

Университет ИТМО

Практическая работа №2
по дисциплине «Визуализация и моделирование»

Автор: Хоанг Минь Тханг
Поток: ВИМ1.1
Группа: К33212
Факультет: ИКТ
Преподаватель: Чернышева А.В.

Санкт-Петербург, 2021 г.

1. Таблица с описанием хранящихся в датасете данных

Датасет слишком большой.

| № | Название столбца | Данные | Тип данных | Шкала |
|----|------------------|--|------------|---------------|
| 1 | Name | Название видеоигры | String | Номинальная |
| 2 | Platform | Предназначенная платформа | String | Номинальная |
| 3 | Year_of_Release | Год выпуска видеоигры | Float | Интервальная |
| 4 | Genre | Жанр видеоигры | String | Номинальная |
| 5 | Publisher | Издатель видеоигры | String | Номинальная |
| 6 | NA_Sales | Количество проданных единиц в Северной Америке (в миллионах) | Float | Относительная |
| 7 | EU_Sales | Количество проданных единиц в Европе (в миллионах) | Float | Относительная |
| 8 | JP_Sales | Количество проданных единиц в Японии (в миллионах) | Float | Относительная |
| 9 | Other_Sales | Количество проданных единиц в других регионах (в миллионах) | Float | Относительная |
| 10 | Global_Sales | Количество проданных единиц в мире (в миллионах) | Float | Относительная |
| 11 | Critic_Score | Оценка критиков | Float | Относительная |
| 12 | Critic_Count | Количество критиков | Float | Относительная |
| 13 | User_Score | Оценка пользователей | String | Относительная |
| 14 | User_Count | Количество пользователей | Float | Относительная |
| 15 | Developer | Разрабатывающая компания | String | Номинальная |
| 16 | Rating | Рейтинг | String | Номинальная |

2. Объекты визуализации

1. Быстрый обзор датасета

```
df = pd.read_csv('video_games_sales.csv')
df.head()
```

| | 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 |
|---|-----------------------------|----------|-----------------|------------------|-----------|----------|----------|----------|-------------|--------------|--------------|--------------|------------|------------|-----------|--------|
| 0 | Wii Sports | Wii | 2006.0 | Sports | Nintendo | 41.36 | 28.96 | 3.77 | 8.45 | 82.53 | 76.0 | 51.0 | 8 | 322.0 | Nintendo | E |
| 1 | Super Mario Bros. | NES | 1985.0 | Platform | Nintendo | 29.08 | 3.58 | 6.81 | 0.77 | 40.24 | NaN | NaN | NaN | NaN | NaN | NaN |
| 2 | Mario Kart Wii | Wii | 2008.0 | Racing | Nintendo | 15.68 | 12.76 | 3.79 | 3.29 | 35.52 | 82.0 | 73.0 | 8.3 | 709.0 | Nintendo | E |
| 3 | Wii Sports Resort | Wii | 2009.0 | Sports | Nintendo | 15.61 | 10.93 | 3.28 | 2.95 | 32.77 | 80.0 | 73.0 | 8 | 192.0 | Nintendo | E |
| 4 | Pokemon Red/Pokemon Blue | GB | 1996.0 | Role- Playing | Nintendo | 11.27 | 8.89 | 10.22 | 1.00 | 31.37 | NaN | NaN | NaN | NaN | NaN | NaN |

Получим типы данных, хранящихся в датафрейме.

```
Name          object
Platform       object
Year_of_Release float64
Genre          object
Publisher       object
NA_Sales       float64
EU_Sales       float64
JP_Sales       float64
Other_Sales    float64
Global_Sales   float64
Critic_Score   float64
Critic_Count   float64
User_Score     object
User_Count     float64
Developer      object
Rating         object
dtype: object
```

Получим размеры датафрейма

```
df.shape
```

```
(16719, 16)
```

```
df.size
```

```
267504
```

Выясним, есть ли в датафрейме пустые значения.

```
cols = list(df.columns)
cols
```

```
['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']
```

```
df_na = {col: list(pd.isna(df[col])).count(True) for col in cols}
df_na
```

```
{'Critic_Count': 8582,
 'Critic_Score': 8582,
 'Developer': 6623,
 'EU_Sales': 0,
 'Genre': 2,
 'Global_Sales': 0,
 'JP_Sales': 0,
 'NA_Sales': 0,
 'Name': 2,
 'Other_Sales': 0,
 'Platform': 0,
 'Publisher': 54,
 'Rating': 6769,
 'User_Count': 9129,
 'User_Score': 6704,
 'Year_of_Release': 269}
```

Пустые ячейки есть, мы удалим соответствующие строки

```
df = df.dropna(axis = 0)
```

```
df_na = {col: list(pd.isna(df[col])).count(True) for col in cols}  
df_na
```

```
{'Critic_Count': 0,  
 'Critic_Score': 0,  
 'Developer': 0,  
 'EU_Sales': 0,  
 'Genre': 0,  
 'Global_Sales': 0,  
 'JP_Sales': 0,  
 'NA_Sales': 0,  
 'Name': 0,  
 'Other_Sales': 0,  
 'Platform': 0,  
 'Publisher': 0,  
 'Rating': 0,  
 'User_Count': 0,  
 'User_Score': 0,  
 'Year_of_Release': 0}
```

```
df.shape
```

```
(6825, 16)
```

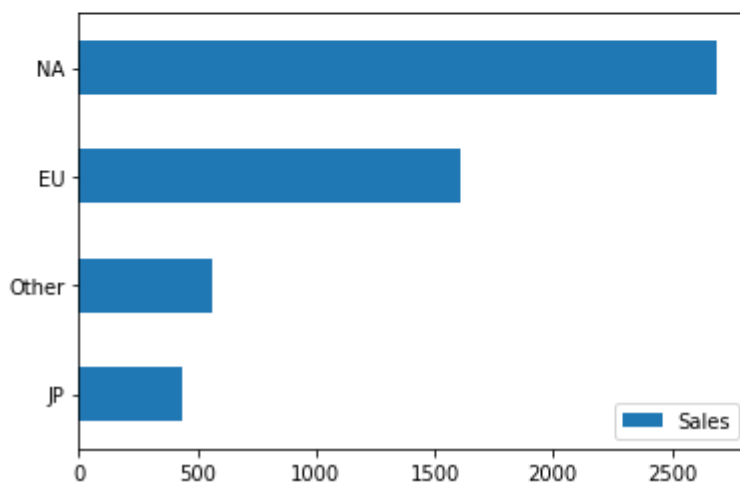
Были рассмотрено обобщение датасета: типы данных, размер датасета. Пустые ячейки существуют. Мы удалили соответствующие строки.

2. Продажа видеоигр по региону

```
NA = df["NA_Sales"].sum()
EU = df["EU_Sales"].sum()
JP = df["JP_Sales"].sum()
Other = df["Other_Sales"].sum()
Global = df["Global_Sales"].sum()
```

```
Region_Sales = pd.DataFrame(
    {"Sales": [Other, JP, EU, NA]},
    index=["Other", "JP", "EU", "NA"])
Region_Sales.sort_values(by='Sales', ascending=True).plot(kind="barh")
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f5710f6cdd0>

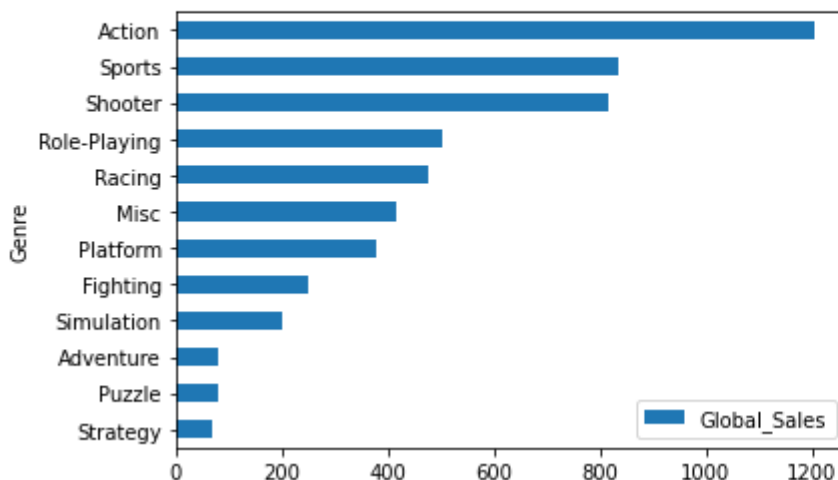


По графике, можем сравнить продажу видеоигр в разных регионах.

3. Продажа видеоигр по жанру

```
df.groupby(['Genre']).sum().sort_values(by='Global_Sales', ascending=True).plot(kind='barh')
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f935bee69d0>

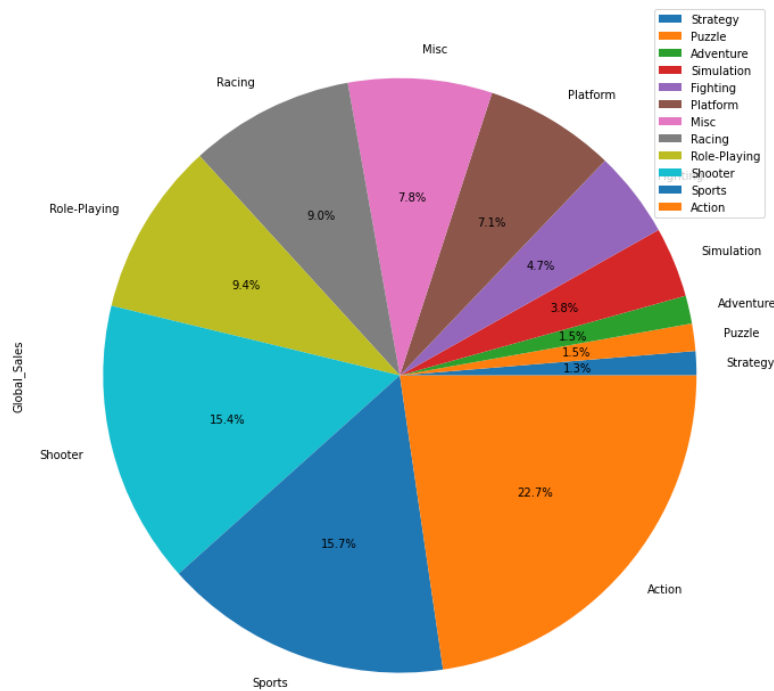


Action является самым популярным жанром видеоигры.

Для наглядности о пропорции каждого жанра построим круговую диаграмму.

```
df.groupby(['Genre']).sum().sort_values(by='Global_Sales', ascending=True).plot(kind='pie', y='Global_Sales', figsize=(15,12), autopct='%1.1f%%')
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f57058976d0>

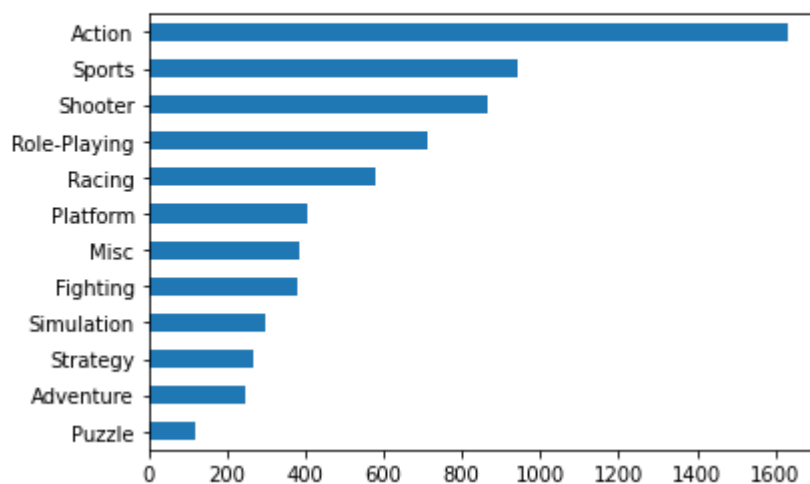


Только 3 из 12 жанров (Action, Sports, Shooter) составляют более 50% мировых продаж видеоигр.

4. Выпуск видеоигр по жанру

```
df["Genre"].value_counts().sort_values().plot(kind='barh')
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f57036e1ad0>

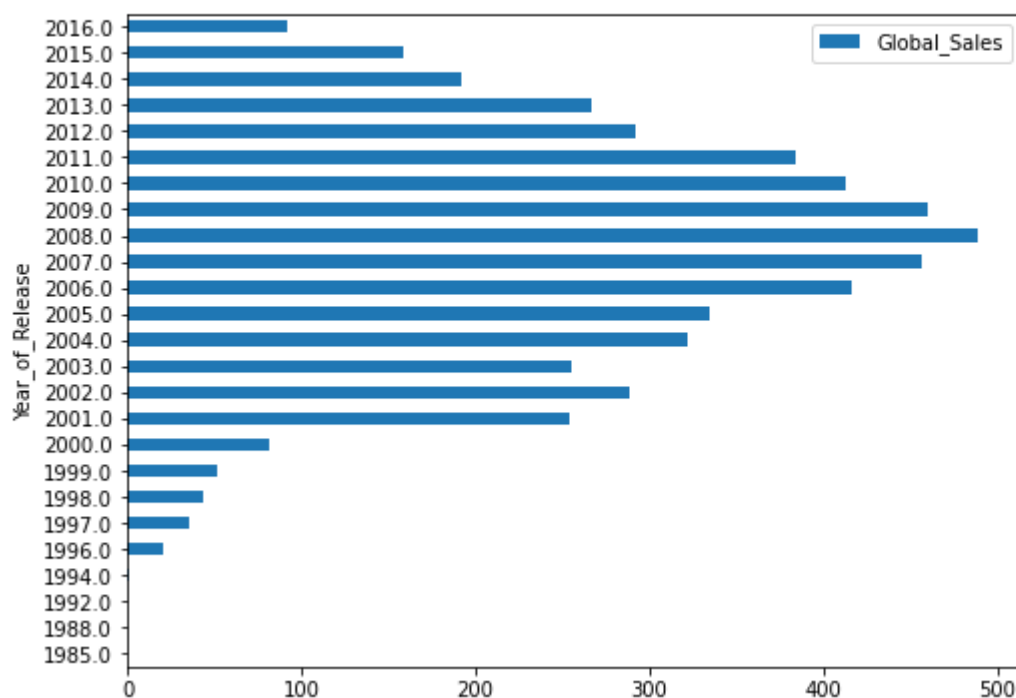


Action, Shooter, Sports - 3 жанра с наибольшим количеством игр.

5. Продажа видеоигр по году выпуска

```
df.groupby(['Year_of_Release']).sum().plot(kind='barh', y='Global_Sales', figsize=(8,6))
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f57039e8f10>
```



Видеоигры, выпущенные после 2000 года имеют резкий рост продаж. Однако, количество проданных копий видеоигр, выпущенных после 2008, уменьшается.

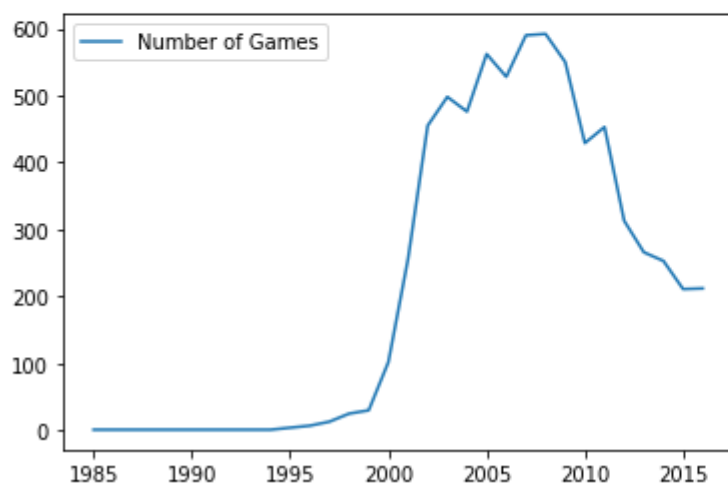
6. Выпуск видеоигр по году

```
year_data = {year: df["Year_of_Release"].to_list().count(year) for year in set(df["Year_
```

```
year_df = pd.DataFrame.from_dict(data=year_data, orient="index", columns=["Number of Gam
```

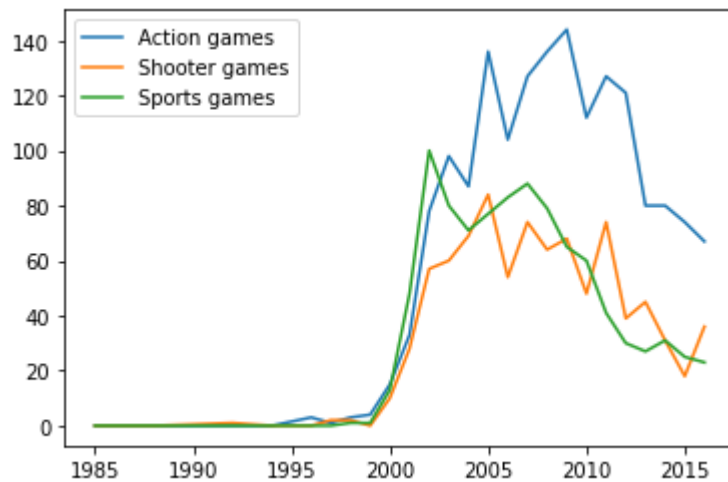
```
year_df.plot()
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f570384a4d0>



Количество видеоигр, выпущенных после 2000 года, резко увеличилось. Это соответствует графику продаж видеоигр по году выпуска.

7. Выпуск популярных жанров по году

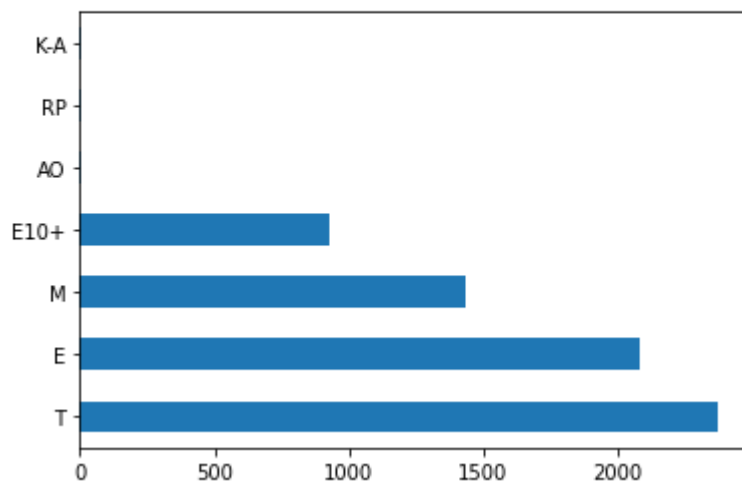


Количество выпущенных Action игр гораздо больше Shooter и Sports, особенно с 2005 по 2016.

8. Количество видеоигр по рейтингу

```
df["Rating"].value_counts().plot(kind='barh')
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f9359dba550>

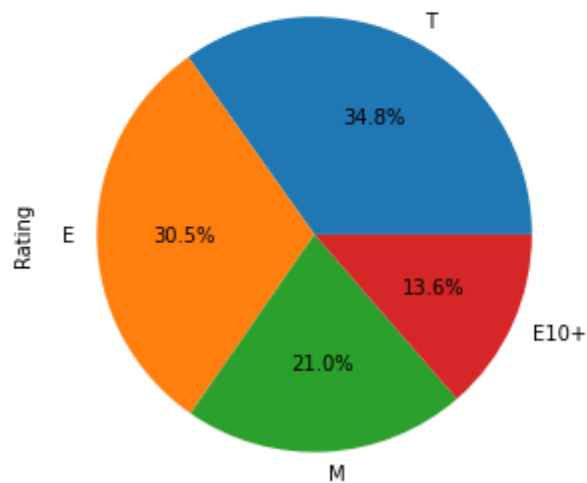


Количество видеоигр с рейтингами K-A, RP, AO очень мало и не стоит говорить об этом. Построим график без этих рейтингов.

```
sub_df = df.loc[(df["Rating"] == "E10+") | (df["Rating"] == "M") | (df["Rating"] == "E")]
```

```
sub_df["Rating"].value_counts().plot(kind='pie', autopct='%1.1f%%', figsize=(5,5))
```

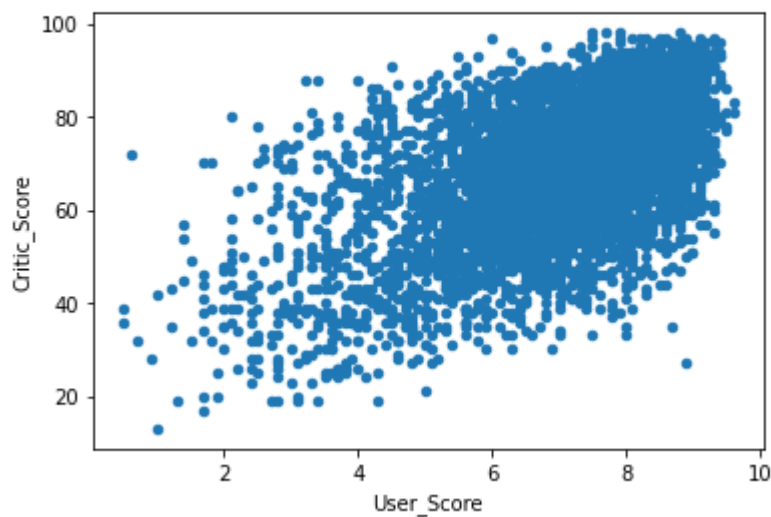
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f5703daa9d0>
```



9. Оценки видеоигр

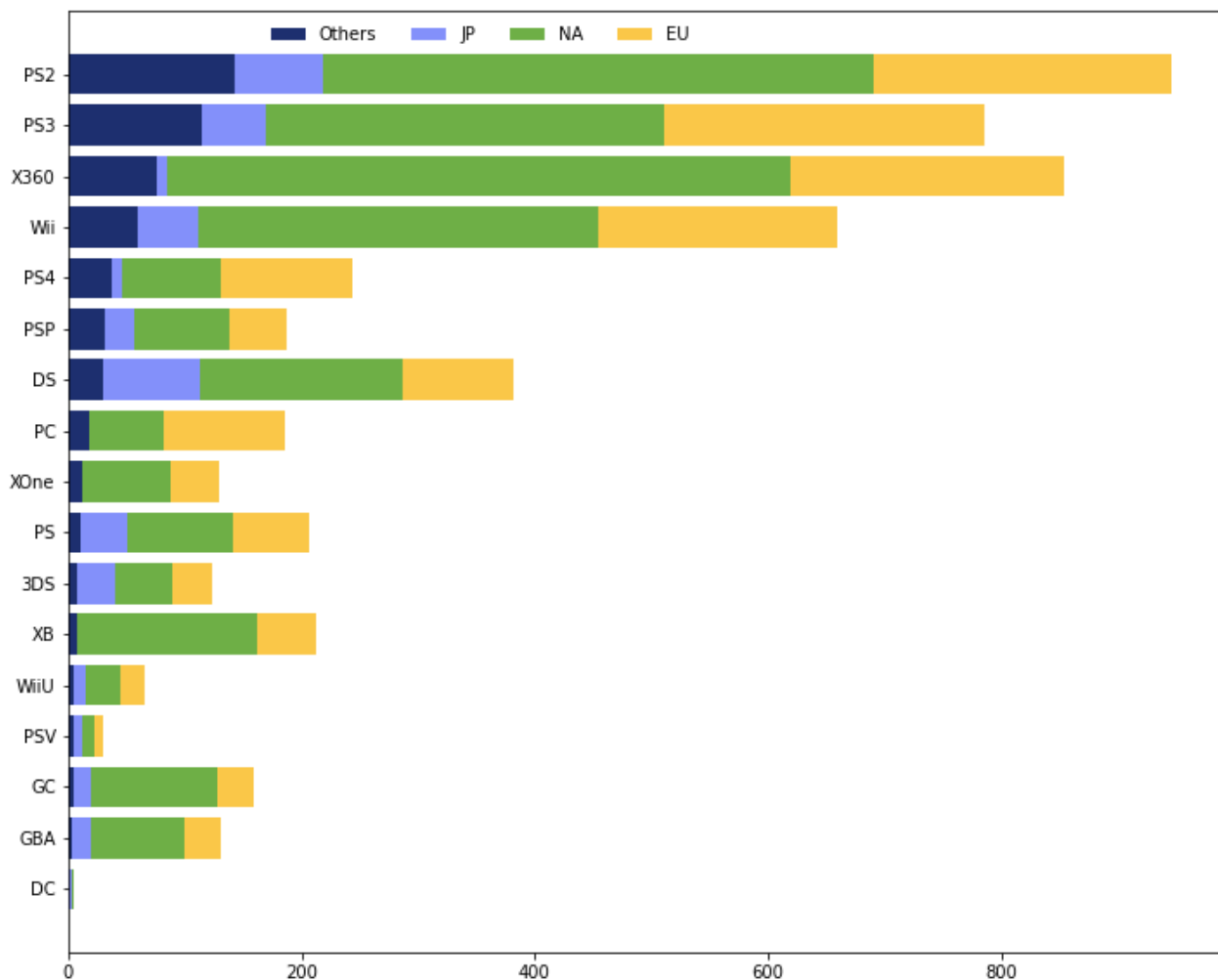
```
df.plot.scatter(x='User_Score', y='Critic_Score')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7a10cd8210>
```



По графикам видимо, что большинство точек находится в правом верхнем углу. Это означает, что в большинстве случаев, видеоигры с хорошей оценкой от критиков получают хорошую оценку от пользователей.

10. Продажи видеоигр по платформам и регионам



PS2-самая популярная платформа. В северной америке и европе принимаются все платформы. А в Японии люди почти не играют PC, XOne, XB игры.