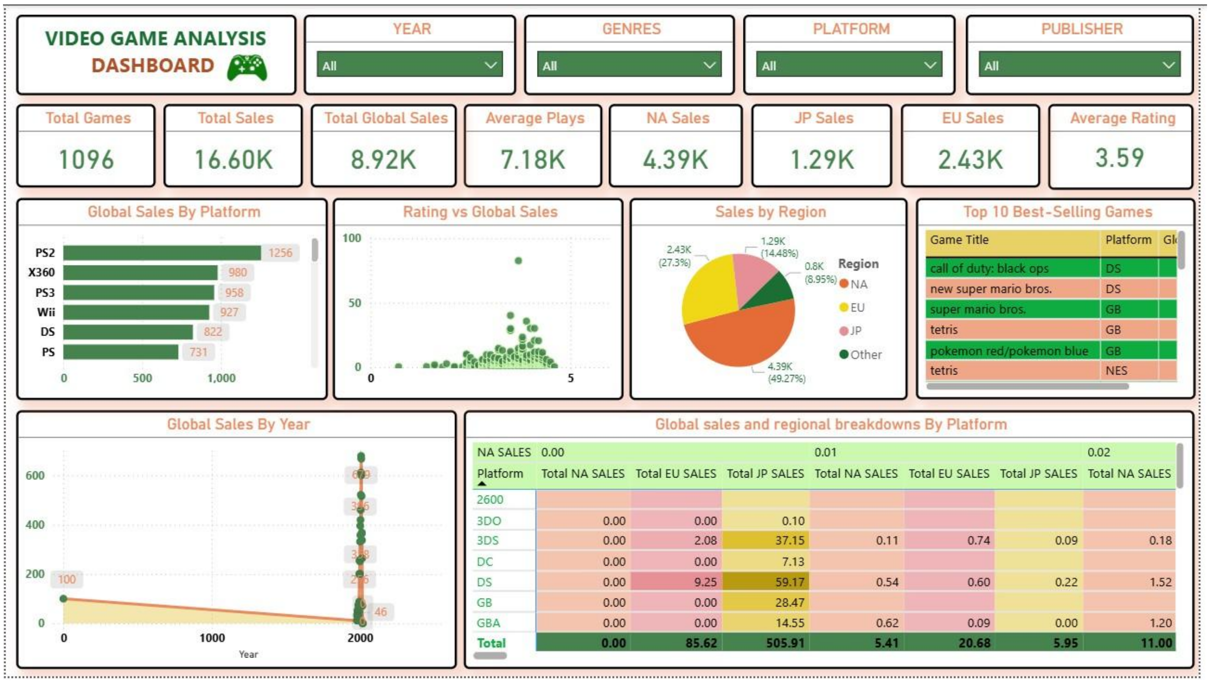
Video Game Data Analytics

Dashboard Screenshots



VIDEO GAME ANALYSIS DASHBOARD

YEAR: All

GENRES: All

PLATFORM: All

PUBLISHER: All

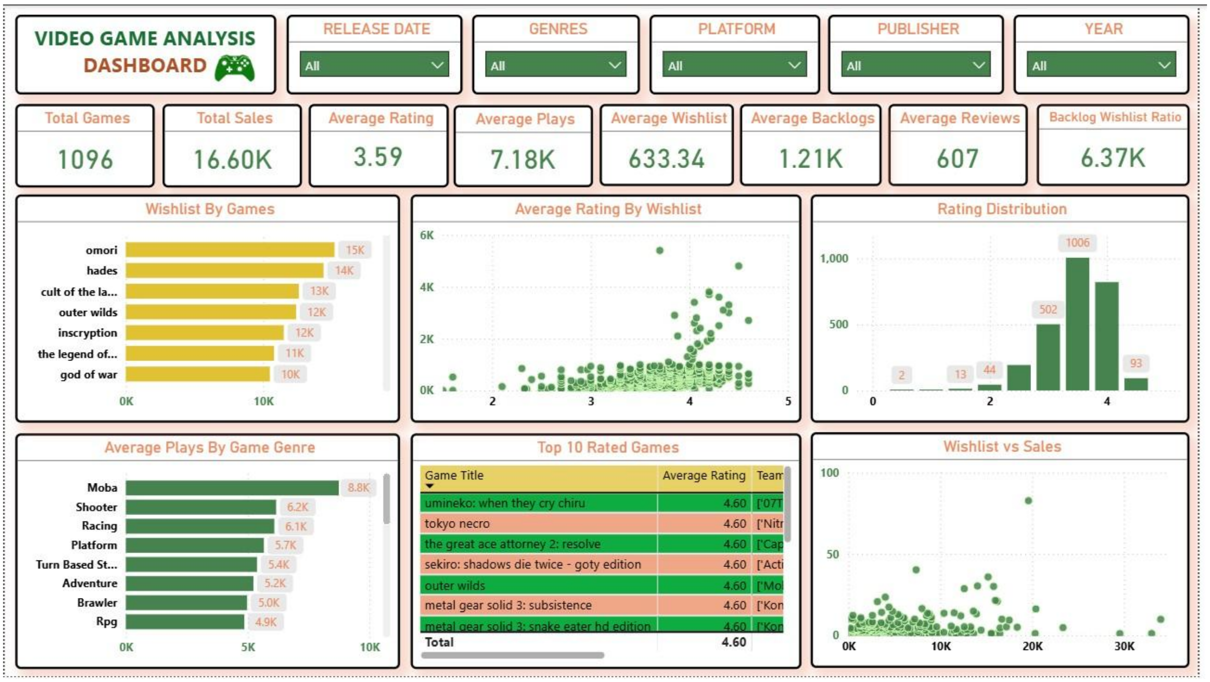
Total Games	Total Sales	Total Global Sales	Average Plays	NA Sales	JP Sales	EU Sales	Average Rating
1096	16.60K	8.92K	7.18K	4.39K	1.29K	2.43K	3.59

Top 10 Best-Selling Games

Game Title	Platform	Gl
call of duty: black ops	DS	
new super mario bros.	DS	
super mario bros.	GB	
tetris	GB	
pokemon red/pokemon blue	GB	
tetris	NES	

Global sales and regional breakdowns By Platform

NA SALES	0.00			0.01			0.02
Platform	Total NA SALES	Total EU SALES	Total JP SALES	Total NA SALES	Total EU SALES	Total JP SALES	Total NA SALES
2600							
3DO	0.00	0.00	0.10				
3DS	0.00	2.08	37.15	0.11	0.74	0.09	0.18
DC	0.00	0.00	7.13				
DS	0.00	9.25	59.17	0.54	0.60	0.22	1.52
GB	0.00	0.00	28.47				
GBA	0.00	0.00	14.55	0.62	0.09	0.00	1.20
Total	**0.00**	**85.62**	**505.91**	**5.41**	**20.68**	**5.95**	**11.00**



VIDEO GAME ANALYSIS DASHBOARD

RELEASE DATE: All

GENRES: All

PLATFORM: All

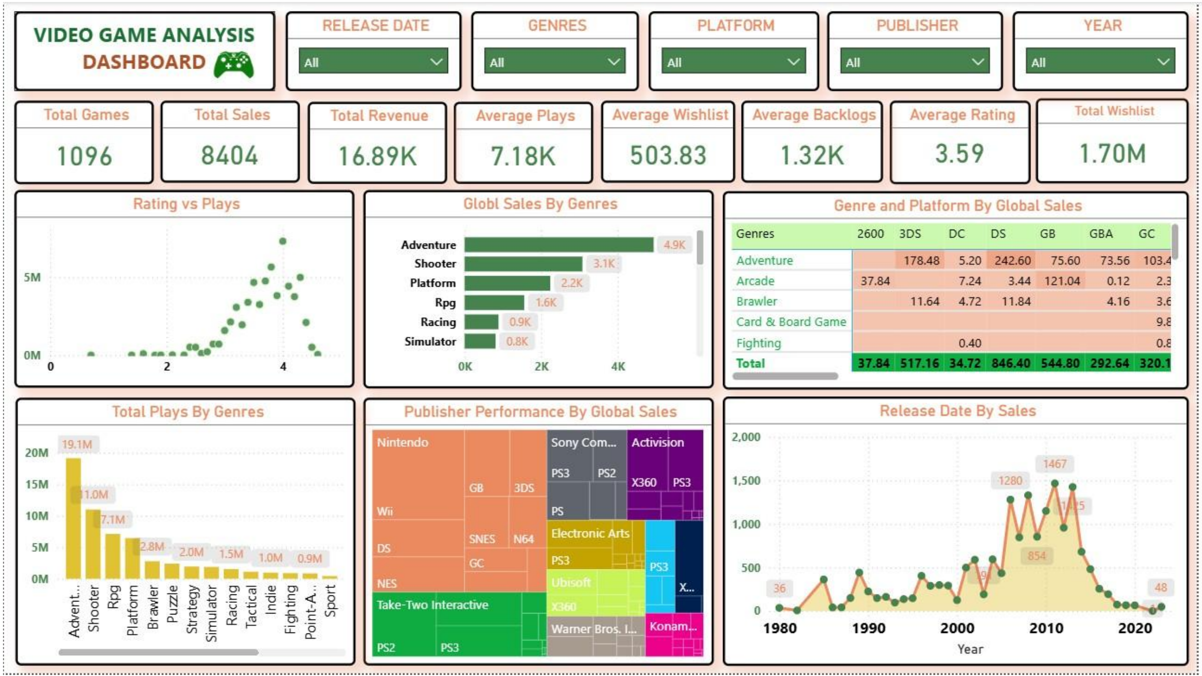
PUBLISHER: All

YEAR: All

Total Games	Total Sales	Average Rating	Average Plays	Average Wishlist	Average Backlogs	Average Reviews	Backlog Wishlist Ratio
1096	16.60K	3.59	7.18K	633.34	1.21K	607	6.37K

Top 10 Rated Games

Game Title	Average Rating	Team
umineko: when they cry chiru	4.60	['07T
tokyo necro	4.60	['Nitr
the great ace attorney 2: resolve	4.60	['Cap
sekiro: shadows die twice - goty edition	4.60	['Acti
outer wilds	4.60	['Mo
metal gear solid 3: subsistence	4.60	['Kon
metal gear solid 3: snake eater hd edition	4.60	['Kon
Total	**4.60**	

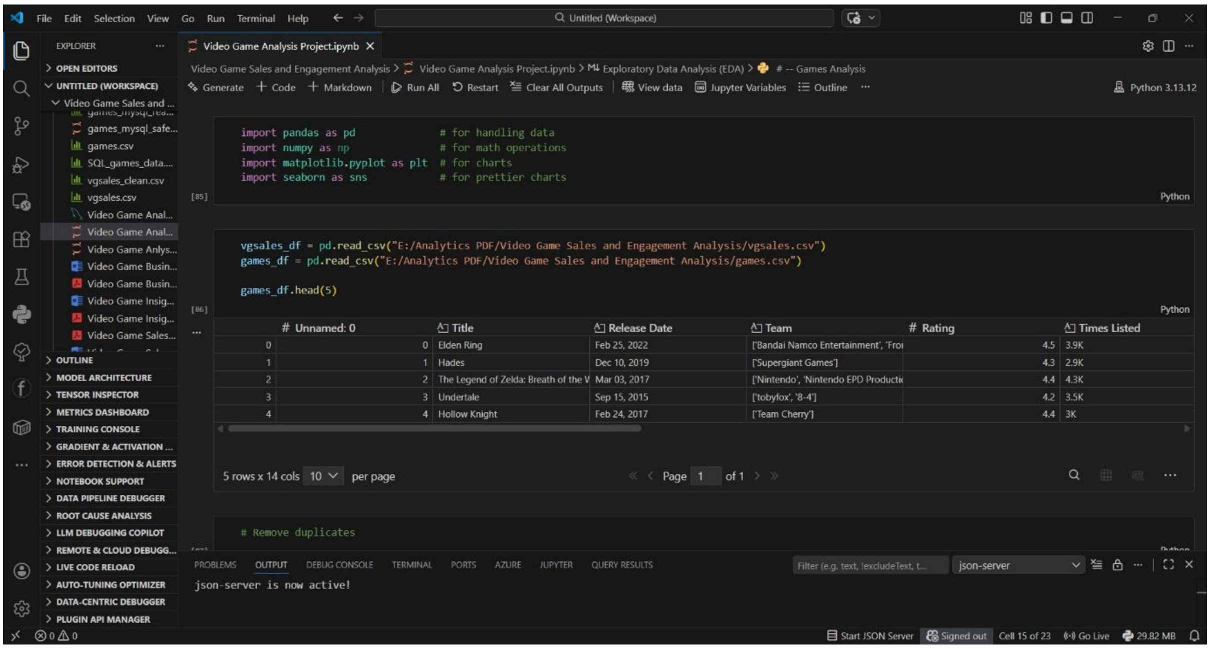


Total Games	Total Sales	Total Revenue	Average Plays	Average Wishlist	Average Backlogs	Average Rating	Total Wishlist
1096	8404	16.89K	7.18K	503.83	1.32K	3.59	1.70M

Genre and Platform By Global Sales

Genres	2600	3DS	DC	DS	GB	GBA	GC
Adventure		178.48	5.20	242.60	75.60	73.56	103.4
Arcade	37.84		7.24	3.44	121.04	0.12	2.3
Brawler		11.64	4.72	11.84		4.16	3.6
Card & Board Game							9.8
Fighting			0.40				0.8
Total	**37.84**	**517.16**	**34.72**	**846.40**	**544.80**	**292.64**	**320.1**

PYTHON CODE EDA SCREENSHOTS



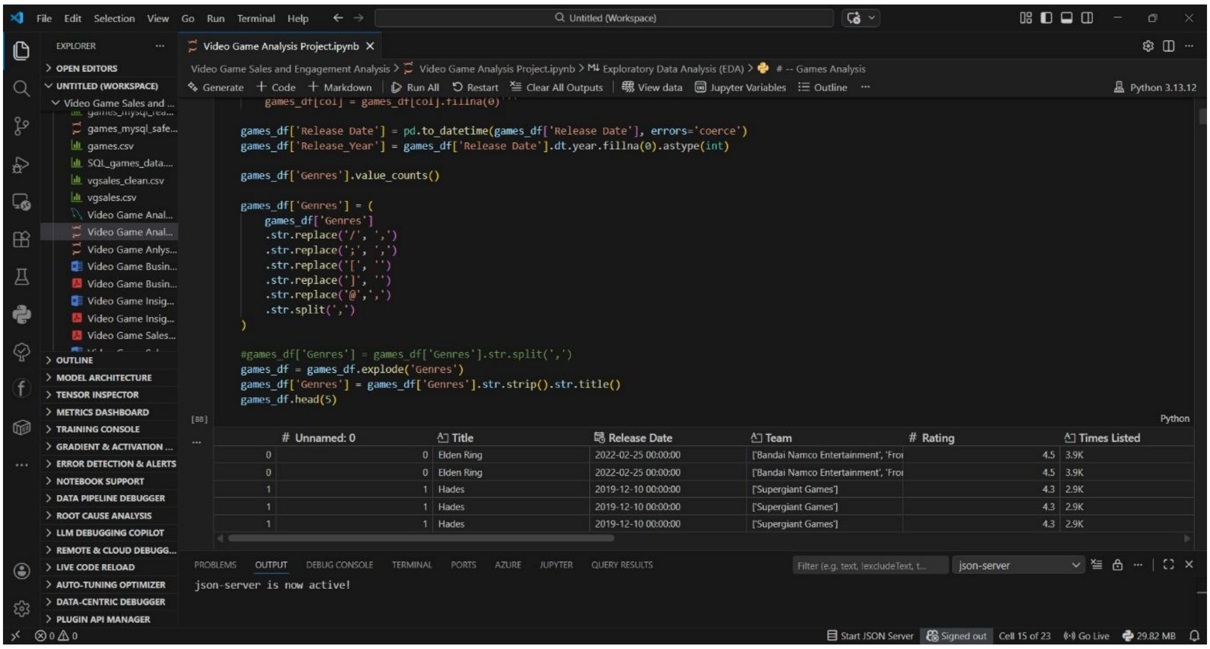
```python
import pandas as pd             # for handling data
import numpy as np              # for math operations
import matplotlib.pyplot as plt  # for charts
import seaborn as sns           # for prettier charts
```

```python
vgsales_df = pd.read_csv("E:/Analytics PDF/Video Game Sales and Engagement Analysis/vgsales.csv")
games_df = pd.read_csv("E:/Analytics PDF/Video Game Sales and Engagement Analysis/games.csv")

games_df.head(5)
```

	# Unnamed: 0	Title	Release Date	Team	# Rating	Times Listed
0	0	Elden Ring	Feb 25, 2022	['Bandai Namco Entertainment', 'Fro	4.5	3.9K
1	1	Hades	Dec 10, 2019	['Supergiant Games']	4.3	2.9K
2	2	The Legend of Zelda: Breath of the V	Mar 03, 2017	['Nintendo', 'Nintendo EPD Productic	4.4	4.3K
3	3	Undertale	Sep 15, 2015	['tobyfox', '8-4']	4.2	3.5K
4	4	Hollow Knight	Feb 24, 2017	['Team Cherry']	4.4	3K

```python
# Remove duplicates
```

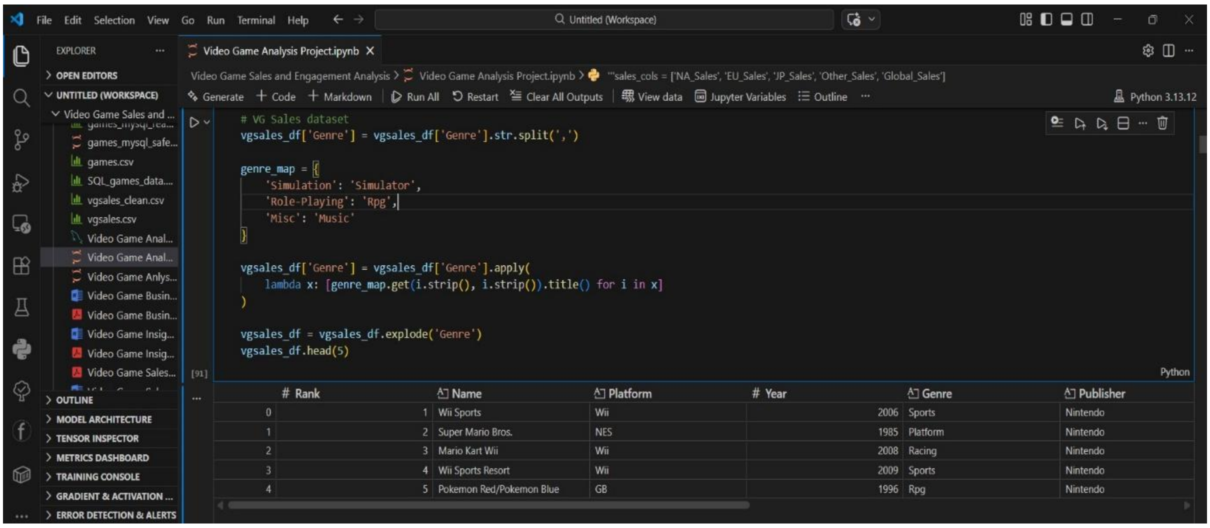
```
json-server is now active!
```



```
    games_df[col] = games_df[col].fillna(0)'''

games_df['Release Date'] = pd.to_datetime(games_df['Release Date'], errors='coerce')
games_df['Release_Year'] = games_df['Release Date'].dt.year.fillna(0).astype(int)

games_df['Genres'].value_counts()

games_df['Genres'] = (
    games_df['Genres']
    .str.replace('/', ',')
    .str.replace(';', ',')
    .str.replace('[', '')
    .str.replace(']', '')
    .str.replace('@',',')
    .str.split(',')
)

#games_df['Genres'] = games_df['Genres'].str.split(',')
games_df = games_df.explode('Genres')
games_df['Genres'] = games_df['Genres'].str.strip().str.title()
games_df.head(5)
```

	# Unnamed: 0	Title	Release Date	Team	# Rating	Times Listed
0	0	Elden Ring	2022-02-25 00:00:00	['Bandai Namco Entertainment', 'Fro	4.5	3.9K
0	0	Elden Ring	2022-02-25 00:00:00	['Bandai Namco Entertainment', 'Fro	4.5	3.9K
1	1	Hades	2019-12-10 00:00:00	['Supergiant Games']	4.3	2.9K
1	1	Hades	2019-12-10 00:00:00	['Supergiant Games']	4.3	2.9K
1	1	Hades	2019-12-10 00:00:00	['Supergiant Games']	4.3	2.9K

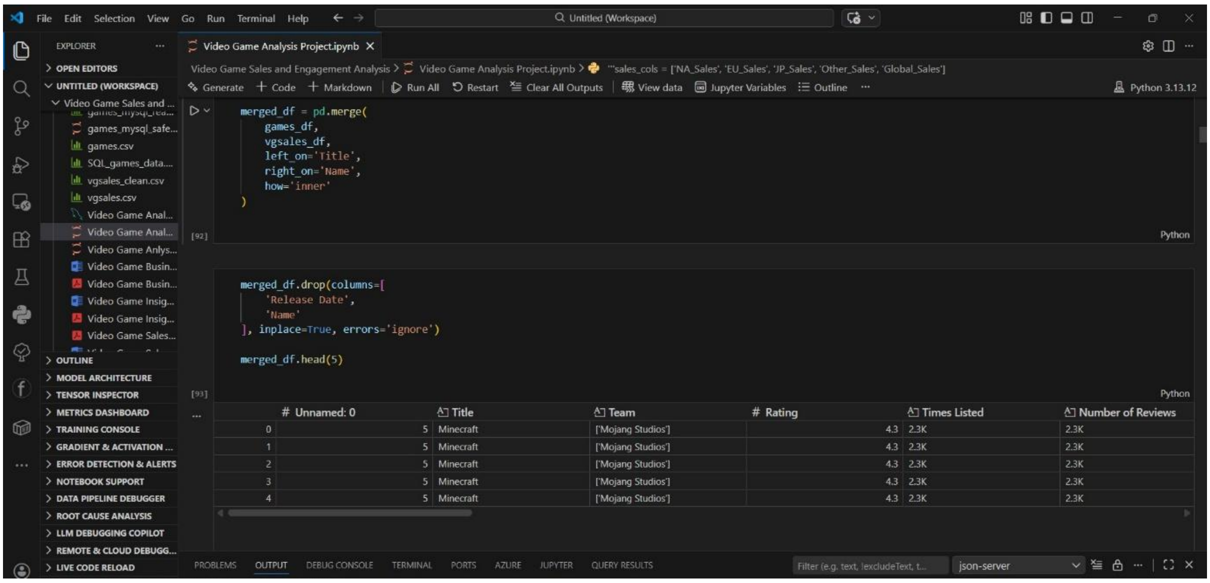
```
json-server is now active!
```



```
# VG Sales dataset
vgsales_df['Genre'] = vgsales_df['Genre'].str.split(',')

genre_map = {
    'Simulation': 'Simulator',
    'Role-Playing': 'Rpg',
    'Misc': 'Music'
}

vgsales_df['Genre'] = vgsales_df['Genre'].apply(
    lambda x: [genre_map.get(i.strip(), i.strip()).title() for i in x]
)

vgsales_df = vgsales_df.explode('Genre')
vgsales_df.head(5)
```

	# Rank	Name	Platform	# Year	Genre	Publisher
0	1	Wii Sports	Wii	2006	Sports	Nintendo
1	2	Super Mario Bros.	NES	1985	Platform	Nintendo
2	3	Mario Kart Wii	Wii	2008	Racing	Nintendo
3	4	Wii Sports Resort	Wii	2009	Sports	Nintendo
4	5	Pokemon Red/Pokemon Blue	GB	1996	Rpg	Nintendo

Video Game Sales and Engagement Analysis > Video Game Analysis Project.ipynb > '''sales_cols = ['NA_Sales', 'EU_Sales', 'JP_Sales', 'Other_Sales', 'Global_Sales']
[91]
Python

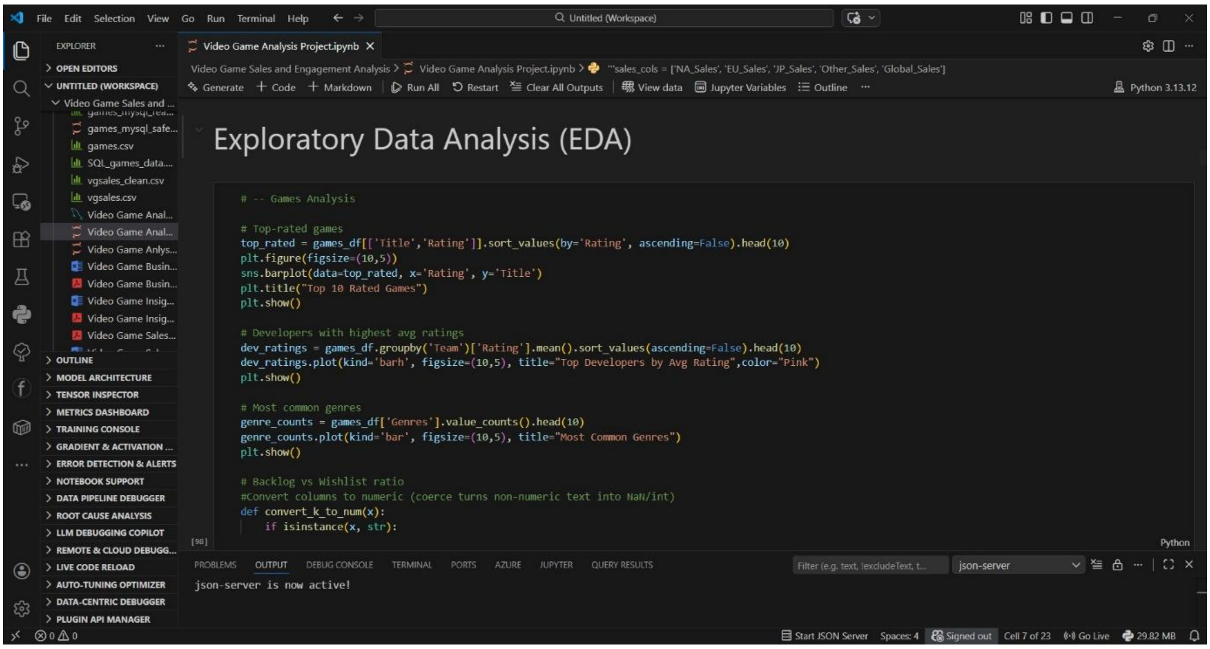


```
merged_df = pd.merge(
    games_df,
    vgsales_df,
    left_on='Title',
    right_on='Name',
    how='inner'
)
```

```
merged_df.drop(columns=[
    'Release Date',
    'Name'
], inplace=True, errors='ignore')

merged_df.head(5)
```

	# Unnamed: 0	Title	Team	# Rating	Times Listed	Number of Reviews
0	5	Minecraft	['Mojang Studios']	4.3	2.3K	2.3K
1	5	Minecraft	['Mojang Studios']	4.3	2.3K	2.3K
2	5	Minecraft	['Mojang Studios']	4.3	2.3K	2.3K
3	5	Minecraft	['Mojang Studios']	4.3	2.3K	2.3K
4	5	Minecraft	['Mojang Studios']	4.3	2.3K	2.3K

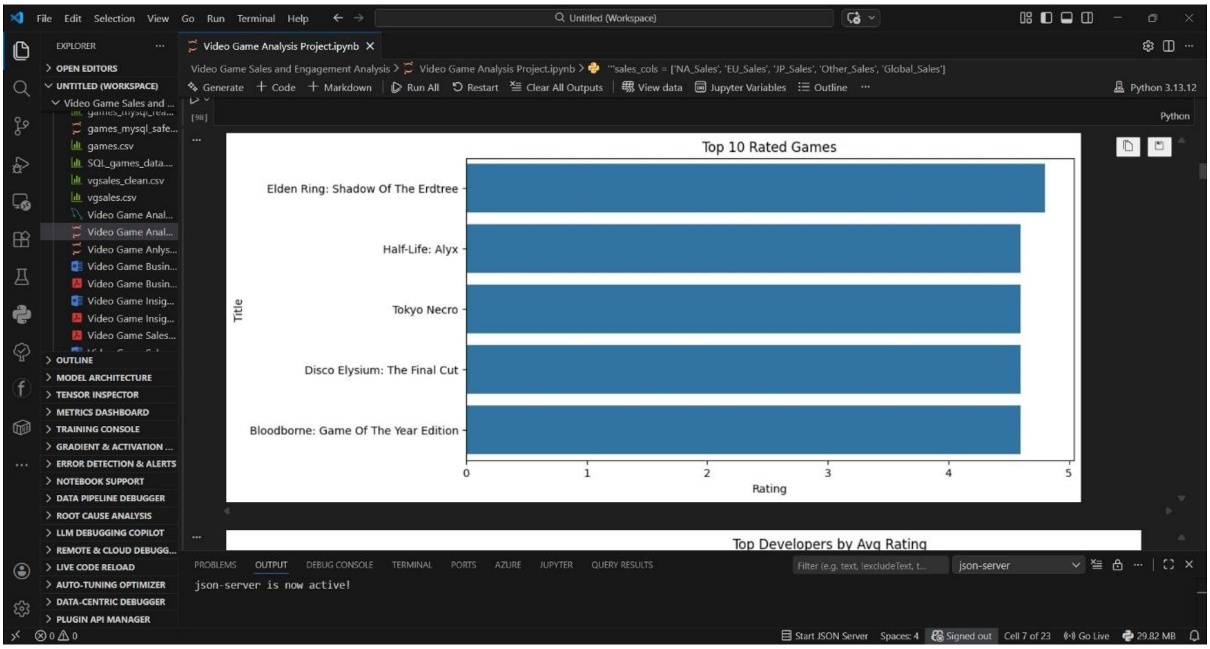
Video Game Sales and Engagement Analysis > Video Game Analysis Project.ipynb > '''sales_cols = ['NA_Sales', 'EU_Sales', 'JP_Sales', 'Other_Sales', 'Global_Sales']
[92]
[93]
Python



Exploratory Data Analysis (EDA)

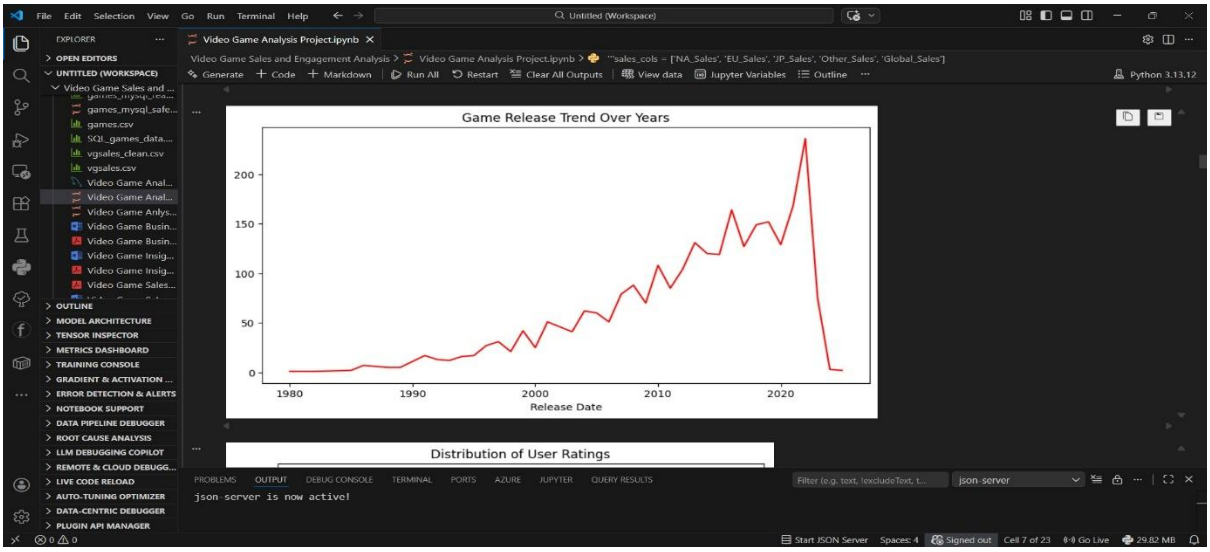
```python
# -- Games Analysis

# Top-rated games
top_rated = games_df[['Title','Rating']].sort_values(by='Rating', ascending=False).head(10)
plt.figure(figsize=(10,5))
sns.barplot(data=top_rated, x='Rating', y='Title')
plt.title("Top 10 Rated Games")
plt.show()

# Developers with highest avg ratings
dev_ratings = games_df.groupby('Team')['Rating'].mean().sort_values(ascending=False).head(10)
dev_ratings.plot(kind='barh', figsize=(10,5), title="Top Developers by Avg Rating",color="Pink")
plt.show()

# Most common genres
genre_counts = games_df['Genres'].value_counts().head(10)
genre_counts.plot(kind='bar', figsize=(10,5), title="Most Common Genres")
plt.show()

# Backlog vs Wishlist ratio
#Convert columns to numeric (coerce turns non-numeric text into NaN/int)
def convert_k_to_num(x):
    if isinstance(x, str):
```

```
json-server is now active!
```

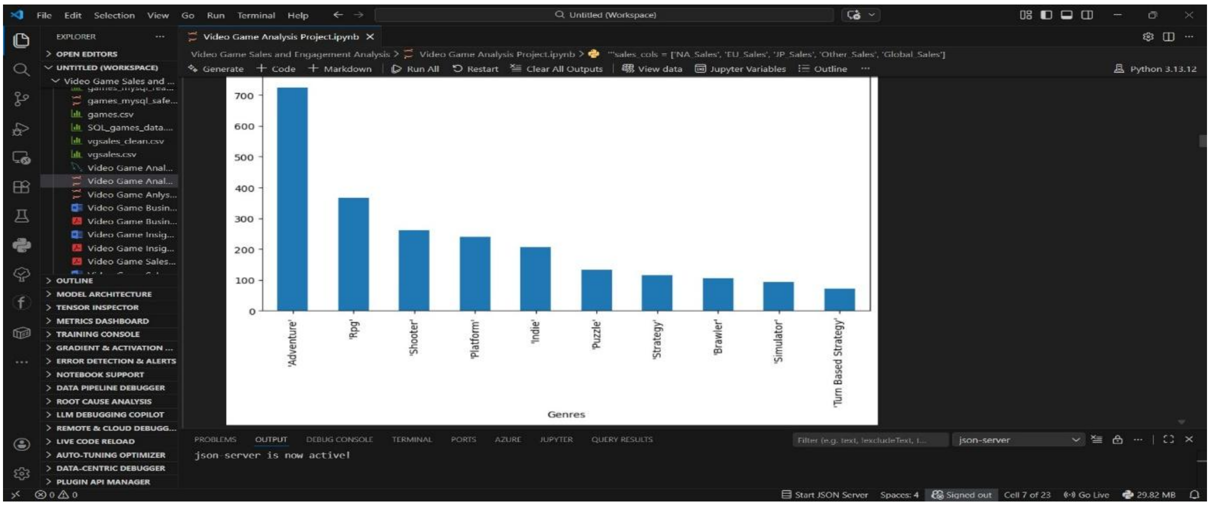


File Edit Selection View Go Run Terminal Help
Untitled (Workspace)
Video Game Analysis Project.ipynb
Video Game Sales and Engagement Analysis > Video Game Analysis Project.ipynb > ""sales_cols = ['NA_Sales', 'EU_Sales', 'JP_Sales', 'Other_Sales', 'Global_Sales']
Generate + Code + Markdown Run All Restart Clear All Outputs View data Jupyter Variables Outline
Python 3.13.12
[98]
Python
Top 10 Rated Games
Elden Ring: Shadow Of The Erdtree
Half-Life: Alyx
Tokyo Necro
Disco Elysium: The Final Cut
Bloodborne: Game Of The Year Edition
Title
0
1
2
3
4
5
Rating
Top Developers by Avg Rating
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS AZURE JUPYTER QUERY RESULTS
json-server
Start JSON Server Spaces: 4 Signed out Cell 7 of 23 Go Live 29.82 MB

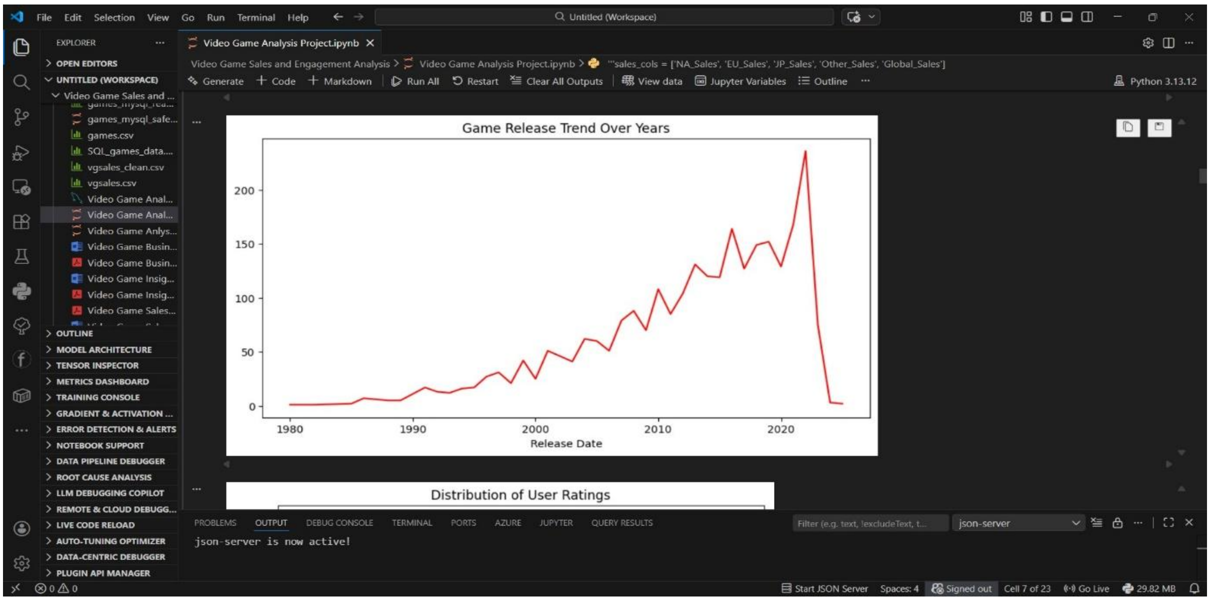
```
json-server is now active!
```



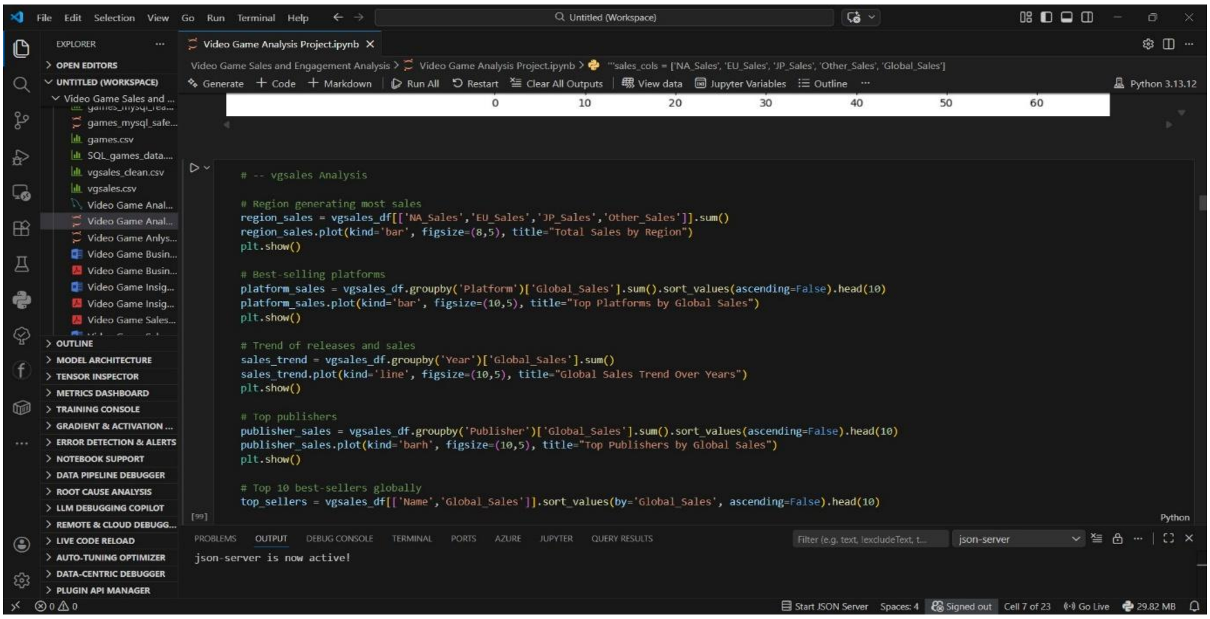
```
json-server is now active!
```



```
json-server is now active!
```

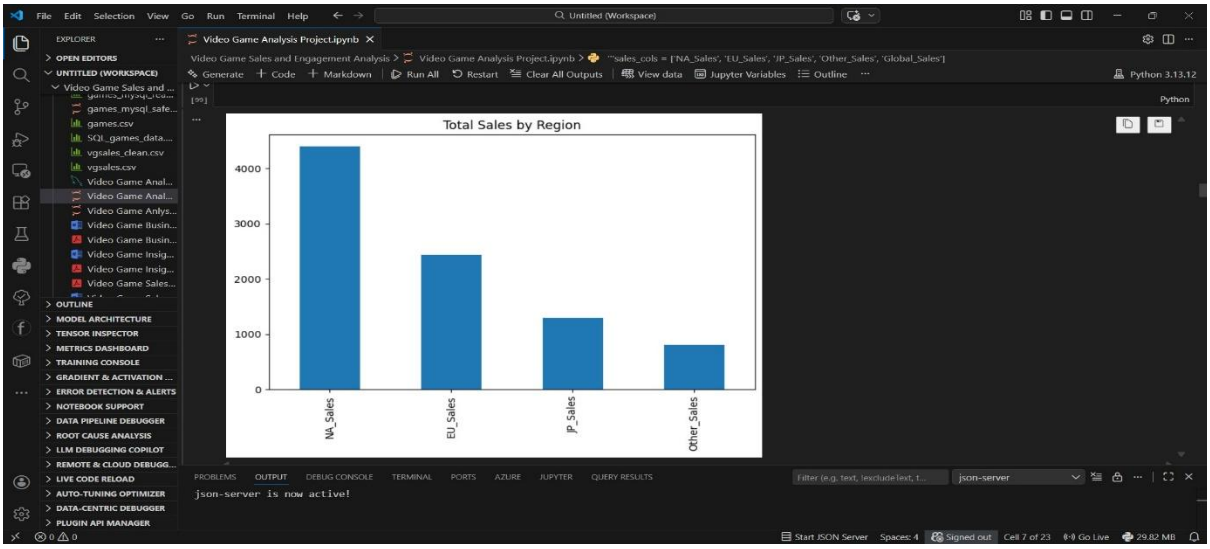


```
json-server is now active!
```



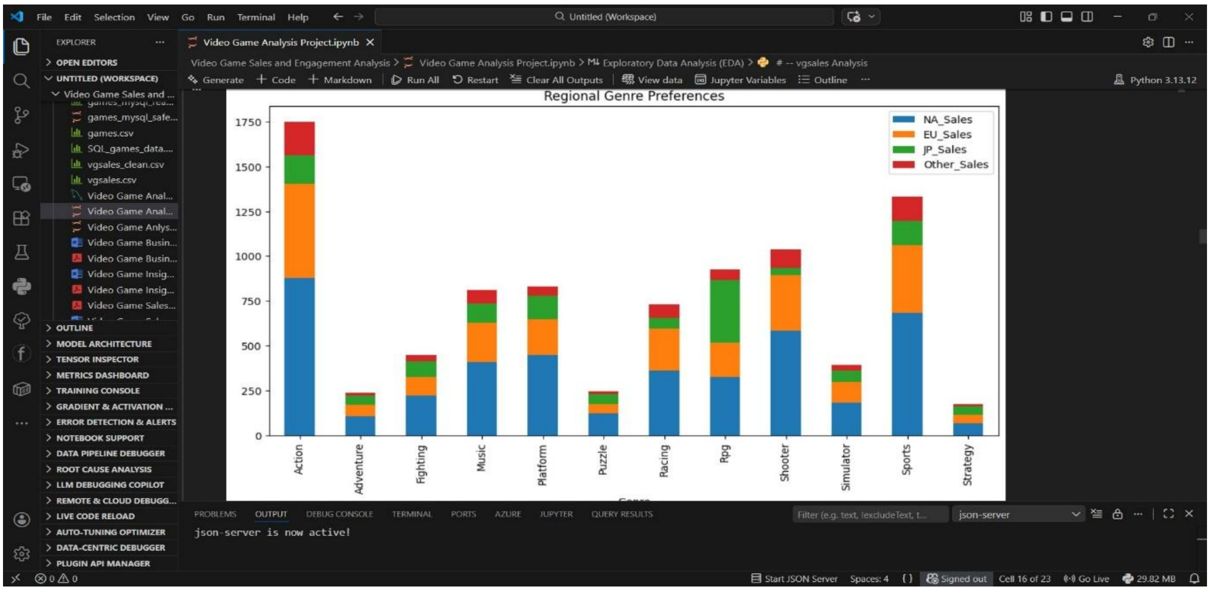
```python
# -- vgsales Analysis

# Region generating most sales
region_sales = vgsales_df[['NA_Sales','EU_Sales','JP_Sales','Other_Sales']].sum()
region_sales.plot(kind='bar', figsize=(8,5), title="Total Sales by Region")
plt.show()

# Best-selling platforms
platform_sales = vgsales_df.groupby('Platform')['Global_Sales'].sum().sort_values(ascending=False).head(10)
platform_sales.plot(kind='bar', figsize=(10,5), title="Top Platforms by Global Sales")
plt.show()

# Trend of releases and sales
sales_trend = vgsales_df.groupby('Year')['Global_Sales'].sum()
sales_trend.plot(kind='line', figsize=(10,5), title="Global Sales Trend Over Years")
plt.show()

# Top publishers
publisher_sales = vgsales_df.groupby('Publisher')['Global_Sales'].sum().sort_values(ascending=False).head(10)
publisher_sales.plot(kind='barh', figsize=(10,5), title="Top Publishers by Global Sales")
plt.show()

# Top 10 best-sellers globally
top_sellers = vgsales_df[['Name','Global_Sales']].sort_values(by='Global_Sales', ascending=False).head(10)
```

```
json-server is now active!
```



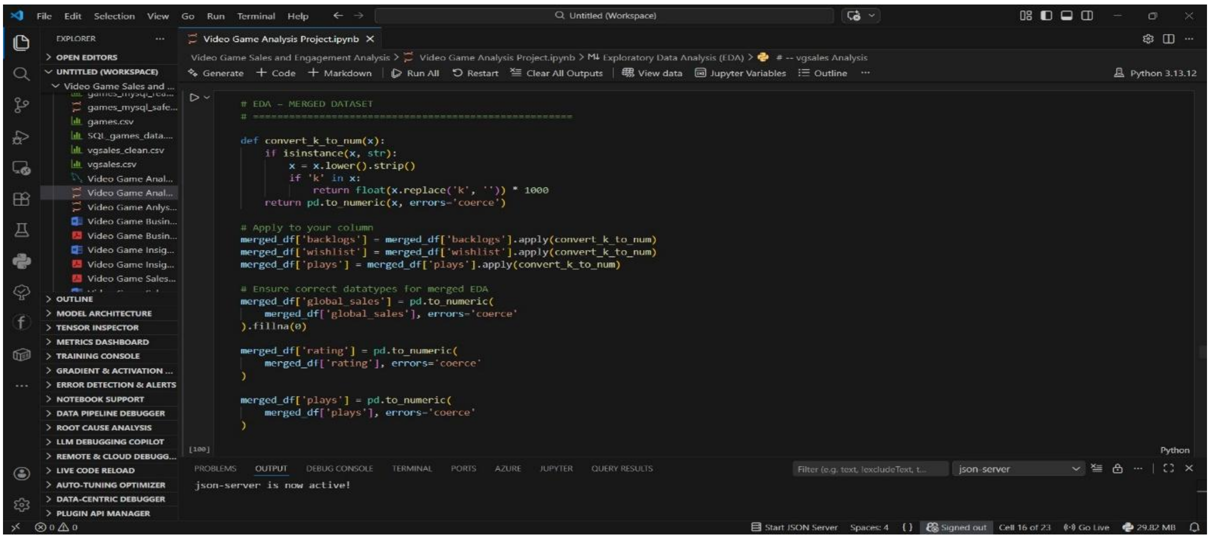
Video Game Analysis Project.ipynb
Video Game Sales and Engagement Analysis > Video Game Analysis Project.ipynb > '''sales_cols = ['NA_Sales', 'EU_Sales', 'JP_Sales', 'Other_Sales', 'Global_Sales']
Python 3.13.12
Total Sales by Region
4000
3000
2000
1000
0
NA_Sales
EU_Sales
JP_Sales
Other_Sales
Cell 7 of 23

```
json-server is now active!
```



Video Game Analysis Project.ipynb
Video Game Sales and Engagement Analysis > Video Game Analysis Project.ipynb > Exploratory Data Analysis (EDA) > # -- vgsales Analysis
Python 3.13.12
Regional Genre Preferences
NA_Sales
EU_Sales
JP_Sales
Other_Sales
1750
1500
1250
1000
750
500
250
0
Action
Adventure
Fighting
Music
Platform
Puzzle
Racing
Rpg
Shooter
Simulator
Sports
Strategy
Cell 16 of 23

```
json-server is now active!
```

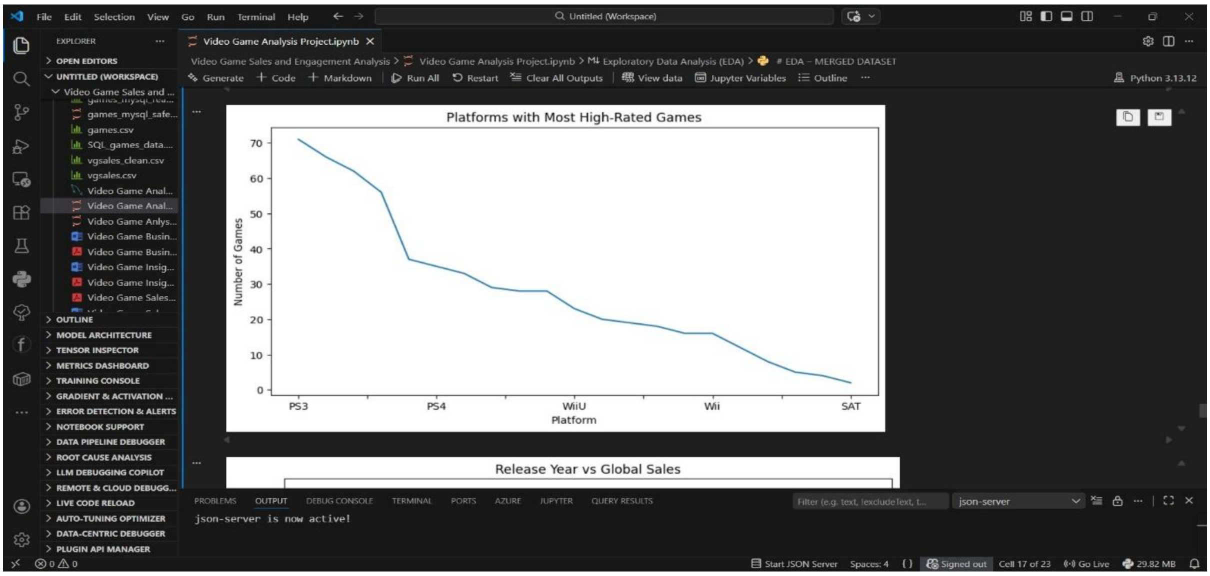


Video Game Analysis Project.ipynb
Video Game Sales and Engagement Analysis > Video Game Analysis Project.ipynb > Exploratory Data Analysis (EDA) > # -- vgsales Analysis
Generate
Code
Markdown
Run All
Restart
Clear All Outputs
View data
Jupyter Variables
Outline
Python 3.13.12

```python
# EDA – MERGED DATASET
# ==========================================================

def convert_k_to_num(x):
    if isinstance(x, str):
        x = x.lower().strip()
        if 'k' in x:
            return float(x.replace('k', '')) * 1000
    return pd.to_numeric(x, errors='coerce')

# Apply to your column
merged_df['backlogs'] = merged_df['backlogs'].apply(convert_k_to_num)
merged_df['wishlist'] = merged_df['wishlist'].apply(convert_k_to_num)
merged_df['plays'] = merged_df['plays'].apply(convert_k_to_num)

# Ensure correct datatypes for merged EDA
merged_df['global_sales'] = pd.to_numeric(
    merged_df['global_sales'], errors='coerce'
).fillna(0)

merged_df['rating'] = pd.to_numeric(
    merged_df['rating'], errors='coerce'
)

merged_df['plays'] = pd.to_numeric(
    merged_df['plays'], errors='coerce'
)
```

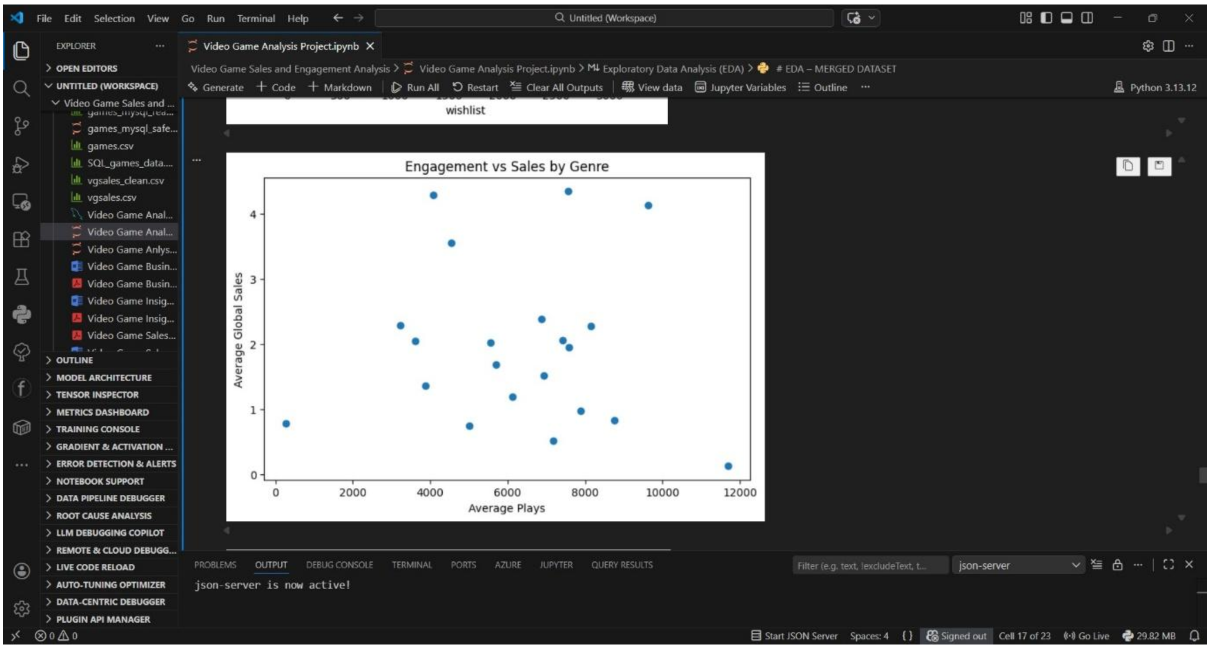
[100]

```
json-server is now active!
```



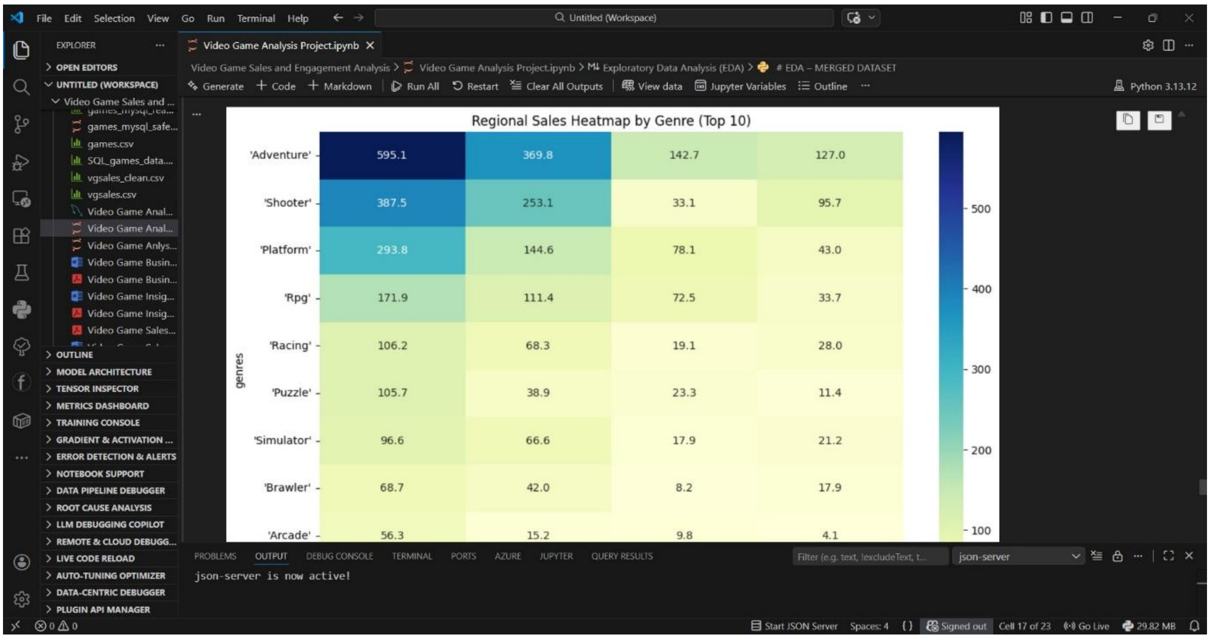
Video Game Analysis Project.ipynb
Video Game Sales and Engagement Analysis > Video Game Analysis Project.ipynb > Exploratory Data Analysis (EDA) > # EDA – MERGED DATASET
Platforms with Most High-Rated Games
Number of Games
70
60
50
40
30
20
10
0
PS3
PS4
WiiU
Wii
SAT
Platform
Release Year vs Global Sales

```
json-server is now active!
```



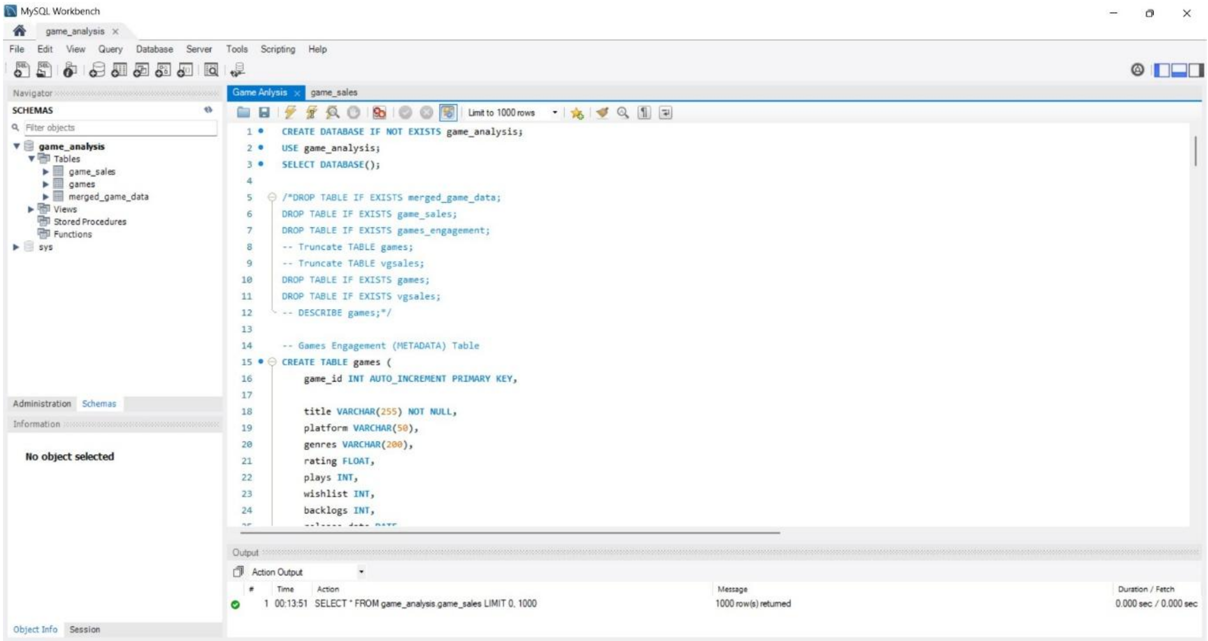
Video Game Analysis Project.ipynb
Video Game Sales and Engagement Analysis > Video Game Analysis Project.ipynb > Exploratory Data Analysis (EDA) > # EDA – MERGED DATASET
wishlist
Engagement vs Sales by Genre
Average Global Sales
4
3
2
1
0
0
2000
4000
6000
8000
10000
12000
Average Plays

```
json-server is now active!
```


Regional Sales Heatmap by Genre (Top 10)
genres
'Adventure' 595.1 369.8 142.7 127.0
'Shooter' 387.5 253.1 33.1 95.7
'Platform' 293.8 144.6 78.1 43.0
'Rpg' 171.9 111.4 72.5 33.7
'Racing' 106.2 68.3 19.1 28.0
'Puzzle' 105.7 38.9 23.3 11.4
'Simulator' 96.6 66.6 17.9 21.2
'Brawler' 68.7 42.0 8.2 17.9
'Arcade' 56.3 15.2 9.8 4.1
500
400
300
200
100
Video Game Analysis Project.ipynb
Video Game Sales and Engagement Analysis > Video Game Analysis Project.ipynb > Exploratory Data Analysis (EDA) > # EDA – MERGED DATASET
Python 3.13.12
Cell 17 of 23

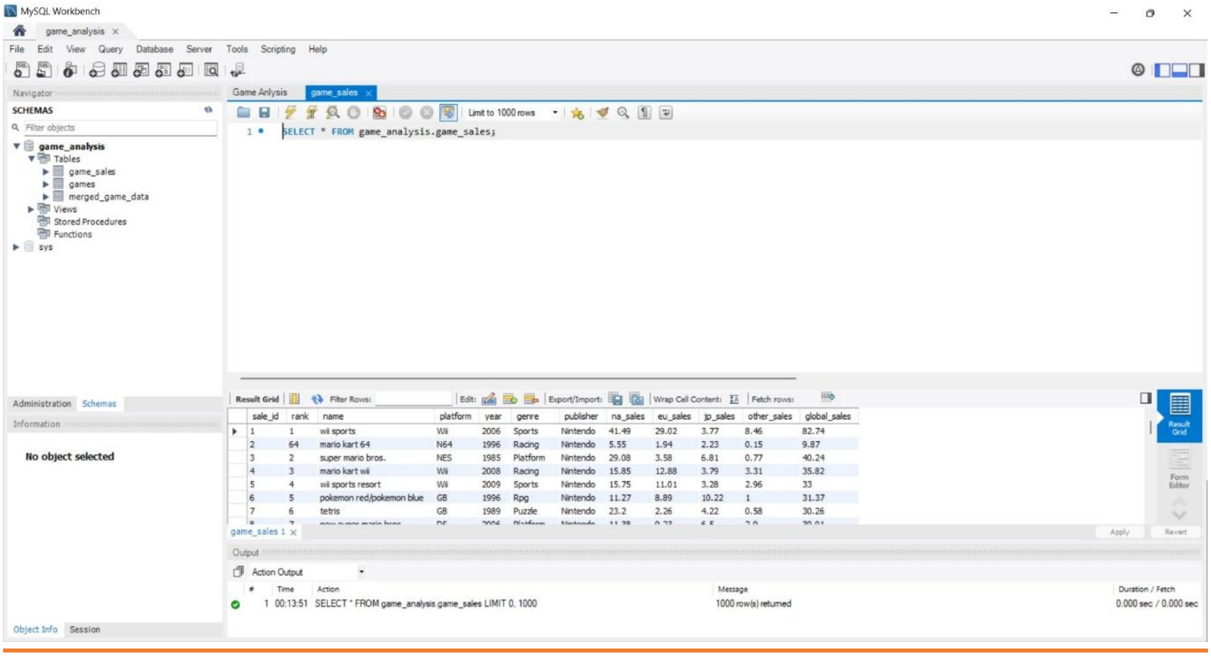
```
json-server is now active!
```

SQL CODE SCREENSHOTS



```sql
CREATE DATABASE IF NOT EXISTS game_analysis;
USE game_analysis;
SELECT DATABASE();

/*DROP TABLE IF EXISTS merged_game_data;
DROP TABLE IF EXISTS game_sales;
DROP TABLE IF EXISTS games_engagement;
-- Truncate TABLE games;
-- Truncate TABLE vgsales;
DROP TABLE IF EXISTS games;
DROP TABLE IF EXISTS vgsales;
-- DESCRIBE games;*/

-- Games Engagement (METADATA) Table
CREATE TABLE games (
    game_id INT AUTO_INCREMENT PRIMARY KEY,

    title VARCHAR(255) NOT NULL,
    platform VARCHAR(50),
    genres VARCHAR(200),
    rating FLOAT,
    plays INT,
    wishlist INT,
    backlogs INT,
```

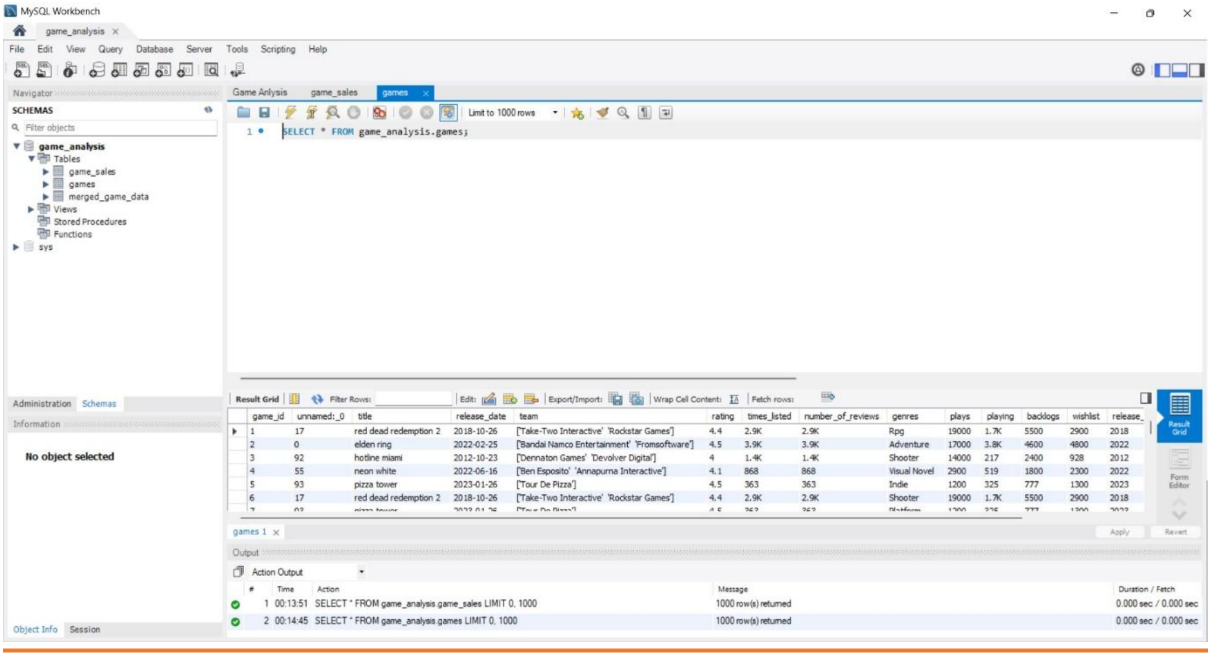
```
#  Time      Action                                           Message                Duration / Fetch
1  00:13:51  SELECT * FROM game_analysis.game_sales LIMIT 0, 1000   1000 row(s) returned   0.000 sec / 0.000 sec
```



```
SELECT * FROM game_analysis.game_sales;
```

sale_id	rank	name	platform	year	genre	publisher	na_sales	eu_sales	jp_sales	other_sales	global_sales
1	1	wii sports	Wii	2006	Sports	Nintendo	41.49	29.02	3.77	8.46	82.74
2	64	mario kart 64	N64	1996	Racing	Nintendo	5.55	1.94	2.23	0.15	9.87
3	2	super mario bros.	NES	1985	Platform	Nintendo	29.08	3.58	6.81	0.77	40.24
4	3	mario kart wii	Wii	2008	Racing	Nintendo	15.85	12.88	3.79	3.31	35.82
5	4	wii sports resort	Wii	2009	Sports	Nintendo	15.75	11.01	3.28	2.96	33
6	5	pokemon red/pokemon blue	GB	1996	Rpg	Nintendo	11.27	8.89	10.22	1	31.37
7	6	tetris	GB	1989	Puzzle	Nintendo	23.2	2.26	4.22	0.58	30.26

```
#   Time      Action                                              Message                Duration / Fetch
1   00:13:51  SELECT * FROM game_analysis.game_sales LIMIT 0, 1000  1000 row(s) returned   0.000 sec / 0.000 sec
```



```
SELECT * FROM game_analysis.games;
```

game_id	unnamed:_0	title	release_date	team	rating	times_listed	number_of_reviews	genres	plays	playing	backlogs	wishlist	release_
1	17	red dead redemption 2	2018-10-26	['Take-Two Interactive' 'Rockstar Games']	4.4	2.9K	2.9K	Rpg	19000	1.7K	5500	2900	2018
2	0	elden ring	2022-02-25	['Bandai Namco Entertainment' 'Fromsoftware']	4.5	3.9K	3.9K	Adventure	17000	3.8K	4600	4800	2022
3	92	hotline miami	2012-10-23	['Dennaton Games' 'Devolver Digital']	4	1.4K	1.4K	Shooter	14000	217	2400	928	2012
4	55	neon white	2022-06-16	['Ben Esposito' 'Annapurna Interactive']	4.1	868	868	Visual Novel	2900	519	1800	2300	2022
5	93	pizza tower	2023-01-26	['Tour De Pizza']	4.5	363	363	Indie	1200	325	777	1300	2023
6	17	red dead redemption 2	2018-10-26	['Take-Two Interactive' 'Rockstar Games']	4.4	2.9K	2.9K	Shooter	19000	1.7K	5500	2900	2018

```
#   Time      Action                                              Message                Duration / Fetch
1   00:13:51  SELECT * FROM game_analysis.game_sales LIMIT 0, 1000  1000 row(s) returned   0.000 sec / 0.000 sec
2   00:14:45  SELECT * FROM game_analysis.games LIMIT 0, 1000       1000 row(s) returned   0.000 sec / 0.000 sec
```

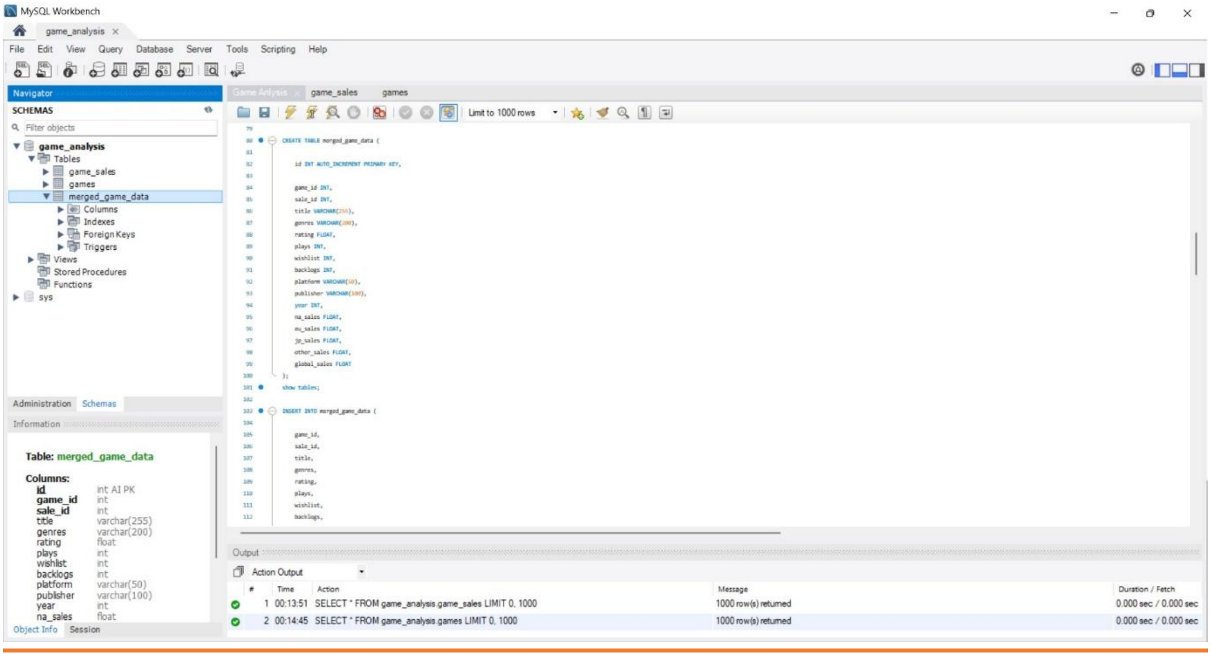


Table: merged_game_data

Columns:

id	int AI PK
game_id	int
sale_id	int
title	varchar(255)
genres	varchar(200)
rating	float
plays	int
wishlist	int
backlogs	int
platform	varchar(50)
publisher	varchar(100)
year	int
na_sales	float

```sql
CREATE TABLE merged_game_data (

    id INT AUTO_INCREMENT PRIMARY KEY,

    game_id INT,
    sale_id INT,
    title VARCHAR(255),
    genres VARCHAR(200),
    rating FLOAT,
    plays INT,
    wishlist INT,
    backlogs INT,
    platform VARCHAR(50),
    publisher VARCHAR(100),
    year INT,
    na_sales FLOAT,
    eu_sales FLOAT,
    jp_sales FLOAT,
    other_sales FLOAT,
    global_sales FLOAT
);
show tables;

INSERT INTO merged_game_data (

    game_id,
    sale_id,
    title,
    genres,
    rating,
    plays,
    wishlist,
    backlogs,
```

Action Output

#	Time	Action	Message	Duration / Fetch
1	00:13:51	SELECT * FROM game_analysis.game_sales LIMIT 0, 1000	1000 row(s) returned	0.000 sec / 0.000 sec
2	00:14:45	SELECT * FROM game_analysis.games LIMIT 0, 1000	1000 row(s) returned	0.000 sec / 0.000 sec

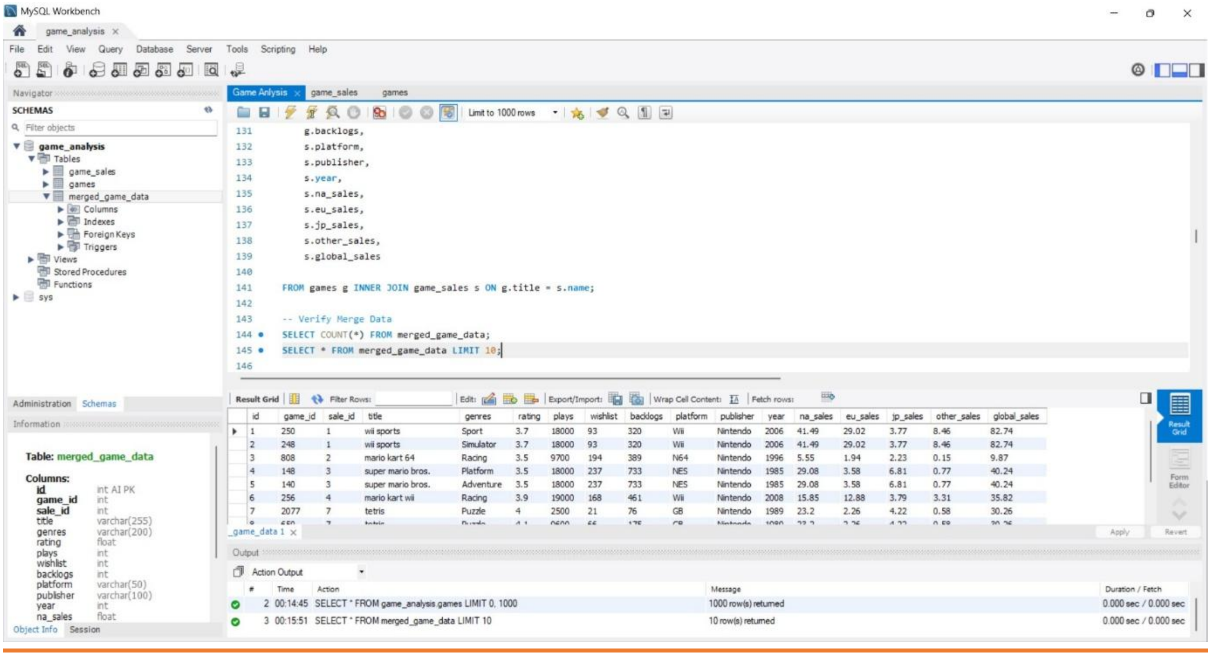


Table: merged_game_data

Columns:

id	int AI PK
game_id	int
sale_id	int
title	varchar(255)
genres	varchar(200)
rating	float
plays	int
wishlist	int
backlogs	int
platform	varchar(50)
publisher	varchar(100)
year	int
na_sales	float

```sql
        g.backlogs,
        s.platform,
        s.publisher,
        s.year,
        s.na_sales,
        s.eu_sales,
        s.jp_sales,
        s.other_sales,
        s.global_sales

    FROM games g INNER JOIN game_sales s ON g.title = s.name;

    -- Verify Merge Data
    SELECT COUNT(*) FROM merged_game_data;
    SELECT * FROM merged_game_data LIMIT 10;
```

id	game_id	sale_id	title	genres	rating	plays	wishlist	backlogs	platform	publisher	year	na_sales	eu_sales	jp_sales	other_sales	global_sales
1	250	1	wii sports	Sport	3.7	18000	93	320	Wii	Nintendo	2006	41.49	29.02	3.77	8.46	82.74
2	248	1	wii sports	Simulator	3.7	18000	93	320	Wii	Nintendo	2006	41.49	29.02	3.77	8.46	82.74
3	808	2	mario kart 64	Racing	3.5	9700	194	389	N64	Nintendo	1996	5.55	1.94	2.23	0.15	9.87
4	148	3	super mario bros.	Platform	3.5	18000	237	733	NES	Nintendo	1985	29.08	3.58	6.81	0.77	40.24
5	140	3	super mario bros.	Adventure	3.5	18000	237	733	NES	Nintendo	1985	29.08	3.58	6.81	0.77	40.24
6	256	4	mario kart wii	Racing	3.9	19000	168	461	Wii	Nintendo	2008	15.85	12.88	3.79	3.31	35.82
7	2077	7	tetris	Puzzle	4	2500	21	76	GB	Nintendo	1989	23.2	2.26	4.22	0.58	30.26

Action Output

#	Time	Action	Message	Duration / Fetch
2	00:14:45	SELECT * FROM game_analysis.games LIMIT 0, 1000	1000 row(s) returned	0.000 sec / 0.000 sec
3	00:15:51	SELECT * FROM merged_game_data LIMIT 10	10 row(s) returned	0.000 sec / 0.000 sec