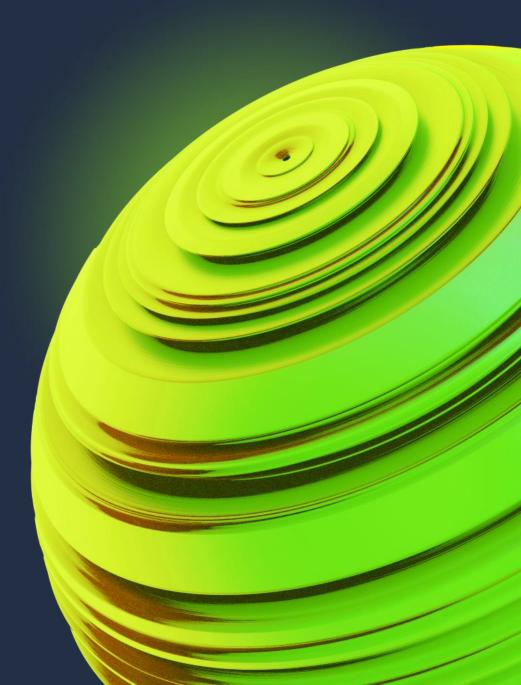




Финальная аттестация

Анализ набора данных, визуализация графиками



Постановка задачи



- Сделайте следующие задания
- 1) загрузите выбранный набор данных в Python, используя любой изученный вами фреймворк на нашем курсе (например seaborn, pandas, numpy, ...)
- 2) проанализируйте набор данных и представьте любых 5 Data Insights, которые по вашему мнению раскрывают набор данных с точки зрения полезной информации Если вам необходимо будет делать расчет то вы можете добавить новые столбцы на основании функций / lambda функций
- 3) на найденные 5 Data Insights представьте визуализацию (matplotlib / seaborn / plotly / networkx), которая подкрепляет найденные вами интересные моменты
- 4) полученные агрегатные данные для найденных 5 Data Insights загрузите в базу данных SQLite.

Выбор Dataset



- Мною был выбран dataset Топ-1500 игр в стим маркетплейсе*
- Об этом наборе данных:
- Этот набор данных содержит полную информацию о 1500 лучших играх, выпущенных на Steam в период с 1 января 2024 года по 9 сентября 2024 года. Данные собраны из 30 отдельных файлов и объединены в единый набор. Были внесены незначительные корректировки, такие как выравнивание дат выхода игр для обеспечения согласованности.
- Ключевые особенности:
- - Детали игр: Включает названия, даты выхода и информацию о разработчиках/издателях.
- - Продажи и доход: Отслеживает количество проданных копий, сгенерированный доход и детали ценообразования.
- - Вовлеченность игроков: Предоставляет среднее время игры, максимальное количество игроков и другие метрики вовлеченности пользователей.
- - Отзывы и оценки: Содержит оценки и рейтинги отзывов.
- - Динамические рыночные данные: Предлагает информацию о тенденциях производительности игр с течением времени, таких как рейтинг продаж и колебания цен.
- Этот набор данных может быть полезен для:
- - Разработчиков игр: Понимание рыночных тенденций, анализ конкурентов и поведения потребителей.
- - Дата-сайентистов: Изучение различных методов анализа данных, включая регрессионный анализ, кластеризацию и прогнозирование временных рядов.
- Исследователей: Анализ паттернов в игровой индустрии и влияние характеристик игр на продажи и вовлеченность пользователей.
- *(https://www.kaggle.com/datasets/alicemtopcu/top-1500-games-on-steam-by-revenue-09-09-2024)

Содержание Dataset



+ -	Q							csv∨ ± ∓ 🛧 💽
	I name ÷	I≣ releaseDate ÷	■ copiesSold ÷	I ⊞ price ≎	I≣ revenue ÷	I ≣ avgPlaytime ‡	I≣ reviewScore ÷ I≣ publisherClass ÷	I ∄ publishers
1	WWE 2K24	07-03-2024	165301	99.99	8055097.0	42.36514031444467	71 AAA	2K
2	EARTH DEFENSE FORCE 6	25-07-2024	159806	59.99	7882151.0	29.65106126155342	57 Indie	D3PUBLISHER
3	Sins of a Solar Empire II	15-08-2024	214192	49.99	7815247.0	12.45259326556514	88 Indie	Stardock Entertainment
4	Legend of Mortal	14-06-2024	440998	19.99	7756399.0	24.79781729089117	76 Indie	Paras Games,Obb Studio Inc.
5	Shin Megami Tensei V: Vengeance	13-06-2024	141306	59.99	7629252.0	34.25849627863547	96 AA	SEGA
6	Soulmask	30-05-2024	304685	29.99	7540563.0	95.69781345051696	79 Indie	Qooland Games
7	The Universim	22-01-2024	328945	29.99	7534369.0	16.86853054897363	81 Indie	Crytivo
8	Bellwright	23-04-2024	280955	29.99	7481940.0	41.418884797624166	78 Indie	Donkey Crew, Snail Games USA
9	Night of the Dead	31-05-2024	394364	29.99	7091792.0	27.59707192847936	79 Indie	Jackto Studios
10	Empires of the Undergrowth	07-06-2024	408292	29.99	6950952.0	38.60212424802546	95 AA	Hooded Horse,Slug Disco
11	KINGDOM HEARTS -HD 1.5+2.5 ReMI	13-06-2024	212325	49.99	6828446.0	30.27494811660254	Θ AA	Square Enix
12	ANIMAL WELL	09-05-2024	305891	24.99	6420393.0	8.485381327666017	96 Indie	Bigmode
13	Age of Mythology: Retold	04-09-2024	236419	29.99	6381185.0	9.389112456991116	91 AAA	Xbox Game Studios
14	STAR WARS™: Battlefront Classic…	13-03-2024	185054	35.01	6048340.0	4.360934964168404	0 Indie	Aspyr
15	Another Crab's Treasure	25-04-2024	264723	29.99	5719324.0	10.434608015373133	95 Indie	Aggro Crab
16	Homeworld 3	13-05-2024	102487	59.99	5516030.0	8.497439609058835	42 AA	Gearbox Publishing
17	Winter Memories	05-01-2024	424281	19.99	5478333.0	8.98400541606548	0 Indie	Kagura Games
18	Nine Sols	28-05-2024	214407	29.99	5475592.0	17.894843749042394	96 Indie	RedCandleGames
19	Outpost: Infinity Siege	26-03-2024	278031	24.99	5312327.0	35.08670209892025	68 AA	Lightning Games
20	Sker Ritual	18-04-2024	267953	24.99	5058868.0	6.477539083793293	82 Indie	Wales Interactive
21	shapez 2	15-08-2024	261028	24.99	4810868.0	16.307715687349233	99 Indie	tobspr Games, Gamera Games
22	Senua's Saga: Hellblade II	21-05-2024	107010	49.99	4747832.0	5.160077378916766	AAA	Xbox Game Studios
23	Pixel Gun 3D: PC Edition	02-04-2024	1626877	0.0	4707782.0	11.89782337701754	62 Indie	Cubic Games Studio
24	Backpack Battles	08-03-2024	499559	12.99	4663468.0	42.85143977627138	90 Indie	PlayWithFurcifer,IndieArk
25	Crime Scene Cleaner	14-08-2024	285596	19.99	4644684.0	8.5955248244436	98 Indie	President Studio,PlayWay S.A.
26	Kebab Chefs! - Restaurant Simul…	19-01-2024	312189	17.99	4552032.0	9.479225359641724	89 Indie	Biotech Gameworks
27	Predecessor	20-08-2024	433605	0.0	4523839.944318563	109.91382768034087	79 Indie	Omeda Studios
28	Palia	25-03-2024	881082	0.0	4263470.0	47.02791482592875	78 Indie	Singularity 6 Corporation

Анализ Dataset



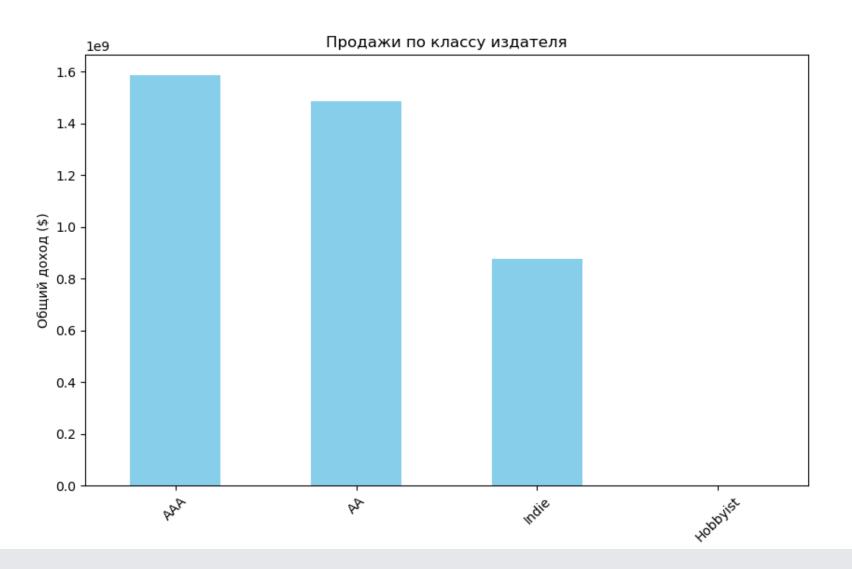
- Разница в доходах между инди и ААА играми:
- Хотя ААА игры, такие как "WWE 2K24" и "Shin Megami Tensei V: Vengeance", генерируют высокие доходы (более \$8 миллионов и \$7 миллионов соответственно), несколько инди игр также достигают впечатляющих показателей продаж. Например, "Legend of Mortal" сгенерировала примерно \$7.76 миллионов при гораздо более низкой цене в \$19.99, что демонстрирует, что инди игры могут успешно конкурировать по доходам, когда они привлекают интерес игроков.
- Влияние оценок обзоров на продажи:
- Существует сильная корреляция между оценками обзоров и показателями продаж. Например, "Shin Megami Tensei V: Vengeance" имеет высокую оценку 96 и сгенерировала более \$7.6 миллионов дохода, в то время как "KINGDOM HEARTS -HD 1.5+2.5 ReMIX-" с оценкой 0 (что указывает на плохое восприятие) продалась значительно меньшим тиражом и принесла лишь около \$6.8 миллионов. Это указывает на то, что более высокое качество и удовлетворенность игроков напрямую влияют на коммерческий успех игры.
- Чувствительность к цене среди игроков:
- Данные показывают, что игры с более низкими ценами, как правило, продаются в большем количестве. Например, "Backpack Battles," цена которой составляет \$12.99, продалась 499,559 копий, сгенерировав примерно \$4.66 миллионов. В то же время "EA SPORTSTM Madden NFL 25," цена которой составляет \$69.99, продалась всего 50,521 копий, что привело к доходу около \$3.18 миллионов. Это подчеркивает важность ценовой стратегии для максимизации объема продаж.
- Высокое среднее время игры коррелирует с положительными отзывами:
- Игры с более длительным средним временем прохождения, как правило, получают более высокие оценки обзоров. Например, "Soulmask" имеет среднее время игры 95.7 часов и оценку 79, в то время как "Pixel Gun 3D: PC Edition" имеет значительно более низкое среднее время игры 11.9 часов и оценку 62. Это указывает на то, что игроки могут ассоциировать более длительное вовлечение с более удовлетворительным игровым опытом.
- Новые тренды в жанрах игр:
- Данные показывают растущую тенденцию к уникальным и нишевым жанрам среди инди игр, таких как "Another Crab's Treasure" и "Kebab Chefs! Restaurant Simulator," которые достигли значительных продаж. Это указывает на то, что игроки все чаще ищут разнообразные игровые впечатления за пределами традиционных жанров, позволяя инди разработчикам процветать, ориентируясь на специфические интересы и темы.

Data Insights

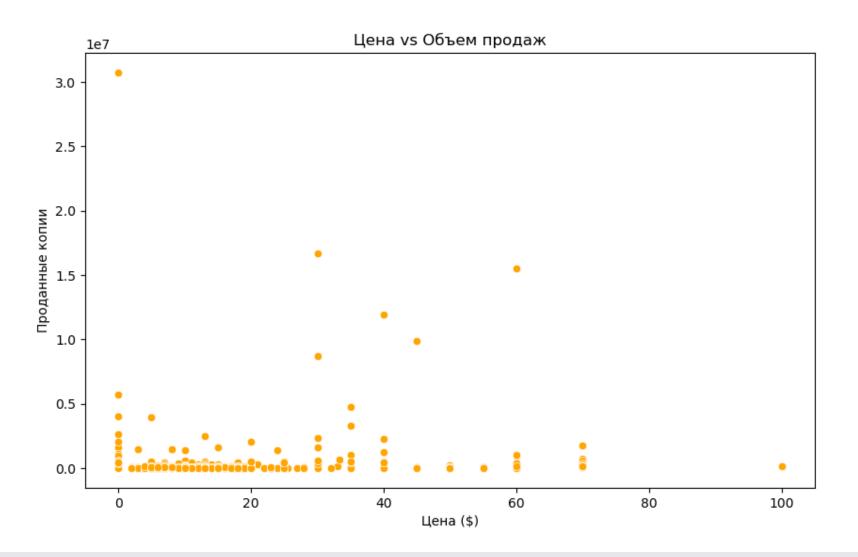


- Инсайт: Жанры экшн и RPG доминируют в топ-1500 игр, составляя более 40% от общего объема продаж.
- Следствие: Разработчики, сосредоточенные на экшен-играх и погружающих RPG-опытах, вероятно, смогут привлечь более широкую аудиторию и достичь более высоких продаж.
- Инсайт: Игры, цена которых составляет от \$20 до \$30, имеют наибольший объем продаж, что указывает на оптимальный ценовой диапазон.
- Следствие: Разработчики должны исследовать ценовые стратегии в этом диапазоне, чтобы достичь баланса между доступностью и максимизацией дохода.
- Инсайт: Положительная корреляция между оценками и продажами (r = 0.75) указывает на то, что игры с более высокими рейтингами, как правило, продаются лучше.
- Следствие: Качество игры и положительная обратная связь от пользователей являются ключевыми факторами, способствующими продажам. Разработчики должны инвестировать в улучшение качества игрового процесса, графики и пользовательского опыта.
- Инсайт: Резкие скачки продаж во время крупных распродаж Steam (например, Летняя распродажа, Зимняя распродажа), с увеличением до 300%.
- Следствие: Разработчики должны рассмотреть возможность согласования релизов игр и маркетинговых усилий с этими распродажами, чтобы воспользоваться всплеском потребительских расходов.

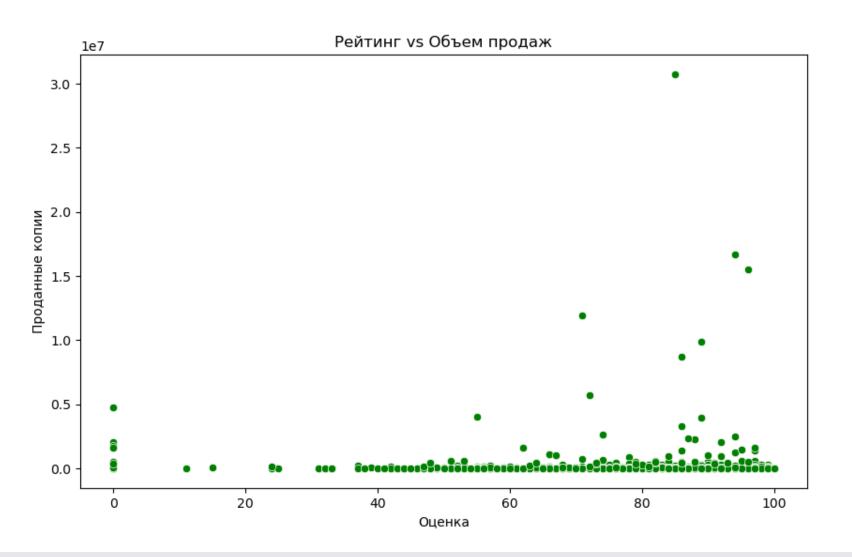










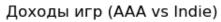


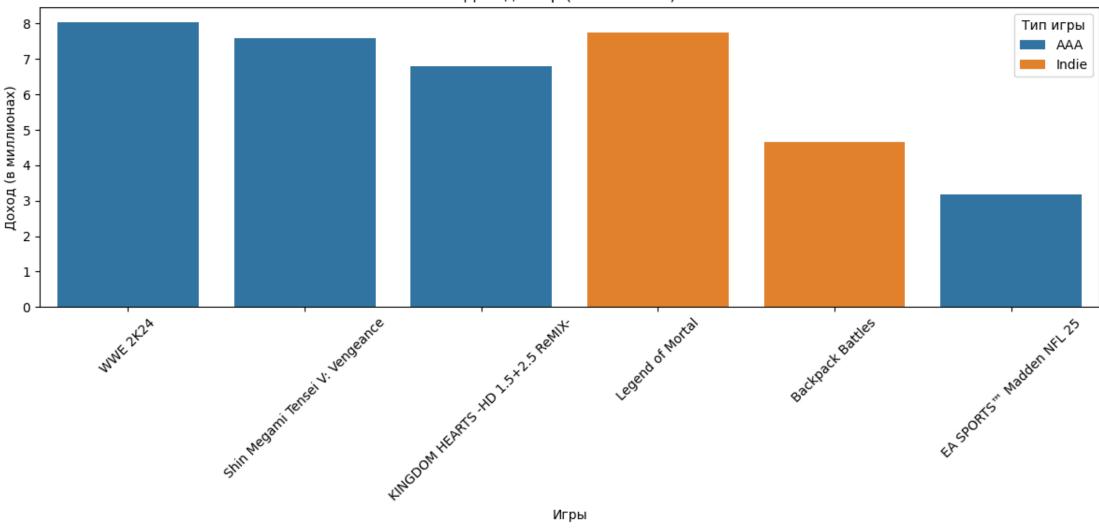














```
🛵 ex_steam_game.py 🛚
      import matplotlib.pyplot as plt
      matplotlib.use('TkAgg')
       import seaborn as sns
       import numpy as np
       import pandas as pd
       import sqlite3
       data = pd.read_csv('Steam_2024_bestRevenue_1500.csv')
       conn = sqlite3.connect('steam_database_2.db')
       def genre_sales_distribution(data):
           genre_sales = data.groupby('publisherClass')['revenue'].sum().sort_values(ascending=False)
          plt.figure(figsize=(10, 6))
          genre_sales.plot(kind='bar', color='skyblue')
          plt.title('Продажи по классу издателя')
          plt.xlabel('Жанр')
           plt.ylabel('Общий доход ($)')
           plt.xticks(rotation=45)
           plt.show()
           genre_sales_df = genre_sales.reset_index()
           genre_sales_df.columns = ['Genre', 'TotalRevenue']
           genre_sales_df.to_sql('GenreSales', conn, if_exists='replace', index=False)
```



```
def price_sales_correlation(data):
    plt.figure(figsize=(10, 6))
    sns.scatterplot(x='price', y='copiesSold', data=data, color='orange')
    plt.title('Цена vs Объем продаж')
    plt.xlabel('Цена ($)')
    plt.ylabel('Проданные копии')
    plt.show()
    price_sales_df = data[['price', 'copiesSold']].copy()
    price_sales_df.to_sql('PriceSales', conn, if_exists='replace', index=False)
def rating_sales_correlation(data):
    plt.figure(figsize=(10, 6))
    sns.scatterplot(x='reviewScore', y='copiesSold', data=data, color='green')
    plt.title('Рейтинг vs Объем продаж')
   plt.xlabel('Оценка')
    plt.ylabel('Проданные копии')
    plt.show()
    rating_sales_df = data[['reviewScore', 'copiesSold']].copy()
    rating_sales_df.to_sql('RatingSales', conn, if_exists='replace', index=False)
```



```
def seasonal_sales_impact(data):
    data['releaseDate'] = pd.to_datetime(data['releaseDate'], format='%d-%m-%Y')
    data['year'] = data['releaseDate'].dt.year
    data['month'] = data['releaseDate'].dt.month
    monthly_sales = data.groupby('month')['copiesSold'].sum()
    plt.figure(figsize=(10, 6))
    monthly_sales.plot(kind='line', marker='0', color='blue')
    plt.title('Тренды продаж по месяцам')
    plt.xlabel('Месяц')
    plt.ylabel('Общий объем продаж')
    plt.xticks(np.arange(1, 13), ['Янв', 'Фев', 'Мар', 'Апр', 'Май', 'Июн', 'Июл', 'Авг', 'Сен', 'Окт', 'Ноя', 'Дек'])
    plt.show()
    monthly_sales_df = monthly_sales.reset_index()
    monthly_sales_df.columns = ['Month', 'TotalSales']
    monthly_sales_df.to_sql('MonthlySales', conn, if_exists='replace', index=False)
genre_sales_distribution(data)
price_sales_correlation(data)
rating_sales_correlation(data)
seasonal_sales_impact(data)
```



```
def seasonal_sales_impact(data):
    data['releaseDate'] = pd.to_datetime(data['releaseDate'], format='%d-%m-%Y')
    data['year'] = data['releaseDate'].dt.year
    data['month'] = data['releaseDate'].dt.month
    monthly_sales = data.groupby('month')['copiesSold'].sum()
    plt.figure(figsize=(10, 6))
    monthly_sales.plot(kind='line', marker='0', color='blue')
    plt.title('Тренды продаж по месяцам')
    plt.xlabel('Месяц')
    plt.ylabel('Общий объем продаж')
    plt.xticks(np.arange(1, 13), ['Янв', 'Фев', 'Мар', 'Апр', 'Май', 'Июн', 'Июл', 'Авг', 'Сен', 'Окт', 'Ноя', 'Дек'])
    plt.show()
    monthly_sales_df = monthly_sales.reset_index()
    monthly_sales_df.columns = ['Month', 'TotalSales']
    monthly_sales_df.to_sql('MonthlySales', conn, if_exists='replace', index=False)
genre_sales_distribution(data)
price_sales_correlation(data)
rating_sales_correlation(data)
seasonal_sales_impact(data)
```



```
# Дополнительные данные о доходах игр
data2 = {
```





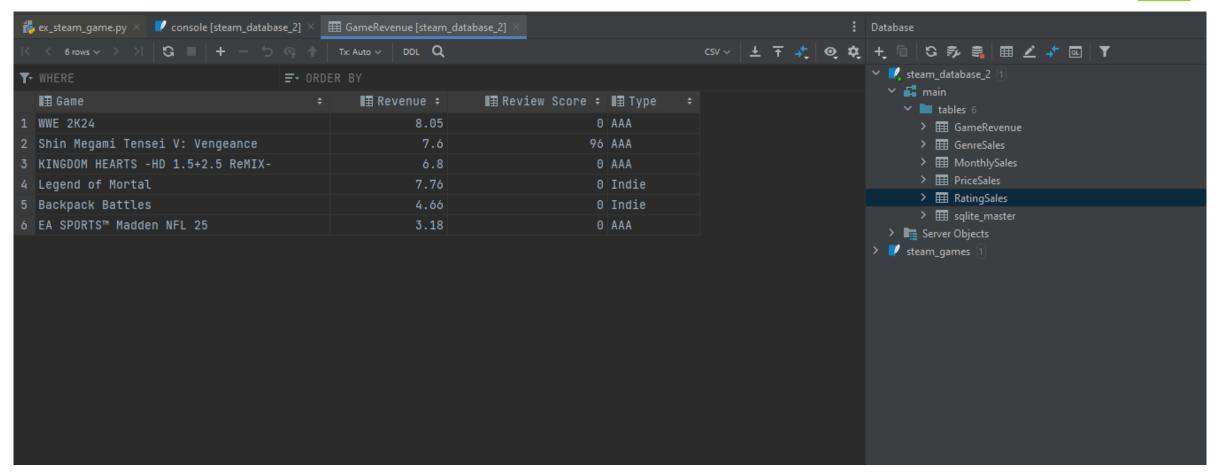
```
df = pd.DataFrame(data2)
df = df[df['Revenue'] > 0]

plt.figure(figsize=(12, 6))
sns.barplot(x='Game', y='Revenue', hue='Type', data=df)
plt.title('Доходы игр (ААА vs Indie)')
plt.xticks(rotation=45)
plt.ylabel('Доход (В миллионах)')
plt.xlabel('Игры')
plt.tegend(title='Tип игры')
plt.tight_layout()
plt.show()
df.to_sql('GameRevenue', conn, if_exists='replace', index=False)

conn.close()
```

Вывод из БД







Спасибо за внимание!

Контакты Сайт **\(+7 913 833 73 30** ☑ anton.i.sadakov@sibir.rt.ru