

STA304 - Fall 2022

Assignment 1

Alexander Tran - 1006314089

October 6, 2022

Part 1

Goal

Genshin Impact is a free-to-play video game developed by miHoYo, first launched on September 28, 2020. Due to its overwhelming success and popularity, it has since then generated a whopping \$3.7 billion in revenue from player spending worldwide across the App Store and Google Play Store.^[4]

As a developer of a live-service game, one of your primary goals is to generate revenue. Thus, it is important to know who your players are and what they are spending on so that you know which products are doing well, and which products to put more development into.

Having said that, the goal of my survey is to offer insight into the spending behaviour of players in Genshin Impact by obtaining data on the demographic of the playerbase, as well as how much they spend and on which items.

Procedure

The survey will be created using Google Forms. In addition to being sent to personal friends, the survey will be sent in several Genshin Impact Discord servers, as well as posted in the Genshin Impact subreddit.

The target population of the survey is the entirety of the Genshin Impact playerbase. The frame population is the set of all Genshin Impact players that are either in a Discord server where the survey was sent, use the Genshin Impact subreddit, or are my personal friends. The sample population consists of those saw the survey message/post, and the sample itself consists of those who volunteered to respond to the survey.

A drawback of this procedure is that the Discord servers and subreddit are English-speaking communities. As Genshin Impact is a game with players from all over the globe, this survey is very likely to miss those that are not from non-English speaking countries, thus being unrepresentative of the entire Genshin Impact playerbase. Similarly, as the survey is written in English, those that are not comfortable with the English language may be less likely to respond.

However, a strength of utilizing Discord servers and the subreddit are that it increases the likelihood that the survey reaches more people. Combined, the servers and the subreddit have over 2 million non-unique members, which have the potential to provide a large amount of data.

Another drawback is that this is volunteer-based sampling. While this method does have its strengths in convenience, time-efficiency, and cost-effectiveness, it is a form of non-probability sampling and is not representative of the entire target population.

I would also like to acknowledge that there maybe bias issues in sending the survey to my personal friends. My personal friends are of a similar demographic to myself, so if the sample size is small enough, they will represent a considerable proportion of the sample data, and may skew the data.

Showcasing the survey.

<https://forms.gle/gAPnogZtw4VdBd3V7>

How much money have you spent in total on Genshin Impact? Please write your answer in USD, to the nearest dollar.

This question asks the survey participant to state how much money they have spent on the game, in USD, rounded to the nearest dollar. The answer must be numerical and greater than or equal to 0. Since we want to understand the spending behaviour of players, knowing how much they spend should be very important, as we can see which players are not spending money, which ones are, and how much they are spending.

One drawback of this question is that it asks for the amount in USD. For many people, converting a currency may be considered too much of a hassle to deal with in a quick survey, which may lead to inaccurate answers. Furthermore, players may not immediately know how much they have spent in total, and are unlikely to review their entire purchase history to answer this question, again leading to inaccurate answers.

What in-game products have you spent money on? (Select all that apply)

This question lists all in-game products that money can be spent on, and asks the survey participant to select all that they have spent money on. Participants can check as many as they like, and an 'I have not spent any money' option, as well as an 'other' option are included. This question allows us to see what items players are spending money on, which allows us to determine which items are most successful, and which ones may need further improvement.

A benefit of this question is that the number of options is not overwhelming, while still covering almost all of the purchasable items, with those not covered being left to the 'other' option. Thus, survey participants are unlikely to be dismayed at the sight of this question, and data can still be collected with minimal downsides, as the excluded items are those that are perceived to be bought by virtually nobody.

A drawback of this question is that some of these items are purchasable with in-game currency that does not require spending any real money, and while each answer was written to be as clear as possible, there may still have been confusion, leading to the overestimation of some of these options.

How satisfied are you with the products that you've spent money on? Please do not answer this question if you have not spent any money.

This question asks the survey participant to rate on a scale of 1 to 7 how satisfied they are with their purchases in Genshin Impact, with 1 being 'very unsatisfied' and 7 being 'very satisfied'. Knowing how a buyer feels about a product after buying it is arguably just as important as knowing whether or not they bought the product in the first place. This question lets us gauge satisfaction with in-game purchases and can indirectly represent how likely they are to purchase it again.

A benefit of this question comes from its simplicity. It is a relatively quick question to answer, and gives us a general idea about player sentiment.

A drawback of this question is that 'very unsatisfied' and 'very satisfied' can be subjective, especially when attributed to a number scale (1-7). Thresholds for different satisfaction levels may vary between people, so we cannot tell what a given response may mean on an absolute level, making comparisons difficult. People may also be hesitant to answer with a 1 or a 7, as they are the answers on the extremes, and may be more likely to choose more neutral options.

Part 2

Data

54 survey responses were collected in total, consisting of Genshin Impact players found on the Genshin Impact subreddit, various Genshin Impact focused Discord servers, and personal friends who were asked to fill out the survey. The data was imported from Google Forms into a spreadsheet.

One drawback of the data is its limited sample size. A sample size of 54 is not a very large sample size, especially since Genshin Impact has millions of players. This sample size is much smaller than what was expected, suggesting that the methods of sending out the survey were not effective. Considering the nature of Discord servers, this is understandable, as messages in a chat quickly become lost when many people are talking. A similar idea applies to Reddit, as it is likely that only a fraction of the community sees your post. Furthermore, while over 2400 people viewed the Reddit post, there were only 54 total responses. Because this survey was relatively quick and simple, I am lead to believe that the lack of response is due to disinterest and lack of incentive.

Another drawback is that there are 0 responses from players playing on the China server, and the Taiwan, Hong Kong, Macau server. These servers represent a large portion of the playerbase, and thus it is highly unlikely that this sample is representative of the global playerbase. The remaining servers (American, Europe, Asia) have a reasonable distribution of responses considering more responses from English-speaking regions were expected.

Cleaning Process

Column names were renamed to represent the indicated variable. Previously the column names were the survey questions themselves.

Genders submitted under ‘other’ were standardized to one spelling of each gender e.g. ‘Nonbinary’ and ‘Non-binary’ were standardized to ‘Non-Binary’. This was done for 4 observations.

When asking how much money a player spent, it was requested that they round to the nearest dollar. Some respondents did not follow this instruction, so their answer was manually rounded the nearest dollar. This was done for two observations.

Important Variables

Year of Birth: What year the respondent was born in. Numerical variable.

Gender: What gender the respondent identifies with. Categorical variable.

Server: Which game server the respondent plays Genshin Impact in. Categorical variable taking on values ‘America’, ‘Europe’, ‘Asia’, ‘China’, and ‘Taiwan, Hong Kong, Macau’.

Play Time: How long the player has been playing Genshin Impact, in months. Numerical variable.

Money Spent: The total amount of money the respondent has spent in-game. Value in USD, rounded to the nearest dollar. Numerical variable.

Items Bought: A list of comma-separated items that the respondent purchased in-game.

Satisfaction: How satisfied the respondent said they were with the items they purchased. Categorical variable taking a numerical value from 1-7, representing ‘very unsatisfied’ to ‘very satisfied’.

Numerical Summaries

Year of Birth: The median year was 2000. The youngest respondent was born in 2007. The oldest respondent was born in 1978. The 25th and 75th quantiles were 1995 and 2003, respectively. The standard deviation was 6.167. The distribution of respondents in terms of age is fairly even, slightly favouring younger players.

Play Time: The median amount of time played was 16 months. The newest respondent has only played for 2 months. The longest-playing respondent has played for 24 months. The 25th and 75th quantiles were 12.2 and 23, respectively. The standard deviation was 6.745. The sample distribution appears to lean towards players who have played a longer period of time.

Money Spent: The mean amount of money spent was \$309.35. The median was \$62.50. The least amount of money spent by a single respondent was \$0. The most amount of money was \$3000. The 25th and 75th quantiles were 0 and 224, respectively. The standard deviation was 653.306. Many players spend little to no amounts of money, but the spread is fairly large among those who do spend money.

Figure 1: Server Distribution Bar Plot

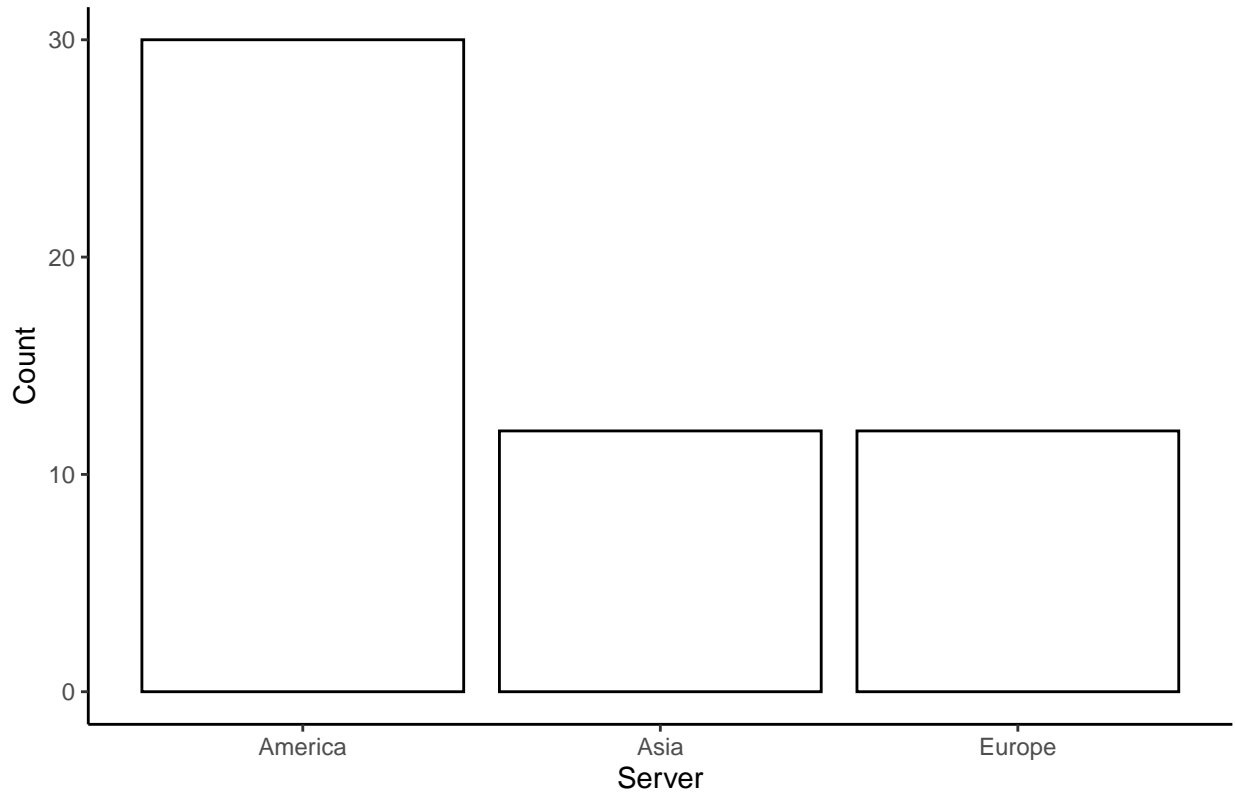


Figure 2: Satisfaction Distribution Plot

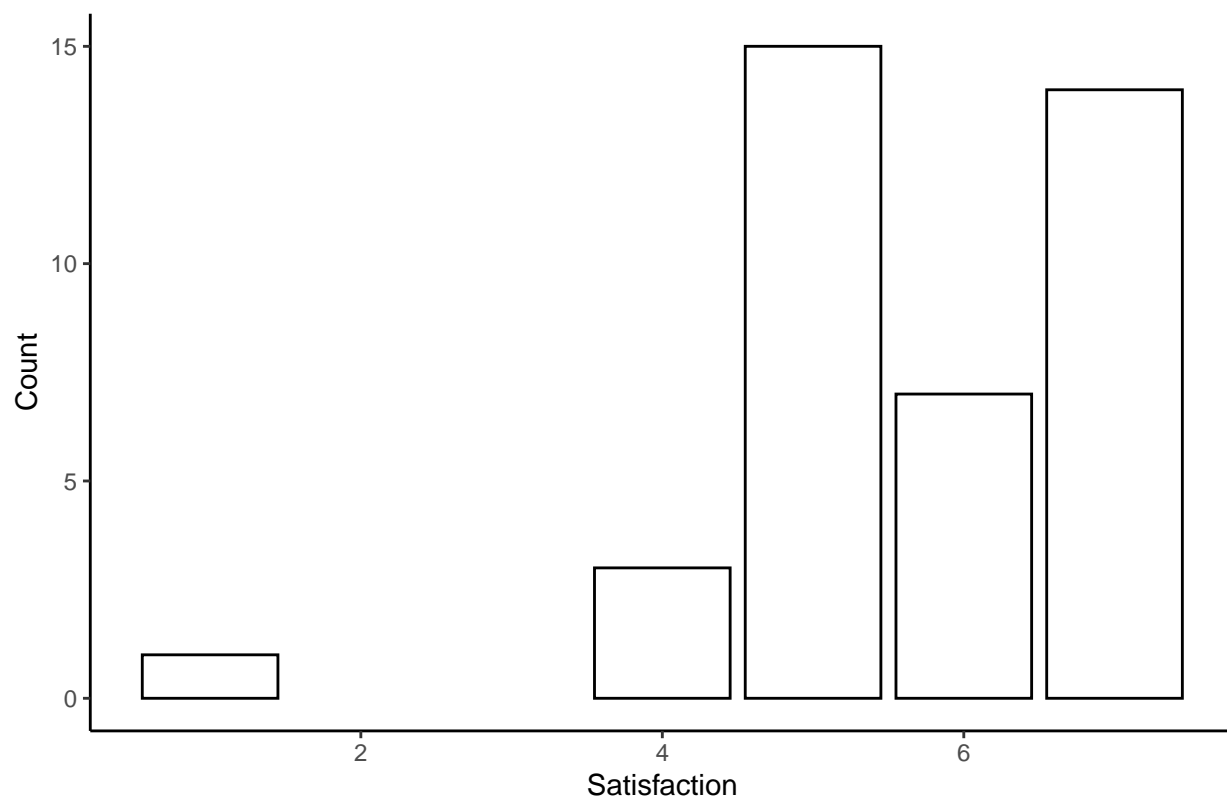


Figure 3: Money Spent Histogram Plot

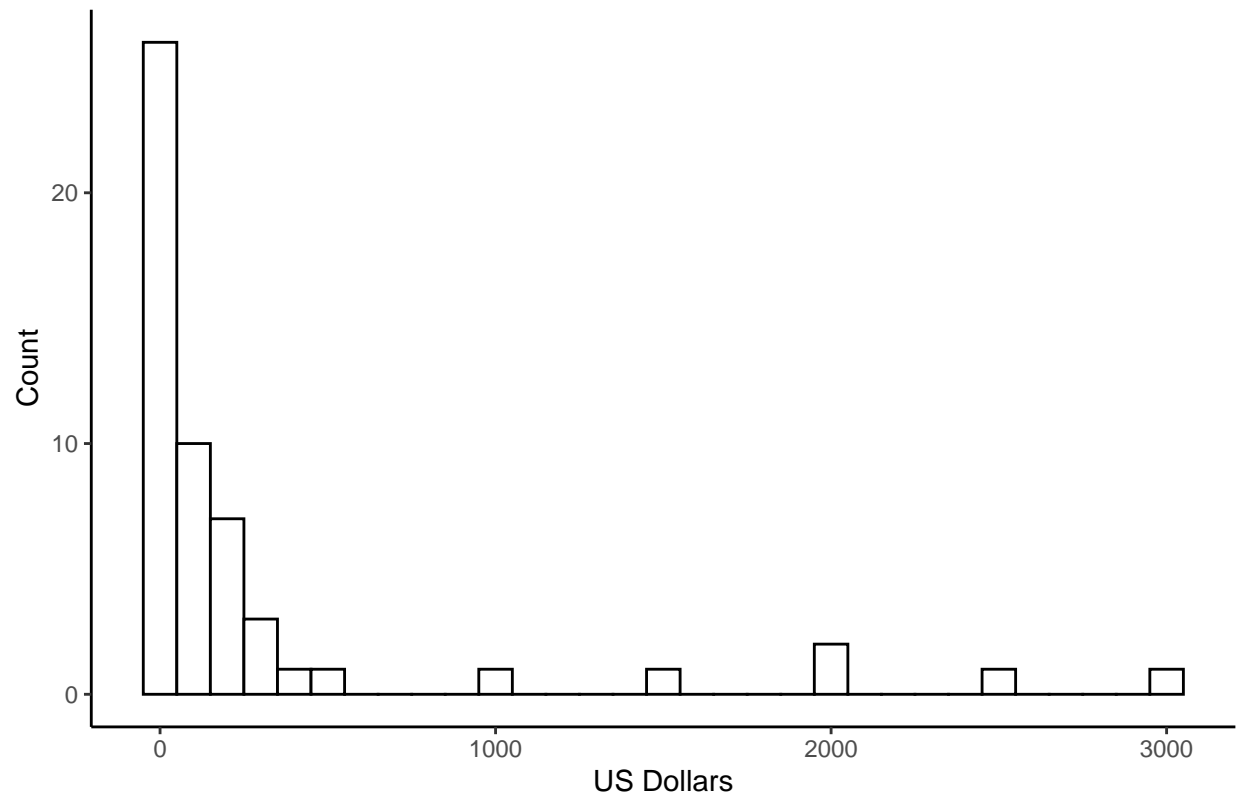
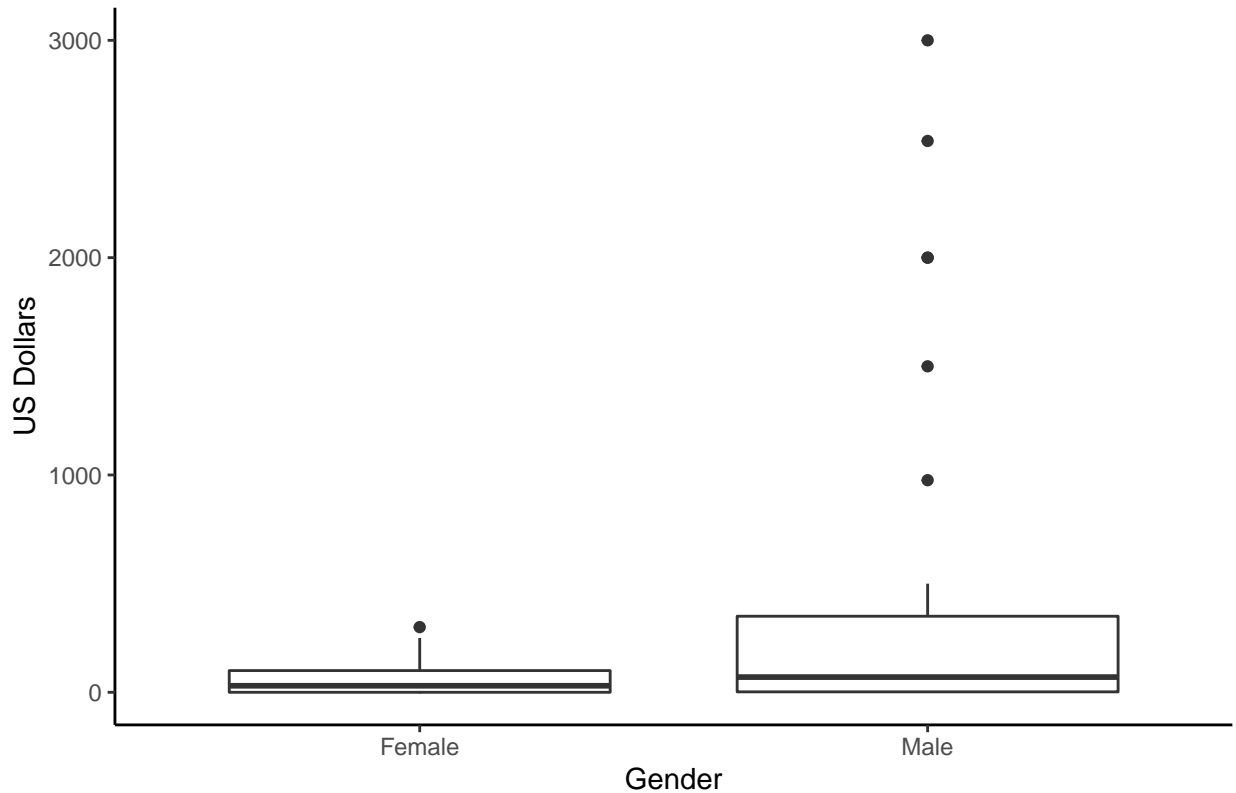


Figure 4: Money Spent Box Plots, by Gender



Looking at *Figure 1*, we see that most players come from the American server, which is what we expected initially. The number of respondents from the America server is greater than the Asia and Europe servers combined. Again, this survey sample is not representative of the entire Genshin Impact community, but how spending behaviour varies between these 3 regions is still something worth looking into.

Figure 2 somewhat goes against the initial belief that people would shy away from choosing the ‘extreme’ option in 1 and 7. While we do see many people responding with 4, 5, and 6, we also see a sizable amount responding with 7. Overall, players seem to be satisfied with their purchases.

Figure 3 and *Figure 4* show that the vast majority of players spend less than \$100, but we observe a large amount of skewness and variance in player spending. This is to be expected, as spending money is not necessary in Genshin Impact, so there is plenty of reason to spend little to no amounts of money on the game. On the other hand, there are players who are willing to spend much more money to get what they desire, which is a demographic of interest. When separated by gender, we observe more spending on average by males, as well as having a larger variance. There are also several outliers to note, particular with male spenders.

All analysis for this report was programmed using R version 4.0.2.

Methods

In this section a hypothesis test and confidence interval will be used to test parameters of interest. To be specific, a Mann Whitney U test will be used to determine whether or not the male and female samples have the same median money spent, and a confidence interval will be calculated through the Wilcoxon signed-rank test to estimate the true population median money spent.

In this hypothesis test, the null hypothesis that will be tested is that the probability is 50% that a randomly drawn member of the male population will be greater than a member of the female population, or in other words, that there is no difference between these two populations. We carry out the test under the assumption

that this is true. The alternative hypothesis is that the populations are in fact different, which is favoured if the test results in the rejection of the null hypothesis.

The test that will be used is the Mann Whitney U test. This test was chosen because the goal is to compare two populations, while avoiding the assumption that the populations are normally distributed when there is evidence suggest that it is not. We do however make the following assumptions required for the test^[5]: 1. All observations from either group are independent of each other. 2. The dependent variable is at least ordinal (any two observations can be compared to see which is greater). 3. Under the null hypothesis, the distributions of both populations are identical. 4. The alternative hypothesis is that the distributions are not identical.

All of these assumptions are justifiable true, as we can easily order the amount of money spent from least to greatest, and the null and alternative hypotheses are constructed as prescribed. It can be argued that one person's spending may affect another person's spending, violating the first assumption, but I find that this is not a very strong argument, and that the first assumption reasonably holds.

For the confidence interval, we can utilize the Wilcoxon signed-rank test to construct a 95% confidence interval by including in the confidence region all those points for which the hypothesis test of the null hypothesis is not rejected at a significance level of 0.05^[6], where the null hypothesis in this case is that median (dubbed pseudomedian) of the distribution from which the samples are drawn is zero. The alternative hypothesis is that the pseudomedian is greater than zero.

We assume that all observations are identically distributed and independent of each other, which we have previously argued is justified.

All analysis for this report was programmed using **R version 4.0.2**.

Results

The Mann Whitney U test results in a test statistic of 210.5, and a p-value of 0.2517115. This means that we do not have evidence against the null hypothesis, and that we fail to reject it. In other words, we interpret this as 'there is not evidence against the idea that the median amount of money spent by males and females is different'.

This result is somewhat surprising, as one may expect that males spend slightly more because Genshin Impact has a history of creating content that is perceived to pander to a male audience.

This result also has several possible implications. One possible implication is that assuming Genshin Impact is profit-driven, they should continue to do as they have been, which is creating content that can be enjoyed by both genders, but favouring that which is geared towards a male audience, which is their primary audience. The alternative option that Genshin Impact has is to experiment with creating more content for the female audience, backed by the fact that they will not be sacrificing too much by doing so, as females still make up a significant fraction of the playerbase, and do not spend differently than males.

The confidence interval results in a lower bound of 127, and an upper bound of infinity. We interpret this as the following: "Were this procedure to be repeated on numerous samples, the proportion of calculated 95% confidence intervals that encompassed the true value of the population parameter would tend toward 95%."^[6] Thus, we are 95% confident that the population median money spent is between 127 and infinity.

This result is very surprising, as the median in the sample was not this large. In addition, from a non-statistic standpoint, one would not expect the average player to spend this much either.

This result suggests that Genshin Impact should continue to do as they are, because a lower bound of \$127 may be interpreted as the average player spending at least \$127, which is an astonishing amount of revenue. If this result was to be taken into account when looking at what items players most frequently spend their money on, Genshin Impact has the option to double down on those options, creating similar items that they know players will be willing to spend on. They may also look at what items players are not spending on, and attempt to improve those so that they are bought in addition to what players are already buying.

Bibliography

1. Golemund, G. (2014, July 16) *Introduction to R Markdown*. RStudio. https://rmarkdown.rstudio.com/articles_intro.html. (Last Accessed: May 5, 2021)
2. Dekking, F. M., et al. (2005) *A Modern Introduction to Probability and Statistics: Understanding why and how*. Springer Science & Business Media.
3. Allaire, J.J., et. el. *References: Introduction to R Markdown*. RStudio. <https://rmarkdown.rstudio.com/docs/>. (Last Accessed: May 5, 2021)
4. *Genshin impact generates \$3.7 billion on mobile in first two years*. Sensor Tower - Market-Leading Digital & Mobile Intelligence. <https://sensortower.com/blog/genshin-impact-mobile-two-years-analysis>. (Last Accessed: October 6, 2022)
5. Conroy, Ronán (2012). *What Hypotheses do “Nonparametric” Two-Group Tests Actually Test?*. Stata Journal. 12 (2): 182–190.
6. Cox, D.R.; Hinkley, D.V. (1974). *Theoretical Statistics*. Chapman & Hall.

Appendix

Here is a glimpse of the data set simulated/surveyed:

```
## Rows: 54
## Columns: 7
## $ Birth.Year    <int> 2001, 2001, 2001, 2001, 2002, 1993, 1993, 1978, 1994, 198~
## $ Gender        <chr> "Male", "Male", "Male", "Male", "Male", "Male", "Male", "~
## $ Server        <chr> "America", "America", "America", "Europe", "Asia", "Ameri~
## $ Playtime      <int> 13, 12, 13, 16, 23, 24, 16, 15, 19, 24, 20, 24, 6, 24, 18~
## $ Money.Spent   <int> 20, 0, 20, 50, 0, 0, 150, 400, 0, 140, 2000, 2000, 250, 1~
## $ Items.Bought  <chr> "Blessing of the Welkin Moon", "I have not spent any mone~
## $ Satisfaction <int> 5, NA, 5, 6, NA, NA, 5, 6, NA, 5, 5, 5, 7, 5, 4, NA, 4, 6~
```