# Social Media & Mental Health

# Exec Summary

# Introduction

Social media often feels like a mainstay of life today, with 59.4% of the population actively engaging with it in some way (*The Changing World of Digital In 2023*, 2023). Social media itself is certainly not new, with numerous social networking sites appearing from the 1980s onwards, such as AmericaOnline in 1985, Friendster in 2001, and Facebook in 2004 (‘The Evolution of Social Media: How Did It Begin and Where Could It Go Next?’, 2020), with daily use increasing from 1 hour 37 minutes in 2013 to 2 hours and 31 minutes in 2022, despite internet use only increasing by 28 minutes in that same time frame (*The Changing World of Digital In 2023*, 2023). Research exploring the impacts of social media on mental health have found both positive and negative outcomes – REF

Impacts on loneliness, links between mental unhealth and loneliness. Anxiety, depression.

Research tends to focus on children/teenagers.

ADHD/attention issues.

Self-esteem.

# Methodology

Raw data was sourced from Kaggle <https://www.kaggle.com/datasets/souvikahmed071/social-media-and-mental-health?select=smmh.csv>. The data was collated via a survey conducted at a University, assessing time spent on social media and answers to questions to 9 questions regarding symptoms of depression (questions 18-20), anxiety (11, 13), and ADHD (9-10, 12, 14), and 3 relating to self-esteem (15-17). These questions were rated on a 5-point Likert scale (Joshi *et al.*, 2015) to quantify mental health outcomes.

The data was imported into Excel for cleaning, raw data above.

The timestamp and affiliated organisation columns were removed as these were not relevant to the analysis. After exploring the responses held in the column “Do you use social media?” it was decided to remove this as well as the time spent on social media column would indicate level of social media use. After exploring the responses under the “Gender” column, one row was removed as the response “There are others???” had the potential to skew results if grouped as “Other” alongside those that had responded with gender identities outside of male and female.

Next, columns containing categorical data were split into separate columns using one-hot encoding (Brownlee, 2017) to avoid any implication of scale between variables that can occur using integer encoding:

* Gender split into “Male”, “Female”, and “Other” columns
* Relationship status = “Single”, “In a Relationship”, “Divorced”
* Occupation status = “Student”, “Employed”, “Retired”
* Social Media platforms = “Discord”, “Facebook”, “Instagram”, “Pinterest”, “Reddit”, “Snapchat”, “TikTok”, “Twitter”, “YouTube”

The column containing hours spent on social media, responses followed the format of “Between x and x hours”. These responses were turned into a number, taking the midpoint of the hours range, i.e. “Between 1 and 2 hours” becomes 1.5. This was done as the lowest option was “Less than 1 hour” which is not necessarily zero hours, so to revert this to 0 did not make sense – is this the right approach or should whole numbers be used?

Columns containing the answers to the Likert scale questions were rearranged so they sat in groups relating to depression, anxiety, ADHD, and self-esteem. While the questions for depression, anxiety, and depression were scored on a basis of 1 meaning symptoms are never experienced and 5 meaning they’re experienced on a very regular bases, self-esteem was measured as 1-2 being high self-esteem and 4-5 being low self-esteem – change scale?

EDA – histograms for gender, age, time spent on social media, and types of platforms.

Exploration of multicollinearity – VIF if in Excel, PLS Regression if in Python.

Multivariate linear regression:

* Independent variables – age, gender, time on social media, which platforms used, and number of platforms used
* Dependent variables – depression, anxiety, ADHD, and self-esteem scores

# Results

# Conclusion

Measures of mental health – questions created by creator of original dataset and may not accurately measure mental health outcomes in the same way as the Beck Depression Inventory (Jackson-Koku, 2016), the Hamilton Anxiety Scale (Bhamra, Naqvi and Arora, 2021), and the Rosenberg Self-Esteem Scale (Ref).

# References

Bhamra, M.K., Naqvi, W.M. and Arora, S.P. (2021) *Assessment of Anxiety using Hamilton Anxiety Scale in Augmented Reality Head Mounted Display User: A Study Protocol*. preprint. Protocol Exchange. Available at: https://doi.org/10.21203/rs.3.pex-1547/v1.

Brownlee, J. (2017) ‘Why One-Hot Encode Data in Machine Learning?’, *MachineLearningMastery.com*, 27 July. Available at: https://machinelearningmastery.com/why-one-hot-encode-data-in-machine-learning/ (Accessed: 27 June 2023).

Jackson-Koku, G. (2016) ‘Beck Depression Inventory’, *Occupational Medicine*, 66(2), pp. 174–175. Available at: https://doi.org/10.1093/occmed/kqv087.

Joshi, A. *et al.* (2015) ‘Likert Scale: Explored and Explained’, *British Journal of Applied Science & Technology*, 7(4), pp. 396–403. Available at: https://doi.org/10.9734/BJAST/2015/14975.

*The Changing World of Digital In 2023* (2023) *We Are Social UK*. Available at: https://wearesocial.com/uk/blog/2023/01/the-changing-world-of-digital-in-2023/ (Accessed: 18 June 2023).

‘The Evolution of Social Media: How Did It Begin and Where Could It Go Next?’ (2020) *Maryville Online*, 28 May. Available at: https://online.maryville.edu/blog/evolution-social-media/ (Accessed: 18 June 2023).