CLUSTER ANALYSIS FOR AVERAGE SCREEN TIME OF UNIVERSITY STUDENTS IN A WEEK

Jhansi priya   21MSD7001   CLASS ROLL:01

CONTRIBUTION:

NAME	REG.NO.	CLASS ROLL N0.	CONTRIBUTION
Jhansi Priya	21MSD7001	01	

Data Collection,Data Visualisation (bar plot for 7 days) and K mean (fitK1, first part average screen time for the week)

Kushi	21MSD7011	10	

Data Collection, K mean (fitK2, part two average screen time for all platforms), Documentation.

Sharanya D.V	21MSD7016	15	

Data Collection, Hierarchical clustering (part 1) and data visualization bar plot and histogram

Srivatsava Gade	21MSD7051	46	

Data Collection, Hierarchical clustering (part 2) and code documentation

CLUSTER ANALYSIS FOR AVERAGE SCREEN TIME OF UNIVERSITY STUDENTS IN A WEEK

ABSTRACT:

Screen time describe all the sedentary activity that happens in front of a screen accounting for all time spent in front of a screen, including mobile device, computer screen for work, watching movies and shows, or playing video games[1].

Background: The screen time as required by the present advanced 21st century education system and the influence of the COVID 19 pandemic has lead to an increase in screen time of university students.The aim in this study is to analyse the change in screen time at the present transitioning period from pandemic lockdown to gradual reopening of physical universities[2].

Methods: A survey to collect the screen time on various electronic devices for 52 university students of MSc. Data Science course in VIT-AP was conducted consider various platforms. This data is then analysed by clustering analysis based on the average screen time for the week and each platform separately.

Conclusion: There are distinct similarities in the platforms most time is spent and the number of students. Browsing and social media having the most number of minutes show clustering and the other platforms tend to mix depending on the student.

INTRODUCTION:

Screen time refers to activities such as watching TV, working on a computer, or playing video games that take place in front of a screen. Every age group is found to spend a significant amount of time on our various digital gadgets on a daily basis[3]. Smart phones and laptops being the warehouse of information and convenience, are a great source of information, communication, and every daily activity at the user's fingertip greatly increases screen time. Social media

is playing a huge role in substituting physical social interactions.The pandemic that started in 2019 has been a catalyst to this, with minimum activities for the people to perform at home, they restored to electronic devices like televisions, laptops and notably, mobile phones for communication, working and learning. The section that used minimal electronic devices turned to this for leisure, whereas the others, who already worked around these devices, added hours of work and leisure into electronic device usage.A study in nine European countries inferred that, approximately 65 percent of the 4,108 participants reported an increased screen time during the pandemic[1].Previous studies have shown that increased levels of screen time is proportional to various health concerns like depression, anxiety, sleep disorder, compromised eyesight and other non-communicable diseases. Furthermore, there is evidence that link increasing screen time progressively lower psychological well being with the rate of rise being twice more than that for someone with less screen time[4].

Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group than those in other groups. In simple words, the aim is to segregate groups with similar traits and assign them into clusters. The well known method K mean and Hierarchical clustering are used in the survey data set, K-means clustering is a type of unsupervised learning, which is used when you have data without defined categories or groups. The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K[5]. On the other hand, hierarchical clustering builds hierarchy of clusters. The algorithm starts with all the data points assigned to a cluster of their own. Then the two nearest clusters are merged into same cluster[6]. The algorithm terminates when there is only one single cluster is left, giving results in a dendrogram.
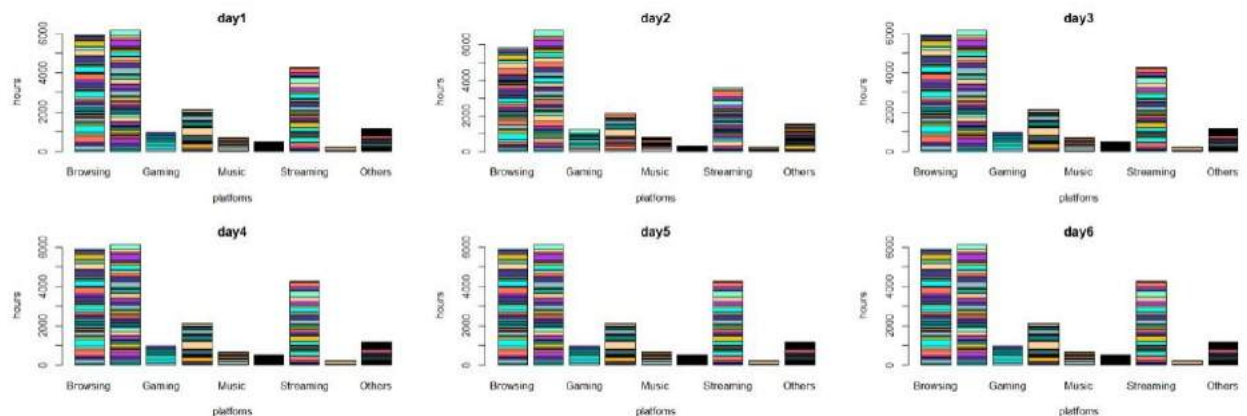
DESCRIPTION:
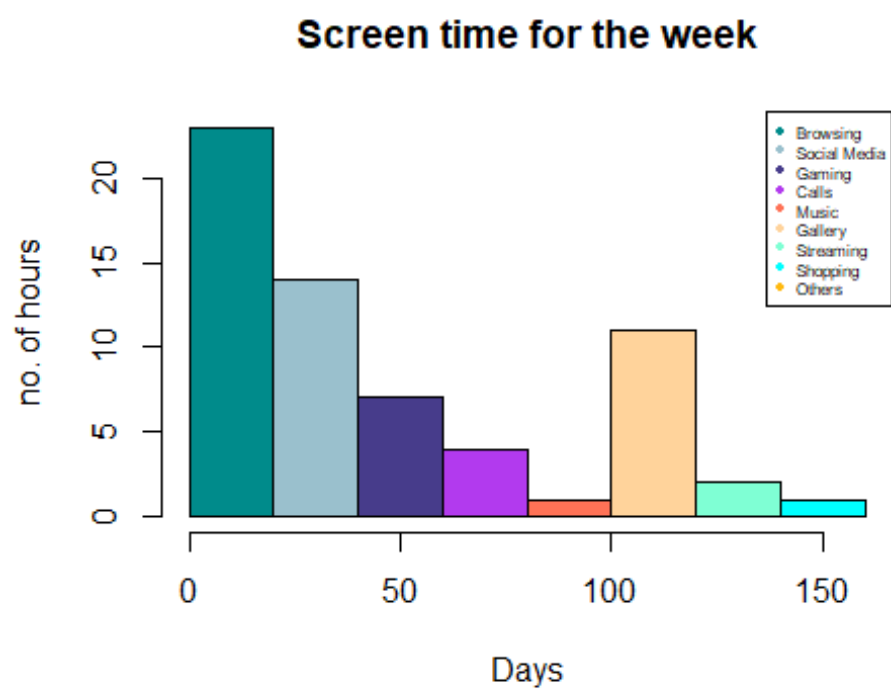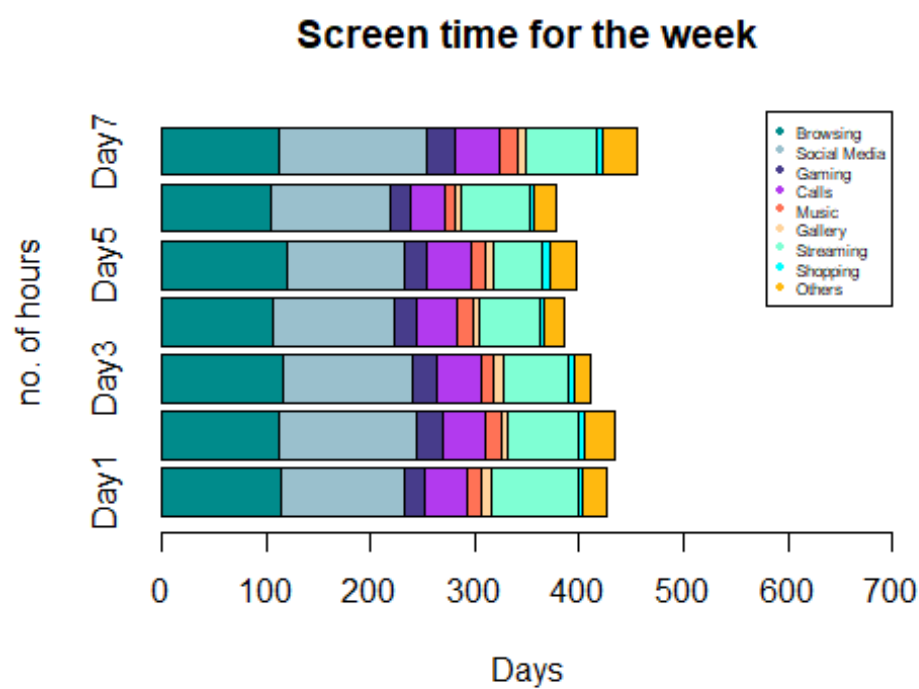
1.Data Collection:

The screen time for the students of MSc Data Science (2021-2023), VIT-AP a class of 52 student, is collected from their phones and laptops, for the time period between December 6,2021 and December 12,2021. The various platforms considered for this criteria were browsing, social media, streaming, gaming, music, video calls, shopping, gallery and others.The data was collected in minutes and two datasets including, average screen time on different platforms and average screen time for different days were made.

2.Data Visualisation:

The data is better understood by analysing the bar plots for screen time on each day distinctively, and the aggregate of the same is represented in bar plot and histogram.


Screen time on each day of the week.

Screen time for the week

Legend:
- Browsing
- Social Media
- Gaming
- Calls
- Music
- Gallery
- Streaming
- Shopping
- Others

Screen time for the week

Legend:
- Browsing
- Social Media
- Gaming
- Calls
- Music
- Gallery
- Streaming
- Shopping
- Others

3.Cluster Analysis:

3.1 Analysing average screen time in a week:

3.1.1 Data Visualisation and Normalisation:

The details for average screen time for seven day of a week based on the
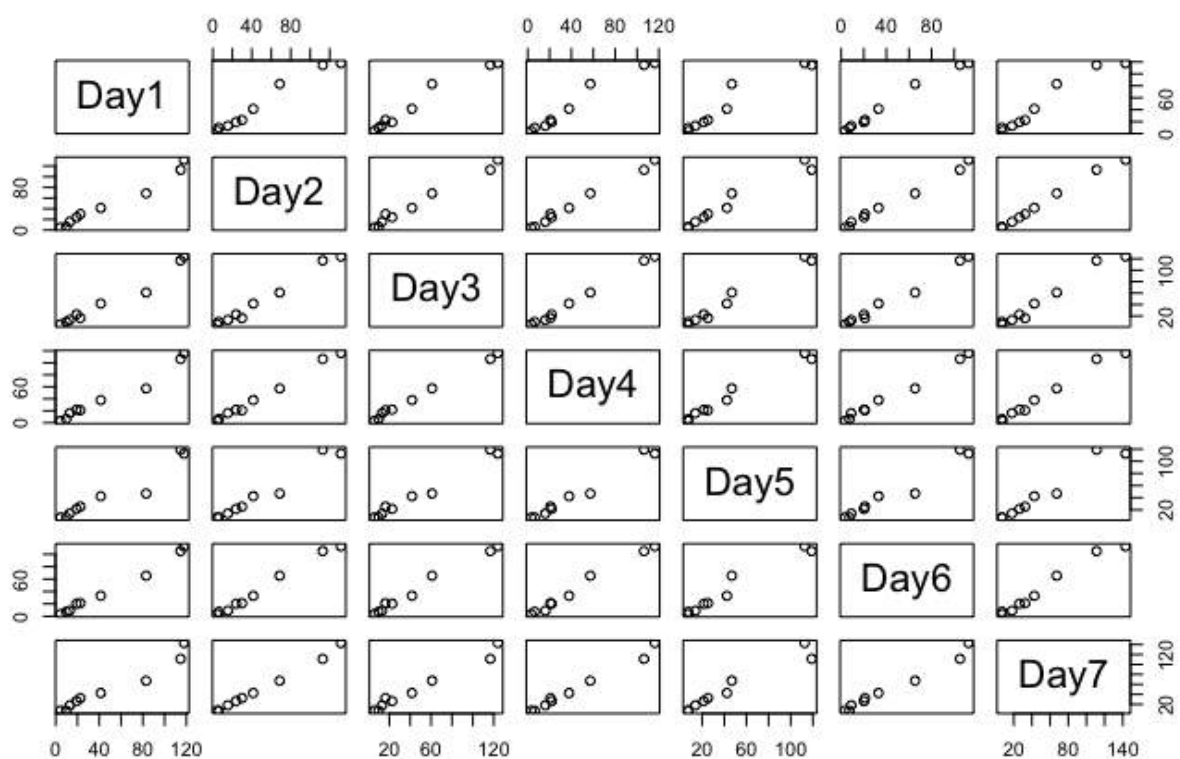
different platforms is understood and the scatter plot for all pairs are made.

Pair wise scatter plot for screen time on different days.

The data is normalised by scaling it and removing the unwanted column that hold the categories of platform and the distance matrix is obtained.

3.1.2 K Mean Clustering:

The optimal number of k is derived by within sum of squares method. According to plot, the best number of clusters is k=3 .
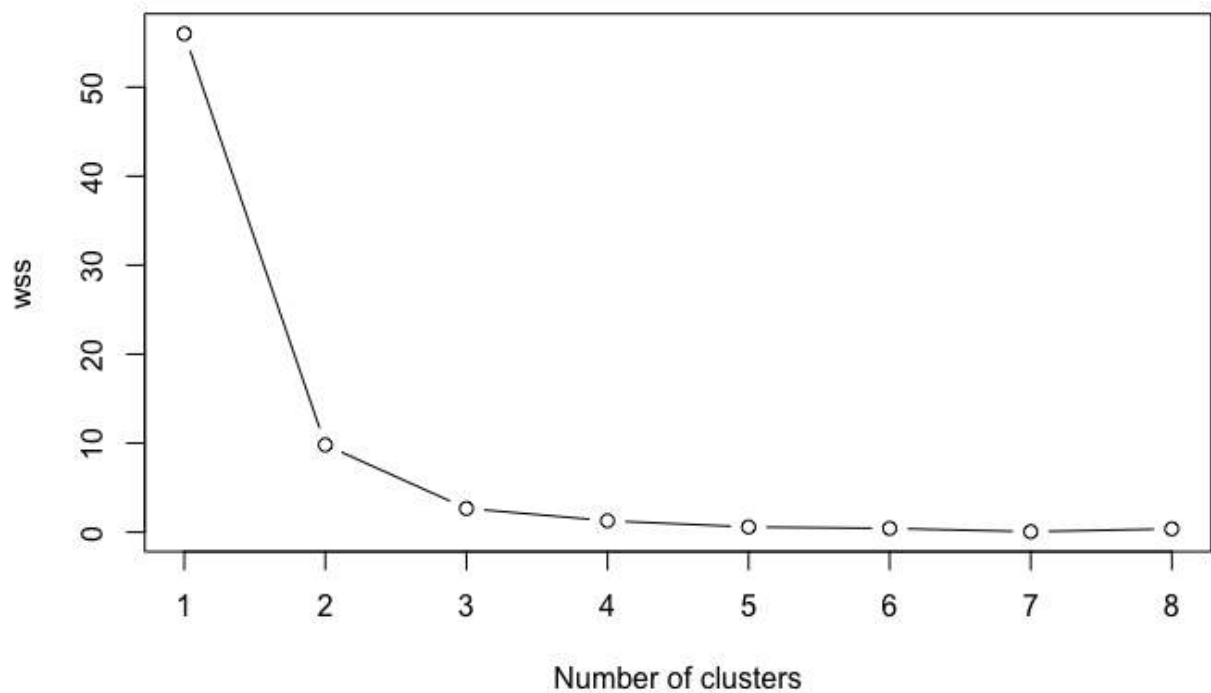

The k mean clustering for 3 clusters gives the between_SS / total_SS of 95.2 % for sum of squares within clusters and the aggregate mean for each of those

clusters is found. These are then plotted using fviz_cluster of facto extra library.
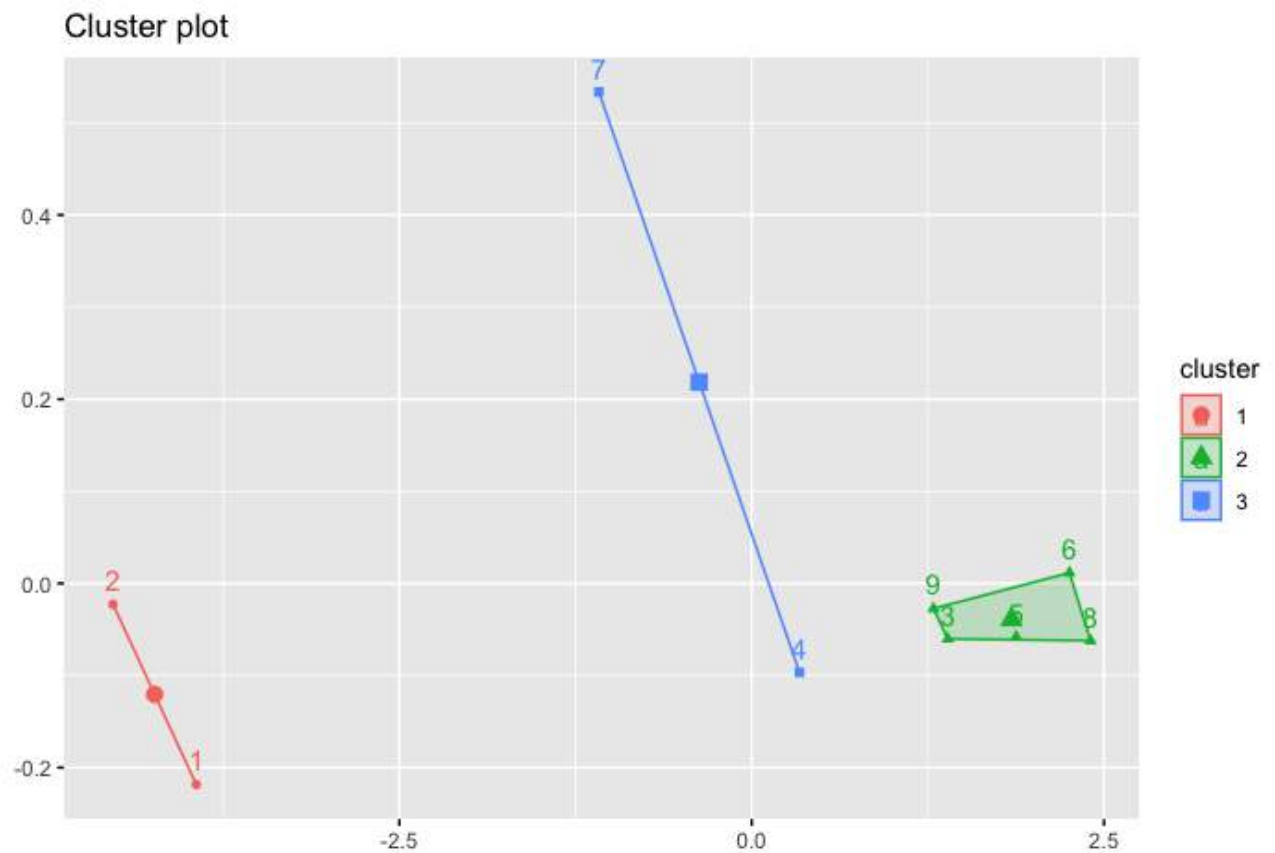
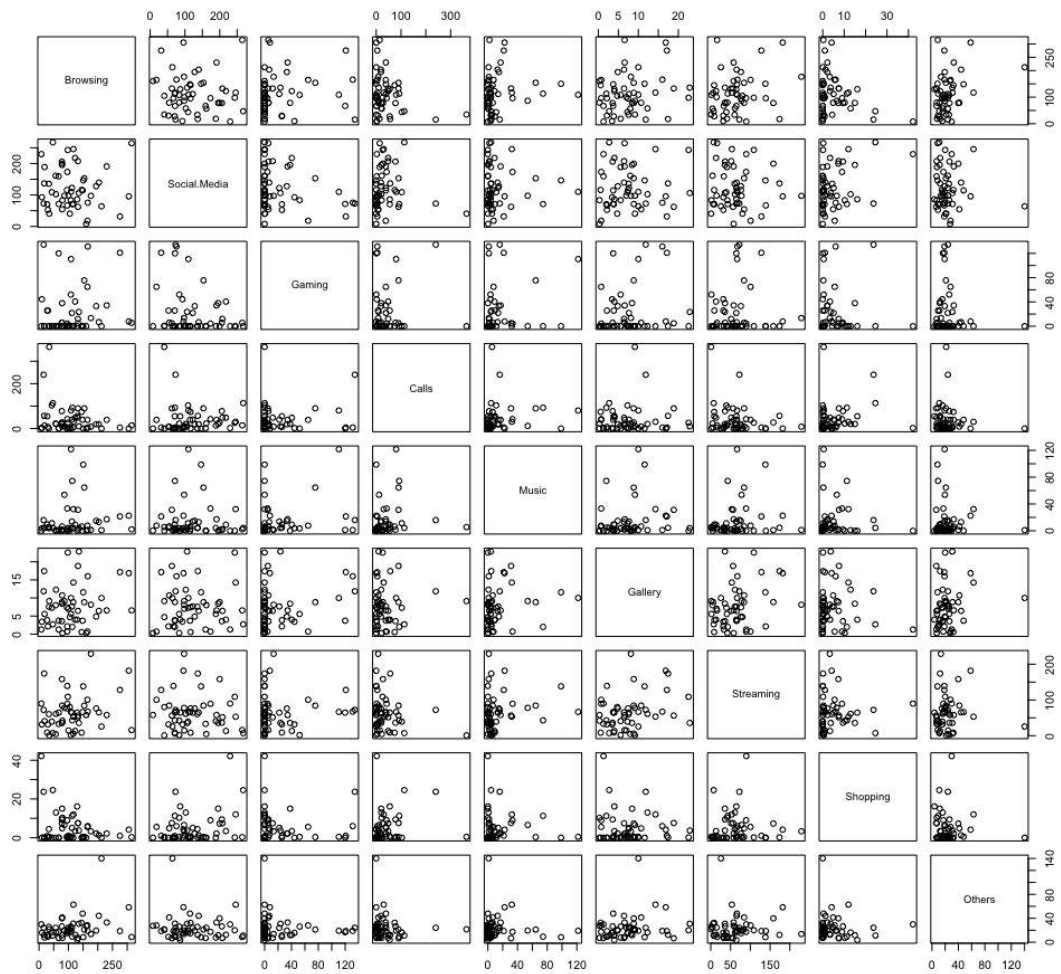

Cluster plot

3.2.3 Hierarchical Clustering Analysis:

The hierarchical clustering is performed using Ward's minimum variance method (ward.D2) on the distance matrix and the cluster dendrogram is plotted. The k=3 clusters are partitioned and marked on the same dendrogram.

Cluster Dendrogram

Distance matrix
hclust (*, "ward.D2")
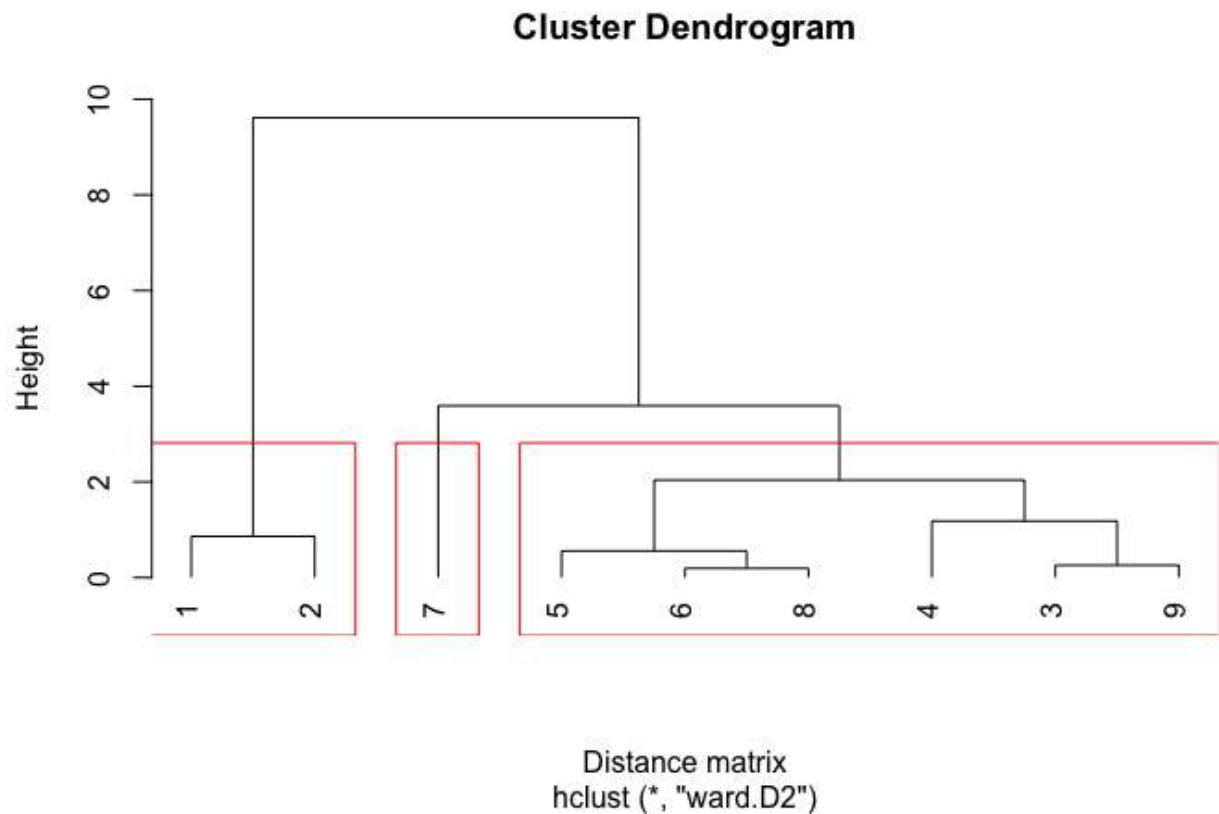
3.2 Analysing average screen time on different platforms for each student:

2.2.1 Data Visualisation and Normalisation:

The details for average screen time on different platforms for seven day of a week is understood and the scatter plot for all pairs are made.

Pair wise scatter plot for screen time on different platforms.

The data is normalised by scaling it and removing the unwanted column that hold the categories of days and the distance matrix is obtained.

3.2.2 K Mean Clustering:

The optimal number of k is derived by comparing 24 different indices provided by NbClust out of which, according to the majority rule, the best number of clusters is 8.

No.of Clusters for 24 Criteria
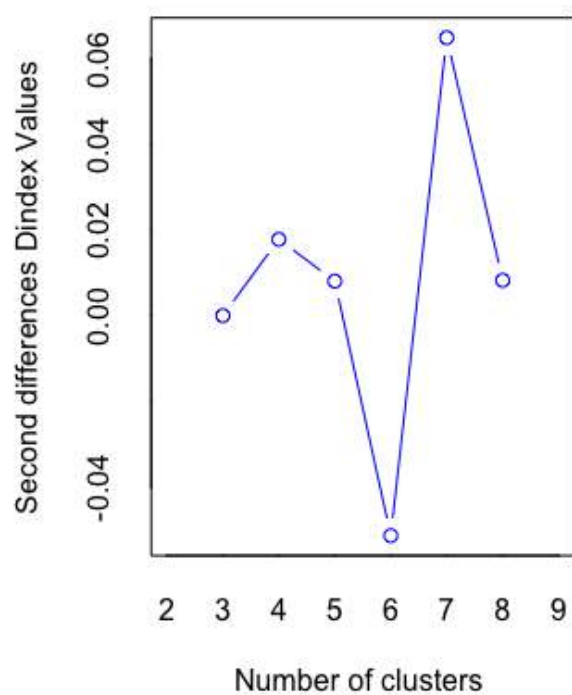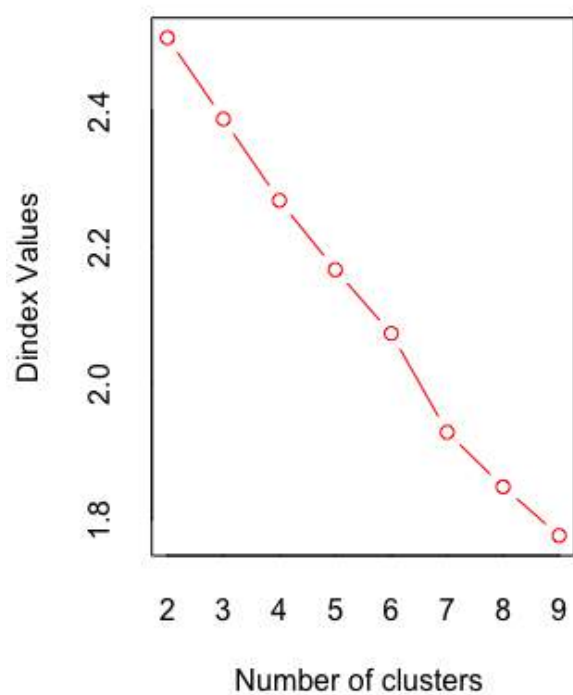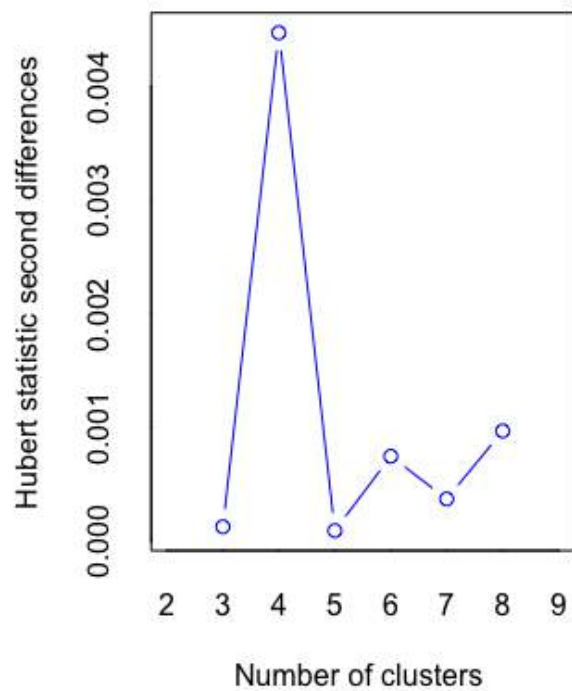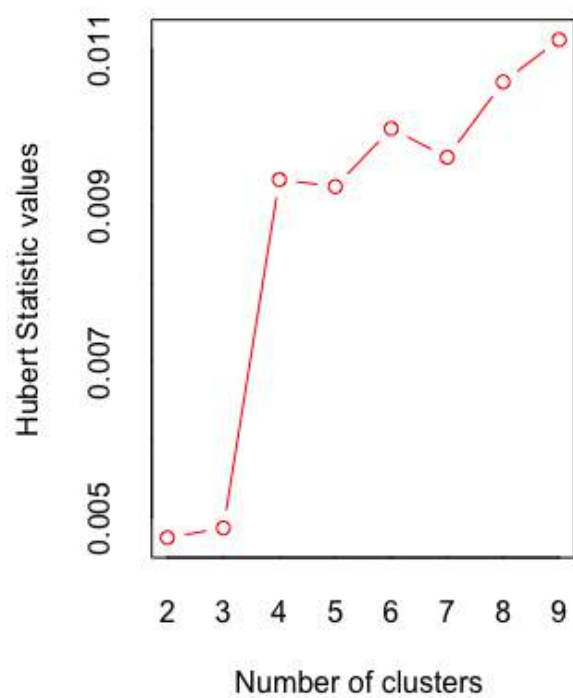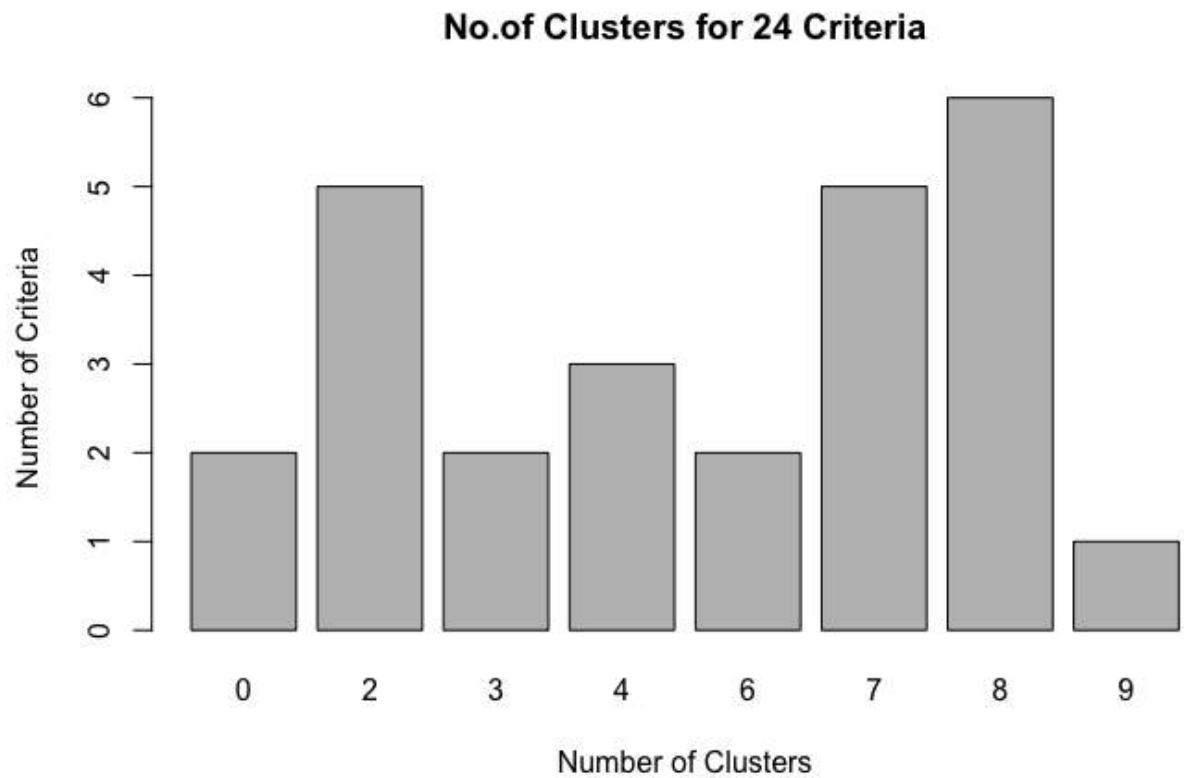
The k mean clustering for 8 clusters gives the between_SS / total_SS of 56.1 % for sum of squares within clusters and the aggregate mean for each of those clusters is found. These are then plotted using fviz_cluster of facto extra library.

Average screen time on different platforms
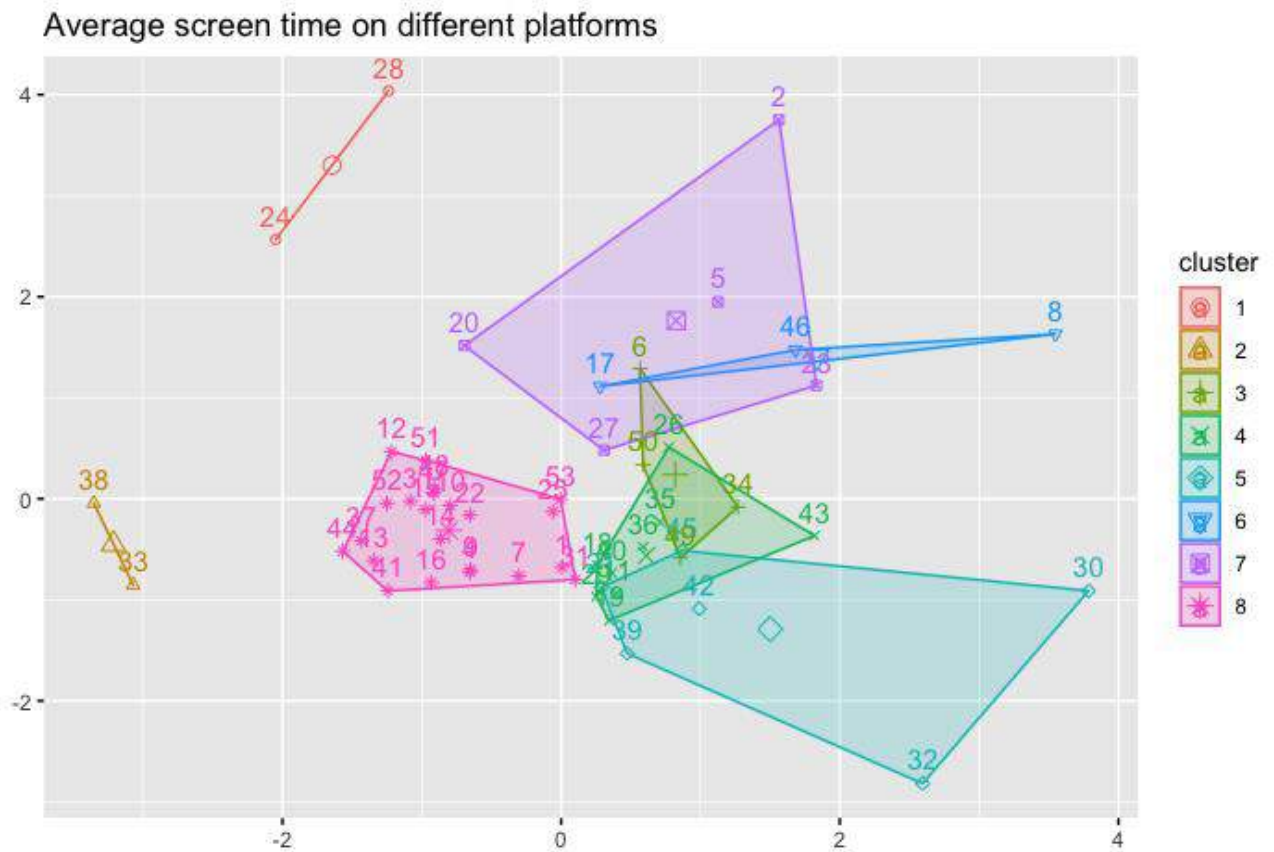
3.2.3 Hierarchical Clustering Analysis:

The hierarchical clustering is performed using Ward's minimum variance method (ward.D2) on the distance matrix and the cluster dendrogram is plotted. The k=8 clusters are partitioned and marked on the same dendrogram.

Cluster Dendrogram

Distance matrix
hclust (*, "ward.D2")
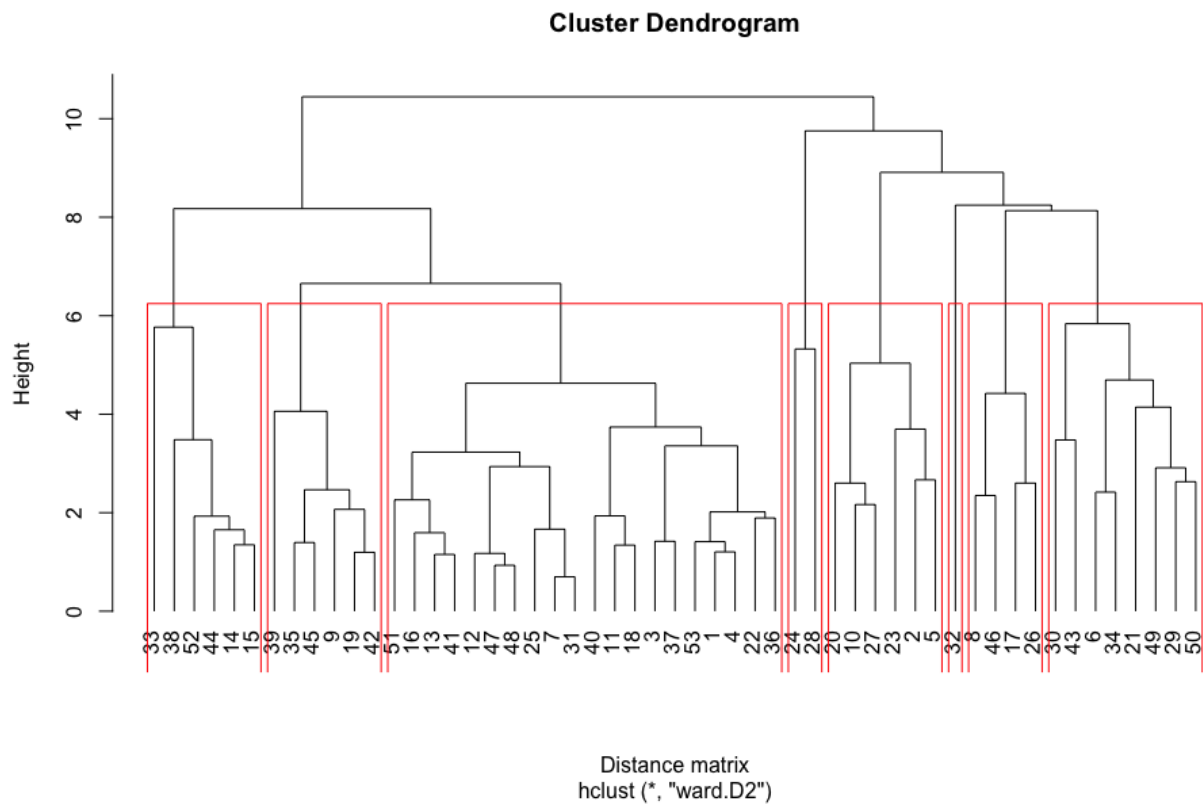
CONCLUSION:

The data when visualised, shows potential clusters and grouping on the bases of the different days and platforms. The plot for wss method for optimal k value for k mean analysis is interpreted using the elbow method that gives k=3, fitting three clusters of size 2,5,2 and between_SS / total_SS value of 95.2 %, suggesting this clustering to be a good fit. It is visibly noted that most students spend a similar amount of time of the week on gaming, music gallery, shopping and others, whereas they have the most screen time in a week on browsing and social media. The weekly screen time for most students is similar for calls and streaming. Hierarchical clustering gives weekly screen time results similar to the first cluster of size two having browsing and social media, whereas the second partition is of size one having platform as others only. The major partition of size 6 has all the other platforms for their screen time in a week.

The second dataset for average screen time based on each platform for every student through out the week is fit in 24 different criteria presented by NbClust out of which 6 proposed 8 as the best number of clusters. With k value as 8 clusters of sizes 2, 2, 4, 9, 6, 3, 5, 22 are obtained. The cluster has between cluster sum of square of 262.3651 and total sum of square of 468, giving between_SS / total_SS value as 56.1 %, suggesting a moderately good fit. The Hierarchical clustering forms 8 partitions at height 6.2 with the largest cluster having a frequency of 8 and the smallest cluster of size 1. This clustering is relatively similar to the k mean clustering, suggesting a good fit.

REFERENCES:

[1]Smartphone Screen Time Among University Students in Lebanon and Its Association With Insomnia, Bedtime Procrastination, and Body Mass Index During the COVID-19 Pandemic: A Cross-Sectional Study, Sajida Fawaz Hammoudi, *Hussein Walid Mreydem , Bayan Tarek Abou Ali, 1 Nada Omar Saleh, Seockhoon Chung, Souheil Hallit and Pascale Salameh.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8473859/

[2]Physical activity, screen time and obesity status in a nationally representative sample of Maltese youth with international comparisons,

Andrew Decelis, Russell Jago, and Kenneth R Fox
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4091762/

[3]The impact of COVID-19 on student equity and inclusion: Supporting vulnerable students during school closures and school re-openings
https://www.oecd.org/coronavirus/policy-responses/the-impact-of-covid-19-on-student-equity-and-inclusion-supporting-vulnerable-students-during-school-closures-and-school-re-openings-d593b5c8/#endnotea0z2

[4]Associations of Sleep Duration and Screen Time with Incidence of Overweight in European Children: The IDEFICS/I.Family Cohort

Guzmán V.a · Lissner L.b · Arvidsson L.c · Hebestreit A.d · Solea A.e · Lauria F.f · Kaprio J.g · Reisch L.A.h · Moreno L.i · Felső R.j · de Henauw S.k · Veidebaum T.l · Ahrens W.d · Hunsberger M.b · on behalf of the IDEFICS and I.Family consortium
https://www.karger.com/Article/FullText/519418

[5] https://www.statology.org/pairs-plots

[6] https://www.r-bloggers.com/2021/04/cluster-analysis-in-r/

[7] https://www.analyticsvidhya.com/blog/2016/11/an-introduction-to-clustering-and-different-methods-of-clustering/