

D214: Data Analytics Graduate Capstone - Assessment 2

Student Information

Student Name: Jackson Goyette

ID Number: 011028154

Date: 2025-01-13

A: Research Question and Hypothesis

Old School RuneScape (OSRS) is a Massively Multiplayer Online Role Playing Game (MMORPG) that is a reboot of RuneScape based on a 2007 backup server of the game. The main gameplay loops of OSRS fall into two categories: Player vs Monster combat encounters (PvM) and levelling non-combat skills (skilling). This is an oversimplification of the game as there are many interconnections between these two playstyles, but these two options fully encompass the gameplay OSRS has to offer. One other thing to note before we delve into the question is that the maximum possible level in any given skill in OSRS is 99, representing the player has achieved the pinnacle of that particular facet of the game.

My question in this analysis is this: Do players who have maxed their combat skills (PvM-ers) have statistically significantly similar total levels to players who have maxed “skilling” skills? In other words, are the game systems of OSRS interlinked to the point where a player focusing on one aspect would need to delve into the other to make substantial progress?

The discussion of interlinked gameplay loops in games (and particularly MMORPGS) is a conversation that has existed from the birth of the genre. One of the main ways OSRS links their gameplay of combat and skilling is by sidetracking you. In OSRS these are referred to as “Distractions and Diversions,” and can include what amount to treasure hunts within the world of the game. Monsters drop “clue scrolls” that have skill requirements to complete, so where you may find yourself killing lizardmen for a slayer task (a PvM encounter), suddenly you’re racing across the world of Gielenor to dig up buried treasure. (Idyl, 2024)

Another important aspect of this question is classifying players by what they do in game. The most holistic grouping I have found is “Bartle’s Taxonomy,” theorized by Richard Bartle back in 1996 when MMORPGs first hit the gaming scene. His discourse on types of players breaks them down into four groups: Achievers, Explorers, Socializers, and Killers. (Bartle, 1996) In the context of our question and analysis, most of these player types fit nicely into the PvM-er and/or skiller basket.

Achievers are all about doing everything they can in the game and optimizing their time played. Since the completionist aspect of OSRS requires both PvM and skilling, Achievers fall into both categories.

Explorers are all about figuring out how the game ticks (that's an OSRS pun), and finding niche interactions within the world. Explorers, like Achievers, tend to fall into both categories simultaneously as gaining access to the entire map requires at least a slight balance of combat and skills. Generally Explorers looking for exploits fall into the PvM-er side of things, and those who are fascinated with the lore and the world are skiller focused.

Socializers are a wild card - they can be part of both groups, just one, or even neither. Those who trend towards the Socializer label value virtual human interaction above all else in their MMORPG. If I had to assign the Socializer a group, I would pin them as skillers as many skills in OSRS have a levelling method that allows for downtime and socialization.

Lastly are the Killers, and I think this is self explanatory. Bartle's original intention for the killer classification is those who impose themselves on other players in player vs player (PvP) combat, but I feel it applies to PvM as well. Killers like to take down threats and prove their worth through combat - a shoe in for the PvM classification here.

Working with my question and knowing these classifications, my hypothesis in regards to my question is: The total levels of skillers and PvM-ers are similar in a statistically significant way. Maxing combat skills has a statistically significant impact on total levels akin to a player maxing their non-combat skills. The null hypothesis in this case would be the opposite: Maxing one's combat skills does not affect total level in a statistically significant way compared to making out skilling skills.

In non-analytic terms, my hypothesis boils down to these interconnected gameplay loops and how they bring in players of all four classifications. Are Killers likely to take a detour to skill and craft themselves superior armor or to go on the clue scroll treasure hunt that dropped from their monster? Will a Socializer level up combat to participate in endgame content with their pals? Are the intricacies of the mechanics and lore of a boss enough to entice an Explorer? Achievers are already spoken for - if there is content, they will do it and they'll do it well!

B: Data Collection and Justification

Jagex, the company behind OSRS, made my data collection very straightforward for me. They have a high score list that they update constantly to house the top two million players. I set out to scrape the data from this site only to find someone had already done it and posted their results. GitHub user "lukearend" uploaded their work to create groupings of OSRS accounts based on total level and account type, and included the raw CSV of the high scores page. I downloaded this 2 GB CSV and got to work!

The biggest advantage of this collection technique is that it comes right from the source. Jagex knows that at its core, OSRS is a game about numbers and stats, and they make sure their high scores update frequently. Scraping directly from the site means the numbers will always be up to date.

The issue is that I did not scrape directly from the site, and I instead found a CSV from a user who had done just that. While there are not usually massive changes in the high scores, a few years from now, this data may be irrelevant as more and more players move up and down the list.

C: Data Extraction and Preparation

Because I was able to find a CSV file with my relevant data, the extraction process was very straightforward.

```
# importing and cleaning our data

osrs_data = pd.read_csv("player-stats.csv")
print(osrs_data.shape)
```

I simply imported it using Pandas' "read_csv" function and checked the shape to ensure it made sense. I found 26 rows: one for username, one for overall rank, one for total level, and one for each of the 23 skills of OSRS.

To clean the data, I started by nixing any empty rows. There shouldn't be any, but we want to make sure.

```
# nix empty rows

osrs_data = osrs_data.dropna(axis = 0, how = "all")
```

The OSRS high scores replace skills lower than a certain value with "N/A," so I had to go through and assign them the lowest value, just as a precaution. Hitpoints starts at level 10, so I did that first so I could just call "fillna" on the rest of the dataframe.

```
# replace "nan" values with minimum level
# hitpoints starts at level 10
osrs_data["hitpoints"] = osrs_data["hitpoints"].fillna(10)
# other skills start at 1
osrs_data = osrs_data.fillna(1)
```

I then printed out the shape again and the columns for future reference. I also exported the full cleaned CSV.

```
# check shape again

print(osrs_data.shape)

# printing summary for reference

print(osrs_data.columns)

# export to csv

osrs_data.to_csv("cleaned_osrs_data.csv")
```

Then I split the data into skillers and PvM-ers. I said that skillers had a 99 in any of the 15 non-combat skills, but no 99s in any of the combat skills.

```
# if a user has a 99 in a "skilling" skill, assign them skiller

skiller = osrs_data[((osrs_data["cooking"] == 99) |
                     (osrs_data["woodcutting"] == 99) |
                     (osrs_data["fletching"] == 99) |
                     (osrs_data["fishing"] == 99) |
                     (osrs_data["firemaking"] == 99) |
                     (osrs_data["crafting"] == 99) |
                     (osrs_data["smithing"] == 99) |
                     (osrs_data["mining"] == 99) |
                     (osrs_data["herblore"] == 99) |
                     (osrs_data["agility"] == 99) |
                     (osrs_data["thieving"] == 99) |
                     (osrs_data["farming"] == 99) |
                     (osrs_data["runecraft"] == 99) |
                     (osrs_data["hunter"] == 99) |
                     (osrs_data["construction"] == 99)) &
                    ~((osrs_data["attack"] == 99) |
                     (osrs_data["defence"] == 99) |
                     (osrs_data["strength"] == 99) |
                     (osrs_data["hitpoints"] == 99) |
                     (osrs_data["ranged"] == 99) |
                     (osrs_data["magic"] == 99)))]
```

Combat skills gain levels almost passively just from playing the game, so many people end up being PvM-ers. To be classified as a PvM-er, the player had to have 99 in any of the combat skills except for prayer (which by design is difficult to train) and slayer (which is technically not a combat skill although it is trained through combat).

```
# if they have 99s in a combat skill, assign them pvmer

pvmer = osrs_data[(osrs_data["attack"] == 99) |
                  (osrs_data["defence"] == 99) |
                  (osrs_data["strength"] == 99) |
                  (osrs_data["hitpoints"] == 99) |
                  (osrs_data["ranged"] == 99) |
                  (osrs_data["magic"] == 99)]
```

Once we had our dataframes of players, I printed out the length and exported them as CSVs.

```
print("Number of skillers: ", len(skiller))
print("Number of pvmers: ", len(pvmer))

skiller.to_csv("cleaned_osrs_skillers.csv")
pvmer.to_csv("cleaned_osrs_PvMers.csv")

Number of skillers: 248703
Number of pvmers: 532595
```

One of the disadvantages I found was with the sheer volume of data I had, my machine took longer than usual to process it all. In talking with other students about this, I found that I should transform the data types from ints to np.ints, which are supposed to take up far less memory. The size issue also caused me to abandon market basket analysis as a previous analysis for this project. I was having issues transforming the database into a workable list of lists to encode, and then with creating the rules table. It was far quicker to work with t-tests where the algorithm scales up linearly as opposed to exponentially with rules tables.

In the same vein, although Pandas and Python are not the quickest when it comes to massive amounts of data, the package and language are very intuitive and easy for me to use. The biggest advantage was that I could use many of Pandas' built-in functions to extract, clean, transform, and export my data. Using bitwise operators to divide players into PvM-er and skiller groups was also made easy with Pandas.

D: Analysis

I used a t-test to compare the means of my two groups: total levels of skillers and total levels of PvM-ers. Because the gameplay loops of OSRS lent itself nicely to two distinct groups, I opted for a t-test because it was designed to compare the means of exactly two groups.

I started by calculating Cohen's D, giving me my expected effect size.

```
# calculate Cohen's d
diff = pvmer.total.mean() - skiller.total.mean()
pooledstdev = ma.sqrt((pvmer.total.std()2 + skiller.total.std()2)/2)
cohend = diff / pooledstdev
print("Cohen's d: ")
print(cohend)

Cohen's d:
0.9269706425705144
```

The 0.93 value seemed good for my hypothesis - the effect size is large and therefore would be easily observable.

Next, I went on to perform the t-test to observe the results.

```
# code used from Luke Hayden's "Performing Experiments in Python" DataCamp course:
# https://app.datacamp.com/learn/courses/experimental-design-in-python

# perform the t test
t_result = stats.ttest_ind(pvmer.total, skiller.total)
print("T test results:")
print(t_result)

T test results:
TtestResult(statistic=381.96033871700257, pvalue=0.0, df=781296.0)
```

With a p-value of under .05, we can determine that there is statistical significance and we can reject the null hypothesis. The means of the total levels of PvM-ers and skillers are statistically significantly similar. This means that OSRS is achieving its goal of intertwining gameplay loops.

The main disadvantage of t-tests is that they are limited to just two groups. If we wanted to go more in depth with this and break down players into all four of Bartle's Taxonomy, we would need to change our method of analysis to ANOVA. We would then also need to figure out ways to categorize players into these four groups.

T-tests' simplicity is a big advantage of them. You simply get your two groups and plug them into the function, and your result tells you if there is statistical significance. It is a great jumping off point to see if there is traction to your hypothesis, or if you should look elsewhere for statistical significance.

E: Data Summary and Implications

The end result of my analysis was finding that there is statistical similarity between the total levels of players considered to be PvM-ers and skillers. This tells us that the game systems are interlinked enough that a player is encouraged to partake in both main aspects of OSRS. A player who focuses on skilling has good reason to dabble in the game's combat, and a player whose preferences are for PvM encounters also have skilling methods they enjoy enough to participate in.

My limitation here is that we cannot tell which exact skills are being focused on by the PvM-ers or which combat scenarios skillers are trending towards. We have determined there is crossover, but the question now is about where that intersection is.

This would be my first course of action for Jagex: find out what skillers enjoy about PvM and vice versa. The community is very engaged in OSRS, so sending out a survey would be effective. Doing this would gain them valuable insights on why people play the game and how they play, allowing Jagex to create content that has massive crossover appeal.

Next, I would recommend further analysis with other techniques. Two that I would like to know the results for would be an ANOVA test and market basket analysis. An ANOVA test would let us measure the variance of levels over more groups, letting the team have a better grasp on how accounts develop and how players move through the game. A market basket analysis of what skills are maxed together would give Jagex insights as to which skills work well together and which might be ready for new content.

Overall, it is clear that there is already a healthy mix of content for folks to do in the game, and that players truly enjoy all that Old School RuneScape has to offer. Jagex has cultivated a gaming environment where players engage with all aspects of the virtual world.

Sources

Idyl. (2024, October 2). ...*why is Runescape so popular?*. YouTube.
<https://www.youtube.com/watch?v=v3T75WvINv4>

Richard Bartle, R. (1996, April). *HEARTS, CLUBS, DIAMONDS, SPADES: PLAYERS WHO SUIT MUDS*. Retrieved January 13, 2025, from <https://mud.co.uk/richard/hcds.htm>

OSRS Hiscores, https://secure.runescape.com/m=hiscore_oldschool/overall

OSRS Hiscore Data Scraping and Grouping Visualization,
<https://github.com/lukearend/osrs-hiscores>, <https://www.osrs-hiscores-explorer.com/>