Play store App
Case Study

Mysql

Presented by:
Yogesh chouhan

Contents of this case study

You can delete this slide when you're done editing the presentation

Infile statement	Load unclean data in mysql using infile statement.
Data cleaning	Using python some data cleaning perform.
Handling database hacking	What should be do when database has been hacked?
Triggers	Creating Triggers
Performance of query	Write efficient , optimize query
Enhancing problem solving skill	Develop problem solving skill using writing complex query.

Presented by:
Yogesh chouhan

Visit my social acounts:
Freepik | Flaticon | Storyset | Wepik | Videvo

Table of contents

01

Data cleaning

02

Importing data

03

Performing Analysis

04

Handling Database Hacking

05

Data cleaning Using mysql

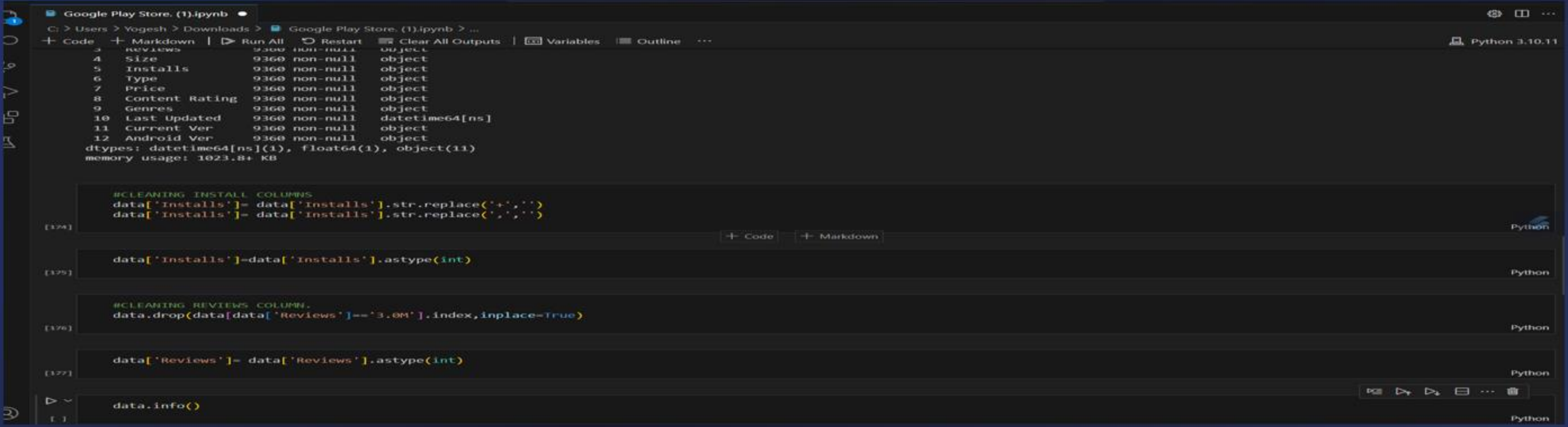
01

Data cleaning

Why Performing Data Cleaning?

- Since our Data is unclean we can't import Data into mysql work bench using table data import wizard
- Because of unclean data we can't get accurate insights from data.
- Because of unclean data we can't perfrom arithmetic operations like,sum,count,avg, Datetime operations etc.
- Missing values are also tough headache for our analysis

Data cleaning Using Python


Google Play Store. (1).ipynb
C: > Users > Yogesh > Downloads > Google Play Store. (1).ipynb > ...
+ Code
+ Markdown
Run All
Restart
Clear All Outputs
Variables
Outline
Python 3.10.11
Python

```
 3   Reviews         9360 non-null   object
 4   Size            9360 non-null   object
 5   Installs        9360 non-null   object
 6   Type            9360 non-null   object
 7   Price           9360 non-null   object
 8   Content Rating  9360 non-null   object
 9   Genres          9360 non-null   object
 10  Last Updated    9360 non-null   datetime64[ns]
 11  Current Ver     9360 non-null   object
 12  Android Ver     9360 non-null   object
dtypes: datetime64[ns](1), float64(1), object(11)
memory usage: 1023.8+ KB
```

```
[174]
#CLEANING INSTALL COLUMNS
data['Installs']= data['Installs'].str.replace('+','')
data['Installs']= data['Installs'].str.replace(',','')
```

```
[175]
data['Installs']=data['Installs'].astype(int)
```

```
[176]
#CLEANING REVIEWS COLUMN.
data.drop(data[data['Reviews']=='3.0M'].index,inplace=True)
```

```
[177]
data['Reviews']= data['Reviews'].astype(int)
```

```
[ ]
data.info()
```

Notebook Link : https://drive.google.com/file/d/1Jm6-Mkjsc2S97_GqIHtZerv8jUWRd-DD/view?usp=drive_link

02

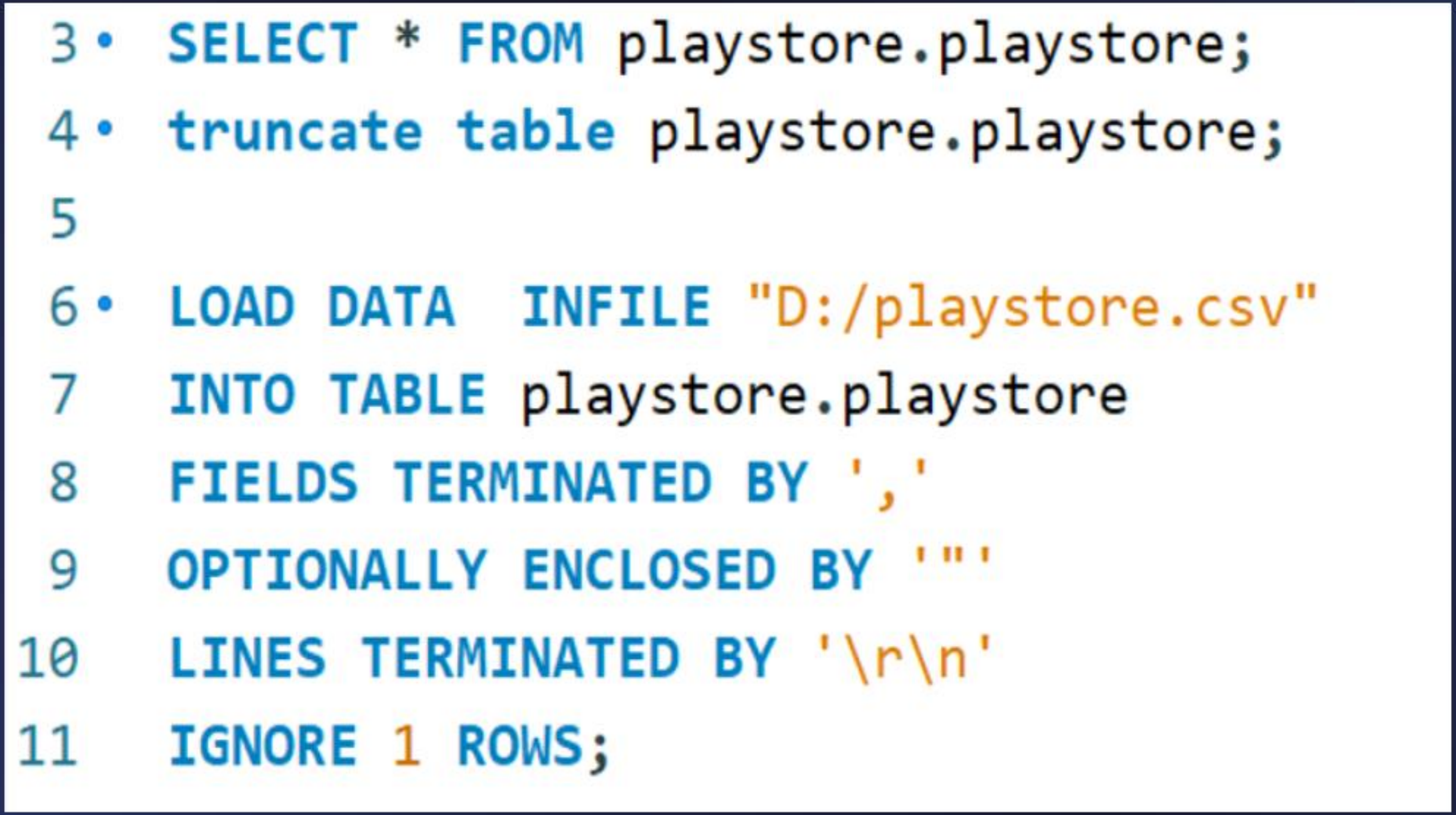
Importing Data

Steps for importing Data using infile

In mysql infile is not active. But we can active infile statement using these following steps:

- Step1. search service on system and turn off mysql task.
- Step2. go to c drive and select unhiden file. Then go to programdata -> Mysql -> Mysqlserver -> open my file in notepad.
- Step3. go to client section in file. Then go to default-character-set= below and write local_infile=ON
- Step4. go to server section in file. Then go to server_type=3 below and write local_infile=ON
- Step5 . Then press ctrl + h and search secure. Then change secure-file-priv="path" to secure-file-priv="path"
- Step6. save the file . And start the mysql . And restart the computer.
- If these steps done properly then infile statement is on your mysql workbench.

After activate infile statement write this query to import data

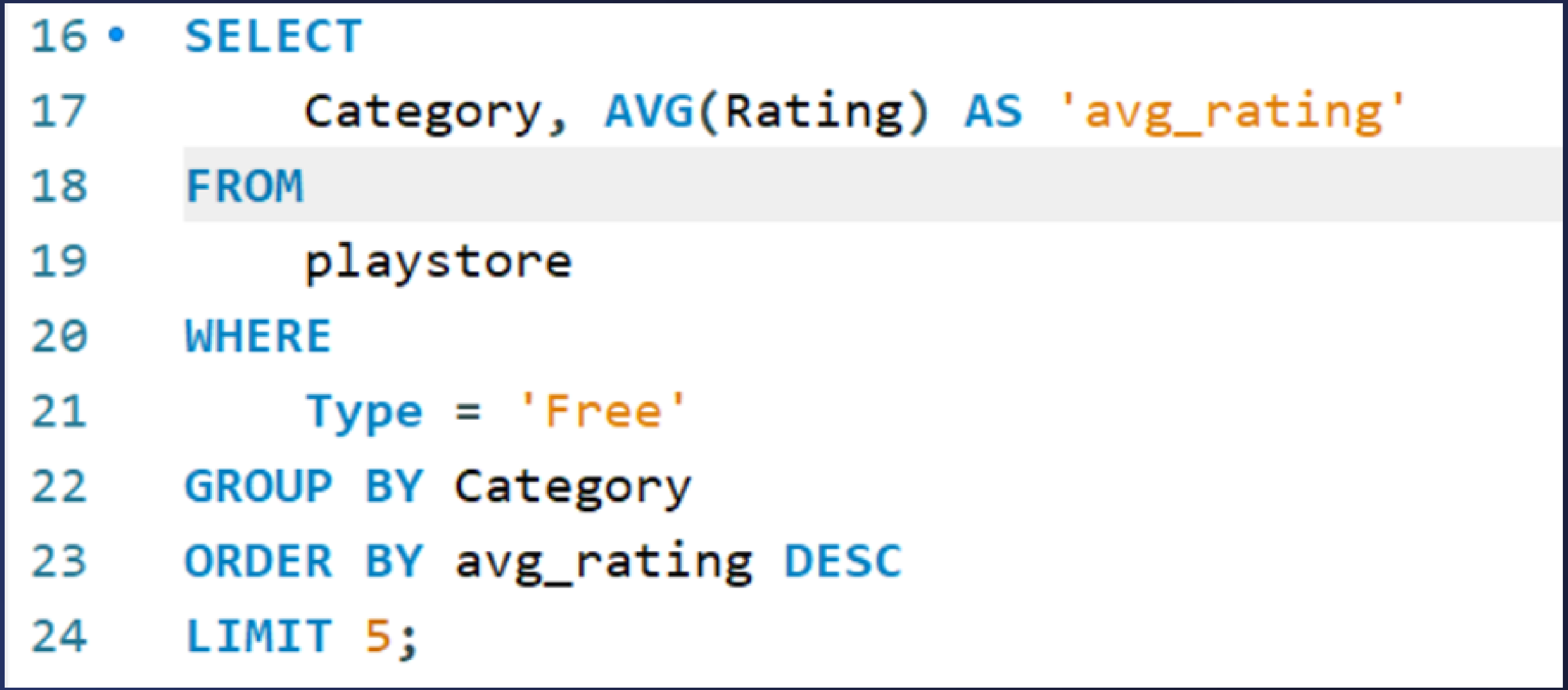


```
SELECT * FROM playstore.playstore;
truncate table playstore.playstore;

LOAD DATA  INFILE "D:/playstore.csv"
INTO TABLE playstore.playstore
FIELDS TERMINATED BY ','
OPTIONALLY ENCLOSED BY '"'
LINES TERMINATED BY '\r\n'
IGNORE 1 ROWS;
```

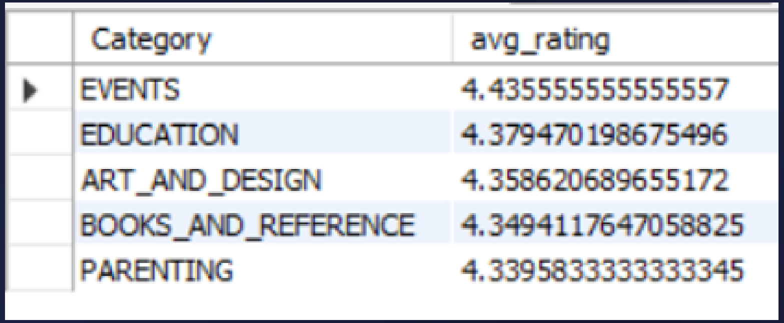
03

Performing analysis

1. You're 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.

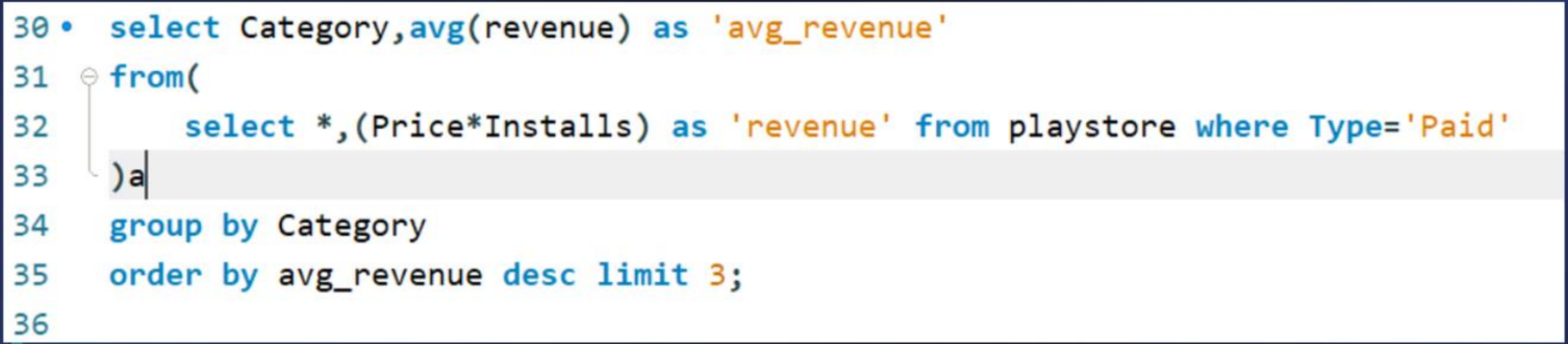


```sql
SELECT
    Category, AVG(Rating) AS 'avg_rating'
FROM
    playstore
WHERE
    Type = 'Free'
GROUP BY Category
ORDER BY avg_rating DESC
LIMIT 5;
```

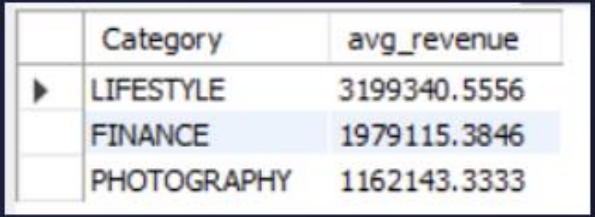


Category	avg_rating
EVENTS	4.435555555555557
EDUCATION	4.379470198675496
ART_AND_DESIGN	4.358620689655172
BOOKS_AND_REFERENCE	4.349117647058825
PARENTING	4.339583333333345

2. 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..

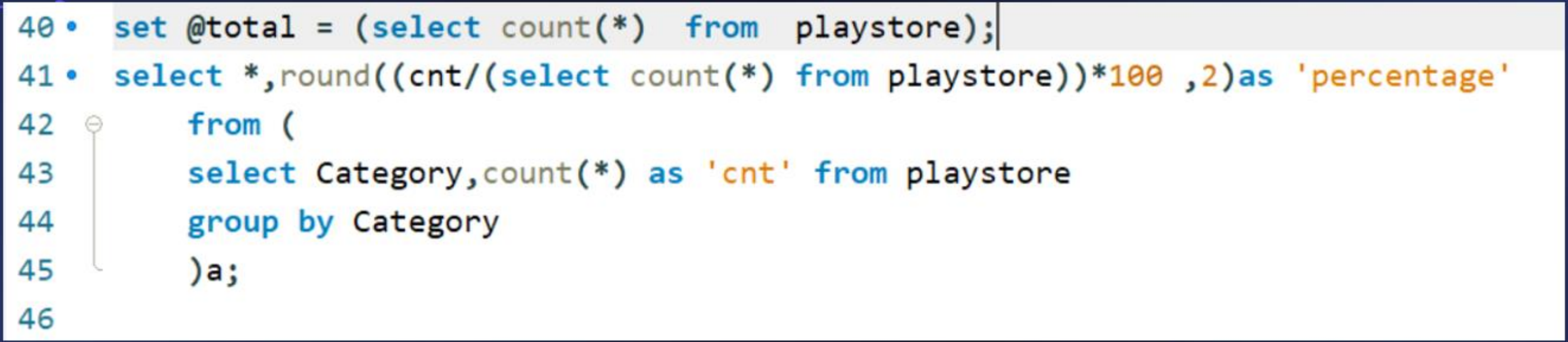


```
select Category,avg(revenue) as 'avg_revenue'
from(
    select *,(Price*Installs) as 'revenue' from playstore where Type='Paid'
)a
group by Category
order by avg_revenue desc limit 3;
```

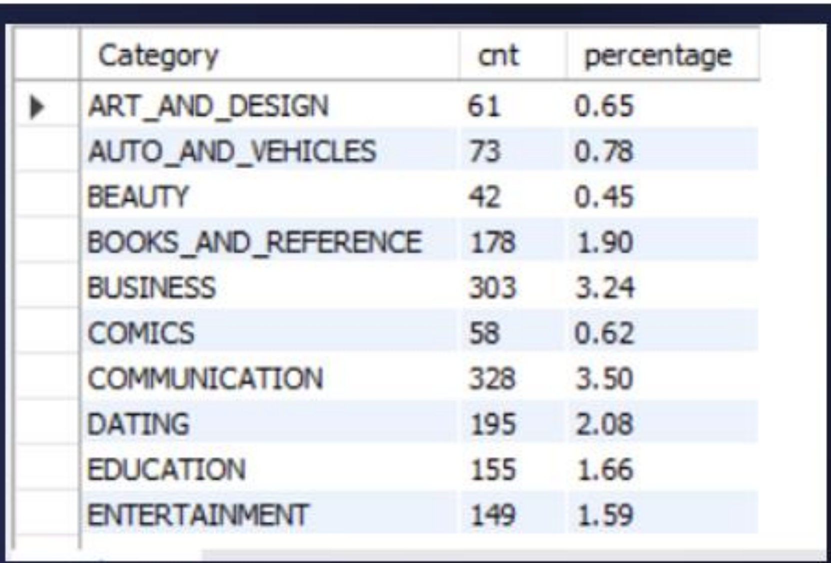


Category	avg_revenue
LIFESTYLE	3199340.5556
FINANCE	1979115.3846
PHOTOGRAPHY	1162143.3333

3. 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.

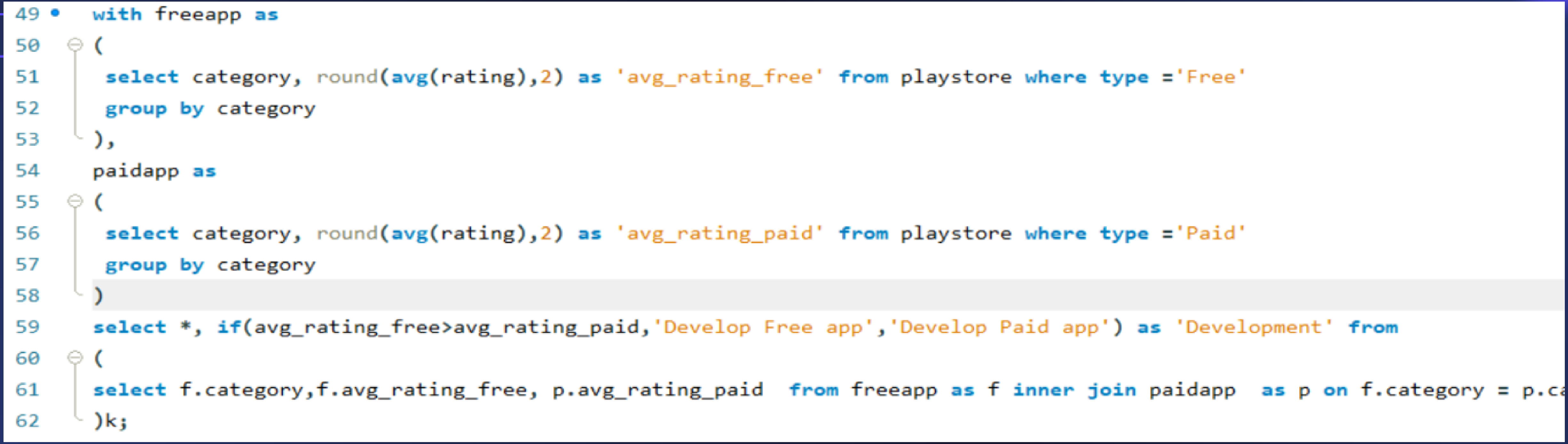


```sql
set @total = (select count(*)  from  playstore);
select *,round((cnt/(select count(*) from playstore))*100 ,2)as 'percentage'
    from (
    select Category,count(*) as 'cnt' from playstore
    group by Category
    )a;

```

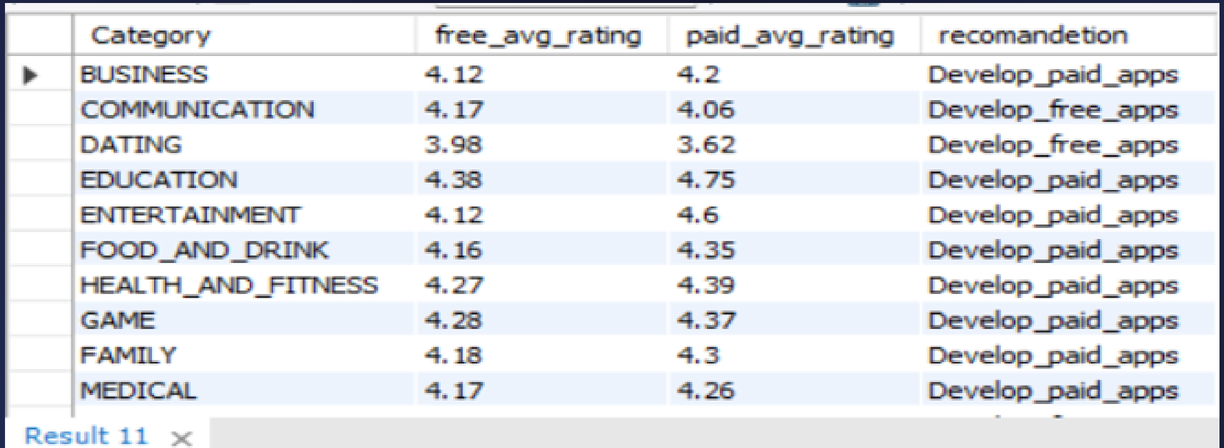


Category	cnt	percentage
ART_AND_DESIGN	61	0.65
AUTO_AND_VEHICLES	73	0.78
BEAUTY	42	0.45
BOOKS_AND_REFERENCE	178	1.90
BUSINESS	303	3.24
COMICS	58	0.62
COMMUNICATION	328	3.50
DATING	195	2.08
EDUCATION	155	1.66
ENTERTAINMENT	149	1.59

4. As a data analyst at a mobile app-focused market research firm, -- you'll recommend whether the company should develop paid or free apps for each category based on the ratings of that category.



```sql
with freeapp as
(
 select category, round(avg(rating),2) as 'avg_rating_free' from playstore where type ='Free'
 group by category
),
paidapp as
(
 select category, round(avg(rating),2) as 'avg_rating_paid' from playstore where type ='Paid'
 group by category
)
select *, if(avg_rating_free>avg_rating_paid,'Develop Free app','Develop Paid app') as 'Development' from
(
select f.category,f.avg_rating_free, p.avg_rating_paid  from freeapp as f inner join paidapp  as p on f.category = p.ca
)k;
```

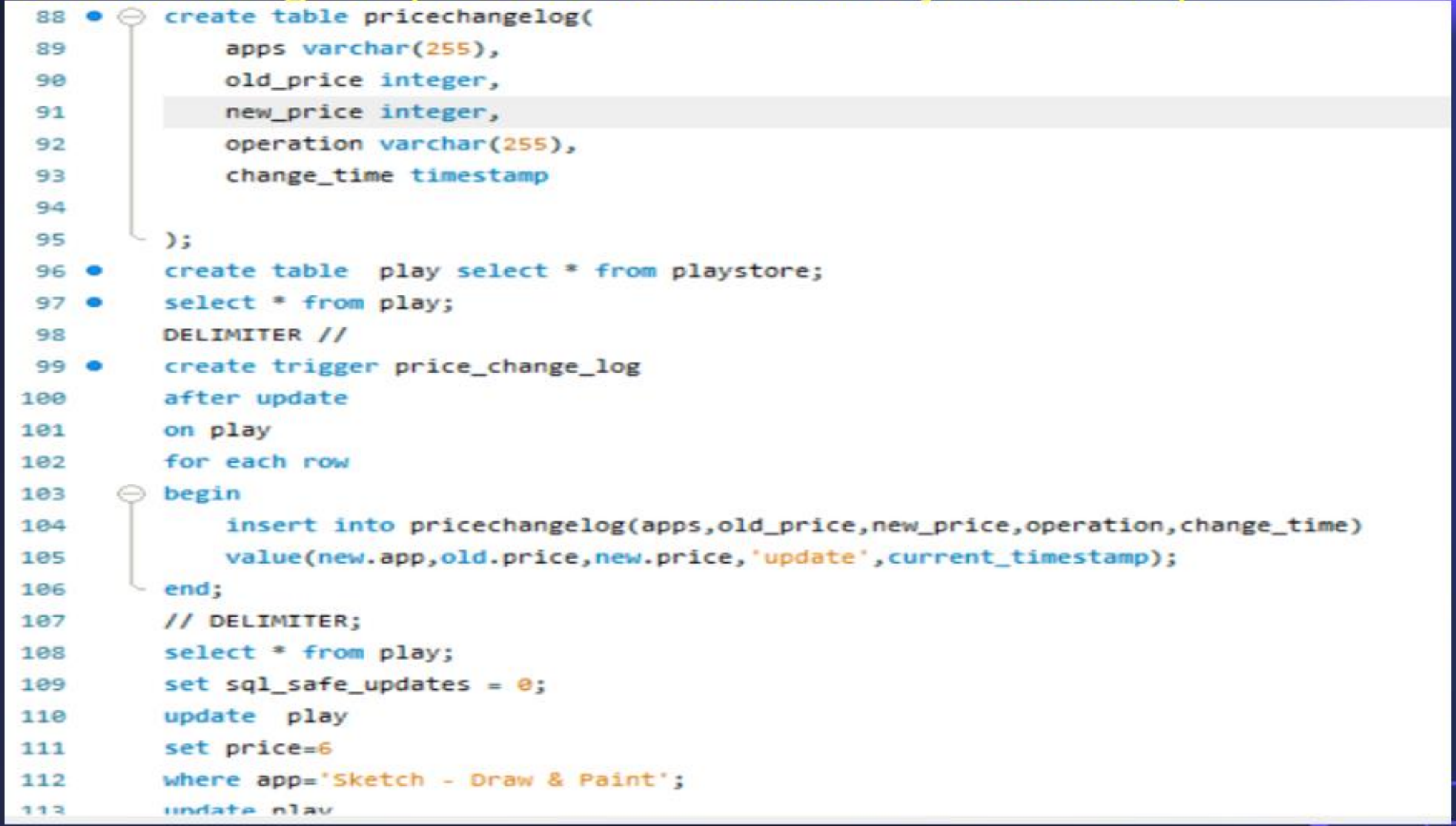


Category	free_avg_rating	paid_avg_rating	recomandetion
BUSINESS	4.12	4.2	Develop_paid_apps
COMMUNICATION	4.17	4.06	Develop_free_apps
DATING	3.98	3.62	Develop_free_apps
EDUCATION	4.38	4.75	Develop_paid_apps
ENTERTAINMENT	4.12	4.6	Develop_paid_apps
FOOD_AND_DRINK	4.16	4.35	Develop_paid_apps
HEALTH_AND_FITNESS	4.27	4.39	Develop_paid_apps
GAME	4.28	4.37	Develop_paid_apps
FAMILY	4.18	4.3	Develop_paid_apps
MEDICAL	4.17	4.26	Develop_paid_apps

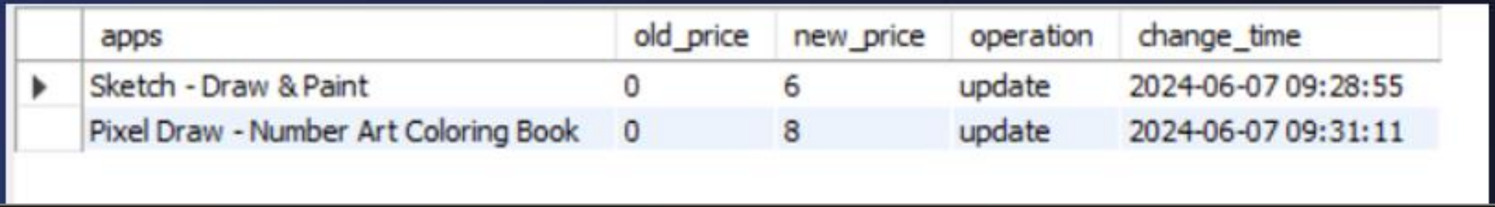
04

Handling Database Hacking

5.Suppose you're a database administrator, your databases have been hacked and hackers are changing price of certain apps on the database , its taking long for IT team to neutralize the hack , however you as a responsible manager dont want your data to be changed , do some measure where the changes in price can be recorded as you cant stop hackers from making changes

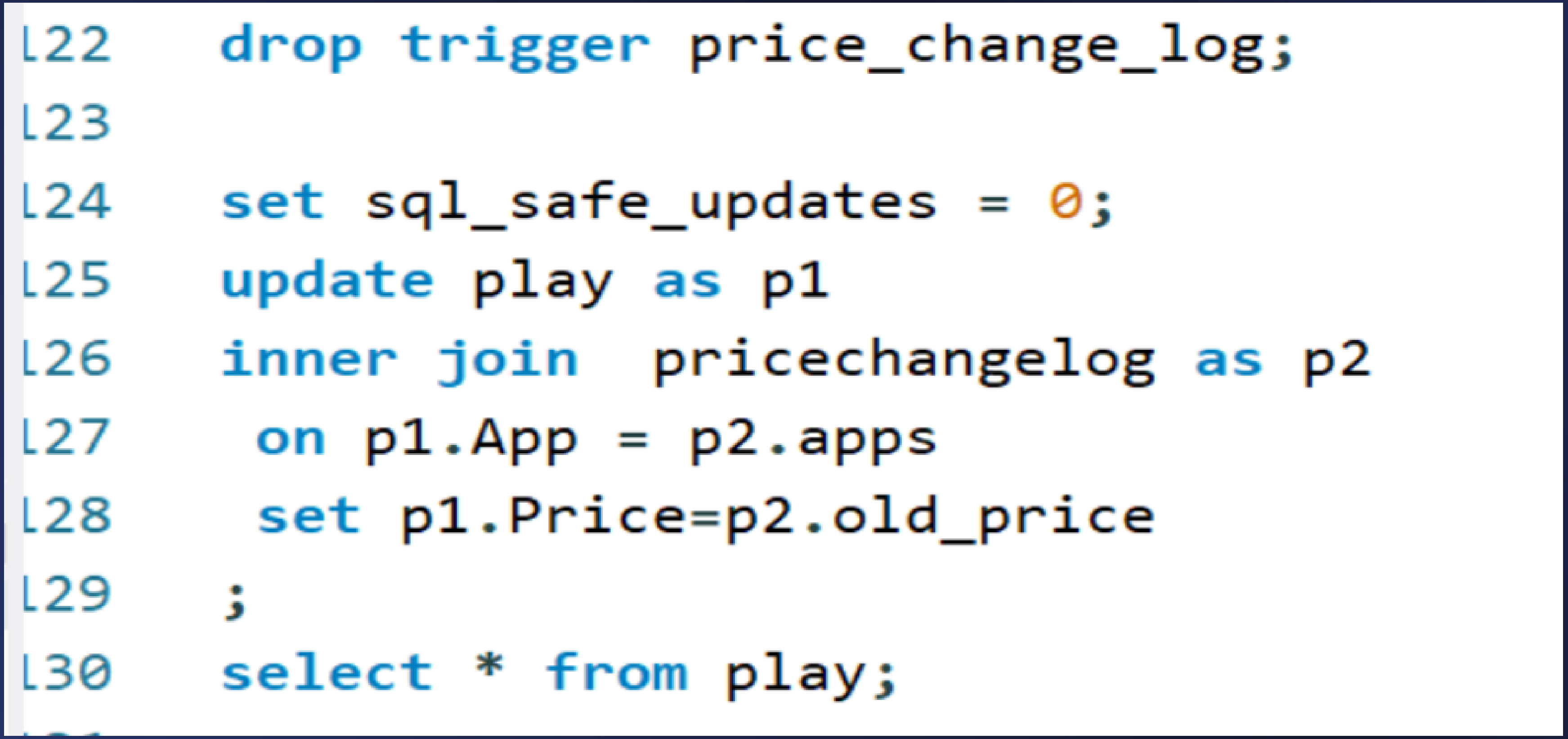


```
create table pricechangelog(
    apps varchar(255),
    old_price integer,
    new_price integer,
    operation varchar(255),
    change_time timestamp

);
create table  play select * from playstore;
select * from play;
DELIMITER //
create trigger price_change_log
after update
on play
for each row
begin
    insert into pricechangelog(apps,old_price,new_price,operation,change_time)
    value(new.app,old.price,new.price,'update',current_timestamp);
end;
// DELIMITER;
select * from play;
set sql_safe_updates = 0;
update  play
set price=6
where app='Sketch - Draw & Paint';
update play
```



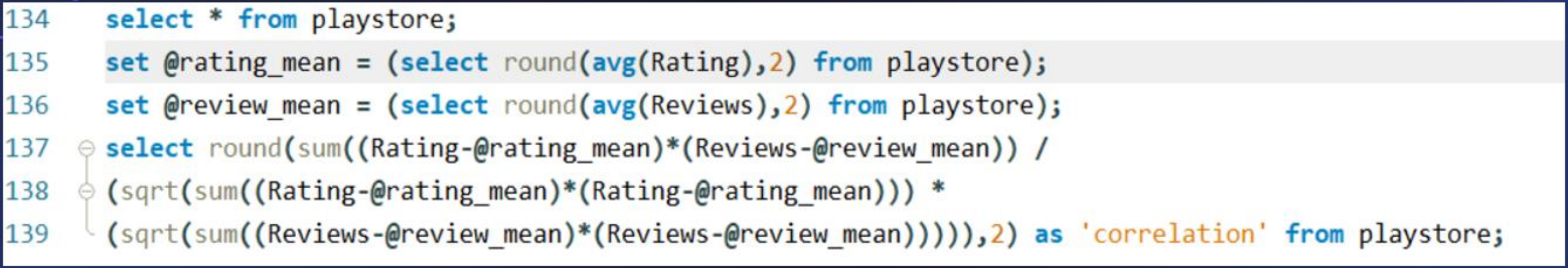
	apps	old_price	new_price	operation	change_time
▶	Sketch - Draw & Paint	0	6	update	2024-06-07 09:28:55
	Pixel Draw - Number Art Coloring Book	0	8	update	2024-06-07 09:31:11

6. your IT team have neutralize the threat, -- however hacker have made some changes in the prices, but becasue of your measure you have noted the changes , now you want-- correct data to be inserted into the database.

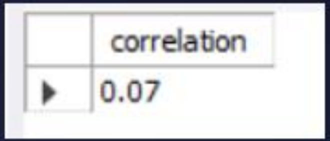


```
drop trigger price_change_log;

set sql_safe_updates = 0;
update play as p1
inner join  pricechangelog as p2
 on p1.App = p2.apps
 set p1.Price=p2.old_price
;
select * from play;
```

7. As a data person you are assigned the task to investigate the correlation between two numeric factors:-- app ratings and the quantity of reviews



```
select * from playstore;
set @rating_mean = (select round(avg(Rating),2) from playstore);
set @review_mean = (select round(avg(Reviews),2) from playstore);
select round(sum((Rating-@rating_mean)*(Reviews-@review_mean)) /
(sqrt(sum((Rating-@rating_mean)*(Rating-@rating_mean))) *
(sqrt(sum((Reviews-@review_mean)*(Reviews-@review_mean))))),2) as 'correlation' from playstore;
```

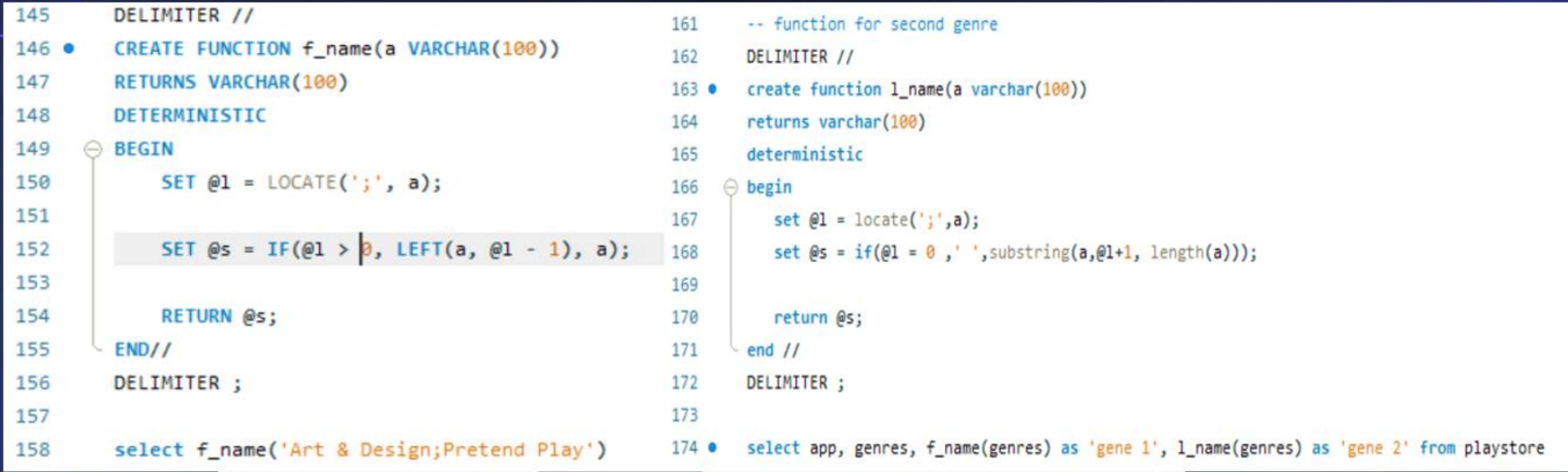


	correlation
▶	0.07

05

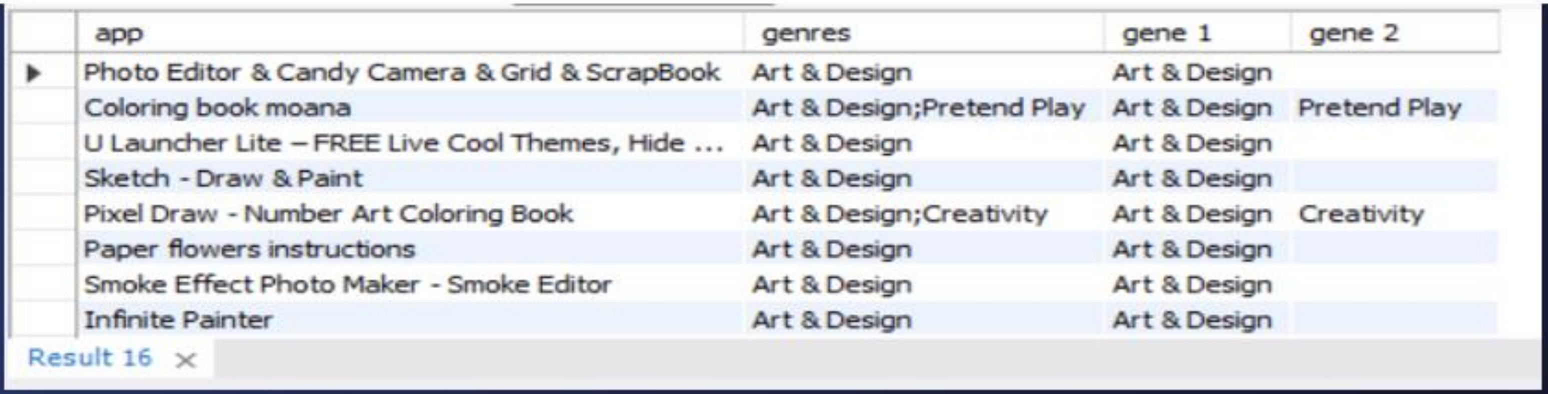
Data cleaning using Mysql

8. Your boss noticed that some rows in genres columns have multiple generes in them, which was creating issue when developing the recommendor system from the data he/she asssigned 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.



```sql
DELIMITER //
CREATE FUNCTION f_name(a VARCHAR(100))
RETURNS VARCHAR(100)
DETERMINISTIC
BEGIN
    SET @l = LOCATE(';', a);

    SET @s = IF(@l > 0, LEFT(a, @l - 1), a);

    RETURN @s;
END//
DELIMITER ;

select f_name('Art & Design;Pretend Play')
```

```sql
-- function for second genre
DELIMITER //
create function l_name(a varchar(100))
returns varchar(100)
deterministic
begin
    set @l = locate(';',a);
    set @s = if(@l = 0 ,' ',substring(a,@l+1, length(a)));

    return @s;
end //
DELIMITER ;

select app, genres, f_name(genres) as 'gene 1', l_name(genres) as 'gene 2' from playstore
```



	app	genres	gene 1	gene 2
▶	Photo Editor & Candy Camera & Grid & ScrapBook	Art & Design	Art & Design	
	Coloring book moana	Art & Design;Pretend Play	Art & Design	Pretend Play
	U Launcher Lite – FREE Live Cool Themes, Hide ...	Art & Design	Art & Design	
	Sketch - Draw & Paint	Art & Design	Art & Design	
	Pixel Draw - Number Art Coloring Book	Art & Design;Creativity	Art & Design	Creativity
	Paper flowers instructions	Art & Design	Art & Design	
	Smoke Effect Photo Maker - Smoke Editor	Art & Design	Art & Design	
	Infinite Painter	Art & Design	Art & Design	