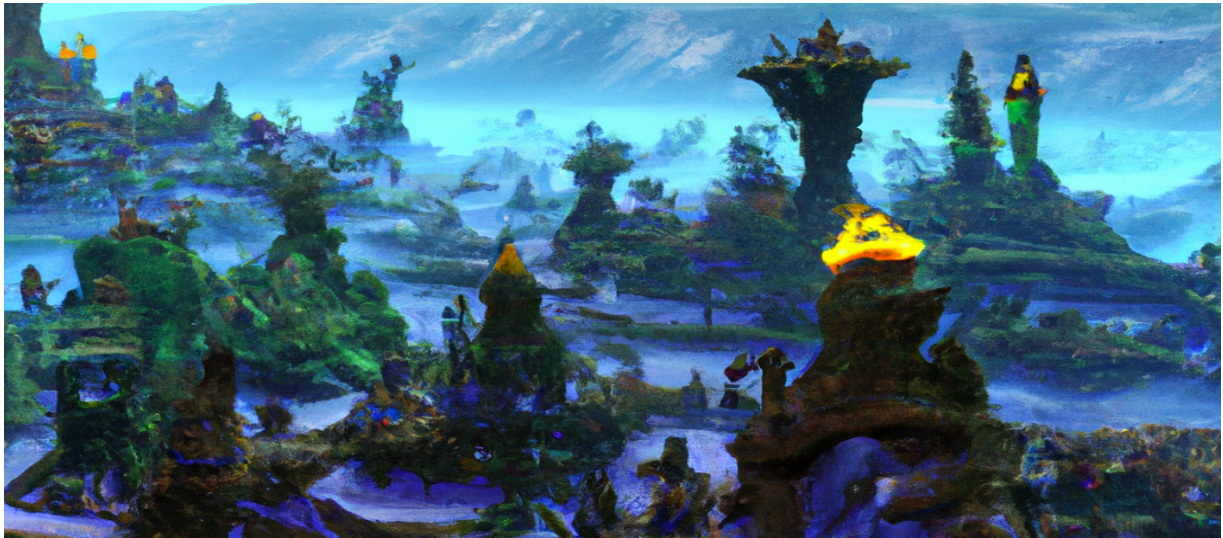


Character Progression: World of Warcraft Data Analysis

World of Warcraft, a massively multiplayer online game developed by Blizzard Entertainment, introduced a rare model upon its release in 2004. Unlike conventional games, which require only an upfront payment, WoW also necessitates a monthly subscription. As well as contrasting traditional games which allow a single protagonist, WoW offered a shared narrative involving all its players.



So I will analyze the dataset and explore how player behavior varies across different levels. Understanding how players' engagement changes throughout when they are online, enable us to identify any specific trends, patterns, or behaviors that are unique to certain levels.

The dataset comprises several features, including unique character identifier (char), current character level (level), selected race (race), character's class (charclass), current location (zone), unique guild identifier (guild), and timestamp of data collection.

The distribution of unique values for each feature is as follows:

Char: 37,354 unique values

Level: 80 unique values

Race: 5 unique values

Charclass: 10 unique values

Zone: 158 unique values

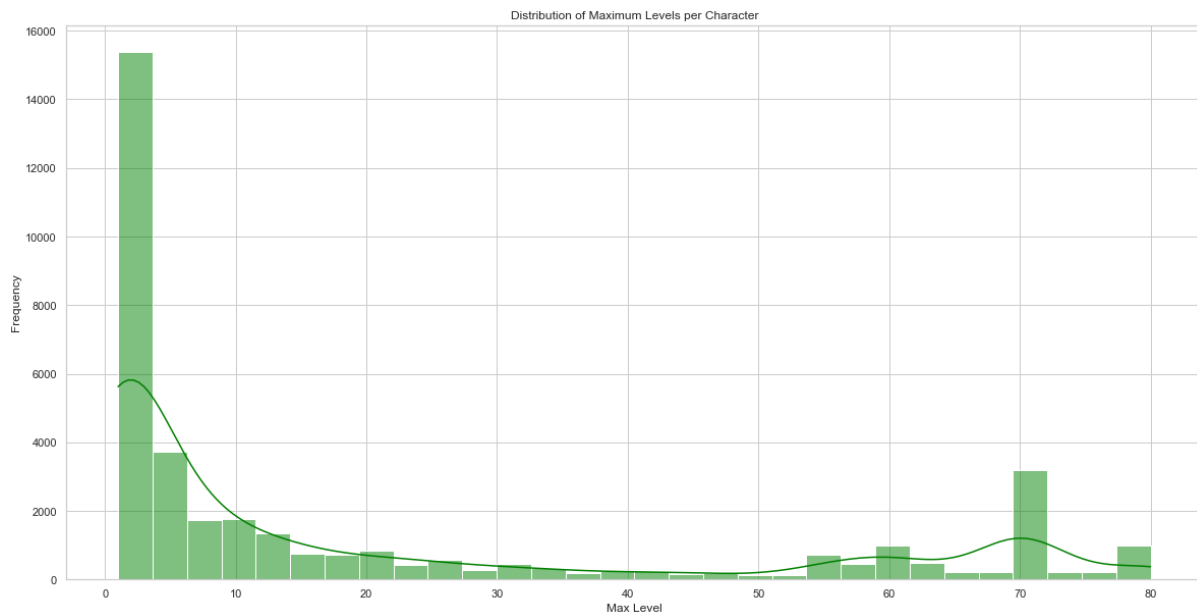
Guild: 420 unique values

Timestamp: 1,826,400 unique values

Exploratory Data Analysis (EDA)

In the Exploratory Data Analysis section, insights into the game's balance and feature exploration for prediction are uncovered.

One notable feature is the character level distribution. The dataset reveals a substantial number of characters at level 1 (11,598 out of 37,354) which can be attributed to players creating alternate characters often used as mules for selling and trading items.

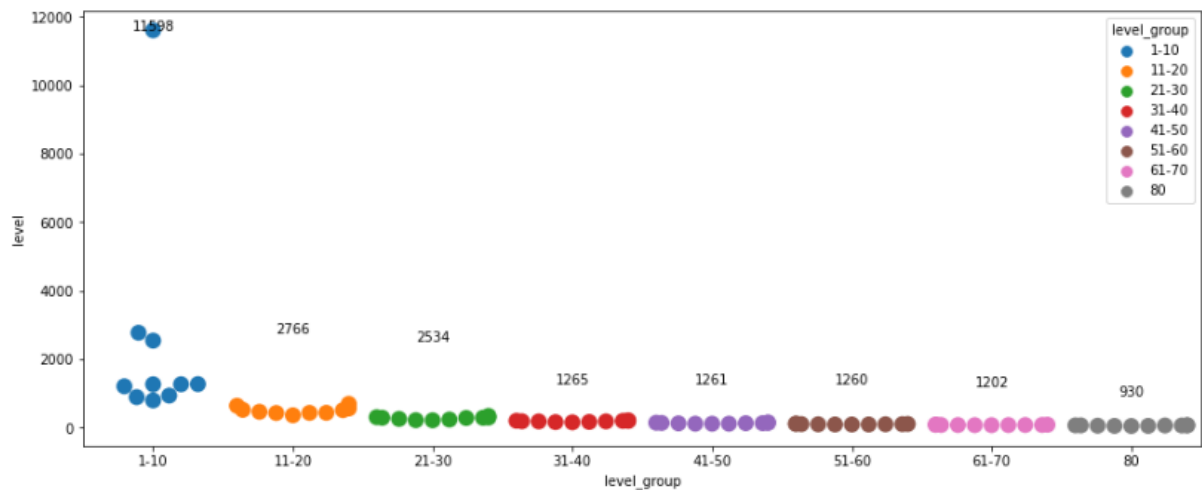


The distribution is right-skewed, large number of players are found within the early levels of the game. There are over 11,598 character entries at level 1, possibly due to players creating alternate characters, which are often used as trading mules in major towns.

Following the peak at level 1, the data presents a steady decline in player count from level 1 through 60. This decline is potentially indicative of the increasing difficulty and time investment required to level up as the game progresses.

Also, there is a wave in the player counts at level 70. It's due to the maximum level limit expansion in 2008. This wave can be a result of a substantial number of players having reached the maximum level by this time.

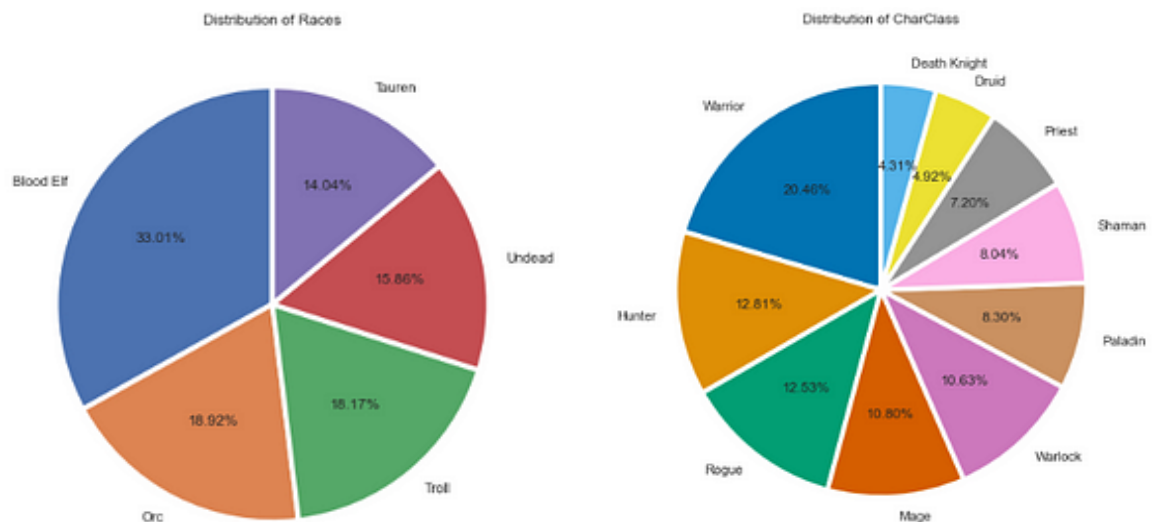
The insights suggest that more resources should be made in the early levels (1–30) where the player base is most concentrated, in order to improve the gameplay experience.



Diving deeper into character-specific distributions, I have analyzed data on character classes, races, and combinations.

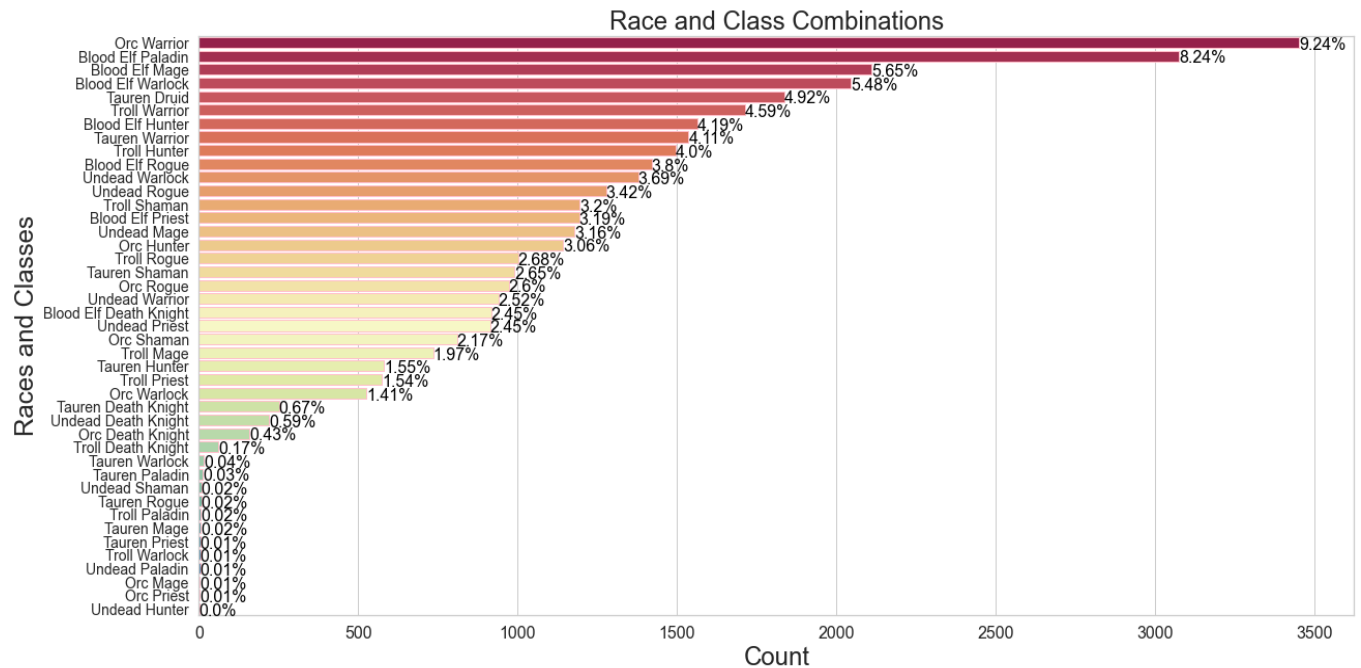
When players create their characters, they get to choose their preferred race and class.

This selection data can provide valuable insights to game developers and balance teams, helping them understand which classes or races may require more attention. For example, if certain classes or races are more popular, it may be because they are perceived as stronger, more enjoyable, or appealing due to their themes. The developers may then decide if additional development time should be allocated to specific starting zones, based on these preferences.



The Blood Elf race appears to be the most popular, accounting for 33.01% of players. The Orc and Troll races are also quite popular, at 18.92% and 18.17% respectively.

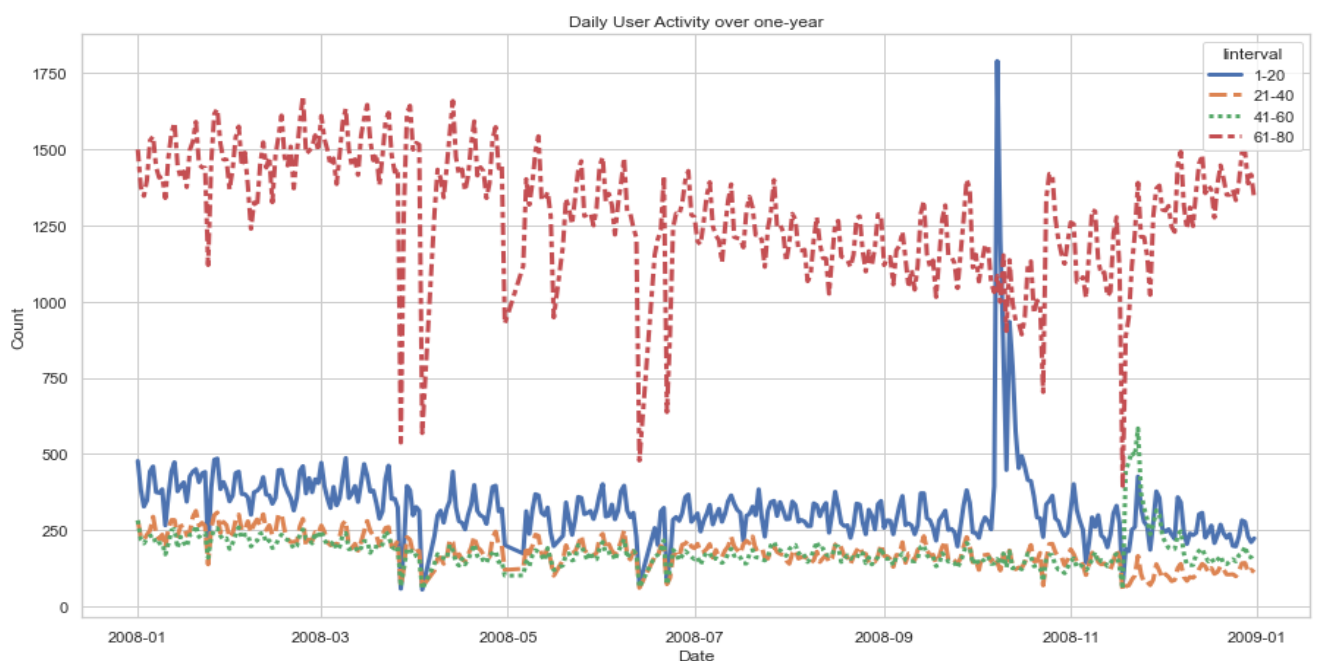
In terms of class selection, Warriors are the most preferred, with a share of 20.46%. Hunters and Rogues are similarly popular, with shares around 12%. On the lower end, the Death Knight class is least preferred at 4.31%.



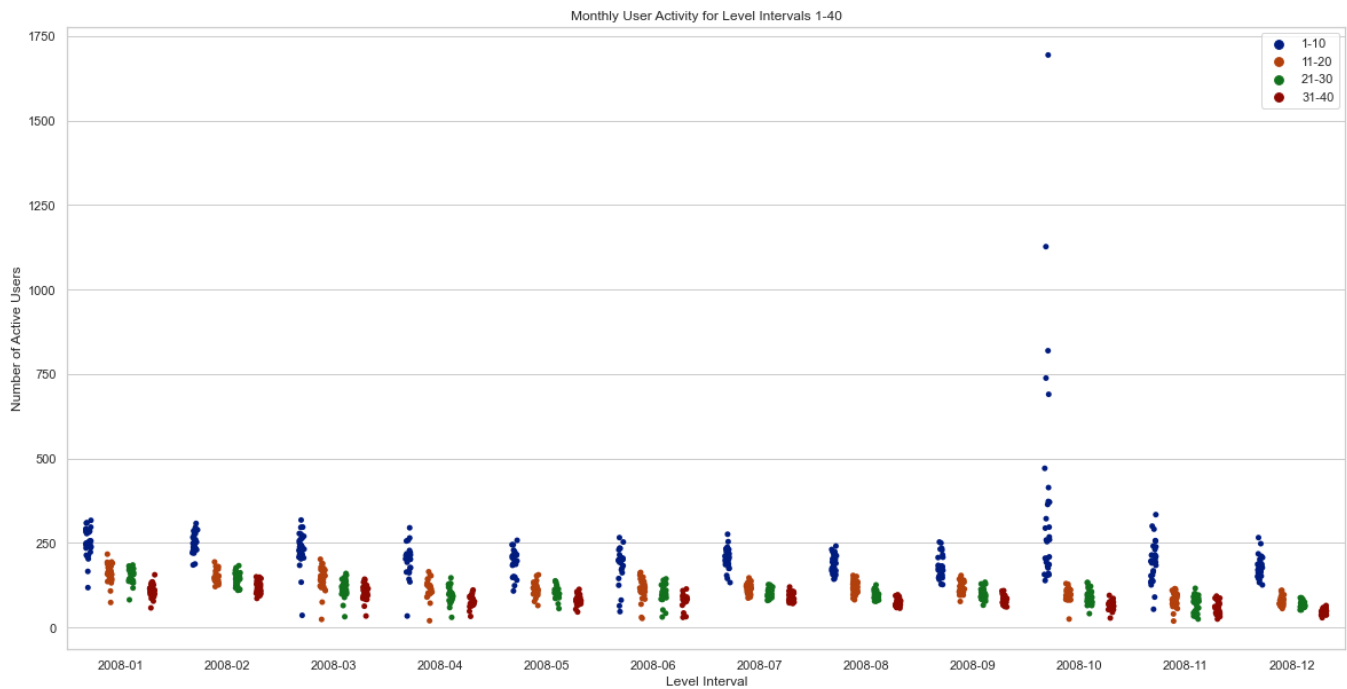
From the analysis, Blood Elves, Warriors, and the combination of Orc Warriors stand out as popular choices among players for races, classes, and race-class combinations, respectively.

Analyzing Player Activity

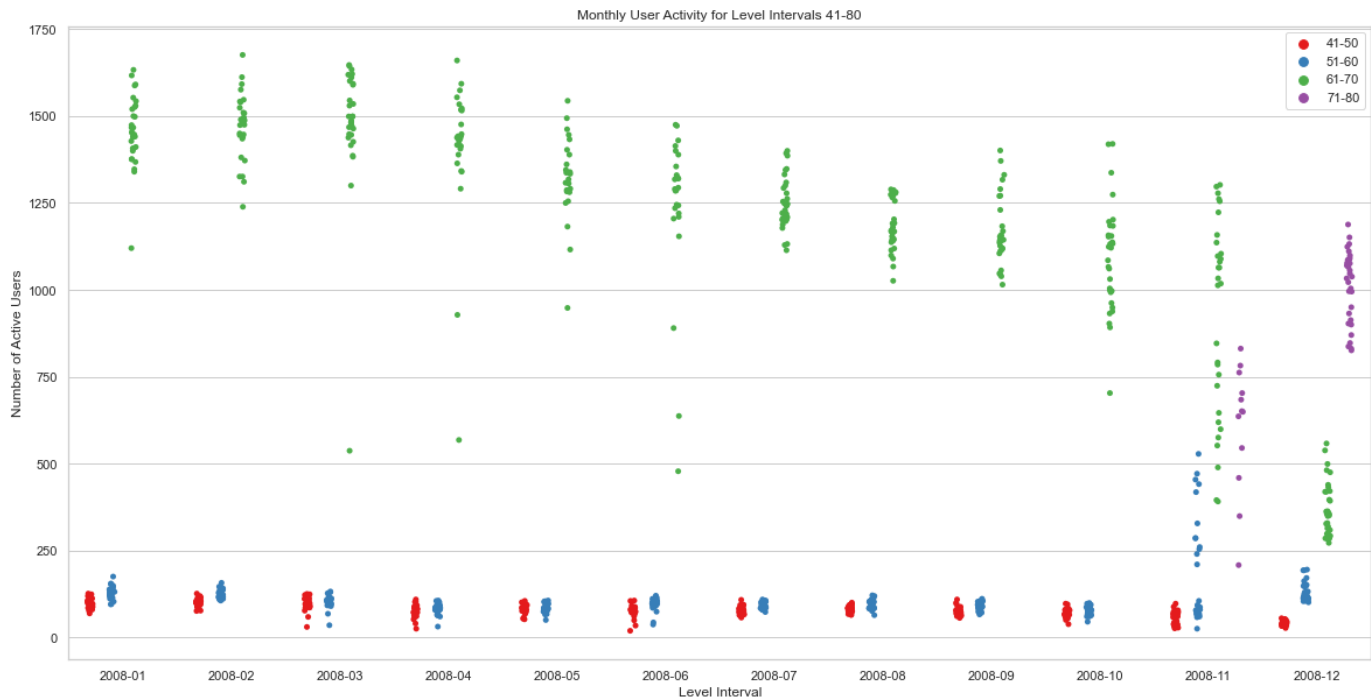
Insight into player activity on hourly, daily and monthly bases, reveals a consistent balance of player concentration in the last level interval.



Similar patterns of fluctuation are observable in the other level intervals from January through September. However, an abrupt surge in the 0–9 level interval is evident in October. This sudden influx could likely be attributed to the arrival of new players in anticipation of the expansion release in November.

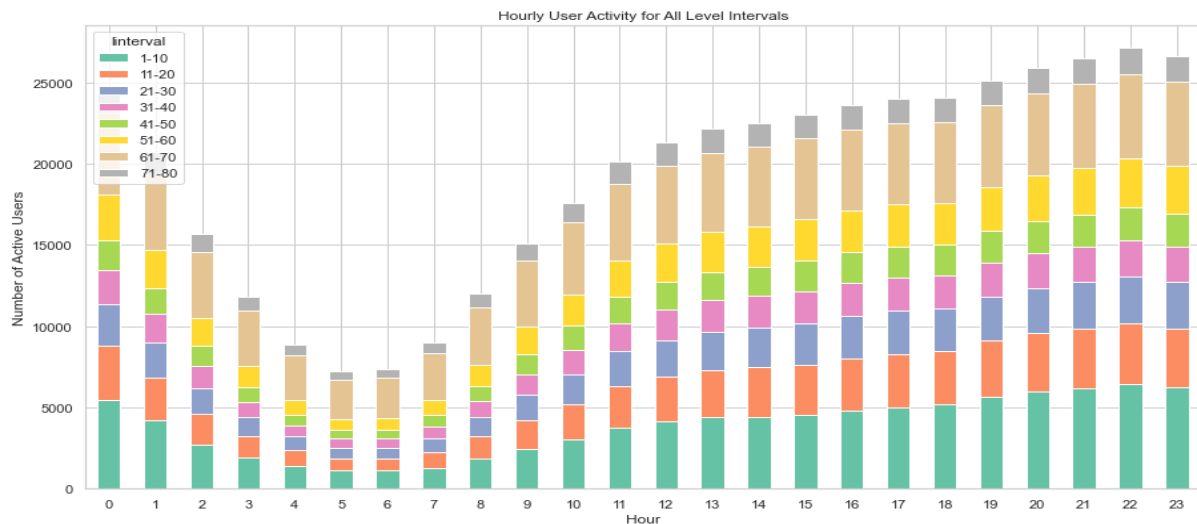


MONTHLY USER ACTIVITY / YEAR FOR 1–40 LEVEL



MONTHLY USER ACTIVITY / YEAR FOR 41–80 LEVEL

The analysis of hourly user activity, daily playtime, and weekday frequency provides further insights into WoW gaming patterns. Notably, there is a consistent decrease in activity until 7:00 am, followed by a significant increase in the afternoon, likely due to younger players returning from school. Activity continues to climb from 5:00 pm until midnight, a pattern attributable to adult gamers joining in post-work. Consequently, the highest user frequency is observed during evening hours.



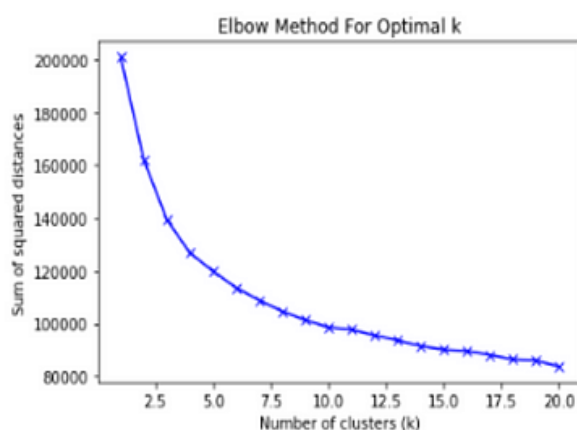
HOURLY USER ACTIVITY / YEAR

The average WoW player is found to engage with the game for about an hour daily, while the top 5% of users play for up to four hours daily.

CLUSTERING

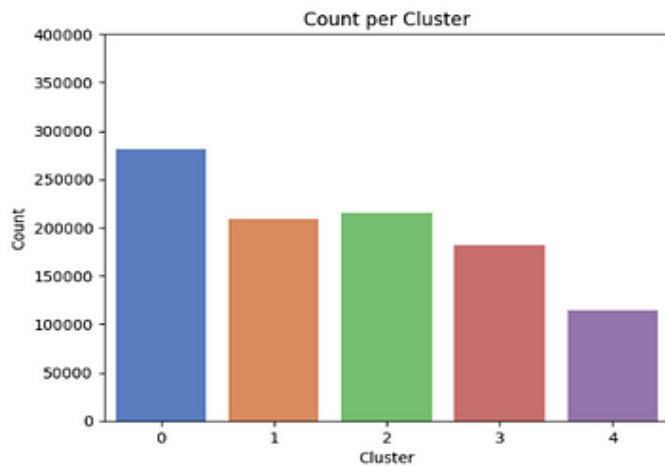
Although I have already segmented the dataset by the “level” column, there are still multiple dimensions not taken into account. For instance, the character’s race, class, zone, and guild, as well as playtime patterns could all potentially add more context. The k-means algorithm can group the data into clusters based on all of these features.

Choosing the Right Number of Clusters



For finding an appropriate K value, I have applied the Elbow Method. This involves plotting the explained variation as a function of the number of clusters and picking the elbow of the curve as the number of clusters to use.

For this case $K == 5$



After applying K-Means, I conducted a cluster analysis to understand the profiles/characteristics of each group. This involves examining the distribution of variables within each cluster and comparing the properties of different clusters.

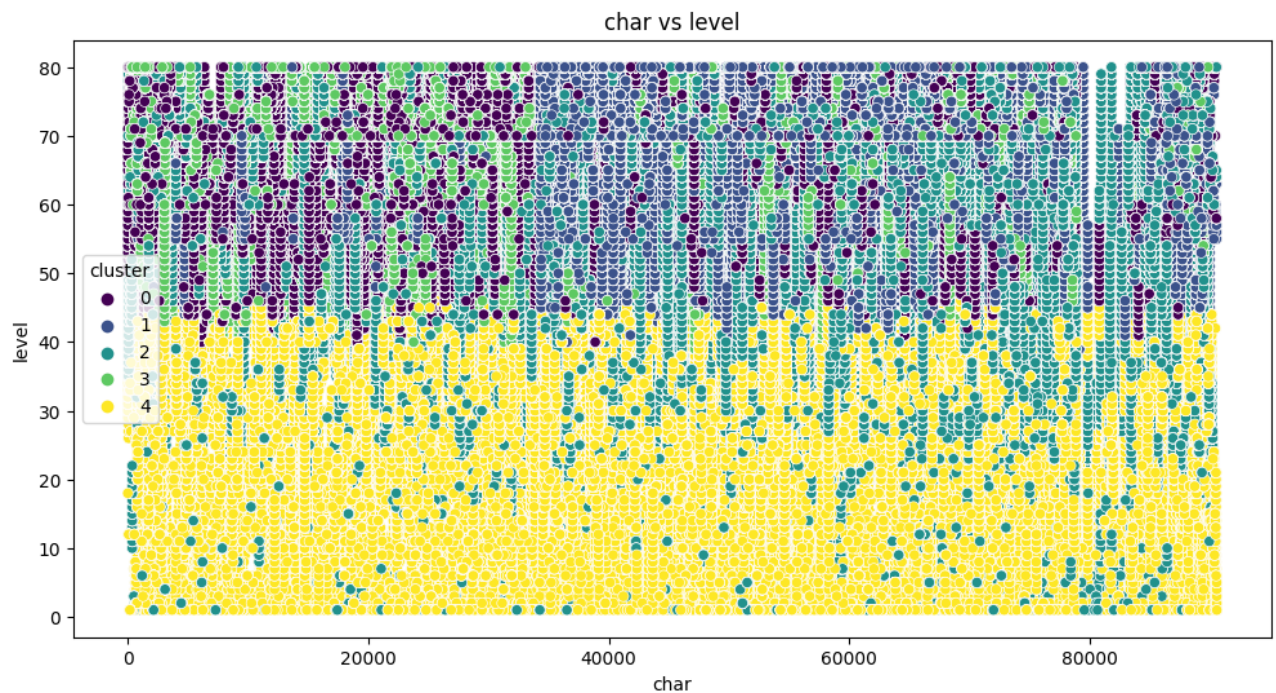
As a start, char counts per cluster ←

In order to understand clusters' behavioral patterns, the distribution of player levels gives a degree of insights.

Clusters 0, 1, and 3 predominantly include players at levels 60–79, indicating they've reached or are near the finish.

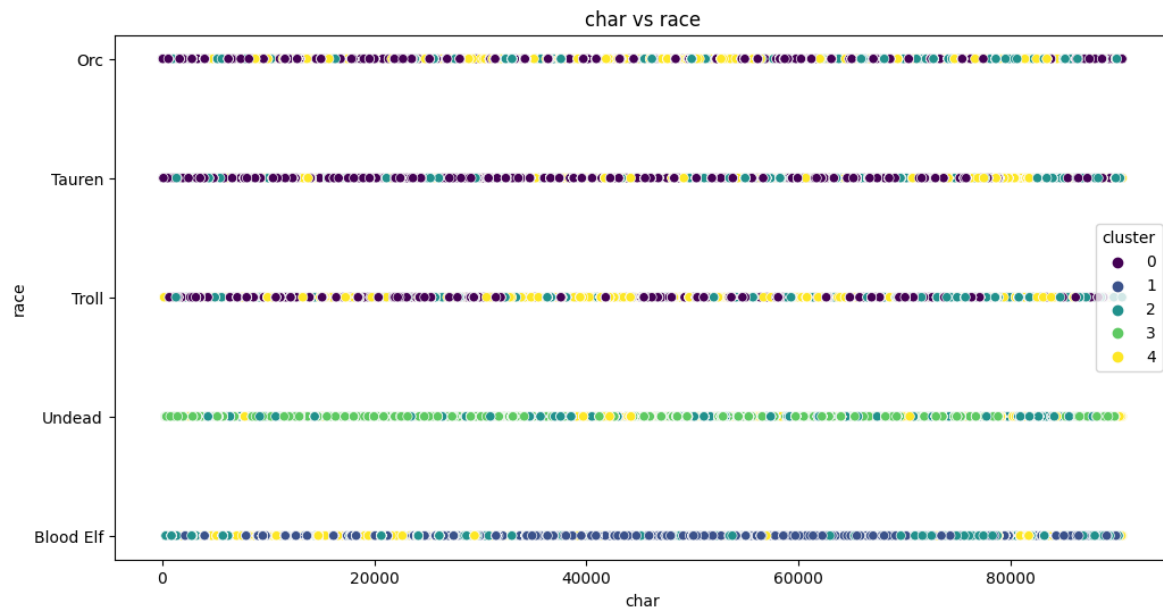
Cluster 2 is more balanced, with significant mid-level (30–59) and endgame (70–79) activity, suggesting actively progressing players.

Cluster 3 also displays some players within the level ranges 30–59, indicative of progression. Unique to the group, cluster 4 consists mostly of players at lower levels (0–39), implying newer or more casual players.



Clusters 0, 1, and 3 represent players near or at endgame, cluster 2 hosts active, progressing players, and cluster 4 accommodates newer or casual players.

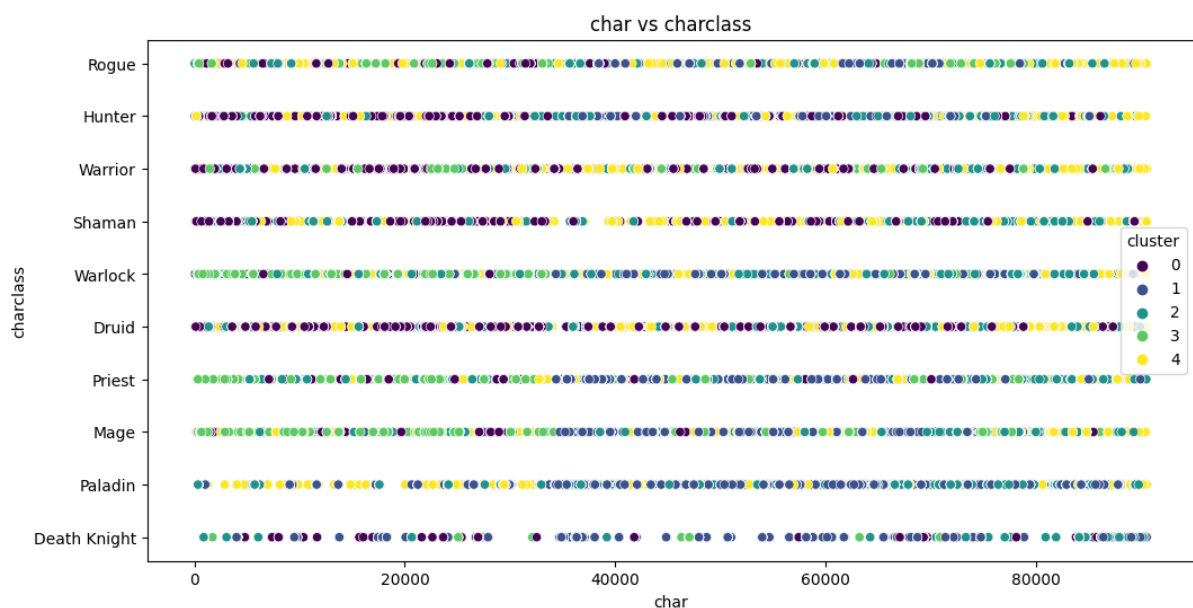
The most common race differs from cluster to cluster. In Cluster 0, the most common race is Tauren, in Cluster 1 and 2, it's Blood Elf, and in Cluster 3, it's Undead.



The choice of the class has a significant impact on the gameplay experience in WoW.

Cluster 0 has a majority of Warriors, who are often the tank role in group content, absorbing damage and protecting other, less durable characters. On the other hand, cluster 1 has mostly Paladins, a versatile class that can fulfill all three roles: tank, healer, and damage dealer, indicating a potentially versatile group of players.

In contrast, cluster 4 has a majority of Hunters, a class often considered beginner-friendly and good for solo gameplay, which we have observed the alignment in the previous plot with the lower levels observed in this cluster.



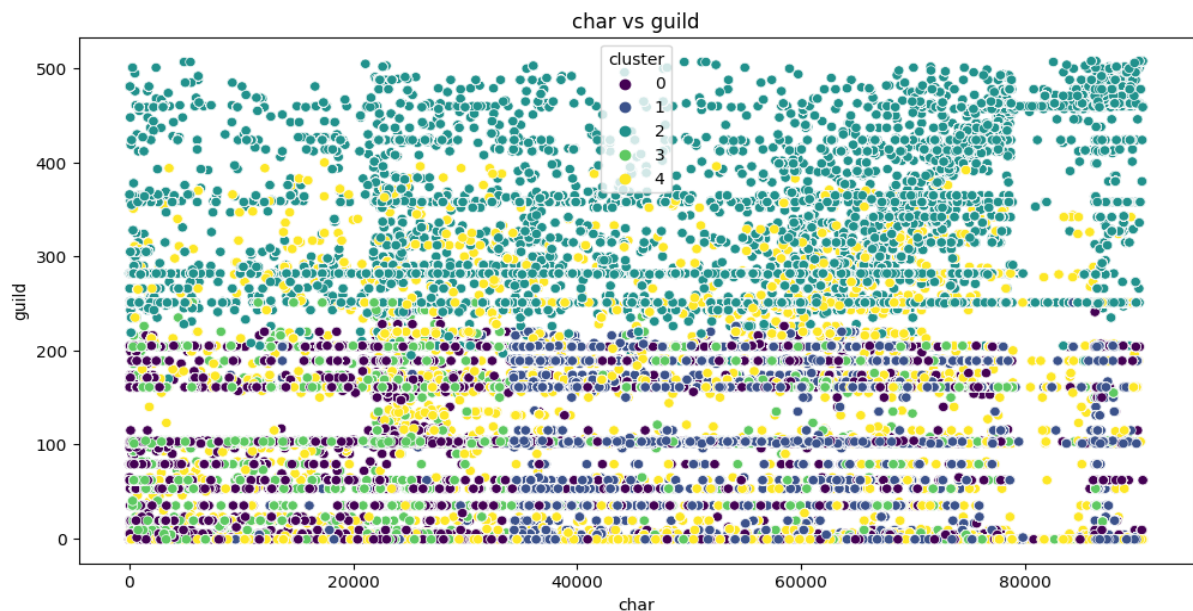
Cluster 0: Players are mainly in guilds with 99–148 members, followed by 49–98 and 149–198, and a small number in smaller guilds (-1 to 48). This suggests a preference for the community without overly populated guilds.

Cluster 1: Similar to cluster 0, cluster 1 has a presence in larger guilds (199–298 members) indicating a preference for smaller to mid-sized guilds.

Cluster 2: Uniquely, this cluster leans towards large guilds, mainly those with 249–298 members, and a substantial portion in 299–398 guilds, suggesting a preference for larger, populated guilds.

Cluster 3: Resembling clusters 0 and 1, most players are in guilds of 99–148 members, followed by smaller (-1 to 48) and mid-sized (149–198) guilds, with few in larger guilds (199–298).

Cluster 4 stands out with a majority belonging to smaller guilds (up to 48 members), albeit with some representation in all sizes.



In summary, players in clusters 0, 1, and 3 lean towards smaller to medium-sized guilds, cluster 2 favors larger guilds, while cluster 4 shows a distinct preference for smaller guilds but with a varied distribution.

Ultimately, I have discovered that players are concentrated in early levels, Blood Elves and Warriors are popular choices, and there's a consistent balance of player concentration in the last level interval. This can inform us, such as improving early-level gameplay and balancing character choices.

Furthermore, through k-means clustering, I have segmented players based on their gameplay aspects like race, class, level, and guild size. This can aid in designing tailored engagement strategies and content to improve player experience. For instance, features can be developed for smaller guilds to promote community engagement and large-scale cooperative content can be created for larger guilds. The data-driven insights from this study are invaluable in shaping the game's future development and improving the gameplay experience.