Project Report - Milestone 1

INST 737 - Introduction to Data Science

Research Study: MindGame Insights: A Deep Dive into Gaming and Psychological Well-Being Relationships

(Areas of Research: Behavioral Data Science with a focus on Cyberpsychology)

(Team: Rajeevan Madabushi, Pranav Adiraju, Asmita Samanta)

1. Introduction:

"Video Gaming is the world's favorite pastime for GenZ and Millennials and fairly so!"

This statement is also backed by a popular research study where **94%** of participants aged between **(13-39)** chose "video gaming" as their favorite hobby over other popular hobbies like reading, cooking, or playing their favorite instrument *"How Video Games Are Influencing Gen Z's Real Life Behavior," (2023)*. Looking back in time, by 2019 there were more than **2.5 billion** players around the world, marking a staggering growth of more than a billion players since 2014 and this rose even more during the COVID-19 pandemic where many individuals turned to gaming to keep themselves occupied during the quarantine and widespread lockdowns. This surge not only heightened player engagement but also bolstered the revenue streams of major gaming businesses to a whole new level. But at what cost?

Although several research studies showed that games could act as stress-relievers, bolster spatial cognitive abilities, and provide a salve for feelings of isolation, with the onset of the pandemic, the landscape of everyday life underwent a drastic shift. As the time and video games became more available, the recreational players who once played just for fun have started to take a path towards the gaming addiction journey ultimately leading to several psychosocial ailments like Anxiety, Depression and Social Phobia. Furthermore, the same WHO, who in 2018, classified gaming disorder as an ***"illness"*** started to champion the use of video games to foster social connections in physically distanced world through ***#PlayApartTogether*** campaign on social media which stands as a testament to this advocacy. But did this backfire? *Jin, Y. et al. (2023)*

Obsessive Gaming leads to mental health disorders such as anxiety, depression and also results in them losing track of their responsibilities and social life. For example, kids missing school and adults missing the office for gaming. This addiction can also lead to distance from their loved ones due to lack of interaction. Moreover, violent video also makes one more aggressive, insensitive and violent. Hence, excessive gaming can not only make you more anxious, depressed, violent and away from friends and family, but it also slowly corrupts your lifestyle and makes you irresponsible.

Hence, we decided to explore the connection of gaming to one's psychological health & further its effect on one's life and associated. This brings us to our research question -

Predict employment status and reasons for gaming using different demographic, social, gaming hours and psychological indicators.

From a technical standpoint, understanding the connection of demographics, social factors, gaming behavior, and psychological indicators in predicting employment status and gaming motivations will incorporate predictive modeling techniques and data analysis methodologies. Developing accurate models can enhance our ability to predict and explain complex human behaviors, contributing to the field of data science and analytics.

From a social impact standpoint, this research can inform policy makers, activists etc. about the potential links between gaming habits and psychological well-being and that can aid in creating educational programs, and policies that regulate the gaming rules and promote healthy gaming behaviors.

2. State of the Art:

Paper 1:

Overview:

Federica Pallavicini et al. (2022) worked on a study that delves into the world of gaming, moving beyond mere playtime, with over **13,000 participants** dedicating an average of **22 weekly hours** to gaming. Through surveys, the research explores life satisfaction, anxiety levels, and social comfort, aiming to uncover how gaming's social aspect, be it solitary, with online friends, or real-life companions, impacts the mental well-being of both casual and serious gamers. Notably, the study challenges the conventional belief that playtime solely signifies a gamer's mental state, highlighting the importance of gaming motivations, particularly achievement-seeking and escapism.

Results:

In examining Internet Gaming Disorder issues, the research finds that the drive for achievements has a more substantial impact than a desire for socializing. Achievement-motivated players tend to prefer real-life friends as gaming partners, while those using gaming as an escape are more inclined to play with online acquaintances. Furthermore, the study uncovers that playtime significantly affects overall well-being, particularly in expansive online role-playing games. Playing with non-gaming friends is linked to higher life satisfaction and reduced social anxiety, contrasting with achievement-driven players who often report lower life satisfaction and heightened anxiety, especially when gaming serves as an escape from daily challenges.

Limitations:

Acknowledging its reliance on survey data, the study recognizes potential gaps in capturing the complete truth. Future research aims to delve deeper into gaming motivations and social elements while continuing to consider playtime and its impact on mental well-being.

Relevance:

It explores how gaming hours and motivations impact mental well-being. This helps understand reasons for gaming and their psychological implications, which can relate to employment status.

Paper 2:

Overview:

Niklas Johannes et al. (2021) in collaboration with Electronic Arts and Nintendo of America, this study harnessed real gameplay data, offering a more realistic perspective compared to survey-based research. Focusing on players of Plants vs. Zombies: Battle for Neighborville and Animal Crossing: New Horizons, the study surveyed players about their well-being and motivations during gameplay. Contrary to concerns about excessive gaming harming mental health, the research uncovered a small positive link between gaming and well-being, challenging the notion that extensive playtime is inherently detrimental.

Results:

Combining precise game data with survey responses, the study analyzed **518** Plants vs. Zombies players **(0.21% response rate, average age 35)** and **6,011** Animal Crossing players **(1.75% response rate, average age 31)**. It found that players often overestimate their playtime, and more playtime generally correlates with enhanced well-being. Interestingly, players' self-reported playtime doesn't always align with their actual engagement. In-game experiences, like autonomy and competence, positively influenced feelings, while external motivations had a negative impact. The study did not consistently link playtime with motivations or in-game feelings.

Limitations:

Relying on self-reported digital behavior and gaming statistics can introduce imprecisions and biases, affecting the interpretation of the relationship between playtime and well-being. Recognizing these potential biases is crucial when drawing conclusions from research on gaming's impact on well-being. Comparatively, our study boasts a larger dataset with a broader range of factors, including reasons for playing and social aspects, which we believe are pivotal in understanding gamers' mental well-being.

Relevance:

It challenges the assumption that excessive gaming is harmful and examines the link between playtime and well-being, which is important for understanding the impact of gaming hours on employment status.

Paper 3:

Overview:

Wei et al. (2012) aimed to investigate the relationship between online gamers' characteristics, gaming hours, social phobia, and depression. An online survey was distributed through popular gaming websites, and the study compared the results with established self-rating scales, including the **Depression & Somatic Symptoms Scale (DSSS)**, **Social Phobia Inventory (SPIN)**, and **Chen Internet Addiction Scale (CIAS).**

Results:

The study involved **722** gamers with **83.2%** male participants with an average gaming time of **28.2** hours overall. The findings revealed a positive link between higher gaming hours and increased internet addiction, symptoms of depression, pain etc. Additionally, participants spending more time on online gaming displayed greater symptoms of anxiety. Notably, female participants, although a minority, exhibited a higher risk for depression.

Limitations:

The study's dependence on survey data introduces bias which would have been more reliable if it were face-to-face interviews. Additionally, the study did not comprehensively capture participants' professions and hobbies, which could influence the correlation between gaming hours and addiction.

Relevance:

It directly investigates the relationship between gaming hours and psychological well-being, providing insights into the potential effects of gaming on employment status.

Paper 4:

Overview:

Stănculescu, et al., (2019) investigates the impact of video gaming habits on various aspects of psychological functioning utilizing correlations and regression analysis. It examines self-esteem, loneliness, satisfaction in life, anxiety etc. in relation to video gaming.

Results:

The data collected was through a point-of-time online survey of **2,732** participants. Excessive gaming was associated with poor mental health, bad coping strategies, low self-esteem, loneliness, and bad school performance. Distraction-motivated gaming was linked to higher symptom ratings, lower self-esteem, and more negative affectivity, while gaming for social relationships in the virtual world was related to more online connections and more positive affect during gameplay. Preferred video game genres had weak associations with psychological functioning, with action games showing the strongest positive affect during gameplay. Positive experiences during gaming and a larger number of online friends were indicative of potentially problematic video game use, leading to psychological symptoms, a lack of offline connections, and poorer school performance.

The study emphasizes that while enjoyable gaming can lead to excessive play, underlying mental health issues play a more significant role. It underscores the need to focus on why people play and their gaming preferences when studying the impact of gaming on mental health.

Limitations:

Studying has several limitations. Firstly, it cannot establish causation as it only examines data from a single point in time, making it challenging to determine whether gaming or mental health problems came first. Secondly, survey data may introduce bias. Thirdly, the study participants were self-selected, potentially not representative of all gamers. Lastly, the research found weak connections between gaming and mental health, indicating a complex and multifaceted relationship that requires further exploration.

Relevance:

It examines the psychological indicators and motivations behind gaming, contributing to understanding reasons for gaming and how these might affect employment status.

3. Dataset Explanation:

As part of our research project, we decided to use the OSF Gaming Habits and Psychological Well-Being Dataset *OSF, n.d.,* which is a publicly available dataset accessible through the **Open Science Framework (OSF)** platform. This dataset was collected through a survey taken by over **13,464 participants** from around **109 countries** in the world. Compared to this size of dataset, most of the other datasets that we came across in other research studies were focused on smaller sample sizes with narrow cultural backgrounds and demographics (younger adults). We expect the OSF Dataset would allow for a better predictive investigation of our research question due to its size and diversity.

The dataset contains **54 columns** of information, covering a range of data points from hours spent on gaming, demographic indicators like age, birthplace etc. to psychological indicators such as **Satisfaction with Life (SWL1-5)**, **Generalized Anxiety Disorder (GAD1-7)**, demographic information, **Social Phobia Inventory (SPIN1-17)**. All participants were asked to provide these details for the past two weeks.

Below are all the columns and their definitions (wherever applicable), the ones we will work with -

- ***Zeitstempel:*** Timestamp when data was recorded.
- ***GADE:*** Response to how difficult anxiety has made it for one at work, taking care of things at home, or getting along with other people. It ranges from "Not difficult at all" to some level of difficulty.
- ***Game:*** Name of the video game being played by the participant.
- ***Platform:*** Gaming platform used, such as PC, console (PS, Xbox, etc.), or other.
- ***Hours:*** Number of hours spent playing the game.
- ***Earnings****:* Earnings related to the game
- ***whyplay****:* Reason why the participant plays the game, such as for fun, improvement, or other reasons.
- ***League****:* Participant's affiliation with a gaming league or level within a game.
- ***streams:*** No. of streaming of the game
- ***Narcissism:*** Score related to narcissism (Scale 1-5)
- ***Work:*** Employment status of the participant.
- ***Degree:*** Highest educational degree attained by the participant.
- ***Birthplace:*** Birthplace of the participant.

- ***Residence:*** Current residence of the participant.
- ***Reference:*** Source or reference through which the participant learned about the survey
- ***Playstyle:*** Preferred playstyle of the participant, such as single-player or multiplayer.
- ***GAD_T, SWL_T, SPIN_T:*** Aggregated scores or totals for the GAD, SWL, and SPIN columns, respectively.

Generalized Anxiety Disorder (GAD) is evaluated through seven questions with each assigned a rating of 0-3 by a participant. These values are then aggregated to a total GAD score, which categorizes individuals into different levels of GAD severity. Total scores range from 0 to 21 where a score of 4 and below means minimal risk, 5-9 means mild risk, 10-14 means moderate risk and 15+ means serious anxiety disorder. *Anxiety and Depression Association of America, n.d.*

GAD-7 Anxiety

Over the last two weeks, how often have you been bothered by the following problems?	Not at all	Several days	More than half the days	Nearly every day
1. Feeling nervous, anxious, or on edge	0	1	2	3
2. Not being able to stop or control worrying	0	1	2	3
3. Worrying too much about different things	0	1	2	3
4. Trouble relaxing	0	1	2	3
5. Being so restless that it is hard to sit still	0	1	2	3
6. Becoming easily annoyed or irritable	0	1	2	3
7. Feeling afraid, as if something awful might happen	0	1	2	3

Column totals ____ + ____ + ____ + ____ =

Total score ______

Scoring GAD-7 Anxiety Severity

This is calculated by assigning scores of 0, 1, 2, and 3 to the response categories, respectively, of "not at all," "several days," "more than half the days," and "nearly every day."
GAD-7 total score for the seven items ranges from 0 to 21.

0–4: minimal anxiety

5–9: mild anxiety

10–14: moderate anxiety

15–21: severe anxiety

The SWLS (Satisfaction with Life Scale) uses a 7-point rating system, where scores can range from 5 to 35. A score of 20 is considered a neutral point. Scores between 5 and 9 suggest the respondent is extremely dissatisfied with life, while scores between 31 and 35 indicate a high level of satisfaction. The scale has shown strong internal consistency, with coefficient alpha values ranging from .79 to .89. Additionally, it demonstrated good test-retest reliability with correlations of .84 and .80 over a one-month period. In our dataset, the information gathered from the SWLS assessment is labeled as SWL1, SWL2, up to SWL5, with each label corresponding to one of the five questions used to assess life satisfaction. *Greenspace Health, n.d., SWLS*

		Strongly Disagree	Disagree	Slightly Disagree	Neither Agree nor Disagree	Slightly Agree	Agree	Strongly Agree
1.	In most ways my life is close to my ideal.	1	2	3	4	5	6	7
2.	The conditions of my life are excellent.	1	2	3	4	5	6	7
3.	I am satisfied with my life.	1	2	3	4	5	6	7
4.	So far I have gotten the important things I want in life.	1	2	3	4	5	6	7
5.	If I could live my life over, I would change almost nothing.	1	2	3	4	5	6	7

Score	Level of Satisfaction with Life
31 - 35	Extremely satisfied
26 - 30	Satisfied
21 - 25	Slightly satisfied
20	Neutral
15 - 19	Slightly dissatisfied
10 - 14	Dissatisfied
5 - 9	Extremely dissatisfied

The **Social Phobia Inventory, often abbreviated as "SPIN"** is a self-assessment tool with 17 questions designed to measure social anxiety disorder, also known as social phobia. This scale evaluates experiences over the past week and covers various aspects of the disorder, including fear, avoidance, and physical symptoms.

The reason for developing the SPIN was that existing self-assessment tools for social phobia didn't comprehensively address all three key areas: fear, avoidance, and physiological symptoms, all of which are crucial in clinical assessment.

The SPIN has proven to be a reliable tool with strong psychometric properties. It holds promise as an effective tool for screening social phobia and assessing treatment response. In our dataset, the information gathered from the SPIN assessment is labeled as SPIN1, SPIN2, all the way up to SPIN17, each corresponding to one of the seventeen questions used to assess social phobia. *Greenspace Health, n.d., SPIN*

		Not at all	A little bit	Somewhat	Very much	Extremely
1.	I am afraid of people in authority	0	1	2	3	4
2.	I am bothered by blushing in front of people	0	1	2	3	4
3.	Parties and social events scare me	0	1	2	3	4
4.	I avoid talking to people I don't know	0	1	2	3	4
5.	Being criticized scares me a lot	0	1	2	3	4
6.	I avoid doing things or speaking to people for fear of embarrassment	0	1	2	3	4
7.	Sweating in front of people causes me distress	0	1	2	3	4
8.	I avoid going to parties	0	1	2	3	4
9.	I avoid activities in which I am the centre of attention	0	1	2	3	4
10.	Talking to strangers scares me	0	1	2	3	4
11.	I avoid having to give speeches	0	1	2	3	4
12.	I would do anything to avoid being criticized	0	1	2	3	4
13.	Heart palpitations bother me when I am around people	0	1	2	3	4
14.	I am afraid of doing things when people might be watching	0	1	2	3	4
15.	Being embarrassed or looking stupid are among my worse fears	0	1	2	3	4
16.	I avoid speaking to anyone in authority	0	1	2	3	4
17.	Trembling or shaking in front of others is distressing to me	0	1	2	3	4

Score	Symptom Severity
0 - 20	None
21-30	Mild
31-40	Moderate
41-50	Severe
51-68	Very severe

Additionally, we have columns labeled GAD_T, SWL_T, and SPIN_T, which represent the overall totals for the respective assessments: GAD (from questions 1 to 7), SWL (from questions 1 to 5), and SPIN (from questions 1 to 17). These, along with the "hours" column, will be integral components of our primary feature matrix for building the model in the future.

4. Data Cleaning Efforts and Exploration:

Once we imported the data into the data frame, we started by describing the numerical and categorical columns to see what we were dealing with. The mean of **Age** in our dataset is **20.9**. And the mean of **Hours** spent playing games is **22.2**. The **standard deviation** for **Hours** spent playing games is **70.28,** which is way too high**.** The max value of Hours spent playing games is **8000** and there is a max of 336 hours in 2 weeks since this dataset is collected for 2 weeks data and this reading explains the extremely high standard deviation for **Hours** spent playing games.

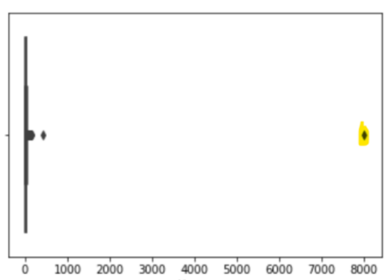
	Age	Hours	GAD_T	SWL_T	SPIN_T
count	13464.000000	13434.000000	13464.000000	13464.000000	12814.000000
mean	20.930407	22.247357	5.211973	19.788844	19.848525
std	3.300897	70.284502	4.713267	7.229243	13.467493
min	18.000000	0.000000	0.000000	5.000000	0.000000
25%	18.000000	12.000000	2.000000	14.000000	9.000000
50%	20.000000	20.000000	4.000000	20.000000	17.000000
75%	22.000000	28.000000	8.000000	26.000000	28.000000
max	63.000000	8000.000000	21.000000	35.000000	68.000000

Descriptive statistics for numerical columns.

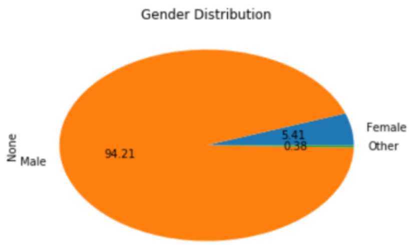
	Gender	Platform	Game	Work	earnings	whyplay	League	Degree	Playstyle
count	13464	13464	13464	13426	13464	13464	11626	13464	13464
unique	3	3	11	4	314	407	1455	5	298
top	Male	PC	League of Legends	Student at college / university	I play for fun	having fun	Gold	High school diploma (or equivalent)	Multiplayer - online - with real life friends
freq	12699	13219	11314	7073	12112	5289	970	8560	5564

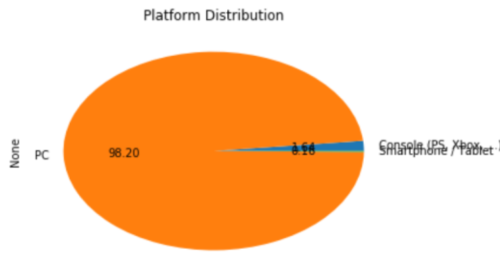
Descriptive statistics for categorical columns.

We started trying to clean up the numerical columns. First, we plotted a box plot for the number of **Hours** played. The boxplot showed one value which is around 8000 hours and one more around 500. So, we sorted the data frame on hours spent playing games in descending order, printed out the top 10, and saw two records 8000 and 420. So, we created a mask to remove these outliers and removed them.

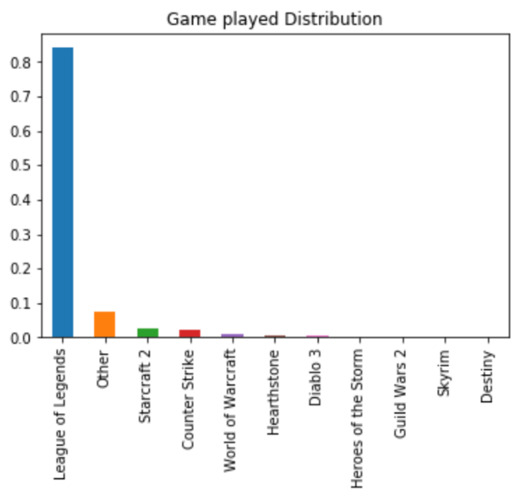


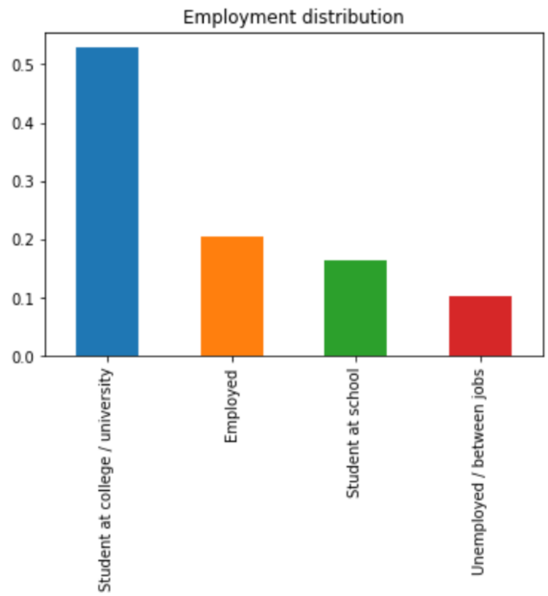
Then we saw there were a few NA in the column **SPIN_T.** So, we removed all those records from our dataset. Then we started plotting a few graphs showing the demographics of our dataset. First, we did a pie chart for Gender Distribution. Around **94.2%** of the sample identified themselves as **men**. We did the same to see the platform in which people played games and **98.2%** of the sample played on **PC**, **1.64%** on **Console**, and only **0.16%** on **tablets/mobile**.

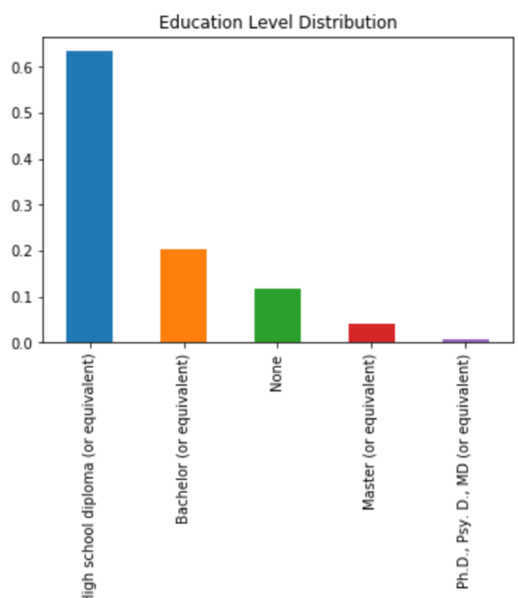




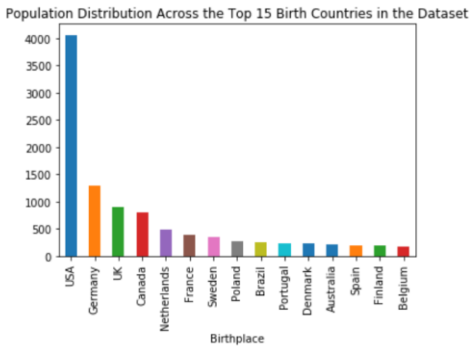
Then we started plotting a few bar charts for Games played, where people work, education level. And we have the screenshots for all the three plots here:

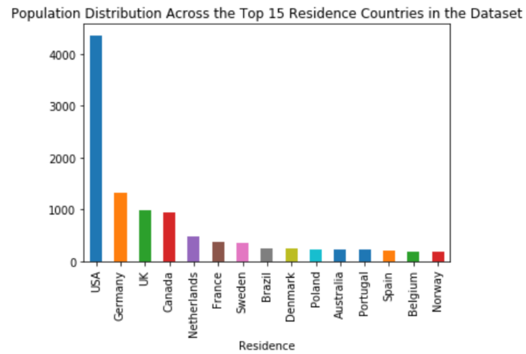






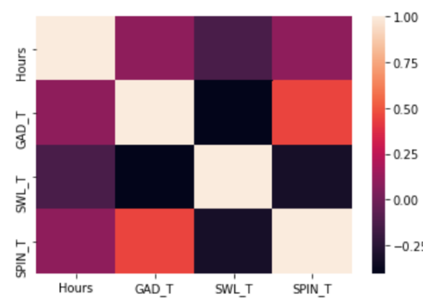
Then we wanted to learn more about the demographic concentration of our dataset. Such as residence country, and birth country. So, before we started plotting it, we made sure there were no duplicate values for the same country such as the USA and the United States of America. Once we checked that, we grounded the complete dataset by birthplace and size, sorted it in descending order, and used this data to plot our top 15 countries of residence and birth. The plots for this are posted below.





Then we wanted to find out the correlation between the main numerical variables that we will be using for future analysis (Hours, GAD_T, SWL_T, and SPIN_T). So, to do this, we got all the numbered columns from our dataset, dropped the columns we didn't need, and created a correlation matrix. Then we plotted a heatmap for this correlation matrix. The results are posted below.

	Hours	GAD_T	SWL_T	SPIN_T
Hours	1.000000	0.099027	-0.130112	0.098099
GAD_T	0.099027	1.000000	-0.403983	0.454435
SWL_T	-0.130112	-0.403983	1.000000	-0.318432
SPIN_T	0.098099	0.454435	-0.318432	1.000000



The correlation matrix shows that there is a negative correlation between the number of Hours spent playing a game and Satisfaction with life total. So as the number of hours of playing increases, satisfaction with life decreases. Both SWL and SPIN are very weakly positively correlated with the number of hours spent playing.

From the previous descriptive statistics of categorical variables that we did in the beginning, we had a few columns that we would need for the project such as **Playstyle**, **whyplay,** and **League**. So, we started to see what kind of data we were dealing with for these categorical variables. The league data we tried to get the details about leagues like gold, silver, platinum, diamond, unranked, bronze, challenger, and master. The actual dataset had values like gold 1, gold 2, diamond 1, etc. But we didn't need these many categories for our analysis. So, we first converted each and every value for the column league into lowercase and checked if each of the above-mentioned league values were in it and simply replaced it with gold, silver, etc. instead of gold 1, gold 2, etc. which helps us for future analysis. We did something similar to clean out whyplay and playstyle and we got it down to a very few categories of each of these variables so that it would be easier for us to work with. The number of records after cleaning the data is as follows.

```
League_clean
gold        3103
platinum    2549
Other       2258
silver      2183
diamond     1513
dtype: int64
```

```
whyplay_clean
fun          5242
improving    4716
winning      1980
relaxing      620
Other         126
dtype: int64
```

```
Playstyle_clean
Multiplayer - online - with real life friends                    5334
Multiplayer - online - with strangers                            3953
Multiplayer - online - with online acquaintances or teammates    2513
singleplayer                                                      749
Other                                                             148
dtype: int64
```

There might be more cleaning that might be needed further in the project but as of now, we cleaned as much data as we could.

5. Other Software Engineering Efforts:

At first, we began the data-cleaning process and generated basic descriptive statistics using R. However, as we found ourselves more comfortable with Python and appreciated the extensive libraries it offers, we made the decision to transition to Python. We employed **Jupyter Notebook** as our platform to work on the code.

We attempted to utilize **Geopandas** for visualizing our demographic data, aiming to display the distribution of individuals across various countries. However, we encountered installation issues with the software. As a result, we decided to abandon the idea of using Geopandas and opted for creating regular bar charts instead. These charts will showcase the top 15 countries with the highest number of individuals in our dataset.

6. Contributions:

a. ***Research Questions*** **-** Rajeevan, Pranav & Asmita
b. ***State of the Art*** **-** Rajeevan, Pranav & Asmita
c. ***Datasets*** **-** Rajeevan & Pranav
d. ***Data Cleaning Efforts*** **-** Rajeevan, Pranav & Asmita
e. ***Other Software Engineering Efforts*** **-** Rajeevan & Asmita
f. ***Presentation & Report*** **-** Rajeevan, Pranav & Asmita

7. References:

1. Draschkow, D., & Sauter, M. (2017, November 18). Gaming habits and psychological well-being: An international dataset about the anxiety, life satisfaction and social phobia of over 13000 gamers. OSF. https://osf.io/vnbxk/
2. Anxiety and Depression Association of America. (n.d.). GAD-7 Anxiety. https://adaa.org/sites/default/files/GAD-7_Anxiety-updated_0.pdf
3. Life Satisfaction | SWLS | Greenspace (CA). (2023, August 30). Greenspace (CA). https://greenspacehealth.com/en-ca/life-satisfaction-swls/
4. Social Anxiety | SPIN | Greenspace (US). (2023, April 3). Greenspace (US). https://greenspacehealth.com/en-us/social-anxiety-spin/
5. Jin, Y., Wang, R., Cook, A., & Fuller, R. D. (2023, September 1). Gaming during the COVID-19 Pandemic: Examining its Effect on Loneliness & Motivation, Playing and Gratification Differences between Competitive and Recreational Gamers. Telematics and Informatics Reports; Elsevier BV. https://doi.org/10.1016/j.teler.2023.100093
6. How Video Games Are Influencing Gen Z's Real-Life Behavior - YPulse. (2023, March 27). YPulse. https://www.ypulse.com/article/2023/03/27/how-video-games-are-influencing-gen-zs-real-life-behavior/
7. Von Der Heiden, J. M., Braun, B., Müller, K., & Egloff, B. (2019, July 26). The Association Between Video Gaming and Psychological Functioning. Frontiers in Psychology; Frontiers Media. https://doi.org/10.3389/fpsyg.2019.01731
8. Wei, H. T., Chen, M. H., Huang, P. C., & Bai, Y. M. (2012, July 28). The association between online gaming, social phobia, and depression: an internet survey. BMC Psychiatry; BioMed Central. https://doi.org/10.1186/1471-244x-12-92
9. Pallavicini, F., Pepe, A., & Mantovani, F. (2022, June 1). The Effects of Playing Video Games on Stress, Anxiety, Depression, Loneliness, and Gaming Disorder During the Early Stages of the COVID-19 Pandemic: PRISMA Systematic Review. Cyberpsychology, Behavior, and Social Networking; Mary Ann Liebert, Inc. https://doi.org/10.1089/cyber.2021.0252
10. Johannes, N., Vuorre, M., & Przybylski, A. K. (2021). Video Game Play Is Positively Correlated with well-being. Royal Society Open Science, 8(2), 202049. https://doi.org/10.1098/rsos.202049

-THE END-