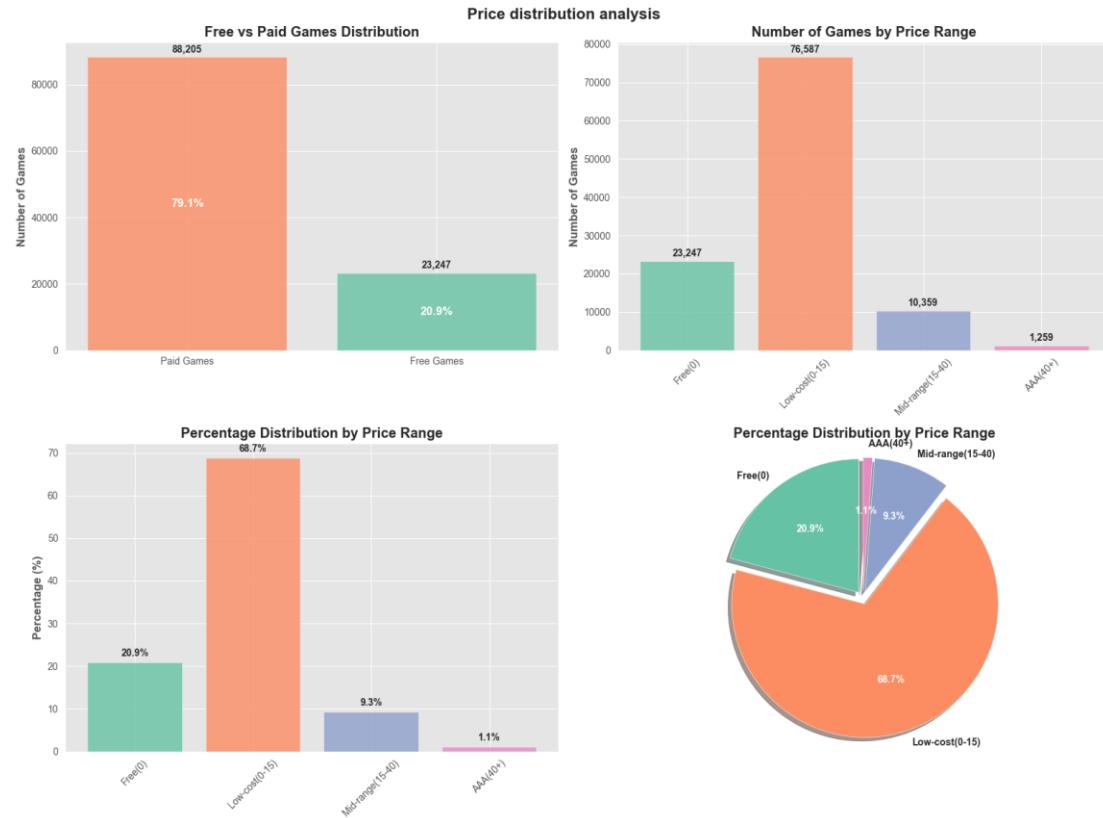
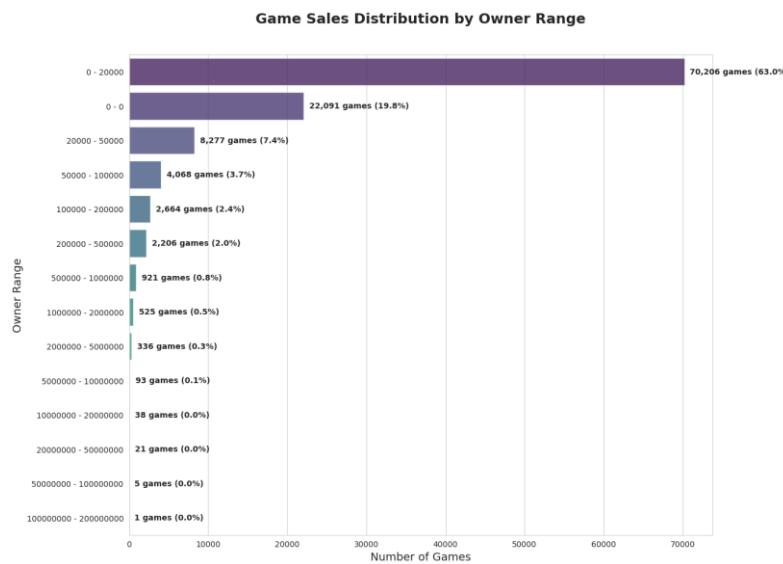


Report on Steam Market Analysis

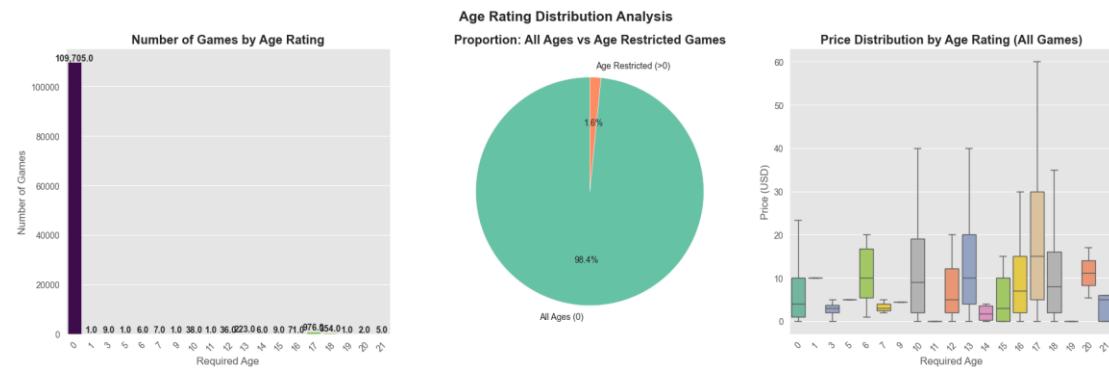
Part I -Some easy intuitive thing to begin with



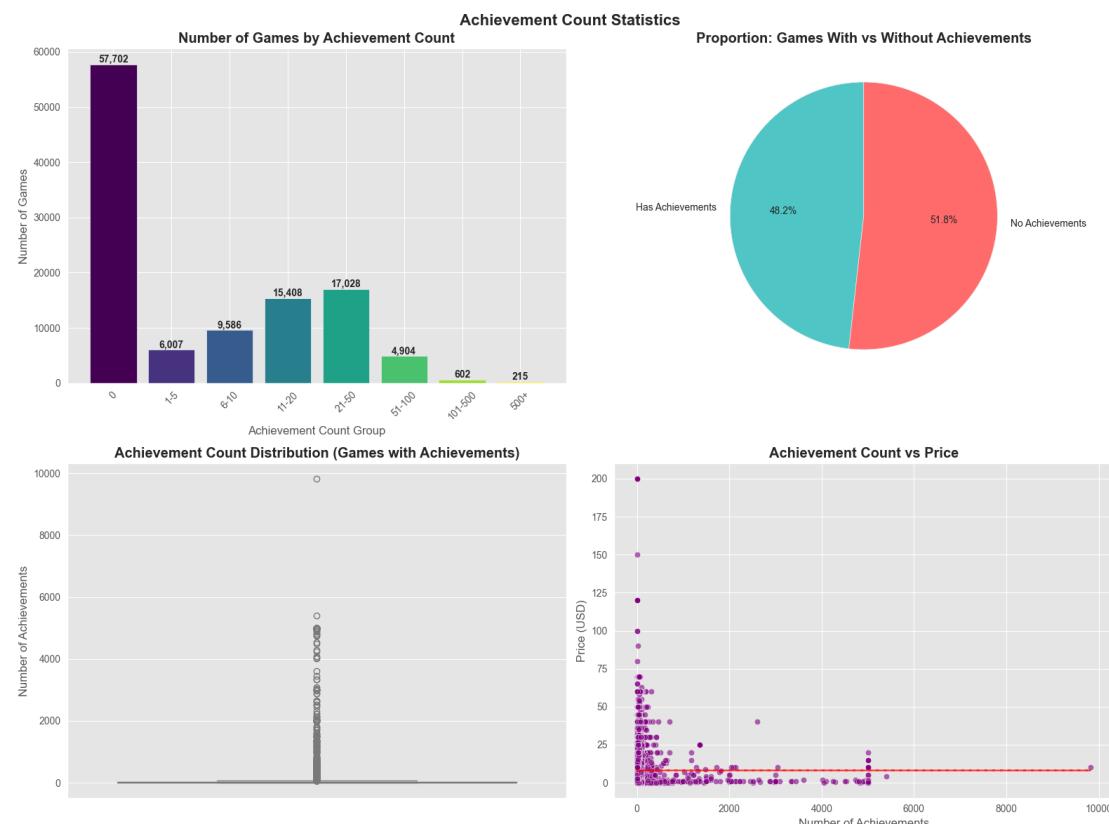
Above is the price distribution of the games in Steam , we can see that 68.7% of the game is between \$0-\$15 , it is the predominant game price of the market this is consist of old AAA games and indie games



Above is the game Sales distribution , this is an estimated range , the true value is hidden inside Valve Corp , if someday I went to work with them or hack into their system I can get the real data, from this data we can see that 82.8% the game is below 20000 copies sold , these games are low quality



Above is the age rating distribution , most of the game(96.4%) is without of age rating or the age restriction is 0. This doesn't conclude anything but watch below!



Above is the distribution of game achievements from there we still conclude nothing ,

but watch this !

==== Correlation Analysis ===

Correlation matrix (games with achievements only):

Variable	Coefficient
achievements	1.000000
average_playtime_forever	0.019320
recommendations	0.012035
positive	0.012016
metacritic_score	0.008623
price	-0.000063

Achievement Group	price	metacritic_score	recommendations	positive	average_playtime_forever
0	N/A	N/A	N/A	N/A	N/A
1-5	4.27	0.31	92.95	127.02	18.21
6-10	6.48	0.90	92.99	130.60	37.04
11-20	6.83	2.97	317.26	412.76	62.07
21-50	10.7	6.67	1194.27	1361.0	145.13
	6			3	
51-100	10.6	8.69	3350.18	3877.9	285.76
	1			3	
101-500	11.2	16.41	17449.45	25279.	1106.99
	0			42	
500+	4.70	2.85	4779.47	9345.0	350.93
				1	

We can see that that overall statistics grows with achievements numbers!

Extremely low correlation (only 0.019) indicates achievement count has almost no relationship with total playtime, **Insight: Simply adding more achievements won't make players play longer**

PRACTICAL RECOMMENDATIONS

For Game Developers:

1. Recommended Achievement Count:

- **Ideal Range: 11-50 achievements**
- This range provides the best balance between playtime, ratings, and player engagement

2. Avoid Extremes:

- Too few achievements (<10): Lower player engagement
- Too many achievements (>100): Diminishing returns, may feel overwhelming

3. Quality Over Quantity:

- Well-designed, meaningful achievements are more valuable than numerous repetitive ones

For Players Choosing Games:

1. **Don't judge games solely by achievement count**
2. **Games with 11-50 achievements typically provide good experiences**
3. **Focus on game genre and personal preferences** rather than achievement statistics

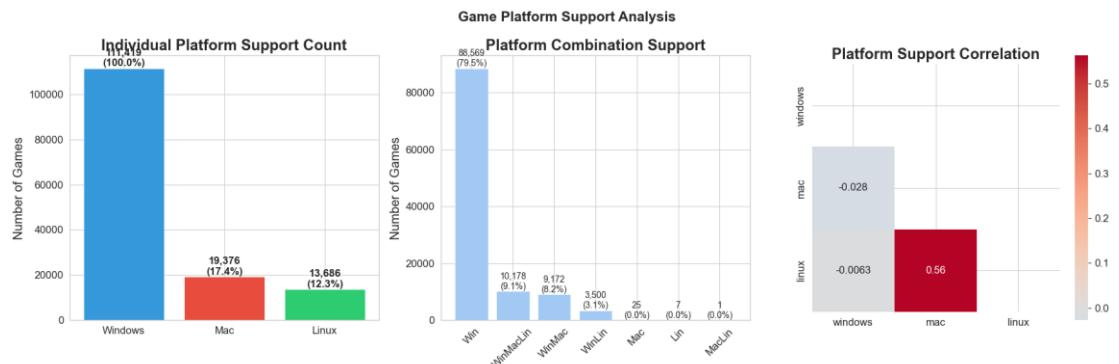
BUSINESS INSIGHTS

- **Pricing Strategy:** Games with 21-500 achievements can support \$10-\$11 price points
- **Player Retention:** Moderate achievement systems (11-50) effectively balance player investment and satisfaction
- **Development Resources:** Don't over-invest in achievement systems; focus on core gameplay instead

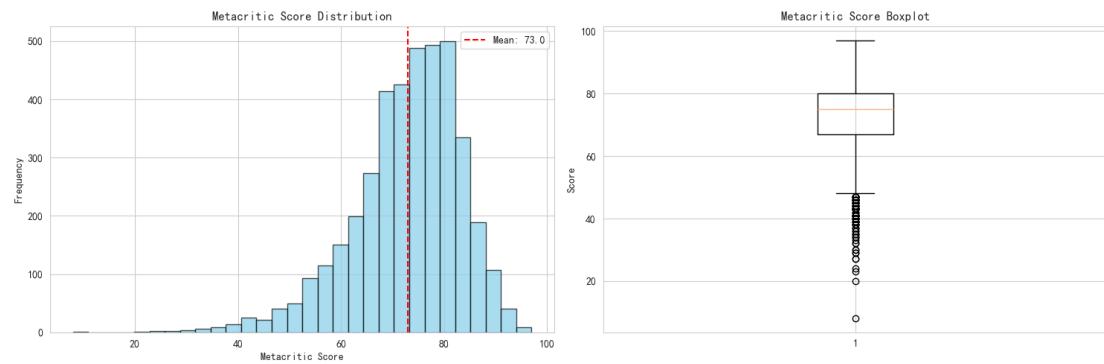
CORE CONCLUSION: In game design, achievement systems should serve as tools to enhance experience, not as primary drivers of player engagement. **11-50 well-designed achievements** represent the most effective configuration.

KEY TAKEAWAY

Achievement systems have minimal impact on core engagement metrics. Focus on creating meaningful, well-paced achievements rather than maximizing quantity. The data suggests that beyond a certain point (approximately 50 achievements), additional achievements provide diminishing returns for both player satisfaction and commercial success.

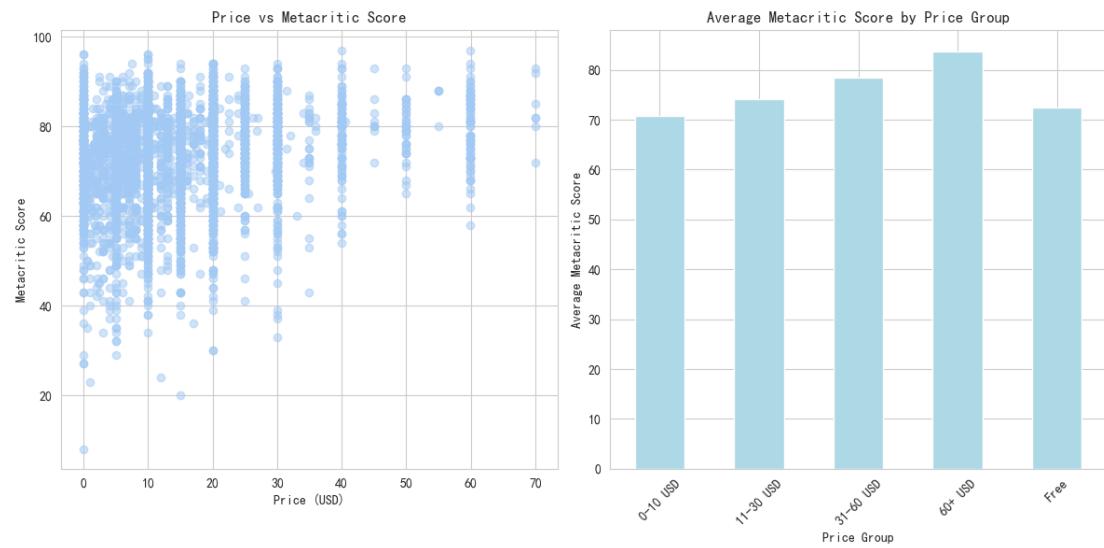


Above is the overall game platform statistics ,we can see that games support mac are likely to support Linux because they are UNIX like system (similar commands, code)



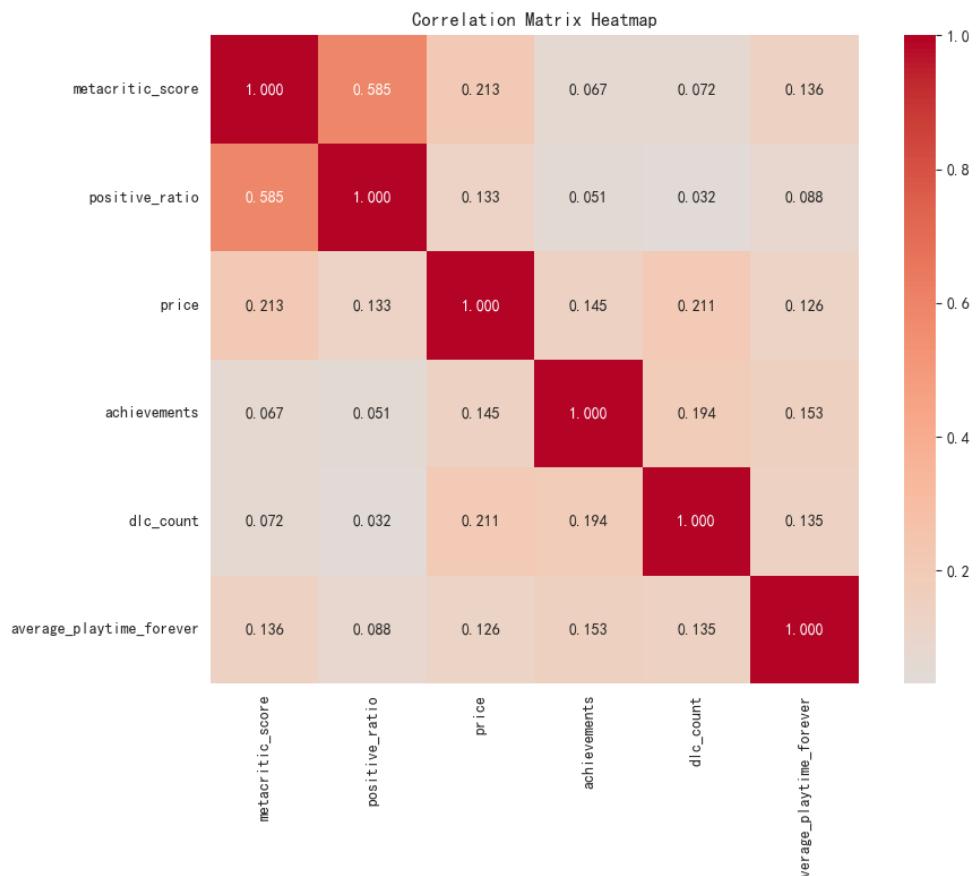
Above is the overall distribution of games' Metacritic score ,this is clearly a normal distribution , the most game ratings are vaguely around 70-80

Score Category Distribution:	
score_category	
Excellent (90-100)	113
Fair (60-69)	850
Good (70-79)	1543
Great (80-89)	1067
Poor (0-59)	432
Name: count, dtype: int64	

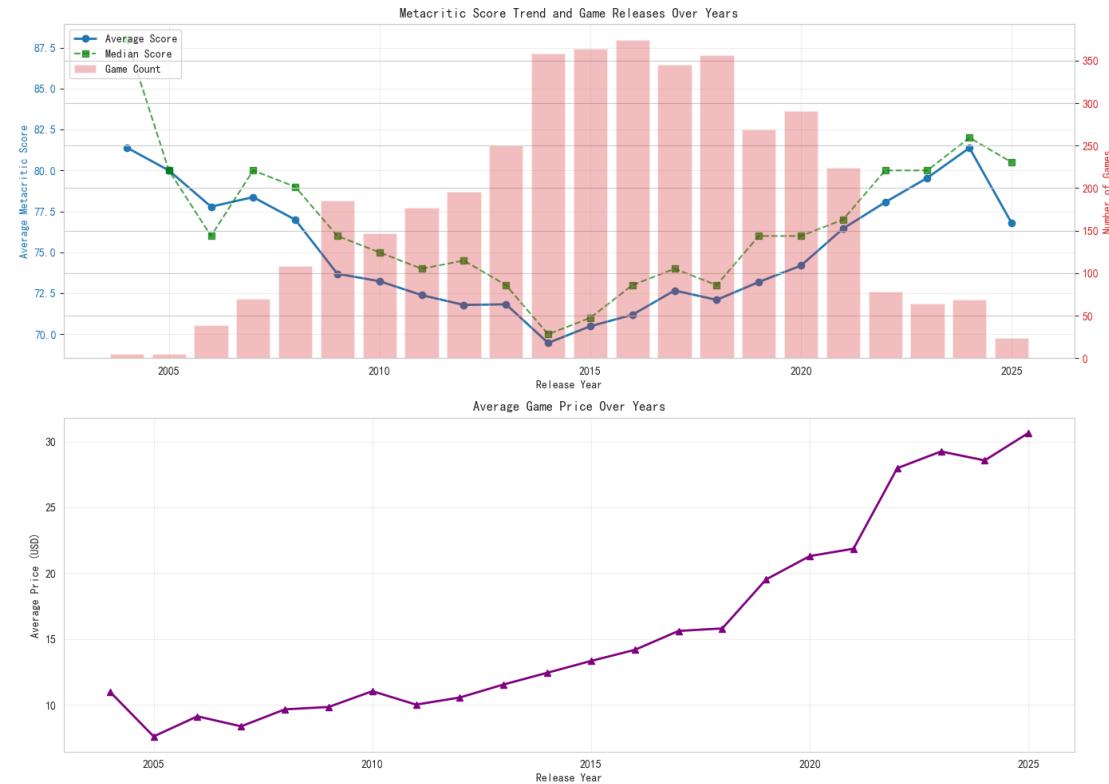


Above is interesting that the price of the game goes with the average Metacritic score ,there is a slight increase in price with higher score as shown in the right graph.

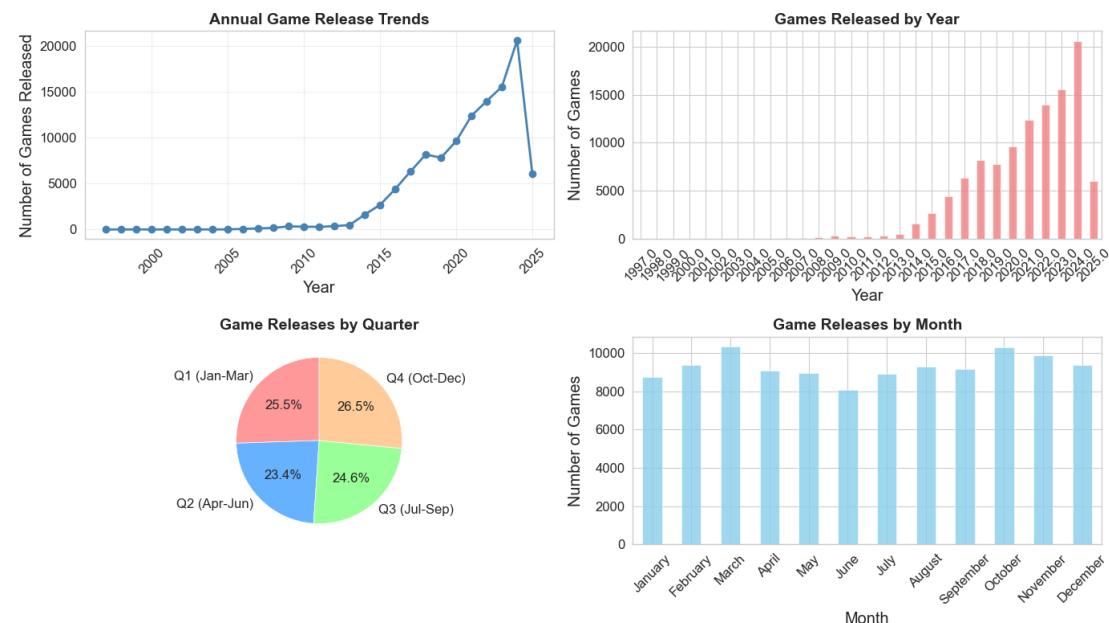
And I want to say a few more things about Metacritic score



Above is a heatmap the positive ratio is clearly influenced by the Metacritic score ,as they are experts in game review

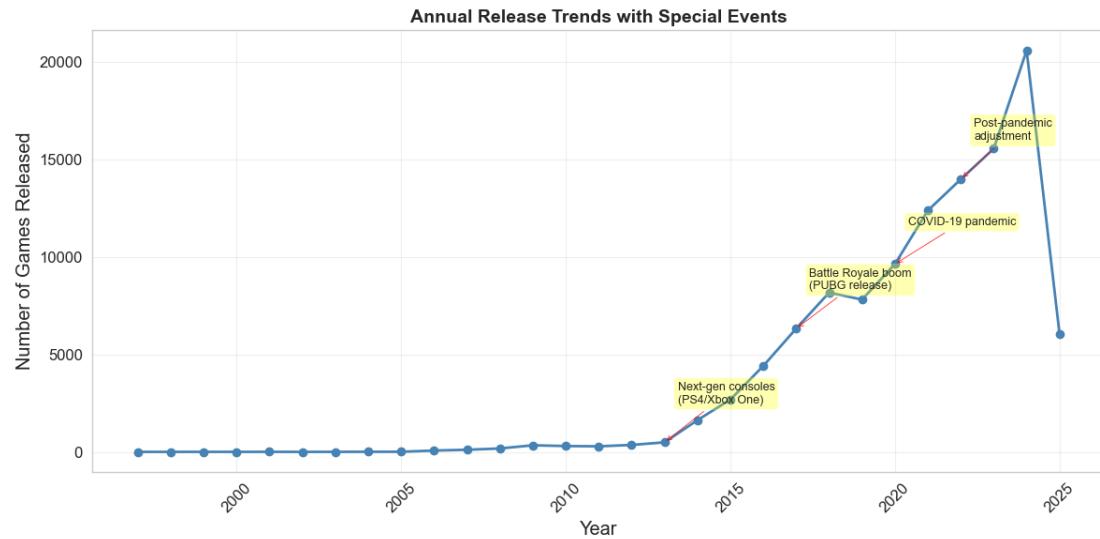


Above is a graph for reference , well , you can see new games are tend to be more expensive



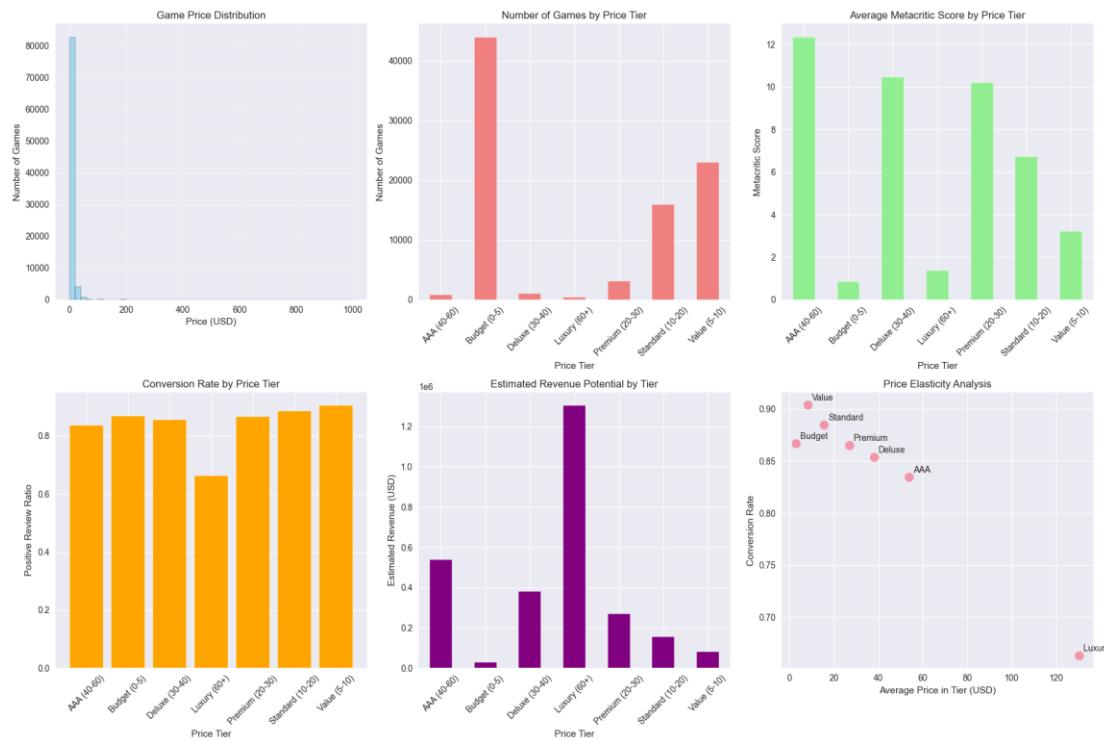
From above ,we can see that with the help of AI and hardware development ,the

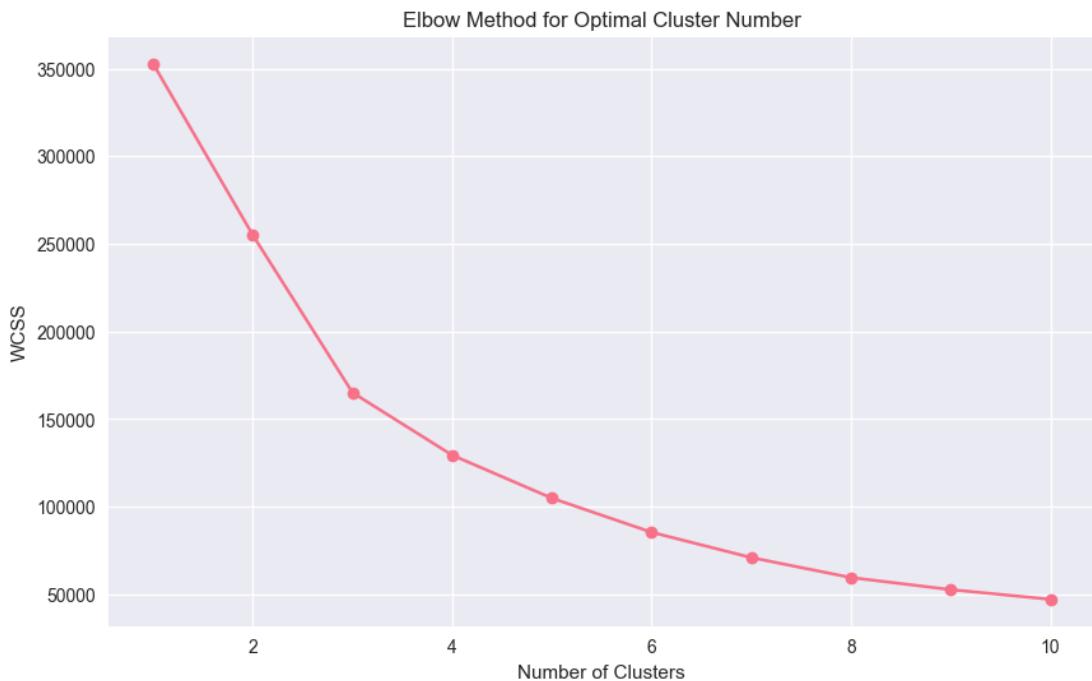
game releases every year is increasing(the data stops at 2025 May)



We can see that after PS4/XBOX ONE released games are explosive growing

Part II -Commercial Value analysis





Analysis Report: Steam Game Pricing Strategy

1. Executive Summary

Based on analysis of 88,205 paid games, an optimal pricing range of USD 15-30 is recommended to balance revenue, player satisfaction, and market positioning. The Value tier (USD 5-10) shows the highest conversion rate (90.4%), while the Luxury tier (USD 60+) demonstrates significant player resistance (66.3% conversion).

2. Key Findings

Price Tier Performance:

- Value (USD 5-10): Highest conversion (90.4%), low sensitivity (0.107)
- Standard (USD 10-20): Strong conversion (88.5%), moderate quality (6.71 avg score)
- Premium (USD 20-30): Good conversion (86.5%), high quality (10.20 avg score)
- Luxury (USD 60+): Poor conversion (66.3%), low quality (1.38 avg score)

Cluster Analysis Reveals Four Market Segments:

- Cluster 0 (84k games): Mass market, low price (USD 8.07), low quality
- Cluster 1 (3.7k games): Premium segment, optimal price (USD 16.65), high quality (72.62 score)
- Cluster 2 (36 games): Blockbuster titles, high engagement (419k avg positives)
- Cluster 3 (274 games): Overpriced segment (USD 163.25), poor quality

Genre-Specific Pricing:

Strategy games command premium pricing (USD 9.99) with highest quality (3.96 score), while Sports games show price-quality mismatch (USD 13.17 price, 1.88 score).

3. Price Sensitivity Analysis

The USD 5-10 range shows lowest sensitivity ratio (0.107), indicating strongest price acceptance. Sensitivity increases significantly above USD 10 (0.155) and below USD 3 (0.190).

4. Recommended Strategy

- **Mass Market Positioning:** Price at USD 5-10 for maximum conversion
- **Premium Positioning:** Target USD 15-30 for quality-value balance
- **Avoid:** Pricing above USD 60 due to severe conversion degradation
- **Genre Considerations:** Adjust pricing based on genre expectations and quality delivery

This analysis demonstrates that optimal pricing requires balancing actual game quality with player value perception, with the USD 15-30 range offering the best equilibrium for quality-focused developers.

And let's take a look at the DLC number

Average indicators grouped by the number of DLCs:			
rating_ratio	positive	owners_estimate	\
dlc_count			
(-1, 0]	0.755208	282.443318	3.953096e+04
(0, 1]	0.814713	2522.299431	1.593887e+05
(1, 5]	0.793004	3929.763782	3.271185e+05
(5, 10]	0.767455	10058.937611	7.591889e+05
(10, 100]	0.765496	26977.807054	1.403537e+06
(100, 1000]	0.720577	18239.730769	1.156731e+06
average_playtime_forever			
dlc_count			
(-1, 0]		48.157923	
(0, 1]		171.174165	
(1, 5]		455.948336	
(5, 10]		690.741533	
(10, 100]		1381.943983	
(100, 1000]		1747.269231	
Having DLC vs without DLC comparison :			
rating_ratio	positive	owners_estimate	
has_dlc			
False	0.755208	282.443318	39530.962035
True	0.804792	4098.115496	274912.579803

Analysis of DLC Count and Game Popularity Relationship

Key Findings from Grouped Analysis

1. Data Distribution Characteristics

From descriptive statistics:

- **75% of games have no DLC** (75th percentile is 0)
- **Average DLC count is 0.45**, but with large standard deviation (12.0), indicating highly uneven distribution
- **Maximum DLC count reaches 2,366**, showing extreme outliers exist

2. Significant Difference: Games With vs Without DLC

Games WITH DLC: Rating ratio 0.805, Positive reviews 4,098, Estimated players 274,912

Games WITHOUT DLC: Rating ratio 0.755, Positive reviews 282, Estimated players 39,531

Conclusion: Games with DLC significantly outperform those without DLC across all metrics:

- **6.5% higher rating ratio** → Higher player satisfaction
- **14.5x more positive reviews** → Much higher popularity
- **7x more players** → Greater commercial success

3. Sweet Spot Analysis by DLC Quantity

DLC Count Range	Rating Ratio	Positive Reviews	Player Count	Playtime
0 DLCs	0.755	282	39,531	48 min
1-5 DLCs	0.793	3,930	327,118	456 min
5-10 DLCs	0.767	10,059	759,189	691 min
10-100 DLCs	0.765	26,978	1,403,537	1,382 min
100+ DLCs	0.721	18,240	1,156,731	1,747 min

Important Business Insights

1. “Having DLC” Matters More Than “DLC Quantity”

- Simply having any DLC correlates with significantly better game performance
- However, excessive DLC quantity can reduce player satisfaction

2. Optimal DLC Strategy: 1-5 DLCs

- **Highest rating ratio** (0.793)
- **Best return on investment** (doesn't require massive development resources)
- **Optimal balance between player satisfaction and engagement**

3. Diminishing Returns Effect

- Beyond 5 DLCs, rating ratios begin to decline
- While player count and playtime continue increasing, player satisfaction decreases
- 100+ DLCs show the lowest rating ratio (0.721), possibly due to:
 - Declining DLC quality
 - Player perception of being “exploited”
 - Overly fragmented content

4. Playtime Positively Correlates with DLC Count

- More DLCs = longer player engagement
- But requires balancing with player satisfaction metrics

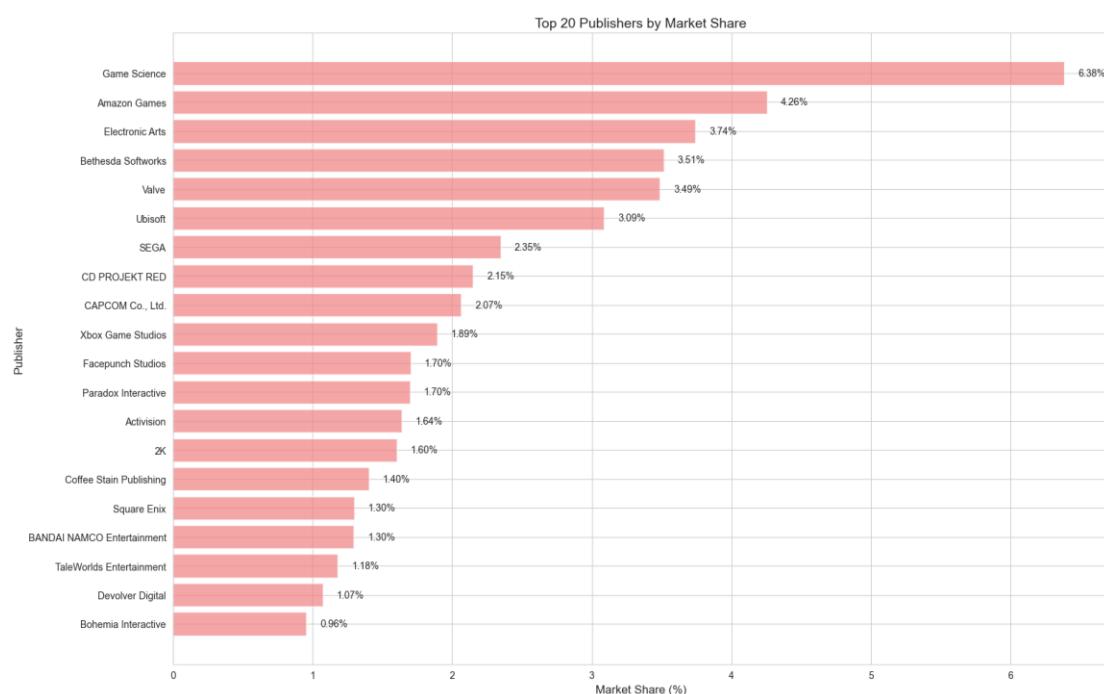
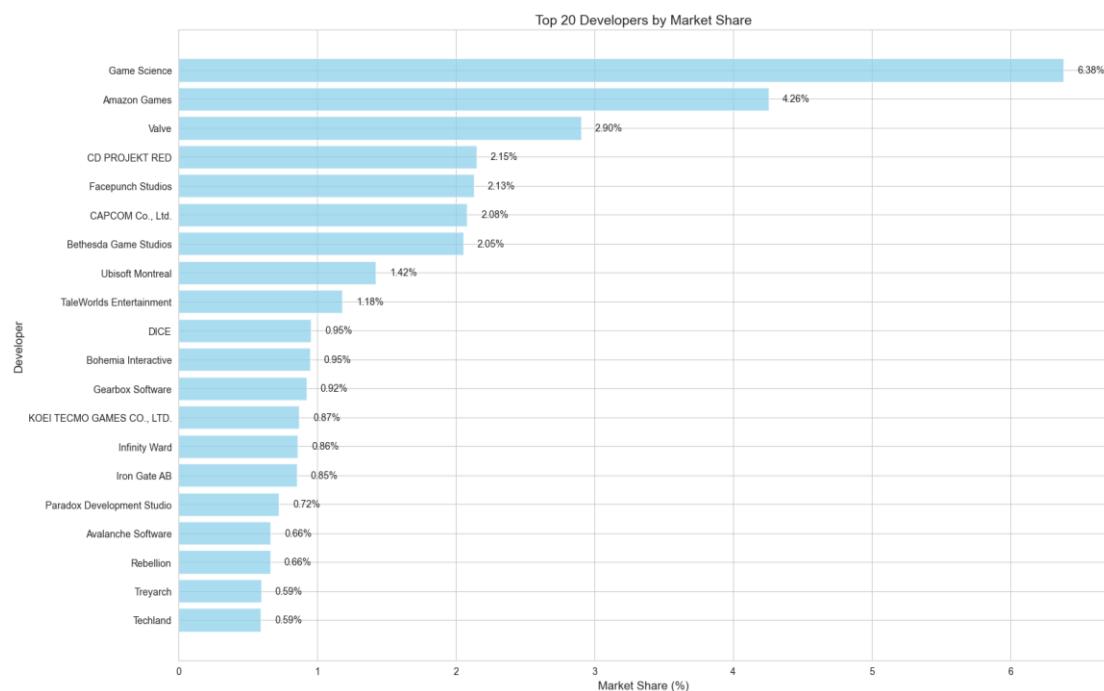
Strategic Recommendations

1. **Prioritize developing 1-5 high-quality DLCs** - this offers the best cost-benefit ratio
2. **Avoid DLC oversaturation** - be cautious when exceeding 10 DLCs
3. **For new games**, DLC strategy can serve as an important success indicator
4. **Focus on DLC quality rather than quantity**

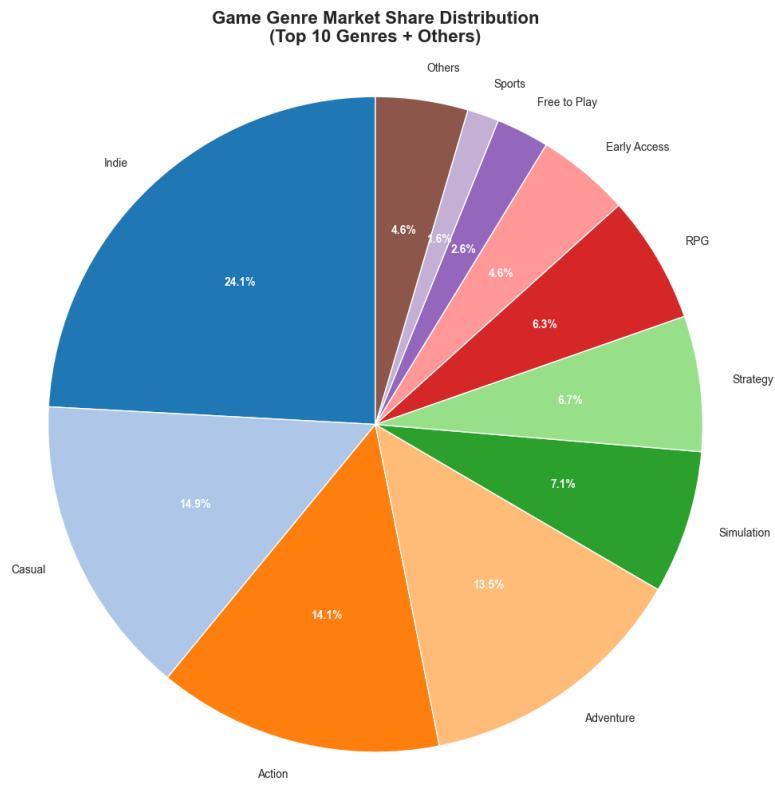
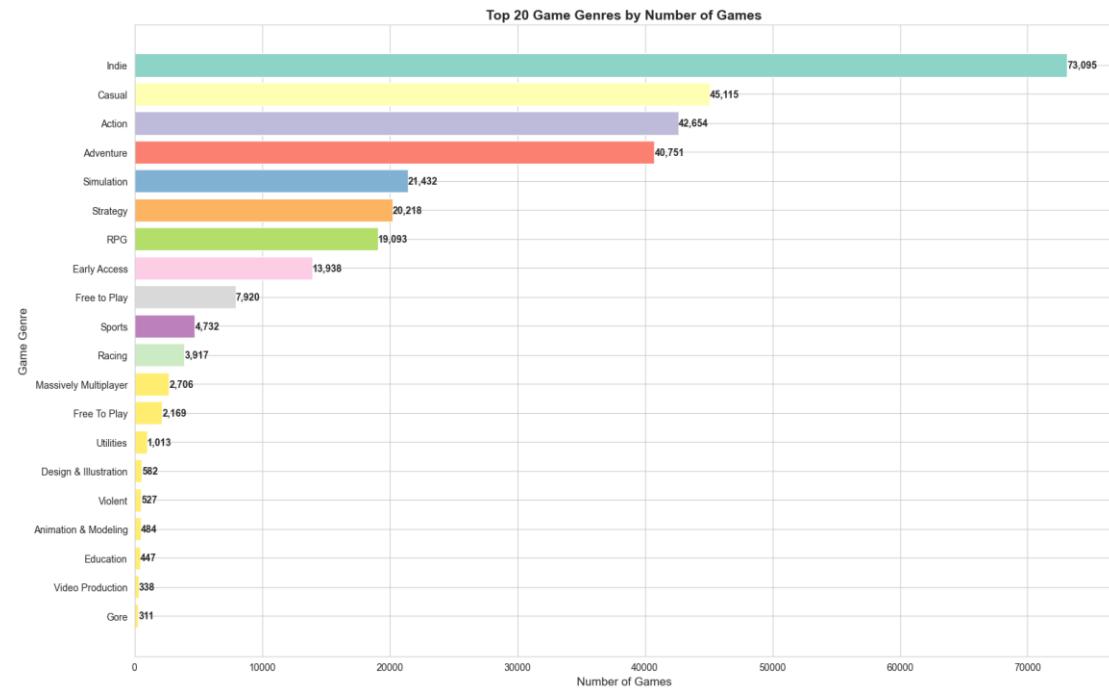
This analysis demonstrates that: **A moderate DLC strategy (1-5 DLCs) maximizes both player satisfaction and commercial success** while maintaining sustainable development practices.

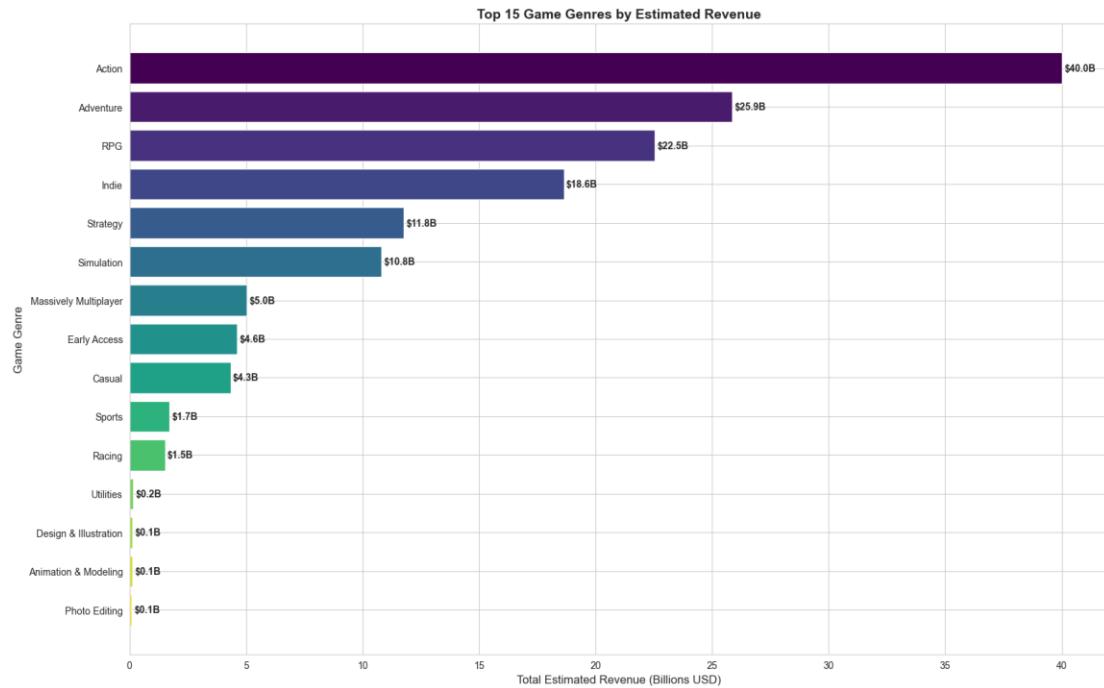
And here is some quick overview of market share with no comments





Next we have genres , which is officially given by steam , here is the statistics

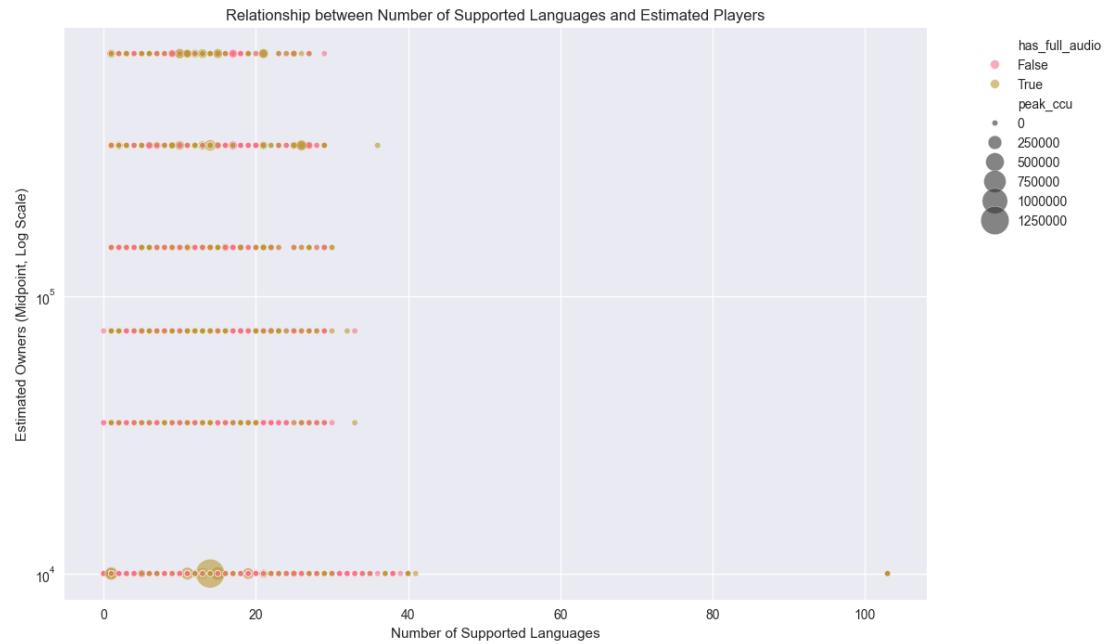




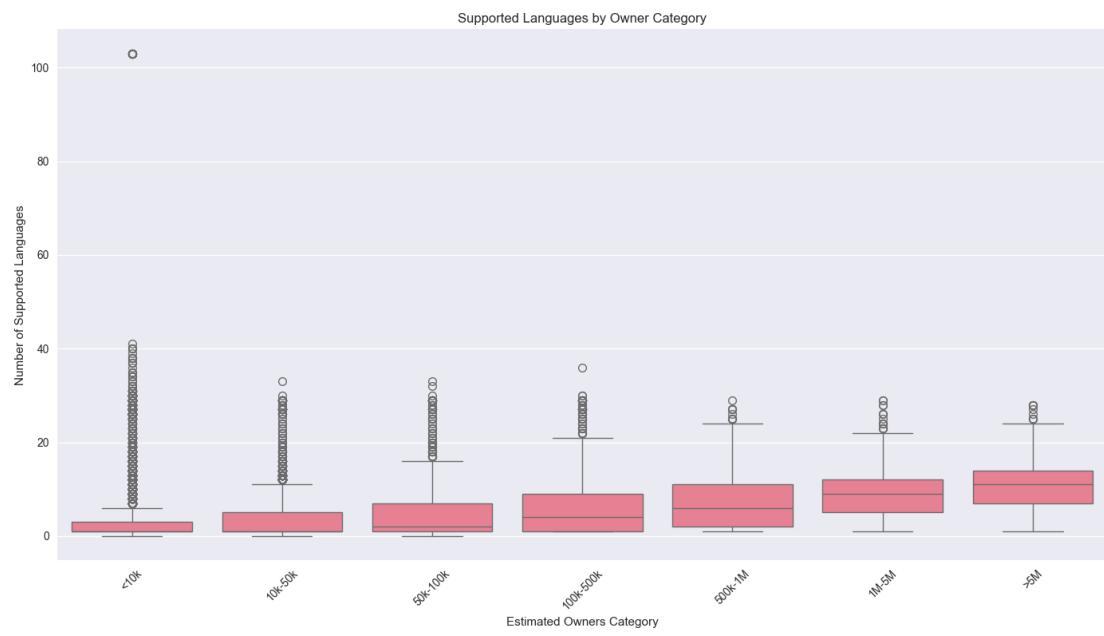
And by the way ,here is a tag cloud which is given by users,players



Because I am using data that is an estimated range(number of game owners), so the following graph might seemed a little bit too rectified , it is using the mid point of the range , the following is about language support and game sales



Pearson Correlation Coefficient: 0.082



Average Supported Languages by Owner Category:

owner_category	mean	median	count
<10k	3.321810	1.0	70206
10k-50k	3.716806	1.0	8277
50k-100k	4.435595	2.0	4068
100k-500k	5.537988	4.0	4870
500k-1M	7.034745	6.0	921
1M-5M	8.743322	9.0	861
>5M	11.145570	11.0	158

Correlation with log(owners): 0.157

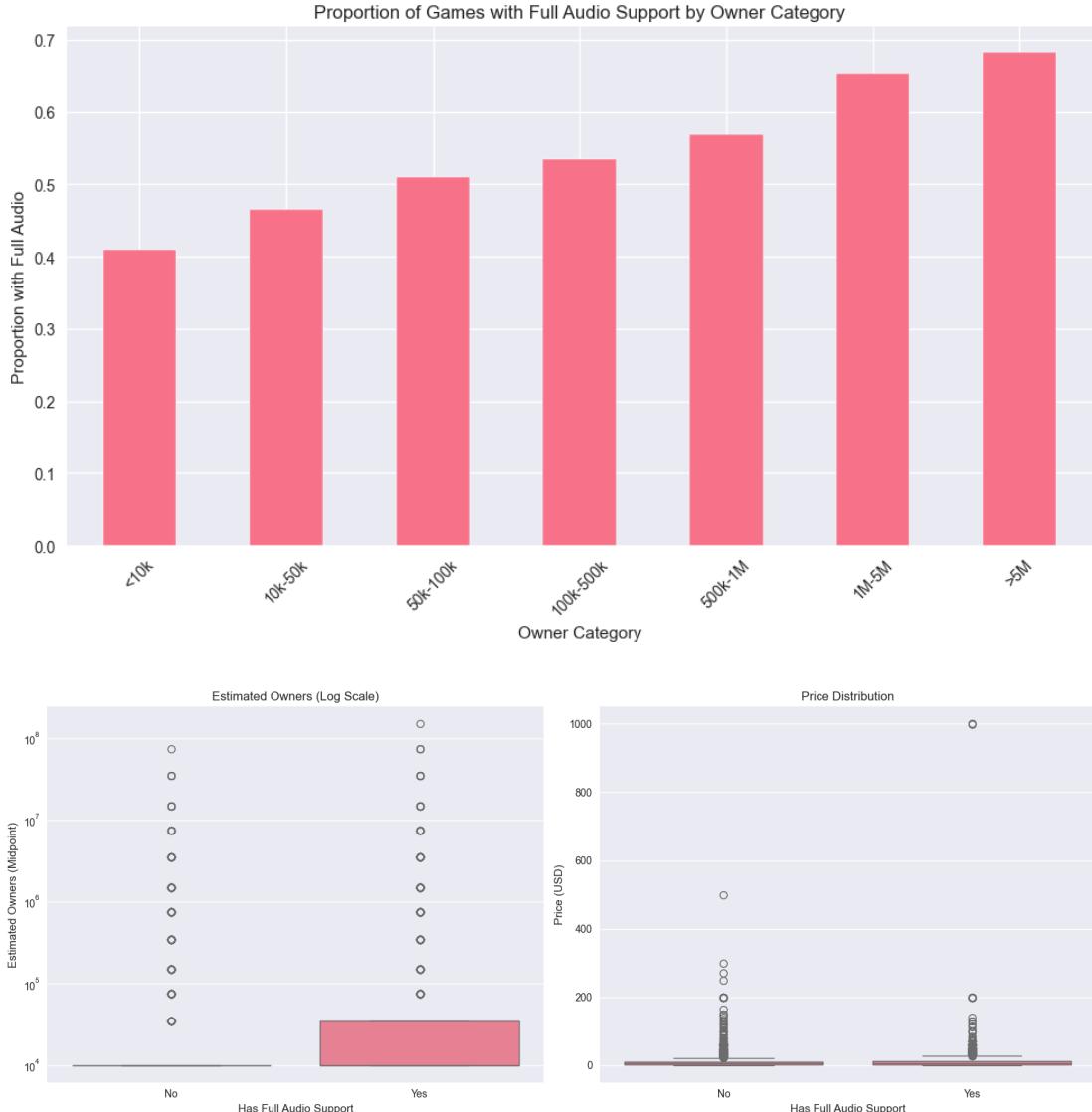
Correlation for small games (<100K): 0.049

Correlation for large games ($\geq 100K$): 0.171

Conclusion

- Adding language support does not appear to cause increased market reach. Language Support Alone is Not a Driver
- For game developers before 100,000 owners, focus on other factors such as game quality ,and maybe invest in 2-3 languages(English,Chinese,Spanish)
- For games with over 100,000 owners, investing in more languages(find which languages have more audience) can help reach a wider audience
- This doesn't mean localization is unimportant - it means it should be treated as an optimization for successful games rather than a growth strategy for unknown games.

Next let's take a look at audio support



Game Full Audio Support and Market

Performance Analysis Report

Executive Summary

Based on analysis of 89,361 games, we found that **42.99% of games provide full audio support**. Games with full audio support significantly outperform those without in key metrics including player base, price, achievement count, and other indicators, with these differences being statistically significant.

Key Findings

1. Basic Distribution

- **Total games analyzed:** 89,361
- **Games with full audio support:** 38,412 (42.99%)
- **Games without full audio support:** 50,949 (57.01%)

2. Key Metrics Comparison

Metric	With Full Audio (Mean)	Without Full Audio (Mean)	Difference
Estimated Player Count	123,645	56,937	+117%
Price (USD)	9.35	7.23	+29%
Total Reviews	1,780	583	+205%
Peak Concurrent Users	393	92	+327%
Average Playtime (minutes)	121.18	86.37	+40%
Achievement Count	25.33	16.58	+53%
DLC Count	0.60	0.47	+28%
Supported Languages Count	3.77	3.53	+7%

3. Statistical Significance Testing

- **Mann-Whitney U Test:** p-value = 4.295e-147
- **Conclusion:** The difference in player count between games with and without full audio support is **highly statistically significant**

In-Depth Analysis

1. Player Scale vs Audio Support Rate

```
# Player scale segmentation and audio support rate
owner_audio_prop = df_clean.groupby('owner_category', observed=False)[['has_full_audio']].mean()
```

Analysis shows that as player scale increases, the proportion of games providing full audio support **significantly rises**. Among games with large player bases (>5M players), the rate of full audio support is substantially higher than in smaller-scale games.

2. Price and Player Count Distribution

Box plot analysis reveals:

- Games with full audio show **more dispersed price distribution**, including more high-priced games
- Games with full audio have **significantly higher player counts** with wider distribution

3. Game Genres and Audio Support

```
# Audio support rate by genre (only genres with >100 games)
genre_audio_stats_filtered = genre_audio_stats[genre_audio_stats['count'] > 100].sort_values('mean', ascending=False)
```

Certain game genres (such as Adventure, RPG, Interactive Fiction) show higher tendency to provide full audio support, likely due to these genres' greater emphasis on narrative and immersion.

Business Insights and Recommendations

1. For Developers

- **Invest in audio production:** Full audio support shows significant correlation with better market performance

- **Competitive differentiation:** High-quality audio can serve as a key differentiator in competitive markets
- **Pricing strategy:** Games with full audio support can command higher price points

2. For Publishers

- **Market positioning:** Full audio support serves as an indicator of game quality
- **Resource allocation:** Investment in localization and audio production may yield better returns
- **User acquisition:** Games with full audio tend to attract and retain players more effectively

3. For Players

- **Quality indicator:** Full audio support may indicate overall higher game quality
- **Immersive experience:** Players seeking deep immersion should prioritize games with full audio support

Analysis Limitations and Future Research Directions

Limitations

1. Estimated player counts are ranges; using midpoint values may introduce errors
2. Audio quality is not differentiated; only binary “presence/absence” is considered
3. Correlation does not imply causation

Future Research Directions

1. Relationship between **number of audio languages** and market performance
2. Impact analysis of **specific language** audio support
3. Effects of **audio quality** (beyond mere presence)
4. Differences in audio importance across **various game genres**

Conclusion

Full audio support serves as a significant predictor of game market performance. Games providing full audio support consistently outperform others across key metrics including player base, revenue potential, and player engagement. For game developers seeking

market success, investing in high-quality audio production and localization represents a strategic decision worthy of serious consideration.

Recommended Action: Medium to large game projects should prioritize including full audio support as a means to enhance product value and market competitiveness.

Well , I must have to say there is a lot of analysis I have done , but they don't have decisive factors , or they are simply finding some particular games (such as Polarizing Review Games, which is games that have both tremendous positive review and negative review)

Here is the thing , since player number growth , games with more positive view seems to have more negative view , but what we should focus on is the ratio of positive and negative reviews , here is some games with ratio close to 50% , most of the reason of these negative review is about BUG,Price,Game optimization

Top 20 Most Polarizing Games (by polarization index):									
	name	positive	negative	total_reviews	positive_ratio	polarization_index	genres	developers	Static Output
77235	NBA 2K20	30430	30445	60895	0.500041	0.250000	['Simulation', 'Sports']	['Visual Concepts']	
11692	Warhammer 40,000: Eternal Crusade	6272	6328	12600	0.497778	0.249995	['Action', 'Adventure', 'Free to Play', 'Massively Multiplayer', 'Strategy']	['Behaviour Interactive Inc.']	
3235	龙之九霄：云中城	10358	103661	209945	0.504421	0.249983	['Action', 'Adventure', 'Indie', 'RPG', 'Simulation']	['游力工作室']	
108235	Sid Meier's Civilization VII	16728	16227	32955	0.507601	0.249942	['Simulation', 'Strategy']	['Paradox Interactive']	
88677	Dragon's Dogma 2	24305	23379	47684	0.509710	0.249906	['Action', 'RPG']	['CAPCOM Co., Ltd.']	
64957	The Callisto Protocol™	6137	5908	12037	0.509845	0.249903	['Action', 'Adventure']	['Striking Distance Studios']	
108111	Monster Hunter: World	29392	30726	60108	0.488820	0.249975	['Action', 'Adventure', 'RPG']	['CAPCOM Co., Ltd.']	
78113	Angry Birds 2	5686	5403	11089	0.512760	0.249837	['Action', 'Free to Play']	['Rovio Entertainment']	
98716	EA SPORTS FC™ 25	5410	5707	11117	0.486642	0.249822	['Simulation', 'Sports']	['EA Canada & EA Romania']	
48933	Bless Unleashed	15047	14837	29864	0.517365	0.249699	['Action', 'Free to Play', 'Massively Multiplayer', 'Strategy']	['VALDEN']	
4040	Richman10	6389	5771	12880	0.522268	0.249594	['Casual', 'Strategy']	['Softstar Technology (Beijing) Co.,Ltd']	
3445	Digger Online	8680	9621	18301	0.474291	0.249339	['Action', 'Adventure']	['Диггер Максим Константинов']	
4123	ATLAS	19204	21291	40495	0.474231	0.249358	['Action', 'Adventure', 'Massively Multiplayer', 'Strategy']	['Digeronot Games', 'Instinct Games']	
8513	Warhammer 40,000: Dawn of War III	9440	10637	20077	0.470190	0.249111	['Action', 'Simulation', 'Strategy']	['Relic Entertainment', 'Instinct Games']	
51606	Torchlight III	4796	5419	10215	0.469586	0.249970	['Action', 'Adventure', 'RPG']	['Echtra Inc.']	
57436	Call of Duty®: Infinite Warfare	13238	10988	23236	0.530956	0.249866	['Action', 'Adventure']	['Infinity Ward']	
56825	Infestation: The New Z	17523	15496	33019	0.530944	0.249958	['Action', 'Free to Play', 'Indie', 'Massively Multiplayer', 'Strategy']	['Freddakis AB']	
32053	Artifact	11582	13245	24807	0.466878	0.248449	['Strategy']	['Valve']	
45258	Wolfenstein: Youngblood	4874	5694	10568	0.461204	0.248495	['Action']	['Machine Games', 'Arkane Studios']	
78173	Survarium	13947	11886	25833	0.539891	0.248409	['Action', 'Free to Play', 'Massively Multiplayer']	['Vostok Games']	

And for the rest of the analysis you can check inside my jupyter notebook!!!!!!!

PartIII-A comprehensive evaluation metric of games

```
    if (is_prime(n)) {  
        cout << "The number is prime." << endl;  
    }  
    else {  
        cout << "The number is not prime." << endl;  
    }  
}  
  
// Function to check if a number is prime  
bool is_prime(int n) {  
    if (n <= 1) {  
        return false;  
    }  
    for (int i = 2; i * i <= n; i++) {  
        if (n % i == 0) {  
            return false;  
        }  
    }  
    return true;  
}  
  
// Main function  
int main() {  
    cout << "Enter a number: ";  
    int num;  
    cin >> num;  
    check_prime(num);  
    return 0;  
}
```

Game Rating System: Comprehensive Scoring Mechanism

⌚ Overall Scoring Architecture

Comprehensive Rating Formula:

```
Final Score = Popularity(25%) + Quality(25%) + Engagement  
(20%) + Content(15%) + Accessibility(10%) + Value(5%)
```

📊 Six Scoring Dimensions Explained

1. Popularity (25%)

Components:

- **Player Base** (40%): `estimated_owners` normalized
- **Approval Rate** (40%): `positive / (positive + negative)`
- **Concurrent Users** (20%): `peak_ccu` normalized

2. Game Quality (25%)

Components:

- **Professional Reviews** (60%): `metacritic_score` normalized to 0-100
- **Achievement System** (20%): `achievements` count normalized
- **User Rating Quality** (20%): Calculated using **Wilson Score Interval** to address small sample bias

3. Player Engagement (20%)

Components:

- **Long-term Play** (40%): `average_playtime_forever` normalized
- **Recent Activity** (40%): `average_playtime_2weeks` normalized
- **Player Retention** (20%): `median_playtime_forever / average_playtime_forever`

4. Content Richness (15%)

Components:

- **DLC Count** (30%): `dlc_count` normalized

- **Language Support** (20%): `supported_languages` count
- **Game Categories** (30%): `categories` diversity
- **Tag Richness** (20%): `tags` count

5. Accessibility (10%)

Components:

- **Platform Compatibility** (40%): Windows + Mac + Linux support count
- **Age Restrictions** (30%): `1 - required_age` normalized
- **Price Barrier** (30%): Free games score high, lower prices get better scores

6. Value Assessment (5%)

Components:

- **Time Value Ratio** (70%): `playtime / price` playtime-to-price ratio
- **Free Content** (30%): Base score for free games

Key Technical Features

1. Data Normalization

```
# All values normalized to 0-1 range
normalized_score = (value - min_value) / (max_value - min_value)
```

2. Wilson Score Interval

Addresses reliability issues between “10 positive reviews vs 1000 positive reviews” using statistical methods for conservative estimates.

3. Scientific Weight Distribution

- Emphasizes **Popularity** and **Game Quality** (25% each)
- Values **Player Engagement** (20%)
- Appropriately considers **Content Richness** and **Accessibility**
- Slightly considers **Value Factors**

Score Output

Final Score Range: 0-100 points

Each dimension's score is first calculated as a normalized 0-1 score, then aggregated by weights, and finally multiplied by 100.

Sample Output (Not actual score) :

```
Grand Theft Auto V:  
Comprehensive Rating: 85.3/100  
  
Dimension Analysis:  
 Popularity: 92.1%  
 Game Quality: 88.5%  
 Player Engagement: 79.2%  
 Content Richness: 81.7%  
 Accessibility: 76.3%  
 Value Assessment: 68.9%
```

Model Advantages

1. **Multi-dimensional Balance:** Not just relying on ratings or sales
2. **Data-driven:** Based on actual player behavior data
3. **Standardized Processing:** Comparable data across different scales
4. **Statistical Methods:** Addresses small sample bias issues
5. **High Interpretability:** Clear visibility of each dimension's score

This model is particularly suitable for **game recommendations, investment analysis, market research** and other scenarios that require comprehensive evaluation of a game's overall performance!

Multi-dimensional Game Comparison

