PLAYSTORE APPS
SQL CASE STUDY

Presented by - DISHA VARSHNEY

ETL - EXTRACT TRANSFORM LOAD

- Data fetch - Kaggle
- Data was uncleaned . Performed data-cleaning on Python
- After the data was cleaned, tried to load the data to MySQL Workbench but the data was failed to import.
- Performed below SQL query and loaded the data into table “playstore”

```
load data  infile "C:/playstore.csv"
into table playstore
fields terminated by ','
optionally enclosed by '"'
lines terminated by '\n'
ignore 1 rows;
```

ABOUT THE SCHEMA

- **Database - projects**
- **Table - playstore**
- **Total entries after cleaning - 9360**

Columns	DataType
id	INT
App	TEXT
Category	TEXT
Rating	DOUBLE
Reviews	INT
Size	TEXT
Installs	INT
Type	TEXT
Price	DOUBLE
Content Rating	TEXT
Genres	TEXT
Last Updated	TEXT
Current Ver	TEXT
Android Ver	TEXT

DATA VIEW

select * from playstore

id	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
0	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
1	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
2	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
3	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
4	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
5	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
6	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
7	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
8	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
9	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

ANSWERING THE QUESTIONS

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.

```
select
    Category,
    round(AVG(Rating),2) as "avg_rating"
from playstore
where Type="Free"
group by 1
order by 2 desc
limit 5;
```

Category	avg_rating
EVENTS	4.44
EDUCATION	4.38
ART_AND_DESIGN	4.36
BOOKS_AND_REFERENCE	4.35
PARENTING	4.34

ANSWERING THE QUESTIONS

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,
   round(sum (Price*Installs), 2) as "revenue"
from playstore
where Type="Paid"
group by 1
order by 2 desc
limit 3
```

Category	revenue
FAMILY	185997086.7
LIFESTYLE	57583939.4
GAME	40986840.88

ANSWERING THE QUESTIONS

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

```
select
      Category,
      count(*) as "cnt",
    round(count(*) *100/ (select count(*) from playstore) ,2) as "%"
from playstore
group by 1
order by 2 desc
```

(Used a subquery here)

Category	cnt	%
FAMILY	1746	18.65
GAME	1097	11.72
TOOLS	733	7.83
PRODUCTIVITY	351	3.75
MEDICAL	350	3.74
COMMUNICATION	328	3.50
FINANCE	323	3.45
SPORTS	319	3.41
PHOTOGRAPHY	317	3.39
LIFESTYLE	314	3.35
PERSONALIZATION	312	3.33
BUSINESS	303	3.24
HEALTH_AND_FI...	297	3.17
SOCIAL	259	2.77
SHOPPING	238	2.54

ANSWERING THE QUESTIONS

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.

```
with t1 as
        (select
        Category, round(avg(rating),2) as "Paid"
        from playstore
        where type="Paid"
        group by 1),
t2 as
       (select
       Category, round(avg(rating),2) as "Free"
       from playstore
       where type="Free"
       group by 1)
```

(Made 2 ctes first)

ANSWERING THE QUESTIONS

```
select *,
    case
    when k.Paid>k.Free then "develop Paid apps"
    else "develop Free apps"
    end as "decision"
from
  (select
    a.Category, Paid, Free
   from t1 a
   inner join t2 b
   on a.Category=b.Category) k
```

(Used a subquery. First this subquery will be executed and then the above query)

Category	Paid	Free	decision
ART_AND_DESIGN	4.73	4.36	develop Paid apps
AUTO_AND_VEHICLES	4.6	4.18	develop Paid apps
BOOKS_AND_REFERENCE	4.28	4.35	develop Free apps
BUSINESS	4.2	4.12	develop Paid apps
COMMUNICATION	4.06	4.17	develop Free apps
DATING	3.62	3.98	develop Free apps
EDUCATION	4.75	4.38	develop Paid apps
ENTERTAINMENT	4.6	4.12	develop Paid apps
FINANCE	3.83	4.14	develop Free apps
FOOD_AND_DRINK	4.35	4.16	develop Paid apps
HEALTH_AND_FITNESS	4.39	4.27	develop Paid apps
LIFESTYLE	4.25	4.09	develop Paid apps

ANSWERING THE QUESTIONS

5 - 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

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

(Creating a duplicate table to perform updates on, as a hacker)

```
create table price_changelog (
    app VARCHAR(255),
    old_price decimal(10,2),
    new_price decimal(10,2),
    operation_type varchar(25),
    operation_date timestamp
);
```

(Creating a table where updates done by hackers will be recorded))

ANSWERING THE QUESTIONS

Everytime the hackers update the original data, this trigger will be executed and all the updated data can be recorded in the table “price_changelog”

```
DELIMITER //
CREATE TRIGGER price_change_hack
AFTER UPDATE
ON play
FOR EACH ROW
BEGIN
   INSERT INTO price_changelog(app, old_price, new_price, operation_type, operation_date)
   VALUES (NEW.app, OLD.price, NEW.price, 'update', CURRENT_TIMESTAMP);
END //
DELIMITER ;;
```

ANSWERING THE QUESTIONS

I tried to update the table “play” as a hacker to check if the updated data is stored in “price_changelog”

```
update play
set price=0
where Type="Paid"
```

```
select * from price_changelog
```

app	old_price	new_price	operation_type	operation_date
TurboScan: scan documents and receipts in PDF	4.99	0.00	update	2024-04-29 21:51:34
Tiny Scanner Pro: PDF Doc Scan	4.99	0.00	update	2024-04-29 21:51:34
TurboScan: scan documents and receipts in PDF	4.99	0.00	update	2024-04-29 21:51:34
Tiny Scanner Pro: PDF Doc Scan	4.99	0.00	update	2024-04-29 21:51:34
Puffin Browser Pro	3.99	0.00	update	2024-04-29 21:51:34
Moco+ - Chat, Meet People	3.99	0.00	update	2024-04-29 21:51:34
Calculator	6.99	0.00	update	2024-04-29 21:51:34

ANSWERING THE QUESTIONS

6 - Your IT team have neutralized the threat; however, hackers have made some changes in the prices, but because of your measure you have noted the changes, now you want correct data to be inserted into the database again.

```
drop trigger price_change_hack;
update play as a
join price_changelog b
on a.App=b.App
set a.price=b.old_price;
```

```
select * from play
where Type="Paid"
```

(Checking)

app	old_price	new_price	operation_type	operation_date
TurboScan: scan documents and receipts in PDF	4.99	0.00	update	2024-04-29 21:51:34
Tiny Scanner Pro: PDF Doc Scan	4.99	0.00	update	2024-04-29 21:51:34
TurboScan: scan documents and receipts in PDF	4.99	0.00	update	2024-04-29 21:51:34
Tiny Scanner Pro: PDF Doc Scan	4.99	0.00	update	2024-04-29 21:51:34
Puffin Browser Pro	3.99	0.00	update	2024-04-29 21:51:34
Moco+ - Chat, Meet People	3.99	0.00	update	2024-04-29 21:51:34
Calculator	6.99	0.00	update	2024-04-29 21:51:34

ANSWERING THE QUESTIONS

7 - 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 avg(Rating) from playstore)
set @y=(select avg(Reviews) from playstore);
```

(Setting the variables)

```
with cte as
  (select *,
   (rat*rat) as "sqr_x",
  (rev*rev) as "sqr_y"
  from
 (select
     Rating,
    round((Rating-@x),2) as 'rat',
    Reviews, round((Reviews-@y),2) as "rev"
 from playstore) k
)
```

(Making a cte)

(Extracting the result from subquery)

corr_coefficient
0.07

ANSWERING THE QUESTIONS

```
select
    @numerator:=sum(rat*rev),
     @deno_1:=sum(sqr_x),
    @deno_2:=sum(sqr_y)
from cte;

select
    round(@numerator/sqrt(@deno_1*@deno_2),2) as "corr_coefficient";
```

(Setting the variables)

corr_coefficient
0.07

ANSWERING THE QUESTIONS

8 - 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.

```
ALTER TABLE playstore
ADD COLUMN Genre_2 VARCHAR(255) AFTER Genres
```

(Adding a new column)

```
UPDATE playstore
SET Genre_2 = CASE
    WHEN Genres = SUBSTRING_INDEX(Genres, ";", -1)  THEN “”
    ELSE SUBSTRING_INDEX(Genres, ";", -1)
    END
```

(Updating the newly added columm)

```
select * from playstore
```

(Viewing the updated table)

id	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Genre_2	Last Updated	Current Ver	Andro Ver
0	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 a
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500000	Free	0	Everyone	Art & Design;Pretend Play	Pretend Play	2018-01-15	2.0.0	4.0.3 a
2	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 a
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50000000	Free	0	Teen	Art & Design		2018-06-08	Varies with device	4.2 an
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100000	Free	0	Everyone	Art & Design;Creativity	Creativity	2018-06-20	1.1	4.4 an
5	Paper flowers instructions	ART_AND_DESIGN	4.4	167	5.6M	50000	Free	0	Everyone	Art & Design		2017-03-26	1	2.3 an
6	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 a
7	Infinite Painter	ART_AND_DESIGN	4.1	36815	29M	1000000	Free	0	Everyone	Art & Design		2018-06-14	6.1.61.1	4.2 an
8	Garden Coloring Book	ART_AND_DESIGN	4.4	13791	33M	1000000	Free	0	Everyone	Art & Design		2017-09-20	2.9.2	3.0 an

ANSWERING THE QUESTIONS

9 -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.

(Creating a stored procedure)

```
CREATE DEFINER=`root`@`localhost` PROCEDURE
`check_app_performance`(IN input_category
VARCHAR(255))
BEGIN
    DECLARE avg DECIMAL(10,2);
    select average into avg from
    (select Category, avg(Rating) as "average" from playstore
    group by 1) t
    where Category=input_category;
 select * from playstore
    where Category=input_category and Rating<@avg;
END
```

(Calculating the average rating for that particular category)

(Retrieving those apps whose rating is lower than avg_rating of that particular category)

ANSWERING THE QUESTIONS

call check_app_performance('business')

id	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Genre_2	Last Updated	Current Ver	Android Ver
187	Visual Voicemail by MetroPCS	BUSINESS	4.1	16129	Varies with device	10000000	Free	0	Everyone	Business		2018-07-30	Varies with device	Varies with device
192	Docs To Go™ Free Office Suite	BUSINESS	4.1	217730	Varies with device	50000000	Free	0	Everyone	Business		2018-04-02	Varies with device	Varies with device
195	USPS MOBILE®	BUSINESS	3.9	16589	9.1M	1000000	Free	0	Everyone	Business		2018-06-14	4.9.10	4.0.3 and up
199	Alba Heaven - Alvarez Job Portal Services	BUSINESS	4	8941	12M	5000000	Free	0	Everyone	Business		2018-07-31	4.6.30	4.0.3 and up
201	Facebook Pages Manager	BUSINESS	4	1279184	Varies with device	50000000	Free	0	Everyone	Business		2018-08-02	Varies with device	Varies with device
207	Jobs in Alabama - Jobs in Alba	BUSINESS	4.1	11622	Varies with device	5000000	Free	0	Everyone	Business		2018-07-26	Varies with device	Varies with device
209	Plugin:AOT v5.0	BUSINESS	3.1	4034	23k	100000	Free	0	Everyone	Business		2015-09-11	3.0.1.11 (Build 311)	2.2 and up
210	Kariyer.net	BUSINESS	3.9	45964	16M	1000000	Free	0	Everyone	Business		2018-07-18	5.1.5	4.1 and up
212	Become a Job - Find a job or advertise	BUSINESS	4.1	6903	14M	1000000	Free	0	Everyone	Business		2018-07-10	2.3.6	4.1 and up
214	Easy Installer - Apps On SD	BUSINESS	4.1	23055	Varies with device	5000000	Free	0	Everyone	Business		2018-06-08	Varies with device	Varies with device
215	Facebook Ads Manager	BUSINESS	4.1	19023	Varies with device	1000000	Free	0	Everyone	Business		2018-08-01	99.0.0.35.75	4.1 and up
225	Secure Folder	BUSINESS	3.8	14760	8.6M	50000000	Free	0	Everyone	Business		2018-01-31	1.1.07.6	7.0 and up
226	UPS Mobile	BUSINESS	3.9	23243	Varies with device	5000000	Free	0	Everyone	Business		2018-06-25	Varies with device	Varies with device

ANSWERING THE QUESTIONS

10 - Calculate the average rating of apps in each category, where the rating is adjusted based on the number of installations. The adjustment factor is defined as the square root of the number of installations. Write a user-defined function to calculate the adjusted rating, and then use it to find the average adjusted rating for each category.

(Creating a user-defined function)

```
CREATE DEFINER=`root`@`localhost` FUNCTION `adjusted_rating`(rating varchar(10), installations int)
RETURNS float
    DETERMINISTIC
BEGIN
 declare adjusted_rating float;
    set adjusted_rating=rating/SQRT(installations);
RETURN adjusted_rating;
END
```

ANSWERING THE QUESTIONS

```
select
   Category,
    ROUND(AVG(adjusted_rating(Rating, Installs)),2) as "adjusted_rating"
from playstore
 GROUP BY 1
```

Category	adjusted_rating
ART_AND_DESIGN	0.03
AUTO_AND_VEHICLES	0.05
BEAUTY	0.04
BOOKS_AND_REFERENCE	0.06
BUSINESS	0.12
COMICS	0.03
COMMUNICATION	0.05
DATING	0.05
EDUCATION	0.01
ENTERTAINMENT	0
EVENTS	0.14
FINANCE	0.05
FOOD_AND_DRINK	0.03
HEALTH_AND_FITNESS	0.05
HOUSE_AND_HOME	0.01