CSCI.4923 Capstone in Interprofessional Informatics

An Insight into Smartphone Obsession and its Adverse Effects

Sachin Karki

Texas Woman's University

Denton, Texas

Author Note

Sachin Karki, Texas Woman's University, Mathematics & Computer Science

Table of Contents

Problem Identification	**3**
History and Basics of Smartphone	3
Smartphone Usage and Engagement Trends	4-6
Smartphone Addiction: Myth or Fact?	6
Generally Speculated Effects of Smartphone Addiction	7-8
Research Questions	8-10
Literature Review	**9**
Search Process	9
Findings	9
Table 1: *Literature Review Articles*	10-11
Methods	**12**
Methods and Procedures	12-13
Data Analysis	**14**
Statistical Analysis (Using R)	14
a. Analysis of Activity.xlsx Dataset	14-23
b. Analysis of Effects.xlsx Dataset	23-29
Visual Analysis (Using Tableau)	29
a. Analysis of Scores.xlsx Dataset	29-32
Interpreting Analysis to Answer Research Questions	**33-34**
Appendix A: Limitation and Potential of the Analysis	**35-36**
Appendix B: Comparing and Contrasting with Primary Research	**36-37**
Appendix C: Conclusion	**38**
References	**39-40**

Problem Identification

As the world is booming and awakening with a surge in smart devices and technology every day, smartphones have become an essential part of human life. A smartphone is an invention of the modern-day that does everything a phone can and much more. Smartphones have ushered in a whole new era of global communication and connectivity. The purpose of this paper is to identify and explore the problem of smartphone addiction and some of the adverse effects it brings. The argument, analysis, and interpretation presented in the paper are meant for readers to grasp a sufficient level of understanding of this problem.

History and Basics of Smartphones

Cellphones have been around for a while. The Motorola DynaTAC 8000X was the first cellphone that was commercially available in 1983. Less than a decade later, the world saw the "IBM SIMON Personal Communicator" in 1992 which became the world's first-ever smartphone. It has been almost 30 years since smartphones have become a household name and as a trivia, there are more smartphones in the world than people. Smartphones are the modern iteration of traditional cellphones that does minor to major communicating tasks and possibly even more. A typical standard smartphone is capable enough to make to call, send a text, write an email, browse web pages, take pictures, shoot videos, and load a wide range of applications, etc. The smartphone is a portable device that one can take anywhere and do anything a phone can do and beyond.

A smartphone generally has a hefty price tag attached to it when it comes to a certain brand and set of features. For an instance, a new iPhone SE has a price tag of $399 for the entry tier model. Other smartphones are available on the market as well that are cheap and do the same thing as an iPhone does but with few restrictions and limited features, but the general idea is that smartphones can be budget-friendly which is what makes them universally affordable and accessible.

Smartphone Usage and Engagement Trends

At this point, the general global smartphone engagement is beyond expectation. A report from PewResearch.org suggests that as of 2020 approximately 5 billion peoples have the access to smartphones and this number is projected to significantly increase within coming years. In the United States, there are more than 290 million smartphone users (Statista, 2021). This number is expected to see exponential growth over the spawn of the next 5 years or so. A graph showing the projection of estimated smartphone users in the United States by 2025 is illustrated below.

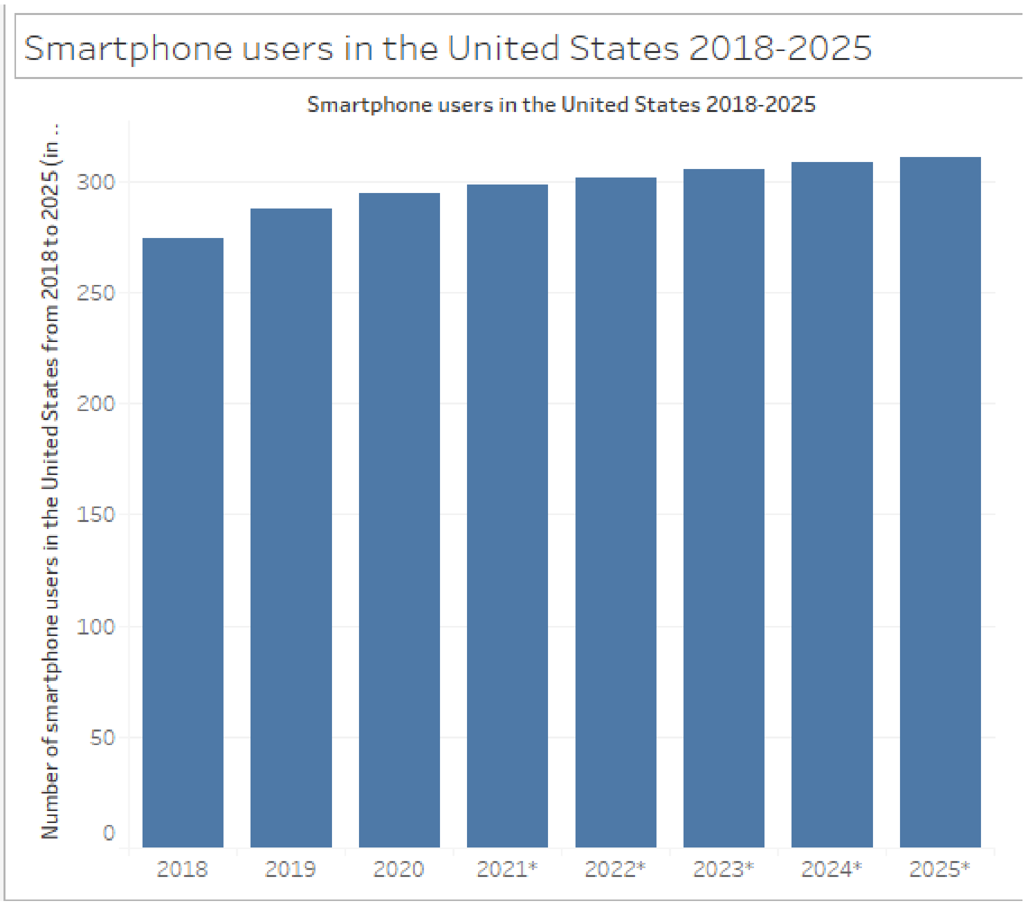

Fig 1: Bar Graph of Smartphone Users in the US by 2025.

Source: Statista

The illustration depicted above is a visual portrayal of data retrieved from Statista which sums up the increment in the number of smartphone users from the year 2018 and how it will be by the year 2025. By 2025 the number of smartphone users is expected to reach 311 million.

Let's retell this fact from a global perspective. Presented below is the bar graph of the estimated billions of smartphone users worldwide by 2025.

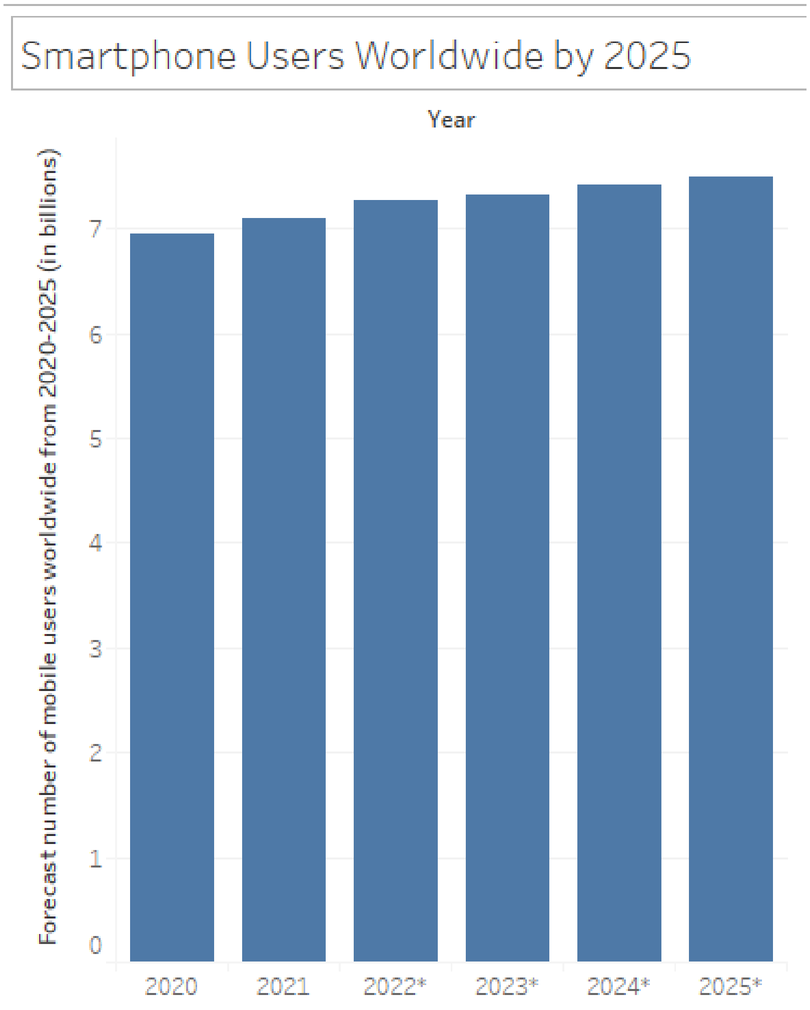

Fig 2: Bar Graph of Smartphone Users Worldwide by 2025.

Source: Statista

The illustrated graph above visually summarizes the data from Statista that predicts the number of smartphone users in the next 4 years. As we can observe that the estimated number of smartphone users worldwide will reach around 7.49 billion by the year 2025. As of 2021, we have around 6.9 billion users. Mathematically, an estimated 8.55% growth within the next 4 years and not to mention that growth represents the population in billion. That indicates a plethora of smartphones in years to come.

Smartphone Addiction: Myth or Fact?

Technological advancements and the invention of smartphones have enabled people to be in touch. Thanks to evolving technology which rose the global connectivity to a promising level. As illustrated in Figure 2, there are about 7.1 billion smartphone users in the world as of 2021. Global smartphone ownership and usage are indeed an exciting insight to look forward to, but the engagement of the smartphone, however, can tell a different story. Smartphone addiction unlike a drug-related addiction is an extreme dependency or fixation with the very devices we use. The problem is not with many people owning a smartphone device, it is with the people being obsessed with these devices to the point it actively as well as passively affects their health. The trend of using smartphones worldwide is increasing day by day. As the technology is rapidly evolving and constantly improving, new ways, features, and functionalities in smartphones are being introduced which adds another reason to own and use a smartphone. These technological changes have led to a revision in the very definition of addiction for it not only refers to drug or substance abuse, but now also includes behavioral addictions such as gambling, internet gaming, or even excessive smartphone use (Boumosleh & Jaalouk, 2017). All the prior research, study, survey, and personal experience of people that we have based our arguments, analysis, and findings upon tells us that the phenomenon of smartphone addiction is very real and has severe effects. These chains of effects have proven to affect an individual in areas of interpersonal habits, mental health, and quality of life. We will analyze datasets to test different variables that explores the nuances of smartphone addiction

Generally Speculated Effects of Smartphone Addiction

The general definition of addiction is to compulsively commit to a habit or a practice. This definition also applies very much in the context of smartphone addiction since an addicted individual spends a majority of their time using a smartphone that triggers detrimental effects. Our initial search, collection of firm evidence, and conclusion based on elaborated studies tell us that smartphone addiction can cause effects that can be distressing, distracting, and depressing. Developers of RescueTime, an iOS, and Android app claim that an average person checks their phone 58 times a day. With that claim, the only way one could be labeled as a smartphone addict is when they continuously use their phone beyond what is considered average. From the preliminary readings of various research papers, journals, and online articles, it was found that smartphone addiction indeed has physical, emotional, mental, psychological, and social effects.

Former research conducted among 106 university students tried to explore the association between sleep quality and late-night smartphone usage and its long-term impact on mental health. The research paper titled ***"The Effects of Smartphone Usage on University Student"*** by authors *Nalini Arumugam, Sivajothy Selvanayagam, Sai Tarishini Sathiyasenan* concluded that late-night smartphone use does have a significant impact on sleep quality. University students' excessive smartphone use during the night has negative effects on their well-being. They are unable to concentrate during lectures the following day, often feeling sleepy during the day, and face difficulties handling the day to day pressures, as well as, thinking rationally whenever problems occur (Arumugam et al., 2020).

A cross-sectional study titled "***Depression, anxiety, and smartphone addiction in university students - A cross sectional study***" conducted by authors *Jocelyne Matar Boumoslehm and Doris Jaalouk* explored the relationship of smartphone addiction with different interpersonal habits and two personality types concluded that addictive smartphone usage pattern leads to a severe decline in quality of life and induces anxiety and various other psychological changes like functional impairment and compulsive behavior.

Another research paper, ***A review on the effects of smartphone usage on attention, inhibition, and working memory*** (Magnus 2020) did an empirical analysis on the effects of smartphone use and found that psychological aspects like memory and ability to pay attention to detail are indeed affected by prolonged use of smartphones.

To sum up, all the prior research that was gathered to serve as a source and the base for our analysis have concluded that smartphone addiction does have the potential to leave a significant impact on a user.

Research Questions

As mentioned earlier, this paper aims at exploring the nuances of smartphone addiction. Using various academic sources, supporting research and studies, and a retrieved dataset, this paper will conduct a new analysis to explore, analyze, and interpret the engagement with smartphone addiction. To be specific our research will focus on answering the following questions.

1) What personal habits (i.e. smoking, alcohol drinking, and religious practices) candidates have that were susceptible to excessive smartphone usage?
2) For what activities (i.e. texting, calling, reading news, entertainment, other reasons) of smartphones led to excessive usage?
3) Is there any relationship between different effects (i.e. decrease in sleep quality, ability to stop worrying, not being able to enjoy a meal) and personality types (i.e. Type A and B)?
4) Which personality type (i.e. Type A or B) is more vulnerable to the effects (i.e. anxiety,

compulsive behavior, functional impairment) of excessive smartphone usage?

The procedure and methods of the dataset creation and modification will be described below.

Literature Review

Analysis of this scope and extent requires various supporting materials. These materials can range from previously proven claims, foundational datasets, strong evidence, and preliminary research and studies. To guide our analysis and research on its right track, several instances of resources were used. Academic journals, research papers, and peer reviews were the main contributing sources to our analysis.

Search Process

Several online resources were used to locate the articles, research papers, and datasets related to smartphone addiction. Different academic portals, news outlets, dataset search engines were utilized to gather and collect a strong set of evidence and information. EBSCOHOST, ScienceDirect.com, Kaggle.com, ResearchGate, PewResearch.org, Statista.com, Data.gov were the key sources to conduct or study. The top five articles are depicted in Table 1.

Findings

The major five articles that were chosen concluded with a common notion that smartphone addiction is very real and is happening as there is a global scale increment in smartphone usage. Smartphone addiction indeed has brought negative health effects among users who tend to spend prolonged periods using their smartphones. These effects are targeted in aspects like personal, mental, emotional, psychological, and quality of life.

Table 1

Literature Review Articles

1. Title **2. First Author** **3. Date** **4. Country**	**5. Data information** **6. Sample Size**	**7. Aspect of Interest** **8. Design (Experiment, observation, analysis, etc.)** **9. Level of Evidence (e.g., meta-analysis, correlation, opinion)**	**10. Comparison of Interest**	**11. Outcome of Interest** **12. Results of Study** **13. Conclusion**	**14. Strengths** **15. Limitations**
A. 1. Depression, anxiety, and smartphone addiction in university students- A cross sectional study 2. Boumosleh 3. 2017 4. Lebanon	5. Notre Dame University (Lebanon) Students 6. 688 undergraduate students	7. Interpretation and exploration of smartphone addiction 8. Analysis 9. Correlation	10. Statistical analysis of a collected dataset	11. Prevalence of smartphone addiction symptoms was substantial among the sample of university students. 12. Identified the problem 13. Smartphone raised stress level among university students	14. The assessment was validated by university students 15. Possible inaccuracies and biases because of self-reported independent variables.
B. 1. The Effects of Smartphone Usage on University Students. 2. Arumugam 3. 2020 4. Malaysia	5. University Students 6. 106 tertiary level students	7. Generate insights into the negative association of excessive smartphone use with the psychological well-being among university students 8. Observation 9. Analysis	10. Descriptive quantitative analysis of effects of smartphone usage	11. Excessive smartphone use has negative effects on the well-being 12. Time spent on smartphones late night affects students' quality of sleep 13. Late night smartphone usage affects students state of mind the following day	14. not listed 15. not listed

C. 1. Smartphones and attention, curse, or blessing? - A review on the effects of smartphone usage on attention, inhibition, and working memory 2. Magnus 3. 2020 4. Germany	5. General smartphone users 6. not listed	7. Claims that smartphone usage may also have beneficial effects on certain processes of attention 8. Observation 9. Opinion	10. Review of different effects of smartphone usage.	11. Effects of smartphone use identified as either a curse or blessing. 12. It's unclear whether effects of smartphone usage affect attentional domains. 13. not listed	14. not listed 15. no argument or evidence for cognitive aspects
D. 1. Problematic smartphone use in a large nationally representative sample: Age, reporting biases, and technology concerns 2. Horwood 3. 2020 4. Australia	5. Australian adults 6. 1164 adults	7.Highlights that age has a linear effect on smartphone use 8. Analysis 9. Correlation	10. Analysis of a collected dataset	11. The trend of excessive smartphone use is high in ages between 18 – 40 and it declines after that. 12. People can estimate their amount of smartphone use. 13. Excessive smartphone use is a subjective experience	14. not listed 15. Data was collected during the corona virus pandemic so it might affect the generalizability of results.
E. 1. Evaluation of mobile phone addiction level and sleep quality in university students 2. Sevil 3. 2013 4. Turkey	5. Students of the Sakarya University 6. 576 University Students	7. Explores the link between smartphone addiction and sleep quality. 8. Analysis 9. Correlation	10. Interpretation of the relationship between excessive smartphone use and sleep quality.	11. Prolonged period of smartphone use indeed related to sleep quality among students. 12. not listed 13. The sleep quality worsens with increasing addiction level	14. not listed 15. Impossibility to establish definitive diagnosis with used scale since the data was collected from only one university.

Methods

Methods and Procedures

After the dataset was acquired, major modifications were made before proceeding with the analysis. The original dataset (Boumosleh & Jaalouk, 2017) was huge with its content as it was raw in its original state. The number of observations (n) was 688 with 26 relating variables/columns. Variables consisted of aspects of socio-demographic, academic, personality trait, lifestyle, and smartphone related factors like usage.

With a total consciousness of the possibility, extent, and scope of the raw dataset, changes and adjustments were made to make it possible and compatible to analyze and extract insights out of them. Three new datasets were created with a relevant set of variables that had the potential to answer our question. First, the questions were defined, and relevant variables were studied to create new datasets. After careful considerations, the main dataset was split into three other datasets where each dataset had variables relating to the activity, personal habits of participants, and their individual score of "Smartphone Addiction Inventory Scale (SPAI)". The SPAI was one of the main focuses of the primary research (Boumosleh & Jaalouk, 2017) which used independent variables like depression/anxiety, age, personality type, and class to generate a scale of measurement. The summative score of SPAI was used for data visualization presented in this paper.

Three datasets, *Activity.xlsx, Effects.xlsx,* and *Scores.xlsx* were created from the original dataset to streamline the process for methodical analysis. The *Activity.xlsx* consisted of variables relating to personal habits (i.e. smoking, alcohol drink, and religious practices) and activity of smartphone (i.e. texting, calling, entertainment, etc.) The *Effects.xlsx* included variables relating to effects that were induced on participants after excessive smartphone usage (i.e. decreased sleep time, not being able to stop worry, etc.) And lastly, *Scores.xlsx* consisted of variables relating to SPAI score on different variables (i.e. anxiety, depression, functional impairment etc.)

This research was mainly focused on bringing results from two types of analysis: statistical and visual. For the statistical part, a multi-layered approach of analysis was used to ensure the accuracy and validity of the results. Procedures like variable dependency check with correlation matrix, mean comparison with Cohen's D, variable selection with Akaike Information Criterion (AIC), and association between dependent and independent variables with multinomial regression were used. The *Activity.xlsx* dataset was used to answer questions 1 and 2 while the *Effects.xlsx* dataset was used to answer question 3. The *Score.xlsx* was used to answer question 4.

One minor and crucial change made to the *Scores.xlsx* was the redefining of two values in the personality type variable. There are two types of personality listed in the dataset (i.e. 0 and 1) Type 0 was categorized as easy-going, laid-back more relaxed, and patient meanwhile Type 1 were categorized as aggressive, competitive, angry, cynical, mistrustful. To make it easy to interpret and comprehend the results the personality type 0 was changed to type A and type 1 was changed to B. In addition to this, to effectively generate the visuals the datatype variables were configured and changed from categorical to a numeric whole number. Few variables were trerated as dimensions in tableau to achieve a right and appealing set of bar graphs. No other changes were administered in the datasets.

Data Analysis *(Using R and Tableau)*

As mentioned before, two types of analysis were used to answer the research questions, i.e. statistical and visual. The statistical analyses were done with R programming language in RStudio meanwhile the visual analyses were done with the help of the visualization tool Tableau. The *Activity.xlsx* and *Effects.xlsx* were used for the statistical analysis while *Score.xlsx* was used for visual analysis. The methods applied and the resulting outcome of each analysis are discussed below with reference to each question they answered.

Statistical Analysis

The *Activity.xlsx* and *Effects.xlsx* were used to answer our research questions 1, 2, and 3. The breakdown of each question and their analysis is further discussed down below.

a. Analysis of Activity.xlsx (To answer questions 1 and 2)

First and foremost a correlation matrix was generated to check the dependency between the variables. It is only after we know for sure if a relationship exists between the variables, we can proceed further. At this point the dataset has not been filtered out for any missing data in the dataset so the empty values were treated as the default "Not Available (NA)" by default R procedure to generate a correlation matrix.

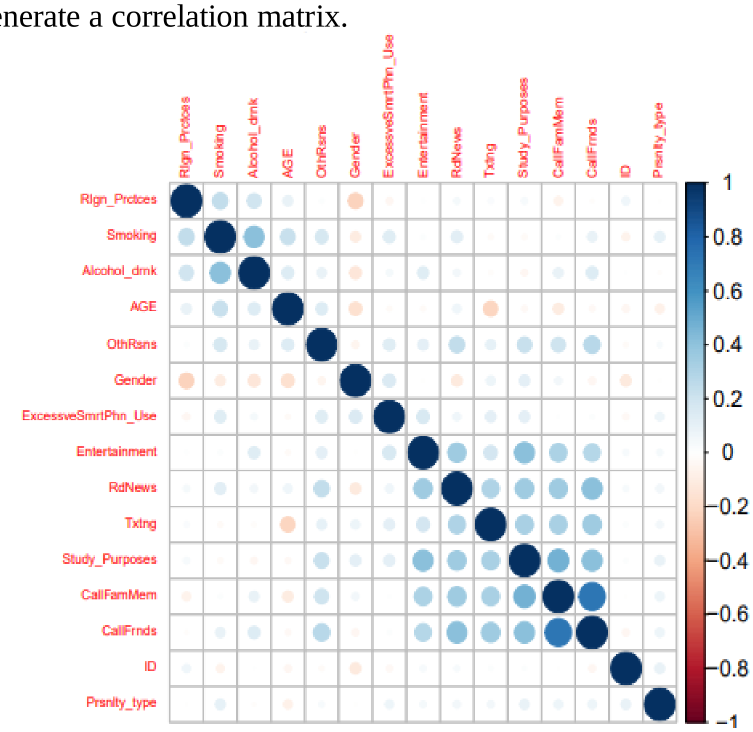

Fig 3: Correlation matrix of activity.xlsx dataset

The given plot above is the summary of correlation coefficients of a different set of variables. It can be noticed that variables "CallFamMem" and "CallFrnds" have a strong positive relationship. The visible and noticeable intensity of the blue dot tells us that they share strong positive relations and logically speaking they attribute somewhat similar activity of smartphone use. Calling friends and calling family members can be considered as a single variable of just calling in general. The rest of the variables however display a curious pattern so we will proceed with a few more steps to extract conclusions. The code for the correlation matrix is depicted below.

```
# read xlsx file # "." will be treated as na
data_activity <- gdata::read.xls("Activity.xlsx", na.strings = c("."))

# removing all na's
data_activity_na_rm <- data_activity %>% na.omit
#Plotting a Correlation
corrplot( cor(data_activity_na_rm),  order = 'hclust',tl.cex =0.5,number.cex=3,title="Activity with NA removed", mar=c(0,0,1,0)) # plotting correlation plot
```

This can be further elaborated by the Spearman coefficient score depicted below

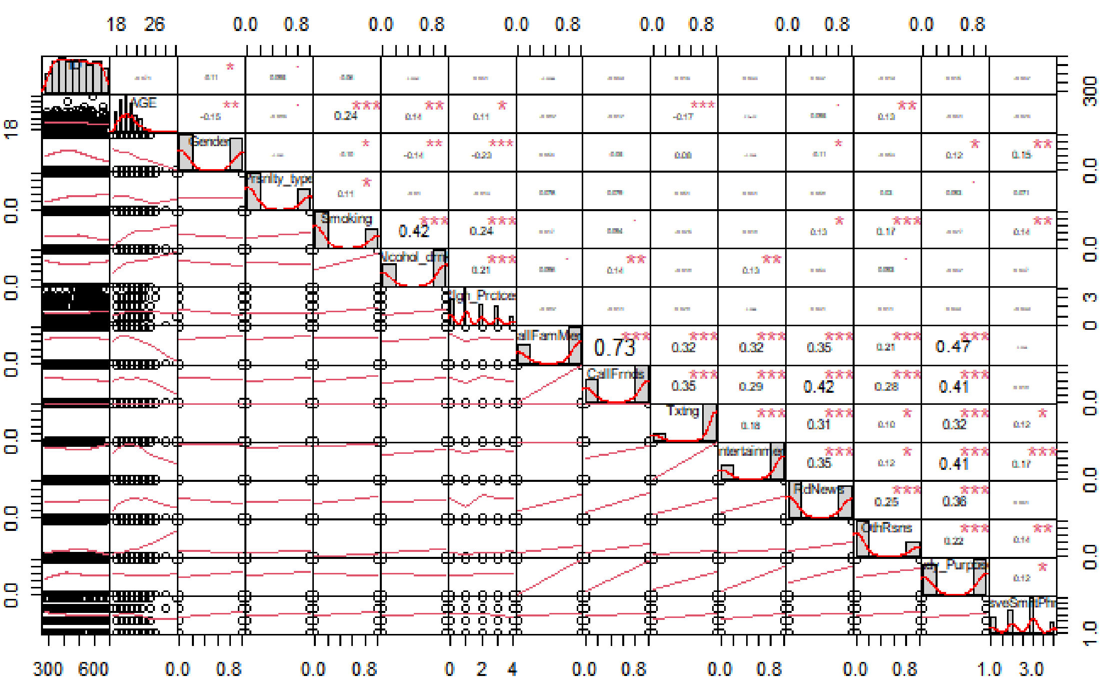

Fig 4: Correlation graph of activity.xlsx dataset

From the correlation graph above, we can notice that variables "CallFamMem" and "CallFrnds" have a Spearman score of 0.73 which represents an above-average number on a scale of 1 so statistically speaking these two variables do not have to be included further for more testing. This was generated using a basic *tidyverse* package in R and the code is depicted below.

```
```{r echo=FALSE,warning=FALSE}
correlation_graph <- chart.Correlation(data_activity_na_rm,
                  method="spearman",
                  histogram=TRUE,
                 pch=32,exact=FALSE)
```
```

In addition to this, our dataset has missing data as well. The number of participants choosing not to respond on a certain topic is very prominent and as a result, we have variables with simply no input. Missing data should always be statistically identified and filtered out to avoid high bias in the results. Using the *ggplot* package in R, missing values were identified, and here's a visual look at the number of missing data.

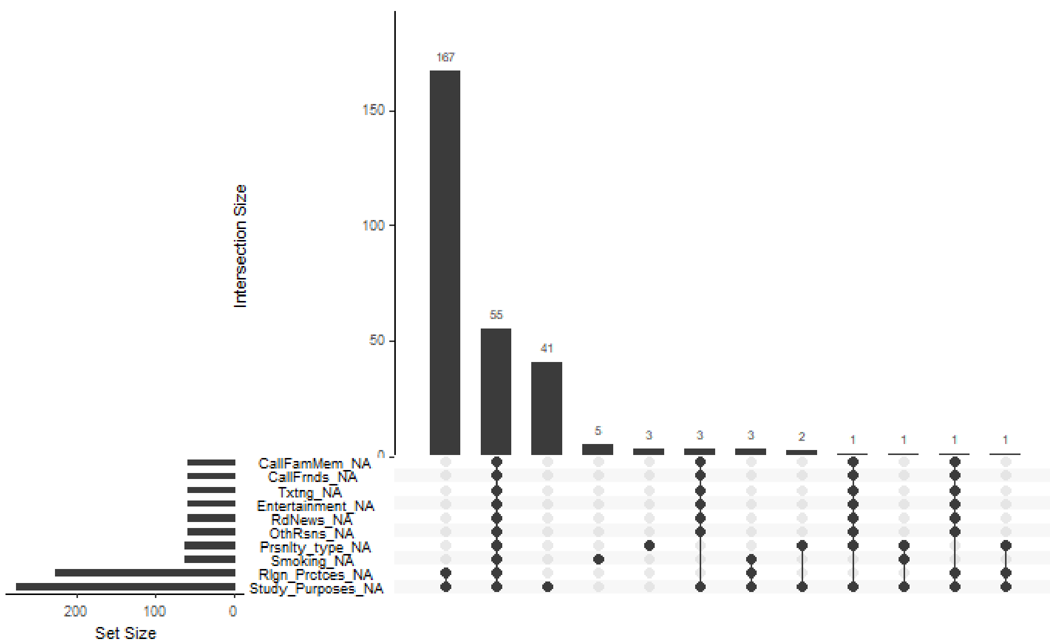

Fig 5: Bar graph of missing data in activity.xlsx datasets

From the graph above it can be verified that the variables like "Rlgn_Prctces" and "Study_Purposes" have the high count of missing data. "Rlgn_Prctces" has a total of 227 and

"Study_Purposes" has a total of 275 missing data. Thus, they will be excluded from the analysis since they are likely to cause bias.

The second step of the analysis is to do a mean comparison with Cohen's D. This will help us determine the effect size between our variables.

```
```{r echo=FALSE}
data_activity_clean <- data_activity %>%
                       mutate(Gender =as.factor(Gender),
                              Prsnlty_type = as.factor(Prsnlty_type),
                              Smoking=as.factor(Smoking),
                              Alcohol_drnk=as.factor(Alcohol_drnk),
                              CallFamMem=as.factor(CallFamMem),
                              CallFrnds=as.factor(CallFrnds),
                              Txtng=as.factor(Txtng),
                              Entertainment=as.factor(Entertainment),
                              RdNews=as.factor(RdNews),
                              OthRsns=as.factor(OthRsns),
                              ExcessveSmrtPhn_Use = as.factor(ExcessveSmrtPhn_Use)) %>%
                       na.omit
Myvariables <- c("AGE" ,  "ExcessveSmrtPhn_Use","Gender"   ,"Smoking" ,"Alcohol_drnk", "Txtng","CallFamMem","CallFrnds" ,"Entertainment" ,
"RdNews" ,"OthRsns")
cateVars <- c("Gender"   ,"Smoking" ,"Alcohol_drnk", "Txtng","CallFamMem","CallFrnds" ,"Entertainment" ,  "RdNews" ,"OthRsns"  )

univarante_1 <- tableone::CreateTableOne(strata = "Prsnlty_type", data = data_activity_clean, vars= Myvariables , factorVars = cateVars ) %>%
               print(noSpaces = TRUE, printToggle = FALSE, exact = "stage",smd = TRUE,quote = FALSE)
univarante_1
```
```

```
                         Stratified by Prsnlty_type
                          0               1               p       test SMD
  n                       "244"           "142"           ""      ""   ""
  AGE (mean (SD))         "20.57 (1.87)"  "20.27 (1.81)"  "0.132" ""   "0.160"
  ExcessveSmrtPhn_Use (%) ""              ""              "0.394" ""   "0.184"
     1                    "48 (19.7)"     "20 (14.1)"     ""      ""   ""
     2                    "69 (28.3)"     "37 (26.1)"     ""      ""   ""
     3                    "95 (38.9)"     "66 (46.5)"     ""      ""   ""
     4                    "32 (13.1)"     "19 (13.4)"     ""      ""   ""
  Gender = 1 (%)          "115 (47.1)"    "66 (46.5)"     "0.986" ""   "0.013"
  Smoking = 1 (%)         "76 (31.1)"     "59 (41.5)"     "0.050" ""   "0.218"
  Alcohol_drnk = 1 (%)    "152 (62.3)"    "87 (61.3)"     "0.927" ""   "0.021"
  Txtng = 1 (%)           "202 (82.8)"    "123 (86.6)"    "0.395" ""   "0.107"
  CallFamMem = 1 (%)      "155 (63.5)"    "101 (71.1)"    "0.158" ""   "0.163"
  CallFrnds = 1 (%)       "142 (58.2)"    "94 (66.2)"     "0.148" ""   "0.166"
  Entertainment = 1 (%)   "171 (70.1)"    "105 (73.9)"    "0.488" ""   "0.086"
  RdNews = 1 (%)          "109 (44.7)"    "72 (50.7)"     "0.299" ""   "0.121"
  OthRsns = 1 (%)         "67 (27.5)"     "43 (30.3)"     "0.634" ""   "0.062"
```

Fig 6: Summary of standard mean difference with Cohen's D by personality type

Cohen's D helps measure the size effect of variables and accompanies additional tests like t-test and ANOVA. A value of 0.2 is considered a small effect size in Cohen's D and from our result above the variable "smoking" has a standard mean difference of 0.2 in personality type. This means that the variable "smoking" is statistically significant in both personality types.

The third step of the analysis is to proceed with AIC with a backward algorithm to select variables and reduce the error rate. AIC is an estimator of prediction error so the lower the AIC model is, the better it is for our analysis. The variable "Smoking" was further tested alongside with an irrelevant variable "AGE" and a biased variable "Study_Purposes" to check the validity of the variable "Smoking". Variables like "ID" and "CallFrnds" were excluded since "ID" is not needed to answer our question and "CallFrnds" had a high correlation. The overall process returned the AIC of 495.90 from the original value of 507.83. The code and results are picture and marked below.

````
```{r  echo=FALSE}
data_activity_for_selection <- data_activity_clean %>% dplyr::select(-ID,-CallFrnds ) %>% na.omit() # excluding not interested var.
model_1 <- glm(Prsnlty_type~.,data=data_activity_for_selection , family = binomial ) # Logistic regression will be applied as our interested variable (personal type has two values)
model_1_null <- glm(Prsnlty_type~1,data=data_activity_for_selection , family = binomial ) # null model, only with intercept
model_01_selection <- stepAIC(model_1,direction="backward",trace = FALSE) # backward stepwise algorithm =  method for var. selection
model_01_selected <- glm(Prsnlty_type ~ AGE + Smoking + Study_Purposes ,data = data_activity_for_selection, # after stepwise AGE + Smoking + Study_Purposes are selected
                  family = binomial(link="logit"))
model_01_selected %>% parameters::parameters(exponentiate = TRUE, df_method = "wald",summary = FALSE) # getting parameters, exponentiate is used of log odds
anova(model_01_selected,
      model_1_null,
      test="Chisq")
```
````

```
Parameter      | Odds Ratio |   SE |        95% CI |     z |     p
------------------------------------------------------------------
(Intercept)    |       4.61 | 5.80 | [0.39, 54.38] |  1.21 | 0.225
AGE            |       0.88 | 0.05 | [0.78,  1.00] | -1.98 | 0.048
Smoking [1]    |       1.78 | 0.41 | [1.14,  2.80] |  2.52 | 0.012
Study_Purposes |       1.50 | 0.33 | [0.98,  2.29] |  1.85 | 0.064
Analysis of Deviance Table

Model 1: Prsnlty_type ~ AGE + Smoking + Study_Purposes
Model 2: Prsnlty_type ~ 1
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       382     495.90
2       385     507.83 -3  -11.937 0.007603 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Fig 7: AIC with a backward algorithm

The final model above suggests that a one-point increment in age will give us a 0.88 odd ratio (OR) increase in both personality types. Smoking is the strongest factor with an odds ratio of 1.78. Each result was accompanied by a p-value of 0.05 significance.

Furthermore, multinomial regression was used on the variable "ExcessveSmrtPhn_Use" as it has four levels of response denoting the general admittance of excessive smartphone usage by the participants. Here's the index of the scale used for four levels of response.

1 = strongly disagree, 2 = somewhat disagree, 3 = somewhat agree, 4 = strongly agree,

To conduct regression, variables like, "ID", "Study_Purposes", and "Rlgn_Prctces", are excluded, because "ID' is irrelevant and "Study_Purposes" and "Rlgn_Prctces" has a high count of missing data. This helps in avoiding biased results.

Multinomial regression requires comparing with a reference in order to test the outcome of different variables. In our case, all response levels i.e. 1, 2, 3, and 4 will be used as a reference to compare with other response levels since response levels 1 and 4 indicate strong denial and admittance respectively of participants with their excessive smartphone usage.

Here's a code and the result from response level 1.

```
```{r  echo=FALSE}
ggplot(data_activity_for_selection, aes(ExcessveSmrtPhn_Use)) + geom_bar() + ggtitle("Checking ExcessveSmrtPhn_Use's distribution
")+ theme_bw() # general overview with smartphone use through data

data_activity_clean$ExcessveSmrtPhn_Use_2 <- relevel( data_activity_clean$ExcessveSmrtPhn_Use , ref= 1) # for multinom
regression, we need reference to compare, here level 2 will be used for comparison to other levels
model_2 <- nnet::multinom(ExcessveSmrtPhn_Use_2~ AGE+ Prsnlty_type + Gender + Smoking  + Txtng +
    Entertainment+RdNews + OthRsns,data=dplyr::select(data_activity_clean, -ID,-Study_Purposes,-Rlgn_Prctces) ) # selecting
something needed
model_2 %>% parameters::parameters(exponentiate = TRUE,df_method = "wald",summary = FALSE) # getting paramet
```
```

Response level: 2

Parameter	Odds Ratio	SE	95% CI	t(359)	p
(Intercept)	1.97	3.83	[0.04, 88.05]	0.35	0.726
AGE	0.96	0.08	[0.80, 1.14]	-0.51	0.608
Prsnlty_type [1]	1.24	0.43	[0.63, 2.43]	0.62	0.536
Gender [1]	2.45	0.85	[1.25, 4.82]	2.59	0.009
Smoking [1]	1.54	0.56	[0.75, 3.16]	1.17	0.241
Txtng [1]	0.74	0.32	[0.32, 1.72]	-0.70	0.485
Entertainment [1]	1.48	0.52	[0.74, 2.96]	1.12	0.265
RdNews [1]	1.18	0.44	[0.57, 2.44]	0.44	0.660
OthRsns [1]	1.50	0.60	[0.68, 3.29]	1.01	0.311

Response level: 3

Parameter	Odds Ratio	SE	95% CI	t(359)	p
(Intercept)	3.76	7.15	[0.09, 156.51]	0.70	0.486
AGE	0.90	0.08	[0.76, 1.06]	-1.26	0.209
Prsnlty_type [1]	1.53	0.50	[0.81, 2.91]	1.30	0.192
Gender [1]	3.29	1.08	[1.72, 6.28]	3.61	< .001
Smoking [1]	1.92	0.68	[0.96, 3.82]	1.84	0.065
Txtng [1]	1.26	0.55	[0.54, 2.95]	0.53	0.598
Entertainment [1]	2.79	0.97	[1.41, 5.51]	2.95	0.003
RdNews [1]	0.82	0.29	[0.41, 1.65]	-0.55	0.584
OthRsns [1]	1.73	0.66	[0.81, 3.67]	1.42	0.156

Fig 8: Multinomial regression on response levels 2 and 3 at 0.05 significance level

Response level: 4

Parameter	Odds Ratio	SE	95% CI	t(359)	p
(Intercept)	0.04	0.10	[0.00, 4.48]	-1.34	0.181
AGE	1.03	0.11	[0.84, 1.27]	0.29	0.774
Prsnlty_type [1]	1.24	0.51	[0.55, 2.79]	0.52	0.605
Gender [1]	3.08	1.27	[1.37, 6.92]	2.73	0.006
Smoking [1]	3.03	1.29	[1.32, 6.99]	2.60	0.009
Txtng [1]	1.85	1.16	[0.54, 6.32]	0.98	0.325
Entertainment [1]	2.46	1.15	[0.99, 6.14]	1.93	0.054
RdNews [1]	0.91	0.41	[0.38, 2.19]	-0.21	0.831
OthRsns [1]	2.64	1.19	[1.09, 6.37]	2.16	0.031

Fig 9: Multinomial regression on response level 4 at 0.05 significance level

Here's the code and outcome from response level 2

```{r echo=FALSE}
ggplot(data_activity_for_selection, aes(ExcessveSmrtPhn_Use)) + geom_bar() + ggtitle("Checking ExcessveSmrtPhn_Use's distribution
")+ theme_bw() # general overview with smartphone use through data

data_activity_clean$ExcessveSmrtPhn_Use_2 <- relevel( data_activity_clean$ExcessveSmrtPhn_Use , ref=2 ) # for multinom
regression, we need reference to compare, here level 2 will be used for comparison to other levels
model_2 <- nnet::multinom(ExcessveSmrtPhn_Use_2~ AGE+ Prsnlty_type + Gender + Smoking  + Txtng +
    Entertainment+RdNews + OthRsns,data=dplyr::select(data_activity_clean, -ID,-Study_Purposes,-Rlgn_Prctces) ) # selecting
something needed
model_2 %>% parameters::parameters(exponentiate = TRUE,df_method = "wald",summary = FALSE) # getting parameter
```

Response level: 1

Parameter	Odds Ratio	SE	95% CI	t(359)	p
(Intercept)	0.51	0.98	[0.01, 22.59]	-0.35	0.726
AGE	1.05	0.09	[0.88, 1.24]	0.51	0.608
Prsnlty_type [1]	0.81	0.28	[0.41, 1.59]	-0.62	0.536
Gender [1]	0.41	0.14	[0.21, 0.80]	-2.60	0.009
Smoking [1]	0.65	0.24	[0.32, 1.33]	-1.17	0.241
Txtng [1]	1.35	0.58	[0.58, 3.14]	0.70	0.485
Entertainment [1]	0.67	0.24	[0.34, 1.35]	-1.12	0.264
RdNews [1]	0.85	0.32	[0.41, 1.76]	-0.44	0.660
OthRsns [1]	0.67	0.27	[0.30, 1.46]	-1.01	0.311

Response level: 3

Parameter	Odds Ratio	SE	95% CI	t(359)	p
(Intercept)	1.90	3.12	[0.08, 47.33]	0.39	0.694
AGE	0.94	0.07	[0.81, 1.09]	-0.85	0.396
Prsnlty_type [1]	1.24	0.33	[0.74, 2.09]	0.80	0.421
Gender [1]	1.34	0.35	[0.81, 2.24]	1.13	0.258
Smoking [1]	1.25	0.35	[0.72, 2.16]	0.78	0.435
Txtng [1]	1.70	0.64	[0.82, 3.54]	1.42	0.156
Entertainment [1]	1.88	0.57	[1.03, 3.42]	2.07	0.039
RdNews [1]	0.70	0.21	[0.39, 1.25]	-1.20	0.228
OthRsns [1]	1.15	0.34	[0.64, 2.07]	0.47	0.637

Fig 10: Multinomial regression on response levels 1 and 2 at 0.05 significance level

Response level: 4

Parameter	Odds Ratio	SE	95% CI	t(359)	p
(Intercept)	0.02	0.04	[0.00, 1.50]	-1.78	0.076
AGE	1.08	0.11	[0.89, 1.31]	0.78	0.437
Prsnlty_type [1]	1.00	0.37	[0.49, 2.05]	3.15e-03	0.997
Gender [1]	1.26	0.45	[0.62, 2.53]	0.64	0.521
Smoking [1]	1.97	0.73	[0.95, 4.07]	1.83	0.067
Txtng [1]	2.50	1.47	[0.79, 7.89]	1.56	0.118
Entertainment [1]	1.66	0.72	[0.71, 3.89]	1.16	0.245
RdNews [1]	0.77	0.31	[0.35, 1.70]	-0.64	0.521
OthRsns [1]	1.76	0.67	[0.84, 3.71]	1.49	0.136

Fig 11: Multinomial regression on response levels 4 at 0.05 significance level

Again, Here's the code and outcome from response level 3

```{r echo=FALSE}
ggplot(data_activity_for_selection, aes(ExcessveSmrtPhn_Use)) + geom_bar() + ggtitle("Checking ExcessveSmrtPhn_Use's distribution
")+ theme_bw() # general overview with smartphone use through data

data_activity_clean$ExcessveSmrtPhn_Use_2 <- relevel( data_activity_clean$ExcessveSmrtPhn_Use , ref=3 ) # for multinom
regression, we need reference to compare, here level 2 will be used for comparison to other levels
model_2 <- nnet::multinom(ExcessveSmrtPhn_Use_2~ AGE+ Prsnlty_type + Gender + Smoking  + Txtng +
    Entertainment+RdNews + OthRsns,data=dplyr::select(data_activity_clean, -ID,-Study_Purposes,-Rlgn_Prctces) ) # selecting
something needed
model_2 %>% parameters::parameters(exponentiate = TRUE,df_method = "wald",summary = FALSE) # getting parameter
```

Response level: 1

Parameter	Odds Ratio	SE	95% CI	t(359)	p
(Intercept)	0.27	0.51	[0.01, 11.07]	-0.70	0.486
AGE	1.12	0.10	[0.94, 1.32]	1.26	0.209
Prsnlty_type [1]	0.65	0.21	[0.34, 1.24]	-1.30	0.192
Gender [1]	0.30	0.10	[0.16, 0.58]	-3.61	< .001
Smoking [1]	0.52	0.18	[0.26, 1.04]	-1.84	0.065
Txtng [1]	0.79	0.35	[0.34, 1.87]	-0.53	0.598
Entertainment [1]	0.36	0.12	[0.18, 0.71]	-2.95	0.003
RdNews [1]	1.21	0.43	[0.61, 2.43]	0.55	0.584
OthRsns [1]	0.58	0.22	[0.27, 1.23]	-1.42	0.156

Response level: 2

Parameter	Odds Ratio	SE	95% CI	t(359)	p
(Intercept)	0.52	0.86	[0.02, 13.05]	-0.39	0.694
AGE	1.07	0.08	[0.92, 1.24]	0.85	0.396
Prsnlty_type [1]	0.81	0.21	[0.48, 1.36]	-0.80	0.421
Gender [1]	0.74	0.19	[0.45, 1.24]	-1.13	0.258
Smoking [1]	0.80	0.23	[0.46, 1.39]	-0.78	0.435
Txtng [1]	0.59	0.22	[0.28, 1.23]	-1.42	0.156
Entertainment [1]	0.53	0.16	[0.29, 0.97]	-2.07	0.039
RdNews [1]	1.43	0.42	[0.80, 2.56]	1.20	0.228
OthRsns [1]	0.87	0.26	[0.48, 1.56]	-0.47	0.637

Fig 12: Multinomial regression on response levels 1 and 2 at 0.05 significance level

Response level: 4

Parameter	Odds Ratio	SE	95% CI	t(359)	p
(Intercept)	0.01	0.02	[0.00, 0.66]	-2.15	0.031
AGE	1.15	0.11	[0.96, 1.38]	1.49	0.135
Prsnlty_type [1]	0.81	0.27	[0.42, 1.57]	-0.63	0.531
Gender [1]	0.94	0.31	[0.49, 1.81]	-0.20	0.845
Smoking [1]	1.58	0.55	[0.80, 3.11]	1.33	0.183
Txtng [1]	1.47	0.85	[0.47, 4.59]	0.67	0.506
Entertainment [1]	0.88	0.37	[0.38, 2.02]	-0.30	0.767
RdNews [1]	1.10	0.41	[0.53, 2.30]	0.26	0.793
OthRsns [1]	1.53	0.54	[0.77, 3.04]	1.21	0.226

Fig 13: Multinomial regression on response levels 4 at 0.05 significance level

Lastly, Here's the code and outcome from response level 4.

```
```{r  echo=FALSE}
ggplot(data_activity_for_selection, aes(ExcessveSmrtPhn_Use)) + geom_bar() + ggtitle("Checking ExcessveSmrtPhn_Use's distribution
")+ theme_bw() # general overview with smartphone use through data

data_activity_clean$ExcessveSmrtPhn_Use_2 <- relevel( data_activity_clean$ExcessveSmrtPhn_Use , ref= 4) # for multinom
regression, we need reference to compare, here level 2 will be used for comparison to other levels
model_2 <- nnet::multinom(ExcessveSmrtPhn_Use_2~ AGE+ Prsnlty_type + Gender + Smoking  + Txtng +
    Entertainment+RdNews + OthRsns,data=dplyr::select(data_activity_clean, -ID,-Study_Purposes,-Rlgn_Prctces) ) # selecting
something needed
model_2 %>% parameters::parameters(exponentiate = TRUE,df_method = "wald",summary = FALSE) # getting parameter
```
```

Response level: 1

Parameter	Odds Ratio	SE	95% CI	t(359)	p
(Intercept)	25.19	60.75	[0.22, 2842.98]	1.34	0.181
AGE	0.97	0.10	[0.79, 1.20]	-0.29	0.774
Prsnlty_type [1]	0.81	0.33	[0.36, 1.82]	-0.52	0.605
Gender [1]	0.32	0.13	[0.14, 0.73]	-2.73	0.006
Smoking [1]	0.33	0.14	[0.14, 0.76]	-2.60	0.009
Txtng [1]	0.54	0.34	[0.16, 1.84]	-0.98	0.326
Entertainment [1]	0.41	0.19	[0.16, 1.01]	-1.93	0.054
RdNews [1]	1.10	0.49	[0.46, 2.65]	0.21	0.831
OthRsns [1]	0.38	0.17	[0.16, 0.91]	-2.16	0.031

Response level: 2

Parameter	Odds Ratio	SE	95% CI	t(359)	p
(Intercept)	49.76	109.50	[0.67, 3714.91]	1.78	0.076
AGE	0.93	0.09	[0.77, 1.12]	-0.78	0.437
Prsnlty_type [1]	1.00	0.36	[0.49, 2.04]	-3.15e-03	0.997
Gender [1]	0.79	0.28	[0.39, 1.60]	-0.64	0.521
Smoking [1]	0.51	0.19	[0.25, 1.05]	-1.83	0.067
Txtng [1]	0.40	0.23	[0.13, 1.26]	-1.56	0.118
Entertainment [1]	0.60	0.26	[0.26, 1.42]	-1.16	0.245
RdNews [1]	1.30	0.52	[0.59, 2.86]	0.64	0.521
OthRsns [1]	0.57	0.22	[0.27, 1.20]	-1.49	0.136

Fig 14: Multinomial regression on response levels 1 and 2 at 0.05 significance level

Response level: 3

Parameter	Odds Ratio	SE	95% CI	t(359)	p
(Intercept)	94.77	200.16	[1.51, 5949.85]	2.15	0.031
AGE	0.87	0.08	[0.72, 1.04]	-1.49	0.135
Prsnlty_type [1]	1.24	0.42	[0.64, 2.41]	0.63	0.531
Gender [1]	1.07	0.36	[0.55, 2.06]	0.20	0.845
Smoking [1]	0.63	0.22	[0.32, 1.24]	-1.33	0.183
Txtng [1]	0.68	0.39	[0.22, 2.12]	-0.67	0.506
Entertainment [1]	1.13	0.48	[0.49, 2.60]	0.30	0.767
RdNews [1]	0.91	0.34	[0.44, 1.89]	-0.26	0.793
OthRsns [1]	0.65	0.23	[0.33, 1.30]	-1.21	0.226

Fig 15: Multinomial regression on response level 3 at 0.05 significance level

Presented results above are done at a significance level $p < 0.05$. The collective summary of our multinomial regression on all three-response levels (i.e. 1, 2, 3, 4) with respect to response levels 1, 2, 3, and 4 indicates the following prominent insights.

From response level 1

- "Entertainment" had the highest odds ratio in all response levels 2, 3, 4 with a value of 1.48, 2.79, and 2.46 respectively.
- "Txtng" had the increasing odds ratio in all response levels 2, 3, and 4 with a value of 0.74, 1.26, 1.85 respectively

From response level 2

- "RdNews" had the second-highest increasing odds ratio in all response levels 1, 3, and 4 with a value of 0.85, 1.15, and 1.86 respectively.
- "Txtng" had the increasing odds ratio in all response levels 1, 3, and 4 with a value of 1.35, 1.70, and 2.50 respectively

From response level 3

- "Txtng" had the highest increasing odds ratio of 0.79, 0.59, 1.47 in response levels 1, 2, and 4 respectively.
- "Entertainment" had the second-highest increasing odds ratio of 0.36, 0.53, 88 in response levels 1, 2, and 4 respectively.

From response level 4

- "RdNews" had the highest odds ratio in response levels 1 and 2 with a value of 1.10 and 1.30 respectively.
- "Txtng" retained the second-highest odds ratio in response level 1 with a value of 0.54.
- "OthRsns" had an increased odds ratio in response levels 1, 2, and 3 with a value of 0.38, 0.57, 0.65.

From the referential analysis on all response levels, it was found that texting had the prominent occurrence of odds ratio in all response levels where it was seen to have either the highest or increasing rate of the odds ratio. Following up, Entertainment retained the highest odds ratio in response level 1 and increased odds ratio in response level 3. On response level 4, reading news has the highest odds ratio with reference to levels 1 and 2. Texting consistently retained the second highest odd ratio in both response levels 1. Lastly, Other reasons showed an increasing trend on both response levels 1 and 4.

Based on the insights generated above, Texting and Entertainment were the two major contributing factors in excessive smartphone usage among participants. Texting retained values of more than 1.0 which shows a strong association among participants and entertainment shows a promising trend as its value was consistently increasing among various response levels. Reading news follows up close after texting and entertainment but other reasons were not seen to have any concerning association among participants.

b. Analysis of Effects.xlsx (To answer question 3)

First of all, a correlation matrix was generated to check the variable dependency. It was found that variables like "TotAddiction_Score", "Compulsive_Behaivor", "Depression_Score", "Anxiety_score" and "Functional_Impairment" correlated with high scores. Here's a look at the code and correlation matrix and correlation graph and that was used in R.

```
data_effects <- read.xls("Effects.xlsx", na.strings = c(".")) #reading data
data_effects_na_rm <- data_effects %>% na.omit # na.omit

corrplot( cor(data_effects_na_rm), order = 'hclust',tl.cex =0.5,number.cex=3,title="Effects with NA removed",mar=c(0,0,1,0))
```

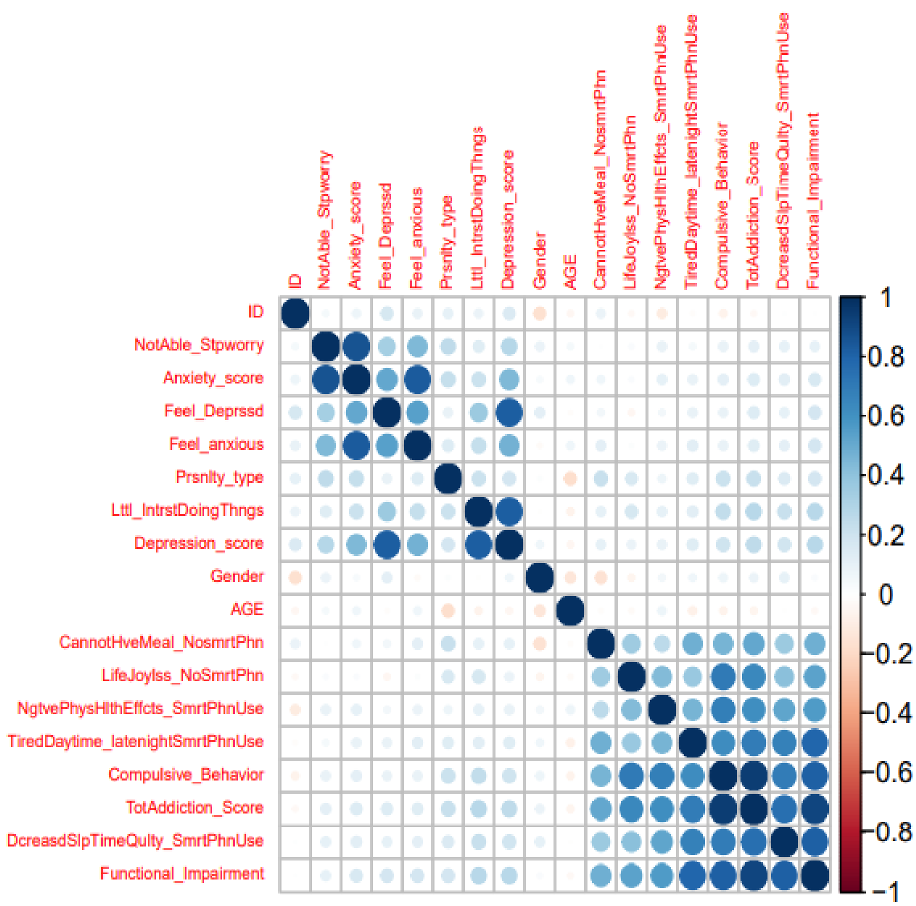

Fig 16: Correlation matrix of effects.xlsx

From the correlation matrix above, it can be noticed that variables like "TotAddiction_Score", "Compulsive_Behaivor", "Depression_Score", "Anxiety_score" and "Functional_Impairment" show a high correlation. The dependency test on these variables can be skipped for further variable selection. This data can also be illustrated with the Spearman correlation score below.

```
correlation_graph_2 <- chart.Correlation(data_effects_na_rm,
                        method="spearman",
                        histogram=TRUE,
                       pch=32,exact=FALSE)
```

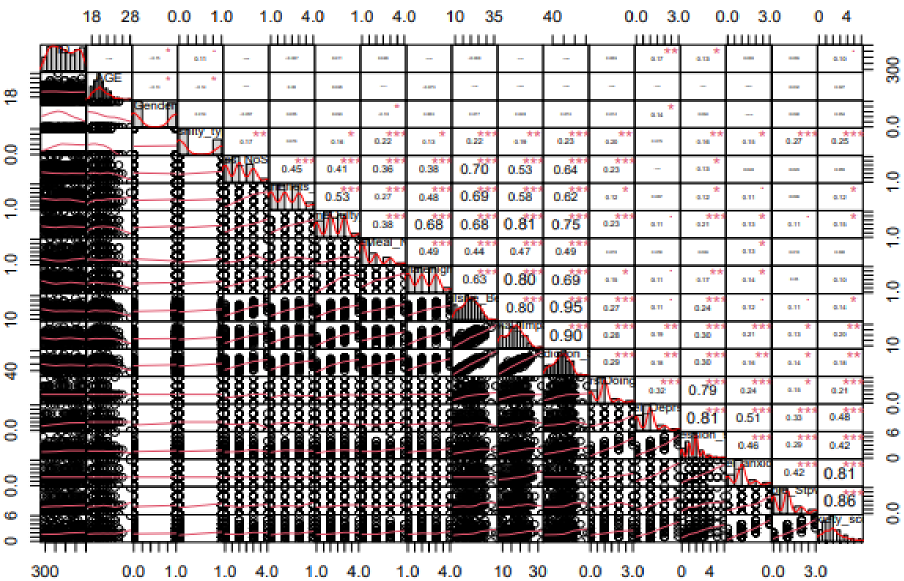

Fig 17: Correlation graph of effects.xlsx dataset

From the correlation graph above, we see a lot of the variables as mentioned earlier show a high correlation with each other. Variables like "TotAddiction_Score", "Compulsive_Behaivor", "Depression_Score", "Anxiety_score" and "Functional_Impairment" have a high correlation with each other. It can be observed that "Compulsive_Behaivor" and "TotAddiction_Score" have a high correlation of 0.95. Similarly, "TotAddiction_Score" and "Functional_Impairment" show a high correlation of 0.90 with each other. These highly correlated variables will not be needed for our analysis since they do not pose any relevance to the question our analysis is trying to answer.

Moving on, the missing values within the dataset were visualized. The *ggplot* package of R was utilized to do so and here are the results.

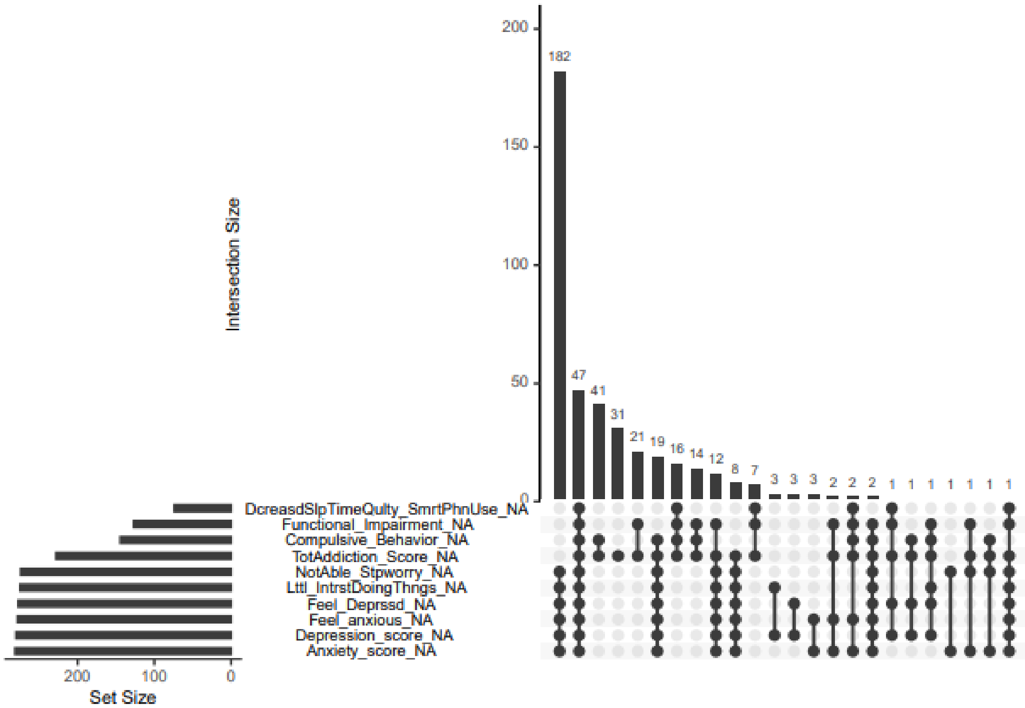

Fig 18: Bar graph of missing values from effects.xlsx dataset

From the graph above, it can be seen that the number of missing data in most of the variables like "Anxiety _score", "Feel_Deprssd", "Lttl_IntrstDoingThngs" etc. are well over 200. We will not exclude them for further analysis since the exclusion will leave us with few variables for the analysis from which the accuracy of the result can be compromised.

Following up, a mean comparison with Cohen's D was performed on the *effects.xlsx* dataset to measure the size effects of variables. This process is similar to what was done with the *activity.xlsx* dataset. The code and results are illustrated below.

```
#creating tableone for generall overview again with cohen's D
Myvariables_2 <- c("AGE", "Gender","LifeJoylss_NoSmrtPhn" ,  "NgtvePhysHlthEffcts_SmrtPhnUse",

"DcreasdSlpTimeQulty_SmrtPhnUse"  ,"CannotHveMeal_NosmrtPhn",
"TiredDaytime_latenightSmrtPhnUse", "Compulsive_Behavior" , "Functional_Impairment" ,
"TotAddiction_Score","Lttl_IntrstDoingThngs", "Feel_Deprssd","Depression_score" ,
"Feel_anxious"  ,
"NotAble_Stpworry")
cateVars_2 <- c("Gender"  ,"LifeJoylss_NoSmrtPhn" ,"NgtvePhysHlthEffcts_SmrtPhnUse",
"DcreasdSlpTimeQulty_SmrtPhnUse","TiredDaytime_latenightSmrtPhnUse","Lttl_IntrstDoingThngs" ,
                "Feel_Deprssd" ,  "Depression_score" ,"Feel_anxious","NotAble_Stpworry",
"Anxiety_score")
#tableOne
univarante_2 <- tableone::CreateTableOne(strata = "Prsnlty_type", data = data_effects_clean,
vars= Myvariables_2 , factorVars = cateVars_2 ) %>%
                print(noSpaces = TRUE, printToggle = FALSE, exact = "stage",smd = TRUE,quote =
FALSE)
univarante_2
```

	Stratified by Prsnlty_type 0	1	p	test	SMD
n	"410"	"215"	""	""	""
AGE (mean (SD))	"20.64 (1.86)"	"20.63 (1.94)"	"0.933"	""	"0.007"
Gender = 1 (%)	"196 (49.1)"	"92 (43.2)"	"0.188"	""	"0.119"
LifeJoylss_NoSmrtPhn (%)	""	""	"0.259"	""	"0.176"
1	"111 (29.9)"	"50 (25.5)"	""	""	""
2	"156 (42.0)"	"78 (39.8)"	""	""	""
3	"84 (22.6)"	"59 (30.1)"	""	""	""
4	"20 (5.4)"	"9 (4.6)"	""	""	""
NgtvePhysHlthEffcts_SmrtPhnUse (%)	""	""	"0.335"	""	"0.166"
1	"108 (29.4)"	"43 (22.4)"	""	""	""
2	"120 (32.7)"	"70 (36.5)"	""	""	""
3	"109 (29.7)"	"64 (33.3)"	""	""	""
4	"30 (8.2)"	"15 (7.8)"	""	""	""
DcreasdSlpTimeQulty_SmrtPhnUse (%)	""	""	"0.132"	""	"0.216"
1	"95 (26.2)"	"34 (17.7)"	""	""	""
2	"134 (37.0)"	"75 (39.1)"	""	""	""
3	"116 (32.0)"	"74 (38.5)"	""	""	""
4	"17 (4.7)"	"9 (4.7)"	""	""	""
CannotHveMeal_NosmrtPhn (mean (SD))	"1.63 (0.80)"	"1.87 (0.91)"	"0.001"	""	"0.288"
TiredDaytime_latenightSmrtPhnUse (%)	""	""	"0.582"	""	"0.124"
1	"118 (32.1)"	"55 (28.4)"	""	""	""
2	"116 (31.5)"	"57 (29.4)"	""	""	""
3	"116 (31.5)"	"70 (36.1)"	""	""	""
4	"18 (4.9)"	"12 (6.2)"	""	""	""
Compulsive_Behavior (mean (SD))	"18.50 (5.46)"	"19.96 (5.46)"	"0.005"	""	"0.266"
Functional_Impairment (mean (SD))	"16.51 (4.89)"	"17.65 (5.07)"	"0.015"	""	"0.228"
TotAddiction_Score (mean (SD))	"53.74 (14.63)"	"59.38 (15.36)"	"<0.001"	""	"0.376"
Lttl_IntrstDoingThngs (%)	""	""	"0.028"	""	"0.306"
0	"81 (30.8)"	"32 (22.1)"	""	""	""
1	"147 (55.9)"	"79 (54.5)"	""	""	""
2	"25 (9.5)"	"21 (14.5)"	""	""	""
3	"10 (3.8)"	"13 (9.0)"	""	""	""
Feel_Deprssd (%)	""	""	"0.383"	""	"0.178"
0	"91 (34.9)"	"43 (29.7)"	""	""	""
1	"137 (52.5)"	"75 (51.7)"	""	""	""
2	"25 (9.6)"	"20 (13.8)"	""	""	""
3	"8 (3.1)"	"7 (4.8)"	""	""	""
Depression_score (%)	""	""	"0.049"	""	"0.359"
0	"41 (15.8)"	"15 (10.5)"	""	""	""
1	"75 (28.8)"	"36 (25.2)"	""	""	""
2	"100 (38.5)"	"49 (34.3)"	""	""	""
3	"29 (11.2)"	"23 (16.1)"	""	""	""
4	"7 (2.7)"	"12 (8.4)"	""	""	""
5	"6 (2.3)"	"5 (3.5)"	""	""	""
6	"2 (0.8)"	"3 (2.1)"	""	""	""
Feel_anxious (%)	""	""	"0.033"	""	"0.314"
0	"74 (28.5)"	"24 (16.6)"	""	""	""
1	"143 (55.0)"	"88 (60.7)"	""	""	""
2	"29 (11.2)"	"19 (13.1)"	""	""	""
3	"14 (5.4)"	"14 (9.7)"	""	""	""
NotAble_Stpworry (%)	""	""	"<0.001"	""	"0.523"
0	"104 (39.7)"	"28 (19.0)"	""	""	""
1	"122 (46.6)"	"77 (52.4)"	""	""	""
2	"22 (8.4)"	"26 (17.7)"	""	""	""
3	"14 (5.3)"	"16 (10.9)"	""	""	""

Fig 19: Summary of standard mean difference with Cohen's D by personality type in effects.xlsx dataset

From the collective results generated by the Cohen's D procedure above, the variables "DecreasdTmeSlpQulty_SmrtPhneUse"," CannotHveMeal_NosmrtPhn" "Compulsive_Behavior" and "Functional_Impairment" retained a perfect size effect of 0.2. This means that these variables are statistically significant in the *effects.xlsx* dataset.

Furthermore, AIC with a backward algorithm was used for variable selection. Variables with a larger size effect were selected alongside a variable with a perfect size effect. Since "Lttl_IntrstDoingthngs" and "NotAble_Stpworry" retained a size effect of 0.306 and 0.523 respectively they were used for AIC alongside "CannotHveMeal_NoSmrtPhn" which retained a perfect size effect of 0.288. The code and outcomes of the procedures are illustrated below.

```r
data_effects_for_selection <- data_effects_clean %>% na.omit() %>%
dplyr::select(-Depression_score, -TotAddiction_Score,
-Compulsive_Behavior,-Functional_Impairment)  # required for independency of data analyse

model_10 <- glm(Prsnlty_type~.,data=data_effects_for_selection , family = binomial )
model_10_null <- glm(Prsnlty_type~1,data=data_effects_for_selection , family = binomial )
model_10_selection <- MASS::stepAIC(model_10,direction="backward",trace = FALSE)
model_10_selected <- glm(Prsnlty_type ~  AGE + CannotHveMeal_NosmrtPhn +
    Lttl_IntrstDoingThngs + NotAble_Stpworry,data = data_effects_for_selection,
                  family = binomial(link="logit"))
model_10_selected %>% parameters::parameters(exponentiate = TRUE, df_method = "wald",summary =
FALSE)
anova(model_10_selected,
      model_10_null,
      test="Chisq")
```

```
Parameter                 | Odds Ratio |    SE |          95% CI |     z |      p
---------------------------------------------------------------------------------
(Intercept)               |      10.14 | 18.20 | [0.30, 341.70]  |  1.29 | 0.197
AGE                       |       0.78 |  0.07 | [0.66,   0.93]  | -2.78 | 0.005
CannotHveMeal_NosmrtPhn   |       1.69 |  0.28 | [1.23,   2.34]  |  3.19 | 0.001
Lttl_IntrstDoingThngs [1] |       1.38 |  0.48 | [0.70,   2.72]  |  0.94 | 0.346
Lttl_IntrstDoingThngs [2] |       2.37 |  1.23 | [0.86,   6.54]  |  1.66 | 0.096
Lttl_IntrstDoingThngs [3] |       4.73 |  3.33 | [1.19,  18.79]  |  2.21 | 0.027
NotAble_Stpworry [1]      |       2.62 |  0.91 | [1.33,   5.17]  |  2.78 | 0.005
NotAble_Stpworry [2]      |       5.30 |  2.46 | [2.13,  13.18]  |  3.58 | < .001
NotAble_Stpworry [3]      |       5.19 |  3.02 | [1.66,  16.21]  |  2.84 | 0.005

Analysis of Deviance Table

Model 1: Prsnlty_type ~ AGE + CannotHveMeal_NosmrtPhn + Lttl_IntrstDoingThngs +
    NotAble_Stpworry
Model 2: Prsnlty_type ~ 1
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1       248     291.26
2       256     338.61 -8  -47.352 1.314e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Fig 20: AIC with a backward algorithm in effects.xlsx dataset

In the AIC analysis done above, a few observations were made.

- 1.69 odd ratio with "CannotHveMeal_NosmrtPhn" by personality 0 to 1
- NotAble_Stpworry shows a notable odd ratio between its level from 2 (OR: 5.3) to Personality type
- Model is significant that in-between final variables, there are relevant predictors to variable "Prsnlty_type"

To summarize our analysis from a visual perspective, here's a resulting violin plot that was generated in R.

````
## Data Visualization with Violinplot

```{r echo=FALSE}
ggplot(data_effects_for_selection,  aes(x= NotAble_Stpworry , y= CannotHveMeal_NosmrtPhn ,
color= CannotHveMeal_NosmrtPhn )) +geom_jitter(alpha= 0.4 )+ geom_violin(alpha= 0.5) +
facet_wrap(~Prsnlty_type) + ggtitle("Comparison within Prsnlty_type", subtitle = "between
NotAble_Stpworry and CannotHveMeal_NosmrtPhn ")
```
````

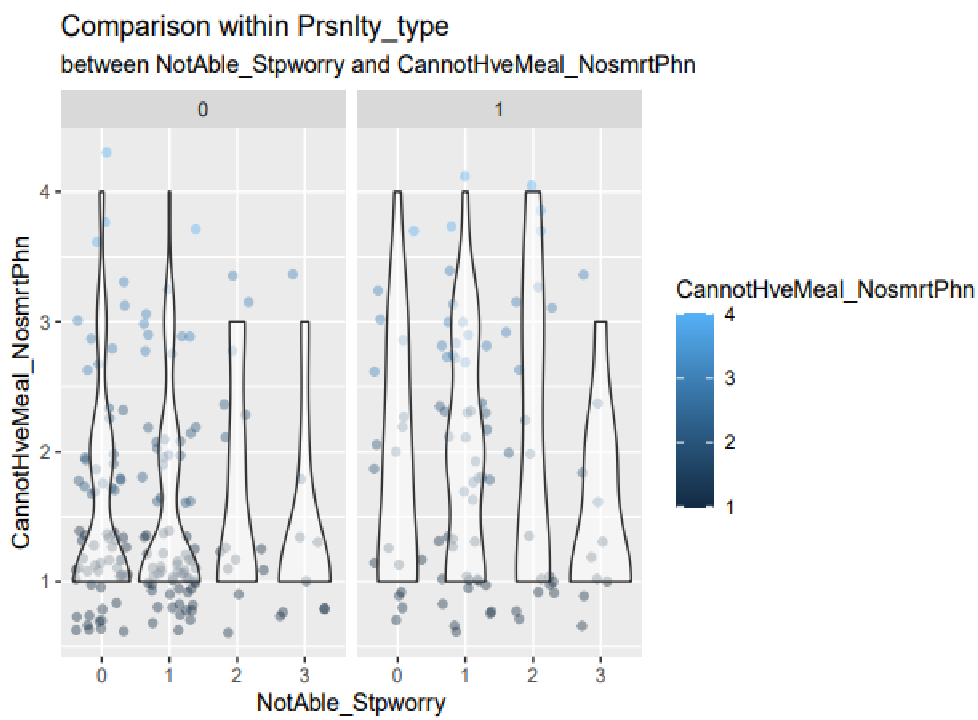

Fig 21: Violin plot of selected variables in different personality types

In the violin plot illustrated above, we can see the x-axis represents the variable "NotAble_Stpworry" meanwhile the y-axis represents the variable "CannotHveMeal_NoSmrtPhn". The violin plot depicts these variables in terms of personality types 0 (A) and 1 (B). Personality type 1 (B) has more thickness and spread out than personality type 0 (A). It would be safe to say that the effect of excessive smartphone usage like not being able to stop worrying and inability to enjoy a meal has more impact on personality type 1 or B.

Visual Analysis

a. Analysis of Scores.xlsx Dataset (To answer question 4)

The visualization software Tableau was used to generate visuals to draw conclusions from them. Scores dataset consists of variables relating to different symptoms that participants admitted of experiencing. Variables like "Compulsive_Behavior", "Functional_Impairment", "Feel_Depressed" are used to represent the psychological aspect of participants. In the original dataset and its analysis (Boumosleh & Jaalouk, 2017) these variables were used to denote the Smart Phone Addiction Indicator (SPAI) scale. SPAI is the cumulative score of a variable derived using the mean difference and Chi-square. The new visualization that was generated using the new dataset is a mere visual portrayal and depiction of the SPAI with a minor modification in the values. The personality type 0 was changed to A and 1 was changed to B to make it easier to comprehend and explain the results. No other changes were made.

In Tableau, the *scores.xlsx* was located, the first and foremost challenge in extracting visuals was to correctly identify the dataset and its attributes. Categorical values like our dataset always need a few modifications to use them properly and effectively. For an instance, even though our dataset has numerical values that denote the count of a number they are not properly configured to be used as the whole number so the data type needs to be changed. In our case, the data type was changed to a numeric whole number and a few variables were converted to dimension to achieve the right and appealing set of bar graphs. From this dataset, we are trying to find effects and the difference in vulnerabilities among two personality types. The resulting bar graphs and a comprehensive dashboard are pictured below.

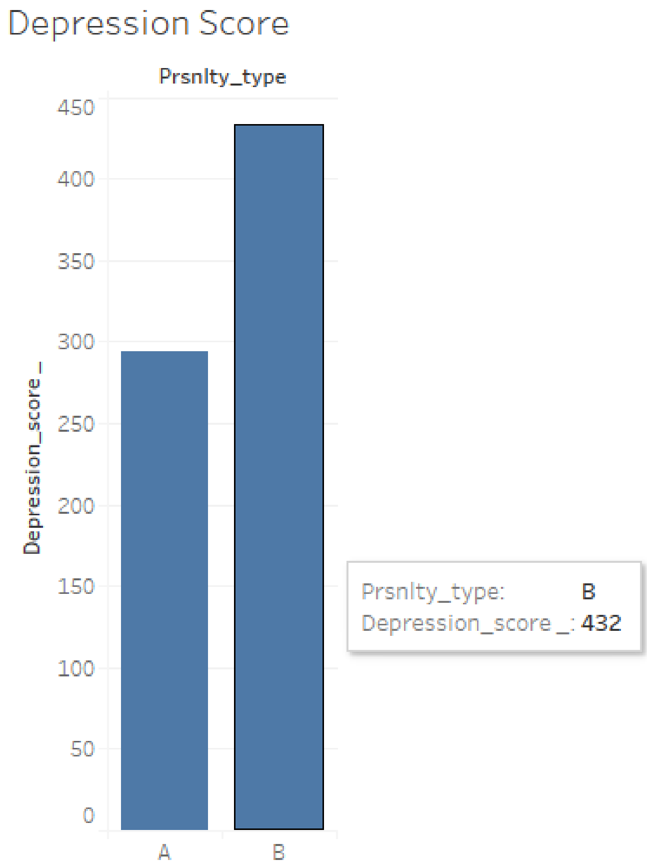

Fig 22: Bar graph of the variable "Depression_score" on two personality types.

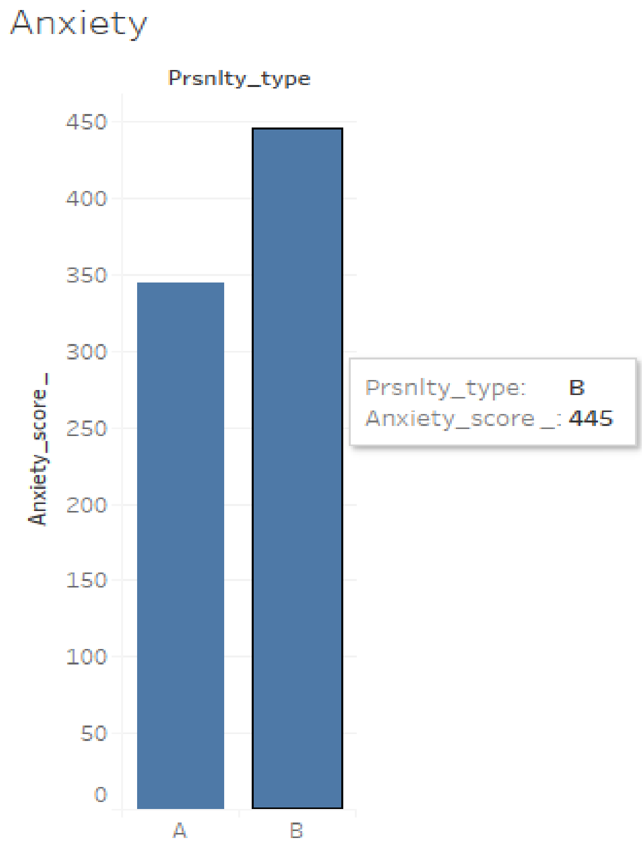

Fig 23: Bar graph of the variable "Anxiety" on two personality types.

In figure 22, we can see that personality type B has a high score on the variable "Depression _score" from the SPAI scale. This hints that personality type A is less susceptible to the effects of excessive smartphone usage.

Similarly, in figure 23 personality type B has a count of 445 in "Anxiety_score" from the SPAI scale. Personality type A following up close but nowhere near to personality B. This makes personality type A less susceptible to the effects of excessive smartphone usage. Meanwhile, personality type B is more susceptible to effects of the excessive smartphone use.

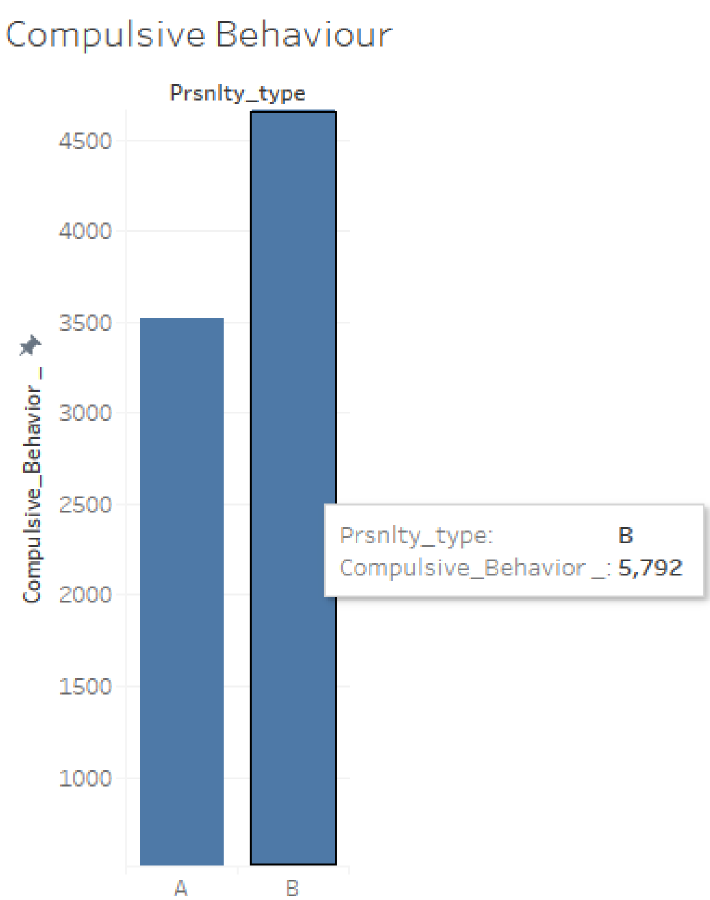

Fig 24: Bar graph of "Compulsive_Behavior" among two personality types

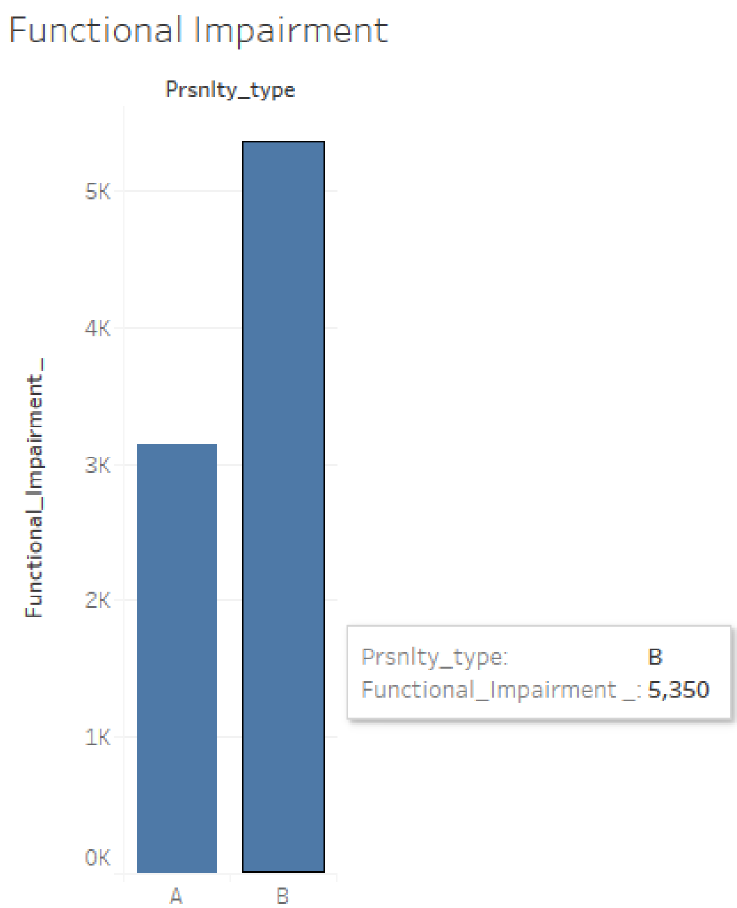

Fig 25: Bar graph of "Functional_impairment" among two personality types

From figure 24, it can be seen that personality type B has a whopping 5792 score from the SPAI scale. Personality type A has a score of 3500. Statistically speaking personality type B is 60% more likely to be affected with compulsive behavior as a result of excessive smartphone usage. Again, figure 25 shows that personality type B has the highest score of 5350 on the SPAI scale. Personality type A sums up to around 3100 making it about 57% less likely to be affected with functional impairment as a result of excessive smartphone usage.

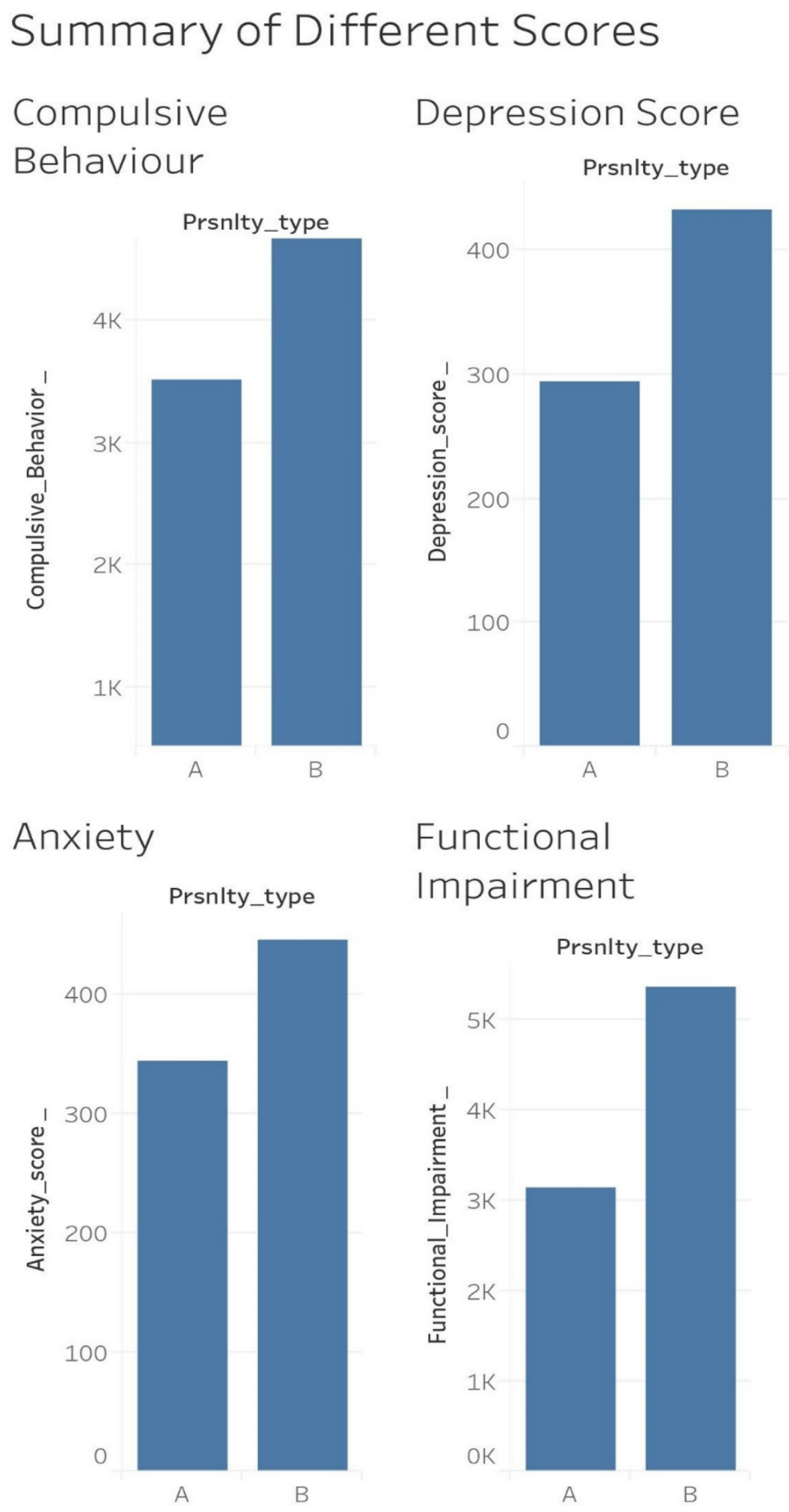

Fig 26: Dashboard of all variables from the scores.xlsx dataset

The dashboard above provides a collective and comprehensive visual summary of the *scores.xlsx* dataset

Interpreting Analysis to Answer Research Questions

From our multi-layered analysis of the datasets, a lot of results were produced. Inspection and filtration were done to extract only the meaningful, accurate, and relevant results. Not every method that was utilized returned expected results.

To begin with, we had three major variables that represented the personal habits of the candidates i.e. "Rlgn_Prctces", "Smoking" and "Alcohol_drnk" from the *activity.xlsx* dataset. Early on "Rlgn_Prctces" was seen to have a high count of missing data so technically the chances of that variable answering one of our research questions was very minimal. When proceeded further with Cohen's D only "Smoking" had a perfect size effect of 0.2 and "Alcohol_drnk" retained a lower value of 0.021. Furthermore, "Smoking" returned the slightest consistently better odds ratio in different response levels. The increasing odds ratio tells us that the change in the "Smoking" variable is likely to cause a shift in the overall analysis.

Therefore, it can be concluded that among the three variables of personal habits, Smoking highly contributes to excessive smartphone usage. Drinking alcohol is less likely to have a higher effect on excessive smartphone usage although it does contribute in a minor way. Religious practices have almost zero or insignificant effect in causing excessive smartphone use. Furthermore, the multinomial regression on the *activity.xlsx* dataset concluded that texting and entertainment showed a consistently increasing pattern on each response level (Fig 8-15). "Txtng" was the most prominent among all. In other words, texting was seen to be strongly associated with excessive smartphone usage. Additionally, from the Cohen's D on the *activity.xlsx* dataset, personality type 1 or B generated higher Standard Mean Difference than type 0 or A which in general can be said that type 1 or B just has a higher effect on the activities than type 0 or A.

Similarly, from the second Cohen's D on the *effects.xlsx* dataset, it was found that variables like "DecreasdTmeSlpQulty_SmrtPhneUse", "Compulsive_Behavior" and "Functional_Impairment" retained a perfect size effect of 0.2 out of all other variables. It can be summed up that the effects of excessive smartphones like a decrease in quality sleep time, compulsive behavior, and functional impairment are higher than any other mentioned effects. The Standard Deviation of personality type 1 or B in three variables was also found to be more causing higher variance. Following up close on the similar pattern are variables like "CannotHveMeal_NosmrtPhn", "Lttl_IntrstDoingThngs" that showed a promising trend. "NotAble_Stpworry" returned a higher size effect value so to further analyze these variables AIC procedure was conducted. Although no significant insights were found with the procedure but generating a violin plot (Fig 21) tells us that personality type 1 or B are likely to have effects like not being able to have a meal without a smartphone and inability to stop worrying. These variables compared to the ones from Cohen's D can be weak since they initially did not show a perfect size effect, but they do have minimal to broad effect over excessive smartphone use.

Lastly, the visualizations generated from the *scores.xlsx* dataset in Tableau are obvious as the illustrations are very precise and simple. The SPAI score of different variables shows how the variable compare against each other in different personality types. Illustrations depicted above, figure 22, 23, 24, 25, and 26 summarize that personality type 1 or B is likely to score more on the SPAI scale of different effects of excessive smartphone usage. This tells us that personality type 1 or B is more vulnerable to the effects of excessive smartphone usage.

Appendix A: Limitation and Potential of the Analysis

As it's known, the original dataset was raw when it was retrieved initially so the possibilities with the original dataset are endless. Our analysis is also an instance of one of those possibilities since a new dataset was created by splitting required variables. This not only helped to conduct the required analysis but also to reach conclusions and answer the research questions. However, even our study and the analysis have a few limitations that were not addressed or explored.

For an instance, the variable "smoking" was found to be strongly linear to excessive smartphone usage. This is to say that group of people who admitted smoking as an interpersonal habit are prone to be addicted to excessive smartphone use. However, it is not clear what category of smokers (i.e. heavy, light, or chain smokers) are more likely to be excessive smartphone users. This is by default one of the limitations of the dataset and analysis we performed.

Another instance of such limitation is that our dataset does not have any information or explanation on some of the effects that participants had because of excessive smartphone use. For example, the original dataset and the one we created contains variables like "compulsive behavior", "depression", "anxiety", "functional impairment" etc. but the values for these variables were generated based on the candidate's admission or denial on the respective topic. A classic yes or no type of response could be ineffective for a broad subject that those variables are representing. We are not given or told the contributing factors of those variables and it can be biased since any candidate on a given day may give a different response on a specific topic depending on their mood.

In addition to this, other forms of analysis can be done with our dataset to answer more questions. For an instance, a possible question can be related to finding a connection between texting and anxiety in both personality types, finding if there exists any relationship between participants who are depressed see using a smartphone as entertainment purpose as a cause of depression, finding the relationship between compulsive behavior and the number of people who admit smoking can lead to excessive smartphone usage and then comparing it against with people who deny, etc.

Using two or more multi-layered approaches of analysis can help in finding the answers to the questions mentioned above. New dataset creation with relevant variables can be a starting point.

Appendix B: Comparing and Contrasting with Primary Research

In a lot of ways, our study concurs with as well as differs from the primary research. First of all, the primary research was specifically focused on finding if any relationship exists between depression, anxiety, and smartphone usage, and our study was specifically based on exploring the adverse effects of smartphone usage by analyzing certain and limited variables from the original dataset. In hindsight, our study is a new iteration of the primary research that was not addressed or explored.

The primary research utilized a different set of tests to achieve mathematical and statistical results. Descriptive statistics, qualitative and quantitative measurements like mean, standard deviation, independent 2 sample T-Test, Chi-Square, and multilinear regression were effectively used whereas our study took a similar approach with added AIC method and multinomial regression for variable selection and variable association respectively.

A common element between the primary research and our study was the use of p-value at less than 0.05 to ensure the computation of strong evidence. Both studies effectively utilized

Spearman correlation coefficients to evaluate the association among the different dependent and independent variables. However, both studies did not use the same tool to analyze and conclude the results, the primary research used Statistical Package for Social Sciences (SPSS) whereas our study used R programming language.

The most notable and mutual similarity in both studies was the conclusion that personality type 1 or B were most susceptible to excessive smartphone use.

In addition to this, there were key differences in terms of the limitation of each study. The primary study does not explicitly describe how candidates were anxious or stressed using smartphones. A typical yes or no survey was carried out to have participants admit or deny if they have anxiety, stress, or depression. This procedure could be ineffective as it is very subjective and not based on any scientific evidence. Self-reporting a psychological aspect is not a precise way of drawing conclusions.

Whereas in our study, a detailed description of a few variables is not given. As mentioned previously, for the variable "smoking" we are not told or given the category of smokers. This is crucial since our analysis found that people who smoke are prone to be addicted to excessive smartphone usage but not all people have the same smoking preference, some are light smokers while some are heavy. This leads to instilling that people who smoke in general are addicted to smartphone usage in general. One of the main differences is that the primary research fully utilizes the original dataset by using methods to test all its variables meanwhile our study partially utilizes and draws new conclusions. In addition to differences, the primary research addresses missing values by treating them as an outlier without using any method, but our study uses the built-in R procedure to treat missing results as "NA" variable to make it compatible for R to run the analysis and to reduce biases

Appendix C: Conclusion

After analyzing the dataset using a multi-layered approach with various methods, we have gleaned results from the analysis to answer our research questions. The main goal of this study was to correctly identify the problem of excessive smartphone use and explore its adverse effects. From ranges of personal habits (i.e. drinking alcohol, smoking, and religious practices) in the *activity.xlsx* dataset, smoking was found to have a strong relationship with excessive smartphone usage smart people. Drinking alcohol follows up as a second personal trait that is somewhat associated with excessive smartphone usage meanwhile religious practices were found to have no association, therefore, posing no effect in excessive smartphone usage.

Similarly, texting and entertainment showed a strong association among participants who admitted to excessive smartphone usage. Reading followed up closely, but other reasons did not have a solid impact compared to any other variable.

From the *effects.xlsx* dataset, it was found that participants who excessively use smartphones highly experience effects like a decrease in sleep time, inability to enjoy a meal, general loss of interest in doing things, and ability to stop worrying. Among which the decrease in sleep time quality overall had a strong association with other independent variables that proved one of the major adverse effects of excessive smartphone use is loss of sleep or insomnia. Other variables of effects mentioned above also had a good level of association with other variables and they were mostly found on participants with personality type 1 or B.

Lastly, the personality type with a noted high majority of psychological changes like compulsive behavior, functional impairment, anxiety, and depression are type B which suggests that personality type B are likely to be more vulnerable and susceptible to excessive smartphone use.

References

Lin, Y. (2021, August 2). *10 mobile usage statistics you should know in 2021*. Orbelo. Retrieved from https://www.oberlo.com/blog/mobile-usage-statistics.

Tocci, M. (2021, July 22). *History and Evolution of Smartphones*. Retrieved from https://simpletexting.com/where-have-we-come-since-the-first-smartphone/

Silver, L. (2019, February 5). *Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally*. PewResearch.org. https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/

Ellis, D., Shaw, H. (2018): *Typical smartphone usage dataset*. Lancaster University. https://doi.org/10.17635/lancaster/researchdata/228

Boumosleh, J. M., & Jaalouk, D. (2017). Depression, anxiety, and smartphone addiction in university students- A cross sectional study. PLOS ONE 12(8): e0182239. https://doi.org/10.1371/journal.pone.0182239

Boumosleh, J. M., & Jaalouk, D. (2017). *Raw data for all participants*. (Version 1). PLOS ONE. https://doi.org/10.1371/journal.pone.0182239.s001

Liebherr, M., Schubert, P., Antons, S., Montag, C. (2020). Smartphones and attention, curse, or blessing? - A review on the effects of smartphone usage on attention, inhibition, and working memory. *Computers in Human Behavior Report.* Vol.1. https://doi.org/10.1016/j.chbr.2020.100005.

Arumugam, N., Selvanayagam, S., & Sathiyasenan, S. T. (2020). The Effects of Smartphone Usage on University Students. *International Journal of Academic Research in Progressive Education and Development*, 9(3), 170–183. http://dx.doi.org/10.6007/IJARPED/v9-i3/7960

S. O'Dea (2021, March 19). *Number of smartphone users in the U.S. 2025*. Statista. Retrieved from https://www.statista.com/statistics/201182/forecast-of-smartphone-users-in-the-us/

S. O'Dea (2021, July 12). *Forecast number of mobile users worldwide from 2020 to 2025*. Statista. Retrieved from https://www.statista.com/statistics/218984/number-of-global-mobile-users-since-2010/

Horwood, S., Anglim, J., Mallawaarachchi, R, S. (2021). Problematic smartphone use in a large nationally representative sample: Age, reporting biases, and technology concerns. *Computers in Human Behavior*. Volume 122. https://doi.org/10.1016/j.chb.2021.106848

Sahin S, Ozdemir K, Unsal A, Temiz N. (2013). Evaluation of mobile phone addiction level and sleep quality in university students *Pakistan journal of medical sciences.* 29(4):913-918. http://dx.doi.org/10.12669/pjms.294.3686