

R Analysis of *Grand Data Auto: Vice City*

Abrar Faruque
Brenda Soljic
Sam Lenet



- Video game success is strongly shaped by marketing and user perception
- Steam reviews offer valuable insight into player behaviour
- GTA: Vice City provides a large dataset of user reviews and analysis
- Our goal: identify behavioural groups based on engagement, writing effort, and platform activity

What are characteristic behaviours in individuals who play Grand Theft Auto: Vice City, based on their reviews?

The *Grand Data Auto Vice City* dataset examines the reviews from the popular video game *Grand Theft Auto: Vice City*, with variables including:

ID	9 digit code for each player
Review	Text content of the review
Voted Up	TRUE = positive review; FALSE = negative review
Votes_Up and Comment_Count	Capture community interaction
Author Playtime Metrics (minutes)	Playtime_forever - total lifetime playtime Playtime_last_two_weeks - recent activity level Playtime_at_review - playtime at the moment the review was posted Values reflect data as of February 2025

	votes_up	comment_count	author_num_games_owned	author_num_reviews	author_playtime_forever	author_playtime_last_two_weeks	author_playtime_at_review	word_count	unique_word_count
votes_up	1	0.247577625	0.03244518	0.028370068	0.010954348	-0.002363583	0.01215683	0.02641987	0.03382086
comment_count	0.247577625	1	0.01785066	0.033160487	0.044846807	0.009388418	0.04311027	0.04961258	0.05958347
author_num_games_owned	0.03244518	0.017850662	1	0.303477984	0.055151656	-0.012149823	0.06425121	0.09699401	0.11810162
author_num_reviews	0.028370068	0.033160487	0.30347798	1	0.005699799	-0.026804472	0.01529067	0.12085607	0.14671054
author_playtime_forever	0.010954348	0.044846807	0.05515166	0.005699799	1	0.219384902	0.967435	0.08304388	0.09461056
author_playtime_last_two_weeks	-0.002363583	0.009388418	-0.01214982	-0.026804472	0.219384902	1	0.10968487	0.01392542	0.01398292
author_playtime_at_review	0.012156833	0.043110272	0.06425121	0.015290672	0.967434997	0.109684875	1	0.08789817	0.10025415
word_count	0.026419873	0.049612583	0.09699401	0.120856067	0.08304388	0.013925422	0.08789817	1	0.90232516
unique_word_count	0.033820861	0.059583467	0.11810162	0.146710541	0.094610559	0.013982924	0.10025415	0.90232516	1

- Strong correlation in Word count versus Unique word count, Play time at review and Playtime Forever (highlighted)
- All other variables show weak correlation

DETERMINING EFFORT

R Analysis of *Grand Theft Auto: Vice City*

Figure 1:

- **Orange(subgroup 1):** <1000 words and <100 unique words
- **Blue(subgroup 2):** 1-500 words and <1000 unique words
- **Green(subgroup 3):** ≥ 1000 words and >500 unique words
- **Red(subgroup 4):** >1000 words and <500 unique words

Overall

- No significant pattern between positive or negative (TRUE or FALSE) review in any subgroup

WORD COUNT VS UNIQUE WORD COUNT

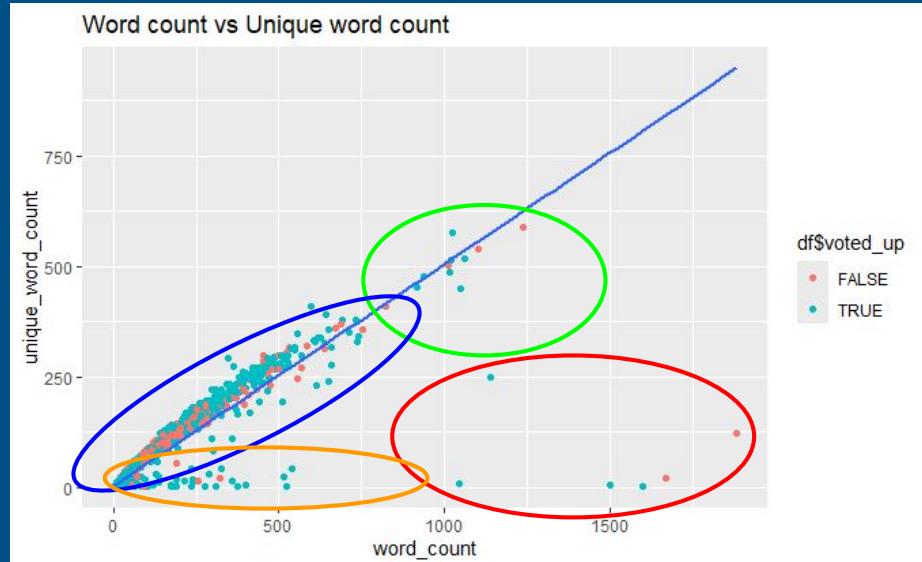


Figure 2:

- Negative (`voted_up = FALSE`)
 - Follows a normal distribution
 - Cluster of short, low-worded reviews
- Positive (`voted_up = TRUE`)
 - Does NOT follow a normal distribution until average Log
 - Unique word count = 1, extremely common
 - Single word reviews making up 32.95% of all pos reviews

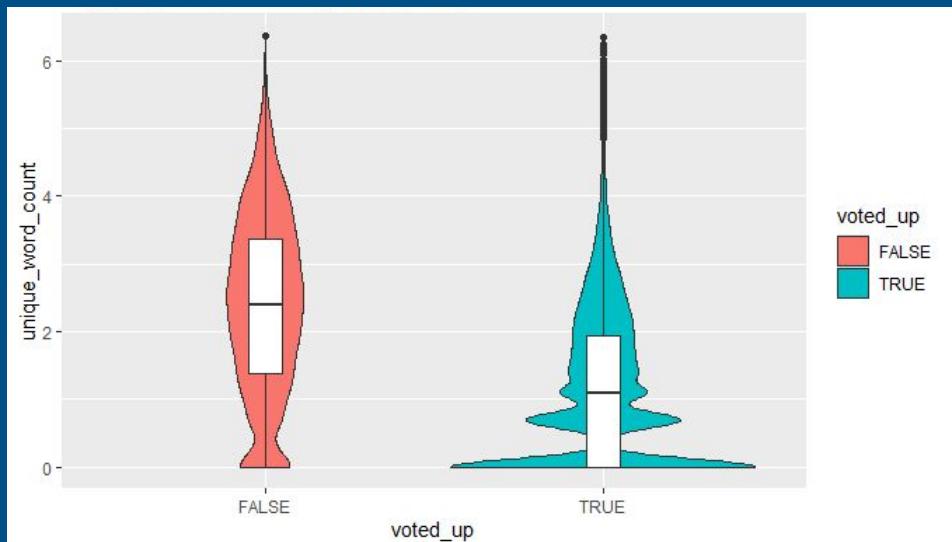
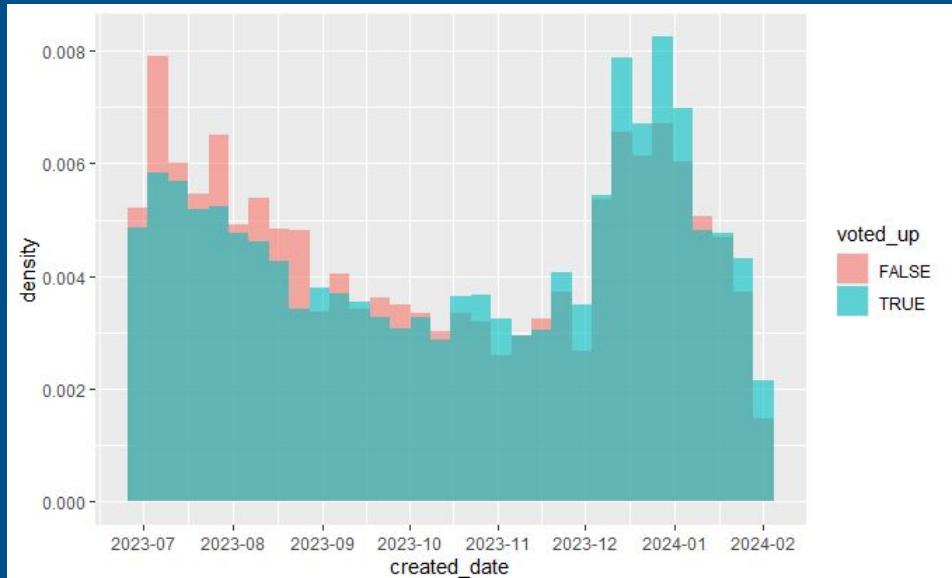
LOG (UNIQUE WORD COUNT) VS VOTED UP

Figure 3:

Number of reviews (June 27, 2023 – Feb 1, 2024) split by voted_up TRUE vs FALSE

- June–October 2023:
 - More negative (FALSE) reviews than positive (TRUE)
- After October 2023:
 - Positive (TRUE) reviews become more frequent than negative (FALSE)
 -
- December 2023 spike:
 - Review count more than doubled compared to November

DISTRIBUTION OF REVIEWS AND TIME

Determining Effort

- No formal “effort” metric, but higher word count + more unique words = higher effort
- Word count and unique word count are strongly correlated

Cluster Characteristics

- **Subgroups 2 & 3:**
 - Most reviews
 - Roughly 2:1 ratio of total words to unique words
 - Brief to moderately expressive reviews
- **Subgroups 1 & 4:**
 - Smaller groups
 - Low unique word diversity relative to length
 - Many contained repeated words/phrases
- **Effort does NOT predict voted_up outcome**
 - No correlation between writing effort and positive vs negative reviews

Reviewer Behaviour Trends

- Positive (TRUE) reviews:
 - Require less effort, extremely short
- Negative (FALSE) reviews:
 - Longer, contain more unique words
- Two reviewer archetypes emerge:
 - “Simpleton” - very short, positive feedback
 - “Critic” - longer, detailed negative feedback

Influence of Canonical Events on Reviews

- **June - November 2023:**
 - Negative reviews outnumbered positive reviews
- **After December 2023:**
 - Positive reviews became more common
 - Coincides with a major game update fixin serval issues
 - Aligns with the GTA 6 announcement

Timing and Playing Relationship

- No significant correlation between:
 - When players bought the game (author_plaaytime_forever)
 - When they wrote their review (author_playtime_at_review)

Guzsvinecz, T., & Szűcs, J. (2023, December). *Length and sentiment analysis of reviews about top-level video game genres on the steam platform*. ELSEVIER.

<https://www.med.upenn.edu/pmi/events/https-www-sciencedirect-com-science-article-abs-pii-s1047847720300046-via-3dihub>

Lutz, B., Proloochs, N., & Neumann, D. (2019, September). *The Longer the Better? The Interplay Between Review Length and Line of Argumentation in Online Consumer Reviews*. arXiv. <http://arxiv.org/pdf/1909.05192>

Mellinas, J. P., & Leoni, V. (2024, September). *Beyond words: unveiling the implications of blank reviews in online rating systems*. SpringerLink. <https://link.springer.com/article/10.1007/s40558-024-00300-4>

Statista. (2025, August). *Grand Theft Auto V Total Unit Sales 2025*. Statista.

<https://www.statista.com/statistics/1247955/gta-v-unit-sales-worldwide-total/#statisticContainer>

Zhong, X., & Xu, J. (2022, June). *Measuring the effect of game updates on player engagement: A cue from DOTA2*. Research Gate.

https://www.researchgate.net/publication/361091263_Measuring_the_effect_of_game_updates_on_player_engagement_A_cue_from_DOTA2

Thank you
Questions?