Анализ продаж Видеоигр PS4

Артем Резников • 25.10.2023





Информация о датасете:

Video Games Sales Dataset

Задача анализа

 Факторы, влияющие на региональные объемы продаж видеоигр для платформы PS4

Вопросы

- ТОП -10 объемы продаж?
- Какие самые продаваемые жанры?
- Влияние возрастного рейтинга?
- Насколько важно происхождение компании-разработчика?

Атрибуты

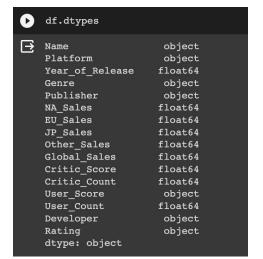
Name - название игры Platform - платформа выпуска Year of Release - год выпуска Genre - жанр Publisher - издатель NA Sales - продажи в Северной Америке EU Sales - продажи в Европе JP Sales - продажи в Японии Global Sales - глобальные продажи Other Sales - остальные продажи Critic_Score - оценки критиков Critic Count - количество критиков User Score - оценки игроков User Count - количество игроков Developer - создатель игры Rating - возрастной рейтинг

Загрузка в Jupiter Notebook



Проверка дубликатов

Проверка типов данных



Вносим изменения в Год выпуска, для правильного отображения даты

Прежде, чем переводить float к int, удаляем NaN значения

```
# Заменяем пропущенные значения в столбце Year_of_Release на ноль

df['Year_of_Release'] = df['Year_of_Release'].fillna(0).astype(int)

# Затем приводим float к int для правильного отображения даты (год)

df['Year_of_Release'] = df['Year_of_Release'].astype(int)
```

Проверка количества значений в колонках для дальнейшей обработки

0	df.count()	
→	Name Platform Year_of_Release Genre Publisher NA_Sales EU_Sales JP_Sales Other_Sales Global_Sales Critic_Score Critic_Count User_Score User_Count Developer Rating dtype: int64	16717 16719 16450 16717 16665 16719 16719 16719 16719 8137 8137 10015 7590 10096 9950

Year_of_Release	User_Score	Critic_Score
2006.0	8	76.0
1985.0	NaN	NaN
2008.0	8.3	82.0
2009.0	8	80.0
1996.0	NaN	NaN

Конвертация значений к удобному формата, удаление бесполезных кортежей

```
] # Меняем str значения на NaN
    # Удаляем строчки, при условии, что у нас нет информации об оценках от критиков и юзеров.
    df['User Score'] = df['User Score'].replace('tbd', np.nan)
    df = df.dropna(subset=['Critic Score', 'User Score'], how='all')
[ ] # Конвертируем Оценки юзеров в 100-бальную систему оценивания.
    df['User Score'] = df['User Score'].astype(float).round(1)
    df['User Score'] = df['User Score'] * 10
   # Конвертируем значения в евро
   columns to convert = ['NA Sales', 'EU Sales', 'JP Sales', 'Other Sales', 'Global Sales']
    for column in columns to convert:
        df[column] = df[column].apply(lambda x: round(float(x) / 1.06, 2))
```

Результат

Мы провели конвертацию валют по текущему курсу в евро (предыдущие данные были в USD).

Год выпуска без плавающей точки.

Оценки критиков и юзеров в одном формате.

ОЦЕНКИ КОИТИК Year_of_Release		Publisher			JP_Sales
2006	Sports	Nintendo	39.02	27.32	3.56
2008	Racing	Nintendo	14.79	12.04	3.58
2009	Sports	Nintendo	14.73	10.31	3.09
2006	Platform	Nintendo	10.64	8.62	6.13
2006	Misc	Nintendo	13.17	8.66	2.76

Critic_Score	Critic_Count	User_Score
76.0	51.0	80.0
82.0	73.0	83.0
80.0	73.0	80.0
89.0	65.0	85.0
58.0	41.0	66.0

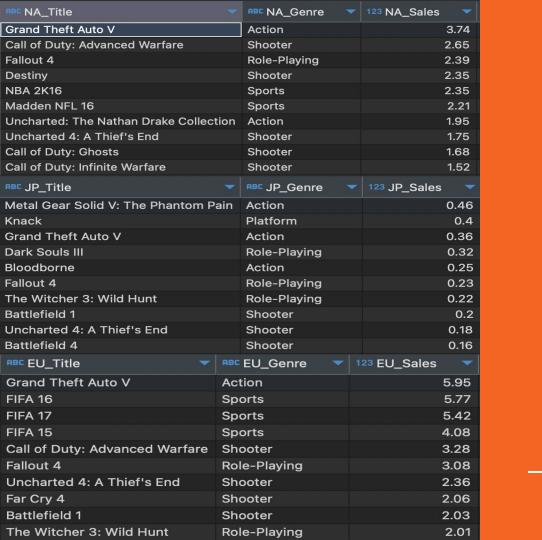
Загрузка в обновленной таблицы базу данных SQL

Региональные продажи ТОП 10

WITH RegionalRanks AS (RBC NA_Title	▼ RBC NA_Genre	▼ 123 NA_Sales ▼
SELECT	Grand Theft Auto V	Action	3.74
Name,	Call of Duty: Advanced Warfare	Shooter	2.65
	Fallout 4	Role-Playing	2.39
NA_Sales,	Destiny	Shooter	2.35
Genre,	NBA 2K16	Sports	2.35
<pre>ROW_NUMBER() OVER (ORDER BY NA_Sales DESC) AS NA_Rank,</pre>	Madden NFL 16	Sports	2.21
EU Sales,	Uncharted: The Nathan Drake Collection		1.95
<pre>ROW_NUMBER() OVER (ORDER BY EU_Sales DESC) AS EU_Rank,</pre>	Uncharted 4: A Thief's End	Shooter	1.75
	Call of Duty: Ghosts	Shooter	1.68
JP_Sales,	Call of Duty: Infinite Warfare	Shooter	1.52
ROW_NUMBER() OVER (ORDER BY JP_Sales DESC) AS JP_Rank	RBC JP_Title	▼ PBC JP_Genre	▼ 123 JP_Sales ▼
FROM "Video_Games_Sales.db" vgsd	Metal Gear Solid V: The Phantom Pai	n Action	0.46
WHERE Platform IN ('PS4')	Knack	Platform	0.4
)	Grand Theft Auto V	Action	0.36
SELECT	Dark Souls III	Role-Playing	0.32
NA.Name AS NA_Title,	Bloodborne	Action	0.25
	Fallout 4	Role-Playing	0.23
NA.Genre AS NA_Genre,	The Witcher 3: Wild Hunt	Role-Playing	0.22
NA.NA_Sales AS NA_Sales,	Battlefield 1	Shooter	0.2
EU.Name AS EU_Title,	Uncharted 4: A Thief's End	Shooter	0.18
EU.Genre AS EU_Genre,	Battlefield 4	Shooter	0.16
EU.EU_Sales AS EU_Sales,	ABC EU_Title	EU_Genre	123 EU_Sales ▼
JP.Name AS JP_Title,	Grand Theft Auto V	Action	5.95
JP.Genre AS JP Genre,	FIFA 16	Sports	5.77
	FIFA 17	Sports	5.42
JP.JP_Sales AS JP_Sales	FIFA 15	Sports	4.08
FROM RegionalRanks NA	Call of Duty: Advanced Warfare	Shooter	3.28
JOIN RegionalRanks EU ON NA.NA_Rank = EU.EU_Rank	Fallout 4	Role-Playing	3.08
JOIN RegionalRanks JP ON NA.NA_Rank = JP.JP_Rank	Uncharted 4: A Thief's End	Shooter	2.36
ORDER BY NA.NA_Rank	Far Cry 4	Shooter	2.06
		Shooter	2.03
LIMIT 10;		Role-Playing	2.01
	The transfer of this right	toro i laying	2.01

Региональные продажи ТОП 10

```
WITH RegionalRanks AS (
  SELECT
   Name,
   NA_Sales,
   Genre,
   ROW NUMBER() OVER (ORDER BY NA Sales DESC) AS NA Rank,
   EU Sales,
   ROW NUMBER() OVER (ORDER BY EU Sales DESC) AS EU Rank,
   JP Sales,
   ROW NUMBER() OVER (ORDER BY JP Sales DESC) AS JP Rank
  FROM "Video Games Sales.db" vgsd
  WHERE Platform IN ('PS4')
SELECT
 NA.Name AS NA Title,
 NA. Genre AS NA Genre,
 NA.NA Sales AS NA Sales,
  EU.Name AS EU Title,
  EU. Genre AS EU Genre,
  EU.EU_Sales AS EU_Sales,
 JP.Name AS JP Title,
 JP.Genre AS JP Genre,
 JP.JP_Sales AS JP_Sales
FROM RegionalRanks NA
JOIN RegionalRanks EU ON NA.NA Rank = EU.EU Rank
JOIN RegionalRanks JP ON NA.NA Rank = JP.JP Rank
ORDER BY NA.NA Rank
LIMIT 10:
```



Продажи ТОП 10 по регионам

- Bo всех регионах предпочтения
 по Action и Shooter похожи
- (раскрученные бренды, ААА игры)
 Следует обратить внимание, что
- спортивные игры отличаются в топе:

NA - регби и баскетбол

EU - футбол

Самые продаваемые жанры

```
SELECT
    Genre,
    SUM(NA_Sales) AS Total_NA_Sales,
    SUM(EU_Sales) AS Total_EU_Sales,
    SUM(JP_Sales) AS Total_JP_Sales
FROM "Video_Games_Sales.db" vgsd
WHERE Platform = 'PS4'
GROUP BY Genre
ORDER BY SUM(NA_Sales) DESC;
```

RBC Genre	1 ² Total_NA_Sales ▼	126 Total_EU_Sales ▼	1 ² Total_JP_Sales ▼
Action	25.01	33.79	3.28
Shooter	21.69	27.35	2.1
Sports	18.64	24.41	0.78
Role-Playing	5.73	7.55	1.52
Fighting	2.69	2.25	0.52
Racing	1.95	5.29	0.2
Platform	1.85	2.68	0.43
Misc	1.2	1.08	0.04
Adventure	1	1.53	0.14
Strategy	0.16	0.21	0.02
Simulation	0.11	0.44	0.06
Puzzle	0	0.02	0

Самые продаваемые жанры

Самые популярные жанры это Action, Shooter и Sports.
В регионах NA, EU они зарабатывают больше всего денег.

В Японии, на третьем месте - RPG.

Влияние оценок критиков и юзеров на глобальные продажи

```
SELECT
    Genre,
    Global_Sales,
    ROUND(AVG(Critic_Score)) AS Average_Critic_Score,
    ROUND(AVG(User_Score)) AS Average_User_Score
FROM "Video_Games_Sales.db" vgsd
WHERE Platform = 'PS4'
GROUP BY Genre
ORDER BY Global_Sales DESC;
```

RBC Genre	123 Global_Sales 🔻	126 Average_Critic_Score V	12 Average_User_Score ▼
Action	11.9	72	69
Sports	8.08	68	58
Shooter	7.23	75	67
Role-Playing	6.75	73	75
Fighting	2.62	73	72
Racing	1.95	72	67
Platform	1.8	72	60
Adventure	1.57	73	77
Misc	0.81	71	63
Simulation	0.39	73	79
Strategy	0.2	70	63
Puzzle	0.03	88	84

Влияние оценок критиков и юзеров на глобальные продажи

 Чаще всего оценки критиков выше, чем у остальных игроков.
 Так как оценки критиков появляются раньше, нежели других.
 Это может напрямую влиять на рост продаж.

Какие жанры самые продаваемые по годам

```
SELECT
    Year of Release,
    Genre,
    SUM(NA Sales) AS Total NA Sales,
    SUM(EU Sales) AS Total EU Sales,
    SUM(JP Sales) AS Total JP Sales,
    Global Sales
FROM "Video_Games_Sales.db" vgsd
WHERE Platform = 'PS4'
GROUP BY Year_of_Release
ORDER BY Year of Release;
```

2,013

2,014

2,015

Shooter

Action

Sports

2,016 Sports

Какие жанры самые продаваемые по годам

0.9

1.96

3.76

2.47

3.61

11.9

8.08

7.16

Виден рост интереса к

10.15

36.12

33.96

26.37

спортивным играм.

9.13

26.23

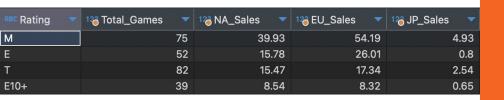
27.21

17.46

123 Global_Sales 23 Year_of_Release **RBC** Genre 126 Total_NA_Sales 126 Total_EU_Sales 12 Total_JP_Sales

Влияние возрастного рейтинга

```
Rating,
COUNT(Rating) AS Total_Games,
ROUND(SUM(NA_Sales), 2) AS NA_Sales,
ROUND(SUM(EU_Sales), 2) AS EU_Sales,
ROUND(SUM(JP_Sales), 2) AS JP_Sales
FROM "Video_Games_Sales.db" vgsd
WHERE Platform = 'PS4'
GROUP BY Rating
ORDER BY NA_Sales DESC;
```



Влияние возрастного рейтинга

Лучше всего продаются игры с
Взрослым возрастным рейтингом.
Из-за того, что во многих странах
регуляция продажи игр
отсутствует - подростки в т.ч
активно покупают игры М.

Выводы

- Рост интереса в спортивным играм.
- При создании спортивного проекта следует учитывать интерес к видам спорта у региона основной аудитории.
- Критики обладают тенденцией к завышению оценок и управлением общественным мнением.
- М рейтинг не отпугивает, а наоборот привлекает игроков.
- Наибольших продаж достигают ААА-игры.
- Такие регионы, как NA и EU имеют схожие предпочтения, тогда как у JP в топ попадают отличающиеся жанры и игры.
 Это говорит о том, что стоит рассматривать Восточный

рынок отдельно от Европейского и Американского.