

Online Gaming Behavior Analysis

Gregory A. Martin

Western Governors University



## Table of Contents

A. Proposal Overview.....	5
A.1 Research Question or Organizational Need.....	5
A.2 Context and Background.....	5
A.3 and A3A Summary of Published Works and Their Relation to the Project.....	5
Review of Work 1 .....	5
Review of Work 2 .....	6
Review of Work 3 .....	6
A.4 Summary of Data Analytics Solution .....	7
A.5 Benefits and Support of Decision-Making Process.....	7
B. Data Analytics Project Plan.....	7
B.1 Goals, Objectives, and Deliverables .....	7
B.2 Scope of Project.....	8
B.2.A Included in Project Scope.....	8
B.2.B Not included in Project Scope.....	8
B.3 Standard Methodology .....	8
B.4 Timeline and Milestones .....	8
B.5 Resources and Costs .....	9
B.6 Criteria for Success.....	9
C. Design of Data Analytics Solution.....	9
C.1 Hypothesis .....	9
C.2 and C.2.A Analytical Method .....	9
C.3 Tools and Environments.....	9
C.4 and C.4.A Methods and Metrics to Evaluate Statistical Significance.....	10
C.5 Practical Significance.....	10
C.6 Visual Communication.....	10
D. Description of Dataset.....	11
D.1 Source of Data.....	11
D.2 Appropriateness of Dataset.....	11
D.3 Data Collection Methods.....	11
D.4 Observations on Quality and Completeness of Data .....	11
D.5 and D.5.A Data Governance, Privacy, Security, Ethical, Legal, and Regulatory Compliances .....	11
References .....	13
Appendix A.....	<b>Error! Bookmark not defined.</b>
Title of Appendix .....	<b>Error! Bookmark not defined.</b>
Appendix B.....	<b>Error! Bookmark not defined.</b>

Title of Appendix .....	<b>Error! Bookmark not defined.</b>
Appendix C.....	<b>Error! Bookmark not defined.</b>
Title of Appendix .....	<b>Error! Bookmark not defined.</b>
Appendix D.....	<b>Error! Bookmark not defined.</b>
Title of Appendix .....	<b>Error! Bookmark not defined.</b>

## **A. Proposal Overview**

I intend to increase revenue of a fictional company by performing data and statistical analysis. By looking at certain metrics, I plan on providing valuable insight derived from my analysis, which can then be used as the focal point of the company to increase revenue.

### **A.1 Research Question**

Is there a correlation between higher levels of player engagement and in-game purchases? And if not, is there some other relevant data that could be used by video game developers to draw conclusions necessary to increase profitability?

### **A.2 Context and Background**

The video game market is saturated and player's attention spans vary rapidly, so knowing which elements to focus on will help developers make informed decisions about how to monetize their game, or which elements of game design to focus on.

### **A.3 and A3A Summary of Published Works and Their Relation to the Project**

#### **Review of Work 1**

My first referenced article *Addiction and Spending in Gacha Games* deals with the spending habits of players in Gacha style games. Games that fall into this category are generally compared to gambling due to their high randomness and flashy gameplay. This article provided lots of data and was more focused on the addictive properties and spending habits of players rather than from the developer's perspective. However, the article did provide some interesting statistics that we may find supportive and relevant. For example, Figures 4 and 5 indicate that the longer a player spends playing the same game, the more money they are willing to spend. *“By taking a look at the columns, it can be seen that most participants who played the game for up to*

*12 months spent a total of USD 0 to USD 50. When looking at medium lengths of playing (2 to 3 years), it can be noticed that these participants had a tendency towards medium/higher consumption. Participants who have been playing the game for more than 4 years show a drastic tendency for higher spending.*” (Lakić, Bernik, & Čep, 2023). This finding corroborates my hypothesis.

## **Review of Work 2**

We can establish the fundamentals of video game design by looking at the next reference *Elements of Game Design* (Robert Zubek, 2020). This book is more from the developer’s perspective and offers valuable insight into the decision making process of game design. On page 7, figure 1.3 explains that the designer can affect player experience by manipulating the underlying mechanics and various systems. What may seem obvious is the first approach of a top-down design: that is the game is designed with an experience already in mind, and the end goal. But what I found interesting is the bottom-up approach: where the design is more exploratory based on what “feels good”. We can conclude that designers may dispatch one, or even both of these strategies, especially when it comes to mobile games or gacha style games. If you have ever played one of these games, they usually include many different types of gameplay, in the hopes that players will enjoy at least one or more of them. In other words - cast a wide net, so that you can attract as many types of players as possible.

## **Review of Work 3**

The last reference goes into the psychology of control, or even perceived control, in video games. What I find most compelling comes from the final reference *Ctrl-Alt-Play: Essays on Control in Video Gaming* (Wysocki, 2013). On page 24, Wysocki notes that some people have a “false perception of having personal control over chance-determined events”. Further reading indicates that when

a player is properly challenged and rewarded, they are more likely to continue engaging with the game. On page 29 - “Players with both high self-efficacy and outcome expectations for a game display high levels of effort, persistence in the face of challenges, and are highly engaged in playing the game.” And we know from the first referenced work that higher engagement from players is more likely to lead to purchases.

#### **A.4 Summary of Data Analytics Solution**

My solution to this analytics problem is to perform exploratory data analysis by using Python and Jupyter Notebooks. I’ll download the dataset from Kaggle, import it into Jupyter Notebooks, and perform any necessary data cleaning. Then, I’ll perform a series of analysis, such as correlation, sorting and statistical analysis. Any relevant findings from the analysis will be visualized using charts and graphs from either the matplotlib or seaborn libraries.

#### **A.5 Benefits and Support of Decision-Making Process**

The benefit gained by this analysis will show what the developer can focus on for maximizing revenue. The analysis will provide valuable insight into player demographics and metrics, giving the developer an edge and making informed decisions about the best way to monetize their product.

### **B. Data Analytics Project Plan**

#### **B.1 Goals, Objectives, and Deliverables**

The goal of this project is to create a Python application that will give insight into player behavior for the purpose of maximizing profit and revenue.

The objective will be to perform exploratory data analysis and visualizing the results.

The deliverable will be the resulting data output and visualizations.

- Goal: create a Python application that will give insight into player behavior for the purpose of maximizing profit and revenue.
  - Objective: perform exploratory data analysis and visualizing the results.
    - Deliverable 1: Data output
    - Deliverable 2: Visualizations

## B.2 Scope of Project

The scope of this project will be a single python script. The script will read in the dataset, perform any cleaning steps, and output formatted data and visualizations using a number of python libraries.

### B.2.A Included in Project Scope

Included in the project scope: Data cleaning, data analysis, data visualization and hypothesis testing.

### B.2.B Not included in Project Scope

We will not be using any machine learning algorithms in this analysis.

## B.3 Standard Methodology

This project will use the Agile methodology. Agile fits this project well due to the exploratory nature of the project. This will help the developer make informed decisions as more information becomes available during analysis, and allow me to make adjustments along the way.

## B.4 Timeline and Milestones

Milestone or deliverable	Duration (hours or days)	Projected start date	Anticipated end date
Task 1	1 day	<i>July 1st</i>	<i>July 2nd</i>



Task 2	3 days	<i>August 1st</i>	<i>August 4th</i>
Task 3	3 days	<i>August 5th</i>	<i>August 8th</i>

### **B.5 Resources and Costs**

1. Computer Laptop: Already owned
2. Anaconda: No cost
3. Jupyter Notebooks: No cost
4. Work Hours: About 14

### **B.6 Criteria for Success**

A positive correlation is found, or else some other data metric is identified as being statistically significant through exploratory analysis.

## **C. Design of Data Analytics Solution**

### **C.1 Hypothesis**

My hypothesis is there is a positive correlation between higher player engagement and higher number of in game purchases.

### **C.2 and C.2.A Analytical Method**

I will use descriptive analysis as the analytical method in this project.

#### **Rubric C2A:**

This kind of test is appropriate simply because it will give valuable insights into the data trends.

### **C.3 Tools and Environments**

Python and Jupyter Notebooks will be the tools used for this project. These tools provide a great environment for the required analysis.

## **C.4 and C.4.A Methods and Metrics to Evaluate Statistical Significance**

My null hypothesis is that there is no correlation between in-game purchases and player engagement. The statistical test that will be used is a correlation and probability test. I will use a correlation coefficient and failing that, I will use probability analysis.

### **Rubric C.4.A:**

This is an appropriate choice for this scenario due to the exploratory nature of this analysis. If my hypothesis is correct, then we can see a positive correlation of two or more data points. For example, as player engagement increases, so does in-game purchases. However, if my hypothesis is not correct, and there is no correlation, then we can rely on probability as a metric for success.

## **C.5 Practical Significance**

The practical significance of these results will create real world benefit to the company in the form of revenue increase. For example, if the results show a positive correlation between player engagement and in-game purchases, then we can conclude that design decisions which involve driving player engagement will also lead to an increase for in-game purchases. Therefore, the results would be shown to be practically significant, because we would see a direct increase in revenue due to the action taken as a result of the analysis.

## **C.6 Visual Communication**

I will use several visualizations in this project. I will include one line chart which shows correlation. Then I will use two histograms which show the distribution of average playtime between two subsets of data.

## D. Description of Dataset

### D.1 Source of Data

I will use the data set from Kaggle.com:

<https://www.kaggle.com/datasets/rabieelkharoua/predict-online-gaming-behavior-dataset>

### D.2 Appropriateness of Dataset

This data set is appropriate for the project because it provides many different data and demographics about players and represents a realistic data set.

### D.3 Data Collection Methods

The data was collected by downloading the .csv file from <https://www.kaggle.com/datasets/rabieelkharoua/predict-online-gaming-behavior-dataset>

### D.4 Observations on Quality and Completeness of Data

The quality and completeness of the data is good. There will be little to no cleaning required which means more time can be spent on doing the analysis. However, one issue to note with the data is that the *InGamePurchases* column is of a binary data type. The data is either a 0 or a 1, which means either the player has spent money, or has not. This may provide a challenge when attempting to plot the data. But we can still derive meaningful insights and will have little impact on the outcome.

### D.5 and D.5.A Data Governance, Privacy, Security, Ethical, Legal, and Regulatory Compliances

The data is public and freely available to anyone without any restrictions. There is no concern for governance, privacy, security, ethical, legal or regulatory issues. There is some demographic information included such as Gender, Age and Location (country), but there is no concern as these are not considered PHI or PII, so each player's privacy is protected.

**D.5.A Precautions**

Governance: There are no precautions required for data governance

Privacy: There is some demographic information in the data which includes Age, Gender and Location (country). But this data does not identify anyone, so privacy is not a concern.

Ethical: There are no precautions required for ethical concerns

Legal: There are no precautions required for legal concerns

Regulatory: There are no precautions required for regulatory concerns

## References

- Lakić, N., Bernik, A., & Čep, A. (2023, July 13). *Addiction and Spending in Gacha Games*. Retrieved from mdpi.com: <https://www.mdpi.com/2078-2489/14/7/399>
- Robert Zubek. (2020). *Elements of Game Design*. Retrieved from books.google.com: [https://www.google.com/books/edition/Elements\\_of\\_Game\\_Design/Lr\\_4DwAAQBAJ?hl=en&gbpv=1&dq=elements+of+a+good+game&pg=PR13&printsec=frontcover](https://www.google.com/books/edition/Elements_of_Game_Design/Lr_4DwAAQBAJ?hl=en&gbpv=1&dq=elements+of+a+good+game&pg=PR13&printsec=frontcover)
- Wysocki, M. (. (2013, February 7). *Ctrl-Alt-Play: Essays on Control in Video Gaming*. Retrieved from books.google.com: [https://books.google.com/books?hl=en&lr=&id=PZdquVHyxUIC&oi=fnd&pg=PA21&dq=psychology+of+video+games&ots=Y\\_wHllzcK-&sig=Cr\\_YpvuuN2Aad-7wGiBkOJ58MM0#v=onepage&q=illusion%20of%20control&f=false](https://books.google.com/books?hl=en&lr=&id=PZdquVHyxUIC&oi=fnd&pg=PA21&dq=psychology+of+video+games&ots=Y_wHllzcK-&sig=Cr_YpvuuN2Aad-7wGiBkOJ58MM0#v=onepage&q=illusion%20of%20control&f=false)