
SQL - Google Play Store Project

— Dipti Soni —

Introduction

This dataset contains information about different apps installed on mobiles. All the app related information is stored in a single table called playstore.

Dataset was analyzed using MySQL

App	Category	Rating	Reviews	Size	Installs	Type	Price	Content_Rating	Genres	Last_Updated	Current_Ver	Android_Ver
Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10000	Free	0	Everyone	Art & Design	2018-01-07	1.0.0	4.0.3 and up
Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500000	Free	0	Everyone	Art & Design;Pretend Play	2018-01-15	2.0.0	4.0.3 and up
U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8.7M	5000000	Free	0	Everyone	Art & Design	2018-08-01	1.2.4	4.0.3 and up
Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50000000	Free	0	Teen	Art & Design	2018-06-08	Varies with device	4.2 and up
Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100000	Free	0	Everyone	Art & Design;Creativity	2018-06-20	1.1	4.4 and up
Paper flowers instructions	ART_AND_DESIGN	4.4	167	5.6M	50000	Free	0	Everyone	Art & Design	2017-03-26	1	2.3 and up
Smoke Effect Photo Maker - Smoke Editor	ART_AND_DESIGN	3.8	178	19M	50000	Free	0	Everyone	Art & Design	2018-04-26	1.1	4.0.3 and up
Infinite Painter	ART_AND_DESIGN	4.1	36815	29M	1000000	Free	0	Everyone	Art & Design	2018-06-14	6.1.61.1	4.2 and up
Garden Coloring Book	ART_AND_DESIGN	4.4	13791	33M	1000000	Free	0	Everyone	Art & Design	2017-09-20	2.9.2	3.0 and up
Kids Paint Free - Drawing Fun	ART_AND_DESIGN	4.7	121	3.1M	10000	Free	0	Everyone	Art & Design;Creativity	2018-07-03	2.8	4.0.3 and up

Most Promising App Categories

You are working as a market analyst for a mobile app development company. Your task is to identify the most promising categories (top 5) for launching new free apps based on their average ratings.

```
select category, round(avg(rating),2) as 'rating'  
from playstore where type = 'Free'  
group by category order by rating desc limit 5;
```



category	rating
EVENTS	4.44
EDUCATION	4.38
ART_AND_DESIGN	4.36
BOOKS_AND_REFERENCE	4.35
PARENTING	4.34

Top revenue generating App Category

As a business strategist for a mobile app company, your objective is to pinpoint the three categories that generate the most revenue from paid apps. This calculation is based on the product of the app price and its number of installations.

```
select category, round(avg(price*installs),2) as 'revenue' from playstore where type = 'paid'  
group by category order by revenue desc limit 3;
```



category	revenue
LIFESTYLE	3199340.56
FINANCE	1979115.38
PHOTOGRAPHY	1162143.33

Distribution of apps across categories

As a data analyst for a gaming company, you're tasked with calculating the percentage of games within each category. This information will help the company understand the distribution of gaming apps across different categories.

```
select category, (100*count(*)/(select count(*) from playstore) as 'percentage' from playstore  
group by category order by percentage desc;
```



category	percentage
FAMILY	18.6538
GAME	11.7201
TOOLS	7.8312
PRODUCTIVITY	3.7500
MEDICAL	3.7393
COMMUNICATION	3.5043
FINANCE	2.4500

Paid or Free Apps

As a data analyst at a mobile app-focused market research firm you'll recommend whether the company

```
with t1 as
(
select category, avg(rating) as 'paid_rating' from playstore where type = 'paid' group by category
),
t2 as
(
select category, avg(rating) as 'free_rating' from playstore where type = 'free' group by category
)
select a.category, a.paid_rating, b.free_rating, if(a.paid_rating > b.free_rating, 'go for paid', 'go for free') as 'decision'
from t1 as a join t2 as b on a.category = b.category;
```



category	paid_rating	free_rating	decision
ART_AND_DESIGN	4.733333333333333	4.358620689655172	go for paid
AUTO_AND_VEHICLES	4.6	4.184722222222223	go for paid
BOOKS_AND_REFERENCE	4.275	4.3494117647058825	go for free
BUSINESS	4.2	4.118493150684932	go for paid
COMMUNICATION	4.0636363636363635	4.1653594771241815	go for free

Database got hacked

Suppose you're a database administrator your databases have been hacked and hackers are changing price of certain apps on the database, it is taking long for IT team to neutralize the hack, however you as a responsible manager don't want your data to be changed, do some measure where the changes in price can be recorded as you can't stop hackers from making changes.

```
-- created a table to record the changes done on the table
```

```
create table pricechangelog(  
  app varchar(255),  
  old_price decimal(10,2),  
  new_price decimal(10,2),  
  operation_type varchar(255),  
  operation_date timestamp  
);
```

```
-- creating backup of playstore table
```

```
create table play as  
select * from playstore;
```

```
-- creating a trigger
DELIMITER //
create trigger price_change_log
after update
on play
for each row
begin
    insert into pricechangelog(app, old_price, new_price, operation_type, operation_date)
    values(new.app, old.price, new.price, 'update', current_timestamp);
end;
//

-- change certain things in play table
update play
set price = 4 where app = 'Infinite Painter';

update play
set price = 5 where app = 'Sketch - Draw & Paint';

-- see the changes in play table
select * from pricechangelog;
```


Correlation between ratings and reviews

As a data person you are assigned the task of investigating the CORRELATION between two numeric factors: app ratings and the quantity of reviews.

```
set @x = (select round(avg(rating),2) from playstore);  
set @y = (select round(avg(reviews),2) from playstore);  
select @x,@y;
```

```
with cte as (  
select *, (rat*rat) as 'ratsq', (rev*rev) as 'revsq' from  
(  
select rating, @x, round((rating-@x),2) as 'rat', reviews, @y, round((reviews-@y),2) as 'rev' from playstore  
)t )
```

```
select @numerator := sum(rat*rev), @deno_1 := sum(ratsq), @deno_2 := sum(revsq) from cte;  
select round((@numerator/(sqrt(@deno_1*@deno_2))),4) as 'correlation_coeff';
```



correlation_coeff
0.0681

Clean the Genres Column

Your boss noticed that some rows in genres columns have multiple genres in them, which was creating issue when developing the recommender system from the data he/she assigned you the task to clean the genres column and make two genres out of it, rows that have only one genre will have other column as blank.

```

Delimiter //
create function f_name (a varchar(255))
returns varchar(255)
deterministic -- give this term when the function returns an output
begin
    set @l = locate(';',a);
    set @s = if(@l>0 , left(a,@l-1), a);
    return @s;
end;
//

```



```

Delimiter //
create function l_name (a varchar(255))
returns varchar(255)
deterministic -- give this term when the function returns an output
begin
    set @l = locate(';',a);
    set @s = if(@l=0 , ' ', substring(a,@l+1,length(a)));
    return @s;
end;
//

```

genres	first_name	last_name
Art & Design	Art & Design	
Art & Design;Pretend Play	Art & Design	Pretend Play
Art & Design	Art & Design	
Art & Design	Art & Design	
Art & Design;Creativity	Art & Design	Creativity
Art & Design	Art & Design	

```

select genres, f_name(genres) as 'first_name', l_name(genres) as 'last_name' from playstore;

```

Apps lower than average rating

Your senior manager wants to know which apps are not performing as par in their particular category, however he is not interested in handling too many files or list for every category and he/she assigned you with a task of creating a dynamic tool where he/she can input a category of apps he/she interested in and your tool then provides real-time feedback by displaying apps within that category that have ratings lower than the average rating for that specific category generalized query(all together)

```
-- procedure
delimiter //
create procedure lessratingsapps(in categ varchar(50))
begin
    set @avg_rating = (select round(avg(rating),2) from playstore p where category = categ);
    select * from playstore where category = categ and rating < @avg_rating;
end
//

call lessratingsapps('TRAVEL_AND_LOCAL');
```

App	Category	Rating	Reviews	Size	Installs	Type	Price	Content_Rating	Genres	Last_Updated	Current_Ver	Android_Ver
Expedia Hotels, Flights & Car Rental Travel Deals	TRAVEL_AND_LOCAL	4.1	136626	14M	10000000	Free	0	Everyone	Travel & Local	2018-08-06	18.30.0	5.0 and up
Trip by Skyscanner - City & Travel Guide	TRAVEL_AND_LOCAL	4.1	5150	Varies with device	500000	Free	0	Everyone	Travel & Local	2018-05-29	Varies with device	Varies with device
Google Trips - Travel Planner	TRAVEL_AND_LOCAL	4.1	26871	Varies with device	5000000	Free	0	Everyone	Travel & Local	2018-07-26	1.10.0.205086730	4.1 and up
GPS Map Free	TRAVEL_AND_LOCAL	4	33782	7.6M	5000000	Free	0	Everyone	Travel & Local	2018-07-11	4.6	4.0 and up
Southwest Airlines	TRAVEL_AND_LOCAL	3.9	24781	8.3M	5000000	Free	0	Everyone	Travel & Local	2018-07-11	5.6.1	6.0 and up
AT&T Navigator: Maps, Traffic	TRAVEL_AND_LOCAL	3.6	32862	14M	10000000	Free	0	Everyone	Travel & Local	2016-12-06	5.10.6.2.8466393	4.1 and up
VZ Navigator	TRAVEL_AND_LOCAL	4	16101	Varies with device	50000000	Free	0	Everyone	Travel & Local	2018-01-22	Varies with device	4.2 and up
KakaoMap - Map / Navigation	TRAVEL_AND_LOCAL	4.1	76779	Varies with device	10000000	Free	0	Everyone	Travel & Local	2018-07-30	Varies with device	Varies with device