**הגורם האנושי בהנדסת נתונים - 096275**

**סמסטר חורף תשפ"ב 2021-2022**

**עבודה 2 - משחקי וידאו**

**מגישים**

שניר לוגסי

תומר זילקה

**הקדמה**

ענף הגיימינג החל כבר בתחילת שנות ה-50 מימי המחשבים הראשונים, מדעני המחשב אז הוכיחו את יכולותיהם ופרצו דרך בעזרת סימולציות ומשחקי מחשב פשוטים.

כיום גיימינג הוא תחביב בלתי נפרד של חלק נרחב מהאוכלוסייה, בשנת 2021 מוערך כי 2.81 מיליארד אנשים משחקים במשחקי וידאו באופן שגרתי, מתוך אזרחי ארה"ב משוער כי מעל 66% מהאוכלוסייה משחקים באופן שגרתי.

משחקי וידאו הם לרב תחביב שנועד לתגמל על התקדמות, לרגש וליצור הנאה אצל השחקנים. אך במקרים רבים הם גם הופכים למקצוע, ישנם שחקני ESports (תחרות באמצעות משחקי וידאו) שהופכים את התחביב לקריירה.

**מאפיינים של משחקי וידאו**

**1 - פלטפורמה**

המכשיר בו ניתן לשחק את המשחק, פלטפורמות פופולריות היום הם: מחשב, סמארטפון, קונסולות ביתיות (Xbox \ Sony Playstation) וקונסולות ניידות (Nintendo Switch).   
הצרכן (שחקן) לרב מתעניין במשחקים המוצעים בפלטפורמה שברשותו אך ישנם מקרים שבו המשחק מבוקש מאוד וגורר רכישה של הפלטפורמה.

**2 - ז'אנר**

הסגנון הכללי של המשחק, למשל: פעולה / מתח / הרפתקה / סימולציה / הישרדות / מדע בדיוני.

**3 - עלילה**

האירועים שהשחקן חווה במהלך המשחק, התקופה, הסביבה והעיצוב, כולם חלק גדול מהתעניינות של השחקן במשחק ובסבירות שירכוש את המשחק מלכתחילה.

**4 - סגנון המשחק (Game Mode)**

סגנון המשחק קובע את האופן בו השחקן משחק, את החוקים והקווים המנחים, ישנם סגנונות משחק המגדירים חוקים כללים שנותנים לשחקנים יותר חופש פעולה (נפוץ יותר עבור משחקים פחות תחרותיים - Casual) וישנם סגנונות עם חוקים נוקשים ומחמירים המגבילים את חופש הפעולה של השחקן אך יוצרים שוויון עם מסגרת ברורה למשחק (נפוץ יותר עבור משחקים תחרותיים).

לדוגמה:

* משחקי הישרדות בסגנון Open-World, נותנים לשחק חופש פעולה במרחב גדול אותו הוא מגלה באופן עצמאי.
* משחק יריות בגוף ראשון (First Person Shooter) בסגנון Search & Destroy, ישנם מספר סופי של סבבים, בכל סבב יש צוות אחד מגן וצוות אחד תוקף.

**5 - משחק אונליין / מרובה משתתפים**

אפשרות של השחקן לשחק עם אנשים אחרים הנמצאים איתו או דרך האינטרנט, במשחקי אונליין המשחק מהווה רשת חברתית ונותן לשחקן אפשרות להכיר אנשים מכל העולם ולהיות חלק מקהילת השחקנים.

**6 - גרפיקה וחווית משתמש**

חווית המשתמש נשענת רבות על איכות וסגנון המשחק מבחינה גרפית, המאפיינים הגרפיים של המשחק משפיעים על הדרך בה השחקן חווה את עולם המשחק.

ישנם משחקי דו-מימד / תלת-מימד / מציאות מדומה (VR), וכל אחד מאלו קובעים את אופי המשחק והחוויה.

**7 - שיטת קלט (Input Device)**

שיטות הקלט המקובלות בעולם הגיימינג הן עכבר ומקלדת או בקר (Gamepad / Controller).

ישנם צרכנים ששיטת הקלט תשפיע על ההחלטה האם לשחק משחק מסויים או לא, וישנם צרכנים שאדישים לשיטת הקלט.

בנוסף, בשנים האחרונות נוסף גם השימוש במסכי מגע כשיטת קלט למשחקי וידאו לאחר התפתחות גדולה במשחקי הוידאו בסמארטפונים וטאבלטים.

**8 - גמול ופרסים**

מרבית המשחקים מתגמלים על השחקנים על הישגים בתוך המשחק (In-Game Achievements) לרב על השלמת משימה שדורשת מאמץ תוך כדי המשחק, וכך נוצר חיזוק חיובי עבור השחקן (לרב גורר רצון להמשיך לשחק).

**9 - מפרט נדרש (System Requirements)**

הדרישות של המשחק מהמערכת עליה הוא פועל, סביר להניח כי משחקים חדשים ידרשו מחשבים / קונסולות חזקות יותר מאשר משחקים ישנים.

דרישות טכניות אלו משפיעות על קהל היעד של המשחק, דרישות גבוהות מדי גוררות קהל יעד מצומצם יותר שבבעלותו מחשבים / קונסולות חזקות יותר.

**10 - מחיר / מודל עסקי**

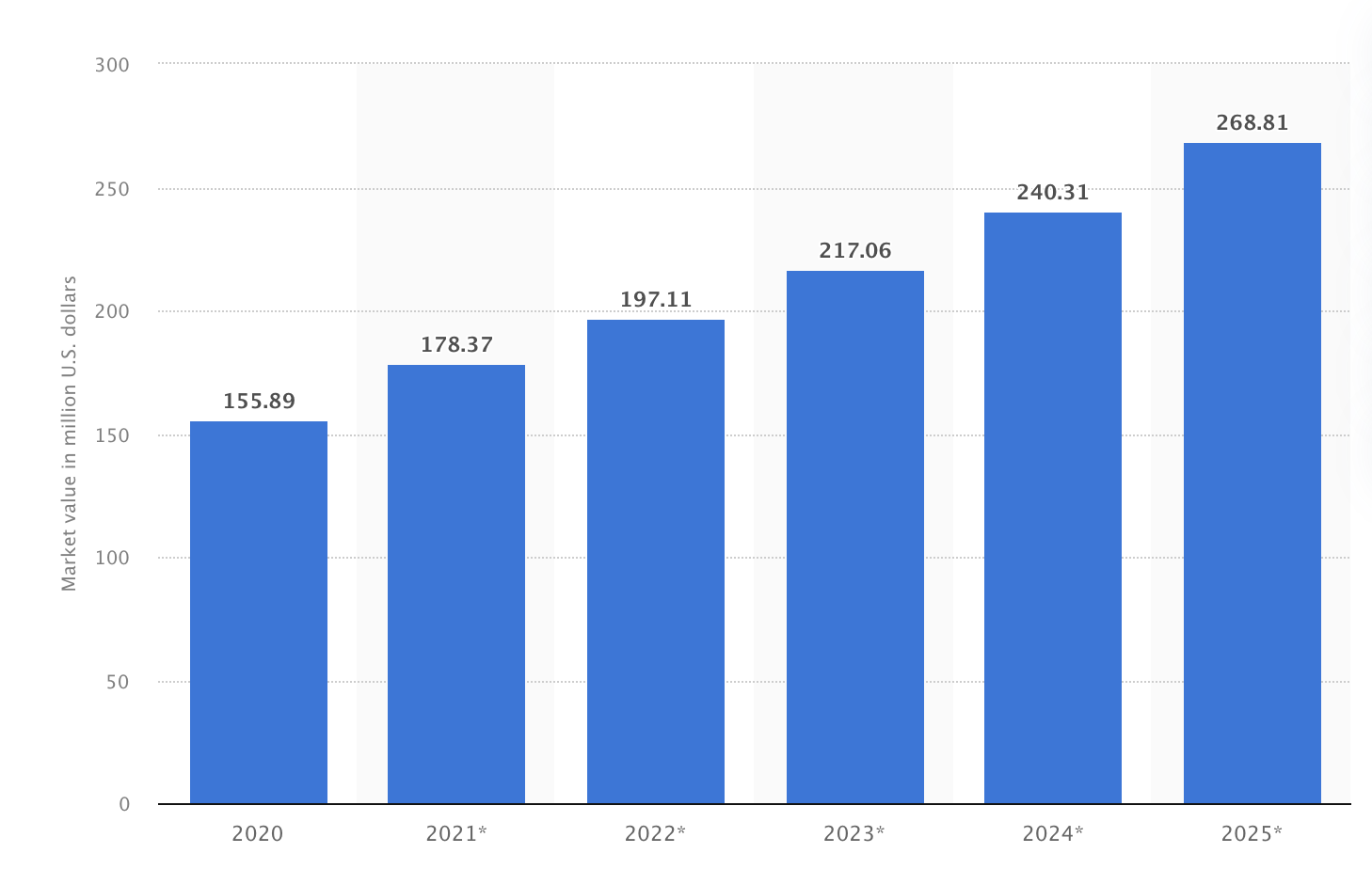
משחקי וידאו פותחים דלת למגוון רחב של אפשרויות עסקיות, למעט מחיר המשחק עצמו ישנם עלויות נוספות אפשריות, למשל:

* מנוי למשחק, לרב תשלום חודשי עבור שימוש בכל האפשרויות של המשחק.
* הרחבות בתשלום של המשחק, למשל: הרחבה של עולם המשחק והעלילה.
* קנייה של פריטים בתוך המשחק (מיקרו-טרנזקציות), למשל: קניית פריטי לבוש לדמות.
* מסלולי פרסים (לרב ידוע בתור Game-Pass / Battle-Pass), במסלול זה השחקן משלם על משימות שמבטיחות פרסים (לרב ידועים מראש), למשל: תשלום חד פעמי עבור 100 שלבים שבכל שלב השחקן מקבל פרס כלשהו, ועליית שלב נעשית ע״י צבירת נקודות תוך כדי משחק.

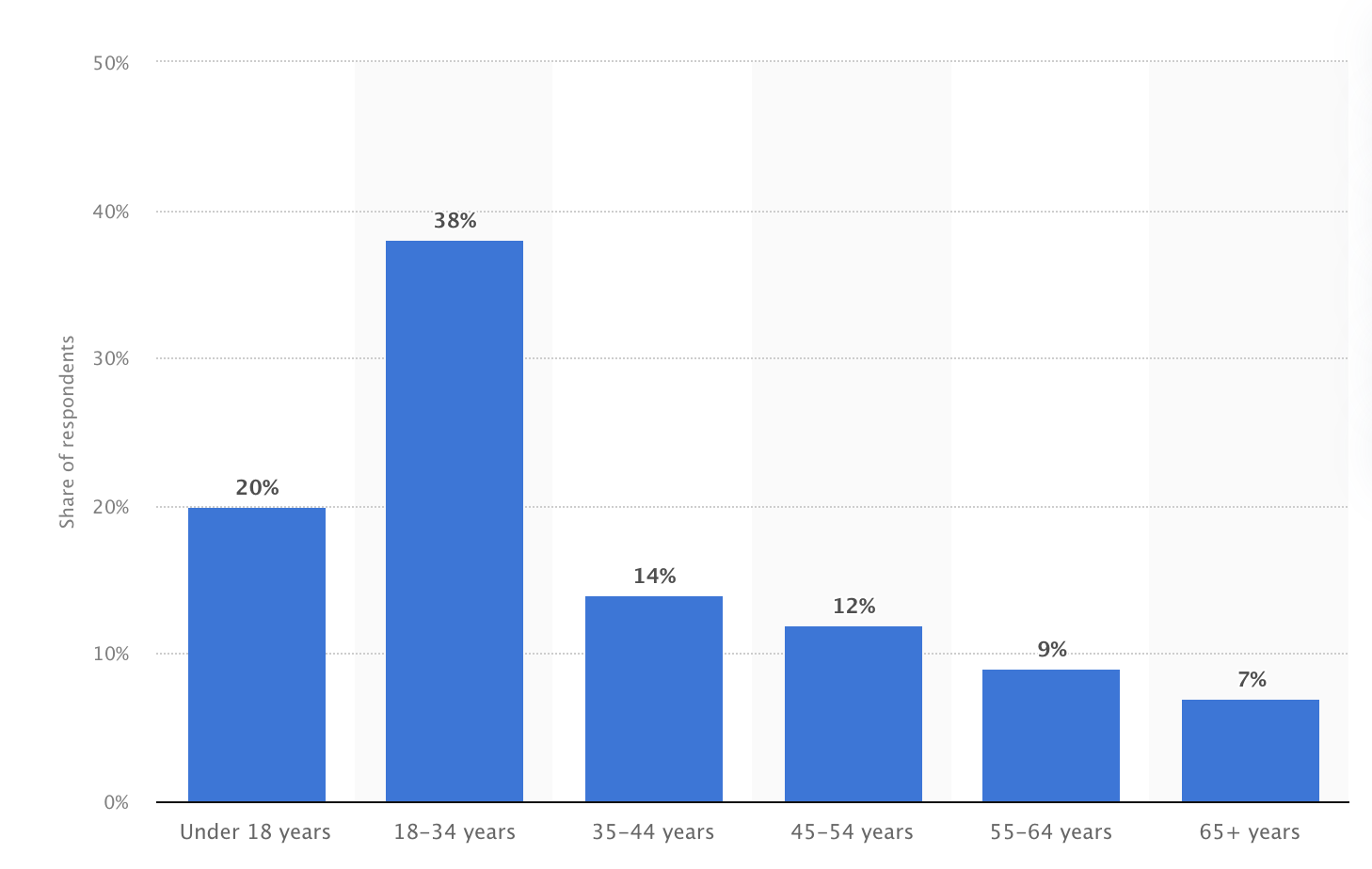
בשנים האחרונות ישנה עלייה משמעותית במודל עסקי הסובב סביב F2P (מודל Free-To-Play) בשילוב עם הכנסות סביב מנוי או מיקרו-טרנזקציות בתוך המשחק. (רווח של 23.1 מיליארד דולר בשנת 2020 עבור משחקים במודל זה)

שוק הגיימינג

**Global video game market value from 2020 to 2025**

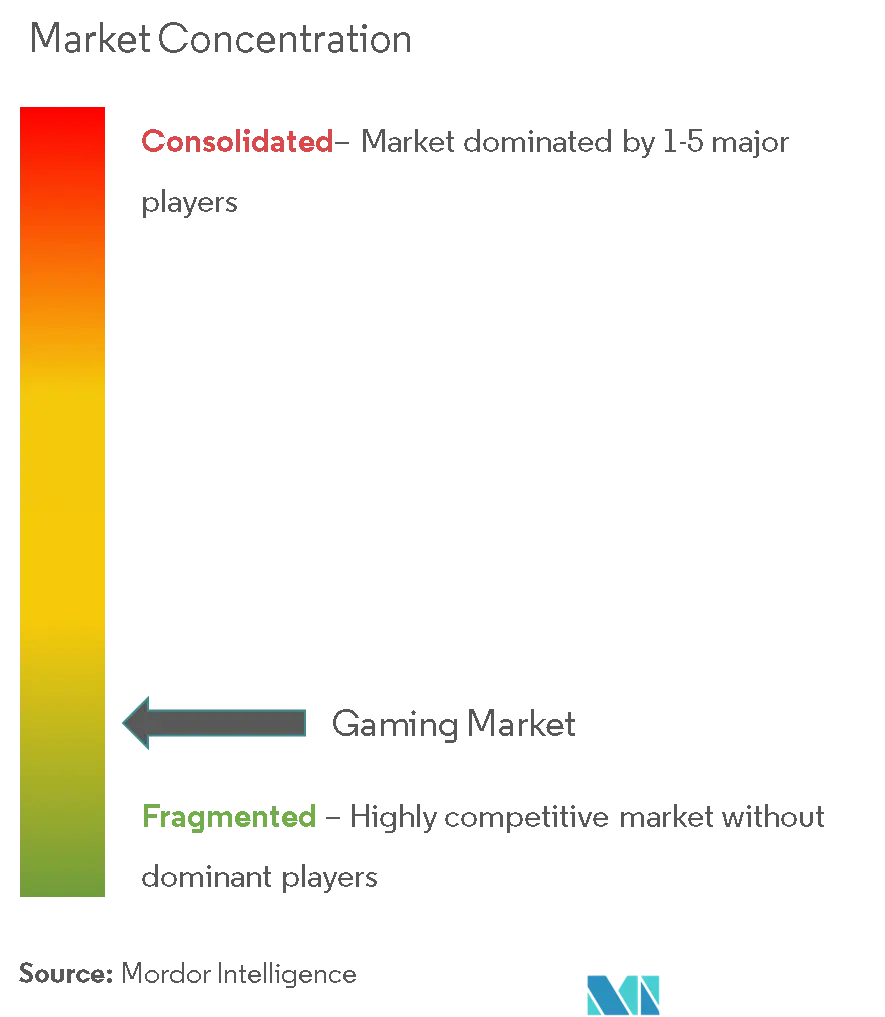


**Distribution of video gamers in the United States in 2021, by age group**



נתונים סטטיסטים ודמוגרפיים בתחום הגיימינג:

* ישנם יותר מ**2.8 מיליארד** אנשים בעולם המשחקים במשחקי וידאו
* גיל השחקן הממוצע הוא **34**, 11% מהשחקנים מעל גיל 50 ורק 30% מתחת לגיל 18
* **45%** מהשחקנים בארה״ב הן נשים
* יותר מ**66%** מאזרחי ארה״ב משחקים באופן **יום-יומי**
* **70%** מההורים מאמינים כי ישנה השפעה **חיובית** של משחקי וידאו על ילדיהם
* **47%** ממפתחי המשחקים משווקים דרך פלטפורמת **Steam**

**Top 10 companies in the gaming industry (by revenue in 2021Q2)**

| Current Market Cap  (in billion $) | Revenue in 2021Q2  (in billion $) | Company | # |
| --- | --- | --- | --- |
| ​​550.144 | 7.932 | Tencent (TCEHY) | 1 |
| 160.86 | 4.147 | Sony (SONY) | 2 |
| 2823 | 3.599 | Apple (AAPL) | 3 |
| 2350 | 2.916 | Microsoft (MSFT) | 4 |
| 1816 | 2.749 | Google (GOOG) | 5 |
| 67.53 | 2.225 | NetEase (NTES) | 6 |
| 49.88 | 2.163 | Activision Blizzard (ATVI) | 7 |
| 54.735 | 1.630 | Nintendo (NTDOY) | 8 |
| 37.212 | 1.551 | Electronic Arts (EA) | 9 |
| 18.946 | 0.813 | Take-Two Interactive (TTWO) | 10 |

מקורות:

<https://techjury.net/blog/video-games-industry-statistics/#gref>

<https://www.statista.com/statistics/292056/video-game-market-value-worldwide/>

<https://www.statista.com/statistics/189582/age-of-us-video-game-players/>

<https://newzoo.com/insights/rankings/top-25-companies-game-revenues/>

<https://finance.yahoo.com/>

<https://www.mordorintelligence.com/industry-reports/global-gaming-market>

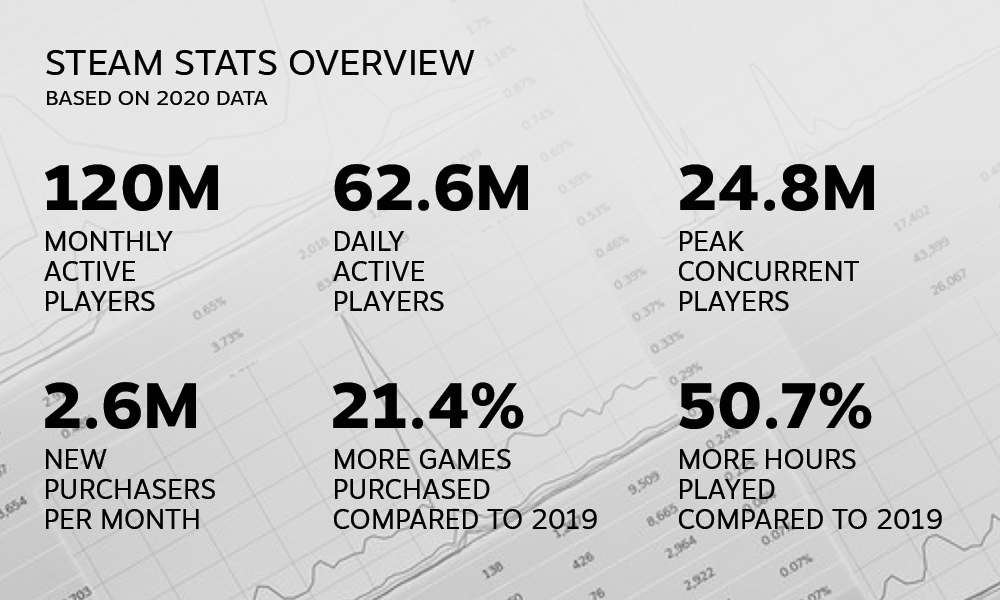
**איסוף נתונים ועיבוד ראשוני**

בחרנו לשלוף מידע מחנות המשחקים הגדולה ביותר בעולם - Steam (נכון לכתיבת עבודה זו ישנם 60,470 משחקים בחנות).

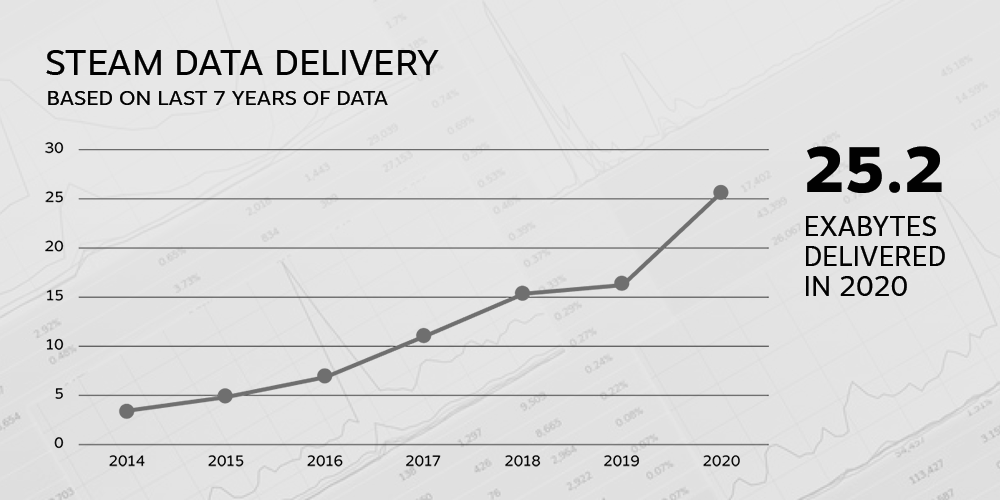
נתונים כלליים על כל משחק נלקחו מהעמוד של המשחק בחנות, וביקורות על המשחק נלקחו מSteam Community, זהי פלטפורמה קהילתית של Steam בה שחקנים מכל העולם יכולים לשתף את החוויה שלהם מהמשחק ואת הביקורת עליו.

בעמוד הביקורת על משחק ישנם אפשרויות רבות לסינונים, אנו בחרנו לקחת את 100 התגובות שSteam סימנו כ״הכי מועילות״. בחירה זו נעשתה מתוך סקרנות לגבי מהי ביקורת חיונית בתחום משחקי הוידאו (ננתח זאת בהמשך).

נתונים מסכמים לשנת 2020 של חנות Steam:



סיכום כמות הנתונים שהועברו דרך שרתי Steam ב-7 השנים האחרונות:



נלקח מ[כתבה רשמית](https://store.steampowered.com/news/group/4145017/view/2961646623386540826) לסיכום שנת 2020 עבור Steam.

כחלק מעבודה זו, פרסמו את הקוד לכריית המידע מSteam Store באופן פתוח.  
קישור לקוד: <https://github.com/snirlugassy/steam-scraping>

בדף הREADME בסמוך לקוד ניתן לראות פירוט מלא של קובץ ה robots.txt עבור הכתובות מהן שלפנו את המידע.

**ניתוח הנתונים**

ניקיון ועיבוד הנתונים: כבר בשלב כריית המידע דאגנו לנקות ולעבד את הנתונים בצורה מיטבית לקראת הניתוח והשאלון, זה נעשה בעיקר ע״י scrape מאוד ממוקד והמרת טיפוסים לאחת שליפה.

לכן, כבר לאחר תהליך ה-scraping הנתונים הגיעו לצורה איכותית ומסודרת.

את הנתונים לגבי כל משחק שמרנו באובייקט עם המבנה:

{

"game": {

"description": String,

"developer": String,

"features": List[String],

"img": URL,

"min\_sys\_req": List[String],

"publisher": String,

"rating": Integer,

"rec\_sys\_req": List[String],

"release\_date": DateString,

"review\_count": Integer,

"tags": List[String],

"title": String,

"video\_src": URL

},

"reviews": List[Review] = [

{

"helpful": String,

"Hrs\_on\_record": Float,

"recommendation": String,

"rewards": Integer,

"Text": String

}

]

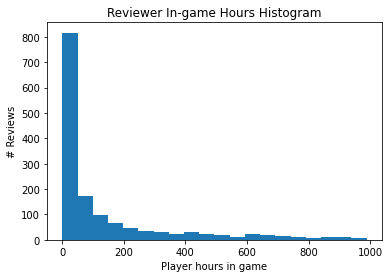
}

את קובץ הנתונים המלא ניתן לראות גם ב-GitHub.

דבר ייחודי לביקורת משחק ב-Steam הוא העובדה שהמבקרים משחקים במשחק אותו הם מבקרים דרך Steam, לכן עבור כל ביקורת ניתן לראות את מספר השעות שהכתוב שיחק במשחק.

נתון זה מעלה את ערך הביקורת עבור הקוראים, ונותן מידע רב אודות אמינות הביקורת והיכולת של המבקר באמת לשפוט את המשחק, למשל: ניתן לייחס חשיבות נמוכה לביקורת מאוד שלילית כאשר המבקר שיחק מעט מאוד במשחק.

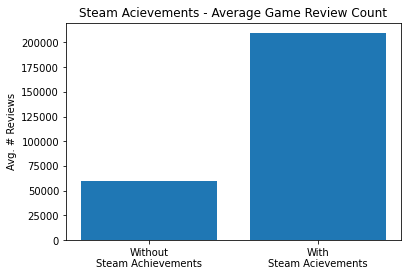
אין ספק כי מידע מסוג זה היה יכול לעזור לקבל החלטות צריכה נכונות יותר ברשת, כאשר אנו מעוניינים לקנות מוצר מסוים, הביקורות עליו לעיתים קרובות מאוד מוטות (Biased) ולרב ללא בסיס כלשהו המגבה את דעת הכותב.



ניתן לראות כי התפלגות שעות המשחק של המבקרים בעלת זנב ארוך, מרבית השחקנים (78.1%) שיחקו פחות מ-200 שעות במשחק אותו הם ביקרו, אך עדיין ביקורות אלו סומנו כמועילות.

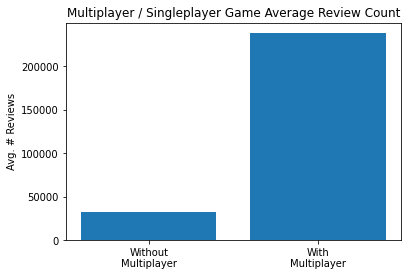
עבור כל משחק אספנו תכונות עיקריות של המשחק (features) שעוזרות לשחקנים למצוא משחקים המתאימים לצרכים והעדפות שלהם, למשל:

* תמיכה בבקר משחק - פיצ׳רים Full / Partial Controller Support
* משחק Single-Player
* משחק Multi-Player
* מעקב אחר הישגים דרך Steam - פיצ׳ר Steam Achievements.



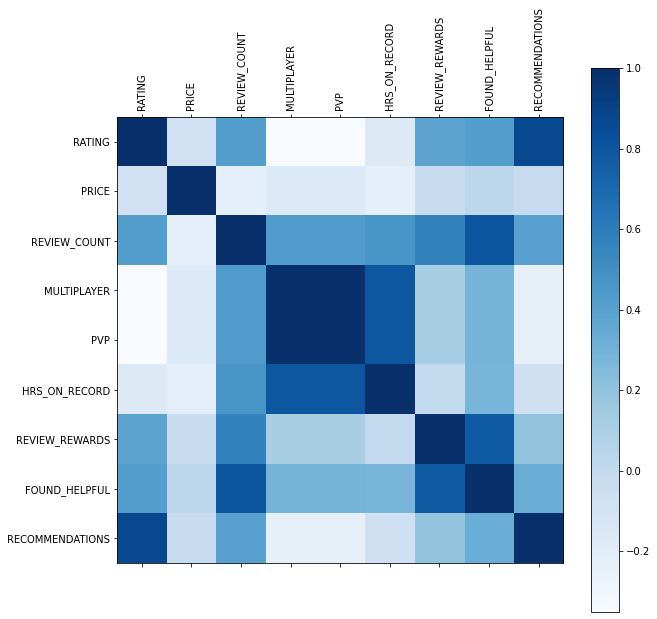
ניתן לראות כי משחקים בהם יש מעקב אחר הישגים דרך הפלטפורמה עצמה (לא רק בתוך המשחק) בעלי מספר ביקורות ממוצע גדול מאשר המשחקים ללא מעקב הישגים.

באופן דומה ניתן לראות זאת עבור משחקי Multiplayer (בעזרת התיוג של המשחק):

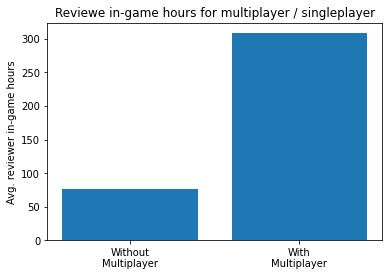


משחקים בעלי אפשרות Multiplayer מהמדגם שלנו בעלי מספר ביקורות ממוצע גבוה מאוד מאשר אלו ללא אפשרות Multiplayer.

 על מנת למצוא מתאמים  משתנים במדגם, יצרנו תמונה של מטריצת הקורלציה:



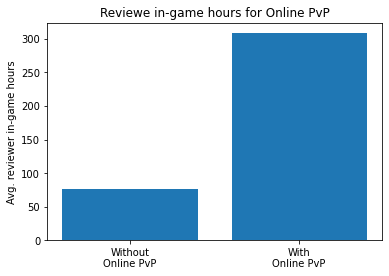
ניתן לראות כי קיימת התאמה בין משחקי מרובה משתתפים לכמות השעות ששיחקו המבקרים:



אנו משערים כי התאמה זו נובעת מהעובדה שמשחקי Single-Player לעיתים עם מספר סופי של אפשרויות או עלילה סופית.

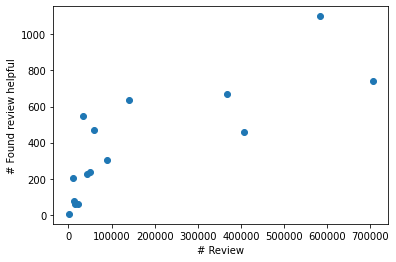
לעומת זאת במשחקי  Multi-Player המוקד הוא לרב על הישגים ופעילות חברתית לעומת עלילה, לכן ישנם מספר רב יותר של אפשרויות והזמן שבילית במשחק הופך למרכיב חשוב בהיכרות של השחקן עם המשחק.

 בדומה להתאמה וזו, ניתן לראות התאמה בין משחקים בהם מאפשרים לשחקנים לשחק אחד נגד השני ברשת (Online PvP - Player vs Player) לבין כמות השעות ששיחקו המבקרים:



גרף זה כמעט זהה לקודם מכיוון שרב משחקי ה-Multiplayer מאפשרים גם PvP, אך זה לא מובן מאליו מכיוון שישנם משחקי Multiplayer בהם אין PvP.

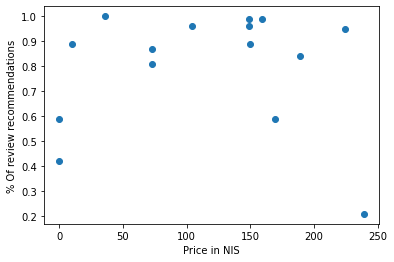
ניתן לראות התאמה נוספת בין מספר הביקורות על משחק לבין המספר המממוצע של אנשים שמצאו ביקורת כמועילה בכל משחק:

​​

אנו משערים כי התאמה זו נובעת מהעובדה שכאשר ישנם יותר ביקורות על משחק, כך גם ישנם יותר אנשים שקוראים את הביקורות ומסמנים האם הביקורת מועילה או לא.

משתנה מעניין שעבורו לא הייתה התאמה עם שאר המשתנים הינו המחיר של המשחק, ניתן לראות במטריצת הקורלציה כי המחיר ללא התאמה עם שאר המשתנים המחושבים, ולכן מסתמן כי איננו משפיע על הביקורות.

למשל, אם נבדוק את אחוז הביקורות הממליצות לשחק במשחק (ממוצע) בין המחירים השונים, נראה כי אין התאמה ואין תלות בין מחיר להמלצה לגבי המשחק:



**איסוף מכוון**

שאלת סיווג בינארי: האם היית משחק במשחק הבא בכדי להירגע?

שאלות נוספות:

1. גיל
2. זכר / נקבה
3. סטטוס (נשוי / רווק / …)
4. כמה את מסכים עם האמירה: ״משחקי וידאו הם דרך טובה לנצל את הזמן הפנוי״? (1=לא מסכים כלל, 5=מסכים מאוד)
5. באיזה גיל לראשונה נחשפת למשחקי מחשב?
6. האם כיום אתה משחק משחקי וידאו באופן קבוע?
   1. במידה וכן, כמה שעות בשבוע אתה משחק?
7. האם לדעתך משחקי וידאו הינם אמצעי טוב להפגת מתח?
8. כיצד לדעתך משחקי וידאו תורמים להתפתחות הילדים? (1=פוגע, 5=תורם)