Game's tags, price and reviews

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

SUMMARY

01 INTRODUCTION AND MOTIVATION

02 SCRAPING AND CLEANING

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.

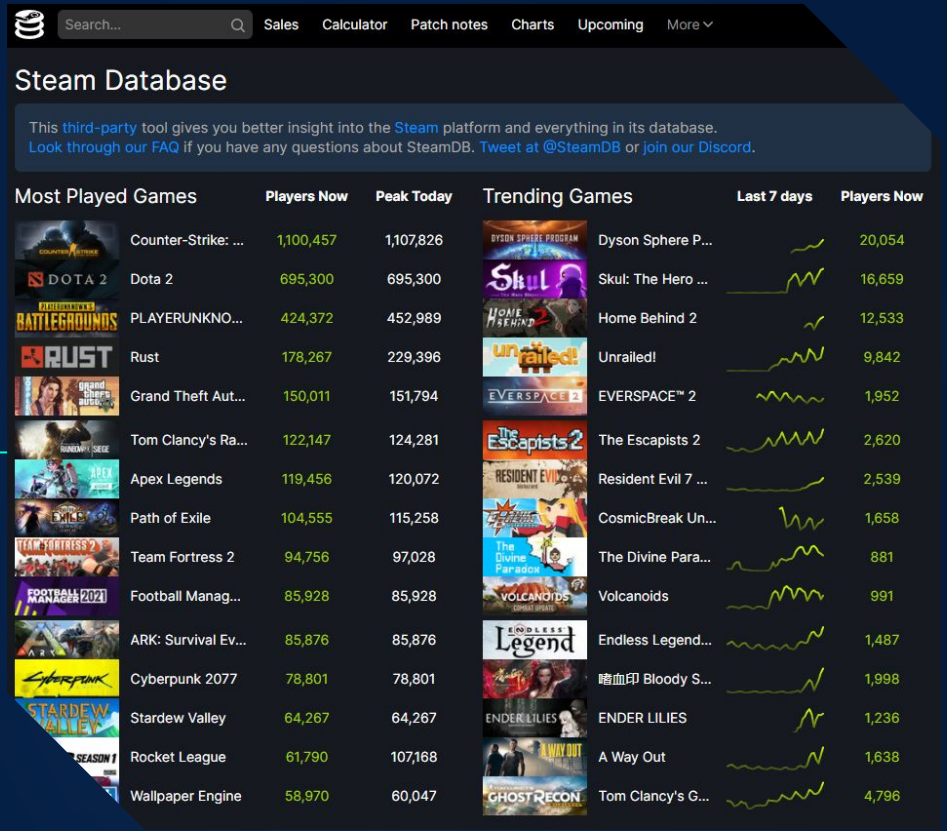

Image [1]: SteamDB website screenshot

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%

Image [2]: SteamDB Top Rated games website screenshot

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

PAGE SAMPLE

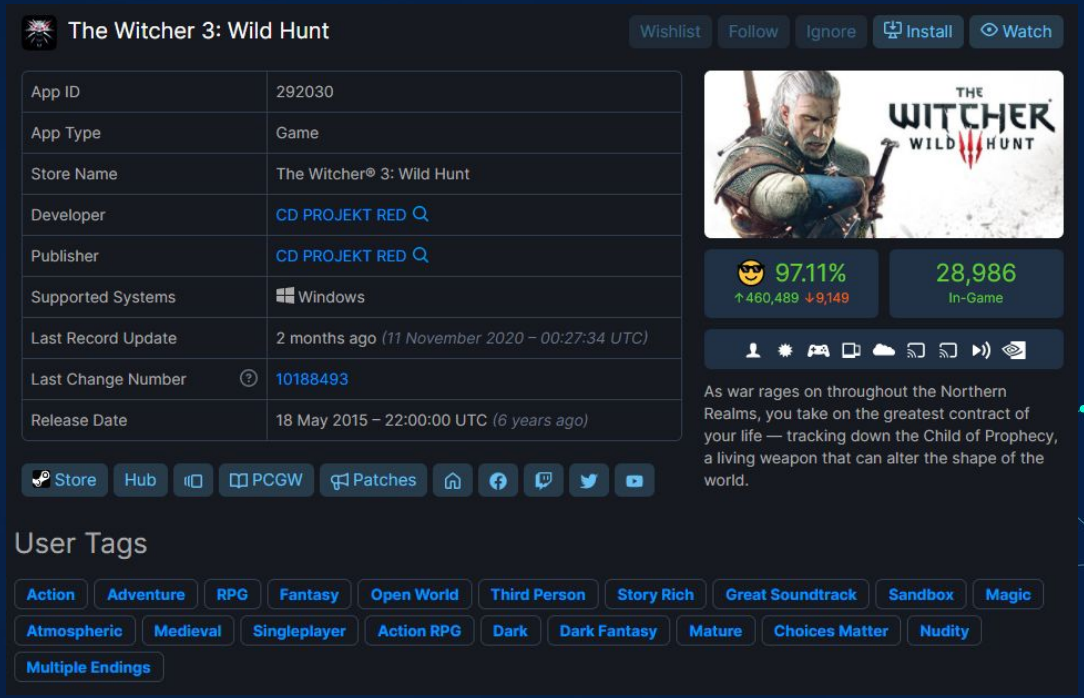

Image [2]: The Witcher 3 page sample screenshot

OUR NUMBERS

4 Days scraping

426 Features

40.000 Samples

DATA CLEANING

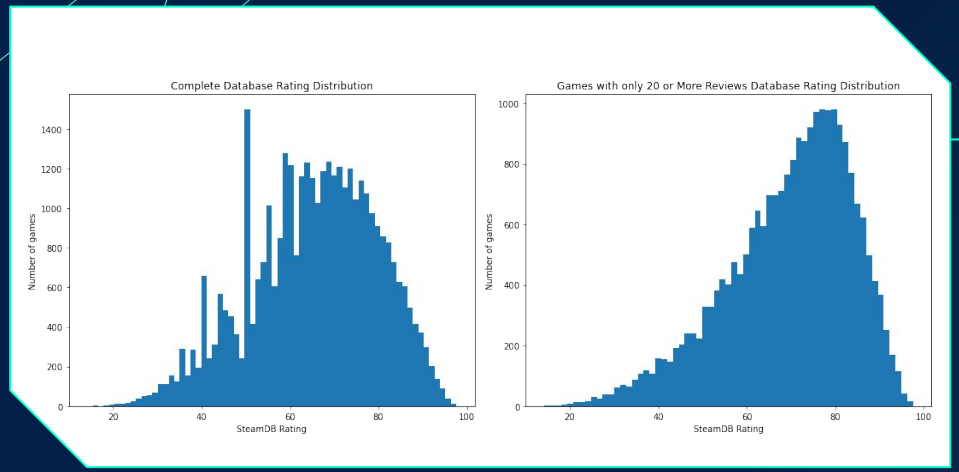

Image [1]: SteamDB website screenshot

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

03

GRAPHICAL ANALYSIS

MOST FREQUENT TAGS

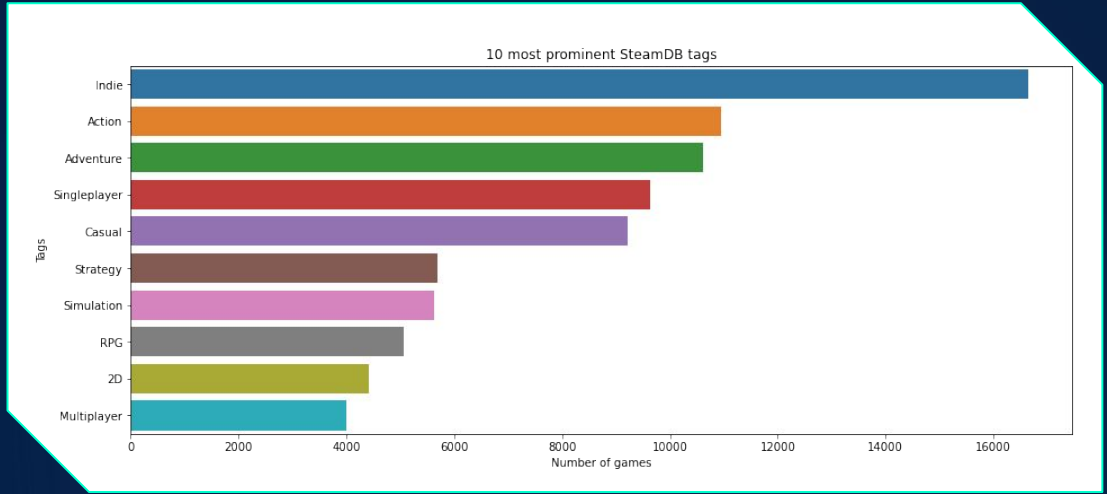

Image [3]: Chart of most prominent SteamDB tags

AVERAGE RATING OF MOST POPULAR

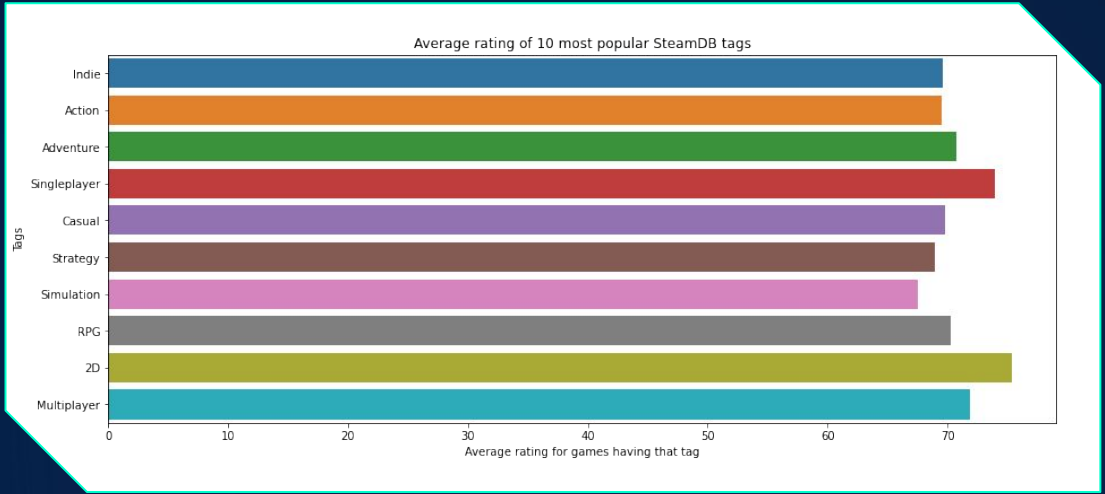

Image [4]: Chart of average rating of 10 most popular tags

AVERAGE RATING OF MOST WELL RANKED TAGS

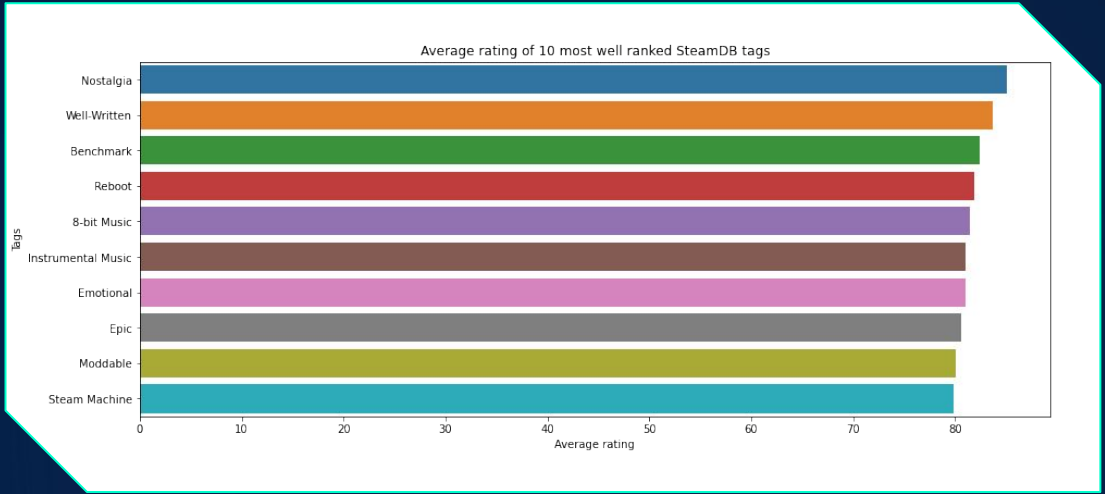

Image [5]: Chart of average rating of most well ranked tags

AVERAGE PRICE FOR MOST POPULAR TAGS

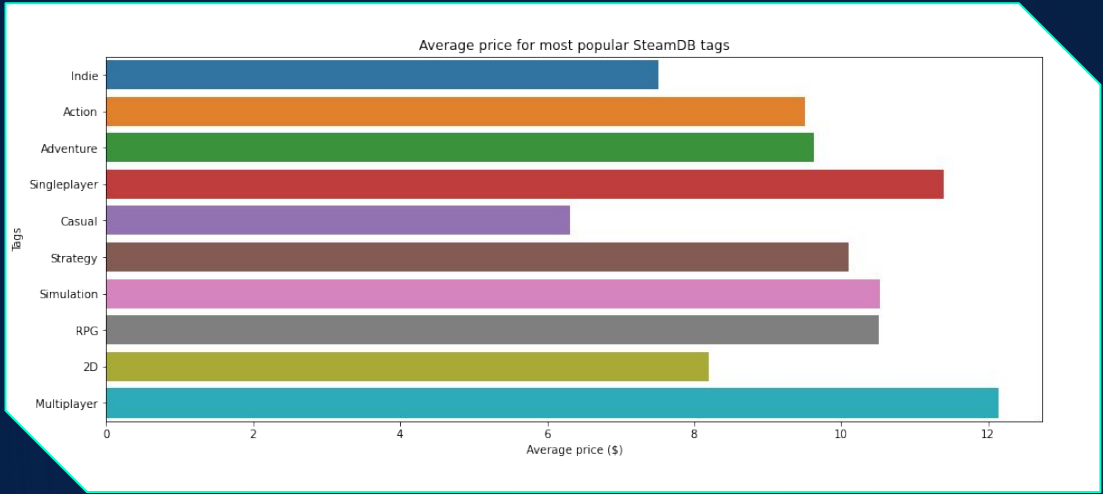

Image [6]: Chart of average price

AVERAGE PRICE FOR MOST EXPENSIVE TAGS

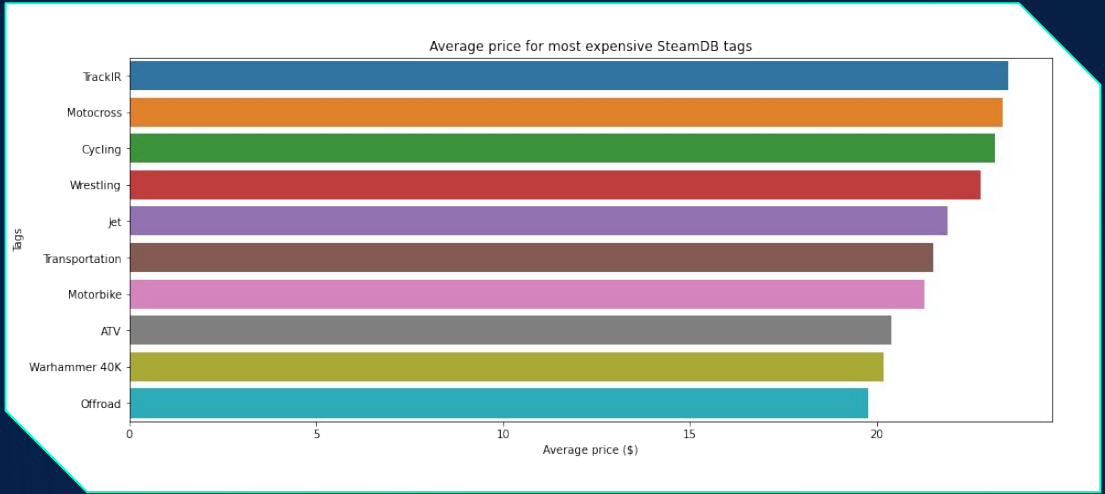

Image [6]: Chart of average price for most expensive tags

RELATION AVERAGE PRICE DEVS AND PUBLISHERS

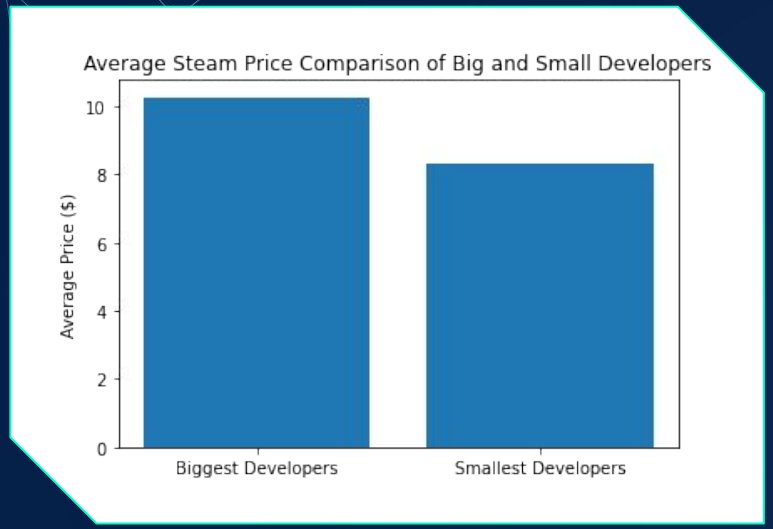

Image [8]: Chart of average price for 30 biggest developers and 30 smallest developers

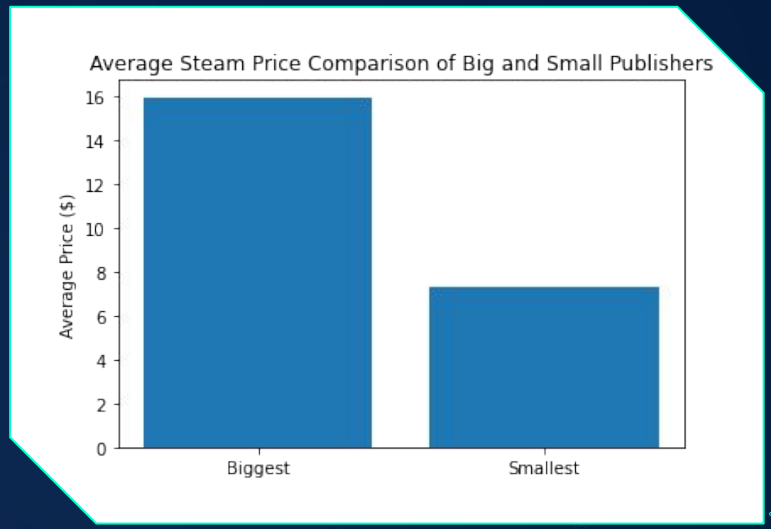

Image [8]: Chart of average price for 30 biggest publishers and 30 smallest publishers

RELATION AVERAGE RATING AND AVERAGE PRICE

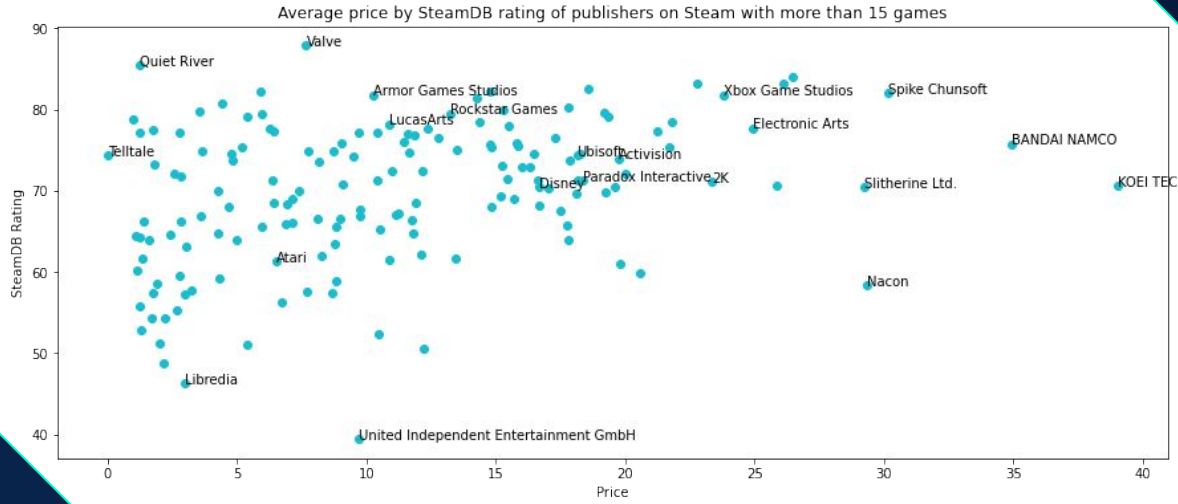

Image [12]: Chart of average price and average rating for each publisher

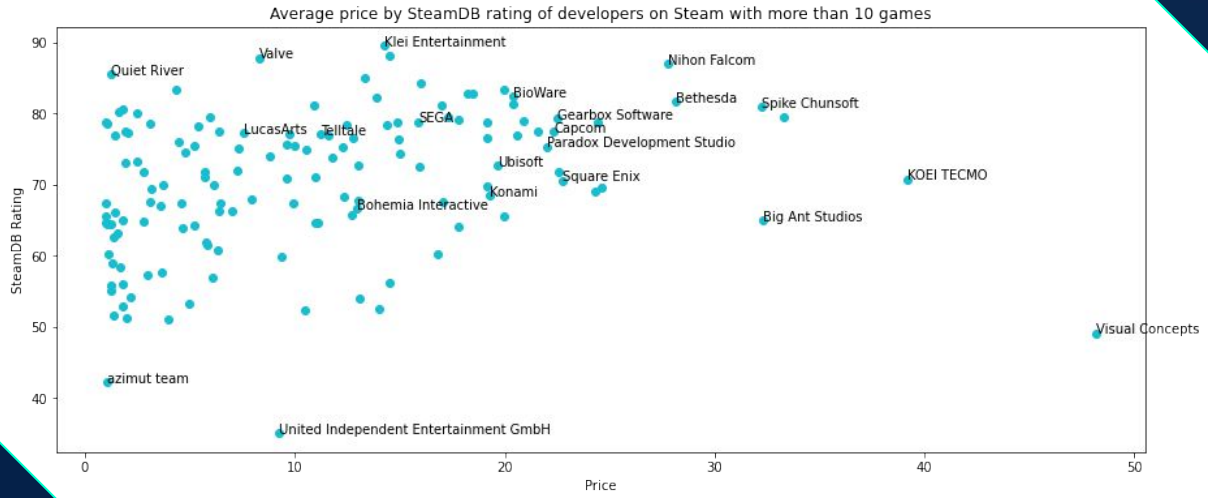

Image [11]: Chart of average price and average rating for each developer

POPULARITY, RATING AND PRICE FOR 200 MOST POPULAR INDIE GAME

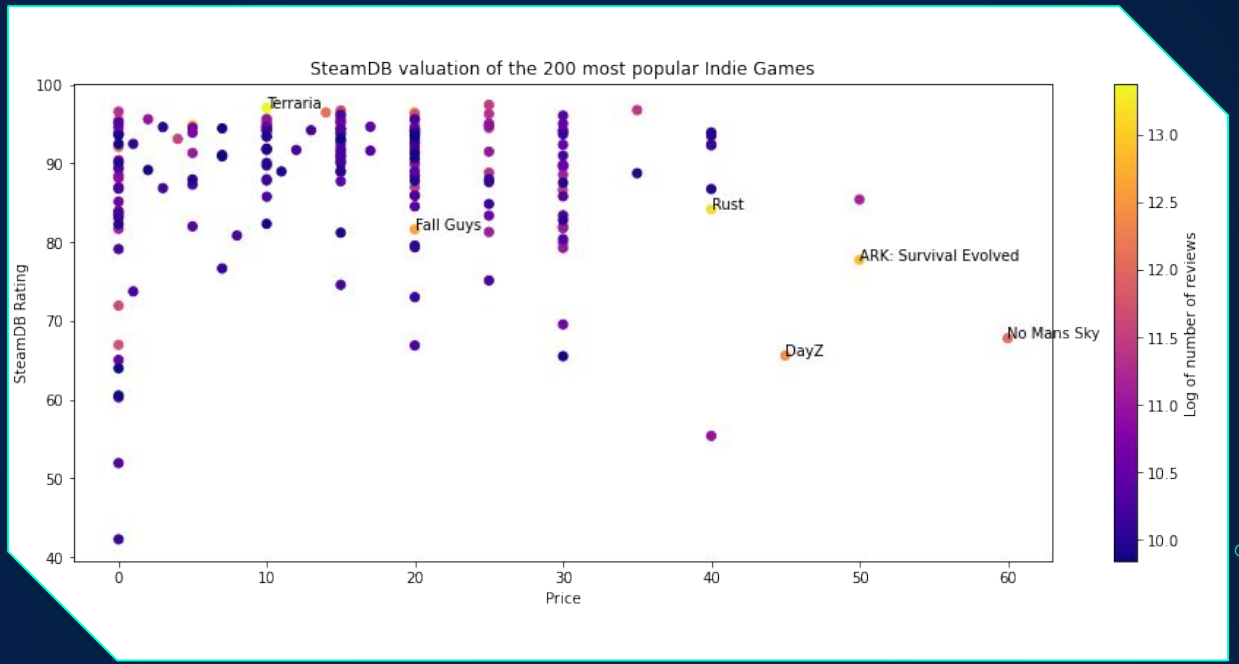

Image [14]: Chart of the relation between rating and price for the 200 most popular games, with an heatmap repreenting the popularity, calculated as log of number of total reviews.

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.

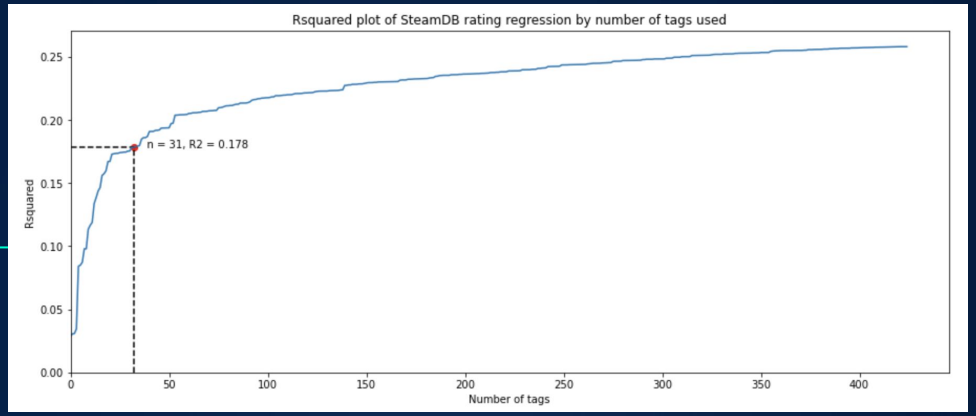

Image [9]: R squared in function of the tags number

LINEAR REGRESSION COEFFICIENTS

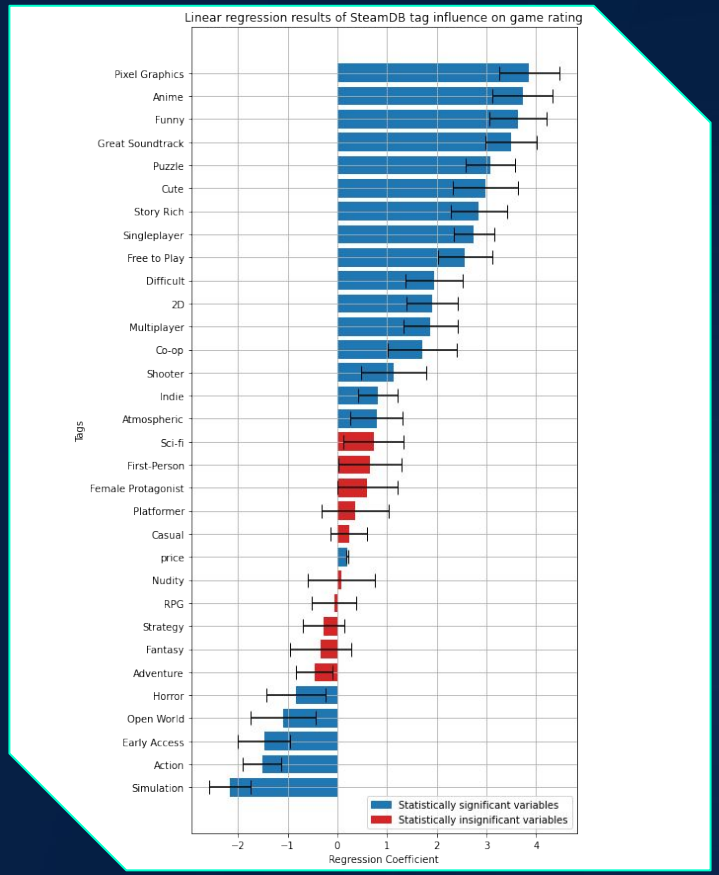

Image [10]: Regression coefficients

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.