Screen Test

Game Development Startup Analysis

Capstone Project 2 on Relational Database and MS Excel Dashboard

My Role

- **Data Analyst**

- **New startup gaming development company**

- **Analyzing the current gaming trend, sales, and popular consoles (from 2018 to 2022)**

- **Design and develop new games for the market**

Data Collection

About Dataset

- **Data source is from Kaggle**
  https://www.kaggle.com/datasets/patkle/video-game-sales-data-from-vgchartzcom

- **This dataset contains ~60k video games data from VGChartz:**

- **It includes:**
  - **Title**
  - **Sales (total, North America, Japan, PAL region + other)**
  - **Total shipped**
  - **Publisher**
  - **Developer**
  - **Release Date**
  - **Platform**
  - **Scores (user, VGChartz & critic score)**

Entity Relationship Diagram
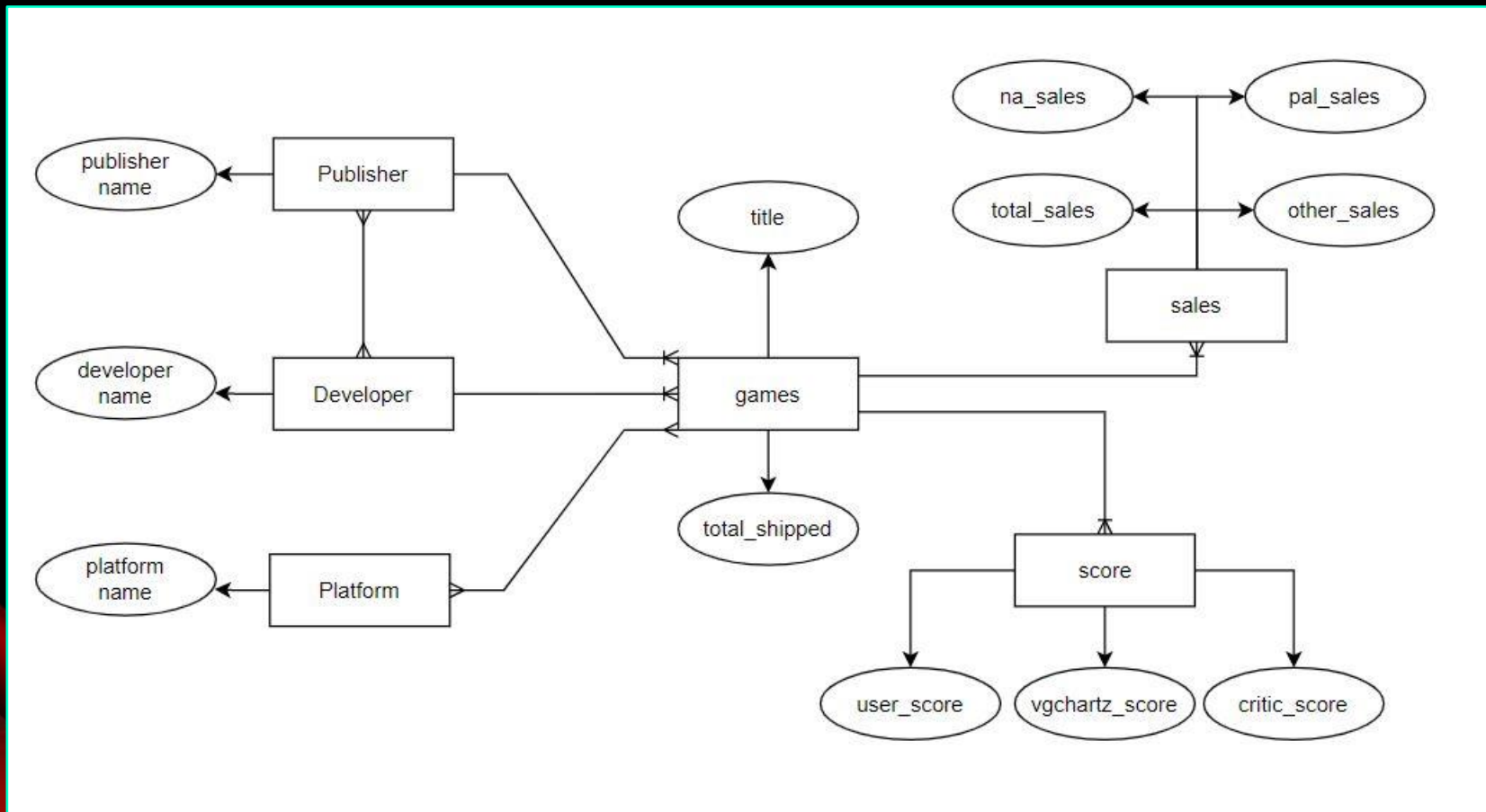
Relational Schema
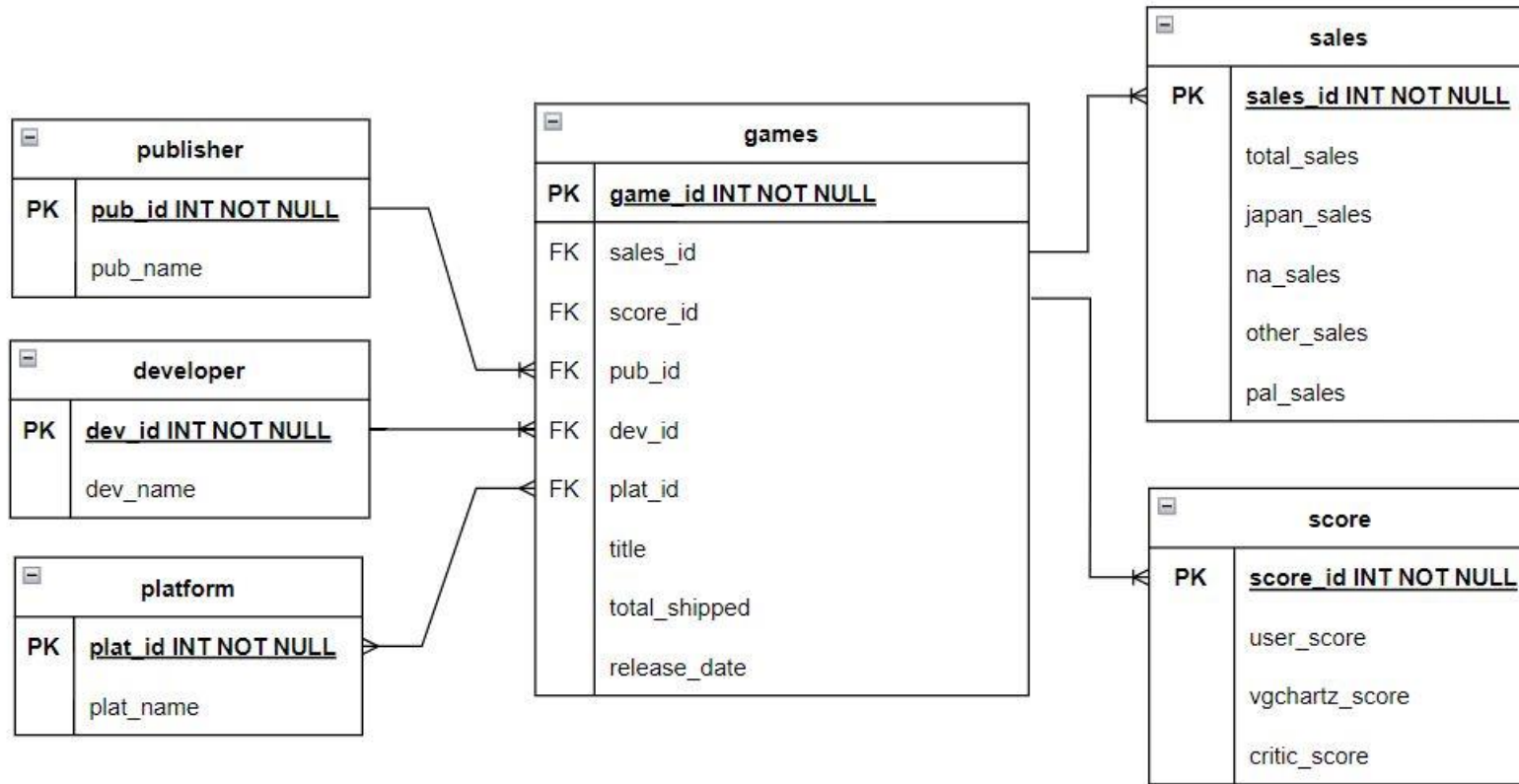
Data Transformation

- Raw data in .csv format

- Using SQLite, convert .csv into .sql

- Import .sql into pgAdmin

title text	total_sales text	total_shipped text	publisher text	developer text	release_date text	platform text	japan_sales text	na_sales text	other_sales text	pal_sales text	pos text	user_score text	vgchartz_score text	critic_score text	last_update text
Grand Th...	10.57m	N/A	Rockstar ...	Rockstar ...	29th Apr 08	PS3	0.44m	4.79m	1.62m	3.73m	365	9	9	10	N/A
Call of Du...	13.53m	N/A	Activision	Infinity W...	10th Nov 09	X360	0.08m	8.54m	1.28m	3.63m	280	9	9	9.5	N/A
Grand Th...	8.72m	N/A	Rockstar ...	Rockstar ...	18th Nov 14	XOne	0.01m	4.70m	0.76m	3.25m	456	9	N/A	9	11th Apr 18
Call of Du...	9.41m	N/A	Activision	Infinity W...	05th Nov 07	X360	0.13m	5.98m	0.91m	2.39m	427	9	N/A	9.6	N/A
Call of Du...	7.50m	N/A	Activision	Treyarch	10th Nov 08	X360	N/A	4.88m	0.71m	1.91m	512	7.8	8.8	8.5	N/A
Uncharte...	6.74m	N/A	Sony Co...	Naughty ...	13th Oct 09	PS3	0.21m	3.28m	1.02m	2.22m	568	9.6	9.2	9.5	N/A
Call of Du...	6.72m	N/A	Activision	Infinity W...	05th Nov 07	PS3	0.28m	3.13m	1.04m	2.27m	570	9.6	N/A	9.5	N/A
Red Dead...	6.50m	N/A	Rockstar ...	Rockstar ...	18th May 10	X360	0.09m	3.77m	0.60m	2.03m	590	10	9.1	9.5	N/A
Assassin'...	5.55m	N/A	Ubisoft	Ubisoft M...	13th Nov 07	X360	0.07m	3.28m	0.56m	1.64m	672	8	7.5	8.2	N/A
Forza Mo...	5.50m	N/A	Microsoft...	Turn 10 S...	27th Oct 09	X360	0.10m	2.99m	0.51m	1.90m	674	9.8	N/A	9.2	N/A
Final Fant...	5.35m	N/A	Square E...	Square E...	09th Mar 10	PS3	1.87m	1.75m	0.51m	1.23m	693	9.2	9	8	N/A
Resident ...	5.10m	N/A	Capcom	Capcom	13th Mar 09	PS3	1.08m	1.96m	0.64m	1.42m	724	8.8	8.4	8.6	N/A
Final Fant...	5.07m	N/A	Square E...	Square E...	29th Nov 16	PS4	1.05m	1.81m	0.68m	1.53m	733	9	N/A	8.1	04th May 18

Data Cleaning 1

- Using SQL to clean data.

- Remove empty rows (if any)

- Remove unwanted columns (example: column "pos")

- Replaced all N/A with NULL

- Create new tables, and primary keys.

Data Cleaning 2

- **Remove the last letter 'm' (as in *millions*)**
- **Multiply (x1000000) into these fields.**
- **Change data type.**

```
UPDATE games SET total_sales = LEFT(total_sales, LENGTH(total_sales)-1);
UPDATE games SET total_sales = (total_sales::numeric *1000000) WHERE total_sales IS NOT NULL;
ALTER TABLE games ALTER COLUMN total_sales TYPE NUMERIC(15,1) USING total_sales::NUMERIC;
```

Before

total_sales text	japan_sales text	na_sales text	other_sales text	pal_sales text
16.15m	0.47m	8.41m	1.78m	5.49m
15.86m	0.06m	9.06m	1.42m	5.33m
15.09m	0.41m	6.18m	2.44m	6.05m
20.32m	0.99m		12m	9.85m
12.67m	0.48m	6.01m	1.78m	4.40m
10.94m	0.12m	1.26m	1.61m	7.95m
19.39m	0.60m	6.06m	3.02m	9.71m
10.61m	0.38m	4.99m	1.59m	3.66m
10.57m	0.44m	4.79m	1.62m	3.73m
11.80m	0.15m	1.27m	1.73m	8.64m

After

total_sales numeric (15)	japan_sales numeric (15)	na_sales numeric (15)	other_sales numeric (15)	pal_sales numeric (15)
750000	180000	80000	90000	410000
7200000	80000	3270000	1020000	2830000
480000	290000	160000	10000	10000
3340000	40000		590000	570000
1460000	40000	1170000	50000	200000
360000	270000	40000	10000	30000
3530000	30000	2870000	250000	390000
6740000	210000	3280000	1020000	2220000
160000	30000	70000	10000	50000
220000	20000	140000	20000	50000

Data Cleaning 3

```
UPDATE games SET release_date = substring("release_date",1,2) || ' ' || substring("release_date",6,6);
UPDATE games SET release_date = to_date("release_date", 'DD Mon YY') WHERE "release_date" is not null;
ALTER TABLE games ALTER COLUMN release_date SET DATA TYPE DATE USING release_date::DATE;
```

- Remove characters at positions 3 & 4.

- Change text format (DD Mon YY) into yyyy-mm-dd.

ISO 8601

- Change Data Type.

release_date text	last_update text
10th Feb 08	04th Feb 20
15th Nov 05	20th Oct 20
25th Feb 22	28th Feb 22
29th	20
12th Feb 98	09th Aug 18
22nd Mar 99	11th Mar 20
01st Nov 96	03rd Feb 20
20th Dec 96	12th Mar 20
12th Dec 03	16th Feb 20
12th Dec 03	30th Nov 20

Before

release_date date	last_update date
2019-03-28	2019-03-23
2020-07-24	2020-07-02
2020-06-24	2019-04-08
2016-	27
2020-12-01	2020-00-17
2019-06-25	2021-06-16
1991-07-19	2017-12-31
1983-01-01	2018-01-13
2018-02-06	2021-07-26
2017-11-09	2018-03-26

After

Data Cleaning 4

After Data Cleaning process:

- 6 tables

- 39,206 game titles

- 3,366 Publishers

- 8,752 Developers

- 80 Platforms

Tables

- **games**

title text	total_shipped numeric (15)	release_date date	game_id [PK] integer	pub_id integer	dev_id integer	plat_id integer	sales_id integer	score_id integer
Test Drive: Off-Road	500000	1997-03-27	6007	1601	7905	71	4810	341
Survivor Squad	120000	2013-11-08	12985	453	219	57	4810	341
Game Corp DX	110000	2015-10-02	13673	453	219	57	4810	341

- **sales**

sales_id [PK] integer	total_sales numeric (15)	japan_sales numeric (15)	na_sales numeric (15)	other_sales numeric (15)	pal_sales numeric (15)
3	750000	180000	80000	90000	410000
4	1270000	0	740000	120000	400000
6	680000	0	300000	60000	320000

- **score**

score_id [PK] integer	user_score numeric (4,1)	vgchartz_score numeric (4,1)	critic_score numeric (4,1)
2	7.8	8.8	8.5
7	8.6	8.2	9.0
9	9.1	8.7	8.2

- **publisher**

pub_id [PK] integer	pub_name text
1	T3 Entertainment
2	gumi Inc.
3	Ghostlight

- **developer**

dev_id [PK] integer	dev_name text
1	Torus Games Pty. Ltd.
2	Just A Pixel Ltd.
3	Smilegate RPG

- **platform**

plat_id [PK] integer	plat_name text
1	PCE
2	PSP
3	GIZ

SQL Queries 1

```sql
SELECT  title            AS game_title,
        SUM(total_sales) AS total_sales
FROM games
INNER JOIN sales
    ON sales.sales_id = games.sales_id
WHERE release_date IS NOT NULL
    AND total_sales IS NOT NULL
    AND release_date >= '2018-01-01'
    AND release_date <= '2022-12-31'
GROUP BY title
ORDER BY 2 DESC
LIMIT 10;
```

- **Game titles with the highest sales (from 2018 to 2022)**

	game_title text	total_sales numeric	publisher text	developer text
1	Red Dead Redemption 2	19710000	Rockstar Gam...	Rockstar Games
2	Call of Duty: Black Ops IIII	14300000	Activision	Treyarch
3	FIFA 19	12180000	Electronic Arts	EA Sports
4	NBA 2K19	4670000	2K Sports	Visual Concepts
5	Assassin's Creed Odyssey	4630000	Ubisoft	Ubisoft Quebec
6	Far Cry 5	4020000	Ubisoft	Ubisoft
7	Battlefield V	3870000	Electronic Arts	EA DICE
8	Madden NFL 19	3720000	EA Sports	EA Tiburon
9	Spyro Reignited Trilogy	2820000	Activision	Toys for Bob
10	Shadow of the Tomb Rai...	2520000	Square Enix	Eidos Montreal

SQL Queries 2

```sql
SELECT pub_name AS publisher,
       count (distinct(title)) AS No_of_game_titles_released
FROM games
INNER JOIN publisher
    ON games.pub_id = publisher.pub_id
WHERE pub_name != 'Unknown'
    AND release_date >= '2018-01-01'
    AND release_date <= '2022-12-31'
GROUP BY pub_name
ORDER BY 2 DESC
LIMIT 10;
```

- **Publisher with the highest title released (from 2018 – 2022)**

	publisher text	no_of_game_titles_release bigint
1	Nintendo	94
2	Square Enix	69
3	Sega	62
4	Bandai Namco Entertainment	59
5	Hamster Corporation	51
6	Ubisoft	49
7	Capcom	49
8	Sony Interactive Entertainment	49
9	THQ Nordic	40
10	Koei Tecmo	39

SQL Queries 3

```sql
SELECT plat_name AS platform,
        count (distinct(title)) AS No_of_game_titles_release
FROM games
--------------------------------------------------------------
INNER JOIN platform
    ON games.plat_id = platform.plat_id
--------------------------------------------------------------
WHERE release_date >= '2018-01-01'
AND release_date <= '2022-12-31'

GROUP BY 1
ORDER BY 2 DESC
LIMIT 10;
```

- **Platform with highest title released, (from 2018 – 2022)**

	platform 🔒 text	no_of_game_titles_release 🔒 bigint
1	NS	1258
2	PC	992
3	PS4	977
4	XOne	666
5	All	195
6	PSV	123
7	PS5	107
8	XS	101
9	And	42
10	OSX	40

SQL Queries 4

```sql
SELECT plat_name          AS platform,
    sum(total_sales)      AS total_sales
FROM games
INNER JOIN sales
    ON sales.sales_id = games.sales_id
INNER JOIN platform
    ON games.plat_id = platform.plat_id
WHERE total_sales IS NOT NULL
AND release_date >= '2018-01-01'
AND release_date <= '2022-12-31'
GROUP BY plat_name
ORDER BY 2 DESC
LIMIT 10;
```

- **Platforms with highest Total Sales, (from 2018 to 2022)**

	platform text	total_sales numeric
1	PS4	86590000
2	XOne	38150000
3	NS	20780000
4	3DS	3630000
5	PC	1620000
6	PSV	1390000
7	PSP	1060000
8	GBA	660000
9	Wii	420000
10	PS3	350000

SQL Queries 5

```sql
SELECT plat_name          AS platform,
    sum(japan_sales)      AS japan_sales,
    sum(na_sales)         AS na_sales,
    sum(pal_sales)        AS pal_sales,
    sum(other_sales)      AS other_sales
FROM games
INNER JOIN sales
    ON sales.sales_id = games.sales_id
INNER JOIN platform
    ON games.plat_id = platform.plat_id
WHERE release_date >= '2018-01-01'
AND release_date <= '2022-12-31'
AND plat_name = 'NS'
OR plat_name = 'XOne'
OR plat_name = 'PS4'
GROUP BY plat_name
ORDER by 2 DESC;
```

- **Popular platforms with total sales in 4 regions, (from 2018 to 2022)**

	platform text	japan_sales numeric	na_sales numeric	pal_sales numeric	other_sales numeric
1	PS4	31740000	199830000	243190000	86930000
2	NS	2870000	10540000	5720000	1540000
3	XOne	380000	165040000	79740000	24630000
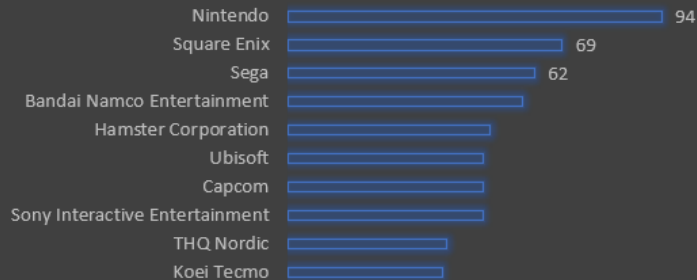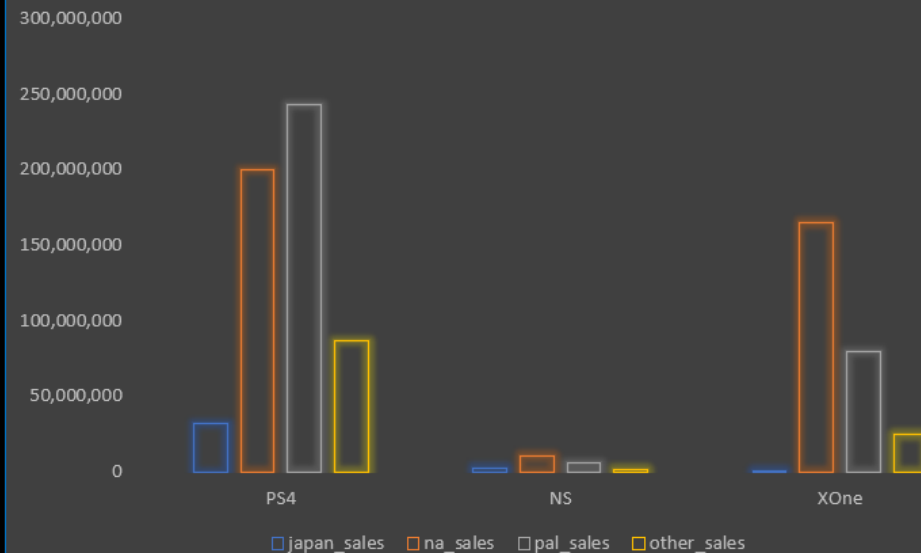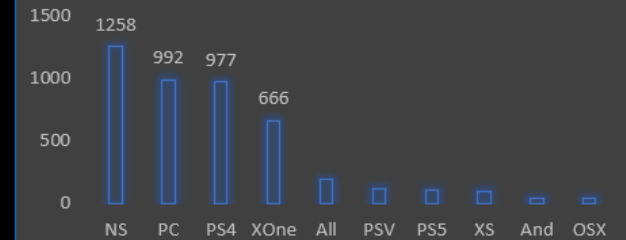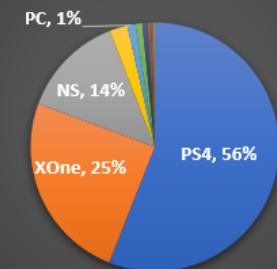
Insights 2 - Conclusion

From the Dashboard, recommendations for the upcoming development are as follows:

- The '*actions*' styles of games are more popular.

- Seek cooperation with **Nintendo** as the game's main publisher.

- Focus on running the new game on *PS (PlayStation)* as the main platform.

- Primary markets are **North America** and the **PAL region**.

Thank you

Q & A

Please be 'kind' to me....