**The Consequences of Video Game Consumption**

Data Analytical Project

Erik Lopez

**Objective of the Project**: The primary aim of this project is to showcase my analytical prowess and proficiency with various tools by examining a dataset generated through Python. There are three main objectives I seek to achieve with this dataset focused on gaming consumption. First, I aim to demonstrate my analytical skills across multiple software platforms, including Microsoft Excel, Microsoft Word, Jupyter Notebook, MySQL, and Looker Studio. Second, I intend to provide comprehensive insights into the dataset, enhancing my audience's understanding of the data. Finally, I strive to derive actionable recommendations based on the insights obtained from the data analysis. These recommendations will be aimed at improving gaming consumption habits and enhancing overall user engagement and satisfaction.

**Organization of Variables**

Demographic Information:

* **UserID**: Identification number unique to each participant (primary key).
* **First Name**: List of first names of participants.
* **Last Name**: List of last names of participants.
* **Age**: 18 – 35 (18 – 25 young adults, 26 – 35 mature adults)
* **Gender**: Female, Male, & Other.
* **Profession**: Engineer, Designer, Teacher, Developer, Student, Manager, Artist, Lawyer, Doctor, Nurse, Writer, & Other.
* **Income**: $0 - $200,000
* **Property** **Owner**: Yes/No (indicated if they own a home or not).

Gaming & Health Information

* **Gaming** **Devices**: PlayStation, Xbox, Nintendo, PC, Cell Phone, Tablet (preferred gaming device)
* **Video Game Genre**: First Person Shooters (FPS), Multiplayer Online Role-Playing Games (MMORPG), Role Playing Games (RPG), Fighting, Racing, Multiplayer Online Battle Arena (MOBA), Sports, Horror (favorite genre)
* **Hours Played Daily**: 1 hour to 12 hours
* **Monthly Purchases of Video Games**: $15 to $1000
* **Hours of sleep Daily**: 4 hours to 12 hours
* **Physical Activity**: Yes/No (Indicates if they work out or play sports)
* **Mental Wellness Score**: 1 – 100 (survey questions measuring mental wellness)
* **Importance Scale**: 1 – 100 (measures how important gaming is to them)
* **Productivity Loss**: 1 – 100 (measures how unproductive they feel due to gaming)
* **Satisfaction**: 1 – 100 (measures satisfaction from playing video games)
* **Self-Control**: 1 -100 (measures self-control in not playing video games)
* **Addiction Level**: 1 – 100 (measures the level of addiction to gaming)

**Random Data Generation Process**

**Jupyter Notebook** – used to generate random data using pandas and numpy (columns: Age, Gender, Profession, Income, Property Owner, Gaming Devices, Video Game Genre, Hours Played Daily, Monthly Purchases of Video Games, Hours of Sleep Daily, Physical Activity, Mental Wellness Score, Importance Scale, Productivity Loss, Satisfaction, Self-Control, & Addiction Level).

**Python IDLE** – used to generate random first and last names using faker package.

**Download** – I downloaded each into a CSV file and imputed them into an Excel Workbook file, which was then manipulated to have errors added to correct and make changes to answer the questions I wanted answered from the dataset.

**Analysis Plan**

**Step 1: Data Cleaning**

The data-cleaning process includes utilizing formulas within Excel to identify various errors found throughout the dataset. Some of the methods used were filtering & sorting data, finding and replacing features, creating pivot tables and graphics, and formulas like TRIM, VLOOKUP, CONCATENATE, and formatting data cells After the cleanup was done, I then moved it to MySQL to double check that it was clean, which I did find two outliers that I originally missed in Excel and missed a few issues with spacing causing two of the same categories, for instance, Female and Female with a space at the end. This caused me to go back to make some minor edits to the original file. After that, I then moved on to the next process, which was data analytics which involved using statistical techniques in Excel, MySQL, and Jupyter Notebook for better insight into the data.

**Step 2: Descriptive Analysis**

The first step is to perform a comprehensive descriptive statistical analysis using various tools like Microsoft Excel, Jupyter Notebook, and MySQL. This involves calculating measures such as minimum, maximum, average, sum, standard deviation, and variance. The purpose is to identify basic patterns and trends in the dataset, which will be visualized in the next steps.

**Step 3: Correlation and Hypothesis Testing Analysis**

Using Jupyter Notebook, I will conduct a correlation analysis focusing on numerical data columns like mental wellness score, hours played daily, hours sleep daily, importance score, addiction level, productivity loss, and self-control. This analysis aims to uncover significant relationships between these variables to understand how they might influence each other. Additionally, I will formulate hypotheses to statistically test the significance of these correlations to validate or refute potential relationships.

**Step 4: Trend Analysis**

To gain deeper insights into the trends between variables, I will utilize visualization tools such as Looker Studio and Excel. This involves creating various charts including pie charts, line charts, and heatmaps to visually represent the data patterns and relationships identified in the previous steps.

**Research Questions & Hypotheses**

**RQ1**: Do high amounts of gaming consumption lead to lower mental health among individuals who play video games for more than 5 hours a day?

**RQ2**: Is there a significant relationship between the average number of hours people play video games daily and their addiction level?

**RQ3**: Do certain types of gaming genres demonstrate a strong association with specific motivations that cause people to play video games, for instance, multiplayer online role-playing games attracting those who play video games for social interaction?

**RQ4**: What gender is shown to have a strong association between high gaming consumption and low sleeping hours?

**H1**: There is a negative correlation between hours played daily and importance score among male gamers.

**H2**: A significant positive correlation exists between hours played daily, hours sleep daily, and mental wellness score.

**H3**: A higher addiction level is associated with lower self-control scores across all types of gamers.

**H4**: Individuals who spend a great amount of money on video games monthly are associated with spending more time playing video games and being less productive.

**Gaming Dataset Insight**

**Descriptive Statistic Analysis**

The data demonstrates that, on average, gaming users are 26 years old, consume 6 hours of gameplay daily, sleep an average of 7 hours daily, spend around $500 monthly on gaming, and have a relatively low mental wellness score of 50. This indicates that while gamers invest a significant amount of their time and money in gaming and maintain close to 8 hours of sleep, their mental wellness scores are unhealthy.

Further investigation using Excel and MySQL reveals that those who use a Nintendo or Tablet for gaming tend to play more hours than users of other gaming devices. Visualizations like pie charts and tables show that the majority of participants are male, with female gamers playing more hours than any other gender. Male Nintendo users play the most hours, while other PlayStation users play the least. The average mental wellness scores by gender and gaming device are not optimal, generally falling around 50, far below the healthy range of 75 to 100. Analysis by hours played, hours slept, and gaming motivation also shows that wellness scores remain low, between 45 and 55. This suggests that gamers generally have unhealthy wellness scores regardless of their gaming or sleeping habits, a conclusion supported by correlations and regression analysis showing no significant relationships between the variables and mental wellness scores. The data also shows that the top five reason as to why people play video games is that they do so for challenging themselves, relaxation, competition, entertainment, and social interaction, which are all positive reasons for playing video games compared to others motivations like boredom, escapism, and loneliness.

**Correlation Analysis**

Based on the analysis conducted in Jupyter Notebook, it demonstrated that between all numerical variables, there was a weak negative to no relationship between them. This means that the variables do not significantly impact each other and that there may be other variables that are not accounted for in this project that can have a great impact. Look at data given and the heatmap displayed in Jupyter, some of the correlations are -0.051, 0.017, and many more close to or worse off demonstrating that they are not even close to the perfect positive correlation (1) or the perfect negative correlation (-1).

**Testing The Hypotheses:**

Testing the first hypothesis revealed no significant correlation between hours played daily and importance score among male gamers. Similarly, the second hypothesis test indicated no significant correlation between hours played, hours slept, and mental wellness scores. This result leads us to fail to reject the null hypothesis, suggesting that the evidence is insufficient to support a relationship between these variables. This could also imply that other unexamined variables might impact the dependent variable. In the third hypothesis test, we also fail to reject the null hypothesis, indicating no significant relationship between addiction level and self-control scores across all types of gamers. Lastly, the fourth hypothesis test demonstrated that we fail to reject the null hypothesis as well, indicating that gamers' spending habits have no significant relationship with the time spent playing video games and productivity loss.

**Research Questions:**

Based on the evidence presented through numerous statistical analyses, the research questions I aimed to answer did not yield positive results. The data consistently indicate that there is no significant correlation between the variables. Regression analysis further demonstrated that the variables have minimal to no effect on each other. The only question that may have been somewhat addressed is the fourth one, with Female Cell Phones users on average having the most hours played with the second least hours of sleep daily. However, even in this case, the association was not statistically significant.

**Regression Analysis**

Independent Variables: hours played daily, hours sleep daily, monthly gaming purchases, productivity loss, satisfaction, addiction level, self-control, importance score. Dependent Variable: mental wellness score.

The regression analysis shows low coefficients, indicating that the predictor variables have minimal impact on the mental wellness score. The high mean square error and low R-squared value suggest inaccuracy in the data, unaccounted variables, and a non-linear relationship between the variables.

**Final Thoughts & Recommendations**

After conducting various statistical analyses and developing multiple visualizations, it can be inferred that the randomly generated dataset is unrealistic, as there is substantial statistical evidence indicating no significant correlations between any of the variables. Consequently, we can conclude that, based on this dataset, there are no negative consequences associated with high levels of video game consumption. For instance, the data shows that variables such as hours played daily, hours slept daily, productivity loss, and many others have no significant impact on mental wellness scores. This is concerning because the data does not lead us to a definitive conclusion.

Although the data was randomly generated, if perceived in a realistic scenario, the analysis process suggests that there may be unaccounted variables. One potential variable to consider is mental health history, such as diagnoses. Additionally, there are limitations regarding certain variables like physical activity, which were not examined in this project. In future studies or projects, it would be beneficial to elaborate on the types of physical activities participants engage in, such as gym workouts, running, walking, or playing basketball. Investigating how different types of physical activity impact gaming consumption would be a valuable direction for future research.

If I were given this dataset by an actual company, I would present the information as detailed in this document. If the company still requested trends and patterns for practical purposes, I would mention that male gamers tend to play longer hours on average, which could inform the development of games targeting this specific gender. Additionally, the data shows that Xbox, PC, and mobile device gamers spend a significant amount of money on video games monthly. Therefore, game developers might consider focusing on these platforms when creating exclusive titles. These are just a few recommendations based on the dataset. However, it would be best to inquire whether the company has additional data available or is willing to use secondary or third-party data sources beyond the primary one provided.

**Cleaning the Dataset Process**

**Removing Duplicates**:

* Select the entire dataset > click data tab > then remove duplicates > in the remove duplicates box select only UserID > then OK.

[Creating a pivot table for each sheet to look for any errors that need to be resolved]

**Misspelling & Typographic Errors:**

* Column Section: Satisfiction
  + Correction: Satisfaction
* Column Section: ProductivityLoss
  + Correction: Productivity Loss
* Gender: Femlee, Mle, O, blank
  + Correction: Corrections: Home tab > find and select > click replace > check match entire cell contents > input the error then input the correction > click replace all > close.
* Profession: Artis, Doc, StudNT, Writre, blank
* Property Owner: Y, N, blank
* First Name: Jack”space”, John “space”, Jonathon, Lily”space”, Lilyy, Lincoin”space”, Logan”space”, Micheal, Nathaniet, Olivia”space”, Penolope”space”, Samual, Sebastian”space”, Shelia, Victoria”space”, Gabrielle, Ethan”space”, Clara”space”, Charles”space”, Carmla, Cameron”space”, Antonio”space”, Arthur”space”, Anthony”space”, Mathew, Valeria
  + Corrections: Home tab > find and select > click replace > check match entire cell contents > input the error then input the correction > click replace all > close.
* Last Name: Adama
  + Corrections: Home tab > find and select > click replace > check match entire cell contents > input the error then input the correction > click replace all > close.
* Gaming Devices: Cell, PlaStation
  + Corrections: Home tab > find and select > click replace > check match entire cell contents > input the error then input the correction > click replace all > close.
* Gaming Motivation: Comp, Lonely, Relax
  + Corrections: Home tab > find and select > click replace > check match entire cell contents > input the error then input the correction > click replace all > close.
* Video Game Genre: Fight, Horrr
  + Corrections: Home tab > find and select > click replace > check match entire cell contents > input the error then input the correction > click replace all > close.

**Removing Blank Cells**: We will remove 13 rows that have blank values to make the data structure more consistent, prevent issues during the analysis phase, and make it more presentable. Also, since we have plentiful data, removing 12 rows will not significantly impact our data.

* Correction: Select entire dataset > home tab > select & find > go to special > check blanks > click OK.
  + Delete the rows with highlighted blank values.
  + UserID: 10006, 10010, 10018, 10091, 10144, 10267, 10402, 10414, 10510, 10522, 10589, 10641, 10616

**Inconsistent Data Types**:

* Monthly Purchases on Video Games
  + Correction: Select data range > right click > format cells > currency > 0 decimal place > click OK.
* Income Column
  + Correction: Select data range > right click > format cells > currency > 0 decimal place > click OK.

Outliers & Anomalies:

* Sorted & Filtered data and manually checked for them.

**Additional** **Edits**:

* Combining first name and last name into one column.
  + Correction: =CONCATENATE(cell 1, “space”, cell 2) : Example =CONCATENATE(B2, “ “, C2)
  + Double-click the bottom right of the cell to copy and paste to the rest of the cells in the new column Name.
* Change Monthly Purchases on Video Games
  + Correction: Monthly Gaming Purchases
* Change Income
  + Correction: Yearly Income

**Combine Spreadsheets**: There are two Excel sheets in the workbook (user\_personal\_information and gaming\_health\_data). It would be best to combine them to have all essential data in a single spreadsheet, which would make it easier to edit and provide insight into the relationships between variables when finally uploaded into Jupyter Notebook for analysis and visualization in Power BI.

* Create a new spreadsheet: [video\_game\_dataset\_v01]
* Transferring Data:
  + Copy and paste straightforward static data that doesn’t need frequent updates: UserID, First Name, Last Name, Age, Gender, Profession, Income, Property Owner, & Physical Activity.
  + Using VLOOKUP for the more dynamic data that may need to be updated in the future: Gaming Device, Video Game Genre, Hours Played Daily, Monthly Purchases of Video Games, Hours of Sleep Daily, Mental Wellness Score, Importance Score, Productivity Loss, Satisfaction, Self-Control, & Addiction Level.
    - Correction:
    - Gaming Devices =LOOKUP(A2, gaming\_health\_data!A:N, 2, FALSE)
    - Video Game Genre =LOOKUP(A2, gaming\_health\_data!A:N, 3, FALSE)
    - Gaming Motivation =LOOKUP(A2, gaming\_health\_data!A:N, 4, FALSE)
    - Hours Played Daily =LOOKUP(A2, gaming\_health\_data!A:N, 5, FALSE)
    - Monthly Gaming Purchases =LOOKUP(A2, gaming\_health\_data!A:N, 6, FALSE)
    - Hours Sleep Daily =LOOKUP(A2, gaming\_health\_data!A:N, 7, FALSE)
    - Mental Wellness Score =LOOKUP(A2, gaming\_health\_data!A:N, 9, FALSE)
    - Importance Score =LOOKUP(A2, gaming\_health\_data!A:N, 10, FALSE)
    - Productivity Loss =LOOKUP(A2, gaming\_health\_data!A:N, 11, FALSE)
    - Satisfaction =LOOKUP(A2, gaming\_health\_data!A:N, 12, FALSE)
    - Self Control =LOOKUP(A2, gaming\_health\_data!A:N, 13, FALSE)
    - Addiction Level =LOOKUP(A2, gaming\_health\_data!A:N, 14, FALSE)

[Note: During my Visualization development, I did end up finding additional errors that needed fixing like one misspelled word and one outlier, which was easily addressed using find and replace option. Other methods used in the future to make sure this does not happen again is to use functions like Inter-Quartile Range (calculate Q1 and Q3 > calculate IQR > determine lower and upper bounds > identify outliers) or using Z-Scores (anything greater than 3 or less than -3 are considered outliers)]

**Data Analytical Process**

After cleaning data, I then moved to developing statistical descriptions in an excel spreadsheet, the purpose of this was to get a better understanding of the data I will be using for this project.

* **Demographic Information**
  + The demographic information that I identified are total participants, average age, gender distribution, average income, and property owner.
  + The formulas used were =COUNTIF and =AVERAGE
  + Example: =AVERAGE(video\_game\_dataset\_v01!C2:C981)
* **Gaming & Health Information**
  + This section included avg hours played daily, avg monthly purchases, avg hours sleep daily, physical activity, avg mental wellness score, avg productivity loss, avg importance score, avg addiction level, avg self-control, and avg satisfaction.
  + The formulas used were =AVERAGE, and =COUNTIF
  + =COUNTIF(video\_game\_dataset\_v01!H2:H981, "Yes")
* **Gender & Gaming Devices**
  + Made a table with information about participants distribution by gender and gaming devices.
  + The formulas used were =COUNTIFS
  + Example: =COUNTIFS(video\_game\_dataset\_v01!$D$2:$D$981, "Female", video\_game\_dataset\_v01!$I$2:$I$981, "PlayStation")
* **No. of Participants by Gaming Devices**
  + This is to see how many participants play video games by gaming devices.
  + Formula used =SUM
  + Example: PlayStation =SUM(K2, K9, K16)
* **Avg Playtime by Gaming Devices**
  + This table contains information that looks at the average hours played daily by gaming devices like PlayStation, Xbox, PC, Nintendo, Cell Phone, and Tablet.
  + The formula used is =SUM/Cell (adding the total of avg playtime by gender and gaming devices from another table made)
  + Example: PlayStation Playtime ==SUM(P2, P9, P16)/K24
* **Avg Gaming Purchases by Gender & Devices**
  + Looks at the average distribution of monthly purchases by gender and gaming devices.
  + The formula is =AVERAGEIFS
  + Example: Female PlayStation Purchases =AVERAGEIFS(video\_game\_dataset\_v01!$M$2:$M$981, video\_game\_dataset\_v01!$D$2:$D$981, "Female", video\_game\_dataset\_v01!$I$2:$I$981, "PlayStation")
* **Avg Gaming Purchases by Gaming Devices**
  + Investigates the average monthly gaming purchases by gaming devices like PlayStation, Xbox, PC, Nintendo, Cell Phone, and Tablet.
  + The formula used is =AVERAGEIFS
  + Example: PlayStation Monthly Purchases =AVERAGEIFS(video\_game\_dataset\_v01!$M$2:$M$981, video\_game\_dataset\_v01!$I$2:$I$981, "PlayStation")
* **Avg Mental Wellness Score by Gaming Devices & Gender**
  + The purpose was to get an idea as to what the average mental wellness scores was when accounting for gaming devices and gender.
  + Formula used =AVERAGEIFS
  + Example: Female PlayStation Wellness Sc =AVERAGEIFS(video\_game\_dataset\_v01!$O$2:$O$981, video\_game\_dataset\_v01!$N$2:$N$981, ">=4", video\_game\_dataset\_v01!$N$2:$N$981, "<=12", video\_game\_dataset\_v01!$L$2:$L$981, ">=1", video\_game\_dataset\_v01!$L$2:$L$981, "<=12", video\_game\_dataset\_v01!$I$2:$I$981, "PlayStation")
* **Avg Mental Wellness Score by Hours Played & Hours Sleep & Gaming Devices**
  + Used as a mean to get an understanding of the mental wellness scores accounting for three variables, like hours played daily, hours sleep daily, and gaming devices.
  + Formula used =AVERAGEIFS
  + Example: Avg Wellness by Hours Played & Sleep & PS =AVERAGEIFS(video\_game\_dataset\_v01!$O$2:$O$981, video\_game\_dataset\_v01!$N$2:$N$981, ">=4", video\_game\_dataset\_v01!$N$2:$N$981, "<=12", video\_game\_dataset\_v01!$L$2:$L$981, ">=1", video\_game\_dataset\_v01!$L$2:$L$981, "<=12", video\_game\_dataset\_v01!$I$2:$I$981, "PlayStation")
* **Avg Mental Wellness Score by Hours Played Daily & Hours Sleep Daily & Gender**
  + Investigate the average mental wellness scores well accounting for hours played daily, hours sleep daily, and gender.
  + Formula used =AVERAGEIFS
  + Example: Avg Wellness Sc by Hours Played & Sleep & Female =AVERAGEIFS(video\_game\_dataset\_v01!$O$2:$O$981, video\_game\_dataset\_v01!$N$2:$N$981, ">=4", video\_game\_dataset\_v01!$N$2:$N$981, "<=12", video\_game\_dataset\_v01!$L$2:$L$981, ">=1", video\_game\_dataset\_v01!$L$2:$L$981, "<=12", video\_game\_dataset\_v01!$D$2:$D$981, "Female")
* Avg Mental Wellness Score by Hours Played Daily & Hours Sleep Daily & Gaming Motivation
  + Investigate the average mental wellness score well accounting for hours played daily, hours sleep daily, and various gaming motivation like entertainment, relaxation, stress relief, etc.
  + Formula used =AVERAGEIFS
  + Avg Wellness Sc by Hours Played & Sleep & GMotivation =AVERAGEIFS(video\_game\_dataset\_v01!$O$2:$O$981, video\_game\_dataset\_v01!$N$2:$N$981, ">=4", video\_game\_dataset\_v01!$N$2:$N$981, "<=12", video\_game\_dataset\_v01!$L$2:$L$981, ">=1", video\_game\_dataset\_v01!$L$2:$L$981, "<=12", video\_game\_dataset\_v01!$K$2:$K$981, "Entertainment")

The next step was simply to develop visualization in Excel to demonstrate that I can make them, but the visualizations that will be used to identify trends and patterns will be developed using Looker Studio. After that, I then moved on to MySQL to basically demonstrate the simple foundation of commands to identify patterns and trends using descriptive statistical analysis. To get futher insight into the dataset, I then moved to Jupyter Notebook (Python IDE) like hypothesis testing, correlation analysis, and regression analysis [The process of each can be seen with the files associated with each of them]