

AI 705.FINAL PROJECT

STEAM GAME RECOMMENDATION SYSTEM

MAYANK CHADHA (IMT2020045)

SHRIDHAR SHARMA (IMT2020065)

DARSHAK JIVRAJANI (IMT2020119)

PREM SHAH (IMT2020044)



01 - NOVELTY OF THE PROBLEM

POPULARITY OF GAMING

The gaming industry has become one of the most popular forms of entertainment around the world. People are always on the lookout for new games to play.

SOCIAL IMPACT

Gaming can have a positive impact on mental health and social connections. A recommendation system can contribute to building a stronger gaming community and improve people's well-being.

PERSONALIZATION

Every gamer has different preferences, and a recommendation system can help personalize their gaming experience by suggesting games that are more likely to appeal to their individual tastes.

02 - DATASET CREATION

We obtained Steam games data by utilizing multiple APIs and performing data manipulation.

Steam Spy API

Steam Official API

The selection of these APIs is based on two key criteria: the abundance and diversity of data available in them, which will be exemplified in upcoming slides, and their frequent updates on a weekly basis.

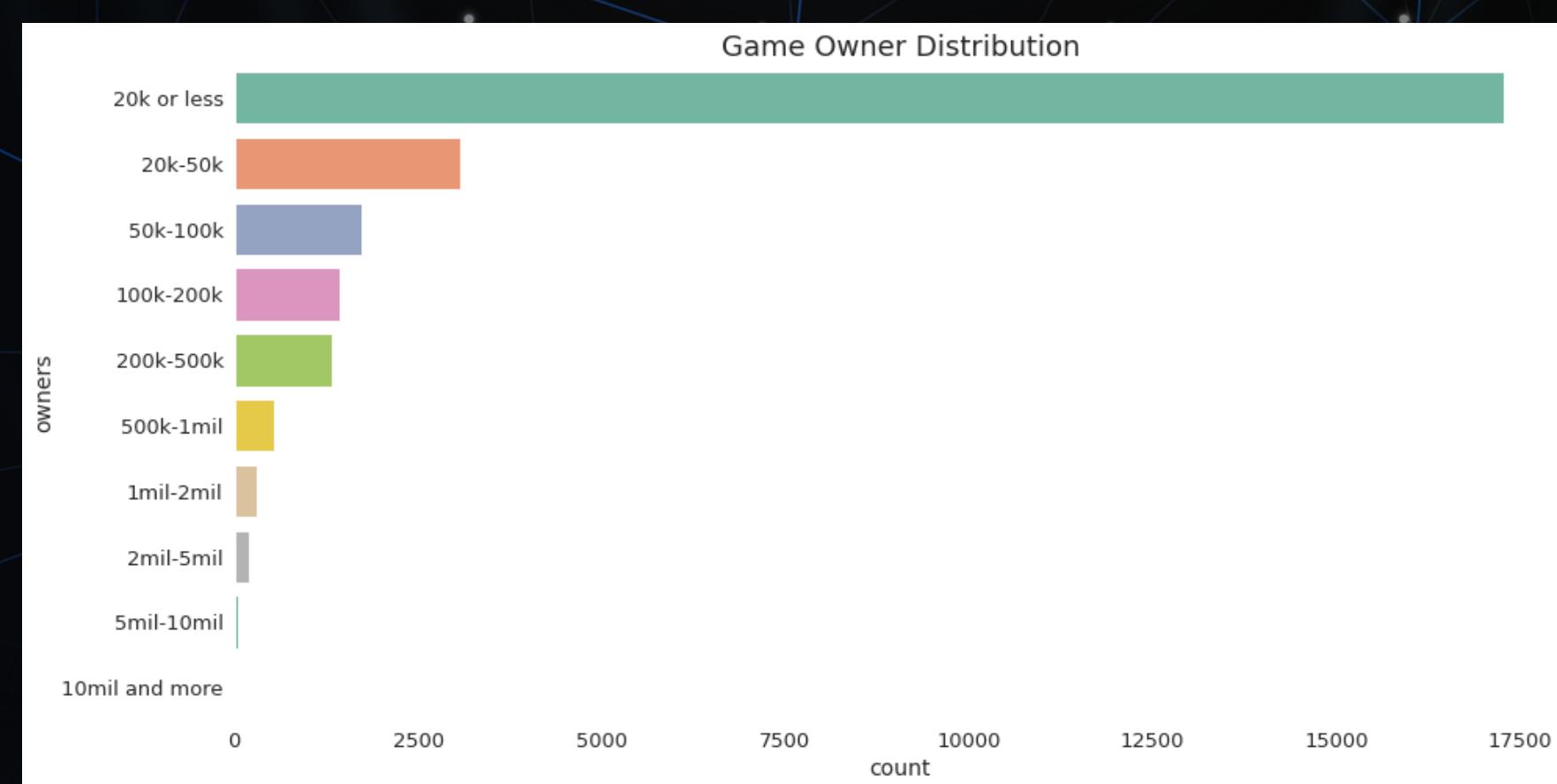
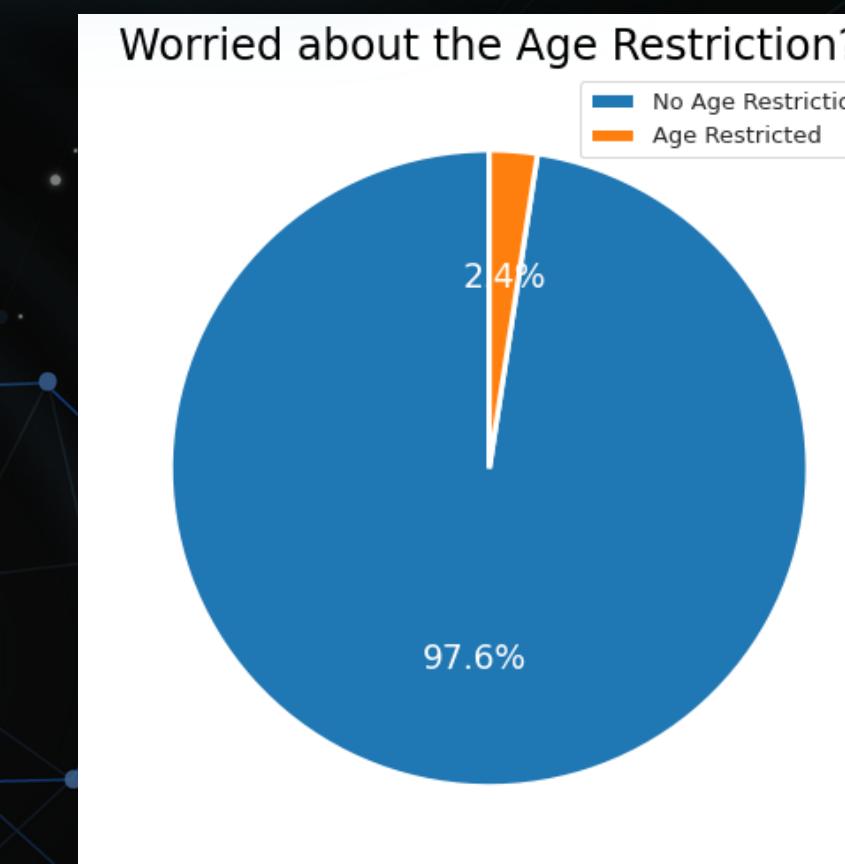
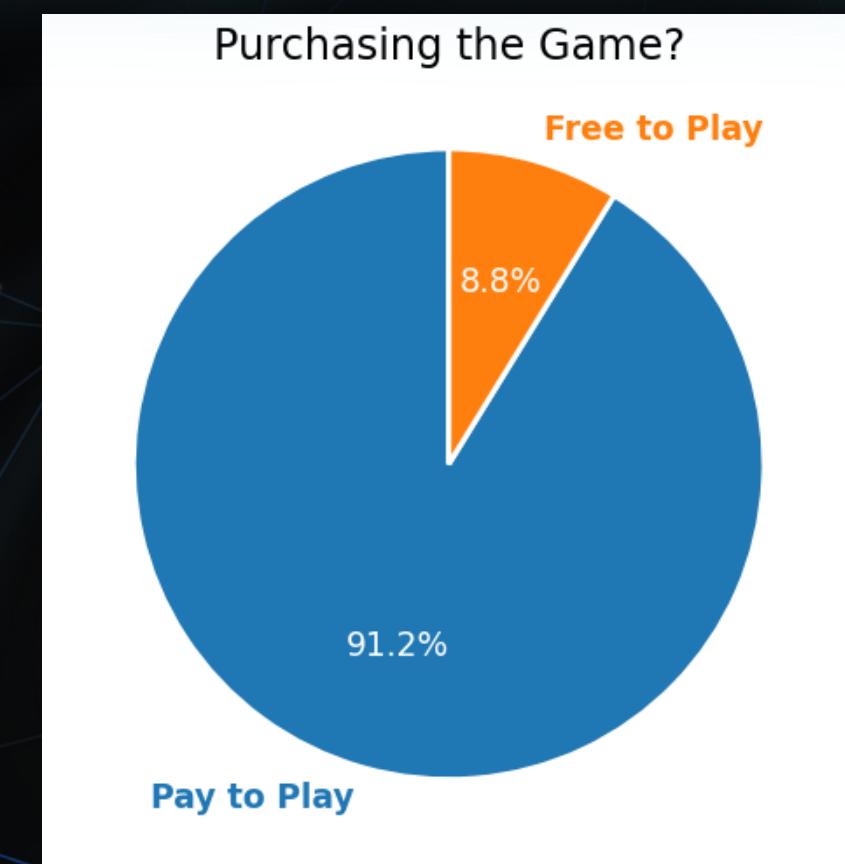
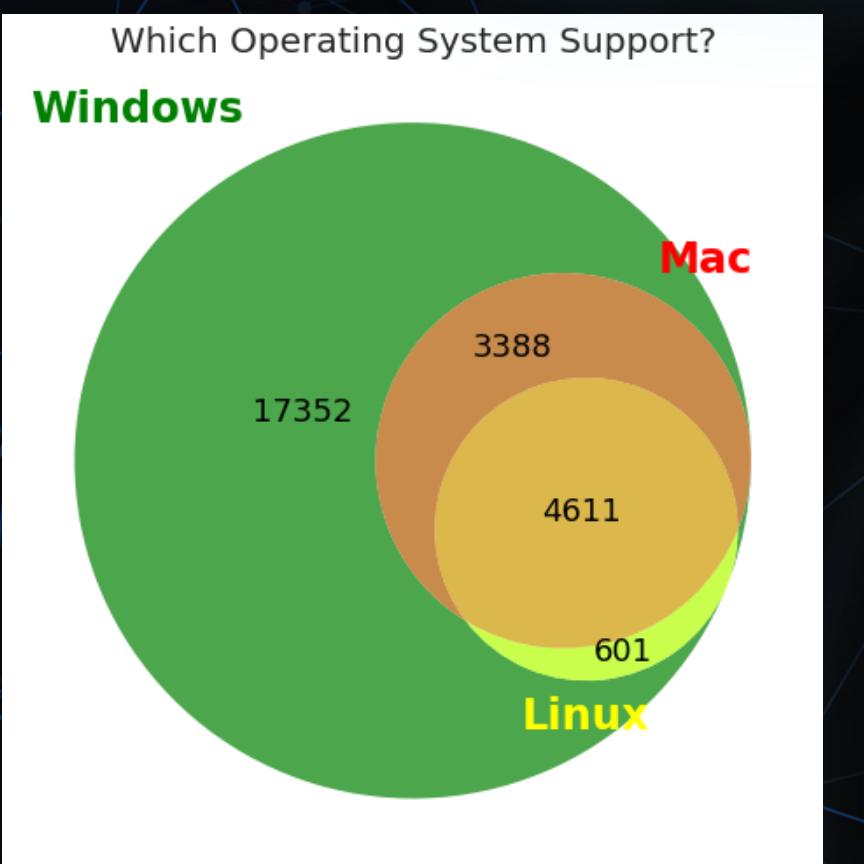
COLUMNS IN THE DATASET

The dataset underwent a thorough cleaning process to remove extraneous columns, prioritizing those relevant to recommendation analysis. The remaining columns deemed useful for exploratory data analysis were retained. The resulting cleaned datasets were merged to obtain the final dataset for further analysis.

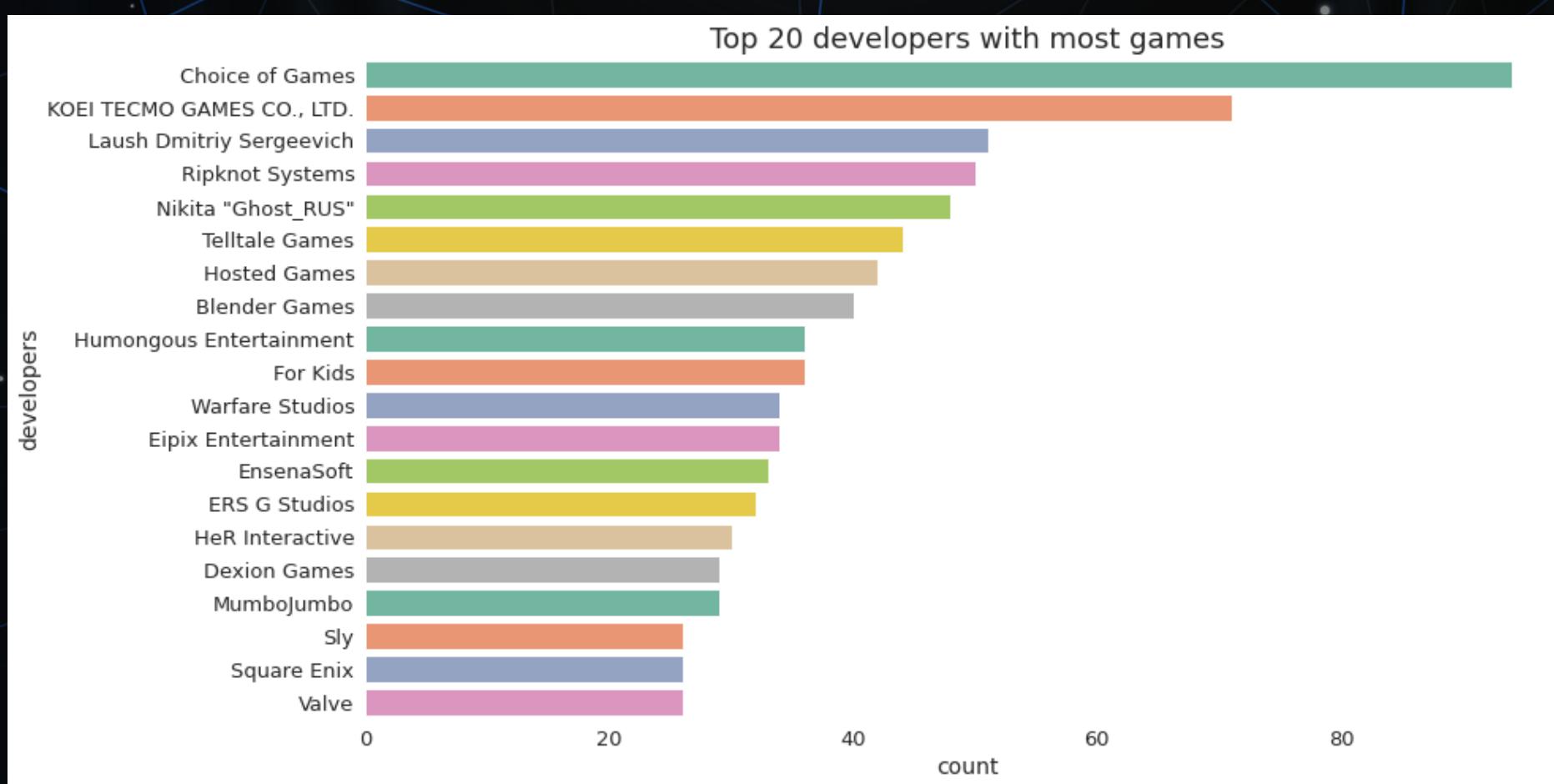
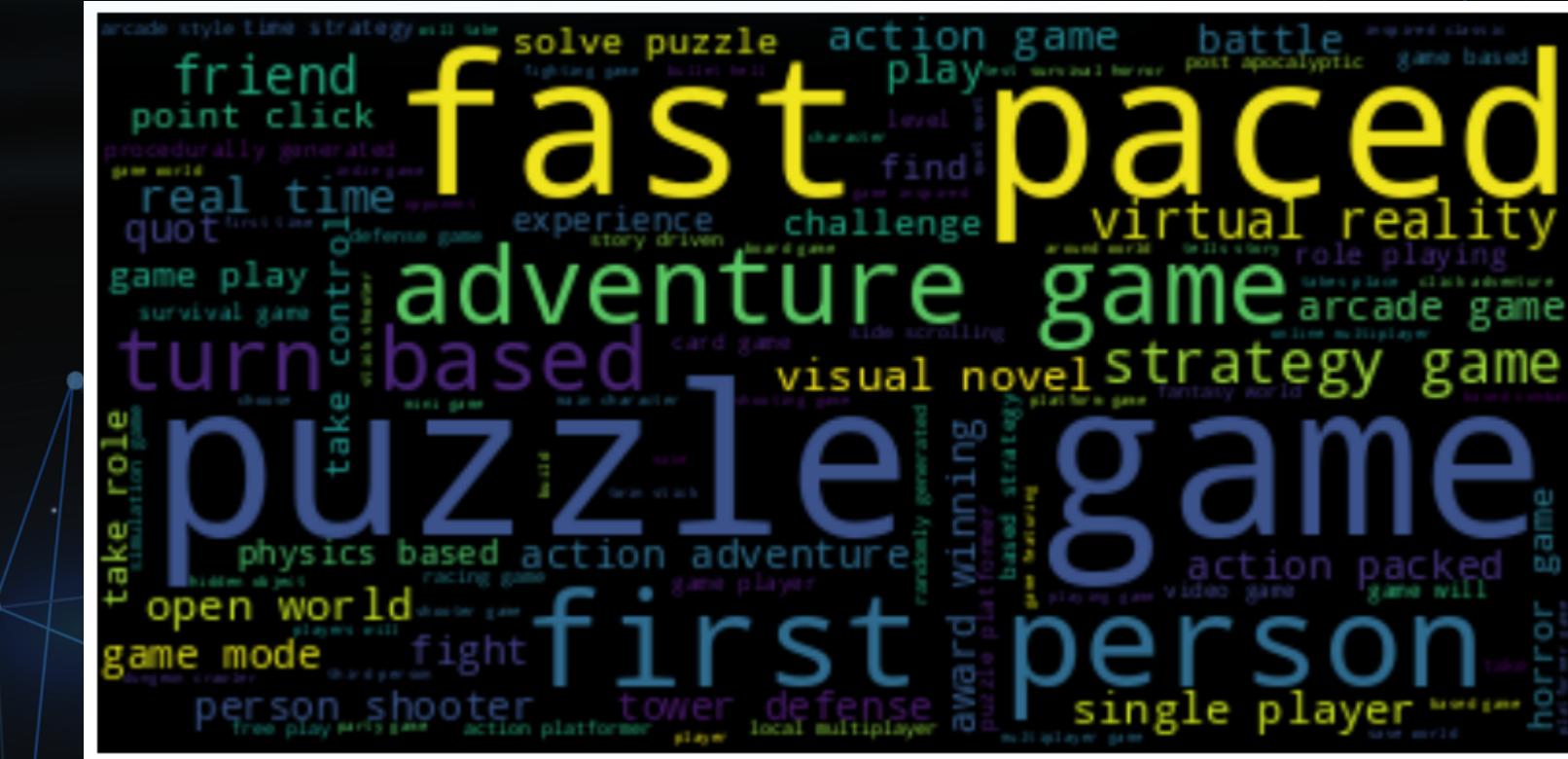
```
1 steam_data.columns
✓ 0.0s
Index(['type', 'name', 'steam_appid', 'required_age', 'is_free',
       'controller_support', 'dlc', 'detailed_description', 'about_the_game',
       'short_description', 'fullgame', 'supported_languages', 'header_image',
       'website', 'pc_requirements', 'mac_requirements', 'linux_requirements',
       'legal_notice', 'drm_notice', 'ext_user_account_notice', 'developers',
       'publishers', 'demos', 'price_overview', 'packages', 'package_groups',
       'platforms', 'metacritic', 'reviews', 'categories', 'genres',
       'screenshots', 'movies', 'recommendations', 'achievements',
       'release_date', 'support_info', 'background', 'content_descriptors'],
      dtype='object')

1 steam_spy.columns
✓ 0.0s
Index(['appid', 'name', 'developer', 'publisher', 'score_rank', 'positive',
       'negative', 'userscore', 'owners', 'average_forever', 'average_2weeks',
       'median_forever', 'median_2weeks', 'price', 'initialprice', 'discount',
       'languages', 'genre', 'ccu', 'tags'],
      dtype='object')
```

EXPLORATORY DATA ANALYSIS - 1



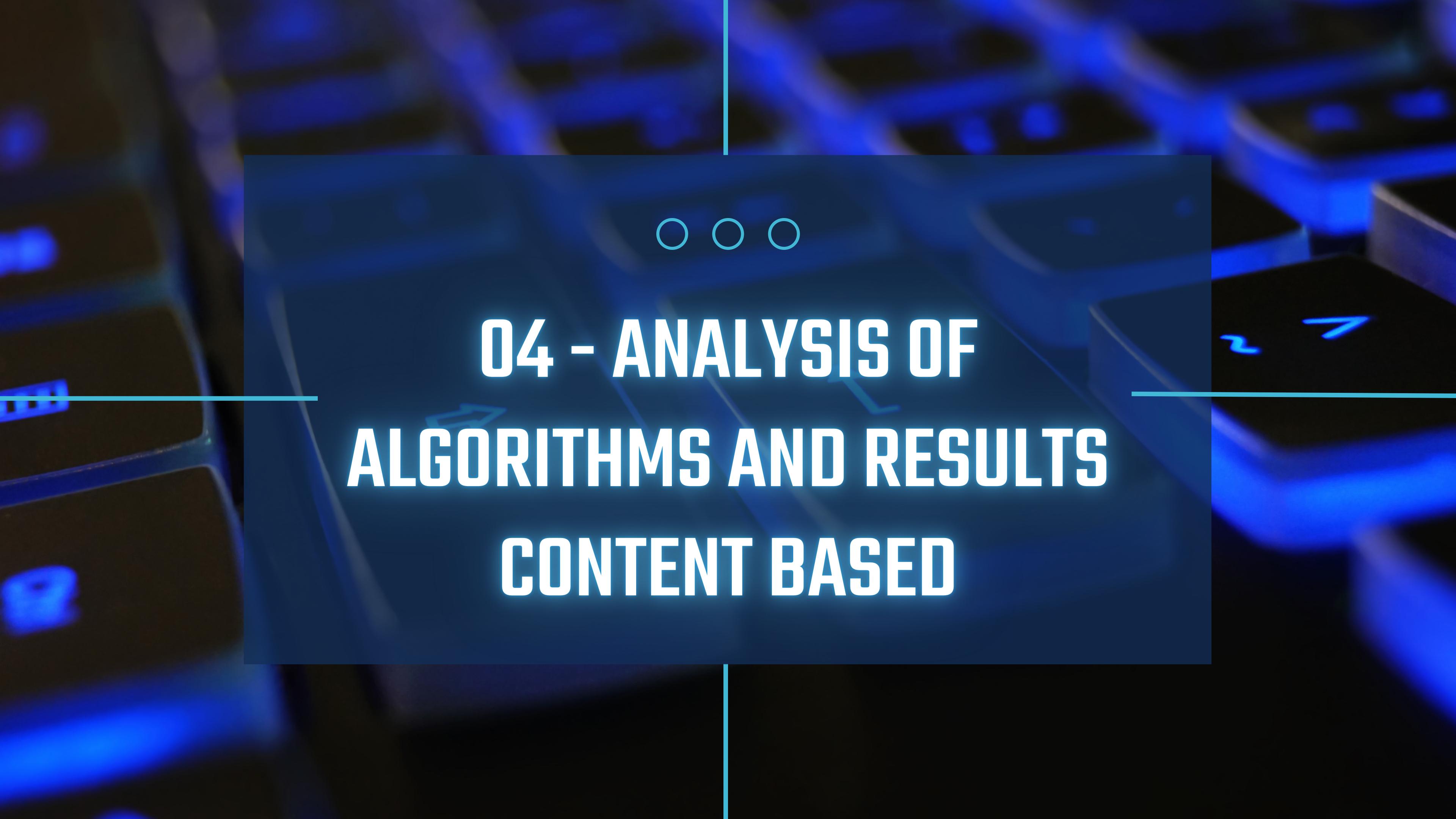
EXPLORATORY DATA ANALYSIS - 2



03 -ALGORITHM SELECTION

The initial impetus was derived from the study of class materials on Word2Vec and the efficacy of transformers.

This led to the consideration of content-based filtering and collaborative-based filtering using various aspects of a game, including its description, genre, description and developer.



000

04 - ANALYSIS OF ALGORITHMS AND RESULTS CONTENT BASED

COUNTVECTORIZER VS TF-IDF

CountVectorizer (based on Bag of Words) simply counts the number of times a word appears in a document (using a bag-of-words approach), while TF-IDF Vectorizer takes into account not only how many times a word appears in a document but also how important that word is to the whole corpus.



TF-IDF

Term x within document y

$$w_{x,y} = tf_{x,y} \times \log \left(\frac{N}{df_x} \right)$$

$tf_{x,y}$ = frequency of x in y
 df_x = number of documents containing x
 N = total number of documents

$$TF - IDF = TF(t, d) * IDF(t)$$

$TF(t, d)$ = Number of Times Term t Appears in doc, d

We know that's just our CountVectorizer

$$TF - IDF = CountVectorizer(t, d) * IDF(t)$$

COUNT VECTORIZER + COSINE SIMILARITY

We experimented with combinations of name, description, publishers, tags, genres and categories to generate embeddings using CountVectorizer.

Recommendation for Grand Theft Auto V	
Recommendations using Count Vectorizer	
14028	Grand Theft Auto: Vice City – The Definitive E...
12420	Grand Theft Auto 2
14027	Grand Theft Auto: San Andreas – The Definitive...
12178	Grand Theft Auto IV: The Complete Edition
3559	Red Dead Online
5062	Super Bomberman R
4110	Dead Rising 4
9438	Grand Theft Auto
14089	The Last of Us™ Part I
6411	Portal 2
4368	Move or Die
4428	Mini World: Block Art
4017	Magicka
9827	Red Dead Redemption 2
Name: Name, dtype: object	

TF-IDF + COSINE SIMILARITY

Recommendation for Grand Theft Auto V	
Recommendations using Tf-idf	
None	
14028	Grand Theft Auto: Vice City – The Definitive E...
12420	Grand Theft Auto 2
14027	Grand Theft Auto: San Andreas – The Definitive...
9438	Grand Theft Auto
12178	Grand Theft Auto IV: The Complete Edition
4074	L.A. Noire: The VR Case Files
3559	Red Dead Online
2332	Tokyo 42
7639	Rustler (Grand Theft Horse)
1917	Manhunt
9827	Red Dead Redemption 2
7854	弹幕音乐绘 ~风雷幻奏曲~ / Barrage Musical ~A Fantasy of...
6266	BROKE PROTOCOL: Online City RPG
10101	Grand Ages: Medieval
Name: Name, dtype: object	

Count Vectorizer and Tf-Idf gave almost similar recommendations. This is because both methods don't account for context of description.

WORD 2 VEC

Next, we tried using Word2Vec to generate embeddings and combined with Tf-Idf.

Recommendation for Grand Theft Auto V	
Recommendations using TFIDF + Word2Vec	
322	Grand Theft Auto: Episodes from Liberty City
311	Grand Theft Auto III
319	Grand Theft Auto 2
20875	Samp RP
321	Grand Theft Auto IV
528	Grand Ages: Rome
16506	AEGYPTUS
1416	Omerta - City of Gangsters
312	Grand Theft Auto: Vice City
1686	Cities in Motion 2
Name: name, dtype: object	

BERT TRANSFORMER + COSINE SIMILARITY

BERT (Bidirectional Encoder Representations from Transformers) is a pre-trained transformer-based language model developed by Google for natural language processing (NLP) tasks.

We got embeddings of the following using BERT -

Description + Name + Categories + Genres + Tags + Company

and then used cosine similarity to find similar games.

We also experimented with weighted cosine similarity to give more weight to popular games.

BERT TRANSFORMER + COSINE SIMILARITY

12859 [NAME]: Grand Theft Auto V, [DESC]: When a young street hustler, a retired bank robber and a terrifying psychopath find themselves entangled with some of the most frightening and deranged elements of the criminal underworld, the U.S. government and the entertainment industry, they must pull off a series of dangerous heists to survive in a ruthless city in which they can trust nobody, least of all each other. Grand Theft Auto V for PC offers players the option to explore the award-winning world of Los Santos and Blaine County in resolutions of up to 4k and beyond, as well as the chance to experience the game running at 60 frames per second. The game offers players a huge range of PC-specific customization options, including over 25 separate configurable settings for texture quality, shaders, tessellation, anti-aliasing and more, as well as support and extensive customization for mouse and keyboard controls. Additional options include a population density slider to control car and pedestrian traffic, as well as dual and triple monitor support, 3D compatibility, and plug-and-play controller support. Grand Theft Auto V for PC also includes Grand Theft Auto Online, with support for 30 players and two spectators. Grand Theft Auto Online for PC will include all existing gameplay upgrades and Rockstar-created content released since the launch of Grand Theft Auto Online, including Heists and Adversary modes. The PC version of Grand Theft Auto V and Grand Theft Auto Online features First Person Mode, giving players the chance to explore the incredibly detailed world of Los Santos and Blaine County in an entirely new way. Grand Theft Auto V for PC also brings the debut of the Rockstar Editor, a powerful suite of creative tools to quickly and easily capture, edit and share game footage from within Grand Theft Auto V and Grand Theft Auto Online. The Rockstar Editor's Director Mode allows players the ability to stage their own scenes using prominent story characters, pedestrians, and even animals to bring their vision to life. Along with advanced camera manipulation and editing effects including fast and slow motion, and an array of camera filters, players can add their own music using songs from GTAV radio stations, or dynamically control the intensity of the game's score. Completed videos can be uploaded directly from the Rockstar Editor to YouTube and the Rockstar Games Social Club for easy sharing. Soundtrack artists The Alchemist and Oh No return as hosts of the new radio station, The Lab FM. The station features new and exclusive music from the production duo based on and inspired by the game's original soundtrack. Collaborating guest artists include Earl Sweatshirt, Freddie Gibbs, Little Dragon, Killer Mike, Sam Herring from Future Islands, and more. Players can also discover Los Santos and Blaine County while enjoying their own music through Self Radio, a new radio station that will host player-created custom soundtracks. Special access content requires Rockstar Games Social Club account. Visit for details., [CATEG]: Single-player, Multi-player, PvP, Online PvP, Co-op, Online Co-op, Steam Achievements, Full controller support, Remote Play on Phone, Remote Play on Tablet, Remote Play on TV, [TAGS]: Open World, Action, Multiplayer, Automobile Sim, Crime, Third Person, First-Person, Shooter, Mature, Adventure, Singleplayer, Third-Person Shooter, Racing, Co-op, Atmospheric, Sandbox, Funny, Great Soundtrack, Comedy, Moddable, [GENRES]: Action, Adventure, [DEVELOPER]: Rockstar North, [PUBLISHER]:Rockstar Games
Name: description, dtype: object

EXAMPLE

Recommendation for Tom Clancy's Splinter Cell Chaos Theory®

Recommendations using Count Vectorizer

2272	Tom Clancy's Splinter Cell®
9222	Tom Clancy's Ghost Recon®
6737	Call of Duty®: Black Ops
1083	Killbot
703	Call of Duty®: Black Ops
13823	Far Cry® 2
927	Tom Clancy's Rainbow Six® 3 Gold
11766	Modern Combat Versus
9018	The Haunted: Hells Reach
7368	Sniper Elite V2 Remastered
6931	Tom Clancy's Splinter Cell Double Agent®
4176	Of Guards And Thieves
1100	Assassin's Creed® Brotherhood
6441	Sniper Elite 5

Name: Name, dtype: object

Recommendations using Tf-idf

None	
2272	Tom Clancy's Splinter Cell®
6931	Tom Clancy's Splinter Cell Double Agent®
2024	Tom Clancy's Splinter Cell Conviction™
927	Tom Clancy's Rainbow Six® 3 Gold
1887	Tom Clancy's Rainbow Six® Siege
9222	Tom Clancy's Ghost Recon®
10066	Tom Clancy's Splinter Cell Blacklist
2697	Tom Clancy's Rainbow Six® Vegas
14020	Tom Clancy's The Division® 2
11204	Tom Clancy's Ghost Recon: Future Soldier™
8508	Assassin's Creed™: Director's Cut Edition
7756	Dark Messiah of Might & Magic
7327	Tom Clancy's Rainbow Six® Vegas 2
7216	Dark Messiah of Might & Magic

Name: Name, dtype: object

Recommendation for Tom Clancy's Splinter Cell Chaos Theory®

Recommendations using BERT embeddings and cosine similarity

	Name	Release date	Estimated owners
2272	Tom Clancy's Splinter Cell®	Apr 1, 2008	200000 - 500000
518	Invisible, Inc.	May 12, 2015	500000 - 1000000
2637	Quantum Replica	May 31, 2018	0 - 20000
3181	The Price of Freedom	Dec 22, 2016	20000 - 50000
2714	Dark Sector	Mar 24, 2009	50000 - 100000
5522	The Bureau: XCOM Declassified	Aug 19, 2013	2000000 - 5000000
13981	Warhammer 40,000: Darktide	Nov 30, 2022	500000 - 1000000
439	Shot In The Dark	Jun 10, 2015	20000 - 50000
8698	Sniper Elite	Jul 16, 2009	500000 - 1000000
794	Phantom Doctrine	Aug 14, 2018	200000 - 500000
12799	Act of War: High Treason	Mar 12, 2008	50000 - 100000
747	Alekhine's Gun	Mar 11, 2016	100000 - 200000
3575	RUINER	Sep 26, 2017	500000 - 1000000
9865	Heavy Fire: Shattered Spear	Oct 23, 2015	0 - 20000

EXAMPLE

Recommendation for EA SPORTS™ FIFA 23

Recommendations using Count Vectorizer

```
1996           FIFA 22
3332           EA SPORTS™ FIFA 21
5138           AO Tennis 2
4417           Superfighters Deluxe
5060           Bloody Trapland
11707          WRC 9 FIA World Rally Championship
10779          Soccertron
8995           Aeolis Tournament
13865          Just Go
8509           Ball 2D: Soccer Online
9869           Madden NFL 22
12693          Gears 5
12695          Knight Squad 2
9825           Super Mega Baseball 2
Name: Name, dtype: object
```

Recommendations using Tf-idf

```
None
1996           FIFA 22
3332           EA SPORTS™ FIFA 21
6129           Battlefield: Bad Company™ 2
8509           Ball 2D: Soccer Online
2307          Captain Tsubasa: Rise of New Champions
541            Head Goal: Soccer Online
5060           Bloody Trapland
553            Warlocks vs Shadows
2445           8-Bit Commando
12693          Gears 5
9869           Madden NFL 22
491            Stick Fight: The Game
5138           AO Tennis 2
4646           Hide and Seek
Name: Name, dtype: object
```

Recommendation for EA SPORTS™ FIFA 23

Recommendations using BERT embeddings and cosine similarity

	Name	Release date	Estimated owners
1996	FIFA 22	Sep 30, 2021	2000000 - 5000000
12527	Soccer Manager	Feb 9, 2016	100000 - 200000
3332	EA SPORTS™ FIFA 21	Oct 8, 2020	1000000 - 2000000
2709	F1® 2021	Jul 15, 2021	1000000 - 2000000
6625	Ball 3D: Soccer Online	Mar 31, 2017	500000 - 1000000
9869	Madden NFL 22	Aug 19, 2021	200000 - 500000
5254	MotoGP™14	Jun 20, 2014	2000 - 50000
13134	EDGE	Aug 11, 2011	200000 - 500000
4398	F1® 22	Jun 30, 2022	100000 - 200000
11991	MX vs. ATV Supercross Encore	Oct 27, 2015	20000 - 50000
2675	NASCAR Heat 4	Sep 13, 2019	100000 - 200000
12201	NASCAR Heat 3	Sep 7, 2018	20000 - 50000
6786	CapRiders: Euro Soccer	Apr 7, 2017	0 - 20000
4789	TrackMania² Stadium	Jun 20, 2013	500000 - 1000000

THANK YOU!

We loved working on this project and would love to continue working on this in our holidays and create a phenomenal GUI.