Analysis of Steam Video Games

Group 17:
Jiaxuan Zhang, Xiaoyang Pan,
Yiming Ren, Xiangtian Li, Mohan Li

Motivation and Objective

- Steam is the world's most popular PC Gaming hub, with over 6,000 games and a community of millions of gamers
- The current recommendation for games is sometimes not very accurate

Goals

- Visualize and analyze the data of Steam games
- Cluster games with methods like k-means
- Recommendation systems for the dataset

Dataset

- Steam Store Games
  - includes categories, genre, price, release date, owners, and more

- Steam Video Game and Bundle Data
  - includes all games purchased by users, and information about which games were bundled together

Data Preprocessing

Data Preprocessing

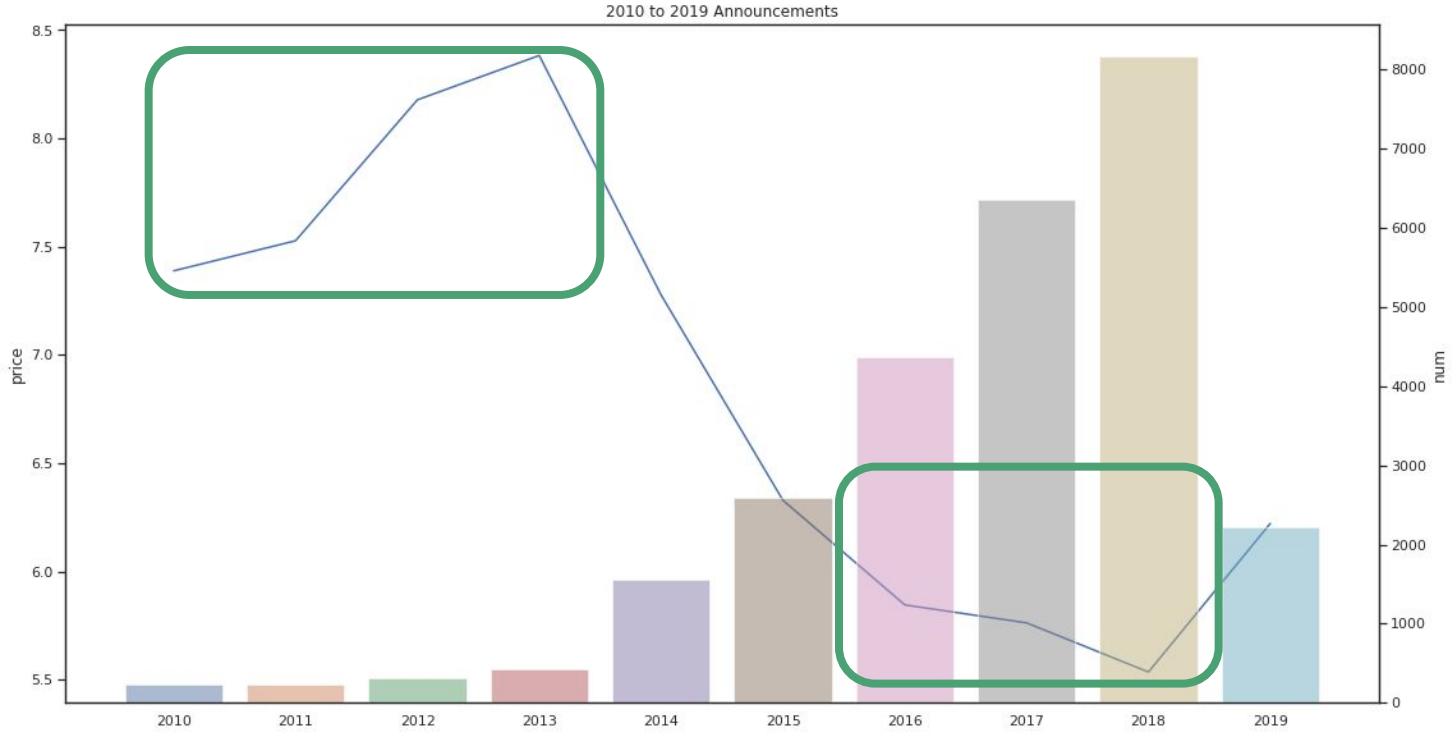
- Change Release Date Format: Only keep year info
  - 2010-07-12 ➜ 2010

- Change Genre/Categories Format: separate values by semicolon
  - Action;Adventure ➜ (Action, Adventure)

- Change Owners Format: replace the range by the middle value of the range
  - 10000000-20000000 ➜ 15000000

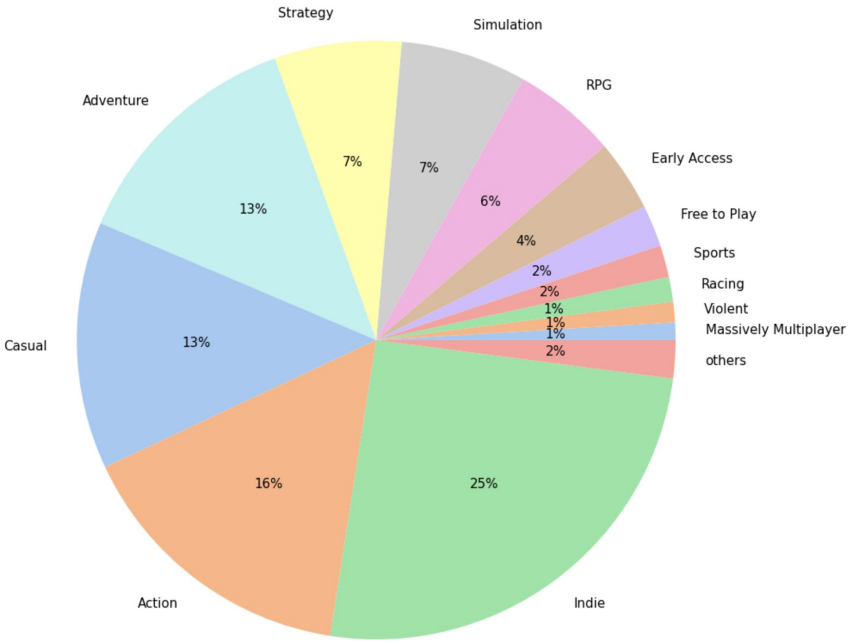
Visualization of the dataset

Average price of game drops over time, and it is negatively correlated to number of games of that year.

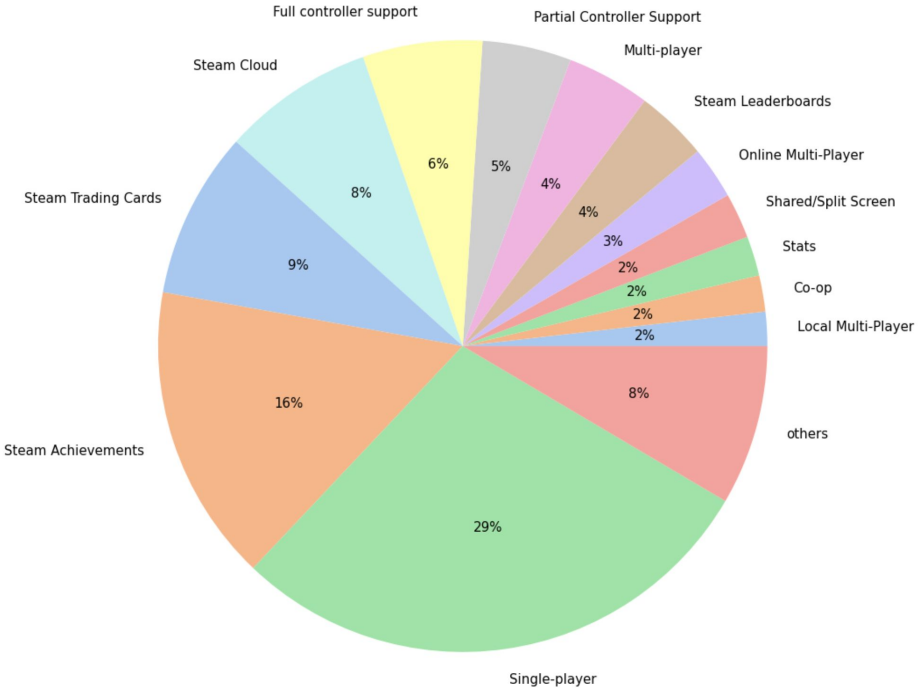


Correlation Matrix

Group Games by Genres



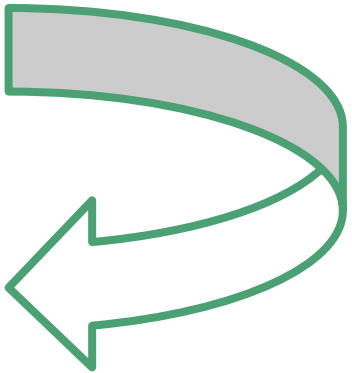
Group Games by Categories



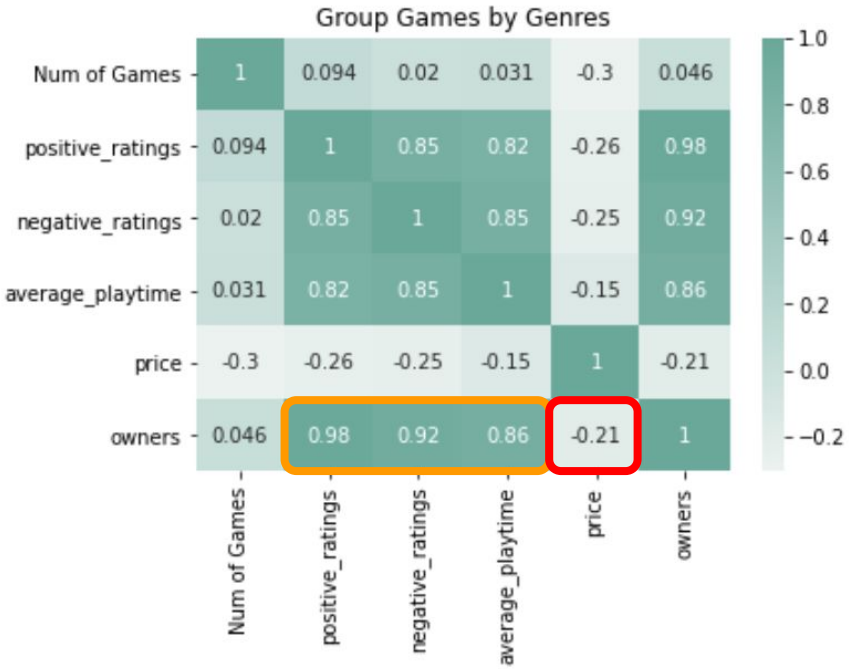
Rearrange Data

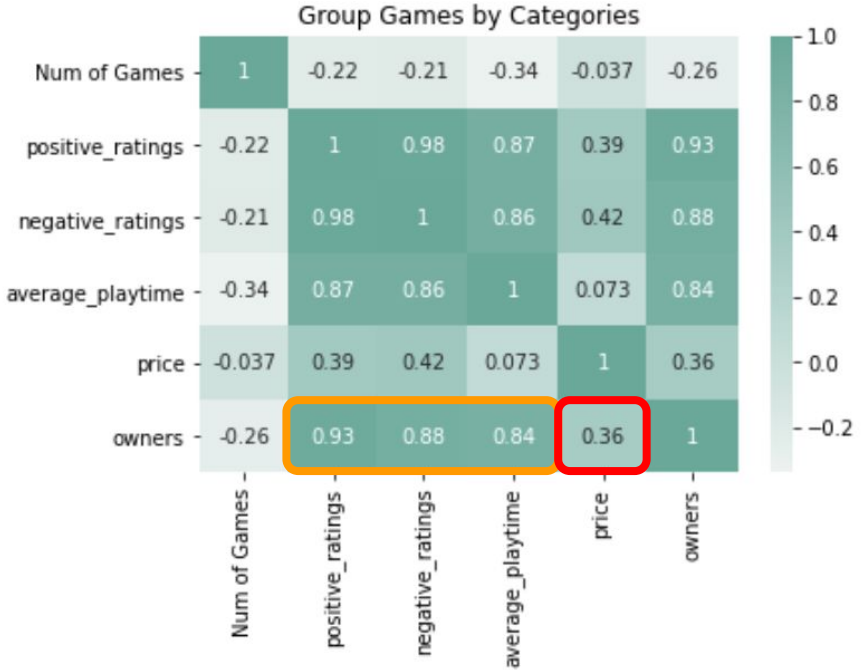
	name	genres	positive_ratings	negative_ratings	average_playtime	price	owners
0	Counter-Strike	Action	124534	3339	17612	7.19	15000000
1	Team Fortress Classic	Action	3318	633	277	3.99	7500000
2	Day of Defeat	Action	3416	398	187	3.99	7500000
3	Deathmatch Classic	Action	1273	267	258	3.99	7500000
4	Half-Life: Opposing Force	Action	5250	288	624	3.99	7500000
...	...	...	...	...	...	...	...
27070	Room of Pandora	Adventure;Casual;Indie	3	0	0	2.09	10000

genres	Num of Games	positive_ratings	negative_ratings	average_playtime	price	owners
Tutorial	1.0	1.000000	1.000000	0.000000	3.000000	10000.000000
Documentary	1.0	1.000000	1.000000	0.000000	3.000000	10000.000000
Accounting	6.0	6.500000	4.166667	0.000000	2.666667	14166.666667
Photo Editing	12.0	84.000000	30.916667	429.583333	14.083333	33333.333333
Game Development	17.0	102.235294	37.529412	116.000000	61.000000	39705.882353



New Correlation Matrices





Findings

- Price is not correlated with number of owners (sales)

- Not only positive rating, but also negative rating are correlated with number of owners

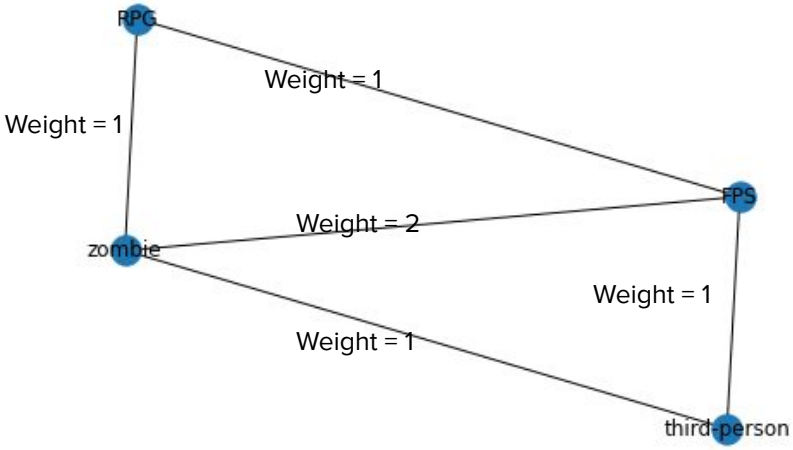
- Genres of game help to determine similarities between games

Clustering of Games

Clustering

Game A: FPS; zombie; RPG

Game B: FPS; zombie; third-person



K-means clustering:

Higher coefficient -> More likely to be in a cluster

K-Clique

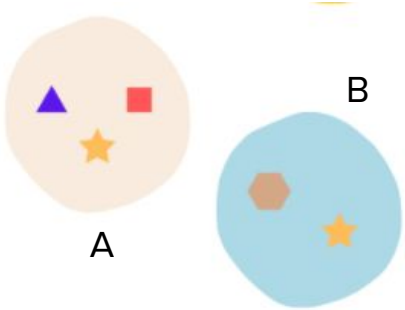

Indie
Casual
Action
Simulation
Early Access
Adventure
Strategy
RPG
Free to Play
Sports
Multiplayer
Puzzle
Racing
Arcade
Open World
VR
Singleplayer
Platformer
Shoot 'Em Up
Survival
Zombies
Massively Multiplayer
Hidden Object
Visual Novel
Sexual Content

Recommendation System

Recommendation System

- Purpose:
  - Recommend users with games they may be more intrigued
  - Targeted advertising is more effective, achieve more selling for steam
- Methods:
  - Find the inner connection of games from all dimensions
  - **Jaccard Similarity** between two games is the base of our analysis

jaccard similarity:



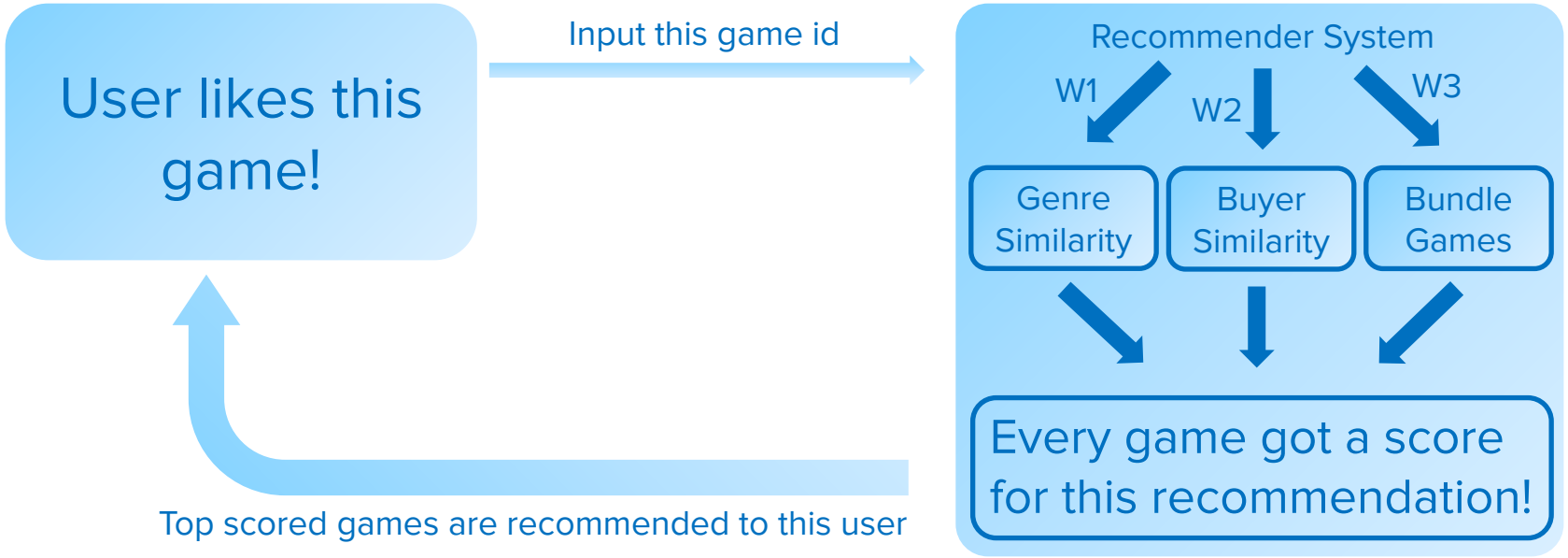
$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

$$\frac{|\bigstar|}{|\blacktriangle \bigstar \blacksquare \hexagon|} = \frac{1}{4}$$

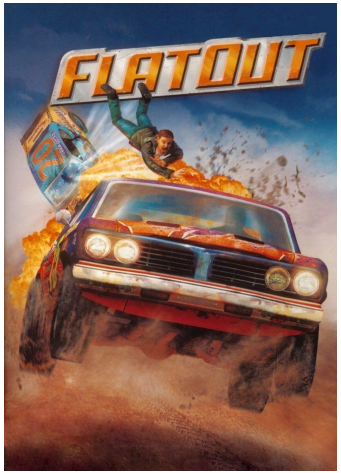
Three features

Similarities of genres	Do these two games have similar game type? (Horror? Action?)
Similarities of buyers	People who bought this game also bought ...
Bundle selling	Steam has bundled them as a pack, maybe they are related

How is a game recommended?



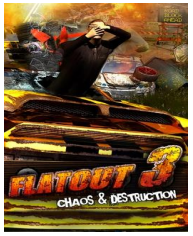
A peek at the result..



FlatOut

1.FlatOut: Ultimate Carnage

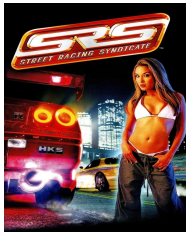
2.FlatOut2



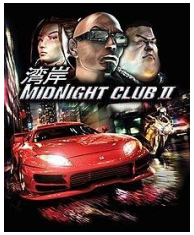
3.FlatOut3

4.Crazy Taxi

5.FlatOut4



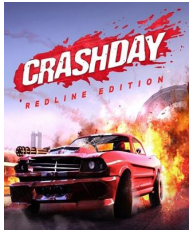
6.Street Racing Syndicate



7.Midnight club 2

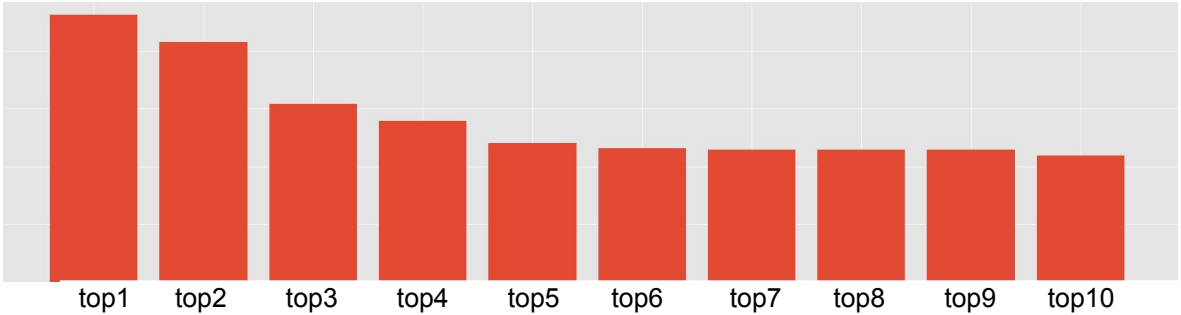
8.Ridge Racer Unbounded

9.Need for Speed Undercover



10.Crashday redline edition

Recommendation score



Crazy car games are recommended!!

Github Page:

https://github.com/lxtrichard/ECE-143-Analysis-of-Steam-Video-Games

Thanks to Prof. Unpingco and Samveed