2020-02-14 22:16:03+07	Not returned

13. The company wants to run a phone call campaing on all customers in Texas (=district). What are the customers (first_name, last_name, phone number and their district) from Texas?

first_name text	last_name text	phone text	district text
JENNIFER	DAVIS	860452626434	Texas
KIM	CRUZ	909029256431	Texas

14. The company wants customize their campaigns to customers depending on the country they are from. Which customers are from Brazil?

first_name text	last_name text	email text	country text
CLAYTON	BARBEE	CLAYTON.BARBEE@sakilacustomer.org	Brazil
JOSEPH	JOY	JOSEPH.JOY@sakilacustomer.org	Brazil
TAMARA	NGUYEN	TAMARA.NGUYEN@sakilacustomer.org	Brazil
NATALIE	MEYER	NATALIE.MEYER@sakilacustomer.org	Brazil

15.Which title has GEORGE LINTON rented the most often?

customer_id integer	first_name text	last_name text	title text	count bigint
314	GEORGE	LINTON	CADDYSHACK JEDI	3

16.Select all of the films where the length is longer than the average of all films

film_id [PK] integer	title text	description text
4	AFFAIR PREJUDICE	A Fanciful Documentary of a Frisbee And a Lumberjack
5	AFRICAN EGG	A Fast-Paced Documentary of a Pastry Chef And a Den
6	AGENT TRUMAN	A Intrepid Panorama of a Robot And a Boy who must Es
11	ALAMO VIDEOTAPE	A Boring Epistle of a Butler And a Cat who must Fight a

Query Task

17.Return all the films that are available un the inventory in store 2 more than 3 times

film_id [PK] integer	title text
1	ACADEMY DINOSAUR
3	ADAPTATION HOLES

18.Return all customer first names and last names that have made payment on 2020-01-25

first_name text	last_name text
ALMA	AUSTIN
ANDREW	PURDY

19, Return all customer first name and email address that have spent a more than $30

first_name text	email text
MARY	MARY.SMITH@sakilacustomer.org
PATRICIA	PATRICIA.JOHNSON@sakilacustomer.org

20.Return all the customer first and last name that are from California and have spent more than 100 in total

first_name text	last_name text
PATRICIA	JOHNSON
BETTY	WHITE

21.What is the average total amount spent per day (average daily revenue)?

daily_rev_avg numeric
1644.31

22.Show all the payment together with how much the payment amount is below the maximum payment amount

payment_id integer	customer_id smallint	staff_id smallint	rental_id integer	amount numeric (5,2)
16050	269	2	7	1.99
16051	269	1	98	0.99
16052	269	2	678	6.99

Query Task

23. Show only those movie titles, their associated film_id and replacement _cost with the lowest replacement _cost for in each rating category - also show the rating

title text	film_id [PK] integer	replacement_cost numeric (5,2)	rating mpaa_rating
ANACONDA CONFESSIONS	23	9.99	R
CIDER DESIRE	150	9.99	PG

24.Show only those movie titles, their associated film_id and the length that have the highest length in each category - also show the rating

title text	film_id [PK] integer	length smallint	rating mpaa_rating
CHICAGO NORTH	141	185	PG-13
CONTROL ANTHEM	182	185	G

25.Show all the payments plus the total amount for every customer as well as the number of payments of each customer

customer_id smallint	staff_id smallint	rental_id integer	amount numeric (5,2)	payment_date timestamp with time zone	sum_amount numeric
1	2	1476	9.99	2020-02-16 02:37:12.996577+07	118.68
1	1	6163	7.99	2020-04-11 14:42:12.996577+07	118.68

26.Show only those films with the highest replacement costs in their rating category plus show the average replacement cost in their rating category

title text	replacement_cost numeric (5,2)	rating mpaa_rating	avg numeric
ARABIA DOGMA	29.99	NC-17	20.1376190476190476
BALLROOM MOCKINGBIRD	29.99	G	20.1248314606741573

27.Show only those payments with the highest payment for each customer's first name - including the payment_id of that payment. How would you solve it if you would not need to see the payment_id?

first_name text	payment_id integer	amount numeric (5,2)
MARY	18497	9.99
PATRICIA	29014	10.99

28.Create a list of all the different (distinct) replacement costs of the films. What's the lowest replacement cost?

min numeric
9.99

Query Task

29.Write a query that gives an overview of how many films have replacements costs in the following cost ranges
low: 9.99 - 19.99
medium: 20.00 - 24.99
high: 25.00 - 29.99
How many films have a replacement cost in the "low“ group?

count bigint	category text
250	medium
236	high
514	low

30.Create a list of the film titles including their title, length, and category name ordered descendingly by length. Filter the results to only the movies in the category 'Drama' or 'Sports’. In which category is the longest film and how long is it?

title text	length smallint	name text
SMOOCHY CONTROL	184	Sports
RECORDS ZORRO	182	Sports

31.Create an overview of how many movies (titles) there are in each category (name). Which category (name) is the most common among the films?

count bigint	name text
74	Sports
73	Foreign
69	Family

32.Create an overview of the actors' first and last names and in how many movies they appear in. Which actor is part of most movies?

first_name text	last_name text	count bigint
SUSAN	DAVIS	54
GINA	DEGENERES	42

33.Create an overview of the addresses that are not associated to any customer How many addresses are that?

count bigint
4

Query Task

34.Create an overview of the cities and how much sales (sum of amount) have occurred there. Which city has the most sales?

city text	sum numeric
Cape Coral	221.55
Saint-Denis	216.54

35. Create an overview of the revenue (sum of amount) grouped by a column in the format "country, city". Which country, city has the least sales?

city text	country text	sum numeric
Tallahassee	United States	50.85
Fuzhou	China	50.86

36.Create a list with the average of the sales amount each staff_id has per customer. Which staff_id makes on average more revenue per customer?

staff_id smallint	avg numeric
2	56.6394657762938230
1	55.9089649415692821

37.Create a query that shows average daily revenue of all Sundays. What is the daily average revenue of all Sundays?

avg numeric
1775.7400000000000000

38.Create a list of movies - with their length and their replacement cost - that are longer than the average length in each replacement cost group. Which two movies are the shortest on that list and how long are they?

title text	length smallint	replacement_cost numeric (5,2)
CELEBRITY HORN	110	24.99
SEATTLE EXPECATIONS	110	18.99

39. Create a list that shows the "average customer lifetime value" grouped by the different districts.

district text	avg numeric
Saint-Denis	216.5400000000000000
Minsk	195.5800000000000000
Skikda	173.6300000000000000

Query Task

43.Write a query that returns the list of movies including film_id,, title, length, category, average length of movies in that category. Order the results by film_id.

film_id integer	title text	length smallint	name text	round numeric
1	ACADEMY DINOSAUR	86	Documentary	108.75
2	ACE GOLDFINGER	48	Horror	112.48

44. Write a query that returns all payment details including the number of payments that were made by this customer and that amount

payment_id integer	customer_id smallint	staff_id smallint	rental_id integer	amount numeric (5,2)
16050	269	2	7	1.99
16051	269	1	98	0.99

45.Rank number one position every film category

title text	name text	length smallint	rank bigint
WORST BANGER	Action	185	1
DARN FORRESTER	Action	185	1
POND SEATTLE	Animation	185	1
GANGS PRIDE	Animation	185	1

46. Write a query that returns the customers' name, the country and how many payments they have. For that use the existing view customer_list. Afterwards create a ranking of the top customers with most sales for each country. Filter the results to only the top 3 customers per country

name text	country text	count bigint	rank bigint
ELEANOR HUNT	Runion	46	1
KARL SEAL	United States	45	1

47. Write a query that returns the revenue of the day and the revenue of the previous day. Afterwards calculate also the percentage growth compared to the previous day

sum numeric	day date	previous_day numeric	difference numeric	percentage_growth numeric
446.92	2020-01-25	[null]	[null]	[null]
724.29	2020-01-26	446.92	277.37	62.06
692.37	2020-01-27	724.29	-31.92	-4.41
815.03	2020-01-28	692.37	122.66	17.72

Query Task

48. Write query that return the sum of the amount for each customer (first name and last name) and each staff_id. Also add the overall revenue per customer

first_name text	last_name text	staff_id smallint	sum numeric
AARON	SELBY	1	63.86
AARON	SELBY	2	46.90

49. Write a query that calculates now the share of revenue each staff_id makes per customer

first_name text	last_name text	staff_id smallint	total numeric	first_value numeric	percentage numeric
AARON	SELBY	[null]	110.76	110.76	100.00
AARON	SELBY	1	63.86	110.76	57.66

50. Write a query that returns all grouping sets in all combinations of customer_id, date and title with the aggregation of the payment amount

customer_id smallint	date date	title text	total numeric
1	2020-01-25	ADAPTATION HOLES	2.99
1	2020-01-25	AMISTAD MIDSUMMER	2.99

51. Find all the pairs of films with the same length

title text	title text	length smallint
HOME PITY	POND SEATTLE	185
HOME PITY	MUSCLE BRIGHT	185
HOME PITY	WORST BANGER	185
HOME PITY	CONTROL ANTHEM	185

52. Create a function that expects the customer's first and last name and returns the total amount of payments this customer has made

name_searchs numeric
127.71

Link Portolio On Github :

https://github.com/nickenshidqia/Online_Movie_Rental_Shop_Data_Analysis_SQL_Project

LinkedIn:

https://www.linkedin.com/in/nickenshidqia/

THANK YOU