



# Game's tags, price and reviews

---

DE MELO SOUSA Isabella  
DE SOUZA OLIVA Gabriel  
MORINA Alessandro  
SOUZA PINTO MEDEIROS Gabriel



INTRODUCTION AND  
MOTIVATION

01

SCRAPING AND  
CLEANING

02

# SUMMARY

03

GRAPHICAL  
ANALYSIS

04

REGRESSION

05

CONCLUSION





# 01

## INTRODUCTION AND MOTIVATION

---

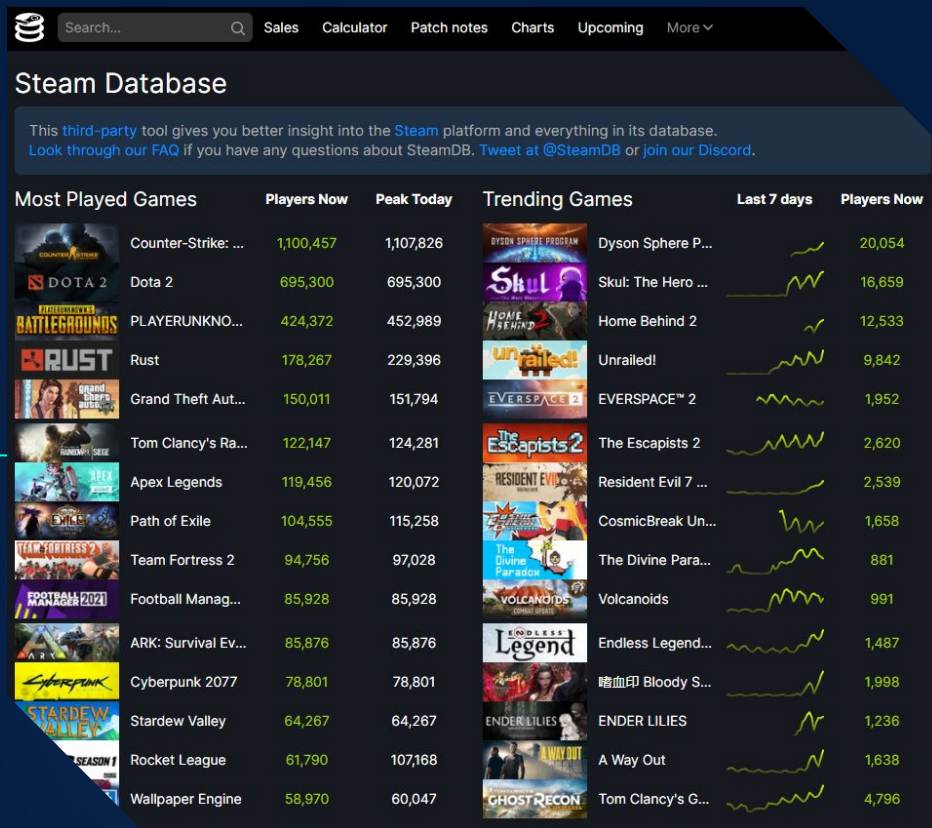


- 
- 
- Context of pandemic, people encouraged to stay at home
  - Understand consumers opinions about games
  - Answer the question: **Can a game's Steam tags and price explain their review score?**
  - Collection of data and analysis
- 
- 

# DATA SOURCE

## STEAM DATABASE

SteamDB is a third-party tool that gives you better insight into the Steam platform and everything in its database.



The screenshot shows the Steam Database website interface. At the top, there's a navigation bar with a search bar and links for Sales, Calculator, Patch notes, Charts, Upcoming, and More. Below the navigation bar, the title 'Steam Database' is displayed. A descriptive paragraph states: 'This third-party tool gives you better insight into the Steam platform and everything in its database. Look through our FAQ if you have any questions about SteamDB. Tweet at @SteamDB or join our Discord.' The main content area is divided into two columns. The left column, titled 'Most Played Games', lists games with their current player counts and peak today counts. The right column, titled 'Trending Games', lists games with their player counts over the last 7 days and current player counts. Each game entry includes a small thumbnail icon, the game title, and the respective player statistics.










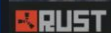









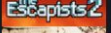

























Most Played Games		Players Now	Peak Today	Trending Games		Last 7 days	Players Now
	Counter-Strike: ...	1,100,457	1,107,826		Dyson Sphere P...		20,054
	Dota 2	695,300	695,300		Skul: The Hero ...		16,659
	PLAYERUNKNO...	424,372	452,989		Home Behind 2		12,533
	Rust	178,267	229,396		Unrilled!		9,842
	Grand Theft Aut...	150,011	151,794		EVERSPACE™ 2		1,952
	Tom Clancy's Ra...	122,147	124,281		The Escapists 2		2,620
	Apex Legends	119,456	120,072		Resident Evil 7 ...		2,539
	Path of Exile	104,555	115,258		CosmicBreak Un...		1,658
	Team Fortress 2	94,756	97,028		The Divine Para...		881
	Football Manag...	85,928	85,928		Volcanoids		991
	ARK: Survival Ev...	85,876	85,876		Endless Legend...		1,487
	Cyberpunk 2077	78,801	78,801		嗜血印 Bloody S...		1,998
	Stardew Valley	64,267	64,267		ENDER LILIES		1,236
	Rocket League	61,790	107,168		A Way Out		1,638
	Wallpaper Engine	58,970	60,047		Tom Clancy's G...		4,796

Image [1]: SteamDB website screenshot



02

# SCRAPING AND CLEANING


---



# GATHERING DATA

Top Rated Games on Steam


Show 25 entries. Hold Shift to sort multiple columns. Search:

#	Name	Positive	Negative	Rating	Positive
1.	 Portal 2	229,370	2,848	97.59%	98.77%
2.	 Hades	117,050	1,329	97.43%	98.88%
3.	 Factorio	107,115	1,166	97.43%	98.92%
4.	 The Witcher 3: Wild Hunt	460,489	9,149	97.11%	98.05%
5.	 Terraria	659,277	14,073	97.07%	97.91%
6.	 Half-Life: Alyx	50,482	672	96.82%	98.69%
7.	 Stardew Valley	309,017	6,859	96.77%	97.83%
8.	 RimWorld	92,464	1,599	96.76%	98.30%
9.	 The Binding of Isaac: Rebirth	122,062	2,307	96.74%	98.15%
10.	 Portal	82,198	1,406	96.72%	98.32%

- SQL Database with all games list
- Crawler accessing app-ids

Image [2]: SteamDB Top Rated games website screenshot


# PAGE SAMPLE

 **The Witcher 3: Wild Hunt**

[Wishlist](#) [Follow](#) [Ignore](#) [Install](#) [Watch](#)

App ID	292030
App Type	Game
Store Name	The Witcher® 3: Wild Hunt
Developer	<a href="#">CD PROJEKT RED</a>
Publisher	<a href="#">CD PROJEKT RED</a>
Supported Systems	Windows
Last Record Update	2 months ago (11 November 2020 – 00:27:34 UTC)
Last Change Number	<a href="#">10188493</a>
Release Date	18 May 2015 – 22:00:00 UTC (6 years ago)

[Store](#) [Hub](#) [PCGW](#) [Patches](#) [Home](#) [Facebook](#) [Discord](#) [Twitter](#) [YouTube](#)



**97.11%**  
↑ 480,489 ↓ 9,149

**28,986**  
In-Game

[User](#) [Sun](#) [Game](#) [PC](#) [Cloud](#) [Screenshot](#) [Video](#) [Screenshot](#) [Screenshot](#)

As war rages on throughout the Northern Realms, you take on the greatest contract of your life — tracking down the Child of Prophecy, a living weapon that can alter the shape of the world.

### User Tags

[Action](#) [Adventure](#) [RPG](#) [Fantasy](#) [Open World](#) [Third Person](#) [Story Rich](#) [Great Soundtrack](#) [Sandbox](#) [Magic](#)

[Atmospheric](#) [Medieval](#) [Singleplayer](#) [Action RPG](#) [Dark](#) [Dark Fantasy](#) [Mature](#) [Choices Matter](#) [Nudity](#)

[Multiple Endings](#)

Image [2]: The Witcher 3 page sample screenshot



## OUR NUMBERS

4

Days scraping

426

Features

40.000

Samples



# DATA CLEANING

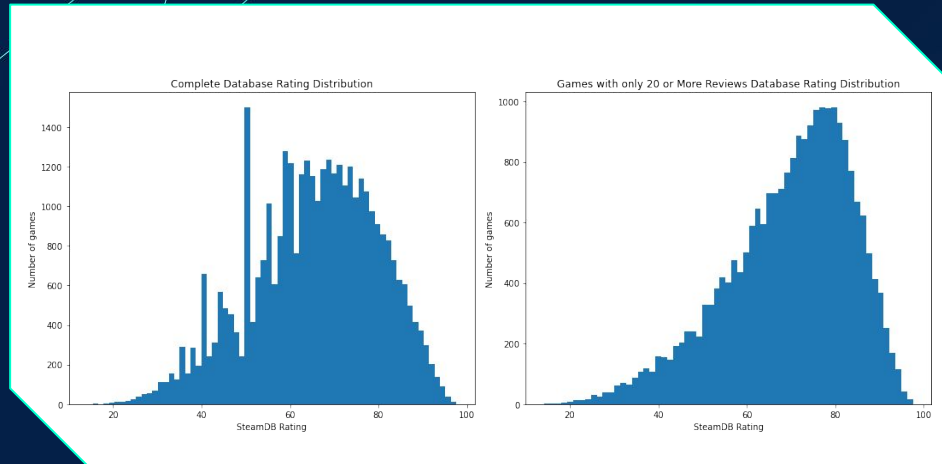


Image [1]: SteamDB website screenshot

- Eliminate rows with missing values
- Data Filtering
- Final sample with 23.000 games



03

# GRAPHICAL ANALYSIS

---

# MOST FREQUENT TAGS

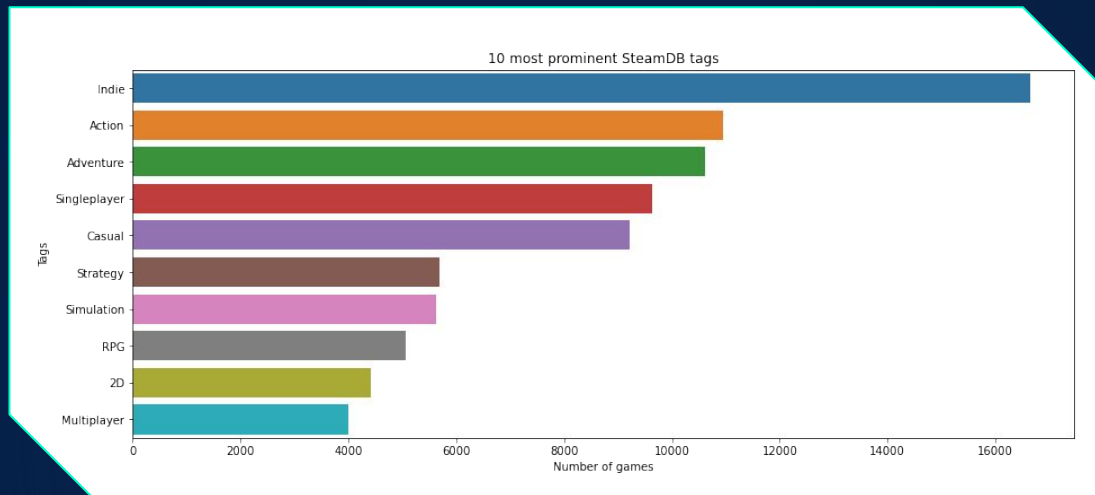


Image [3]: Chart of most prominent SteamDB tags

# AVERAGE RATING OF MOST POPULAR

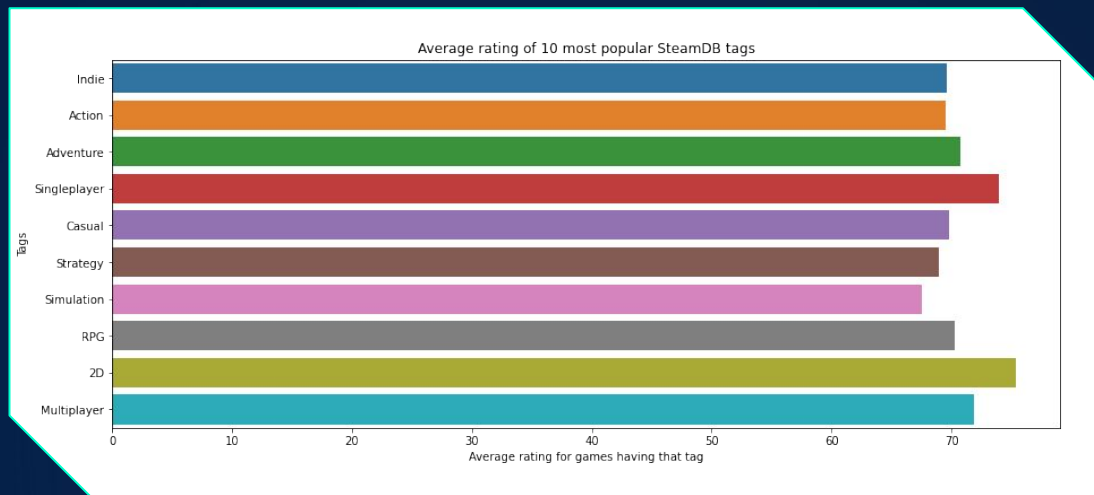


Image [4]: Chart of average rating of 10 most popular tags

# AVERAGE RATING OF MOST WELL RANKED TAGS

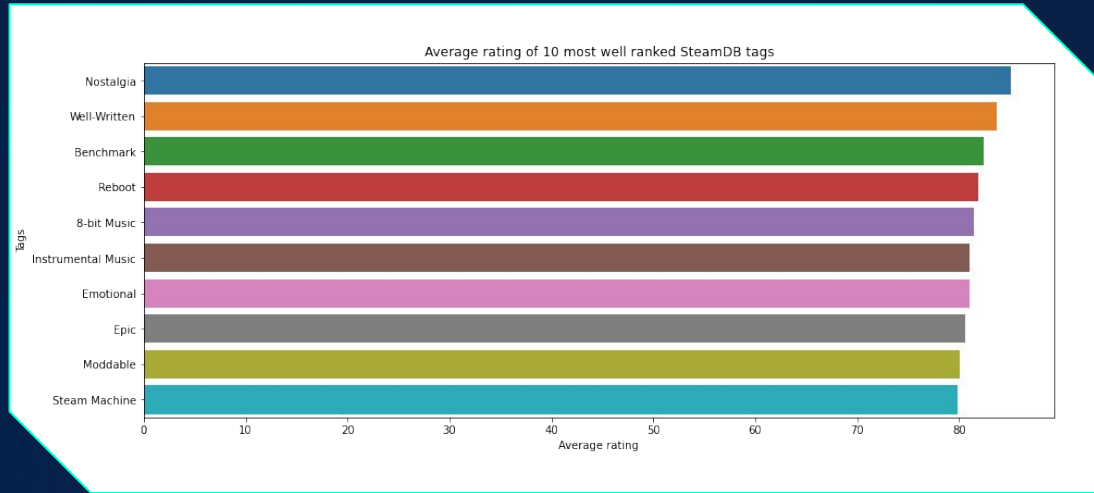


Image [5]: Chart of average rating of most well ranked tags



# AVERAGE PRICE FOR MOST POPULAR TAGS

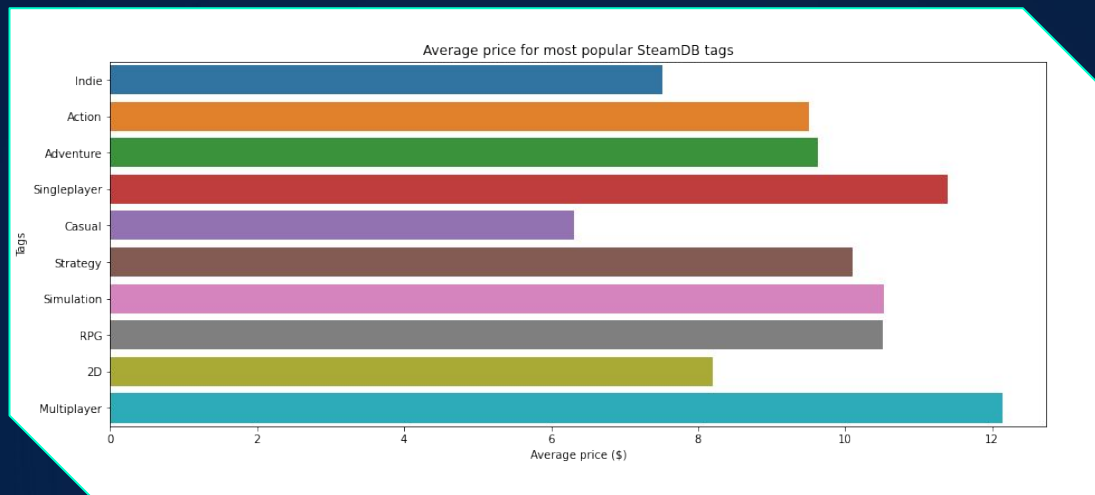


Image [6]: Chart of average price

# AVERAGE PRICE FOR MOST EXPENSIVE TAGS

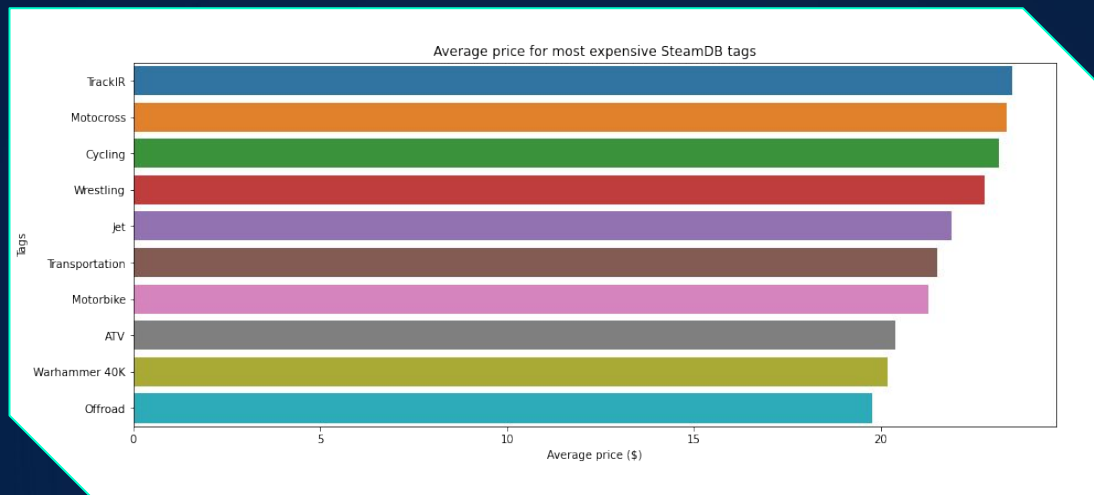


Image [6]: Chart of average price for most expensive tags

# RELATION AVERAGE PRICE DEVS AND PUBLISHERS

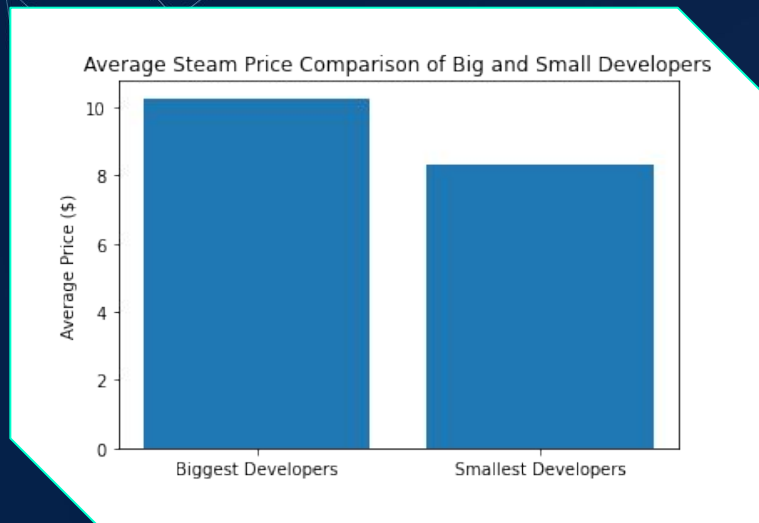


Image [8]: Chart of average price for 30 biggest developers and 30 smallest developers

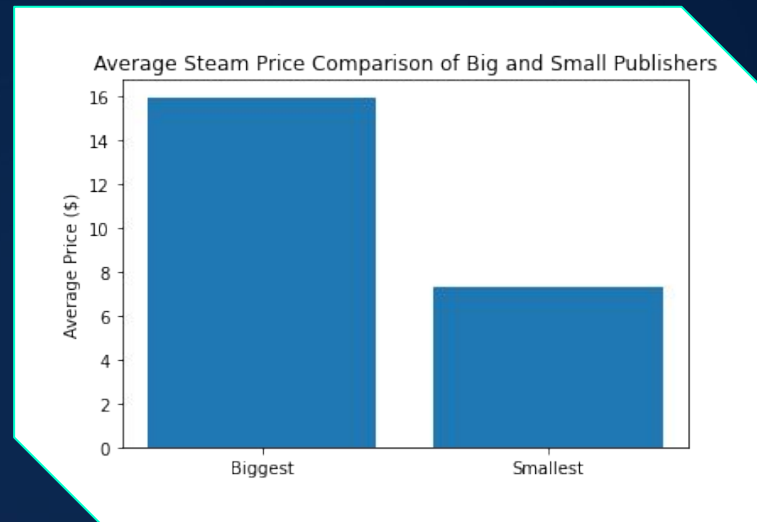


Image [8]: Chart of average price for 30 biggest publishers and 30 smallest publishers

# RELATION AVERAGE RATING AND AVERAGE PRICE

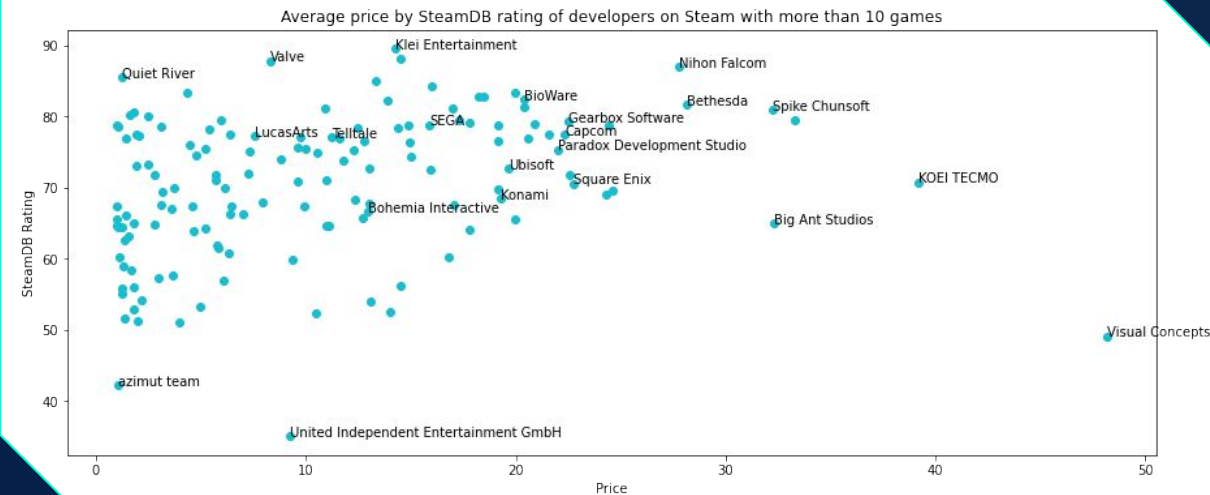
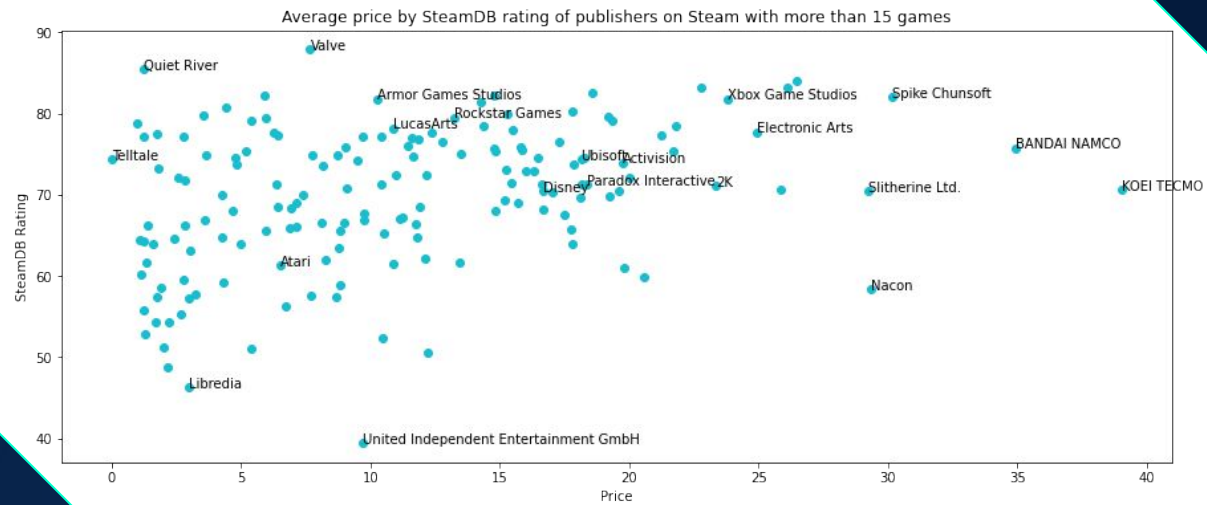


Image [12]: Chart of average price and average rating for each publisher

Image [11]: Chart of average price and average rating for each developer

# POPULARITY, RATING AND PRICE FOR 200 MOST POPULAR INDIE GAME

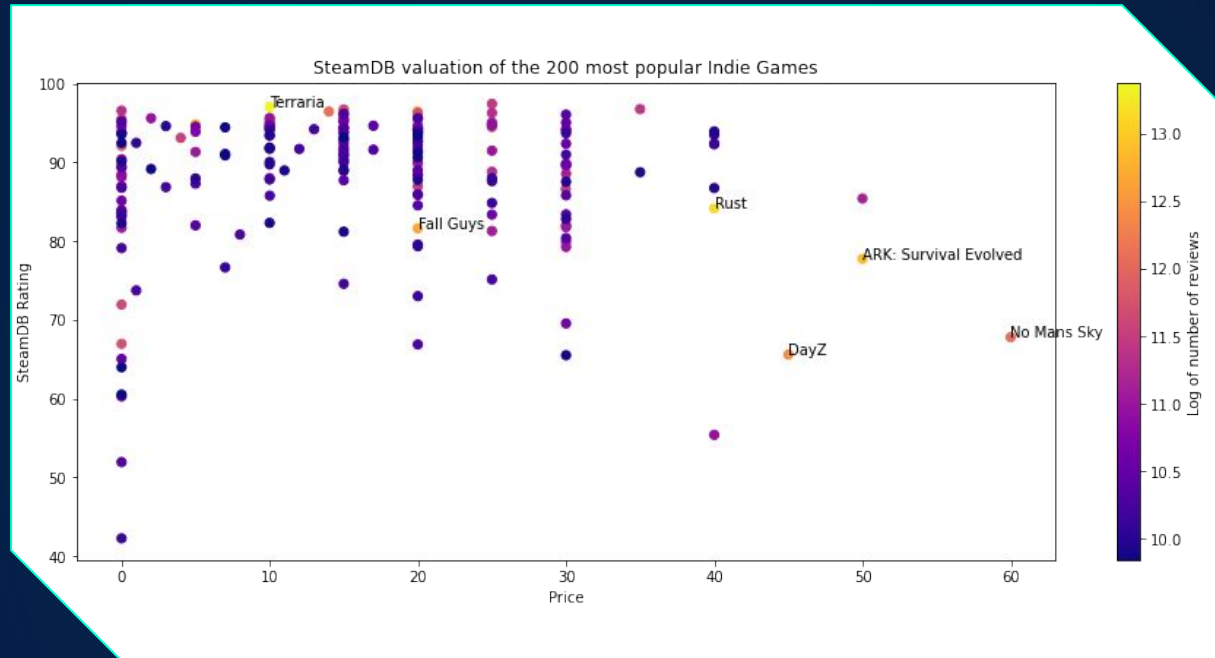


Image [14]: Chart of the relation between rating and price for the 200 most popular games, with an heatmap representing the popularity, calculated as log of number of total reviews.

# 04

## REGRESSION

---





# CHOOSING THE NUMBER OF TAGS

## R SQUARED ANALYSIS

In order to linearly regress we had to first decide on how many tags were to be selected as features, and to determine this we used  $R^2$  measures.

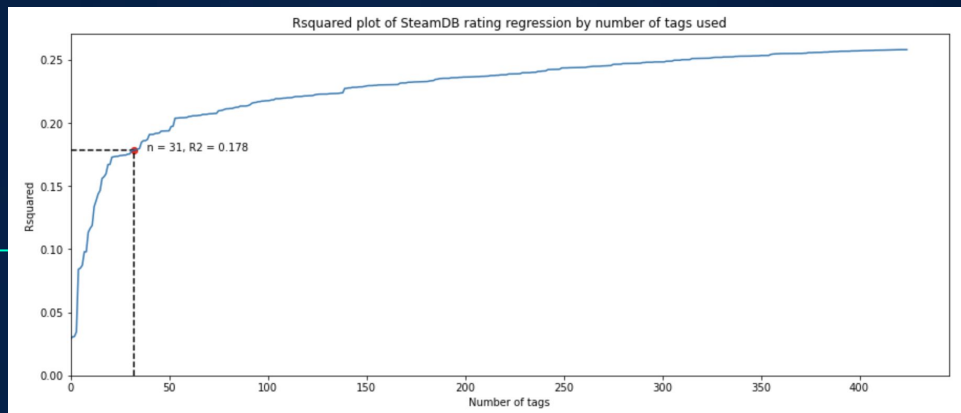


Image [9]: R squared in function of the tags number

# LINEAR REGRESSION COEFFICIENTS

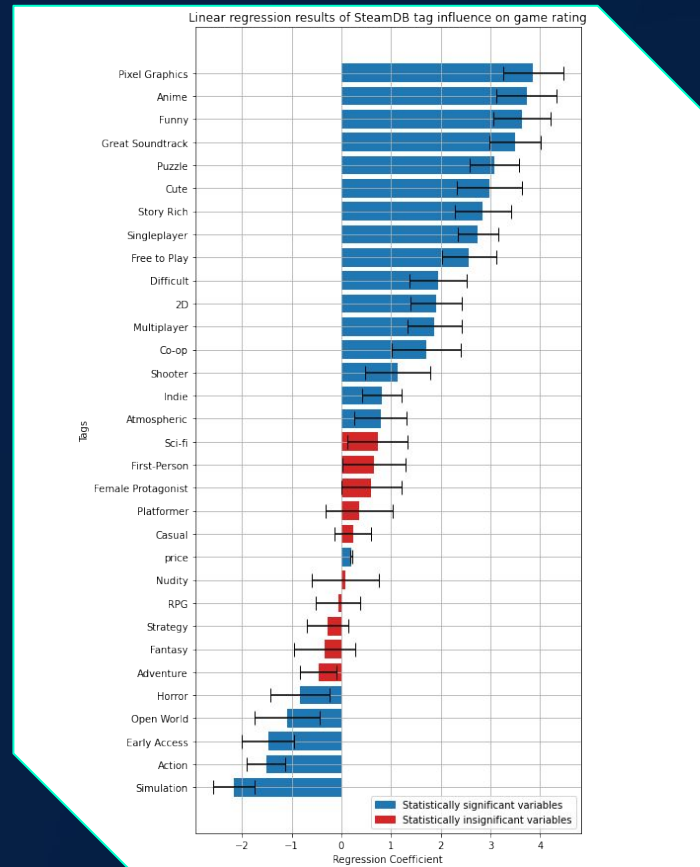




Image [10]: Regression coefficients

# 05

## CONCLUSION



- 
- 
- Many useful pieces of information: most popular tags, highest rated tags, highest priced tags
  - Price contributes weakly to rating
  - The model is not predictive, but helps to understand a tag influence
- 
- 



# ROOM FOR IMPROVEMENT

---

- Current model brings  $R^2$  to only 0.25
  - Rating histogram and regression residuals not normally distributed
  - Disconsidered games' rating and price evolution over time
  - Tags are user attributed
- 





# QUESTIONS?

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**.

Please keep this slide for attribution.