BBC iPlayer Cluster Analysis – Technical Report

November 15th, 2020

Table of Contents

1. ABSTRACT

This report examines, discovers and reveals distinct customer segments of the BBC iPlayer based on usage behaviour and preferences by means of cluster analyses. The BBC aims to engage the public by offering their contents on their own streaming platform which allows for not only more convenient usage for customers but also for better data collection on these customers. We investigate a data set with several thousand iPlayer users characterised by the frequency of usage, the time of day of usage as well as the preferred genres. We start the analysis with a basic cluster algorithm and find that distinct clusters exist; more specifically, we find three to be the optimal number of clusters. We conclude the analysis with a great confidence in the three clusters that we have found as we are able to validate both our method and our model. The resulting customer groups of the BBC iPlayer that we discover are depicted by a first large cluster mainly watching Drama shows during the evening or night as well as mostly finishing their shows; a second small cluster depicted by children that watch Children and Learning shows, mainly during the afternoon; and a third medium cluster mainly interested in watching parts of News, Weather and Sports shows. We outline business implications and recommendations to the BBC concerning targeted advertising, tailored content and potential premium services. Lastly, we describe the limitations of our analysis and how it may be improved in the future.

2. INTRODUCTION

The following report describes a cluster analysis of the BBC iPlayer's users and their watching behaviour. Hereby, we use data mining techniques to establish a data-based understanding of the BBC iPlayer's content and its customers. We aim to identify meaningful clusters that group the iPlayer's users into distinct customer segments based on their viewing behaviour and preferences. These will allow us to gain insight into the existing customer groups of the BBC iPlayer and, consequently, enable us to derive business implications for the BBC to better tailor the product to their customers and improve customer acquisition and/or retention.

The cluster analysis and this report are set up in the following structure:

- We first clean the data and then prepare it for the analysis
- We start our clustering analysis with the k-means method
- We decide on the optimal number of clusters and then validate this by using alternative clustering methods, namely K-medoid (PAM) and Hierarchical clustering
- We validate our final model by training the model on half of our data sample and then testing it on the other subsample
- We compare the different results and derive conclusions on which clusters are most meaningful
- We make business recommendations for the BBC based on these results

The underlying data of this report is derived from the BBC's iPlayer customer base and the attributes of their viewing sessions. While the BBC traditionally reached its audience through TV broadcasting, the BBC iPlayer was recently introduced as its own digital TV streaming service to deliver content on demand. The iPlayer is not only a more modern and convenient vehicle to provide the BBC's audiences with content, but it also allows the BBC to collect data on its viewers as well as their engagement and preferences in an easier and more efficient way.

The original data file for this cluster analysis was extracted from the BBC iPlayer database and collected data on approximately 10,000 random viewers determined in January and their viewing behaviour from January to April. While this limits the customers included in the subsample to customers that watched in January, we must keep in mind that the data is no longer a random sample of the iPlayer viewers after January.

3. METHODOLOGY

3.1. Phase 1: Data Processing & Data Exploration

After the raw data is thoroughly cleaned and structured, we begin to prepare the data set so that it is ready to be used in the cluster analysis. As the initial data is engagement-based data, meaning each observation represents one viewing session, we convert this data into user-based data. In order to do so, we group together all recorded viewing sessions of one user. This will enable us to understand different viewing habits per user rather than per viewing event.

We finally generate a data set that shows a summary of each user's viewing behaviour per observation (row) in the data set. Here, each user is identified by a unique user ID and further described by the number of shows watched and length of all viewing sessions as well as the proportions of different day times and genres that a user watched. We further add a variable that indicates how likely a user is to watch more than 90% of any show respectively. This allows us to understand how likely people are to complete a show.

As a next step, we derive the correlations between all variables to understand how these interact with each other. We find perfect negative correlation between the weekend and weekday variable (any single viewing can only be either on the weekend or on a weekday) as well as very high positive correlation between the total time watched and number of shows per user (the more shows watched, the higher the total time watched). Therefore, we later on eliminate the number of shows variable and the weekday variable in our data set for the model to not double weigh these influences.

After further investigation of the distributions of the number of shows and total time variables, we observe many outliers in the data set. While the overall median number of shows per user is around 3, there are multiple users that watched more than 100 and one even up to 782 shows. Similarly, the median minutes watched per user is below 20 minutes although multiple viewers record total watching times of up to 10,000 minutes and few even higher than that. We conclude that we need to account both for the low-frequency users that do not add any contextual value to our data set as well as the very high-frequency users that do not show viewing behaviours representative of the entire population and may be mistakes in the data. We do so by firstly, eliminating all low engagement users, meaning users with a total watching time below 5 minutes and/or a number of shows watched lower than 5. As this removes a significant share of 61% the observations, the resulting data set only includes 3,625 user observations. Secondly, we log transform the total time variable (replacing each value total_Time with log(total_Time)) to reduce the skewedness of the distribution and, thereby, reduce weight of outliers on our cluster analysis.

As additional final steps, we remove variables that we identified above as irrelevant for our analysis. Lastly, we standardize the resulting data to make all variables with different unit scales comparable.

3.1.1 Phase 1: Outcome

We derive a user-based data set that includes only dimensions relevant to the following cluster analysis and is now ready to be processed in the cluster analysis. A more detailed list of the variables included can be found in *Appendix 1*.

3.2. Phase 2: Training a simple k-means model

After sufficiently exploring and adjusting the data, we are ready to analyse possible clusters by means of training a simple k-means model. We begin by using the k-means method to search for 2 possible clusters, setting k = 2. K-means clustering is our initial choice for the analysis as it is a universal and commonly applied clustering method which is, additionally, less computationally intensive than other clustering methods.

The k-means clustering algorithm aims to partition observations into k clusters (for a specified number of clusters k) in which each observation belongs to the cluster with the nearest induced mean (cluster centres). Each center gives us insight into what the dynamics of its representative cluster are and acts as a prototype for its respective cluster from which conclusions about all observations in that cluster can be drawn.

Using the k-means algorithm for k = 2, yields two distinct clusters with a size of 2,404 and 1,221 respectively, each cluster embodying a unique set of characteristics. These clusters are unique in terms of both the time of viewing, genres viewed and whether or not they are likely to finish watching shows (weight_pct_90).

These characteristics are summarized by plotting the cluster center per dimension (variable) in *Figure 1* as well as describing the resulting differences between the clusters in *Table 1*.

	Cluster 1	**Cluster 2**
Size	**2404** (66%)	**1221** (34%)
Favours	▪ High total watch time (may also correspond to more shows watched) ▪ Evening/night-time watching ▪ Drama shows ▪ Watching complete show (>90%)	▪ Daytime/afternoon watching ▪ Children, Learning, News shows
Adverse to	▪ Afternoon/daytime watching ▪ Learning, News shows	▪ High total watch time ▪ Evening watching ▪ Drama shows ▪ Watching complete show (>90%)

Table 1. Derived cluster characteristics for simple k-means modelling of two clusters

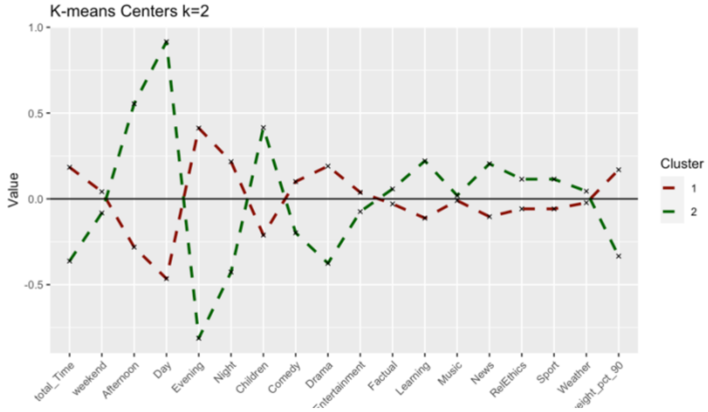


Figure 1. K-means centers plotted per dimensions for two clusters

This simple 2 cluster model allows us to already gain first insights into the dynamics of BBC iPlayer users. Cluster 1 is more representative of evening- and night-time viewers who perhaps have more time to watch and complete their shows and who have a strong preference towards Drama programs. While cluster 2 is representative viewers, who primarily watch Children, Learning, News and Sport programs during the day and do not have time or interest to complete the full show on account of their afternoon/daytime viewing.

As a next step, we can visualize these two clusters by mapping them over two principal components (dimensions), established through principal component analysis (PCA) techniques. This visualization supports us in determining whether the clusters are distinct enough or significant overlap occurs.

While we do notice a fair amount of overlap between these two cluster, we can also see a fair share of distinction between them. However, it is important to note that this PCA only portrays roughly 20% of all variables under consideration. Since the clusters are based across greater dimensions, this is only the best representation in a 2D space.

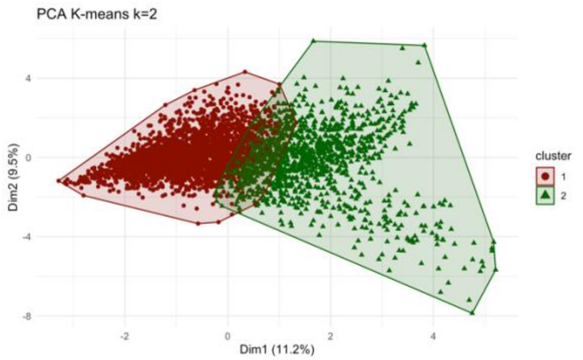


Figure 2. K-means PCA for two clusters

3.2.1 Phase 2 Outcome

Phase 2 helped us to determine if there are indeed differences among BBC iPlayer users so that these users could potentially be divided into multiple, different segments. The fact that two distinct groups emerge is promising and allows us to already see the characteristics that users in each cluster seem to embody. However, we cannot be sure if these two clusters can be further broken down into more meaningful, distinct clusters from which we can draw better conclusions.

3.3. Phase 3: Deriving optimal number of clusters

We proceed our analysis by completing the same process as described above across for greater numbers of clusters. Consequently, we increase k to 3, 4, and 5 and make use of further clustering tools to determine the optimal number of clusters. The resulting center plots and PCA visualisations can be found in *Figure 3*, *Appendix 2* and *Appendix 3*.

The cluster analysis for k = 3 reveals the existence of a third and relatively small cluster (Cluster 2, 5% of users) which can be still be found throughout the subsequent k = 4 and k = 5 analysis. Despite being small in size, this cluster is relevant as it is very distinguished from the remaining clusters in its attributes, depicting new insights into the dynamics of daytime/afternoon viewers which can be broken down into multiple customer groups with different genre preferences. This small cluster most likely represents children viewers (e.g., family accounts where the children use the BBC iPlayer) as it is characterized by not only afternoon-watching but also the Children and Learning genres.

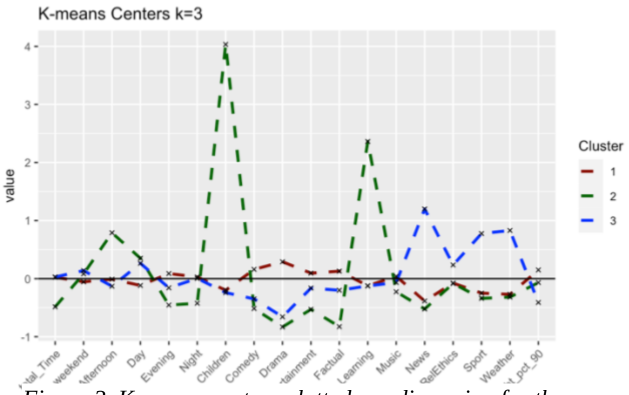


Figure 3. K-means centers plotted per dimension for three clusters

Considering the center plots and the different PCA analysis only, increasing the number of clusters k to 4 or higher does not create separate or distinguishable clusters as there is significant overlap of cluster centers and the clusters themselves. Increasing k to 4 or higher, therefore seems to hinders our ability to draw meaningful conclusions.

To validate this hypothesis and determine the optimal number of clusters (i.e., make the right choice of k), we use a series of tools and techniques that do not necessarily provide definitive answers but serve as a guideline for optimal choice of k.

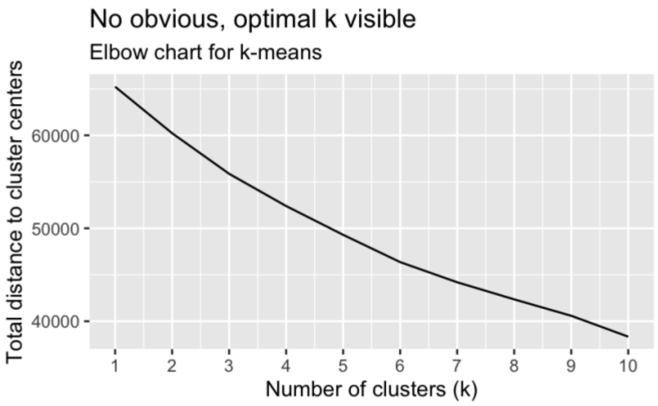


Figure 4. Elbow chart depicting optimal choice of number of clusters for k-means

Elbow chart. Elbow charts plot the total distance from observations to their cluster centers as a function of number of clusters. In k-means the total distance continuously decreases with increasing number of clusters with a minimized distance for when each observation is in its own cluster. Therefore, there is no optimal stopping point, meaning, we cannot choose the number of clusters that minimizes the total distance as this would be meaningless. However, we can look for points which show the greatest decrease in distance for increases in k.

Plotting the elbow chart for k-means in *Figure 4* does not yield much information as the resulting graph is a rather smooth line without the significant kink/"elbow" that we expected to find (no significant concavity). Although we seem to observe a smaller rate of decrease after k = 6, we cannot identify one obvious optimal k. Nevertheless, we seem to observe a slightly greater rate of decrease in distance between k = 1 and k = 3, implying that we will most probably end up with two to three but definitely no more than six clusters. We will consider additional techniques to narrow down this choice of k.

Silhouette Analysis. The Silhouette analysis provides a concise graphical representation of how well each observation has been classified into a relevant cluster. The silhouette value is a measure of how similar an object is to its own cluster (cohesion) compared to other clusters (separation). The silhouette ranges from −1 to +1, where a high value indicates that the object is well matched to its own cluster and poorly matched to neighbouring clusters. By comparing the silhouette width averages over all clusters for multiple k, we can find k that maximizes this average. The individual silhouette graphs can be seen in *Appendix 4*. *Figure 5* summarizes the silhouette width averages for up to k = 8 clusters, deciding to limit k at 8 as individual cluster sizes otherwise become too small to draw meaningful conclusions on.

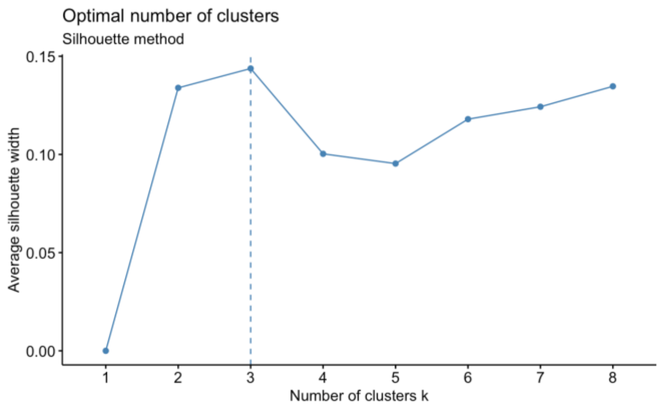


Figure 5. Silhouette analysis depicting optimal choice of number of clusters for k-means

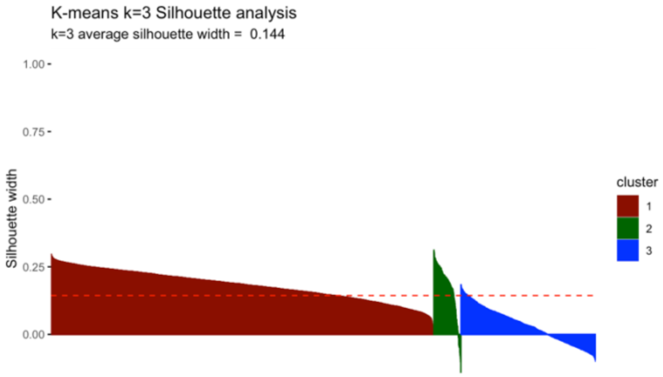


Figure 6. Silhouette analysis for k-means with optimal three clusters (k=3)

This method shows an optimal choice of three clusters, meaning we should set k = 3 in our k-means clustering model in order to most correctly classify observations into their respective clusters. Reviewing individual silhouettes plots (*Appendix 4*) for two to five clusters, we find the greatest average silhouette width at 0.144 for k = 3 (compared to 0.134 for k = 2). Increasing k to 4 or 5, reduces this width, and thereby the desired separation between clusters, again. Although we see that the average silhouette width increases again with k >5, we disregard these higher number of clusters for simplicity sake. Looking at the individual silhouette plots, k=3 (*Figure* 6) has the least observations with negative or below average silhouette width, meaning points are most accurately assigned to respective clusters.

3.3.1 Phase 3 Outcome

This phase allowed us to select the optimal number of clusters that we should consider reviewing. Using a variety of tools and techniques we derive three clusters (k = 3) as the optimal choice to represent BBC iPlayers users.

3.4. Phase 4: Method validation using alternative clustering methods

In the next phase, we make use of two common alternative clustering algorithms to validate the results from phase 3 by means of using alternative techniques to k-mean.

Alternative method 1: Partitioning Around Medoids (PAM)

- The PAM algorithm is very similar to k-means, in that both algorithms are partitional and attempt to minimize the distance between points assigned to a cluster and a point designated as the center of that cluster. While in k-means each cluster center is represented by the mean of the observations in a cluster, PAM determines one actual observation in a cluster as a center, namely a medoid, which minimizes the distance to the remaining points of that cluster. Since the PAM does not rely on means, its algorithm is more robust to noises and outliers than k-means which is highly influenced by outliers.

Alternative method 2: Hierarchical Clustering (H-clust)

- H-clust is a useful method to visualize the impact of changing the number of clusters. It differs from k-means and PAM in that it starts by treating each observation as an individual cluster, then identifies the two clusters that are closest together, merges these two most similar clusters and then repeats these steps until all the clusters are merged together. We make use of the *Ward's distance* (ward.D) variation of this method which merges cluster pairs whose joint cluster minimizes the increase in the total sum of squares within-cluster error.

We complete the cluster analysis for k = 3 through PAM, and H-clust and derive all relevant plots and evaluations to compare with the outputs of k-means. Please refer to *Appendix 5-8* for this.

After completing the alternative methods, we notice great validation of the dynamics from the three clusters established in k-means (some recurring similarities amongst all methods). Especially the H-clust analysis almost replicates our results from k-means. However, we also find a few noticeable differences in center characteristics, cluster sizes and silhouette averages from the PAM outputs that make us more hesitant in the robustness of our results and the conclusions that can be drawn. A detailed analysis of the attributes and sizes of the derived clusters comparing different methods can be found in *Appendix 9*.

While the k-means method with an optimal number of 3 clusters revealed three distinct customer segments, the Hierarchical clustering (ward.D) method validated the attributes and sizes of these clusters. Although the clusters determined by k-means and H-clust are very similar, we observe significant differences in the clusters determined by PAM. While cluster 1 inhibits mostly similar attributes over all three methods, their sizes differ significantly in PAM and cluster 2 and 3 show differences in both size and relevant attributes. Firstly, k-means and H-clust assign each assign around 5% of the users to cluster 2 while H-clust assigns 16%. Secondly, k-means and H-clust define cluster 2 to be a niche, children's cluster, whereas PAM does not identify similar preferences at all. As PAM seems to flip the attributes of previously defined cluster 2 and 3 while keeping cluster 2 as the smallest and the silhouette plot of the PAM clusters indicates that cluster 3 is not meaningful at all, we conclude that PAM is not able to identify the niche children cluster due to its underlying principles of avoiding small clusters.

3.4.1 Phase 4 Outcome

This section compared alternative clustering methods to k-means for determining the attributes and size of each cluster for k = 3 (three clusters). Thereby, we investigated the validity of k-means with three clusters and the meaningfulness of these results across methods. The H-clust algorithm using the ward.D method, reassured us of our results as center plots, PCA as well as silhouette analysis all yield very similar results to those derived from the k-means method. At the same time, PAM yields somewhat varying results to those in k-means, particularly in terms of cluster size and the characteristics of the smallest cluster.

As we do still notice recurring combinations of relevant cluster attributes throughout all three methods, we are relatively confident in the k-means method and the three identified clusters.

3.5. Phase 5: Model validation

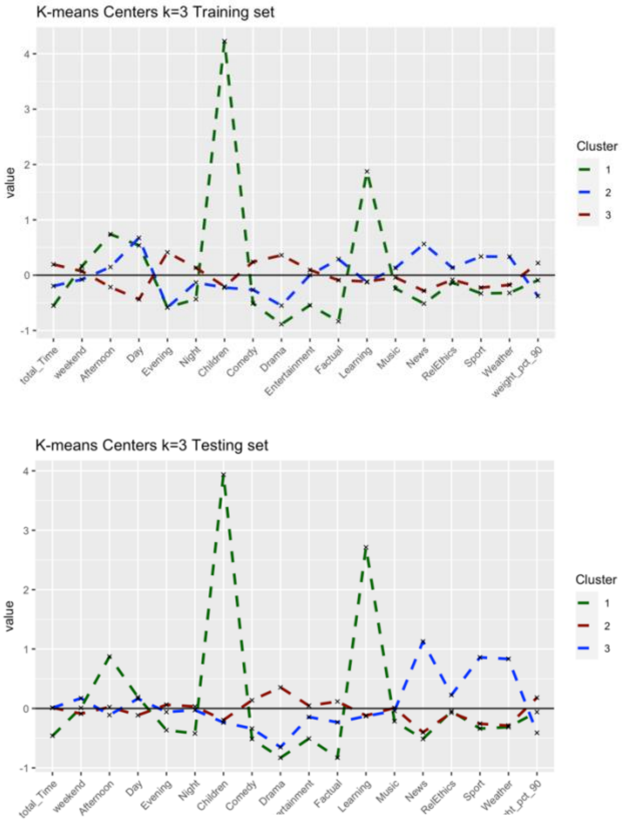


Figure 7. Comparison of center plots of k-means training and testing set for k=3

After validating our method, we, eventually, try to validate our model to reinforce conclusions and verify that our results are not due to chance. We do this by dividing the data into two equal parts, meaning 50% of the data to train the model and 50% of the data to test that model so that we can compare the results and understand if our model is reproducible (*Figure 7*).

We find that the subsample check validates our previously identified three clusters (under all data points) with very similar distinctive characteristics. When plotting the center plots of the training and testing data we find great similarities between both these two plots as well as these two plots and our original k-means model. The fact that these plots are almost identical, reassures us of the validity of resulting clusters attributes, and that these resulting dynamics do not occur by chance or random seed setting but do actually exist.

3.5.1 Phase 5 Outcome

The validation of similar results under testing and training data reassures us of the meaningfulness of the characteristics identified amongst these 3 clusters. The fact that they are not only similar to each other but also to the parent population (model under all data) assures us that these are indeed unique characteristics/attributes that distinguish 3 clusters of BBC iPlayer users.

4. CONCLUSION

By means of the cluster analysis we were able to discover and validate distinct clusters that summarise typical viewing behaviours of similar customer groups of the BBC iPlayer. Roughly outlined, these are the following:

1. A **large cluster** (~70%) that mainly watches **Drama** shows, prefers to watch in the **evening or night** and **finishes** their shows; these users are adverse towards News and Weather shows
2. A **small cluster** (~5%) that watches **Children** and **Learning** shows in the **afternoons**; these users don't watch for a long time or during the evening or night time and are not interested in any other genre
3. A **medium-large cluster** (~25%) that is mainly interested in **News**, **Weather** and **Sports** shows; these users often do not finish their shows and are adverse towards **Drama** and **Comedy** shows

We were hesitant about the importance of the 'Children' cluster at first as it is such a small cluster in size relative to the other two. When looking at the PCAs we do, however, find that this small cluster takes up a similar area as the other clusters which means that we do not violate one of the underlying assumptions of k-means clustering that all clusters should be of similar area size. Further, we conclude that we observed deviating cluster characteristics and sizes from PAM as it does not allow for niche clusters and, therefore, mixed the actually distinct Children's cluster (2) with the other observations.

Considering the validation through H-clust in phase 4 as well as the training and testing set in phase 5, we are confident to conclude the existence of the three identified clusters. The fact that two methods resulted in almost identical clusters and the 50:50 split in training and testing again reflecting almost identical center attributes makes us confident that our resulting clusters did not occur by chance but do exist in the underlying population.

After using different clustering data mining techniques throughout the analysis, we succeed in establishing a quantitative understanding of the BBC iPlayer's customer segments. We are able to identify three meaningful clusters, significantly distinct in their size as well as viewing behaviour and preferences. Finally, this allows us to derive business implications and recommendations for the BBC.

5. RECOMMENDATIONS

Identifying three distinct clusters in the data set, we can now make business recommendations based on these customer segments. As we do not guarantee that the size and the attributes of these three clusters is both 100% representative of the population and representative of 100% of the population, we recommend the BBC uses these implications as a guideline rather than strict rules to follow.

Firstly, we find that the majority of iPlayer users watched in the evening and seems to enjoy Drama and other Entertainment shows rather than news or weather reports. As this customer segment is the largest, we recommend putting most resources into improving and customizing the content for these customers. As these customers may be working during the day and streaming movies or series in the evening, the BBC should focus their advertising for these users on the evening and night hours. Since these users seem to mostly finish their shows, ads can also be played during the show without fearing that the user stops watching.

Secondly, we find a niche cluster (2) that seems to reflect children and perhaps their parents watching Children and Learning shows during the day and afternoon. Although this cluster is small in size, families often have a large purchasing power and children can have great influence on their parents so that targeted and timed advertisement of, for example, toys could lead to children asking their parents for these products and, consequently, valuable ad space that the BBC can sell.

Lastly, a quarter of users seems to be interested in the news, weather and sports while not usually finishing their shows or watching entertainment programs. For these users, it may be profitable to lock them into the rest of the BBC ecosystem and steer them towards the BBC News and BBC Sports online channels. Here, their interest is answered with other means of information delivery while these users will stick to the offerings of the BBC.

Overall, these three segments have very different interests and, therefore, the BBC iPlayer could offer customized offerings to better meet these. For example, a BBC iPlayer for kids that allows parents to not worry about the possibility of other non-children-friendly shows as well as a separate BBC iPlayer interface for News or Sports that is connected to the rest of the content the BBC delivers online on these topics and advertises products specific to the context. Although the iPlayer is free for people that pay a TV license in the UK, the BBC may be able to monetize premium content that is specifically tailored to these customer segments' desires. For example, an add-on package that a user can subscribe to that allows them to watch other shows than the basic iPlayer provides (e.g., newer movies, or additional sports games).

Limitations. Again, these recommendations serve as a guideline for the BBC iPlayer to improve their offering. However, we recognize that the data and the cluster analysis is subject to several limitations that must be taken into account and should be improved in a future analysis.

The iPlayer counts more than 300 million monthly requests while approximately a third of all British internet users use the iPlayer in some form [1]. In 2019, the BBC reported that more than 1 million users sign into the iPlayer weekly over 13 weeks [2]. A random sample of 10,000 users that was reduced to 3,625 users to eliminate outliers may not be representative of the underlying population. Increasing the overall sample size may give us more meaningful clusters in another analysis as well us enable our clusters to be less sensitive to outliers skewing the data. This may also allow us to get more similar results from the PAM method. Further, due to the nature of how the sample was drawn, the sample was not random for three out of four analysed months. This, again, may distort the picture and should be accounted for. Lastly, testing the trained k-means model on another subset of the population would allow us to better validate the model and derive conclusions on the underlying population.

Despite usage behaviour being reduced to viewing time, timings and genres, these variables alone may be too simplistic and ignorant of other influential factors. For example, demographics may be highly indicative of viewing behaviour when it comes to online streaming services. The age, gender as well as the location of a user may give us more insight into the iPlayer's customer segments as well as how and where to best target them. Lastly, in this analysis, we draw conclusions about iPlayer accounts rather than iPlayer users as one family or household that pays one TV license fee will use only one BBC iPlayer account. With the available data, we cannot distinguish the small children watching kids' shows in the afternoon from their teenage siblings watching reality TV on the weekend from the parents watching movies at night as all these sessions would fall into one user account. Here, the iPlayer could introduce sub-accounts per person into one household account to improve the quality of the data so that more meaningful clusters can be derived.

Significance. As its mission to "inform, educate and entertain" [3] their audiences, the BBC is less focused on financial profit as a performance indicator but rather how well the public is engaged in its content offering. Consequently, the BBC is particularly interested in understanding its customer segments and how it can drive their engagement and increase customer satisfaction. Using this data for predictions of customer behaviour or clustering of customer segments (as here!) allows the BBC to better "inform, educate and entertain.“

6. APPENDIX

Appendix 1

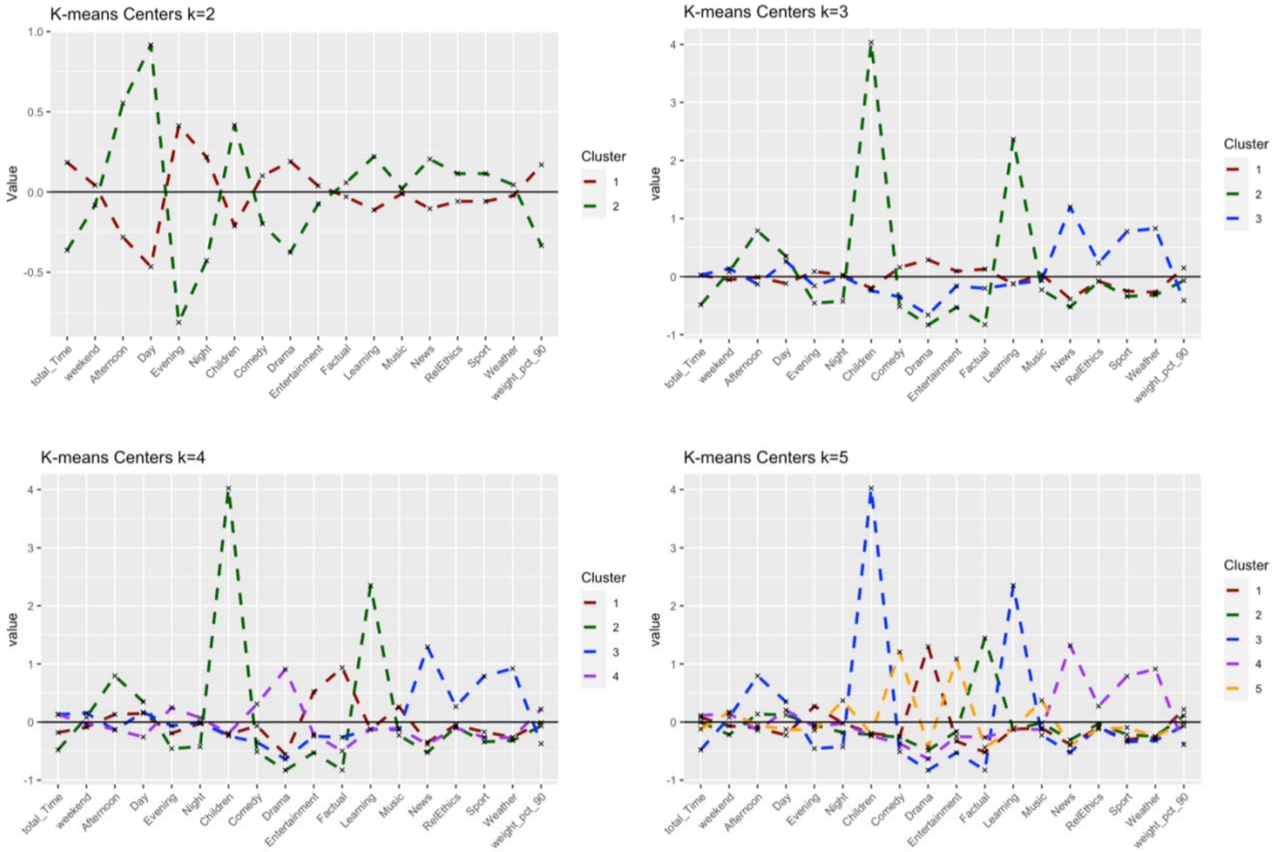
List of variables in the final data set used for the cluster analysis:

- **total_Time:** The total time watched by a user
- **weekend:** proportion of viewings on the weekend as compared to a weekday
- **Day:** proportion of viewings occurred between 6am and 2pm
- **Afternoon:** proportion of viewings occurred between 2pm and 5pm
- **Evening:** proportion of viewings occurred between 5pm and 10pm
- **Night:** proportion of viewings occurred between 10pm and 6am
- **Children:** proportion of shows viewed from the Children's genre
- **Comedy:** proportion of shows viewed from the Comedy genre
- **Drama:** proportion of shows viewed from the Drama genre
- **Entertainment:** proportion of shows viewed from the Entertainment genre
- **Factual:** proportion of shows viewed from the Factual genre
- **Learning:** proportion of shows viewed from the Learning genre
- **Music:** proportion of shows viewed from the Music genre
- **News:** proportion of shows viewed from the News genre
- **RelEthics:** proportion of shows viewed from the Religion and Ethics genre
- **Sport:** proportion of shows viewed from the Sports genre
- **Weather:** proportion of shows viewed from the Weather genre
- **weight_pct_90:** proportion of shows watched more than 90% (considered finished shows)

In the final data set, all variables are standardized and, therefore, the actual values do not represent the true observed value but just a relative scale.

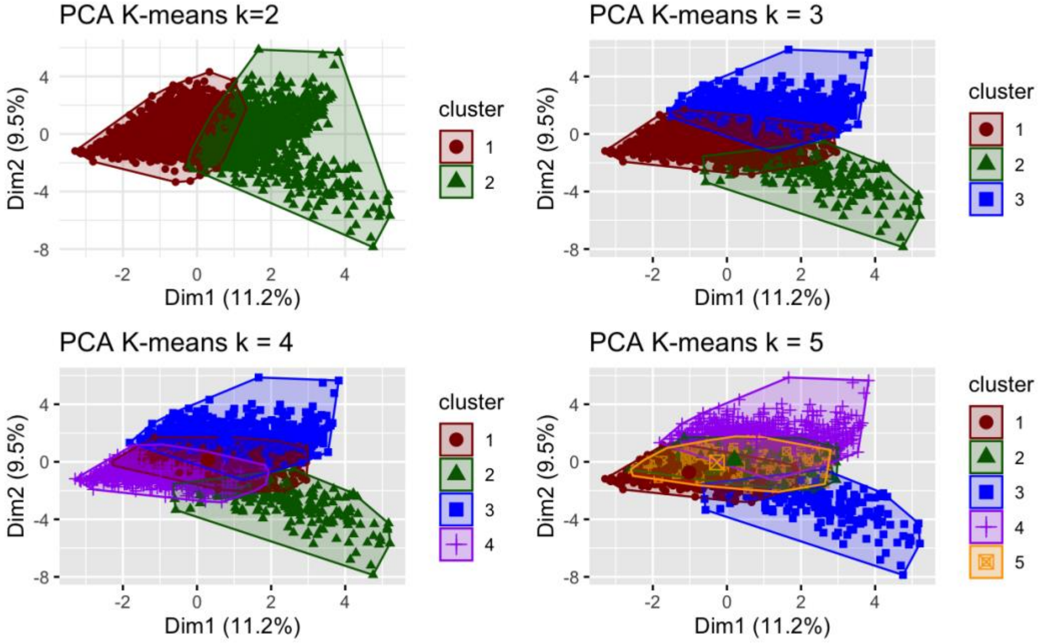
Appendix 2

K-means center plots for k = 2, 3, 4 & 5



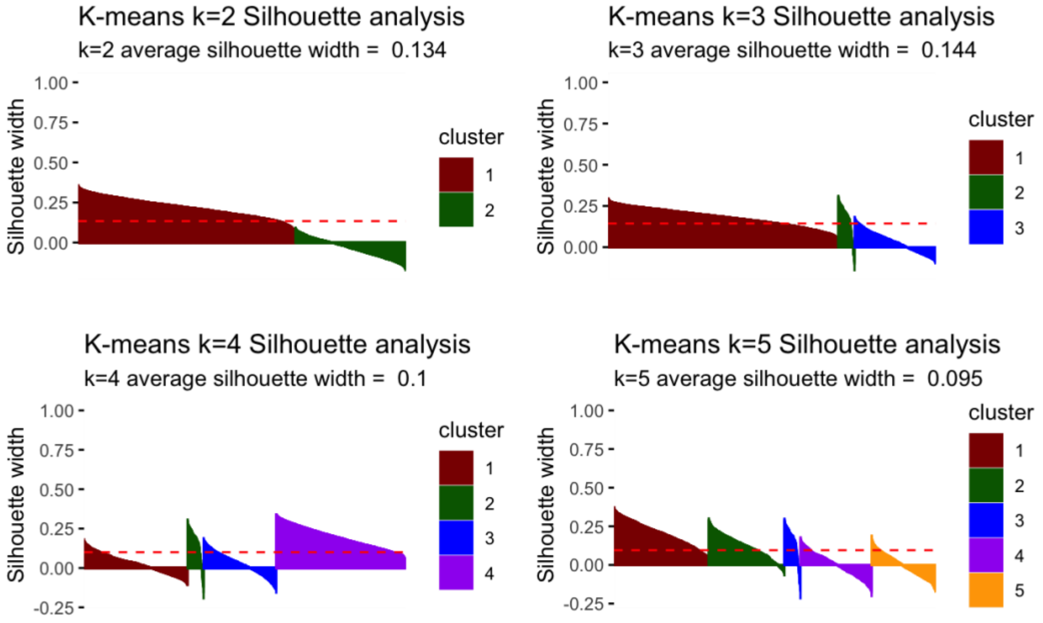
Appendix 3

K-means PCA for k = 2, 3, 4 & 5



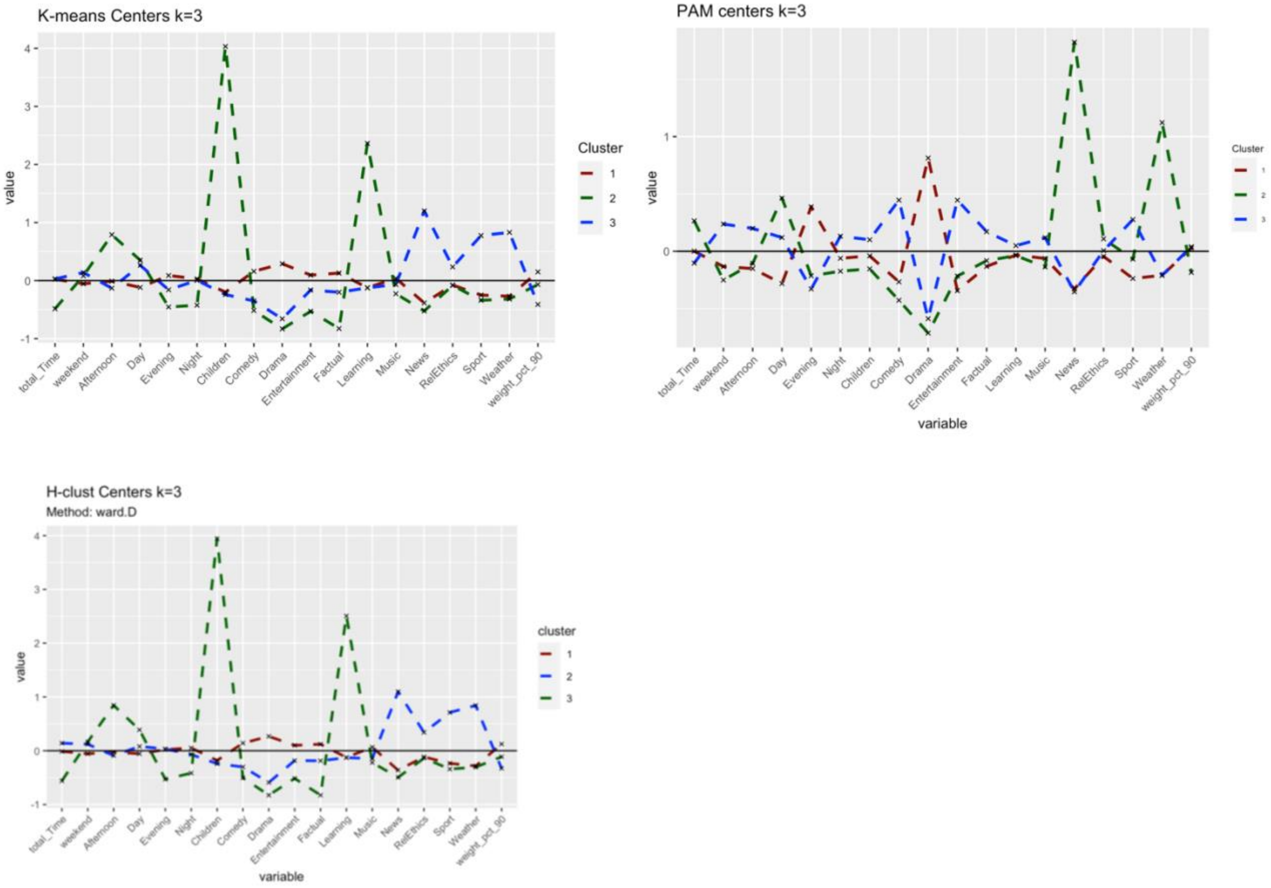
Appendix 4

K-means Silhouette analysis for k = 2, 3, 4 & 5



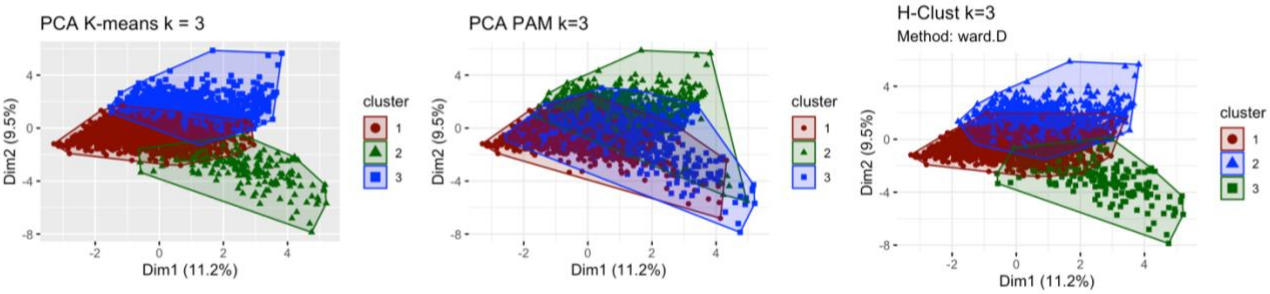
Appendix 5

Comparing center plots for methods k-means, PAM and H-clust



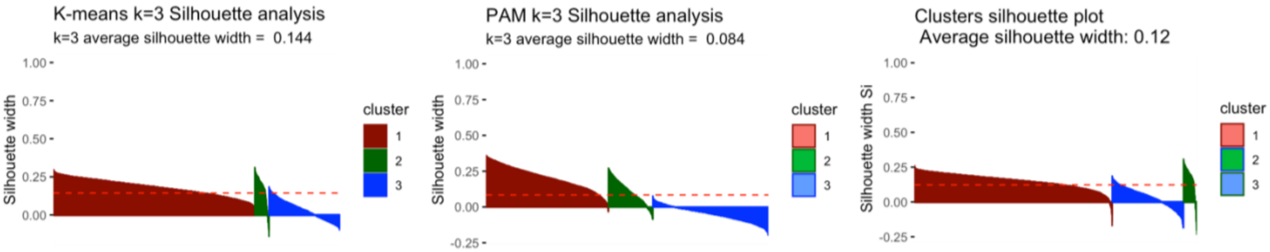
Appendix 6

Comparing PCA plots for methods k-means, PAM and H-clust



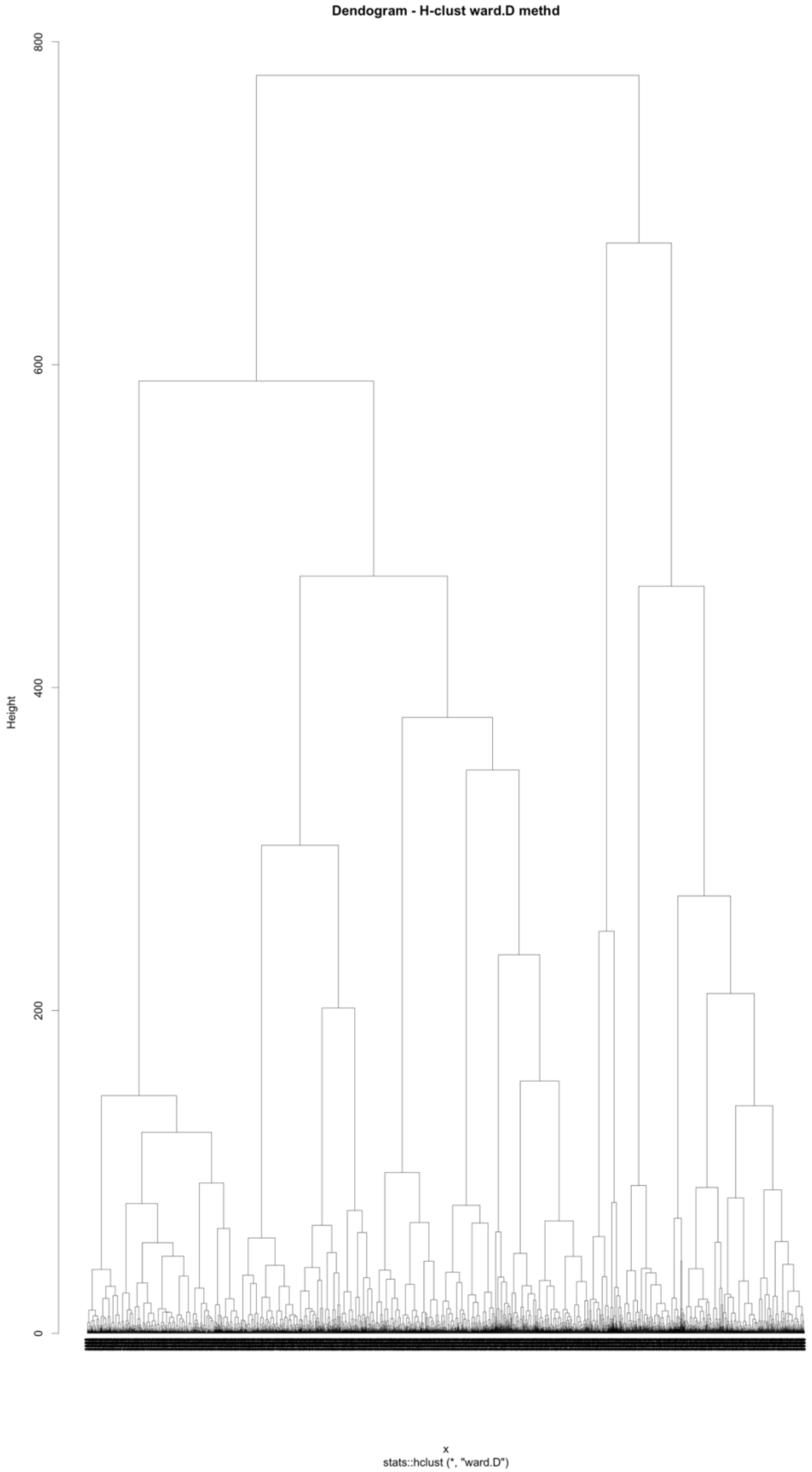
Appendix 7

Comparing Silhouette analysis plots for methods k-means, PAM and H-clust



<u>Appendix 8</u>

Dendogram from H-clust, visualising distances between clusters



Appendix 9

Rough comparison of different, significant cluster attributes over k-means, PAM and H-clust for k=3

		Method		
Cluster	**Attribute evaluation**	**K-means**	**PAM**	**H-clust (ward.D)**
1	Size	**2548** (70%)	**1571** (43%)	**2538** (70%)
	Favour	▪ **Drama** ▪ Factual ▪ Entertainment ▪ Evening ▪ weight_pct_90	▪ **Drama** ▪ **Evening**	▪ **Drama** ▪ Factual ▪ Entertainment ▪ Night ▪ weight_pct_90
	Adverse	▪ **News** ▪ **Weather** ▪ Sport ▪ Children ▪ Learning ▪ Day	▪ **News** ▪ **Day** ▪ **Entertainment**	▪ **News** ▪ **Weather** ▪ Children ▪ Learning
2	Size	**180** (5%)	**570** (16%)	**172** (5%)
	Favour	▪ **Children** ▪ **Learning** ▪ **Afternoon** ▪ Day	▪ **News** ▪ **Weather** ▪ total_Time ▪ Day	▪ **Children** ▪ **Learning** ▪ **Afternoon** ▪ Day
	Adverse	▪ All other genres (especially, **Drama, Factual, Entertainment, News, Comedy**) ▪ **total_Time** ▪ **Evening** ▪ **Night**	▪ **Drama** ▪ Comedy ▪ Children ▪ Weekend ▪ Evening ▪ weight_pct_90	▪ All other genres (especially, **Drama, Factual, Entertainment, News, Comedy**) ▪ **total_Time** ▪ **Evening** ▪ **Night**
3	Size	**897** (25%)	**1484** (41%)	**915** (25%)
	Favour	▪ **News** ▪ **Weather** ▪ **Sport** ▪ RelEthics ▪ Day	▪ **Comedy** ▪ **Entertainment** ▪ **Sport**	▪ **News** ▪ **Weather** ▪ **Sport** ▪ RelEthics
	Adverse	▪ **Drama** ▪ **weight_pct_90** ▪ **Comedy** ▪ Factual ▪ Entertainment	▪ **Drama** ▪ **News** ▪ **Evening** ▪ Weather	▪ **Drama** ▪ **weight_pct_90** ▪ **Comedy** ▪ Factual ▪ Entertainment

7. REFERENCES

[1] https://www.statista.com/topics/3836/the-bbc/
accessed: 13.11.2020

[2] https://www.statista.com/statistics/1042466/bbc-iplayer-visits-frequency-united-kingdom-uk/
accessed: 13.11.2020

[3] https://www.bbc.com/aboutthebbc/governance/mission
accessed: 13.11.2020