Dungeons and Dragons and Data

Gabriel Casillas

University of Denver

LIS4235 Scripting for Large Databases

Doctor Organisciak

June 12, 2021

**Dungeons and Dragons and Data**

**Scope**

The scope of the project was to analyze a dataset that looked at the multiple fields in player character creation for the tabletop roleplaying game Dungeons and Dragons (DND). It was to look at some of the correlations between different choices players may make and what that may look like visualized in graphic form.

**Questions**

1. What are the correlations between choice in class, race, and background of player characters?
2. What are the differences in player character choice between a single race, class, and subclass with weapons?
3. What are the most popular types of weapons for a single race and class?

**Data**

The data was structured and collected by B. Ogan Mancarci (oganm) on github in his dnddata repository. Based on some of the documentation Ogan provided in his repository, he received all of his data via upload to his printSheetApp and interactiveSheet sites where users were taking a mobile app character sheet and exporting to pdf.

Other sources of data used were Fandom’s wikidata, DNDbeyond, and Wizards of the Coast’s *Player’s Handbook*. These were used to help determine if content entered into the dataset by users was official DND material, Unearthed Arcana playtest material, or if it was homebrew—unofficial content created by players.

**Data Cleaning**

The original dataset was brought into Colab and cleaned with regex and pandas. The dataset included fields that were not used: ip, finger, hash, date, country, countryCode, processedRace, processedSpells, alias, processedAlignment, good, lawful, choices, processedWeapons, and levelGroup. ‘skills’, ‘spells’, and ‘Class’ were later dropped during cleaning in order to focus on the most important fields of data.

Race, classes, subclass, background, feats, weapons, and alignment were all partially cleaned and normalized across variants of the same type. With aid from DND sources, the data was called by unique values, compared in a separate .csv to DND sources, and normalized within the dataset through string replacement.

**Software and Packages**

The data was worked with in Colab with the programming language Python. The packages regex and pandas were used to clean, explore, analyze, and visualize the data. Matplotlib and seaborn were added to help visualize and chart the data.

**Exploratory Data Analysis**

Through the use of unique, value counts, and basic math functions, the final semi-cleaned data set was explored and analyzed across the major categories within the table.

***Race***

When looking at the player selected race of characters, it is important to note that the data does not define whether the selected race was from a dropdown menu with an option for 'other' or was a blank field for users to generate their own data.

There is a wide variety of instances where homebrew options have been inserted, but mostly the content comes from the wide body of official DND sourcebooks. Some of the player races have even been selected from older editions of DND systems, such as Pathfinder as opposed to the most current player system of DND 5e. Within the various player races, some may have moved from Unearthed Arcana to Official; Unearthed Arcana is beta content that may or may not be published. There were some instances were player races were identified to be in other languages and were merged with the English version, such as 'Meio-Orc'--Portuguese--to 'Half-Orc'.

Out of the data, there are 5949 individual player characters. Some of these players may have created multiple characters within the app. Looking at the index, players created multiple characters. Of those multiple characters, some of them contained multiple data points separated by the delimiter ‘|’; these are considered multiclass player characters. Of the 5949, player characters, there are 199 different races from official DND sourcebooks and adventure modules, Unearthed Arcana playtest materials, and homebrew creation. Humans, Half-Elves, Wood Elves, Half-Orcs, and Dragonborns are the top player character choices.

***Class***

Class is the type of fighter a player character will select. Player characters are not stuck with their choice in class when choosing their characters. At any point--with Dungeon Master approval--a player character could begin to multiclass. Multiclassing is where a player character learns levels of a new class as they level up their characters. Multiclassed player characters have access to the previous number of levels in their starting class, as well as levels they unlock in their new class.

During the cleaning and normalizing of the data, it was necessary to switch analysis from the column 'class' to 'justClass'. 'Class' kept each player character class with its corresponding level in tact between the delimiter ' | '; Rogue 4 | Bard 2. 'classes' removed the corresponding level from each player character class; Rogue | Bard.

Instead of looking at multiclass levels, the only instance of the level of a player character is their overall level recorded in 'level'; 'Rogue 4 | Bard 2' with overall level being 6, became 'Rogue | Bard' with overall level being 6. When calling for instances of a particular class and charting it across levels, it no longer gives an accurate representation of how many player characters are at a particular level for an individual class; it instead charts the overall progress of a player character's level across the field of 'level'.

There are 19 different class types that players had the option to choose from when selecting their player characters. Of the 19, 2 were Unearth Arcana, 4 were homebrew, and 13 were official DND races. Fighter, Rogue, Barbarian, Cleric, and Monk were the most common choices in class.

The 'subclass' field refers to the type of path or variant a player character could choose for a single class. Players selected 148 different subclass variations across the 19 different class types. Battlemaster (Fighter), Champion (Fighter), Assassin (Rogue), and Path of the Totem Warrior (Barbarian) were the most popular choices across the classes.

***Background***

The background category for a player character helps to define extra skill sets, proficiencies, and narrative elements. The main DND sourcebook, the Player's Handbook, offers a body of different backgrounds players can choose from. Some of the backgrounds even come with narrative options in the form of a chart that can be rolled on or selected from. The dataset did not come with any descriptions for the various fields, so some instances within 'background' may have reflected two separate background choices labeled under the same name of 'Pirate', for instance. The amount of background choices players made could be exponentially higher.

There were just as many variations in background types that players selected for their characters: 144. Criminal, Outlander, Soldier, Acolyte, and Charlatan were the most popular selected backgrounds characters chose from; many of these were normalized from variants or narrative options within a background choice—Exile from the Outlander subclass.

***Level***

Player characters ranged from levels 0-261 across 35 different levels. The greatest number of player characters tended to be around levels that have major milestone progressions based on character class types. Level 1 saw 4856, level 3 saw 3508, level 5 saw 3403, level 4 2588, level 8 2009, and level 6 at 1973 to look at the top 6 levels ranked.

Level 1 is where most players begin and thus would see the highest number of player characters. Level 3 is the point at which player characters can pick their subclass type and gain new abilities based on their choices. Level 5 is another point at which players gain new abilities that can affect game play; for example, within the Barbarian class, a character's proficiency bonus goes from +3 to +4 which means when rolling a 20 sided dice, the player will be able to add +4 to their roll in a skill that they are proficient at. Level 4 is where many classes get their first ability score improvement or the option to take a feat; increasing ability scores can allow for players to increase damage dealt with strength based weapons or can increase dexterity in order to evade danger; feats allow the player character to improve a specific area, such as 'Alert' which allows the player character to gain these benefits: +5 Initiative, can't be surprised, no advantage for hidden attackers. Level 8 is the next ability score improvement or feat option. Level 6 offers a variety of improvements across the different classes, so some may benefit more than others. It is important to note that at each level players do receive some sort of ability or feature improvement; however, there are some classes that may not gain any features or abilities but may have the ability to choose more spells and at higher levels.

***Feats***

The feats that were selected by many player characters totaled at 70. DND official sourcebooks and adventure modules provide a large variety to choose from, and only a few of the feats were homebrew or Unearth Arcana content. Of the 70, Sharpshooter, War Caster, Lucky, Dual Wielder, and Alert were the top 5 selected. Sharpshooter is important for characters that make use of long range weapons, such as Gunslingers and Rangers. War Caster allows those with magic to have improved spell casting abilities during combat. Lucky allows players to reroll a dice roll a specific number of times a day, which could potentially change the outcome of an encounter or attack. Dual Wielder gives player characters the advantage of being able to use two weapons during an attack round as opposed to one that many races only have access to. Alert, as described above, puts others at a disadvantage for surprise attacks.

***Health***

HP--Hit Points--is the amount of health a player character has access to based on the amount they roll with each level increase. During combat, players deal and receive damage in turn. The app only recorded the HP attached to each unique player id at the time of compiling the data; the dataset does not have the max HP for each player character recorded. A player character may have a recorded '12 HP' in the app, but in actuality the player character could have a max HP of '42'. The lowest amount of HP a player character had was -24, which implies the character was dead or unconscious; 6894 was the max, which could point at a DM using the app as a way to record the profiles of the boss or BBEG--Big Bad Evil Guy--the characters face at the end of their adventures. On average, players had about 197 HP; this may lean heavily in favor of player characters that have HP above 200 and are at higher levels.

***Armor Class***

AC--Armor Class--is how durable a player character is. The higher the AC, the harder it is to deal damage to that character. Some classes have higher AC's than others due to the roles they take in combat, such as a Barbarian versus a Wizard. There are multiple factors that can affect an AC score such as the type of armor being worn, if the character is using a shield and has it equipped, if the player character has any feats, if the player has any special items or weapons equipped, or their ability scores. The minimum AC score a player character had was 7, whereas the maximum AC score a player character had was 66. Having such a low score means that nearly every attack will hit the player character; the opposite can be said about having a score of 66. With a 20 sided dice, proficiency bonuses, race and class bonuses, feat bonuses, and weapon bonuses, it can make it nearly impossible to roll high enough to actually do any damage to that player character, or potential BBEG. But on average, most players had about an AC of 16.

***Weapons***

Player characters have access to a large body of weapons and spells. Within the dataset, some of the spells were found within 'weapons'. A large portion of the weapons were also of homebrew creation. Cross referencing the individual weapons before cleaning the data with official DND sourcebooks and adventure modules would be important in understanding where a large portion of the weapons are coming from. Making use of digital apps like DNDbeyond can only aid so far since some of their content is behind a paywall; users have to purchase digital sourcebooks in order to gain access to the content of the book to make use of in their digital character sheets. Sometimes weapons may be written in a book but are not included in a weapons table within a sourcebook or are not given a name. Players will create a large body of homebrew weapons items in order to defeat BBEG's that have an AC of 66 because there is only so much that can be done with bonuses and a 20 sided dice. That being said, the unique number of weapons that players had in the app were 519--including any homebrew, spells, other items, and other column information within 'weapons'. Overall, player characters stored Daggers 4446, Light Crossbow 2099, Quarterstaff 1937, Rapier 1900, and Shortsword 1820. Some of these items include special instances of those weapons or if player characters included a number of items they possessed, such as 3 Daggers. Instances like '3 Daggers' was normalized to 'Dagger' in order to better count across the multiclassed characters.

**Conclusion**

Overall, the original data set was under ten thousand rows, but once the data was split and expanded, it turned the dataset into nearly thirty thousand rows. Since the dataset was also only semi-cleaned upon analysis, it made calling for visualizations from Colab nearly impossible. Computing times took upwards of minutes to plot a visualization. Further cleaning would be necessary in order to further look at the questions of correlation between class, race, and background, those correlations with weapon choice, and popularity of weapon choice per race and class. Other tables would need to be created per race type or class type in order to analyze and create visualizations of the data without calling for tens of thousands of instances.

References

Fandom (n.d.). Feats. *Unofficial D&D 5e Wiki*. https://d-n-d5e.fandom.com/wiki/Feats.

Fandom (n.d.). List of dungeons & dragons 5th edition character classes. *D&D Lore Wiki*.

https://dungeonsdragons.fandom.com/wiki/List\_of\_Dungeons\_%26\_Dragons\_5th\_edition\_character\_classes.

Fandom (n.d.). Races. *Unofficial D&D 5e Wiki*. https://d-n-d5e.fandom.com/wiki/Races.

Fandom Games (n.d.). *D&D Beyond*. https://www.dndbeyond.com/.

Mancarci, B.O. (2021). dnddata. *Github.* https://github.com/oganm/dnddata.

Wizards of the Coast (2014). *Player’s Handbook*. Wizards of the Coast.