

# Анализ рынка игровой индустрии

## Введение

Хлеба и зрелищ хотели древние римляне, нравы с тех пор изменились мало, хлеб все тот же, а вот гладиаторские бои уже не в моде. Поистине компьютерные игры это композиция наиболее любимых человеком искусств. Искусство во все времена оказывало огромное влияние на человека во многих аспектах, но игры смогли подтолкнуть нас к чему то большему.

Индустрия компьютерных игр уже сейчас является важной частью индустрии развлечений, и ее значение продолжает расти. Все больше людей в мире отдают предпочтение вечерам проведенным за любимой игрой, а не перед экраном телевизоров. Телевизор не позволяет почувствовать себя героем любимого романа или гладиатором на песке. Но самое важное на мой взгляд это то, что игровая индустрия дала толчок развитию вычислительной мощности компьютеров, взять например историю компании Nvidia, а это уже оказало прямое влияние на весь мир.

Очевидно за играми будущее, а также за теми кто их продает. Для планирования рекламной компании по продвижению различных игр в интернет-магазине "Стримчик", нужно провести анализ популярности различных жанров и платформ в доступных регионах, чтобы определить потенциальные целевые аудитории и наиболее эффективные способы продвижения. Цель данной работы понять, какие факторы влияют на успешность игры, какие жанры любят пользователи разных частей света и как продать больше игр.

## Ознакомление с данными

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats as st

data = pd.read_csv('/datasets/games.csv')
data.head()
```

	Name	Platform	Year_of_Release	Genre
NA_sales \				
0	Wii Sports	Wii	2006.0	Sports
41.36				
1	Super Mario Bros.	NES	1985.0	Platform
29.08				
2	Mario Kart Wii	Wii	2008.0	Racing
15.68				
3	Wii Sports Resort	Wii	2009.0	Sports
15.61				
4	Pokemon Red/Pokemon Blue	GB	1996.0	Role-Playing
11.27				

	EU_sales	JP_sales	Other_sales	Critic_Score	User_Score	Rating
0	28.96	3.77	8.45	76.0	8	E
1	3.58	6.81	0.77	NaN	NaN	NaN
2	12.76	3.79	3.29	82.0	8.3	E
3	10.93	3.28	2.95	80.0	8	E
4	8.89	10.22	1.00	NaN	NaN	NaN

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 16715 entries, 0 to 16714
```

```
Data columns (total 11 columns):
```

#	Column	Non-Null Count	Dtype
0	Name	16713 non-null	object
1	Platform	16715 non-null	object
2	Year_of_Release	16446 non-null	float64
3	Genre	16713 non-null	object
4	NA_sales	16715 non-null	float64
5	EU_sales	16715 non-null	float64
6	JP_sales	16715 non-null	float64
7	Other_sales	16715 non-null	float64
8	Critic_Score	8137 non-null	float64
9	User_Score	10014 non-null	object
10	Rating	9949 non-null	object

```
dtypes: float64(6), object(5)
```

```
memory usage: 1.4+ MB
```

Доступны данные о продажах игр в 3 областях и для различных платформ. Большинство игр также имеют оценку пользователя по 10-бальной шкале, критика по 100-бальной, а также жанр и возрастной рейтинг.

Всего в датафрейме 16715 строк, в 5 столбцах имеются пропуски. Возможно удастся заменить 2 пропуска в названии игры и 2 пропуска в жанре руками, остальные пропуски можно заполнить данными из различных источников, но так как такой возможности я не имею придется оставить как есть. В данных о продажах нет пропусков. Тип данных в столбце User\_score необходимо перевести в float64. Названия столбцов следует привести к нижнему регистру. Столбцы с категориальными данными стоит проверить на неявные дубликаты.

## Подготовка данных

Прежде чем приступить к анализу данных необходимо выполнить предобработку. Возможно удастся обнаружить некорректные значения в категориальных столбцах или дубликаты. После этого мы сможем приступить к заполнению пропущенных значений и изменению типов данных.

## Замена названий столбцов

```
columns_lower = []

for column in data.columns:
    columns_lower.append(column.lower()) #Все названия в нижний индекс

data.set_axis(columns_lower,axis=1,inplace=True)# Переименуем столбцы
```

## Преобразование типов данных

```
data.loc[data['user_score']=='tbd','user_score'] = np.nan # Значение подлежащее определению заменим на nan
data['user_score'] = pd.to_numeric(data['user_score']) # Преобразуем столбец из object во float64
```

```
# Комментарий ревьюера 0
data[data[['name', 'platform', 'year_of_release', 'na_sales']].duplicated(keep=False)]
```

Empty DataFrame

Columns: [name, platform, year\_of\_release, genre, na\_sales, eu\_sales, jp\_sales, other\_sales, critic\_score, user\_score, rating]  
Index: []

Тип данных в столбце user\_score изначально был установлен как object из-за строкового значения "tbd" (подлежит уточнению), которое можно считать отсутствием значения. Типы данных в столбцах year\_of\_release и critic\_score нельзя заменить на тип данных int из-за пропусков, которые являются типом float.

## Обработка пропусков

Данные могут иметь явные пропуски, то есть абсолютно одинаковые строки и неявные пропуски, например ошибка в названии игры создаст 2 фактически одинаковые строки.

```
data.duplicated().sum() # Явных дубликатов нет
data['name'].value_counts() # Кажется, что это 12 дубликатов
```

Need for Speed: Most Wanted	12
Madden NFL 07	9
LEGO Marvel Super Heroes	9
Ratatouille	9
FIFA 14	9
..	..
Battlefield 2142	1
Family Jockey	1
Super Robot Taisen Impact	1
Batman: Gotham City Racer	1
Barbie and the Magic of Pegasus	1
Name: name, Length: 11559, dtype: int64	

На первый взгляд, кажется, что тот, кто составлял выборку, очень любил Need for Speed: Most Wanted и поэтому добавил одну и ту же игру 12 раз, но это только на первый взгляд. Игра вышла на разных платформах, а в 2012 году была переиздана под тем же названием, выходит это 12 равноправных строк таблицы.

```
data[data['name']=='Need for Speed: Most Wanted'] # Ничего
подозрительного
```

	name	platform	year_of_release	
genre \				
253	Need for Speed: Most Wanted	PS2	2005.0	Racing
523	Need for Speed: Most Wanted	PS3	2012.0	Racing
1190	Need for Speed: Most Wanted	X360	2012.0	Racing
1591	Need for Speed: Most Wanted	X360	2005.0	Racing
1998	Need for Speed: Most Wanted	XB	2005.0	Racing
2048	Need for Speed: Most Wanted	PSV	2012.0	Racing
3581	Need for Speed: Most Wanted	GC	2005.0	Racing
5972	Need for Speed: Most Wanted	PC	2005.0	Racing
6273	Need for Speed: Most Wanted	WiiU	2013.0	Racing
6410	Need for Speed: Most Wanted	DS	2005.0	Racing
6473	Need for Speed: Most Wanted	GBA	2005.0	Racing
11715	Need for Speed: Most Wanted	PC	2012.0	Racing

	na_sales	eu_sales	jp_sales	other_sales	critic_score
user_score \					
253	2.03	1.79	0.08	0.47	82.0
9.1					
523	0.71	1.46	0.06	0.58	NaN
NaN					
1190	0.62	0.78	0.01	0.15	83.0
8.5					
1591	1.00	0.13	0.02	0.10	83.0
8.5					
1998	0.53	0.46	0.00	0.05	83.0
8.8					
2048	0.33	0.45	0.01	0.22	NaN
NaN					
3581	0.43	0.11	0.00	0.02	80.0
9.1					

5972	0.02	0.23	0.00	0.04	82.0
8.5					
6273	0.13	0.12	0.00	0.02	NaN
NaN					
6410	0.24	0.01	0.00	0.02	45.0
6.1					
6473	0.19	0.07	0.00	0.00	NaN
8.3					
11715	0.00	0.06	0.00	0.02	82.0
8.5					

	rating
253	T
523	NaN
1190	T
1591	T
1998	T
2048	NaN
3581	T
5972	T
6273	NaN
6410	E
6473	E
11715	T

В начале я предполагал найти игры без названия по жанру, году выпуска и продажам, но данных недостаточно, поэтому удалим эти 2 строки.

```
display(data[data['name'].isna()])# Похоже найти названия игр не
получится придется удалить
data.dropna(subset=['name'],inplace=True)
data[data['name'].isna()]
```

	name	platform	year_of_release	genre	na_sales	eu_sales
jp_sales \						
659	NaN	GEN	1993.0	NaN	1.78	0.53
0.00						
14244	NaN	GEN	1993.0	NaN	0.00	0.00
0.03						

	other_sales	critic_score	user_score	rating
659	0.08	NaN	NaN	NaN
14244	0.00	NaN	NaN	NaN

Empty DataFrame

Columns: [name, platform, year\_of\_release, genre, na\_sales, eu\_sales, jp\_sales, other\_sales, critic\_score, user\_score, rating]  
Index: []

Значения категориальных данных могут быть записано с ошибкой, поэтому проверим все столбцы.

```
data['platform'].unique() # В названиях игровых платформ нет ошибок
array(['Wii', 'NES', 'GB', 'DS', 'X360', 'PS3', 'PS2', 'SNES', 'GBA',
      'PS4', '3DS', 'N64', 'PS', 'XB', 'PC', '2600', 'PSP', 'XOne',
      'WiiU', 'GC', 'GEN', 'DC', 'PSV', 'SAT', 'SCD', 'WS', 'NG',
      'TG16',
      '3D0', 'GG', 'PCFX'], dtype=object)

data['genre'].unique() # Жанры записаны правильно
array(['Sports', 'Platform', 'Racing', 'Role-Playing', 'Puzzle',
      'Misc',
      'Shooter', 'Simulation', 'Action', 'Fighting', 'Adventure',
      'Strategy'], dtype=object)

data['rating'].unique() # Значение K-A устарело, заменим его более
современным E
data.loc[data['rating']=='K-A', 'rating'] = 'E'
data['rating'].unique()
data.duplicated().sum() # Дубликатов не прибавилось
data.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 16713 entries, 0 to 16714
Data columns (total 11 columns):
#   Column                Non-Null Count  Dtype
---  -
0   name                  16713 non-null  object
1   platform              16713 non-null  object
2   year_of_release       16444 non-null  float64
3   genre                 16713 non-null  object
4   na_sales              16713 non-null  float64
5   eu_sales              16713 non-null  float64
6   jp_sales              16713 non-null  float64
7   other_sales           16713 non-null  float64
8   critic_score          8137 non-null   float64
9   user_score            7590 non-null   float64
10  rating                9949 non-null   object
dtypes: float64(7), object(4)
memory usage: 1.5+ MB

data.loc[data['rating'].isna(), 'rating'] = 'Another rating system'
```

Просмотреть 16 тысяч названий игр и найти в них дубликаты своими руками и глазами сложно(невозможно), но можно попробовать найти значения индентичные по всем остальным критериям, возможно так удастся найти дубликаты.

```
display(data[data.duplicated(subset=['genre','platform','year_of_release','user_score','critic_score','rating','na_sales','eu_sales','jp_sales'])])#Все равно слишком много значений для анализа
```

	name	platform	\
1906	Ready 2 Rumble Boxing	PS	
3042	Tiger Woods PGA Tour 2000	PS	
3159	SpongeBob SquarePants: Game Boy Advance Video ...	GBA	
3633	Fighting Force 2	PS	
3774	Winning Post	SNES	
...	...	...	
16703	Strawberry Nauts	PSV	
16707	Aiyoku no Eustia	PSV	
16710	Samurai Warriors: Sanada Maru	PS3	
16712	Haitaka no Psychedelica	PSV	
16713	Spirits & Spells	GBA	

	year_of_release	genre	na_sales	eu_sales	jp_sales
other_sales \					
1906	1998.0	Sports	0.60	0.41	0.00
0.07					
3042	1998.0	Sports	0.37	0.25	0.00
0.04					
3159	2004.0	Misc	0.46	0.17	0.00
0.01					
3633	1999.0	Action	0.31	0.21	0.00
0.04					
3774	1993.0	Sports	0.00	0.00	0.53
0.00					
...	...	...	...	...	...
...					
16703	2016.0	Adventure	0.00	0.00	0.01
0.00					
16707	2014.0	Misc	0.00	0.00	0.01
0.00					
16710	2016.0	Action	0.00	0.00	0.01
0.00					
16712	2016.0	Adventure	0.00	0.00	0.01
0.00					
16713	2003.0	Platform	0.01	0.00	0.00
0.00					

	critic_score	user_score	rating
1906	NaN	NaN	Another rating system
3042	NaN	NaN	Another rating system
3159	NaN	NaN	Another rating system
3633	NaN	NaN	Another rating system
3774	NaN	NaN	Another rating system
...	...	...	...
16703	NaN	NaN	Another rating system

16707	NaN	NaN	Another rating system
16710	NaN	NaN	Another rating system
16712	NaN	NaN	Another rating system
16713	NaN	NaN	Another rating system

[1030 rows x 11 columns]

```
display(data[data.duplicated(subset=['name','platform','year_of_release'],keep=False)])# Вот это уже похоже на дубликат
data.drop(index=16230,inplace=True) # 0.01 млн продаж в Европе
выглядят неубедительно
```

	name	platform	year_of_release	genre	na_sales
eu_sales \					
604	Madden NFL 13	PS3	2012.0	Sports	2.11
0.22					
16230	Madden NFL 13	PS3	2012.0	Sports	0.00
0.01					

	jp_sales	other_sales	critic_score	user_score	rating
604	0.0	0.23	83.0	5.5	E
16230	0.0	0.00	83.0	5.5	E

```
data.reset_index(drop=True) #Пронумеруем заново отредактированные
данные
```

	name	platform	year_of_release	
genre \				
0	Wii Sports	Wii	2006.0	
Sports				
1	Super Mario Bros.	NES	1985.0	
Platform				
2	Mario Kart Wii	Wii	2008.0	
Racing				
3	Wii Sports Resort	Wii	2009.0	
Sports				
4	Pokemon Red/Pokemon Blue	GB	1996.0	Role-
Playing				
...	...	...	...	
...				
16707	Samurai Warriors: Sanada Maru	PS3	2016.0	
Action				
16708	LMA Manager 2007	X360	2006.0	
Sports				
16709	Haitaka no Psychedelica	PSV	2016.0	
Adventure				
16710	Spirits & Spells	GBA	2003.0	
Platform				
16711	Winning Post 8 2016	PSV	2016.0	
Simulation				



	na_sales	eu_sales	jp_sales	other_sales	critic_score
0	41.36	28.96	3.77	8.45	76.0
1	29.08	3.58	6.81	0.77	NaN
2	15.68	12.76	3.79	3.29	82.0
3	15.61	10.93	3.28	2.95	80.0
4	11.27	8.89	10.22	1.00	NaN
...	...	...	...	...	...
16707	0.00	0.00	0.01	0.00	NaN
16708	0.00	0.01	0.00	0.00	NaN
16709	0.00	0.00	0.01	0.00	NaN
16710	0.01	0.00	0.00	0.00	NaN
16711	0.00	0.00	0.01	0.00	NaN

	rating
0	E
1	Another rating system
2	E
3	E
4	Another rating system
...	...
16707	Another rating system
16708	Another rating system
16709	Another rating system
16710	Another rating system
16711	Another rating system

[16712 rows x 11 columns]

## Суммарные продажи

Количество копий проданных во всем мире.

```
data['total_sales'] = data['na_sales']+data['eu_sales']
+data['jp_sales']+data['other_sales']
data
```

		name	platform	year_of_release		
genre \						
0		Wii Sports	Wii	2006.0		
Sports						
1		Super Mario Bros.	NES	1985.0		
Platform						
2		Mario Kart Wii	Wii	2008.0		
Racing						
3		Wii Sports Resort	Wii	2009.0		
Sports						
4		Pokemon Red/Pokemon Blue	GB	1996.0	Role-	
Playing						
...		...	...	...		
...						
16710		Samurai Warriors: Sanada Maru	PS3	2016.0		
Action						
16711		LMA Manager 2007	X360	2006.0		
Sports						
16712		Haitaka no Psychedelica	PSV	2016.0		
Adventure						
16713		Spirits & Spells	GBA	2003.0		
Platform						
16714		Winning Post 8 2016	PSV	2016.0		
Simulation						
		na_sales	eu_sales	jp_sales	other_sales	critic_score
user_score \						
0		41.36	28.96	3.77	8.45	76.0
8.0						
1		29.08	3.58	6.81	0.77	NaN
NaN						
2		15.68	12.76	3.79	3.29	82.0
8.3						
3		15.61	10.93	3.28	2.95	80.0
8.0						
4		11.27	8.89	10.22	1.00	NaN
NaN						
...		...	...	...	...	...
...						
16710		0.00	0.00	0.01	0.00	NaN
NaN						
16711		0.00	0.01	0.00	0.00	NaN
NaN						
16712		0.00	0.00	0.01	0.00	NaN
NaN						
16713		0.01	0.00	0.00	0.00	NaN
NaN						
16714		0.00	0.00	0.01	0.00	NaN
NaN						

	rating	total_sales
0	E	82.54
1	Another rating system	40.24
2	E	35.52
3	E	32.77
4	Another rating system	31.38
...	...	...
16710	Another rating system	0.01
16711	Another rating system	0.01
16712	Another rating system	0.01
16713	Another rating system	0.01
16714	Another rating system	0.01

[16712 rows x 12 columns]

data.info()

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 16712 entries, 0 to 16714
Data columns (total 12 columns):
#   Column                Non-Null Count  Dtype
---  -
0   name                   16712 non-null  object
1   platform               16712 non-null  object
2   year_of_release        16443 non-null  float64
3   genre                  16712 non-null  object
4   na_sales               16712 non-null  float64
5   eu_sales               16712 non-null  float64
6   jp_sales               16712 non-null  float64
7   other_sales            16712 non-null  float64
8   critic_score           8136 non-null   float64
9   user_score             7589 non-null   float64
10  rating                 16712 non-null  object
11  total_sales            16712 non-null  float64
dtypes: float64(8), object(4)
memory usage: 1.7+ MB
```

Названия столбцов были приведены к нижнему регистру. Игры, в которых пользовательская оценка не была установлена, были помечены как "tbd". После замены всех "tbd" на "np.nan" удалось изменить тип данных столбца с object на float. Явных дубликатов в данных не обнаружено, но имелся один не явный дубликат.

В столбцах с оценками пользователей, критиков, возрастным рейтингом и годом выпуска имеются пропуски. Это может быть связано с неполнотой данных в источнике, ошибками ввода данных, отсутствием информации о конкретном игровом проекте или недоступностью данных на момент сбора информации, поскольку данные собирались из открытых источников.

Названия платформ и жанров были записаны верно и остались без изменений, а значение возрастного рейтинга K-A (Kids to Adults) было заменено на E (Everyone). Было обнаружено 1030 потенциальных дубликатов при сравнении всех параметров, за исключением имени, но ближайший анализ показал, что эти строки не являются дубликатами, а просто содержат одинаковые значения. Только один дубликат был обнаружен после поиска дубликатов по трём столбцам.

## Исследовательский анализ данных

Чтобы продать товар, надо сначала продать борозду в мозгах.

Для грамотного построения маркетинговой программы платформы по продаже игр, необходимо знать на какую игру сделать ставку. Какой жанр любят японцы? На чем они играют? Сколько еще продержится та или иная платформа? Выявить эти закономерности и определить наиболее перспективные направления для размещения продукции поможет исследовательский анализ.

### Сколько игр выпускалось в разные годы

```
data.pivot_table(index='year_of_release', values='name', aggfunc='count')
)
```

year_of_release	name
1980.0	9
1981.0	46
1982.0	36
1983.0	17
1984.0	14
1985.0	14
1986.0	21
1987.0	16
1988.0	15
1989.0	17
1990.0	16
1991.0	41
1992.0	43
1993.0	60
1994.0	121
1995.0	219
1996.0	263
1997.0	289
1998.0	379
1999.0	338
2000.0	350
2001.0	482
2002.0	829
2003.0	775
2004.0	762
2005.0	939

2006.0	1006
2007.0	1197
2008.0	1427
2009.0	1426
2010.0	1255
2011.0	1136
2012.0	652
2013.0	544
2014.0	581
2015.0	606
2016.0	502

```
data['year_of_release'].hist(bins=data['year_of_release'].nunique())
plt.title("Динамика продажи игр")
plt.xlabel("Год")
plt.ylabel("Количество проданных копий");
```



На гистограмме видно, что в 80-90 годах количество выпускаемых игр не росло, да и в целом игр тогда выпускалось мало. Так как мы строим прогноз на 2016 год, нас интересуют актуальные данные, поэтому данные за 80-90 можно объединить в группу, удалить или оставить как есть, они обещали не мешать.

## Как менялись продажи по платформам

Количество платформ в исходных достаточно велико поэтому возьмем топ 5 платформ по общим продажам, и хотя некоторые из наиболее продаваемых платформ уже не актуальны, это поможет выявить "смену поколений".

```
data.pivot_table(index=['platform'],values='total_sales',aggfunc='sum')
.sort_values(by='total_sales',ascending=False).head()
```

	total_sales
platform	
PS2	1255.77
X360	971.42
PS3	939.64
Wii	907.51
DS	806.12

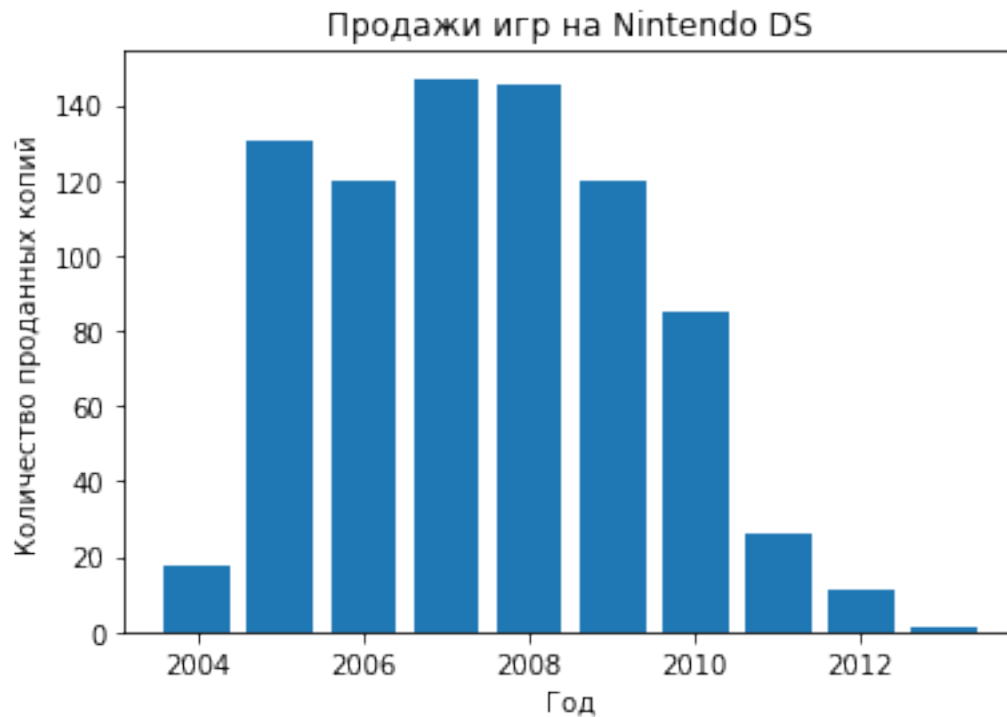
3 поколения PlayStation попали в лидеры по количеству проданных игр, логично предположить что и 4 поколение скоро их догонит.

```
data[data['name']=='Strongest Tokyo University Shogi DS']# Strongest
Tokyo University Shogi DS для DS вышла в 85 году, и это при том что DS
вышла в 2004
```

```
data.loc[15957,'year_of_release'] = 2007
```

```
top5 =
data.pivot_table(index='year_of_release',values='total_sales',columns=
'platform',aggfunc='sum').loc[:,
['PS2','X360','PS3','Wii','DS']].dropna(how='all')
```

```
plt.bar(x = top5.index, height = top5['DS'])
plt.title("Продажи игр на Nintendo DS")
plt.xlabel("Год")
plt.ylabel("Количество проданных копий");
```



Продажи Nintendo DS были значительными, особенно на рынке Японии, где имела широкую популярность среди игроков всех возрастов. Резкое снижение количества игр в 2011 году может быть связано с выходом приставки 3DS, следующего поколения приставок Nintendo.

```
plt.bar(x = top5.index, height = top5['Wii'])  
plt.title("Продажи игр на Wii")  
plt.xlabel("Год")  
plt.ylabel("Количество проданных копий");
```



Еще одна японская игровая платформа, которая добилась успеха не только в японии, но и смогла покорить сердца множества игроков из Америки и Европы. В отличии от DS, Wii является портативной консолью. Поддержка платформы была прекращена в 2013 году.

```
plt.bar(x = top5.index, height = top5['PS2'])  
plt.title("Продажи игр на Play Station 2")  
plt.xlabel("Год")  
plt.ylabel("Количество проданных копий");
```

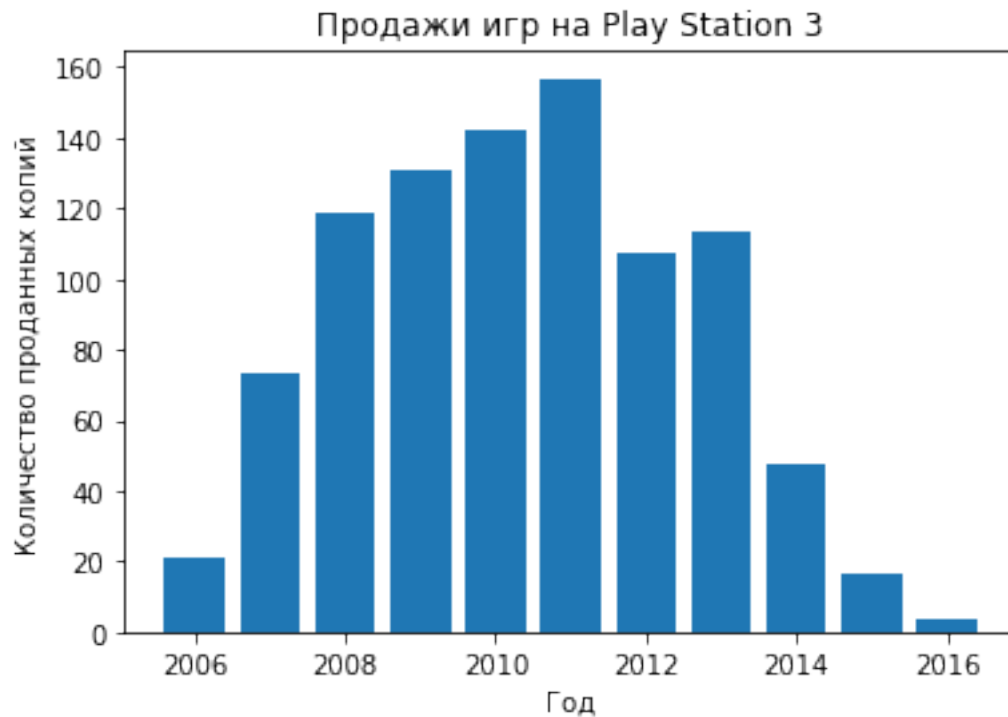




Игровая приставка стала наиболее быстро продаваемой и самой популярной игровой консолью в истории. На третий квартал 2011 года в мире было продано почти 155 миллионов экземпляров PS2.

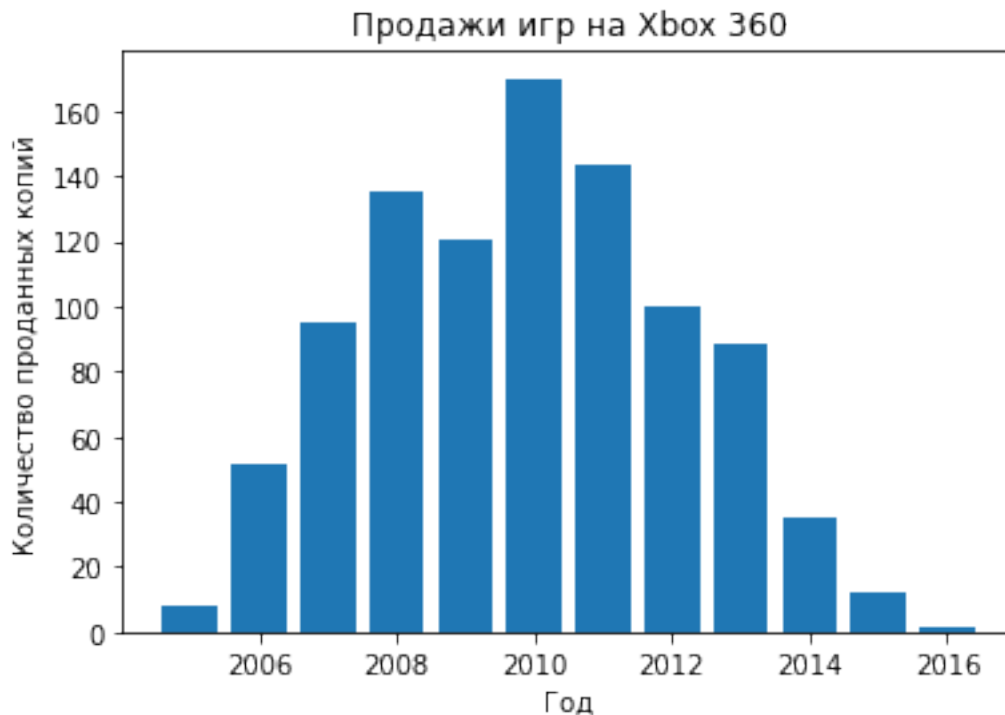
28 декабря 2012 года было прекращено производство в Японии, а 4 января 2013 года Sony подтвердила, что производство PS2 прекращено по всему миру. Тем самым, PS2 выпускалась почти 13 лет, это один из самых длинных жизненных циклов в истории игровой индустрии.

```
plt.bar(x = top5.index, height = top5['PS3'])
plt.title("Продажи игр на Play Station 3")
plt.xlabel("Год")
plt.ylabel("Количество проданных копий");
```



Производство PS3 было прекращено 29 мая 2017 года. И снова приставка от компании Sony производилась больше 10 лет.

```
plt.bar(x = top5.index, height = top5['X360'])  
plt.title("Продажи игр на Xbox 360")  
plt.xlabel("Год")  
plt.ylabel("Количество проданных копий");
```



```
temp = data.copy()
time_life = temp.pivot_table(index='platform',
                              values='year_of_release', aggfunc=['min', 'max']).reset_index()
time_life['life'] = time_life['max'] - time_life['min'] + 1 # в срок
жизни платформы, должны входить границы диапазона, # поэтому
+1
time_life['life'].median()
```

Игровая приставка от компании Microsoft не уступила конкурентам и была снята с производства в 2016 году, через более чем 11 лет с момента выхода.

Игровые приставки PlayStation 2, PlayStation 3 и Xbox продолжали оставаться популярными еще долгое время после выпуска следующих поколений консолей. Продолжение продаж этих платформ может быть связано с широким каталогом игр, которые были доступны для них, а также с их способностью воспроизводить контент в достаточно высоком качестве.

Срок устаревания консолей составляет около 10 лет, после этого консоль перестают поддерживать. С выходом нового поколения консолей, разработчики игр переключаются на следующее поколение, происходит снижение числа выпускаемых игр для предыдущего поколения консоли игр. Это связано с устареванием железа в консоли. Согласно закону Мура производительность железа за 2 года растет в 2 раза, то есть консоль которая выйдет через 10 лет будет в 32 раза производительней. Ничего удивительного в том что игроки и разработчики идут в ногу со временем.

Несмотря на то, что консоли выходят из употребления после 10 лет, многие люди продолжают использовать их для игр и развлечений. Некоторые даже коллекционируют

старые модели консолей, так как они имеют историческую ценность и являются частью игровой культуры.

## Прогноз по росту продаж на различных платформах для 2017 года

Самые популярные консоли живут около 10 лет и так уж совпало что имнно они нас и интересуют, поэтому актуальным будем считать период за последние 5 лет.

```
current_platforms
=data.pivot_table(index='year_of_release',values='total_sales',columns
='platform',aggfunc='sum').loc[2012.0:].dropna(axis='columns',how='all
')
#current_platforms.index = current_platforms.index.astype('int64')

plt.plot(current_platforms)
plt.title("Годовые продажи игр на разных платформах")
plt.xlabel("Год")
plt.ylabel("Количество продаж")
plt.legend(current_platforms.columns.tolist(),bbox_to_anchor=(1,1),loc
="upper left");
```



За прошедшие 5 лет, наиболее актуальными платформами являются PS4 и Xbox One. В 2016 году обе платформы были обновлены, вышли версии slim и pro для PlayStation, Microsoft же презентовали Xbox One S и анонсировали Xbox One X. Это говорит о том, что обе платформы не потеряют актуальности в ближайшее время.

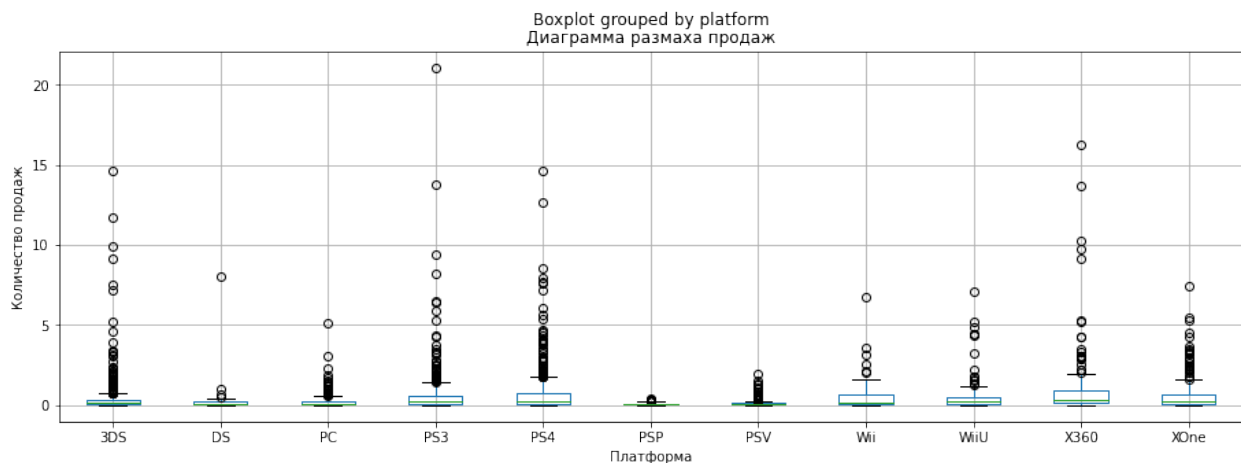
## Глобальные продажи игр в разбивке по платформам

Различные игровые платформы обладают уникальными характеристиками, которые используются для создания и выпуска эксклюзивных игр. Каждая консоль имеет свои особенности и возможности, которые разработчики игр используют для создания уникального контента, который будет доступен только на определенной платформе.

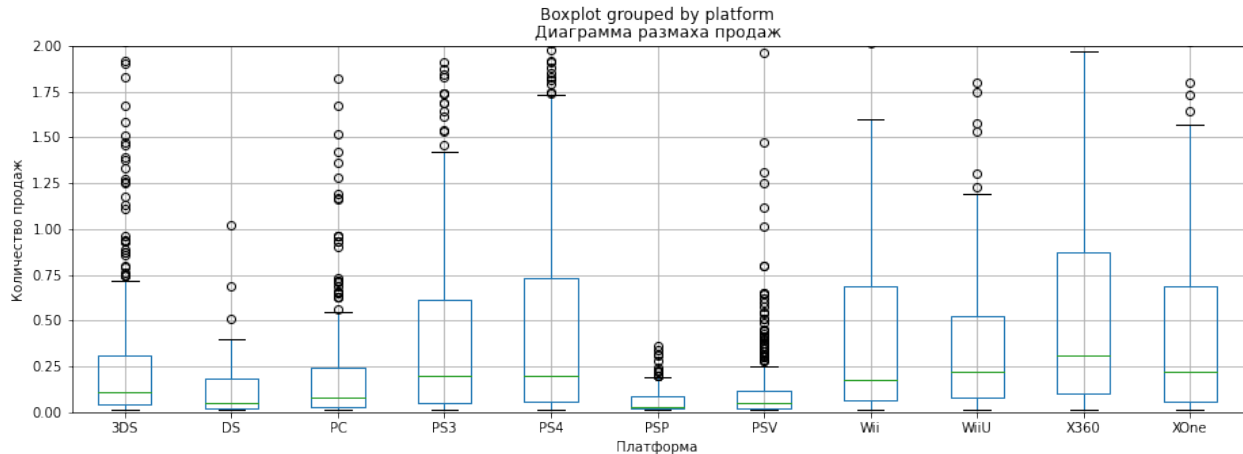
```
data = data[data['year_of_release']>=2012]

data[data['platform']=='Wii'].sort_values(by='total_sales',ascending=False) #Wii Sports вложена в коробку и появилась в продаже вместе с приставкой
data[data['platform']=='X360'].sort_values(by='total_sales',ascending=False)#Kinect Adventures! входит в комплект поставки контроллера

data.boxplot(column='total_sales',by='platform',figsize=(15,5))
plt.title("Диаграмма размаха продаж")
plt.xlabel("Платформа")
plt.ylabel("Количество продаж");
```



```
data.boxplot(column='total_sales',by='platform',figsize=(15,5))
plt.ylim((0,2))
plt.title("Диаграмма размаха продаж")
plt.xlabel("Платформа")
plt.ylabel("Количество продаж");
```



У многих платформ есть свои хиты выпущенные специально для того, чтобы использовать возможности платформы по максимуму. Они часто считаются "флагманскими" играми платформы и могут иметь большое влияние на популярность консоли. Однако, вне зависимости от эксклюзивов, каждая платформа обычно имеет широкий выбор игр, от самых популярных до менее известных, чтобы удовлетворить различные предпочтения игроков. В целом можно сказать о том что есть очень популярные игры, а есть все остальные.

## Как влияют на продажи внутри одной популярной платформы отзывы пользователей и критиков

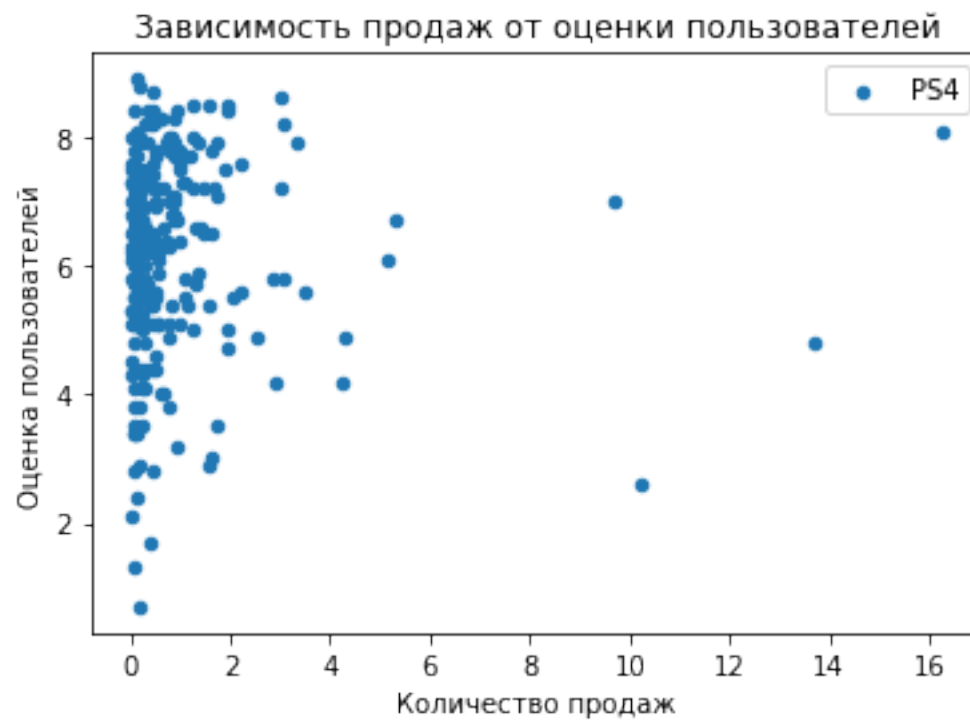
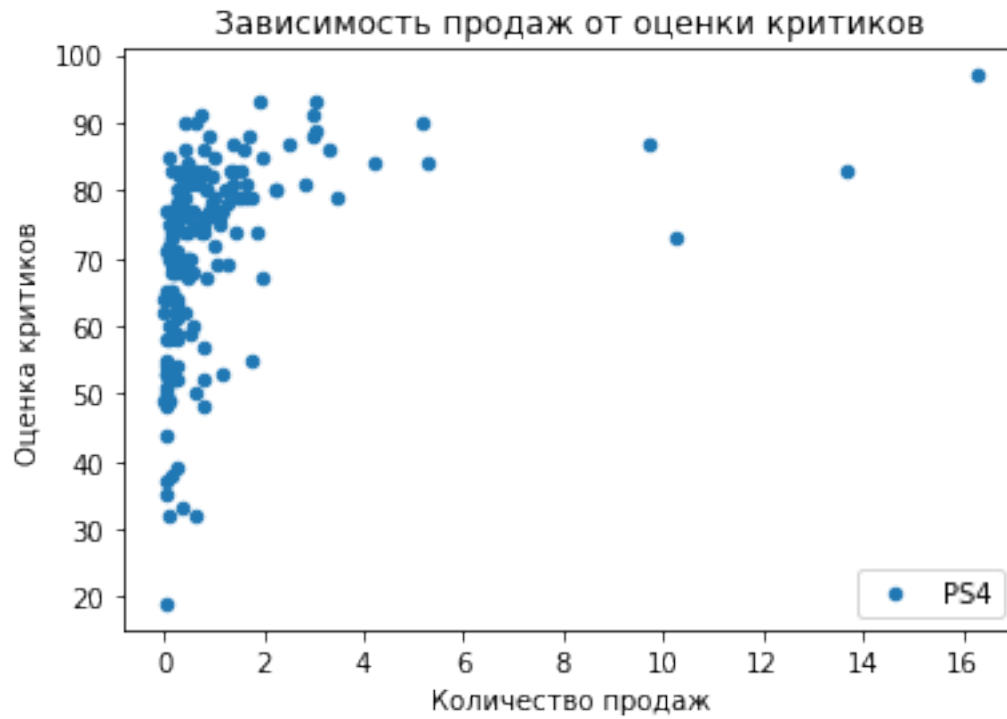
Отзывы игр от игроков важны для продажи игр, потому что они помогают покупателям принять решение о покупке, основываясь на реальном опыте других игроков. Положительные отзывы увеличивают доверие к игре, а негативные могут предупредить о проблемах. Как еще узнать о качестве игры если сам не играл.

```
data[data['platform']=='X360'].plot.scatter(x='total_sales',y='critic_score',label='PS4')
plt.title("Зависимость продаж от оценки критиков")
plt.xlabel("Количество продаж")
plt.ylabel("Оценка критиков")
plt.legend()
```

```
data[data['platform']=='X360'].plot.scatter(x='total_sales',y='user_score',label='PS4')
plt.title("Зависимость продаж от оценки пользователей")
plt.xlabel("Количество продаж")
plt.ylabel("Оценка пользователей")
plt.legend()
```

```
data[data['platform']=='PS4'].corr().loc['total_sales','critic_score':  
'user_score']
```

```
critic_score    0.406568  
user_score      -0.031957  
Name: total_sales, dtype: float64
```



```
other_platform =
data.pivot_table(index='platform',values='total_sales',aggfunc='sum').
sort_values(by='total_sales',ascending=False).iloc[1:6].index.to_list(
)
```

```
for platform in other_platform:
    display(f'Коэффициенты корреляции продаж и оценок для платформы
{platform}',
data[data['platform']==platform].corr().loc['total_sales','critic_score':
'user_score'])
    plt.figure(figsize=[10, 10])
    plt.subplot(2,2,1)
    plt.scatter(x=data[data['platform']==platform]
['total_sales'],y=data[data['platform']==platform]
['critic_score'],label=platform)
    plt.title("Зависимость продаж от оценки критиков")
    plt.xlabel("Количество продаж")
    plt.ylabel("Оценка критиков")
    plt.legend();

    plt.subplot(2,2,2)
    plt.scatter(x=data[data['platform']==platform]
['total_sales'],y=data[data['platform']==platform]
['user_score'],label=platform)
    plt.title("Зависимость продаж от оценки пользователей")
    plt.xlabel("Количество продаж")
    plt.ylabel("Оценка пользователей")
    plt.legend();
```

'Коэффициенты корреляции продаж и оценок для платформы PS3'

```
critic_score    0.334152
user_score      -0.006206
Name: total_sales, dtype: float64
```

'Коэффициенты корреляции продаж и оценок для платформы X360'

```
critic_score    0.360573
user_score       0.006164
Name: total_sales, dtype: float64
```

'Коэффициенты корреляции продаж и оценок для платформы 3DS'

```
critic_score    0.320803
user_score       0.197583
Name: total_sales, dtype: float64
```

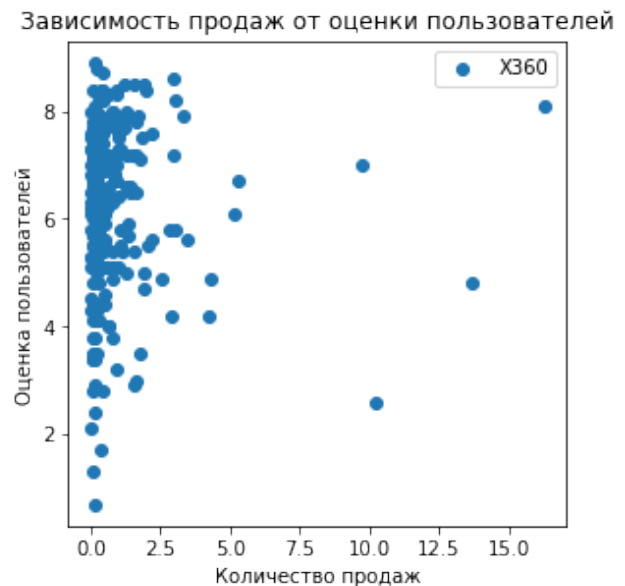
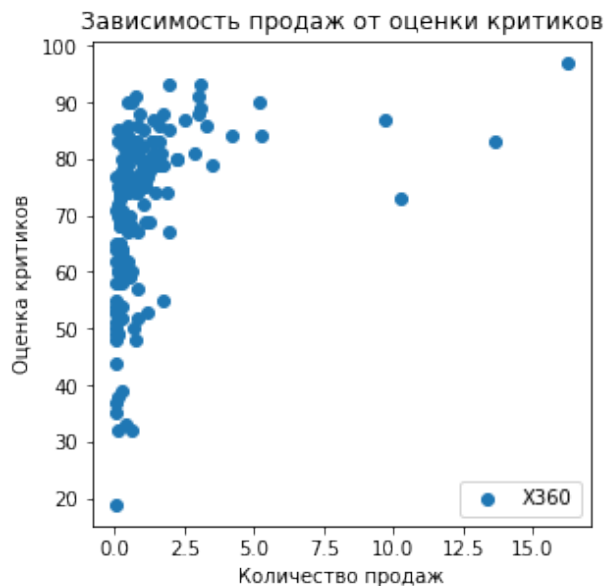
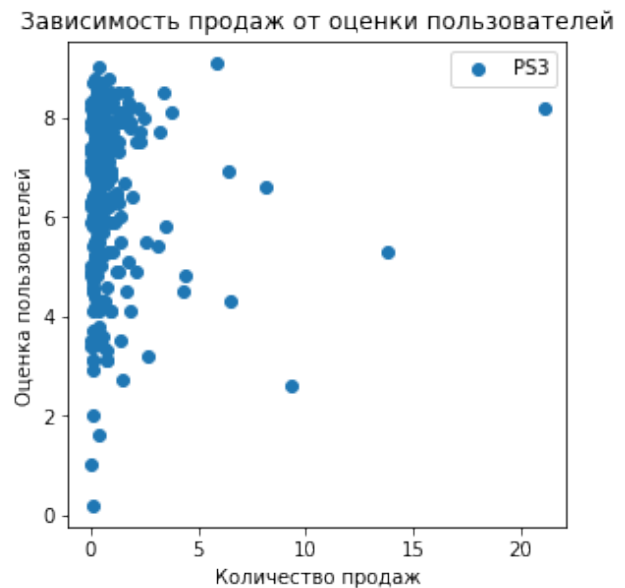
'Коэффициенты корреляции продаж и оценок для платформы XOne'



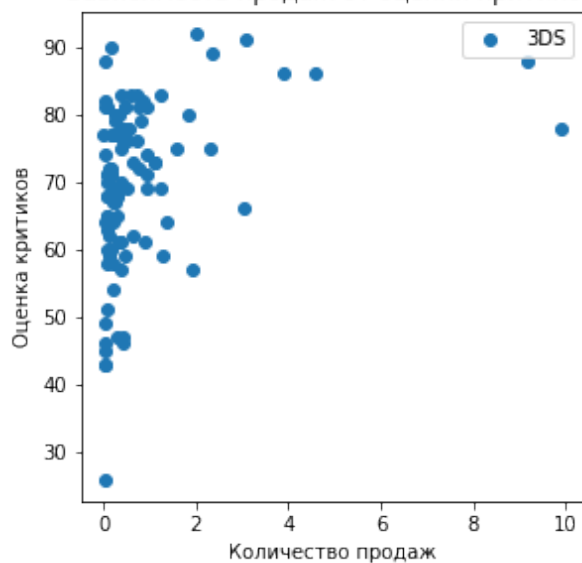
```
critic_score    0.416998
user_score      -0.068925
Name: total_sales, dtype: float64
```

'Коэффициенты корреляции продаж и оценок для платформы WiiU'

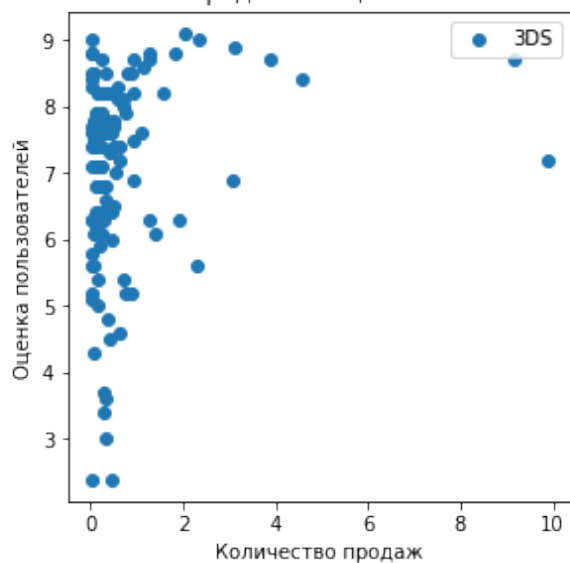
```
critic_score    0.348380
user_score      0.400219
Name: total_sales, dtype: float64
```



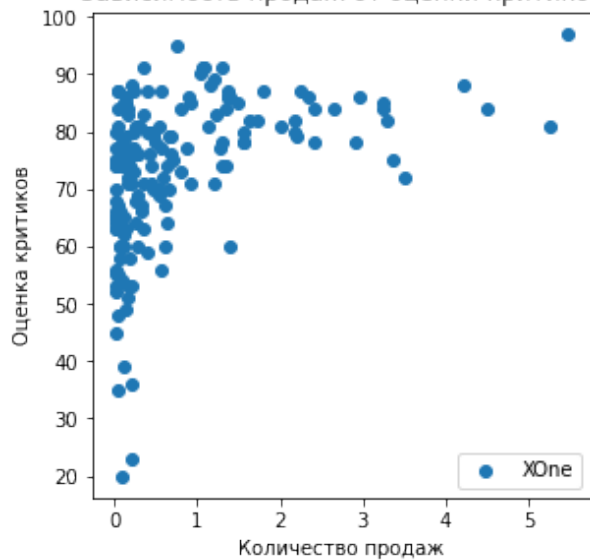
Зависимость продаж от оценки критиков



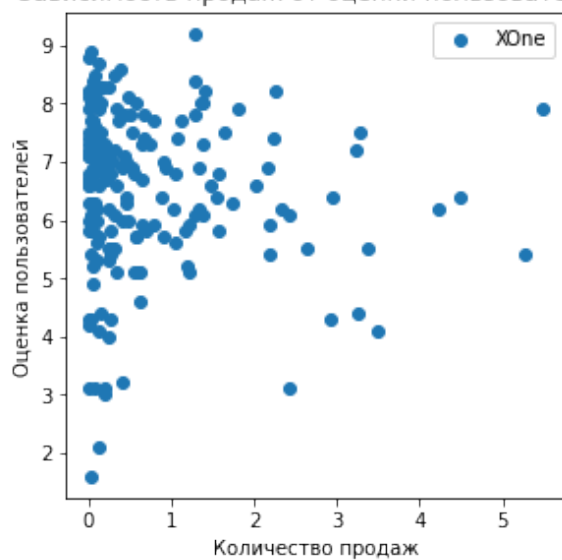
Зависимость продаж от оценки пользователей



Зависимость продаж от оценки критиков



Зависимость продаж от оценки пользователей





Положительные отзывы критиков, такие как известные игровые ресурсы или журналы, могут вызвать доверие у покупателей и привлечь их внимание к игре. Отзывы пользователей также могут иметь сильно влияние, так как они представляют мнение тех, кто уже приобрел и сыграл игру. Однако коэффициент корреляции между продажами и отзывами критиков довольно низок, это говорит о том, что скорее всего зависимость продаж от отзывов имеет более сложную форму. Мнения игроков могут не совпадать с отзывами критиков, корреляция с продажами практически отсутствует. Люди могут покупать игры на основе других факторов, таких как маркетинговые кампании, предыдущий опыт с похожими играми или влияние друзей и сообщества. Все эти факторы могут привести к низкой корреляции между отзывами и продажами.

## Распределение игр по жанрам

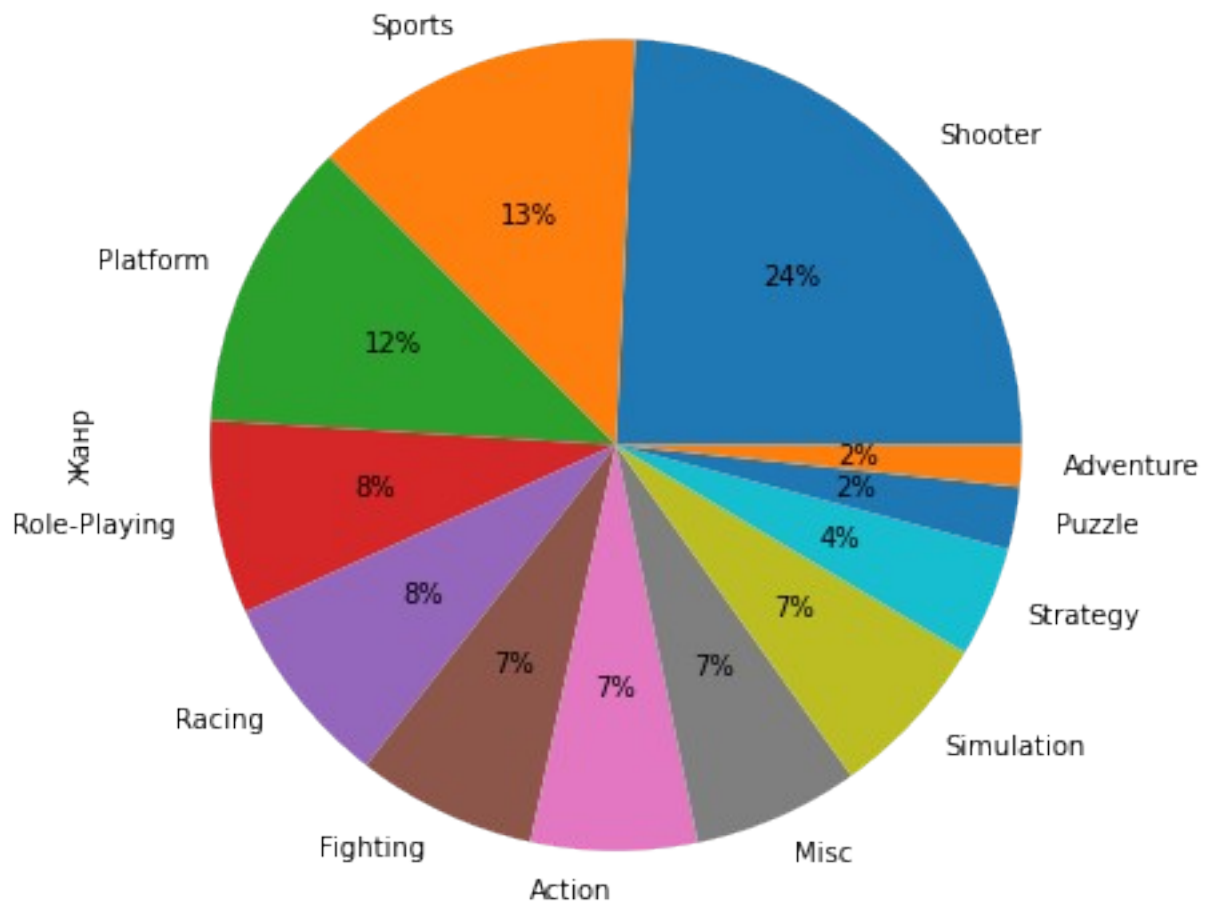
```
genre_sales =
data.pivot_table(index='genre', values=['eu_sales', 'jp_sales', 'na_sales', 'total_sales'], aggfunc='median').sort_values(by='total_sales', ascending=False)
display(genre_sales)

genre_sales['total_sales'].plot.pie(figsize=(7,7), autopct='%1.1f%%')
plt.title("Наиболее продаваемые жанры")
plt.ylabel("Жанр");
```

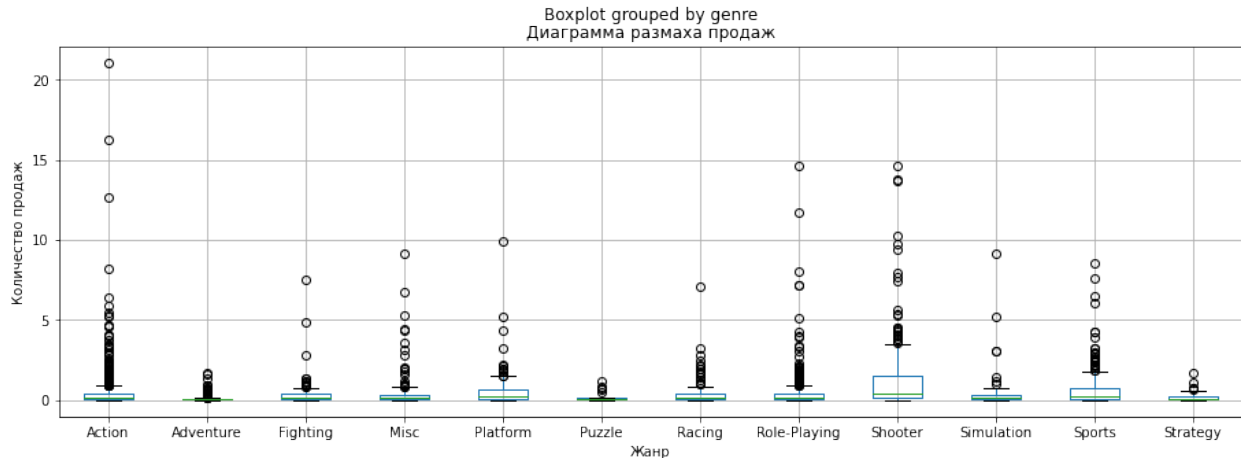
	eu_sales	jp_sales	na_sales	total_sales
genre				
Shooter	0.170	0.00	0.20	0.440
Sports	0.050	0.00	0.08	0.240
Platform	0.080	0.00	0.09	0.210
Role-Playing	0.010	0.06	0.02	0.140
Racing	0.090	0.00	0.04	0.140
Fighting	0.020	0.03	0.05	0.130
Action	0.020	0.01	0.02	0.120

Misc	0.015	0.02	0.01	0.120
Simulation	0.030	0.00	0.00	0.120
Strategy	0.010	0.00	0.00	0.080
Puzzle	0.010	0.00	0.00	0.045
Adventure	0.000	0.01	0.00	0.030

Наиболее продаваемые жанры



```
data.boxplot(column='total_sales',by='genre',figsize=(15,5))
plt.title("Диаграмма размаха продаж")
plt.xlabel("Жанр")
plt.ylabel("Количество продаж");
```



Наибольшей популярностью пользуются шутеры, игры этого жанра в среднем притягивают большее количество пользователей на разных платформах. Ролевые игры и экшены могут пользоваться огромным спросом, однако большая часть игр этого жанра остается незамеченными. Если взглянуть на таблицу, то можно заметить что в японии шутеры не пользуются особой популярностью, это может быть связано с цензурой в регионе или региональными предпочтениями. Региональные особенности могут оказать значительное влияние на продажу тех или иных игр.

Исследовательский анализ позволил сформировать общие рекомендации и выявить закономерности, но как стало понятно в процессе, предпочтения пользователей в различных регионах могут сильно отличаться, поэтому в дальнейшем необходимо подробно остановиться на каждом регионе.

В целом можно сказать что рынок видеоигр на текущий момент является перспективным направлением индустрии развлечений. Крупные технологические корпорации уверенно вкладывают деньги в это направление создавая свои платформы и эксклюзивные игры. Различные платформы предлагают разные возможности и привлекают разные аудитории.

## Портрет пользователя каждого региона

Портрет пользователя должен быть актуальным, так как вкусы людей могут меняться со временем. Новые игры, тренды и технологии могут влиять на предпочтения пользователей. Поэтому для прогноза на следующий год важно учитывать текущие предпочтения и интересы пользователей, чтобы предложить им актуальные игры и контент.

## Самые популярные платформы

Региональные особенности могут значительно влиять на спрос игр для различных платформах. Причины могут быть различными начиная от желания поддержать отечественного производителя, до культурных особенностей или экономических показателей.

```
platform_sales_eu =
data.pivot_table(index='platform', values=['eu_sales', 'jp_sales', 'na_sa
```

```

les', 'total_sales', 'other_sales'], aggfunc='sum').sort_values(by='eu_sales', ascending=False) # Это будут самые популярные платформы за 6 лет

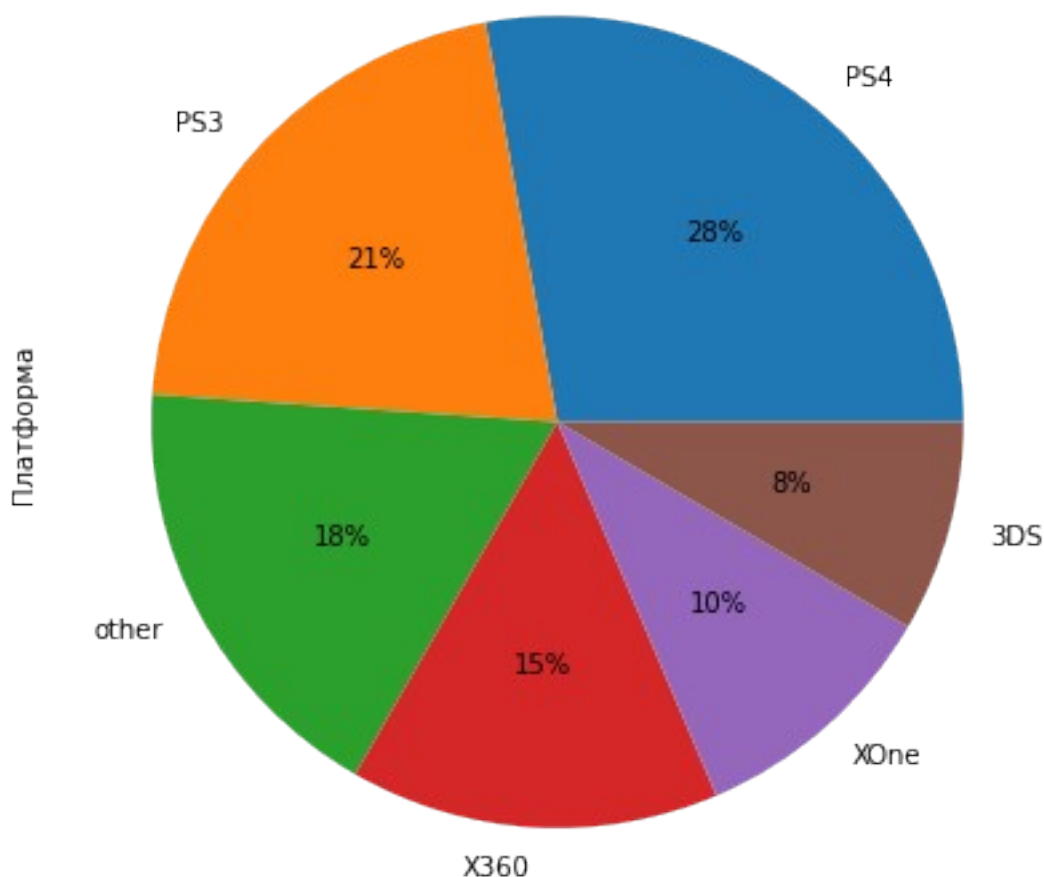
platform_sales_eu.loc['other'] = platform_sales_eu.iloc[5:].sum() #
Запишем все что меньше первой 5ки в остальные
platform_sales_eu =
platform_sales_eu.sort_values(by='eu_sales', ascending=False).head(6) #
удалим остальные
display(platform_sales_eu)

platform_sales_eu['eu_sales'].plot.pie(figsize=(7,7), autopct='%1.1f%%')
plt.title("Наиболее популярные платформы в Европе")
plt.ylabel("Платформа");

```

	eu_sales	jp_sales	na_sales	other_sales	total_sales
platform					
PS4	141.09	15.96	108.74	48.35	314.14
PS3	106.85	35.29	103.38	43.26	288.78
other	90.12	51.63	90.37	21.01	253.13
X360	74.52	1.57	140.05	20.40	236.54
XOne	51.59	0.34	93.12	14.27	159.32
3DS	42.64	87.79	55.31	8.87	194.61

## Наиболее популярные платформы в Европе



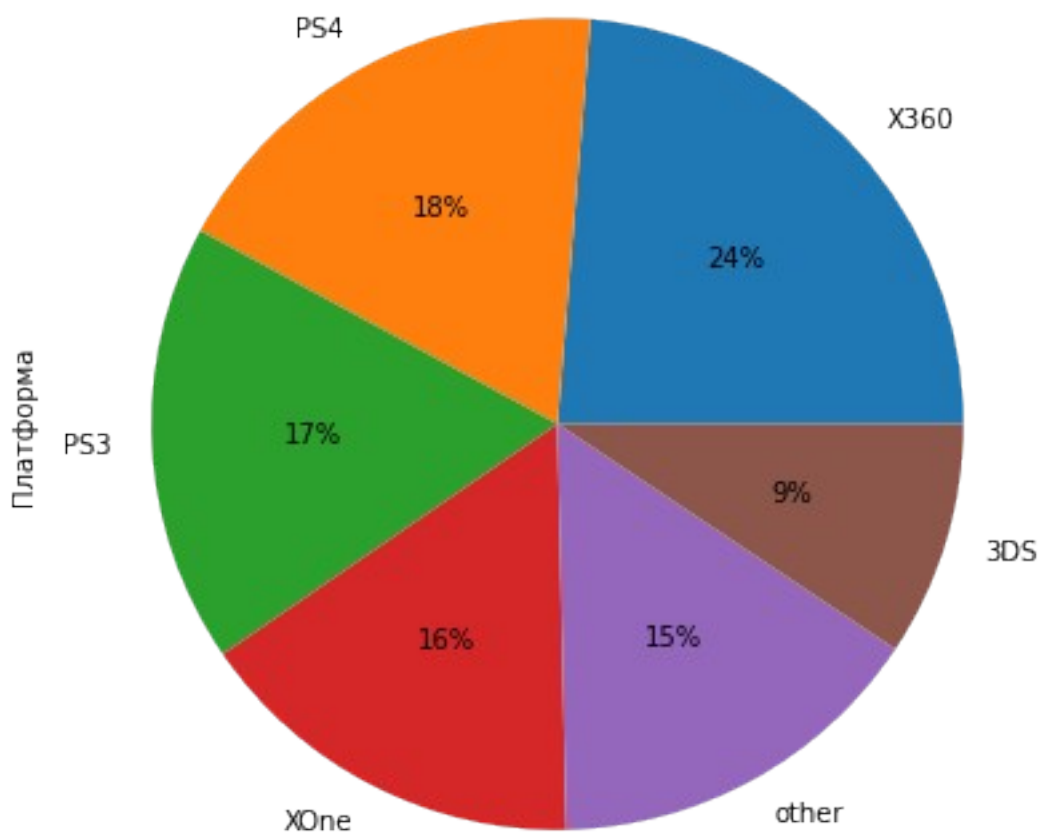
Пользователи Европы предпочитают стационарные консоли от компании Sony, консолями компании Microsoft, но обе компании имеют большую долю рынка. Портативные консоли и компьютеры также имеют заметное присутствие на рынке игровых устройств в Европе.

```
platform_sales_na =  
data.pivot_table(index='platform', values=['eu_sales', 'jp_sales', 'na_sales', 'total_sales', 'other_sales'], aggfunc='sum').sort_values(by='na_sales', ascending=False) # Отличается только сортировка по региону  
  
platform_sales_na.loc['other'] = platform_sales_na.iloc[5:].sum() #  
Запишем все что меньше первой 5ки в остальные  
platform_sales_na =  
platform_sales_na.sort_values(by='na_sales', ascending=False).head(6) #  
удалим остальные  
display(platform_sales_na)  
  
platform_sales_na['na_sales'].plot.pie(figsize=(7,7), autopct='%1.1f%%')
```

```
plt.title("Наиболее популярные платформы в Северной Америке")
plt.ylabel("Платформа");
```

	eu_sales	jp_sales	na_sales	other_sales	total_sales
platform					
X360	74.52	1.57	140.05	20.40	236.54
PS4	141.09	15.96	108.74	48.35	314.14
PS3	106.85	35.29	103.38	43.26	288.78
XOne	51.59	0.34	93.12	14.27	159.32
other	90.12	51.63	90.37	21.01	253.13
3DS	42.64	87.79	55.31	8.87	194.61

Наиболее популярные платформы в Северной Америке



В Америке же все наоборот, и хотя PS4 обгоняет по продажам Xbox One, в целом компания Microsoft владеет большей долей рынка чем ее основной конкурент. Скорее всего это связано с тем, что Microsoft является американской компанией, и ее продукты часто имеют большую популярность на родном рынке. Кроме того, у Microsoft есть долгая история на американском рынке, и различные факторы, такие как маркетинг, партнерства с

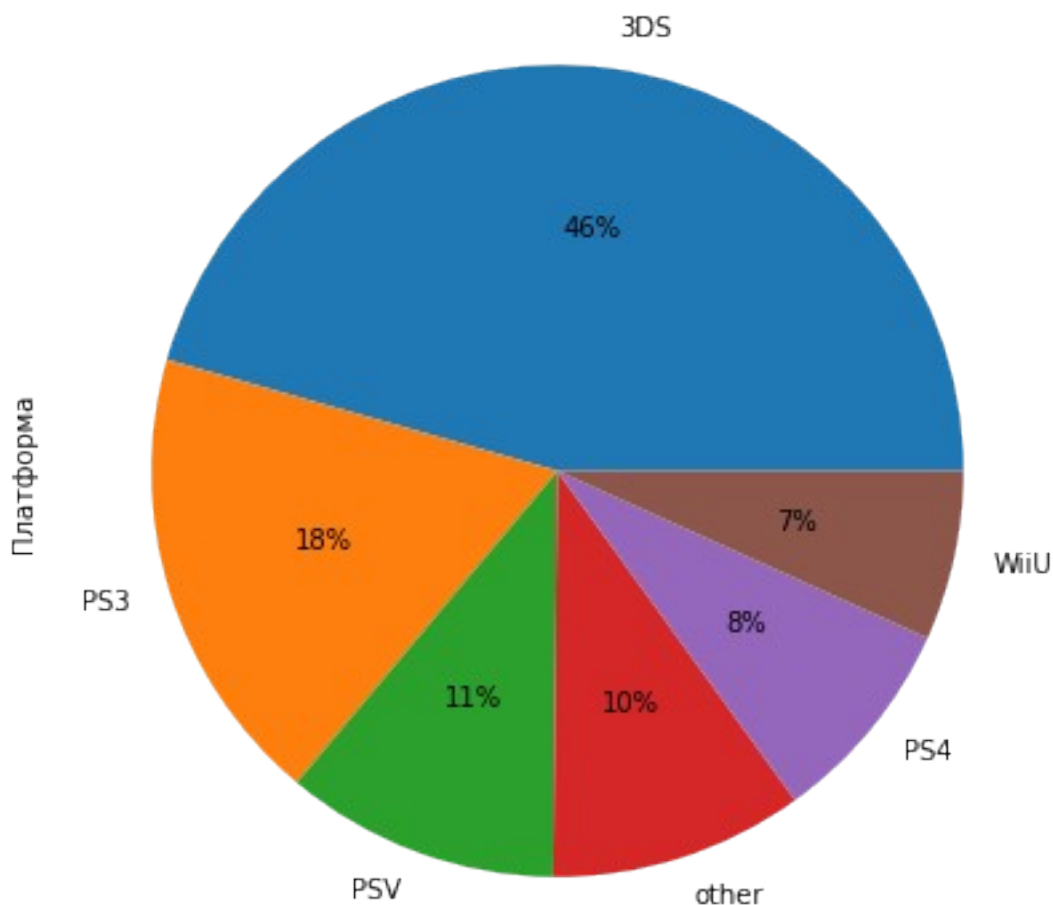


разработчиками игр и доступность эксклюзивных контентов, могут также оказывать влияние на предпочтения потребителей этого региона.

```
platform_sales_jp =  
data.pivot_table(index='platform',values=['eu_sales','jp_sales','na_sa  
les','total_sales','other_sales'],aggfunc='sum').sort_values(by='jp_sa  
les',ascending=False)# Отличается только сортировка по региону  
  
platform_sales_jp.loc['other'] = platform_sales_jp.iloc[5:].sum()#  
Запишем все что меньше первой 5ки в остальные  
platform_sales_jp =  
platform_sales_jp.sort_values(by='jp_sales',ascending=False).head(6)#  
удалим остальные  
display(platform_sales_jp)  
  
platform_sales_jp['jp_sales'].plot.pie(figsize=(7,7),autopct='%1.f%%')  
plt.title("Наиболее популярные платформы в Японии")  
plt.ylabel("Платформа");
```

	eu_sales	jp_sales	na_sales	other_sales	total_sales
platform					
3DS	42.64	87.79	55.31	8.87	194.61
PS3	106.85	35.29	103.38	43.26	288.78
PSV	11.36	21.04	10.98	5.80	49.18
other	179.74	19.49	274.46	43.93	517.62
PS4	141.09	15.96	108.74	48.35	314.14
WiiU	25.13	13.01	38.10	5.95	82.19

## Наиболее популярные платформы в Японии



Ситуация на рынке игровых консолей в Японии отличается от других регионов. В Японии портативные игровые консоли имеют огромную популярность, и Nintendo и Sony традиционно являются лидерами на этом рынке. Культура видеоигр в Японии имеет долгую историю, и многие японские игроки предпочитают портативные устройства для игр в дороге, в общественном транспорте или дома.

Стационарные игровые консоли имеют значительно меньшую популярность в Японии, поэтому продажи таких устройств относительно невелики. Кроме того, компания Sony, как японская компания, имеет преимущество на своем родном рынке, что еще больше укрепляет ее доминирующее положение в сфере стационарных игровых консолей в Японии.

## Самые популярные жанры

Природа не храм, и уж тем более не мастерская. Природа — тир, и огонь в нём надо вести на поражение.

Как уже было выяснено особенности национального гейминга в каждом регионе носят свой характер. Различные игровые платформы часто ориентированы на определенные жанры игр.

```
genre_sales_eu =
data.pivot_table(index='genre', values=['eu_sales', 'jp_sales', 'na_sales',
', 'total_sales'], aggfunc='sum').sort_values(by='eu_sales', ascending=False)
# Самые популярные жанры за 6 лет
display(genre_sales_eu)

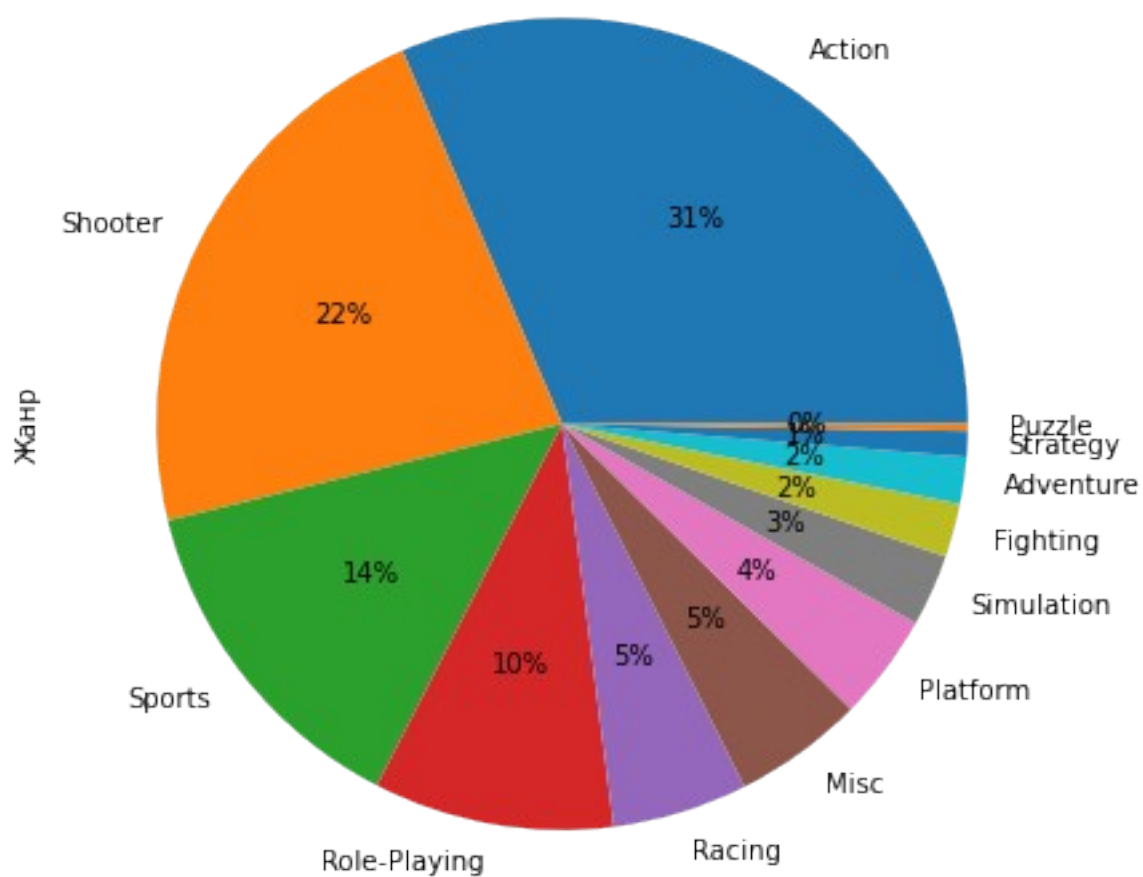
genre_sales_eu['eu_sales'].plot.pie(figsize=(7,7), autopct='%1.1f%%',
normalize=True)
plt.title("Наиболее популярные жанры в Европе")
plt.ylabel("Жанр")
plt.show();

genre_sales_eu.loc['other'] =
genre_sales_eu.sort_values(by='total_sales', ascending=False).iloc[5:].
sum()
# а не выдумать ли новый жанр
genre_sales_eu =
genre_sales_eu.sort_values(by='eu_sales', ascending=False).head(6)

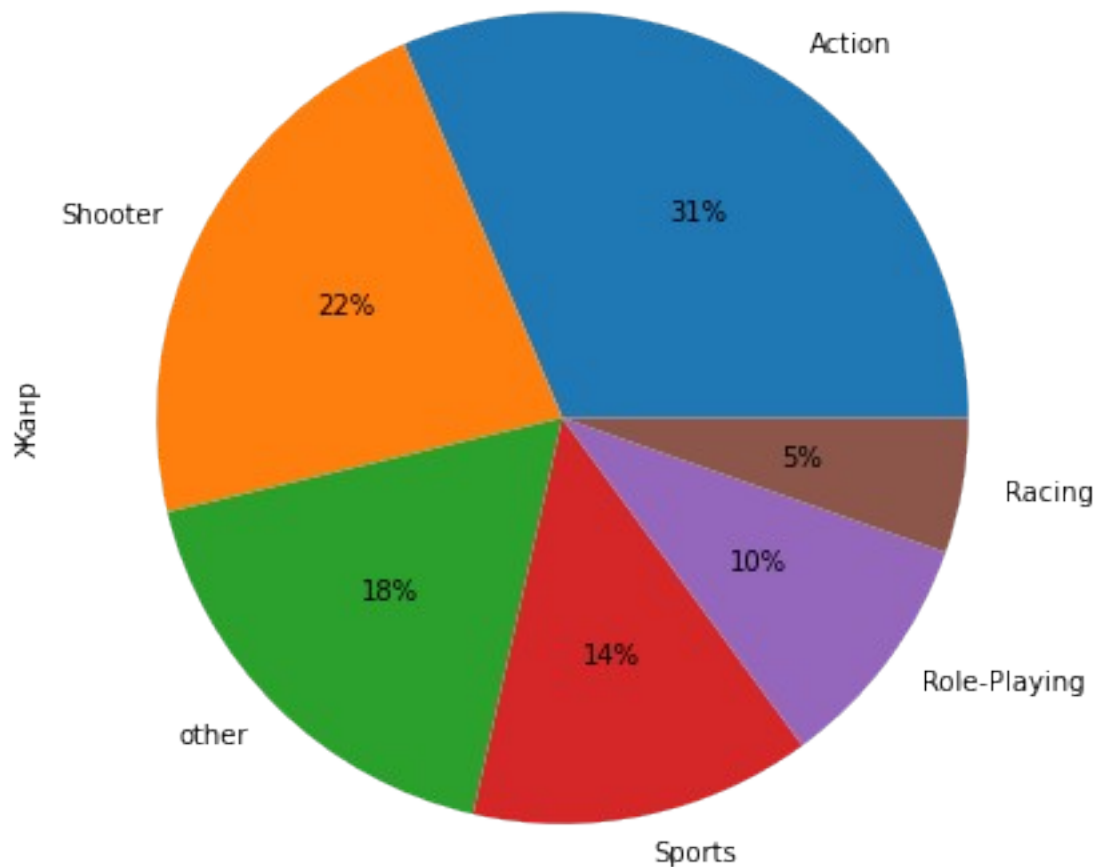
genre_sales_eu['eu_sales'].plot.pie(figsize=(7,7), autopct='%1.1f%%',
normalize=True)
plt.title("Наиболее популярные жанры в Европе топ 5")
plt.ylabel("Жанр");
```

	eu_sales	jp_sales	na_sales	total_sales
genre				
Action	159.34	52.80	177.84	441.12
Shooter	113.47	9.23	144.77	304.73
Sports	69.08	8.01	81.53	181.06
Role-Playing	48.53	65.44	64.00	192.80
Racing	27.29	2.50	17.22	53.50
Misc	26.32	12.86	38.19	85.04
Platform	21.41	8.63	25.38	61.00
Simulation	14.55	10.41	7.97	35.12
Fighting	10.79	9.44	19.79	44.49
Adventure	9.46	8.24	8.92	29.43
Strategy	5.17	2.88	4.23	13.34
Puzzle	1.40	2.14	1.13	4.89

## Наиболее популярные жанры в Европе



### Наиболее популярные жанры в Европе топ 5



Жители Европы отдают предпочтение экшенам. Вторыми по популярности являются шутеры, а замыкают тройку лидеров спорт.

```
genre_sales_na =
data.pivot_table(index='genre', values=['eu_sales', 'jp_sales', 'na_sales', 'total_sales'], aggfunc='sum').sort_values(by='na_sales', ascending=False) # Самые популярные жанры за 6 лет
display(genre_sales_na)

genre_sales_na['na_sales'].plot.pie(figsize=(7,7), autopct='%1.1f%%', normalize=True)
plt.title("Наиболее популярные жанры в Северной Америке")
plt.ylabel("Жанр")
plt.show();

genre_sales_na.rename({'Misc': 'other'}, axis=0, inplace=True) # Так как Misc попадает в топ 5 заранее отнесем его к other
```

```

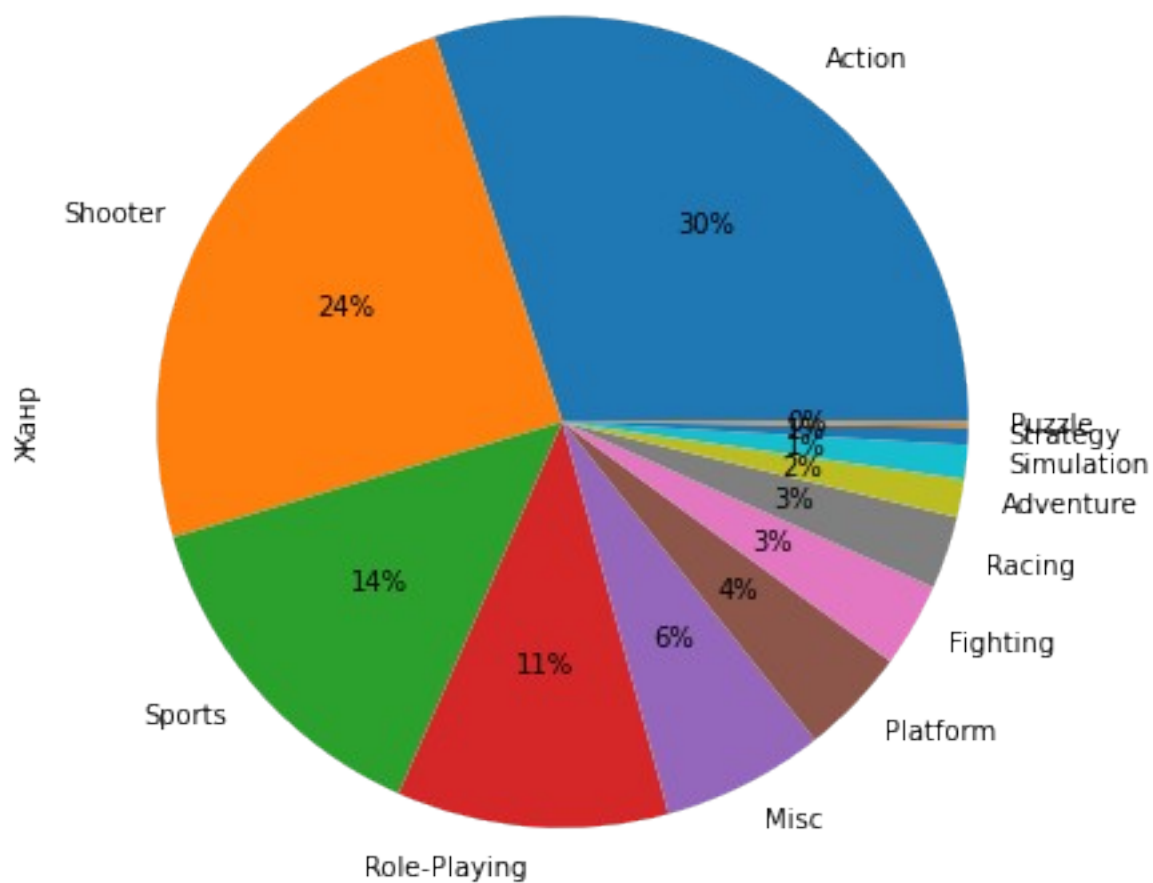
genre_sales_na.loc['other'] = genre_sales_na.loc['other'] +
genre_sales_na.sort_values(by='na_sales',ascending=False).iloc[5:].sum
()# а не выдумать ли новый жанр
genre_sales_na =
genre_sales_na.sort_values(by='na_sales',ascending=False).head(6)

genre_sales_na['na_sales'].plot.pie(figsize=(7,7),autopct='%1.f%
%',normalize=True)
plt.title("Наиболее популярные жанры в Северной Америке топ 5")
plt.ylabel("Жанр");

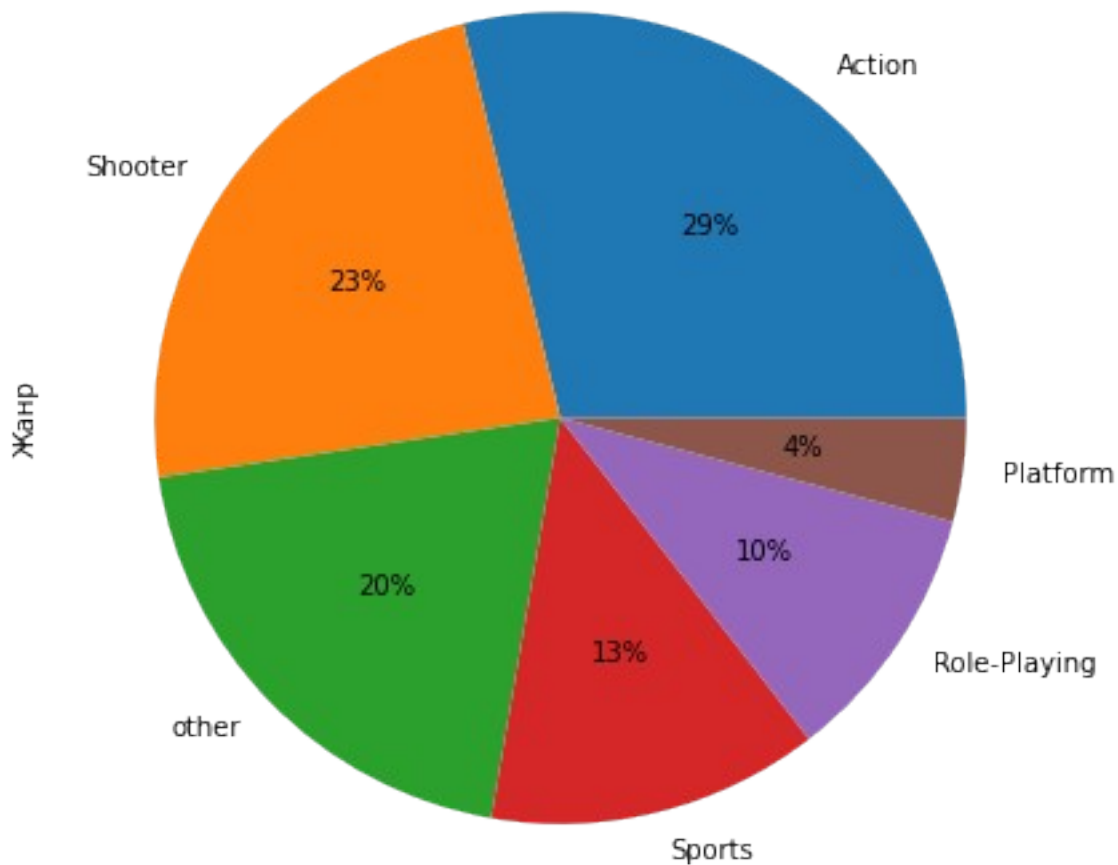
```

	eu_sales	jp_sales	na_sales	total_sales
genre				
Action	159.34	52.80	177.84	441.12
Shooter	113.47	9.23	144.77	304.73
Sports	69.08	8.01	81.53	181.06
Role-Playing	48.53	65.44	64.00	192.80
Misc	26.32	12.86	38.19	85.04
Platform	21.41	8.63	25.38	61.00
Fighting	10.79	9.44	19.79	44.49
Racing	27.29	2.50	17.22	53.50
Adventure	9.46	8.24	8.92	29.43
Simulation	14.55	10.41	7.97	35.12
Strategy	5.17	2.88	4.23	13.34
Puzzle	1.40	2.14	1.13	4.89

## Наиболее популярные жанры в Северной Америке



## Наиболее популярные жанры в Северной Америке топ 5



В Америке первая четверка лидеров аналогична европейской, однако любовь европейцев к гонкам американцы не разделяют, предпочитая платформеры.

```
genre_sales_jp =
data.pivot_table(index='genre', values=['eu_sales', 'jp_sales', 'na_sales', 'total_sales'], aggfunc='sum').sort_values(by='jp_sales', ascending=False) # Самые популярные жанры за 6 лет
display(genre_sales_jp)

genre_sales_jp['jp_sales'].plot.pie(figsize=(7,7), autopct='%1.1f%%', normalize=True)
plt.title("Наиболее популярные жанры в Японии")
plt.ylabel("Жанр")
plt.show();

genre_sales_jp.rename({'Misc': 'other'}, axis=0, inplace=True)
```



```

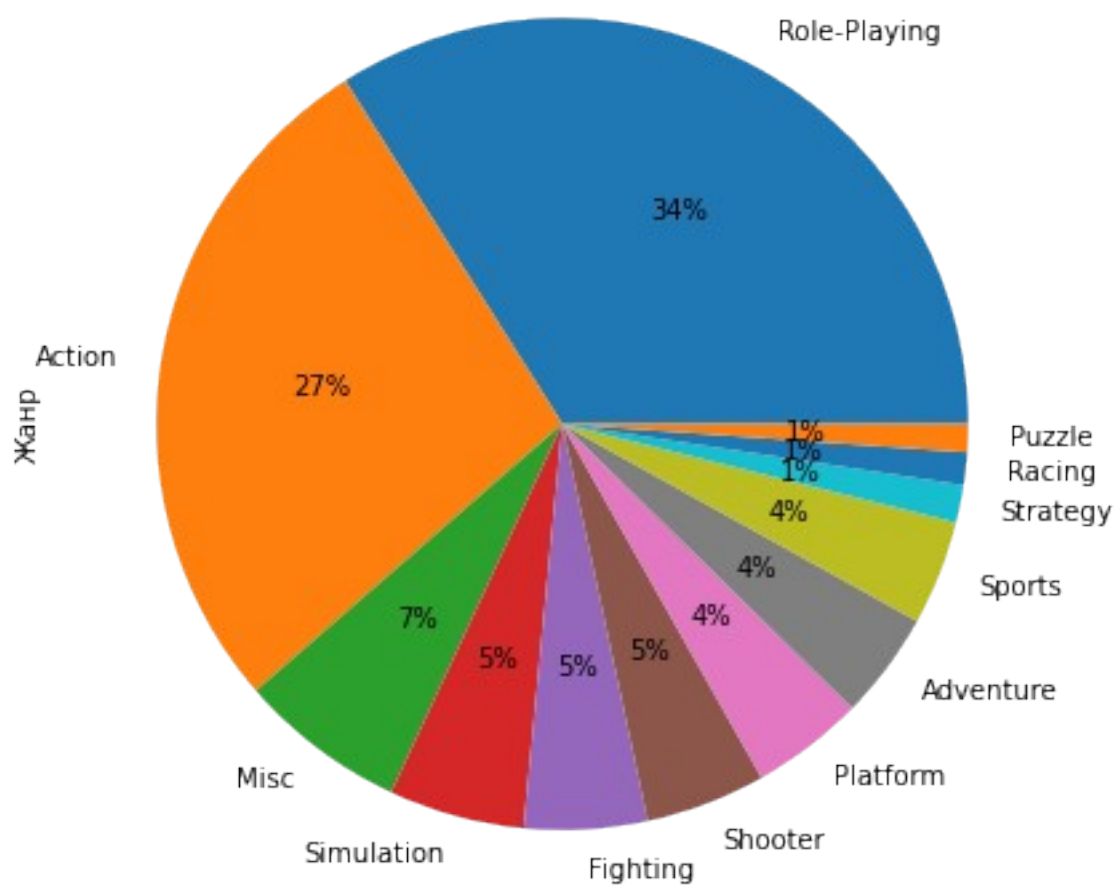
genre_sales_jp.loc['other'] = genre_sales_jp.loc['other'] +
genre_sales_jp.sort_values(by='jp_sales',ascending=False).iloc[7:].sum
()# Другой другой
genre_sales_jp =
genre_sales_jp.sort_values(by='jp_sales',ascending=False).head(5)

genre_sales_jp['jp_sales'].plot.pie(figsize=(7,7),autopct='%1.f%
%',normalize=True)
plt.title("Наиболее популярные жанры в Японии топ 5")
plt.ylabel("Жанр");

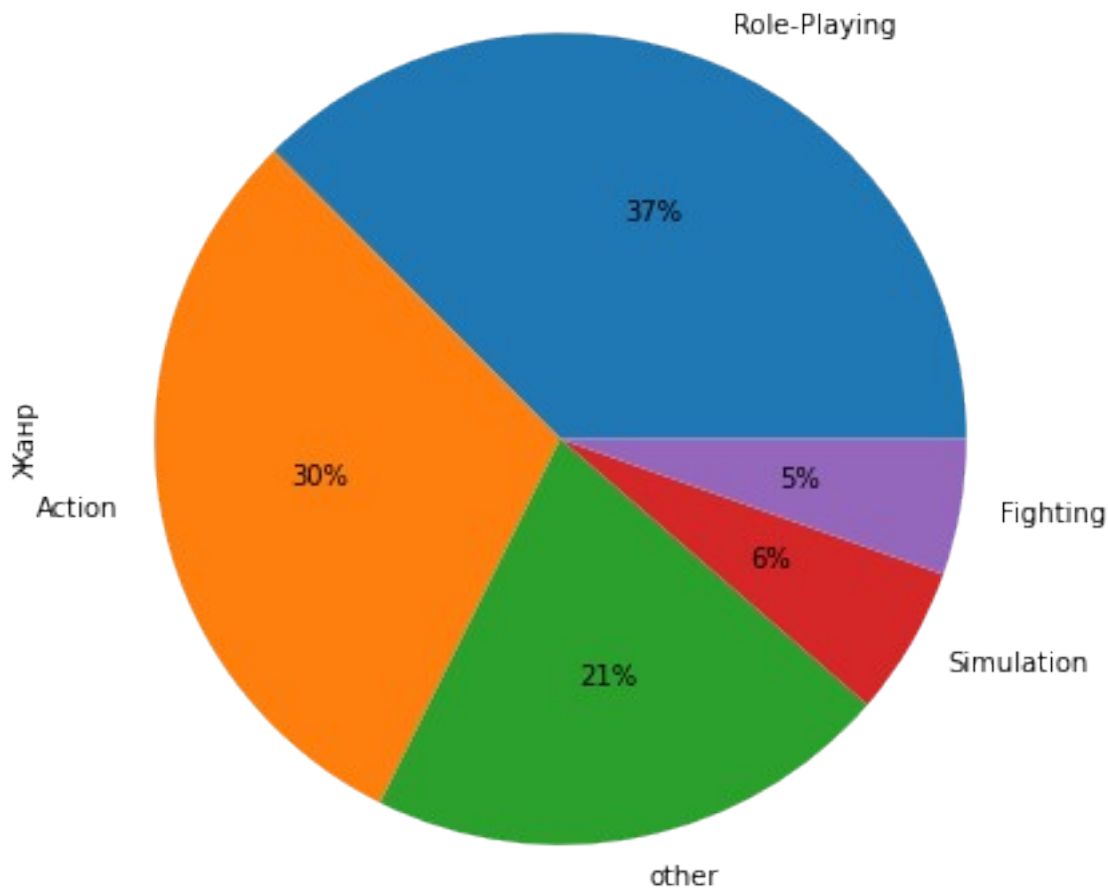
```

	eu_sales	jp_sales	na_sales	total_sales
genre				
Role-Playing	48.53	65.44	64.00	192.80
Action	159.34	52.80	177.84	441.12
Misc	26.32	12.86	38.19	85.04
Simulation	14.55	10.41	7.97	35.12
Fighting	10.79	9.44	19.79	44.49
Shooter	113.47	9.23	144.77	304.73
Platform	21.41	8.63	25.38	61.00
Adventure	9.46	8.24	8.92	29.43
Sports	69.08	8.01	81.53	181.06
Strategy	5.17	2.88	4.23	13.34
Racing	27.29	2.50	17.22	53.50
Puzzle	1.40	2.14	1.13	4.89

## Наиболее популярные жанры в Японии



## Наиболее популярные жанры в Японии топ 5

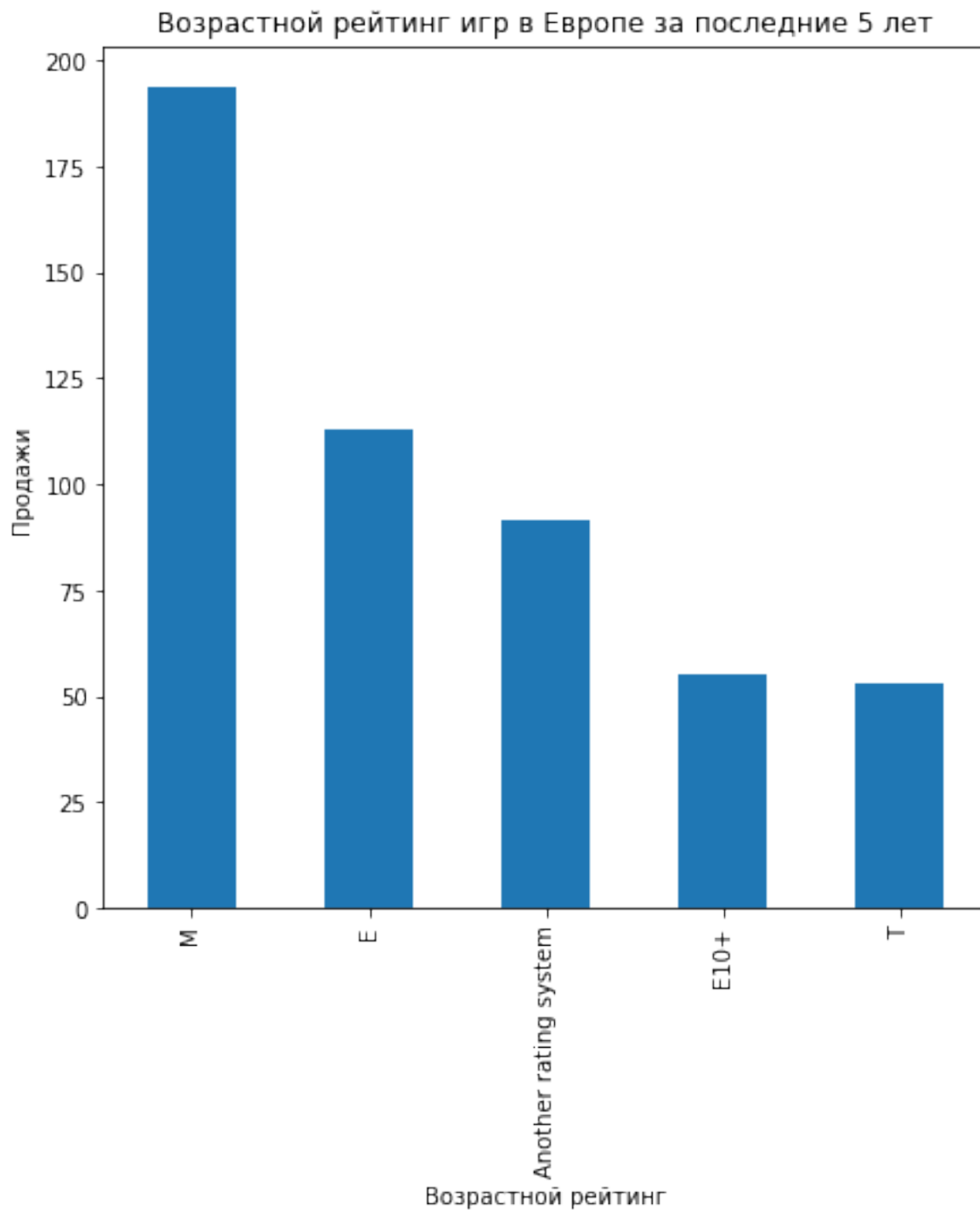


Большую часть рынка видеоигр в Японии занимают ролевые игры, а если точнее японские ролевые игры JRPG. Большинство таких игр было разработано в Японии и в первую очередь для внутреннего японского рынка для игровых приставок. Жанр экшен, по видимому, любим игроками со всего света. Такие жанры как симулятор, файтинг, шутер, платформер, приключения и спорт в равной степени востребованы пользователями, а остальные жанры остаются практически незамеченными.

## Влияет ли рейтинг ESRB на продажи в отдельном регионе

Возрастной рейтинг игр может оказать значительное влияние на их продажи. В различных странах существуют различные системы классификации игр по возрастным рейтингам, и рейтинг ESRB (Entertainment Software Rating Board), PEGI (Pan European Game Information) и CERO (Computer Entertainment Rating Organization) имеют разные значимости в разных регионах. Некоторые игры могут иметь ограниченное количество потенциальных покупателей из-за своего возрастного рейтинга, поэтому разработчики и издатели игр должны учитывать этот фактор при планировании маркетинговых стратегий.

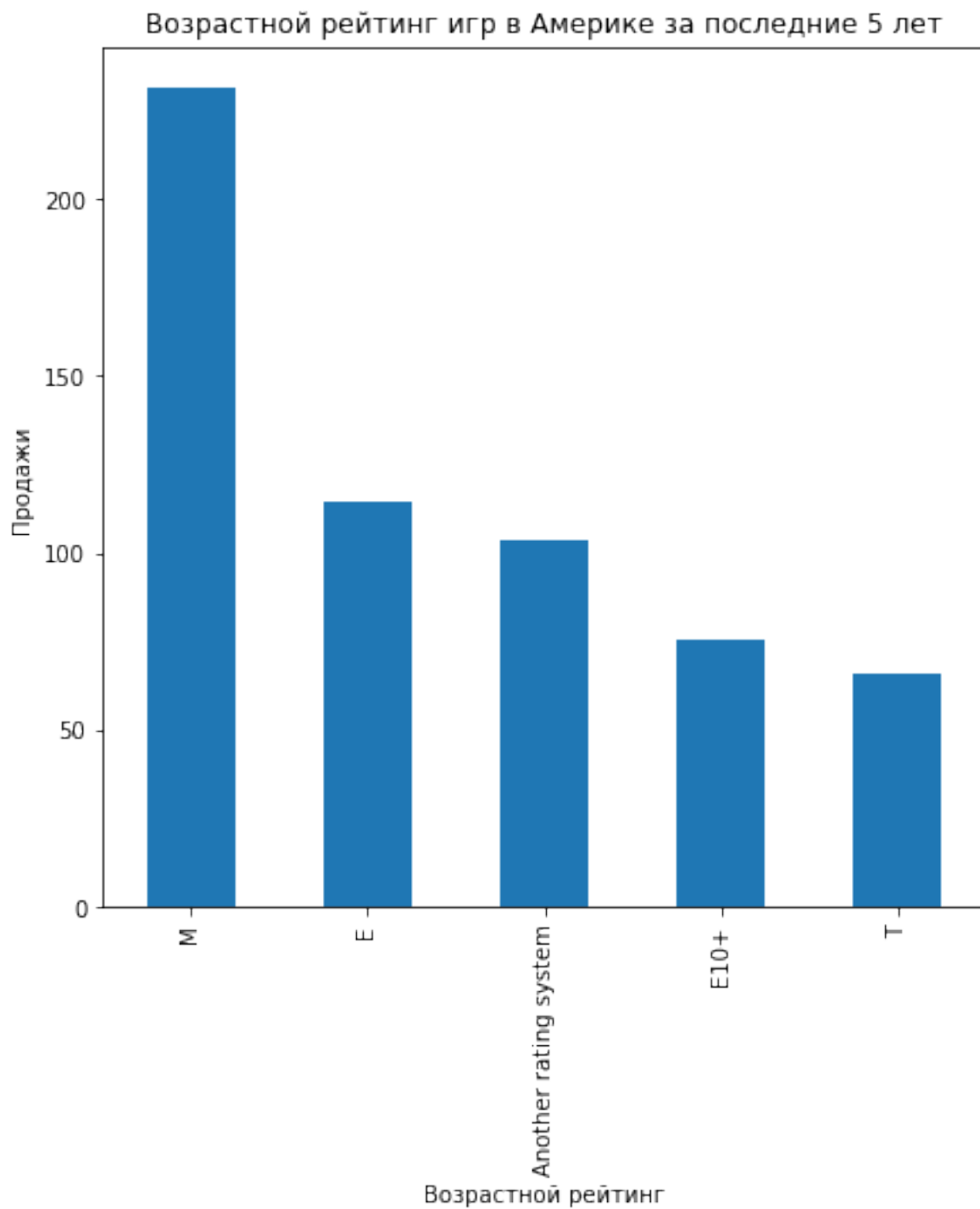
```
esrb_sales =  
data.pivot_table(index='rating', values=['eu_sales', 'jp_sales', 'na_sales', 'total_sales'], aggfunc='sum') # Это будут самые популярные платформы за 6 лет  
  
esrb_sales['eu_sales'].sort_values(ascending=False).plot.bar(figsize=(7,7))  
plt.title("Возрастной рейтинг игр в Европе за последние 5 лет")  
plt.xlabel("Возрастной рейтинг")  
plt.ylabel("Продажи");
```



Игры для взрослых (17+) в Европе пользуются значительной популярностью, привлекая внимание широкого круга игроков.

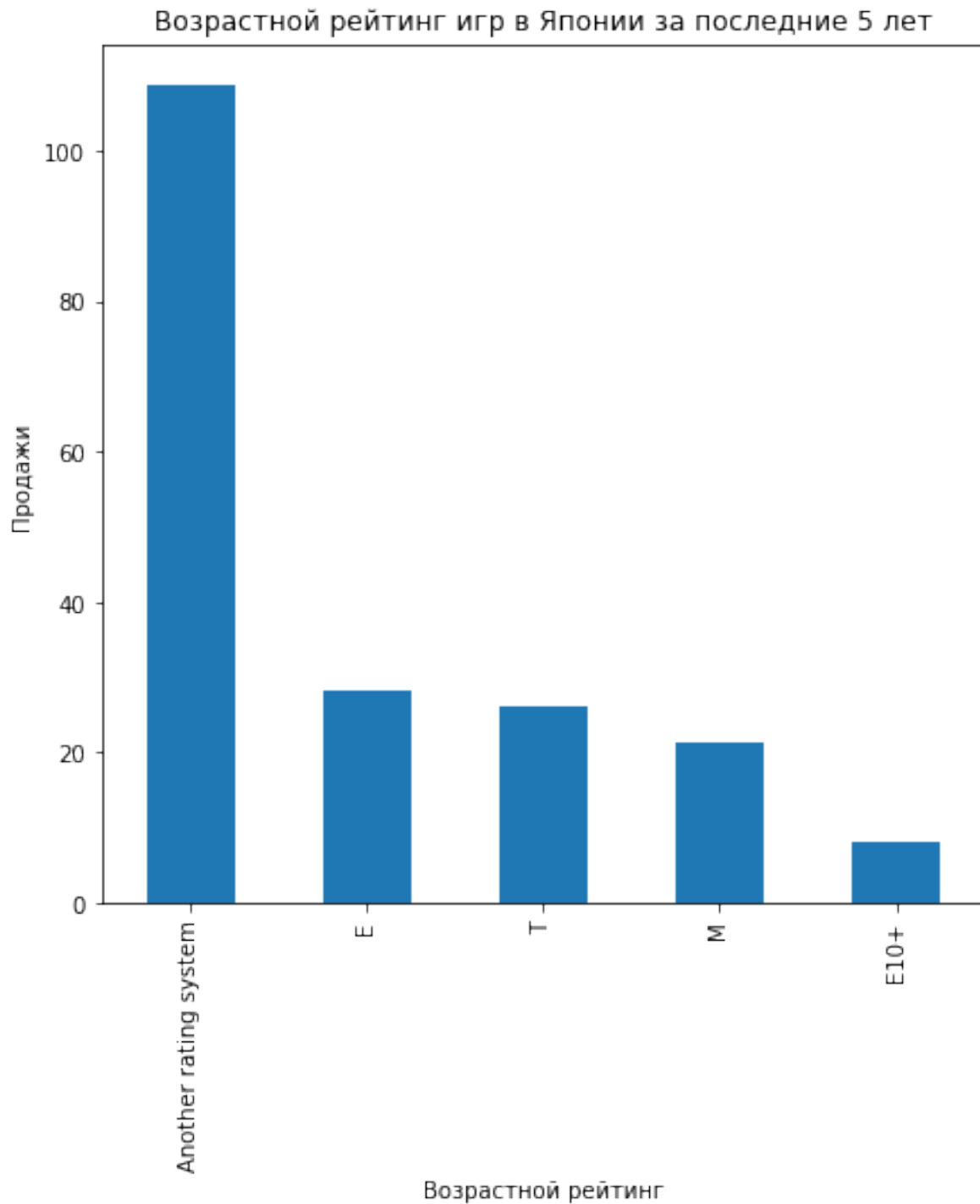
```
esrb_sales['na_sales'].sort_values(ascending=False).plot.bar(figsize=(7,7))  
plt.title("Возрастной рейтинг игр в Америке за последние 5 лет")
```

```
plt.xlabel("Возрастной рейтинг")  
plt.ylabel("Продажи");
```



В Северной Америке наблюдается схожая тенденция.

```
esrb_sales['jp_sales'].sort_values(ascending=False).plot.bar(figsize=(7,7))  
plt.title("Возрастной рейтинг игр в Японии за последние 5 лет")  
plt.xlabel("Возрастной рейтинг")  
plt.ylabel("Продажи");
```



Для оценки возрастного рейтинга игр, в Японии используется своя система рейтинга(CERO), однако анализируя популярность игр с европейской рейтинговой системой проданных на территории Японии можно сделать вывод, что в Японии популярность игр в зависимости от рейтинга отличается. Здесь наибольшей популярностью пользуются игры для всех возрастов и без ограничений E, а игры для взрослых занимают только 3 место.

В Европе и Америке, игры с категорией ESRB M (для взрослых) могут быть популярными из-за наличия широкого спектра игроков, среди которых взрослые и подростки, а также влияния западной культуры и традиций. В этих регионах ценится свобода зрелищ, стоит также вспомнить, где возникли гладиаторские бои.

С другой стороны, в Японии картина может быть иной. Японская культура значительно отличается от западной. Игры для всех возрастов (ESRB E) могут быть более популярными из-за разнообразной аудитории, включая семьи и детей, которые могут предпочитать более яркие, креативные и дружественные сюжеты а также из-за культурных особенностей.

Таким образом, различия в предпочтениях могут быть связаны как с составом аудитории, так и с культурными традициями каждой из рассматриваемых регионов.

```
# Комментарий ревьюера
temp = data.copy()
print(temp.rating.isna().sum(), temp.rating.isna().sum()/len(temp))
temp.rating.value_counts(dropna=False)

0 0.0

Another rating system    1275
M                        498
T                        412
E                        394
E10+                     306
Name: rating, dtype: int64
```

## Проверка гипотез

Для того чтобы узнать, какие игры пользователи ценят больше всего, необходимо провести анализ данных, используя статистические методы. Один из таких методов - это t-тест, который позволяет определить, есть ли статистически значимые различия между двумя группами данных.

Тест предполагает что математические ожидания обеих выборок равны. Параметр `alternative` по умолчанию установлен как `two-sided`, это означает, что альтернативная гипотеза является двусторонней и предполагает, что математические ожидания обеих выборок не равны. В данном тесте используется `p-value` или `p-критерий`, поскольку необходимо оценить вероятность получить наблюдаемые различия в средних пользовательских рейтингах предполагая, что нулевая гипотеза верна и таким образом проверить гипотезы.

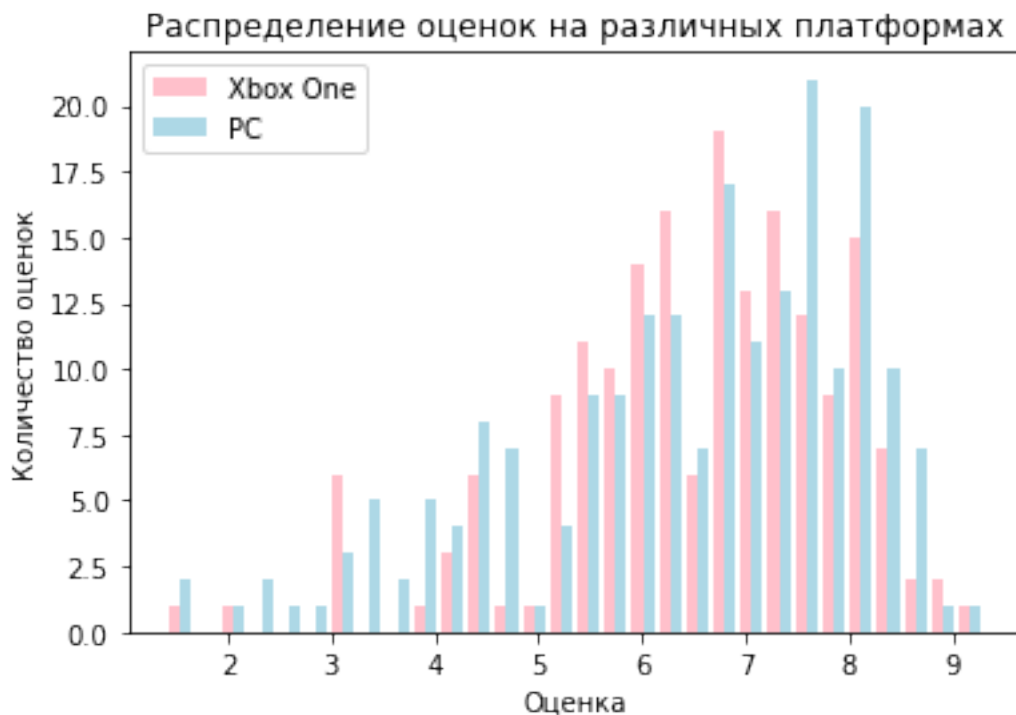


## Средние пользовательские рейтинги платформ Xbox One и PC одинаковые

**Нулевая гипотеза ( $H_0$ ):** Средние пользовательские рейтинги платформ Xbox One и PC одинаковые.

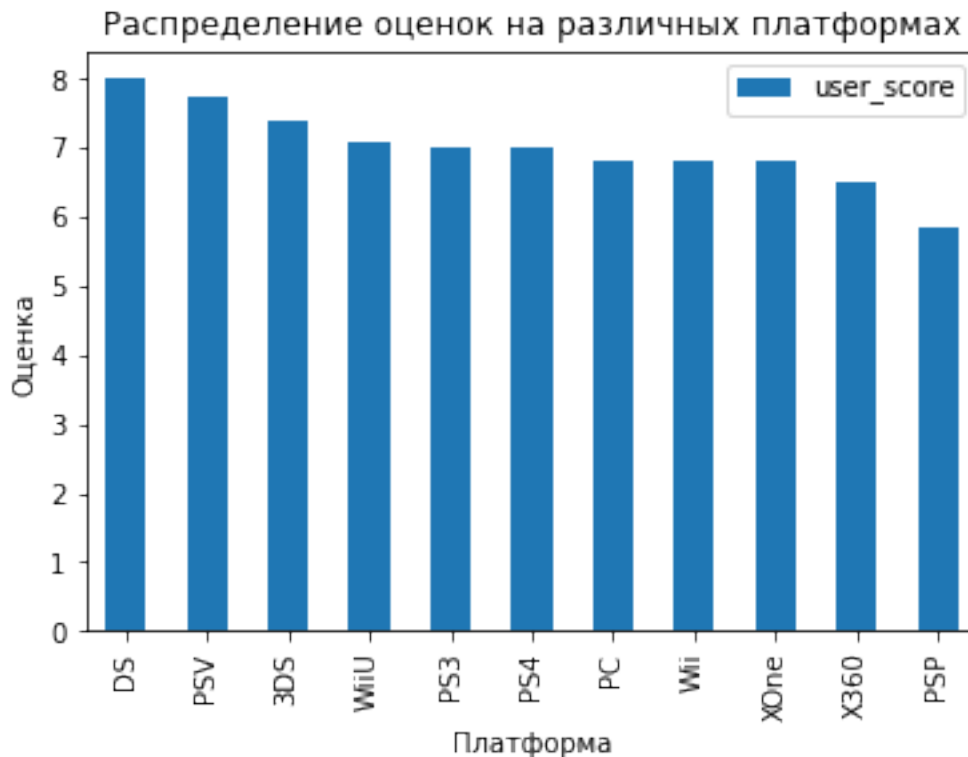
**Альтернативная гипотеза ( $H_1$ ):** Средние пользовательские рейтинги платформ Xbox One и PC различаются.

```
xo_score= data[data['platform']=='XOne']['user_score'].dropna() #  
Пользовательские оценки игр на платформе Xbox One  
pc_score= data[data['platform']=='PC']['user_score'].dropna() #  
Пользовательские оценки игр на ПК  
  
results = st.ttest_ind(xo_score, pc_score)  
results.pvalue  
  
0.5535080500643661  
  
plt.hist(x=[xo_score,pc_score],bins=30,color=['pink','lightblue'],label=[  
'Xbox One','PC'])  
plt.title("Распределение оценок на различных платформах")  
plt.xlabel("Оценка")  
plt.ylabel("Количество оценок")  
plt.legend();
```



Средние пользовательские рейтинги по платформам будут иметь следующее распределение:

```
data.pivot_table(index='platform',values='user_score',aggfunc='median')
.sort_values(by='user_score',ascending=False).plot.bar()
plt.title("Распределение оценок на различных платформах")
plt.xlabel("Платформа")
plt.ylabel("Оценка")
plt.legend();
```



Полученное значение p-value, выше общепринятого уровня значимости  $\alpha=5\%$ , поэтому мы не можем отвергнуть нулевую гипотезу. Скорее всего между пользовательскими оценками игр на ПК и Xbox One нет существенных различий.

Средние пользовательские рейтинги жанров Action и Sports разные

**Нулевая гипотеза ( $H_0$ ):** Средние пользовательские рейтинги жанров Action и Sports одинаковые.

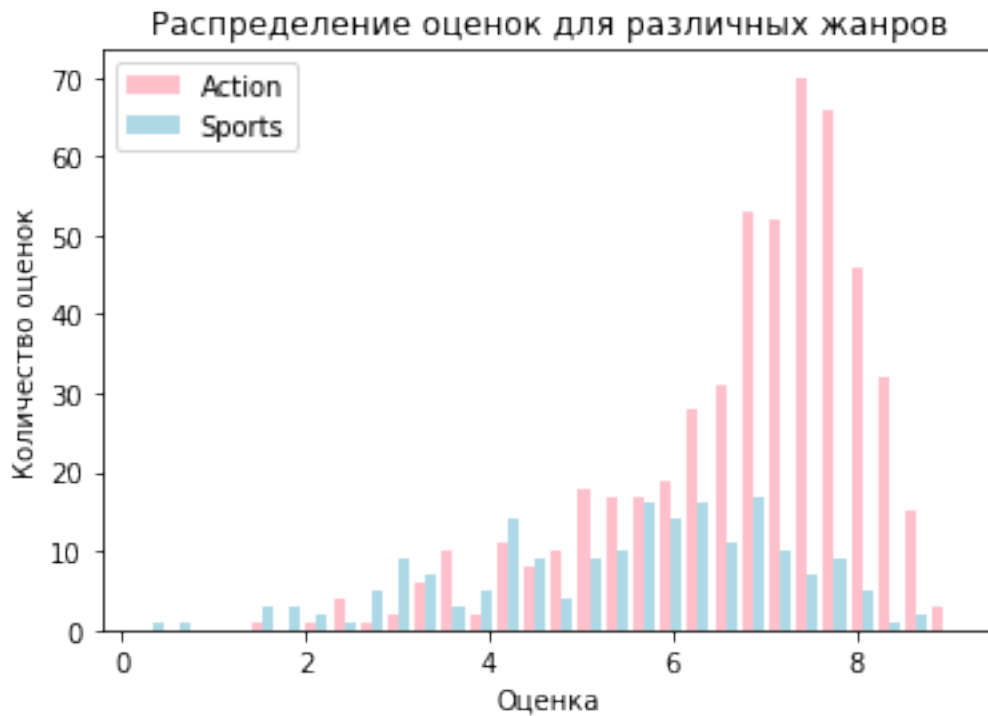
**Альтернативная гипотеза ( $H_1$ ):** Средние пользовательские рейтинги жанров Action и Sports различаются.

```
action_score= data[data['genre']=='Action']['user_score'].dropna() #
Пользовательские оценки игр жанра Action
sports_score= data[data['genre']=='Sports']['user_score'].dropna() #
Пользовательские оценки игр жанра Sports

results = st.ttest_ind(action_score, sports_score)
results.pvalue
```

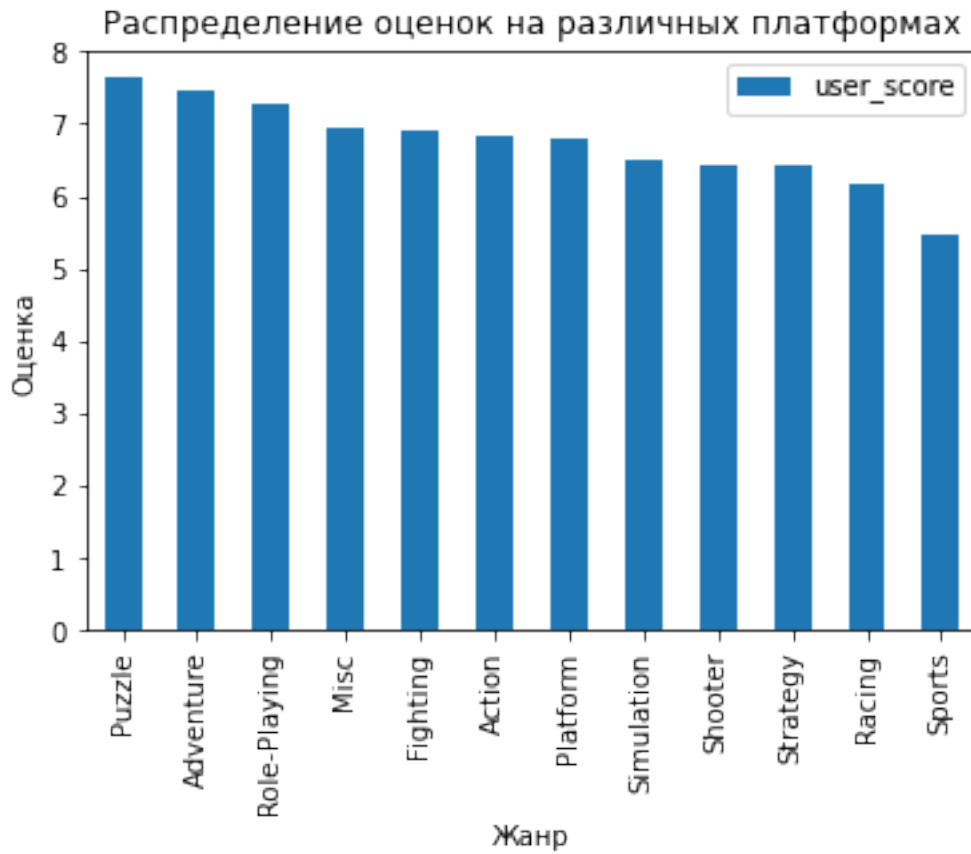
4.8045233804629176e-26

```
plt.hist(x=[action_score,sports_score],bins=30,color=['pink','lightblue'],label=['Action','Sports'])
plt.title("Распределение оценок для различных жанров")
plt.xlabel("Оценка")
plt.ylabel("Количество оценок")
plt.legend();
```



Средние пользовательские рейтинги по жанрам будут иметь следующее распределение:

```
data.pivot_table(index='genre',values='user_score',aggfunc='mean').sort_values(by='user_score',ascending=False).plot.bar()
plt.title("Распределение оценок на различных платформах")
plt.xlabel("Жанр")
plt.ylabel("Оценка")
plt.legend();
```



p-value существенно меньше общепринятого уровня статистической значимости  $\alpha=5\%$ , поэтому мы можем отвергнуть нулевую гипотезу и принять альтернативную. Есть основания полагать, что средние оценки жанров Action и Sports могут существенно различаться.

## Вывод

Исходный датафрейм для анализа представляет собой данные о продажах игр в различных регионах. В таблице содержится информация о 16713 играх, в большинстве случаев присутствуют данные о жанре, годе выпуска, игровой платформе и возрастном рейтинге. Чуть меньше чем для половины игр также имеются оценки критиков и пользователей.

Исходные данные были обработаны: названия столбцов были приведены к змеиному регистру; значение tbd было заменено на pr.nap, что позволило изменить тип столбца с оценками пользователей с object на float; явных дубликатов в данных обнаружено не было, однако был обнаружен 1 неявный дубликат который был удален; игры с пропусками вместо названия (всего 2 штуки) были удалены, а пропуски в столбце с возрастным рейтингом были заменены значением 'Another rating system'; встреченная аномалия связана несоответствующим годом выпуска игры была обработана; добавлен столбец с суммарными продажами по все регионам.

Количество выпускаемых каждый год увеличивалось вплоть до 2010 года, примерно в этот момент крупные компании выпускающие игры изменили свою политику и вместо

множества разнообразных игр, решили сосредоточиться на крупных проектах, которые в потенциале могли бы принести прибыль большую чем суммарные продажи более мелких проектов. Однако в целом динамика по количеству выпускаемых игр положительная, не только крупные компании продолжают осваивать рынок видеоигр, также и мелкие группы разработчиков объединяются для создания новых продуктов.

Динамика продаж игр на различных платформах показывает, что с момента анонса или выхода платформы следующего поколения интерес разработчиков к текущей платформе начинает ослабевать. Средний срок жизни платформы составляет 7 лет, за это время платформу перестает поддерживать производитель. В данный момент наиболее перспективными платформами являются PlayStation 4 и Xbox One, которые судя по выходу обновленных версий, останутся актуальными еще несколько лет.

Многие платформы обладают флагманскими продуктами, раскрывающими заложенные разработчиками возможности. Такие игры имеют высокие продажи, однако как правило коммерциализировать их маловероятно, так как большинство таких игр выпускаются компанией разработчиком платформы напрямую. Помимо флагманских игр, заметно выделяется несколько десятков хитов, причем для каждой платформы хиты могут быть различными. Продажи таких игр могут превосходить продажи среднепопулярных игр в несколько сотен раз и приносят огромную прибыль.

Отзывы критиков могут ощутимо влиять на продажи игр на тех или иных платформах, однако зависимость между средней оценкой критиков и продажами имеет сложную форму и не всегда может оказывать желаемое воздействие на рост продаж. Для построения грамотного взаимодействия с игровой прессой необходимо провести более глубокий анализ.

Оценки пользователей также могут оказывать влияние на продажи игр, хотя это влияние может быть менее прямым, чем оценки критиков. Некоторые потребители могут принимать решение о покупке, ориентируясь на мнения других игроков, высказанные на площадках по продаже игр, специальных сайтах с рецензиями, или в социальных сетях. Однако влияние оценок пользователей неоднозначно и может зависеть от различных факторов, таких как активность сообщества игроков, наличие скандальных ситуаций с игрой, репутация разработчика и другие. Поэтому для понимания реального воздействия оценок пользователей на продажи также необходимо провести более глубокий анализ и учитывать различные контексты.

Правильно понимать целевую аудиторию и рынок игр - это критически важно для успешной маркетинговой программы. Аналитика данных позволяет понять потребности и предпочтения целевой аудитории, что оказывает важное влияние на выбор платформы и рекламных стратегий. Учитывая растущее количество игр и изменения в платформах, анализ данных о продажах и трендах может помочь определить наиболее перспективные платформы для маркетинговых усилий.

Очень важную роль в составлении маркетинговой компании играет региональная направленность рекламы. Пользователи различных регионов обладают различной культурой гейминга, и если в Европе и Америке популярны стационарные платформы, то в Японии отдают предпочтение портативным консолям. Наиболее популярными платформами в Европе и Америке являются платформы двух компаний: Sony и Microsoft. В данный момент наиболее актуальными версиями платформ являются PlayStation 4 Pro и Xbox One S вышедшие в 2016 году. В Японии же лидерами рынка являются компания

Nintendo с их портативными консолями и компания Sony выпускающая как портативные так и стационарные девайсы, игровые приставки компании Microsoft на территории Японии не пользуются популярностью.

Различие в жанровых предпочтениях тесно связано с популярными в том или ином регионе платформами и должно быть учтено маркетологами. Из-за предпочтения портативных консолей, обусловленного культурными различиями, самым популярным жанром видеоигр в Японии является JRPG, это жанр предназначен специально для портативного гейминга и как правило ориентирован на внутренний рынок Японии. Жители Европы и Америки обладают похожей культурой, возможно поэтому различия в любимых жанрах у этих двух регионов минимальны. Жители обоих регионов предпочитают экшены, шутеры, спортивные и ролевые игры. Жанр файтинг в последние годы набирает популярность как в Японии так и в Северной Америке и может стать универсальным продуктом.

Возрастная группа людей играющих в видеоигры включает в себя людей всех возрастов, однако и здесь есть региональные различия. Игры категории М (для взрослых) пользуются популярностью в Америке и Европе, в то время как в Японии самой популярной является категория Е (для всех). Это может быть обусловлено культурными различиями, а также различием в составе аудитории и законодательстве.

Отзывы критиков и медийных персон также играют важную роль в формировании маркетинговой стратегии, поскольку они оказывают влияние на мнение потенциальных покупателей. Игры с низким рейтингом, как правило не достигают высоких продаж. Тут важно понимать относительность оценки игр на разных платформах. На разных платформах оценка игр может отличаться и если средняя игра на платформе PSP получает 7 то у DS это может быть 8, этот же принцип применим и к различным жанрам.

Также важно учитывать культурные особенности каждого региона в рекламной кампании. В Европе и Америке, где игровая культура более распространена, можно использовать более дерзкие и агрессивные методы рекламы, в то время как в Японии, где игры тесно связаны с японской анимацией и культурой, следует учитывать эти особенности и делать акцент на них в рекламе.

Подводя итог можно выделить следующие рекомендации для проведения рекламной кампании в различных регионах:

**Европа :** В Европе популярностью пользуются стационарные платформы, такие как PlayStation 4 и Xbox One. Эти платформы являются наиболее перспективными и в ближайшие годы не утратят своей актуальности. При продвижении игр в этом регионе стоит сосредоточиться на наиболее популярных жанрах а именно: экшены, шутеры ,спортивные и ролевые игры, а также гонки и платформеры (в порядке убывания популярности).

**Америка :** Американские игроки также предпочитают стационарные консоли, а жанр шутер является даже более популярным чем в Европе. Также как и европейские игроки, они отдают предпочтение экшенам, спортивным и ролевым играм, а также платформерам. Любовь к файтингам (хотя и меньшая чем у японцев) роднит их с японскими игроками.

**Япония :** Исключительно дела обстоят в Японии. Японские игроки в отличие от игроков других регионов предпочитают портативные консоли. Платформы таких компаний как Sony и Nintendo пользуются в Японии огромной популярностью. Наиболее перспективной консолью является Nintendo 3DS, однако с выпуском нового поколения портативных

консолей все может измениться. Предпочтение портативных консолей стационарным накладывает определенный жанровый отпечаток. Наибольшей популярностью пользуется жанр JRPG, игры этого жанра как правило нацелены на внутренний рынок японских портативных консолей. Вторым по популярности жанром является универсальный жанр экшен. Следующие жанры имеют равную долю рынка: симуляторы, файтинги, шутеры, платформеры, приключения, спорт.

Независимо от региона стоит учесть тот факт, что некоторые игры могут выходить за рамки жанровых или платформенных предпочтений пользователей обретая огромную популярность. Трудно предсказать какая игра станет хитом заранее, однако сделав ставку в нужный момент можно получить огромную прибыль.