ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ

УЧРЕЖДЕНИЕ ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ

«МОСКОВСКИЙ АВИАЦИОННЫЙ ИНСТИТУТ

(национальный исследовательский университет)» (МАИ)

Направление подготовки: 27.03.05 «Инноватика»

**Лабораторная работа №7**

по дисциплине «Прикладные методы обработки и защиты данных»

Выполнил:

Студенты гр. М3О-236Б-22

Попов М.А

Кузьмин Д.Е

Преподаватель:

Александрова С.С.

Москва 2023

Оглавление

[Цель работы: 3](#_Toc153668186)

[Раздел 1. Изучение информации о данных. 4](#_Toc153668187)

[Раздел 2. Подготовка данных. 5](#_Toc153668188)

[Раздел 3. Исследовательский анализ данных. 6](#_Toc153668189)

[Раздел 4. Составление портрета пользователя региона. 7](#_Toc153668190)

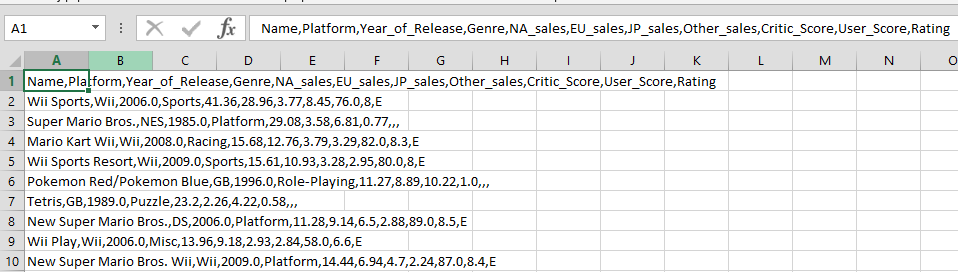
[Раздел 5. Вывод. 8](#_Toc153668191)

# Цель работы:

Проанализировать данные продажи игр для интернет-магазина «Стримчик» за период до 2016-го года для последующего построения рекламной компании и выявления перспективных направлений развития ассортимента. Построить графики, согласно заданию лабораторной работы.

# Раздел 1. Изучение информации о данных.

Данные в файле представлены следующим образом



# Подключим необходимые библиотеки для проведения анализа

**import** **pandas** **as** **pd**

**import** **numpy** **as** **np**

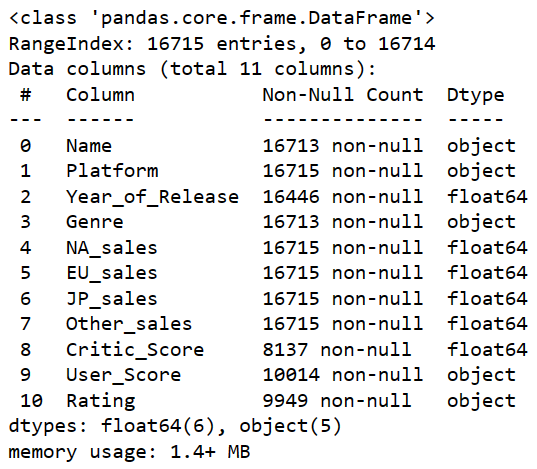
**import** **seaborn** **as** **sns**

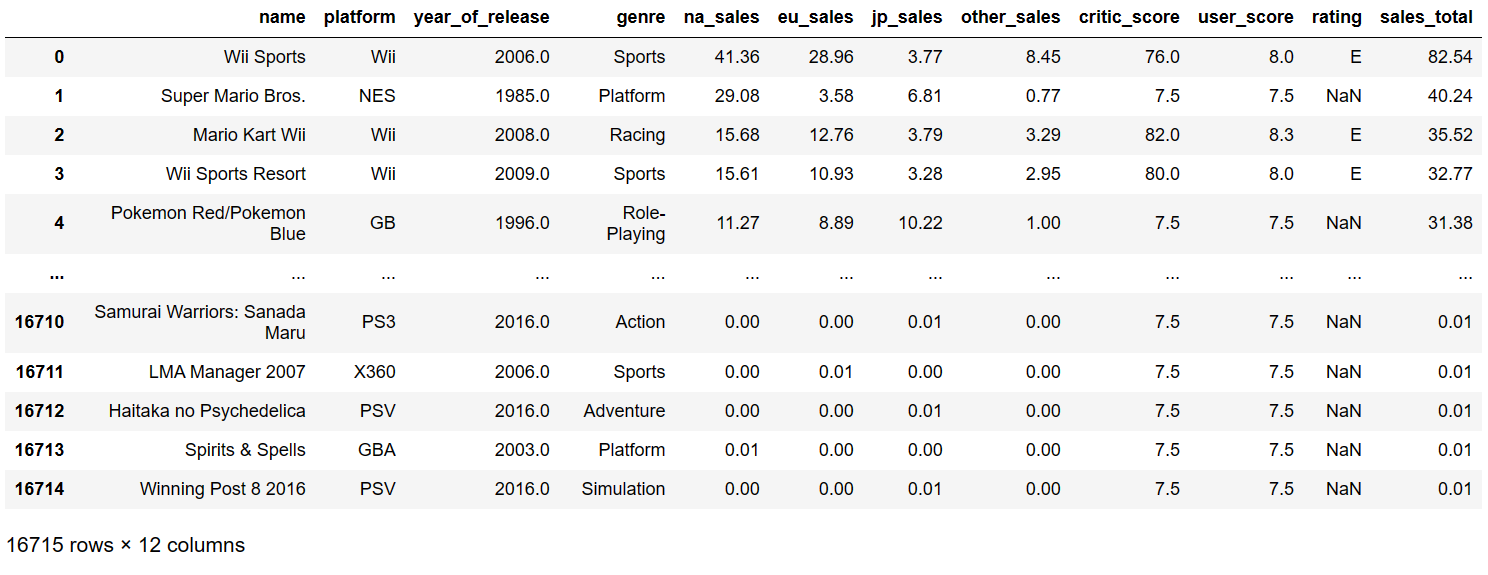
**from** **scipy** **import** stats **as** st

**import** **matplotlib.pyplot** **as** **plt**

df = pd.read\_csv('games.csv')

df.info()





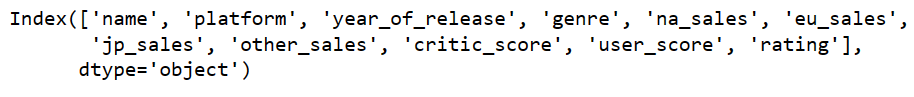
Ознакомившись с данными и обозначив библиотеки, которые могут нам пригодится для последующего анализа ДатаСета.

# Раздел 2. Подготовка данных.

Приведение к нижнему регистру названия столбцов.

df.columns = df.columns.str.lower()

df.columns



Преобразование и изменение типов данных в нужных для этого столбцах.

# Заменяем тип данных для столбца 'user\_score' - также заменяем там значение tbd, перед изменением типа данных

df['user\_score'] = df['user\_score'].replace('tbd', np.NaN)

df['user\_score'] = df['user\_score'].astype('float')

df.head(**10**)

Заменил значения в 'user\_score' и 'critic\_score' пустые значения на медианные по столбцу, чтобы не потерять значения. Для 'genre' на местах пропусков поставил "Не опеределен", чтобы было более понятно, с точки зрения пользователя.

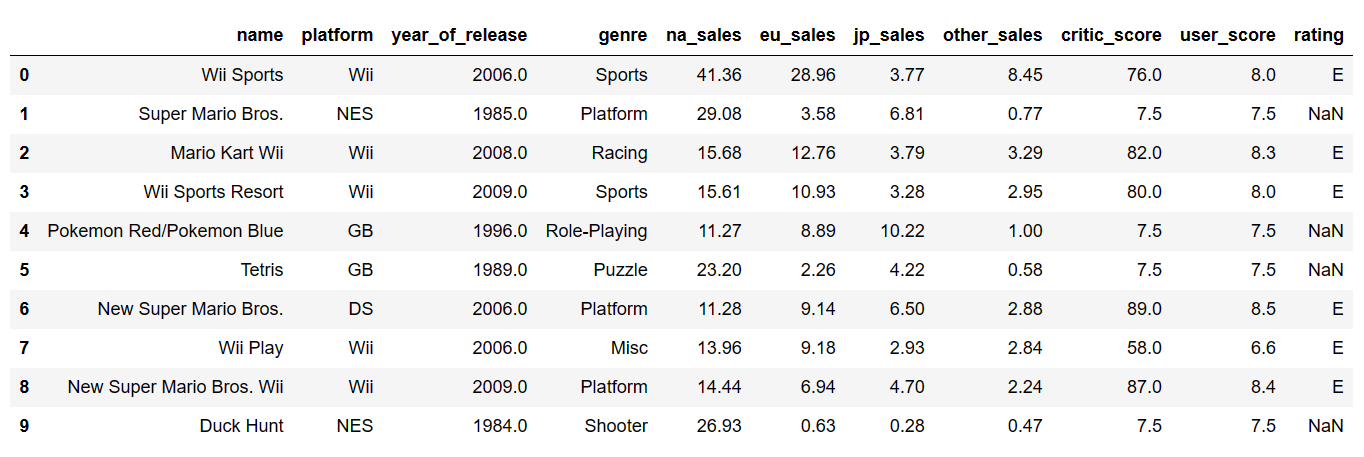
df['genre'] = df['genre'].fillna('undefinded')

median\_critic\_score = df['critic\_score'].median()

median\_critic\_score = df['user\_score'].median()

df['critic\_score'] = df['critic\_score'].fillna(median\_critic\_score)

df.head(**10**)

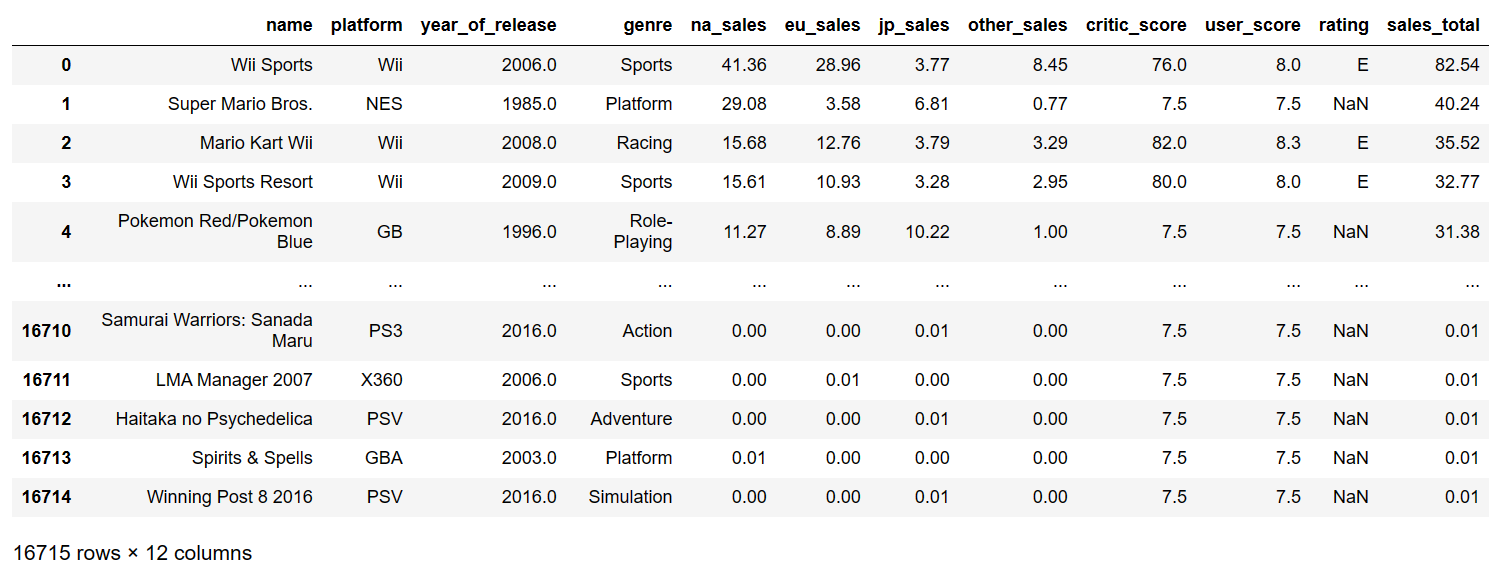


Подсчёт суммарных продаж по всем регионам

df['sales\_total'] = **0**

df['sales\_total'] = df[['na\_sales', 'eu\_sales', 'jp\_sales', 'other\_sales']].sum(axis=**1**)

df



# Раздел 3. Исследовательский анализ данных.

Составим график сопоставления количества проданных игр по годам.

sns.set(style="whitegrid")

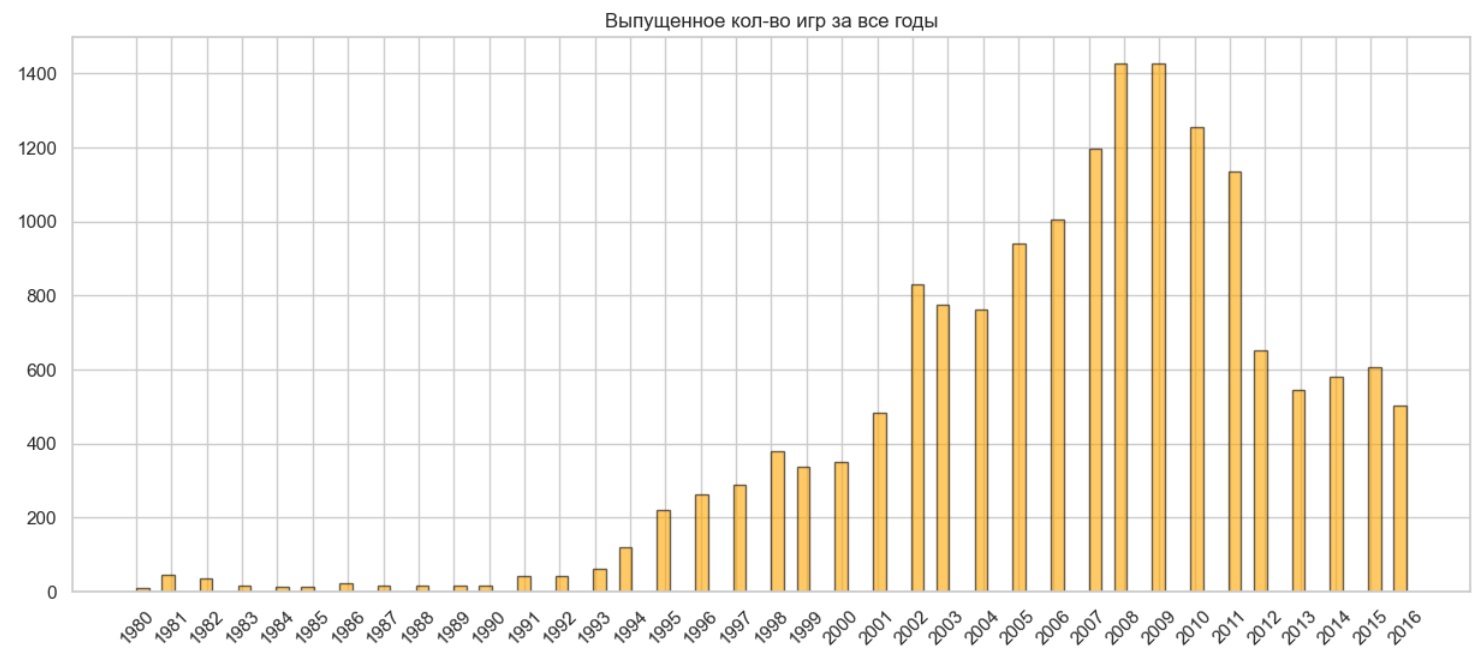
plt.figure(figsize=(**15**, **6**))

plt.hist(df['year\_of\_release'], bins=**100**, color='orange', alpha=**0.6**, align='mid', edgecolor='black', linewidth=**1**)

plt.title('Выпущенное кол-во игр за все годы')

plt.xticks(df['year\_of\_release'][pd.notna(df['year\_of\_release'])].unique(), rotation=**45**)

plt.show()



До 1994 года игры практически отсутствовали, особенно на фоне первого поколения процессоров Pentium. Однако, всё изменилось с появлением новых технологий и рынок игр начал развиваться бурными темпами. Вместе с этим, в последние годы можно заметить серьёзное уменьшение игр, представленных на рынке. Я предполагаю, что это связано с увеличением закрытости рынка, развитием больших проектов (AAA) и уменьшением количества нишевых, самодельных проектов. На разработку AAA проектов требуются большие денежные вливания и большое количество времени (по сравнению с нишевыми проектами).

Построение графика продаж по системе распространения игр, консоли, пк, прочие.

sorted\_df = df['platform'].value\_counts().reset\_index()

sorted\_df.columns = ['platform', 'count']

sorted\_df = sorted\_df.sort\_values(by='count', ascending=False)

sns.set(style="whitegrid")

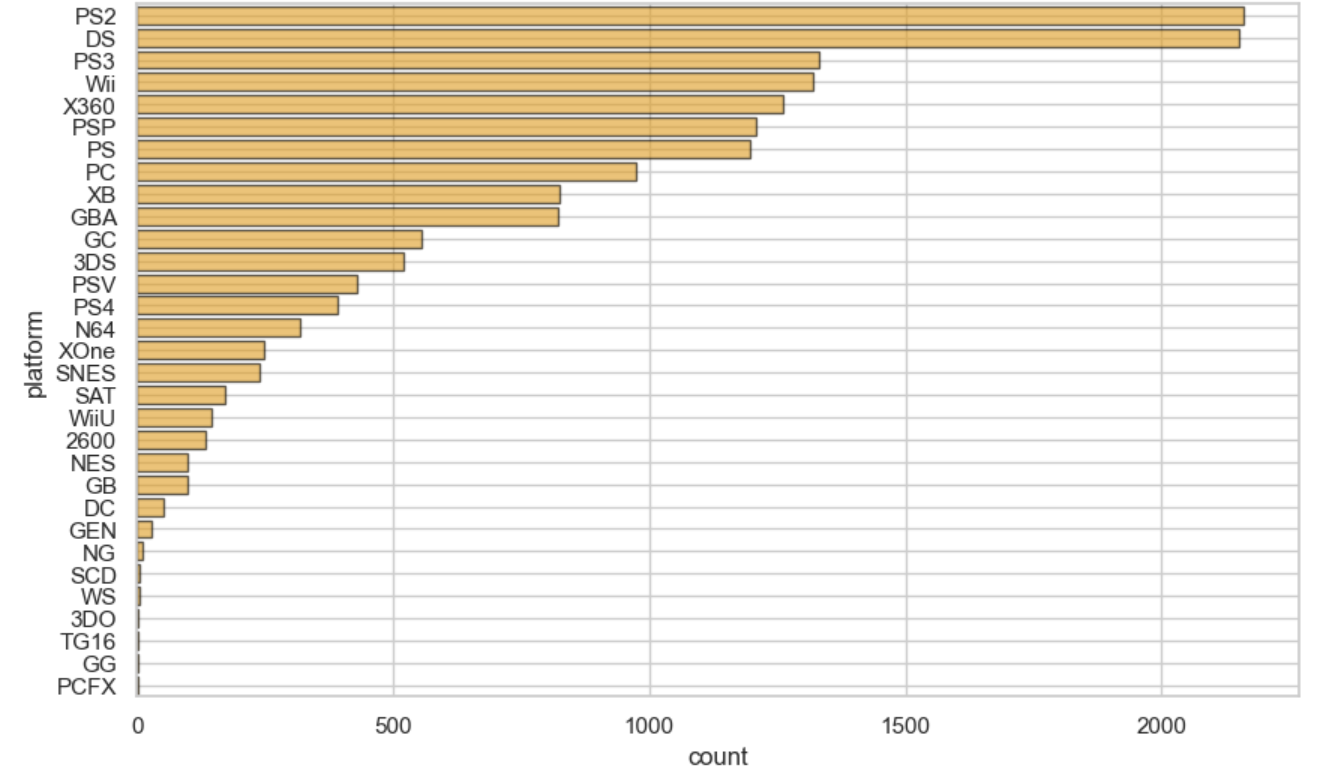
plt.figure(figsize=(**10**,**6**))

sns.barplot(x='count', y='platform', data=sorted\_df, color='orange', alpha=**0.6**, edgecolor='black', linewidth=**1**)

plt.grid ( True )

plt.show()

Из графика видно, что самые популярные игровые платформы за весь период это PS2, X360, PS3, WII, DS.



Из графика видно, что больше всего продаж игр приходится на технологию консолей, нежели на пользовательские ПК.

График «ящик с усами» динамики продаж консолей в виде

platform\_sales = df.pivot\_table(index=['platform'], values='sales\_total', aggfunc='sum').sort\_values('sales\_total', ascending=False)

platform\_sales = platform\_sales.reset\_index()

box\_total = np.zeros((**2**, len(platform\_sales)))

**for** i **in** range(len(platform\_sales)):

box\_total[**0**][i] = i

box\_total[**1**][i] = platform\_sales['sales\_total'].loc[i]

box\_total

plt.grid()

plt.boxplot(box\_total)

plt.xticks(rotation=**90**)

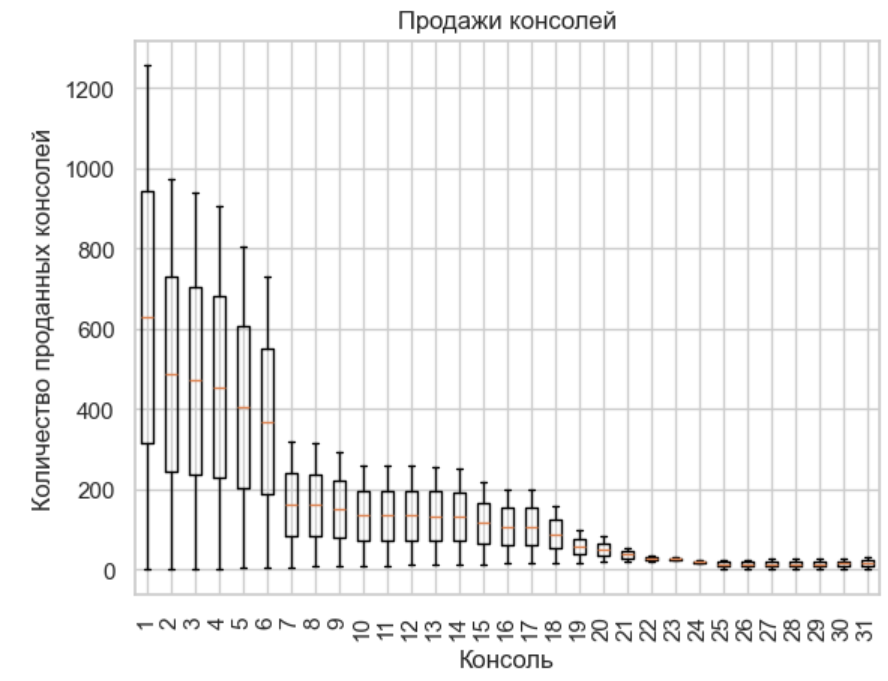
plt.title('Продажи консолей')

plt.xlabel('Консоль')

plt.ylabel('Количество проданных консолей')

plt.grid ( True )

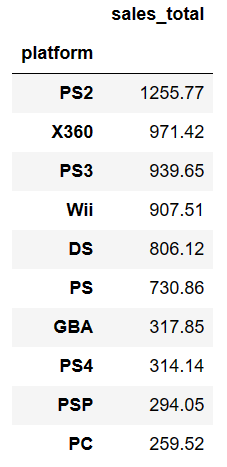
plt.show()



Продажи топ-10 консолей

df.groupby('platform')['sales\_total'].sum().to\_frame('sales\_total').sort\_values(by='sales\_total', ascending=False).head(**10**)

Из анализа данных понятно, что самая популяная платформа консолей - это Play Station, а самая популярная модель - это PS2. Вторая Консоль, которая занимает второе место - это XBox с моделью X360.



Самые прибыльные жанры по продажам игр.

df.groupby('genre')['sales\_total'].sum().to\_frame().sort\_values(by='sales\_total', ascending=False)

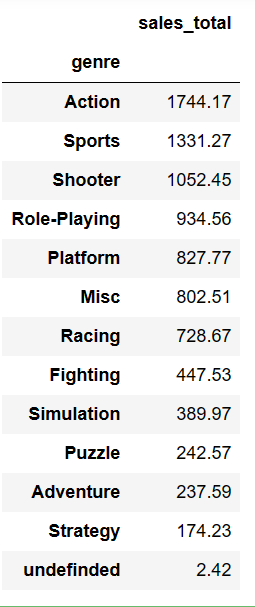


График количества игр по жанрам

game\_genre = df.pivot\_table(index=['genre'], values='name', aggfunc='count').sort\_values('name', ascending=False)

game\_genre = game\_genre.reset\_index()

plt.grid()

plt.barh(game\_genre['genre'], game\_genre['name'], color ='orange', alpha=**1**, edgecolor='black', linewidth=**1**)

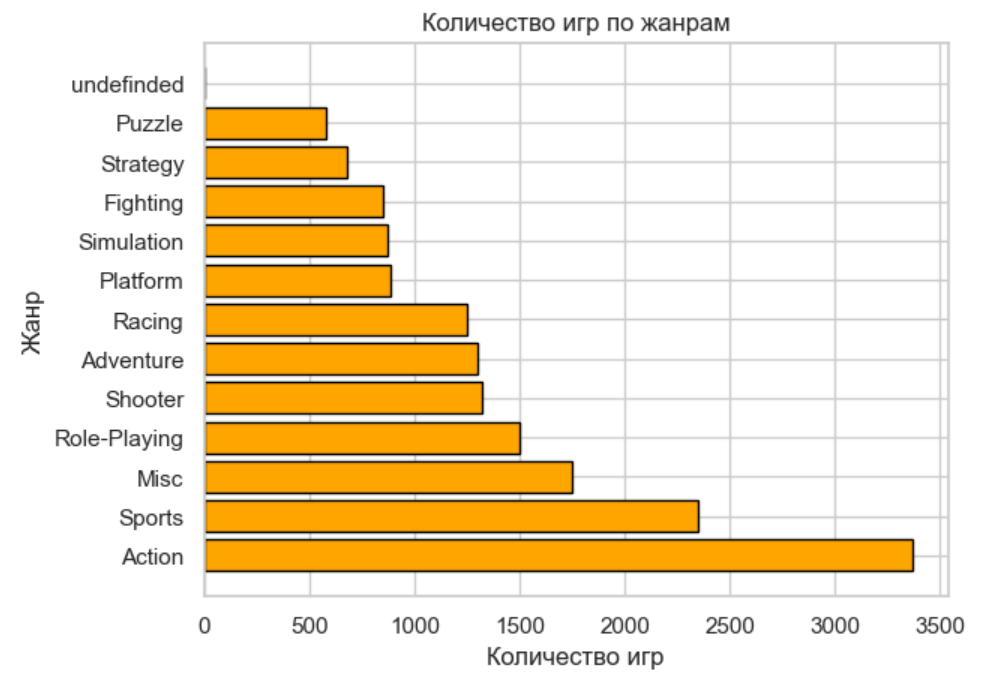
plt.title('Количество игр по жанрам')

plt.xlabel('Количество игр')

plt.ylabel('Жанр')

plt.grid ( True )

plt.show()



Исходя из графика, можно сделать вывод, что самый популярные жанры это Экшен, Спорт, Музыка, Ролеплей, Шутер.

Скорее всего, в ближайшее время эта тенденция продолжиться. Следовательно, можно увеличить количество игр, этих жанров, так как они популярны у игроков и пользователей компьютеров.

# Раздел 4. Составление портрета пользователя региона.

# сортируем игровые платформы по регионам Северная Америка, Европа, Япония

top\_platforms\_na = df.groupby('platform')['na\_sales'].sum().sort\_values(ascending=False).head(**5**)

top\_platforms\_eu = df.groupby('platform')['eu\_sales'].sum().sort\_values(ascending=False).head(**5**)

top\_platforms\_jp = df.groupby('platform')['jp\_sales'].sum().sort\_values(ascending=False).head(**5**)

# сортируем игровые жанры по регионам Северная Америка, Европа, Япония

top\_genres\_na = df.groupby('genre')['na\_sales'].sum().sort\_values(ascending=False).head(**5**)

top\_genres\_eu = df.groupby('genre')['eu\_sales'].sum().sort\_values(ascending=False).head(**5**)

top\_genres\_jp = df.groupby('genre')['jp\_sales'].sum().sort\_values(ascending=False).head(**5**)

# сортируем продажи по регионам Северная Америка, Европа, Япония

sales\_by\_rating\_na = df.groupby('rating')['na\_sales'].sum()

sales\_by\_rating\_eu = df.groupby('rating')['eu\_sales'].sum()

sales\_by\_rating\_jp = df.groupby('rating')['jp\_sales'].sum()

# составление профиля пользователя, исходя из полученных выше данных

average\_user = {

'Среднестистический пользователь в Северной Америке': {

'Самые популярные платформы': top\_platforms\_na.index.tolist(),

'Самые популярные жанры': top\_genres\_na.index.tolist(),

},

'Среднестистический пользователь в Европе': {

'Самые популярные платформы': top\_platforms\_eu.index.tolist(),

'Самые популярные жанры': top\_genres\_eu.index.tolist(),

},

'Среднестистический пользователь в Японии': {

'Самые популярные платформы': top\_platforms\_jp.index.tolist(),

'Самые популярные жанры': top\_genres\_jp.index.tolist(),

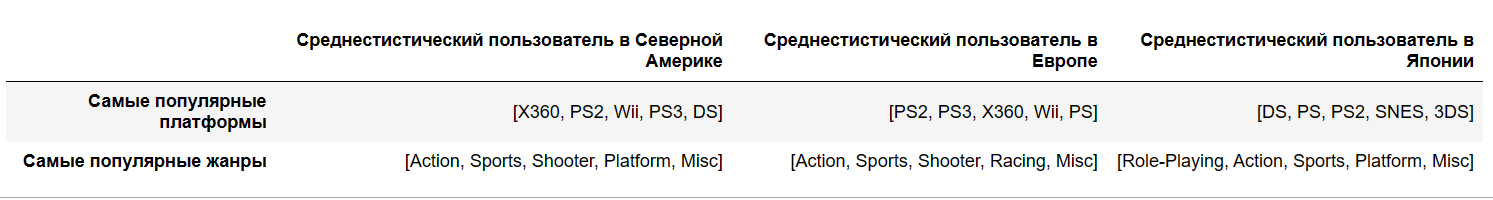
}

}

# формирования датасета-таблицы

df\_average = pd.DataFrame(average\_user)

df\_average

****

Исходя из полученных данных о пользователях по регионам и типам консолей и жанрам, которые в них популярны. Можно сделать вывод, что на территории региона Северная Америка игроки предпочитают: X360, PS2, Wii, PS3, DS, а их любимые жанры: Action, Sports, Shooter, Platform, Misc.

В тоже время, игроки из региона Европа выбирают: PS2, PS3, X360, Wii, PS, а среди жанров они придерживаются: Action, Sports, Shooter, Racing, Misc.

Игроки же в регионе Япония голосуют йеной за: DS, PS, PS2, SNES, 3DS, а жанры же: Role-Playing, Action, Sports, Platform, Misc.

# Раздел 5. Вывод.

В ходе выполнения работы был проанализирован файл с данными о продажах игр до 2016-го года. По итогам анализа можно сказать, что у рынка видеоигр присутствуют тенденции к дальнейшему развитию. Рынок десктопных игр уступает место технологическим новинкам, вызывая рост бюджетов. Большие компании гонятся за графикой, в то время как меньшие сталкиваются с трудностями. Изменение ценностей приводит к снижению интереса к глубоким играм, что влияет на общий спад рынка.