Weekly Seasonal Player Population Patterns in Online Games: A Time Series Clustering Approach

Dulakshi Vihanga, Michael Barlow, Erandi Lakshika, Kathryn Kasmarik
School of Engineering and IT
University of New South Wales
Canberra, Australia
d.wannigamage@student.unsw.edu.au, {m.barlow,e.henekankanamge,k.merrick}@adfa.edu.au

***Abstract*—With the continuous technological advancement in the game industry, millions of players engage in various online games everyday. Player population size of games ebb and flow through time as a complex series. Analyzing these player population numbers in a shorter time window, such as weekly, could help generate insights that enrich the understanding about low-level population fluctuation patterns of online games. However, this area of game data analytics still has space for further enhancement. This study focuses on discovering patterns of weekly player population fluctuations, that could aid in comprehending how the population of various kind of games change within a framing window of a week. We use player population time series of 1963 games available on Steam. Utilizing several trend removal techniques and conducting seasonality detection we identify that 77% of games display a recurring weekly pattern in player population fluctuations. Moreover, our dynamic time warping based cluster analysis shows that there are 9 diverse weekly player population fluctuation patterns. Among these 9, the governing pattern visible in the majority of games displays that the player population is higher towards the weekend. Finally, we scrutinize the tags, age requirements and overall population size of games in each cluster associated with the diverse patterns to generate insights about the characteristics of games associated with each weekly population pattern.**

***Index Terms*—Game data analytics, Steam, Weekly Patterns, Time Series Clustering**

I. Introduction

The game industry, being a prominent source of entertainment, has been rapidly growing over the years to become a multi-billion-dollar industry. In fact, the number of active players in the world in the year 2018 is more than 2.3 billion [1]. Achieving such growth is not an easy task, but can be assisted if player behaviour is properly monitored and games are reformed accordingly to supplement the innovations in the game industry. Thus, game data analytics has emerged and become prominent within recent years among game developers and game researchers alike [2].

Most of the game related studies that have been conducted so far are dependent on individual player behavioural telemetry. These studies help developers to understand players' expectations from games [3], player retention strategies [4], player churn rationale [5] and even common playing patterns among players [6]. Some are focused on recognizing motivations for game play [7] while some attempts to analyse play time and corresponding player demographics [8] to better understand players. Apart from these aspects about players, further understanding about games can be achieved by analysing the player population dynamics with respect to time. Specifically, insights about short-term player population fluctuation patterns of games, such as weekly patterns, could help strengthen the game industry in various means. However, currently there is little knowledge on weekly player population fluctuation of games.

The main aim of this paper is to explore player population fluctuations within online games using a framing window of one week to identify the existence of weekly seasonality, archetypal weekly population patterns and the relative frequency of these patterns. Additionally, we explore the common characteristics of games that display these diverse patterns in order to enhance the comprehension of these patterns. For this purpose, we gathered extensive player population data from 1963 online games from *Steam*[1], the popular game distribution platform, for a period of 6 months. First and foremost, to address the question of, "do games exhibit weekly seasonality in player population fluctuation", we adopt an autocorrelation based seasonality detection technique. Next, the games that exhibit weekly seasonality are further explored through a dynamic time warping based cluster analysis process in order to discover diverse patterns of weekly player population fluctuation. Finally, we analyse the characteristics of the games in each cluster such as tags, age requirement and overall population size.

Our key contributions and findings in this paper are fourfold,

1) We analyse a significant time-series dataset of some 1963 games across a period of play of 6 months sampled at 5 minutes or 1 hour.
2) We apply several trend removal techniques for weekly seasonality detection in player population time series, thereby revealing that 77% of games display weekly seasonality in player population fluctuations.
3) We conduct a cluster analysis on the population time series of games and reveal 9 diverse patterns of weekly player population dynamics.
4) We investigate the tags, age requirements and overall population size of games in each cluster and generate

[1]https://store.steampowered.com/

978-1-7281-1884-0/19/$31.00 ©2019 IEEE

further insights about the characteristics of games associated with each distinct weekly population pattern.

The outcomes and methods used in the study would be beneficial to various parties. For instance, the insights could aid in scheduling weekly events for maximum player reach, resource allocation, in assessing the anticipated population of games beforehand, for business intelligence and in several other means. We believe that this study sheds light on the diversity of short term player population fluctuation patterns of online games. The remainder of the paper is organized as follows. We first discuss the related work. Then we provide specifics of the data collection process. Next, we present the details of weekly seasonality detection followed by weekly population pattern discovery. We then present the results and discussion and finally provide the conclusion and future work.

II. Related Work

Understanding game players has been an integral part of the game development process. Moreover, research conducted to examine game player behavior have diverse motives. Discovering player behavioural patterns is one such prominent motive. Menéndez, Vindel and Camacho [9] have identified different player behavioural profiles based on the player behaviour and its evolution. Using time series clustering techniques, they have identified three gamer profiles. Similarly, using time series clustering, Saas, Guitart and Periáñez [10] have attempted to discover diverse patterns of playing activity during game events and purchase behaviour of players. Further specializing, the study of Baumann et al. [11] identifies 6 groups of hardcore gamer profiles based on their playtime and characteristics of the games played. Apart from these examples, there are various other studies conducted to reveal diverse player behavioural patterns using game telemetry [12] [13].

While behavioural profiling is a driving force in the arena of gaming, it is not the sole approach game behavioural telemetry can be used. It could also be utilized to derive various insights about games. For instance, Sifa, Bauckhage and Drachen [14] have conducted a large scale analysis using player behaviour data and have identified 4 archtypes of games, each representing a different playtime frequency distribution pattern. Moreover, analysing the playtime dataset of 5 games, Bauckhage et al. [15] have revealed that higher player activity occurs towards weekend in two single player games. Specially in one game, it's higher on Saturday. However, they claim that in general, average playing activity of players seems to be similarly distributed during all days of the week among all 5 games they have studied. Apart from playtime data of individual players as used by above studies, player population numbers can also aid in distilling insights about games with respect to player behaviour. This, however, has not been well addressed by prior studies. A main exception is the study of Chambers et al. [16] focused on game players and server workloads. Analysing the player population of different sets of 550 online games, they state that player population display daily and weekly periodicity. However, it does not further investigate on the types of weekly patterns those games display or common game properties. Moreover, *Steamcharts* [2], an online visual tool, displays the player population variation through time of numerous games in Steam. However, the analyses offered in it does not include insights about common player population variation patterns. Studying how player population numbers ebb and flow with time could serve as a means of comprehending player behavioural patterns of diverse games. Thus, in this study we focus on uncovering weekly player population patterns among games, which, to the best of our knowledge, has not been studied extensively.

III. Data Collection and Pre-processing

Steam is used as the sole data source for this study. Steam is the largest digital distribution platform for video games which currently has around 125 million users and more than 20,000 games of various genres [17]. We collected population data of 2000 applications in Steam from 14th December 2017 to 13th June 2018, which covers a time stretch of 6 months. For our study, it is vital that we select games which have a strong player base as indicated by a high average number of players in order to determine the existence of seasonality. Thus, we used *SteamDB*[3], a third party application that provides stats about Steam games in order to select the 2000 applications with the highest player population within last 24 hours on 11th December 2017. After excluding all non-game applications, the dataset reduced to 1963 games. Moreover, current player population size was recorded in 5 minutes interval for 982 games (first half of the selected list of games) and in 1 hour interval for the rest of the games due to storage and performance limitations. The Steam API was used to record the number of current players in each game.

To further emphasize the underlying patterns in population timeseries, Median Filtering data smoothing procedure was applied over the population dataset. *Median Filtering* replaces a data point with the median of its neighboring data points [18].

In the rest of the article, the smoothed data will be addressed as follows when separate addressing to the dataset is required.

5mData : Player population data of 982 games collected at 5 minutes intervals

60mData: Player population data of 981 games collected at 60 minutes (1 hour) intervals

Additionally, for each game, we extracted the 20 most applied tags in Steam along with the number of users who have applied each tag, on 1st of October 2018 using the *SteamSpy*[4] API. Moreover, we manually collected the age restrictions of a selected set of games applied by the Australian Classification Board[5] based on the game content.

IV. Weekly Seasonality Detection

Before exploring diverse weekly population patterns, we need to identify what games, if any, display regular player

[2]https://steamcharts.com/
[3]https://steamdb.info/graph/
[4]https://steamspy.com
[5]http://www.classification.gov.au/Pages/Home.aspx

population patterns that recur every week. In other words, we need to identify games that exhibit weekly seasonality. For this study, we use autocorrelation to detect weekly seasonality as our data is collected in time domain in short intervals. *Autocorrelation* is defined as the correlation of a variable with a lagged version of itself [19]. If there is a seasonality, the autocorrelation value at the seasonal lag would be positive and high [20] compared to the neighboring lags.

The first step of recognizing games that display weekly seasonality is the calculation of autocorrelation values for each game for lags up until 2030 and 180 for *5mData* and *60mData* respectively. These boundaries were chosen as the lag that represents a week, which is lag 2016 and lag 168 of *5mData* and *60mData*, appear within the selected lag range. As the next step, the lag corresponding to the maximum autocorrelation value has to be identified for each game. However, lags 1-432 and 1-36 of *5mData* and *60mData*, which represent lags upto one and half day, are excluded in this step as these lags have higher autocorrelation values than that of the week's lags since data points that are closer together have a high autocorrelation. Thus, for the proper detection of weekly seasonality of games those lags are excluded in maximum autocorrelation identification. The final step is the recognition of games that exhibit weekly seasonality. Ideally, if a game exhibit weekly seasonality, its lag corresponding to maximum autocorrelation would be the lag number representing a week. However, it is not always exactly the case due to distortions in the dataset. As games attract players from all over the world, time zone differences could result in some distortions. Thus, a game is labelled as a game displaying weekly seasonality not only if its maximum autocorrelation occurs at exactly a week's lag, but also if it occurs in a lag within a certain distance to the week's lag. In order to determine the most appropriate distance, a number of games identified with increasing lag ranges were recorded. The number of identified games saturated and didn't increase when the lag range $2016 - 7 \leq l \leq 2016 + 7$ was used for *5mData* and $168 - 3 \leq l \leq 168 + 3$ for *60mData*, where l represent the maximum autocorrelation lag number of the game in focus. Thus, these lag ranges were used for the recognition of games with weekly seasonality.

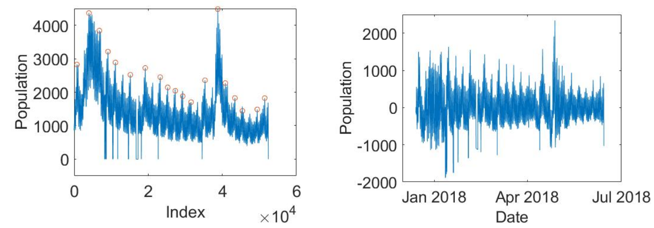
A. *Trend Removal*

Time series can display continuously increasing or decreasing values, which is recognized as a trend. Trend of each series were removed prior to seasonality detection for accuracy. Trend is removed by subtracting the value calculated by the fitted trend functions provided below from the original value, at a given point in the population time series.

Linear Trend Removal

Polynomial Trend Removal: An order 8 polynomial function is used as mini trends of some series are not captured by linear functions.

Piece-wise Linear Trend Removal: In order to remove the smaller trends caused by irregular player population fluctuations we devised a piece-wise trend removal process. Pieces were determined by identifying the globally maximum peaks in the data using *findPeaks*[6] implementation in Matlab within a windows size of 7 days. Size 7 was chosen so as the window size is not too small that it will loose the patterns inside a week and not too large that it will overlook the low level trends. Then linear trend of each piece was removed. For instance, see Fig. 1.



(a) Boundary peaks of pieces (b) Trend removed time series

Fig. 1: Before and After Piece-wise Linear Trend removal of population series of the game 'theHunter: Call of the Wild'

B. *Weekly Seasonality Detection Results*

We recorded the number of games, out of 1963, that display weekly seasonality by applying autocorrelation based seasonality detection on raw data and trend removed data of the 6 months population timeseries. The results are depicted in Table I.

TABLE I: No. of Games displaying Weekly Seasonality under different trend removal techniques

NoTrendRem	Linear	Polynomial	Piecewise
649	668	861	1508

As indicated in Table I, each improvement to the trend removal process has resulted in identifying more games with weekly patterns. If we compare the results of *NoTrendRem*, which uses the data as is without any trend removal process, and *Linear*, which removes a single linear trend in data, we can observe that the increment in *Linear* is not very significant, which is only 0.96%. On the other hand, *Polynomial* shows a 10% increase in the results compared to *NoTrendRem* indicating that more games in the dataset display mini trends in population that can be represented by a high order polynomial. However, most games, in fact, 77% of games have been identified by *Piecewise* which used a piecewise linear trend removal technique. Apart from indicating that most games display a weekly pattern, *Piecewise* also implies the highly irregular nature of player population fluctuation in games. Due to various reasons such as competitions, discounts and updates the player populations in games varies extensively throughout time. Thus, it is harder for *Linear* or *Polynomial* to remove the trend.

We discovered that 77% of games exhibit weekly seasonality. However, the results only implies that games display weekly patterns in how player population fluctuate, but it

[6] https://au.mathworks.com/help/signal/ref/findpeaks.html

does not indicate if all games have the same weekly pattern or diverse patterns. Thus, it is of interest to explore what different/similar weekly patterns games exhibit with respect to how the player population change. Whilst it is also interesting to investigate the 23% of games that did not display a weekly pattern, it is beyond the scope of this paper.

V. Weekly Population Pattern Discovery

In this section, we present the specifics of the clustering approach undertaken to discover the existence of common weekly player population fluctuation patterns among games.

A. Data Selection

As previously identified, not all games display weekly seasonality in terms of player population fluctuations. In fact, out of the 1963 games of our dataset, only 77% of games displayed weekly seasonality. Thus, only those 1508 games are used in this clustering process. Moreover, to reveal diverse weekly player population fluctuation patterns commonly displayed by games, population data representing how player population normally fluctuate within a week needs to be extracted from each game's time series. As a result of various external events, such as sales, competitions, updates and etc player population of some weeks will display trends. Thus, averaging the timeseries to extract a representative week would not be an ideal method. Hence, for each game, we extract a week from the segment of the time series that has the lowest trend, which is closest to zero, since it represents a fragment in time where the game is not impacted by external events. Moreover, it represents how player population fluctuate within a normal week that is not impacted by external events.

To extract the required data, piece-wise linear trend is calculated as before and the piece with the lowest trend is located. Next, a week starting from Monday and ending on Sunday is identified from within or either side of the selected piece. Finally, the time series is detrended by piecewise linear trend removal and population data corresponding to the identified week is extracted.

It is apparent that the extracted player population data of a week in games of *5mData* and *60mData* are different in size with respect to the number of data points. Thus, hourly population data is extracted from games of *5mData*, so as to reduce the complexity in further computations, yet the population data of each game is still high dimensional. However, this data is used as is in the clustering process, without further dimensional reductions, since our goal is to discover the common weekly player population patterns based on the shape itself.

B. Data Normalization

Data normalization linearly transforms raw data to a specific range of values, such as 0-1, in order to standardize raw data. Normalizing the extracted weekly player population data before clustering is necessary due to several reasons. The overall population size of each game in our dataset is different and range from hundreds to several thousands. Since our goal for the clustering process is to discover diverse weekly player population patterns based on the shape alone, irrespective of the size of overall playerbase of games, extracted weekly population data should be normalized. Moreover, normalization is also compulsory due to our choice of distance measure [21]. Thus, data is normalized to scale the population data to a range of 0 - 1.

C. Distance Measure

Dynamic Time Warping (DTW) is an algorithm capable of recognizing shape-based similarity between two time series that may vary in speeds and length [22]. Due to data collection issues or other reasons there could be slight distortions in player population time series of games. Thus, utilizing point to point distance measures, such as Euclidean distance, is inappropriate in measuring distance between the extracted weekly population patterns of games based on shape. On the other hand, DTW computes the optimal global alignment between two time series utilizing temporal warps, to aid better measure of the similarity between them. We calculate and record the DTW distance for each pair of games in a distance matrix and use it as an input to the clustering process.

D. Clustering Technique and Parameters

We use agglomerative hierarchical clustering technique to perform the cluster analysis. When forming clusters using hierarchical methods, linkage methods are used to determine how the clusters should be formed based on the distance between the data objects [23]. We explored 3 linkage methods to select the best method suited for this clustering task.*Single Linkage* considers the distance between two clusters as the smallest distance between any 2 data objects in the two clusters. As opposed to single linkage, *Complete Linkage* uses the largest distance between any two data objects in the two clusters to represent the distance between the two clusters. Moreover, *Average Linkage* calculates the average distance between all pairs of data objects in the two clusters to determine the distance between two clusters.

In order to determine the best linkage method for our clustering process, hierarchical clustering was conducted using each linkage method on the pre-calculated DTW distance matrix. We evaluated each linkage method using *Cophenetic correlation coefficient*[7], which is a measure that indicates how well the distances between data objects, as provided in the distance matrix, are represented in the dendrogram [24]. Equation (1) represents the Cophenetic correlation coefficient.

$$c = \frac{\sum_{i<j}(D_{ij} - \overline{D})(T_{ij} - \overline{T})}{\sqrt{\sum_{i<j}(D_{ij} - \overline{D})^2 \sum_{i<j}(T_{ij} - \overline{T})^2}} \tag{1}$$

In (1), D_{ij} represents the distance between i and j data objects in the distance matrix D which was used to build the dendrogram. $\overline{D}$ is the average value in D. In the same way, T_{ij} represents the cophenetic distance in the dendrogram

[7]https://www.mathworks.com/help/stats/

between i and j data objects. It's the height of the link at which theses data objects are first joined together in the dendrogram. $\overline{T}$ is the average value of T which has the heights of links connecting all pairs of data objects in the dendrogram.

The results depicted in Table II indicates that Average Linkage method performs best in representing the original distance values between each pair of weekly population patterns of games in the dendrogram. Thus, our hierarchical cluster analysis process is continued with Average linkage. The resulting dendrogram is depicted in Fig. 2.

TABLE II: Cophenetic Correlation Coefficient for different Linkage methods

	Linkage		
Cophenetic Corr.	Single	Complete	Average
Coeffcient	0.44	0.65	0.79

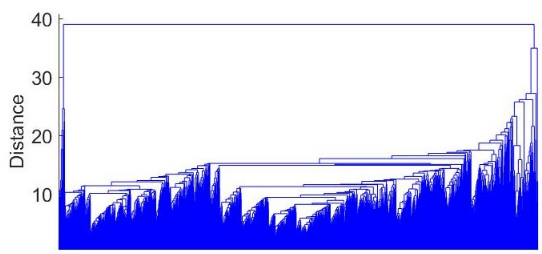


Fig. 2: Dendrogram(1508Games):Clustering of weekly patterns based on the DTW distance by Average Linkage

One of the most important tasks in cluster analysis is deciding the optimal number of clusters. Dendrogram can be used to assist in deciding the number of clusters, as it represents the natural divisions in the data. Our dendrogram, as depicted in Fig. 2 is quite skewed to the right and is not displaying any significant divisions of clusters. At first glance, this may indicate that there are no remarkable differences among the weekly patterns of games. But it is important to explore the types of different weekly patterns the player population data may exhibit, nonetheless how minor the differences are, as these could lead to interesting findings. Hence, we experimented with different upper boundaries for the number of clusters. Specifically, the upper boundaries we used are 5, 10,15, 25, 50, 75 and 100. Only the clusters with 10 or more games were accepted as meaningful clusters as clusters with low number of elements are not appropriate to make generalized claims about weekly patterns of games. Moreover, we noticed that when 5, 10,15,25 and 50 are used as the upper bound, a single dominating cluster with more than 1200 games is generated. This negatively impacts the identification of diverse weekly patterns as more than 80% of games belong to a single cluster. However, at 75 and 100 boundaries the dominating cluster gets separated. Hence, 75 was chosen as the optimum number of clusters for the cluster analysis process. Also, out of the 75 clusters, 10 clusters consisted of at least 10 elements. However, among those 10 clusters, one consisted of 10 games displaying unusual population fluctuations due to low population. Thus, only the 9 meaningful clusters are further explored.

E. Representative Weekly Patterns Generation

Once the clusters are generated, it is necessary to construct weekly player population patterns representing each cluster, in order to visualize the patterns each cluster represents. Thus,for this purpose we propose a DTW based averaging procedure. This procedure also acknowledges the hierarchical clustering approach of joining elements one by one in a bottom up fashion based on the distance between elements. It is presented in *Algorithm 1*

Algorithm 1 Visualizing Representative Weekly Pattern of a Cluster

Inputs: $Data_{NxW}$, $Order_K$
$gameA \leftarrow Data[Order[1]][:]$
$gameB \leftarrow Data[Order[2]][:]$
$[iA, iB] \leftarrow DTW(gameA, gameB)$
$avgAlign_{sizeOf(iA)}$
for $i = 1 : sizeOf(avgAlign)$ **do**
  $avgAlign(i) \leftarrow \frac{1}{2} * gameA[iA[i]] + \frac{1}{2} * gameB[iB[i]]$
end for
for $k = 3 : sizeOf(Order)$ **do**
  $tmpAvgAlign \leftarrow avgAlign$
  $gameB \leftarrow Data[Order[k]][:]$
  $[iA, iB] \leftarrow DTW(tmpAvgAlign, gameB)$
  $avgAlign_{sizeOf(iA)}$
  for $i = 1 : sizeOf(avgAlign)$ **do**
    $avgAlign(i) = \frac{k-1}{k} * tmpAvgAlign[iA[i]] + \frac{1}{k} * gameB[iB[i]]$
  end for
end for

Algorithm 1 takes two inputs. The first input is a NxW matrix named $Data$ where N is the total number of games used in the clustering process, which is 1508, and W is the number of data points representing a week. It also takes the connection order of games within the cluster in the dendrogram as represented in Fig. 2 as an input named $Order$. It is an array of size K where K is the number of games within the cluster. It contains the indexes of games that belong to the cluster currently being considered, ordered based on how they were joined during the clustering process. Initially, *Algorithm 1* identifies the first two games joined in the cluster from $Order$ and uses DTW to align them. The resulting iA and iB contains new alignment indexes of the two games. Next, it generates an average weekly pattern, represented by $avgAlign$, by iteratively averaging each pair of data points in the newly aligned pair of games. Next, the weekly pattern of the next joined game of the cluster and $avgAlign$ is aligned using DTW and averaged. It is iteratively continued until all games in the cluster are used to create the final representative weekly pattern. Moreover, a weighing mechanism is used to correctly account for the number of weekly patterns involved in creating the average pattern. Thus, in each iteration more weight is offered to $tmpAvgAlign$, which represent the average weekly pattern of multiple games.

VI. Results and Discussion

Our clustering process revealed 9 weekly player population fluctuation patterns games exhibit. Fig. 3 presents these discovered patterns where each peak in a pattern corresponds to a day of a week from Monday to Sunday. Among the clusters, Cluster 1 has the highest number of games, which is 760, making it the dominating cluster. As per that cluster, in most games, players tend to play games highly towards the end of the week. It can be seen that the population has started to increase on Friday and has continued to increase till Sunday. Further Cluster 2 which has 446 games, making it another dominant cluster, displays a similar pattern. However, the player population difference between weekdays and weekend tends to be higher in Cluster 1 than Cluster 2. Apart from Cluster 1 and Cluster 2, all other clusters have a comparatively lower number of games in them. In general, even these non-dominating clusters represent a weekly pattern where population is higher during the weekend than weekdays. However, there exists some significant differences in how population vary during the weekend among these clusters. For instance, in Cluster 3, 4, 6 and 8, player population is greater on Saturday than Friday and Sunday. This rise is higher in Cluster 8. This suggests to us that players tend to involve with some games mostly on Saturdays. By contrast, players are also likely to play some other games heavily on Sundays as depicted in Cluster 7. It is interesting to observe that the player population has intensely escalated on Sunday in Cluster 7 compared to all other clusters. On the other hand, Cluster 5 has the same player population in both Saturday and Sunday. Lastly, Cluster 9 seems to be quite interesting as it represents a weekly pattern where population is consistent throughout the whole week as presented. We could also observe from the clusters that in some games the player population within weekdays is significantly lower than that of during the weekends, specially in Cluster 1,4,7 and 8. However in other clusters, the difference between weekdays and weekends seems to be quite moderate.

A. Extraction of Game Characteristics

In order to provide further insights about the discovery of weekly population patterns, we explore the common characteristics of games such as tags, age requirements and overall population size of games in each cluster.

Tags: Tags are used for characterizing games. Each game has multiple tags of different sizes where *tag size* means the number of players who have assigned a certain tag to the game. We use a weighing criteria to determine how each game in a cluster contributes towards a certain tag. The weighing criteria should be made independent of the overall population size of a game so as to allow fair inclusion of tags of games with less population within a cluster. As per the definition, tag size has an indirect correlation with the overall population size of a game. Thus, the weight contribution of a game G towards a certain tag is calculated by (2).

$$Weight_{G,Tag_x} = \frac{Size_{G,Tag_x}}{\max_{i=1..T}(Size_{G,Tag_i})} \qquad (2)$$

Since we divide the size of Tag_x by the maximum tag size value out of all T tags of that game, the weight a game contributes towards a tag is now independent of the overall player population of a game. Finally, the percentage of a certain Tag in a cluster is determined by calculating the sum of weights contributed by all games in the cluster and dividing it by the number of games in the cluster and multiplying by 100. The top 10 tags with the highest percentage in each cluster is presented in Table III.

Age Restrictions: Based on the time availability, interest and other factors, weekly playing patterns could differ among players of different age groups. The rating system of ‘Australian Classification Board’ assigns each game an age based class based on the content of the game. We explore the age restriction based classes of all games in the smaller clusters only, which are Cluster 3 -9, as this required manual collection of data. We calculate the percentage of games of each age class in each cluster. Results are depicted in Fig.4.

Overall Population Size: For each game, the average player population throughout the 6 months time period the data were collected is calculated. Next, for each cluster, Minimum, Average and Maximum of Average player population of all games is calculated. These value ranges provide a perception about the population sizes of games in each cluster as presented in Fig. 5.

B. Discussion

Extraction of game characteristics in each cluster revealed several perceptions about the games that display diverse weekly player population patterns. Games displaying the pattern of Cluster 1 are mainly Action, Singleplayer, Multiplayer and Strategy games where the mean population ranges between 9 to 41742. Such a high population range is observable as Cluster 1 consists of the highest number of games. However, it is interesting to note that the pattern of Cluster 2 is observable in games of much larger range of population although Cluster 2 has lesser games than Cluster 1. It could be be due to the popular titles in Cluster 2 such as *PUBG, Dota2* and *Counter Strike:Global Offensive*. Games in Cluster 2 are mainly Multiplayer, Action, FreeToPlay and Strategy combinations. While the major tags in Cluster 3 are Indie, Action and Multiplayer, its mean population range is quite smaller. In Cluster 4, Action tag is leading and its percentage is quite higher than the next tags, which are Multiplayer and Adventure. Its mean population range between 23 and 54927. Most games displaying Cluster 5 pattern are FreetoPlay, Mutiplayer and OpenWorld. Interestingly, OpenWorld appears among top 3 tags only in this cluster. Also, its maximum mean population is 17513, which is somewhat smaller than Cluster 4. Almost all games displaying Cluster 6 pattern are FreeToPlay, as its percentage is 98%. Also, the existence of Massively Muliplayer, Multiplayer and MMORPG describes that most games in this cluster are multiplayer to some level.

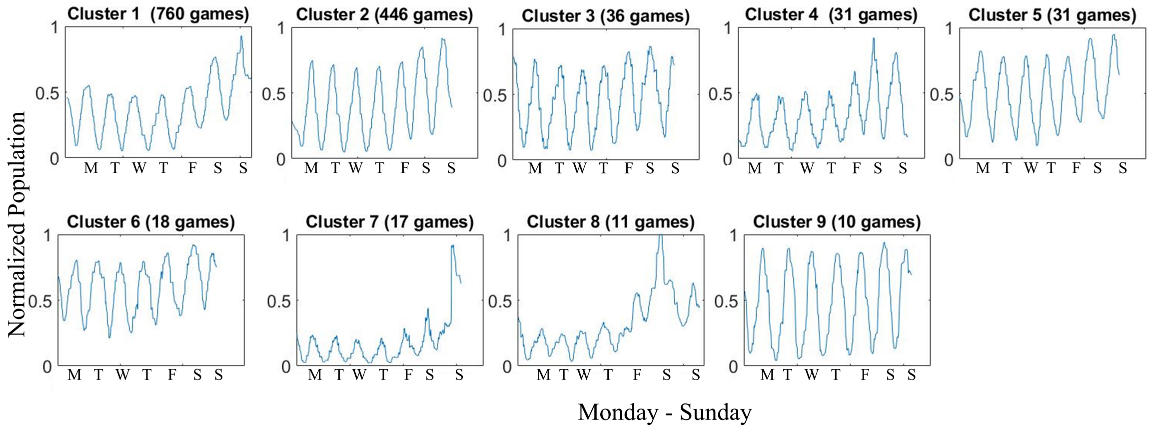


Fig. 3: Representative Weekly Player Population patterns of Clusters: Represents how player population change from Monday to Sunday. Each peak corresponds to a day of a week

TABLE III: Top 10 Tags of each Cluster sorted by the rounded-up Percentage

Clust1	Action[40%], Singleplayer[35%],Multiplayer[31%],Strategy[31%],Adventure[29%], Open World[24%],Indie[24%],RPG[23%],Simulation[23%],Survival[16%]
Clust2	Multiplaye[36%], Action[34%],Free to Play[30%],Strategy[29%],Simulation[25%], Singleplayer[22%], Open World[18%],Adventure[18%],RPG[18%],FPS[13%]
Clust3	Indie[40%], Action[37%],Multiplayer[36%],RPG[31%],Strategy[27%], Adventure[24%],Singleplayer[21%], Massively Multiplayer[20%],Open World[20%],Simulation[20%]
Clust4	Action[51%],Multiplayer[36%],Adventure[36%],Indie[33%], Singleplayer[27%] , RPG[25%], Open World[22%],Simulation[21%],Survival[21%],Strategy[21%]
Clust5	Free to Play[51%],Multiplayer[43%],Open World[40%],RPG[32%],Action[30%] Massively Multiplayer[28%],Survival[27%],Adventure[26%],Strategy[26%], Simulation[22%]
Clust6	Free to Play[98%],Massively Multiplayer[56%],Action[52%],Multiplayer[51%],RPG[37%], MMORPG[35%], Anime[34%],FPS[29%],Open World[26%],Adventure[26%]
Clust7	Action[52%],Indie[51%],Adventure[32%],Casual[30%],Singleplayer[25%], Multiplayer[24%],Strategy[23%],Funny[23%],RPG[20%],Great Soundtrack[17%]
Clust8	Action[65%],VR[39%],Adventure[30%],Multiplayer[29%],First-Person[29%], Co-op[28%],Singleplayer[24%],FPS[22%],Indie[20%],Strategy[19%]
Clust9	Action[38%],Strategy[36%],Free to Play[30%],Singleplayer[25%],Adventure[20%], Platformer[19%],Multiplayer[18%],Great Soundtrack[18%],Simualtion[15%],Casual[13%]

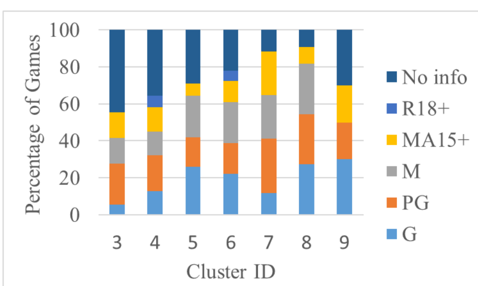


Fig. 4: Age based Class Distribution of Games in Clusters: G -General(Suitable for everyone),PG -Parental Guidance recommended for persons under 15, M -Mature(For persons aged 15 or more),MA15+ -Mature Accompanied(Legally Restricted for persons aged 15 or more), R18+ -Restricted for Adults.

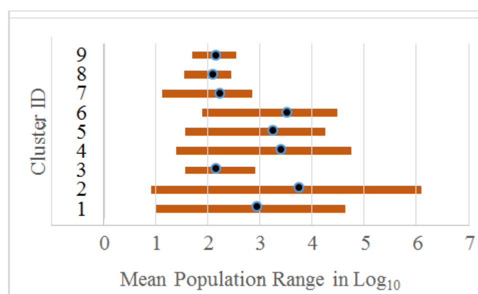


Fig. 5: Range of Mean Population of games in Clusters: Per each cluster, mean population of each game over 6 months is calcluated. The minumum, mean and maximum of those values are converted to $\log_{10}$ and presented

Probably as a result, the minimum mean population is highest in this cluster. Cluster 7 pattern is displayed in Action and Indie games. Interestingly, games tagged as Funny also exists in Cluster 7. The mean population of games have a smaller range and also a small minimum value of 12. Interestingly, this cluster contains some popular titles, Far Cry Primal and Assassin's Creed Syndicate from Ubisoft developers. Also, 2 games from the same sequel, The Jackbox Party Pack 3 and 4 also appears. Since these 2 are party games, their appearance in cluster 7 which has a pattern of very lower population during the weekdays and a significantly high population on

Sunday is acceptable. Cluster 8 has mainly Action and VR games and VR did not appear in other clusters. Thus, it is arguable whether there is any relationship with access to VR equipment and time to play VR games has any relationship with the pattern of population increasing highly on Saturday compared to the weekdays. Also, Cluster 8 has the lowest mean population which may indicate that a small number of players play VR games. Cluster 9 games are mainly Action, Strategy and FreeToPlay. The small range of mean population indicates that all games in this cluster have somewhat similar population size.

Analysis of tags across clusters indicate that there is no unique tag as the most frequent tag per cluster. Thus, it is evident that the discovered weekly patterns are not associated with a single tag but with different combinations of tags as presented in Table III. Moreover, the distribution of age based classes do not appear to be significantly different among each cluster. However. the mean values of the mean population of games in clusters are quite different among clusters. This may indicate the influence of overall population of a game to a weekly player population pattern. In general, these game characteristics provide some insights about the diverse weekly population patterns. However, these major characteristics alone seems to be insufficient to clearly distinguish each weekly pattern. Hence this leads us to further investigate the impact of other game characteristics such as, price, sale event frequency, churn rate and different tag classifications(game genres, game elements).

VII. Conclusion and Future Work

In this paper, we focused on exploring the player population fluctuations of online games in order to identify whether games display recurring weekly population fluctuation patterns and to discover such archetypal patterns along with how frequently those are observed. We identified that most popular games, in fact 77%, display weekly seasonality. Moreover, we were able to detect 9 different weekly population fluctuation patterns that online games exhibit. Among these there were 2 dominant patterns which were highly similar and it indicated that most games display a weekly pattern where player population increases towards the weekend. Also, our study reveals that generating archetypal game characteristics for each weekly pattern based on tags, age requirement and overall population is not straightforward. However, some weekly patterns can be distinguished based on tags and overall population size to some extent. The outcomes of this study provides assistance in understanding the different patterns of the number of players, games with different characteristics attract every week. As future work, we expect to explore the diverse patterns games exhibit with respect to player population fluctuation throughout the lifetime of a game. Moreover, we plan to investigate the impact of external factors such as social media influence and sales on game player population.

References

[1] T. Wijman, "Newzoo's 2018 Report: Insights Into the $137.9 Billion Global Games Market," 2018.

[2] M. S. El-Nasr, *Game Analytics: Maximizing the Value of Player Data*. New York: Springer, 2013.

[3] S. Rezaei and S. S. Ghodsi, "Does value matters in playing online game? An empirical study among massively multiplayer online role-playing games (MMORPGs)," *Computers in Human Behavior*, vol. 35, pp. 252–266, Jun. 2014.

[4] T. Debeauvais, B. Nardi, D. J. Schiano, N. Ducheneaut, and N. Yee, "If You Build It They Might Stay: Retention Mechanisms in World of Warcraft," in *Proceedings of the 6th International Conference on Foundations of Digital Games*, ser. FDG '11. New York, NY, USA: ACM, 2011, pp. 180–187.

[5] J. Runge, P. Gao, F. Garcin, and B. Faltings, "Churn prediction for high-value players in casual social games," in *2014 IEEE Conference on Computational Intelligence and Games*, Aug. 2014, pp. 1–8.

[6] R. Sifa, A. Drachen, and C. Bauckhage, "Large-Scale Cross-Game Player Behavior Analysis on Steam," in *Proceedings Of the eleventh Artificial Intelligence and Interactive Digital Entertainment International Conference (AIIDE 2015)*. AAAI, 2015, pp. 198–204.

[7] X. Liu, K. Merrick, and H. Abbass, "Toward Electroencephalographic Profiling of Player Motivation: A Survey," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 10, no. 3, pp. 499–513, Sep. 2018.

[8] D. Williams, N. Yee, and S. E. Caplan, "Who plays, how much, and why? Debunking the stereotypical gamer profile," *Journal of Computer-Mediated Communication*, vol. 13, no. 4, pp. 993–1018, Jul. 2008.

[9] H. D. Menéndez, R. Vindel, and D. Camacho, "Combining Time Series and Clustering to Extract Gamer Profile Evolution," in *International Conference of Computational Collective Intelligence*, 2014, pp. 262–271.

[10] A. Saas, A. Guitart, and A. Periáñez, "Discovering Playing Patterns: Time Series Clustering of Free-To-Play Game Data," *arXiv:1710.02268 [cs, stat]*, pp. 1–8, Sep. 2016.

[11] F. Baumann, D. Emmert, H. Baumgartl, and R. Buettner, "Hardcore Gamer Profiling: Results from an unsupervised learning approach to playing behavior on the Steam platform," *Procedia Computer Science*, vol. 126, pp. 1289–1297, Jan. 2018.

[12] C. Bauckhage, A. Drachen, and R. Sifa, "Clustering Game Behavior Data," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 7, no. 3, pp. 266–278, Sep. 2015.

[13] A. Drachen, J. Green, C. Gray, E. Harik, P. Lu, R. Sifa, and D. Klabjan, "Guns and guardians: Comparative cluster analysis and behavioral profiling in destiny," in *2016 IEEE Conference on Computational Intelligence and Games (CIG)*, Sep. 2016, pp. 1–8.

[14] R. Sifa, C. Bauckhage, and A. Drachen, "The Playtime Principle: Large-scale cross-games interest modeling," in *2014 IEEE Conference on Computational Intelligence and Games*, Aug. 2014, pp. 1–8.

[15] C. Bauckhage, K. Kersting, R. Sifa, C. Thurau, A. Drachen, and A. Canossa, "How players lose interest in playing a game: An empirical study based on distributions of total playing times," in *2012 IEEE Conference on Computational Intelligence and Games (CIG)*, Sep. 2012, pp. 139–146.

[16] C. Chambers, W.-c. Feng, S. Sahu, and D. Saha, "Measurement-based characterization of a collection of on-line games," in *Proceedings of the 5th ACM SIGCOMM conference on Internet Measurement*. USENIX Association, 2005, pp. 1–1.

[17] C. Smith, "34 Interesting Steam Statistics — By the Numbers," https://expandedramblings.com/index.php/steam-statistics/, Dec. 2018.

[18] I. Pitas and A. N. Venetsanopoulos, "Median Filters," in *Nonlinear Digital Filters*, ser. The Springer International Series in Engineering and Computer Science. Springer, Boston, MA, 1990, pp. 63–116.

[19] G. E. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control, Revised Ed*. Holden-Day, 1976.

[20] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*. OTexts, 2018.

[21] P. Montero and J. A. Vilar, "**TSclust** : An *R* Package for Time Series Clustering," *Journal of Statistical Software*, vol. 62, no. 1, 2014.

[22] A. Mueen and E. Keogh, "Extracting Optimal Performance from Dynamic Time Warping," in *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '16. New York, NY, USA: ACM, 2016, pp. 2129–2130.

[23] O. Maimon and L. Rokach, Eds., *Data Mining and Knowledge Discovery Handbook*, 2nd ed. New York: Springer, 2010.

[24] R. R. Sokal and F. J. Rohlf, "The Comparison of Dendrograms by Objective Methods," *Taxon*, vol. 11, no. 2, pp. 33–40, 1962.