Social Media's Impact on Shopping Behavior: An Exploratory Data Analysis

Ma. Addine Anne T. Carreon
Sydney Anne V. Lucas
Dale Russel Quipit

DS119 – Exploratory Data Analysis
Term, SY 2022-2023

I. INTRODUCTION

In today's digital age, the increasing use of social media platforms has greatly influenced various aspects of our lives, including shopping behavior. Millennials and Gen Z members have been at the forefront of this wave of change, being the most active users of social media. This exploratory data analysis aims to explore the social influence on shopping behavior based on a social survey conducted among 300,000 Millennials and Gen Z members. The survey data provides valuable insights into the impact of social media on their shopping decisions and the factors that contribute to their purchasing choices. Various social media platforms such as Instagram, Facebook, and Twitter offer consumers numerous opportunities to make informed decisions when purchasing products. These platforms play a crucial role in influencing consumers' online shopping behaviors. As social media continues to thrive, it becomes an invaluable source of information for buyers, offering insights into the utility and features of products (Diwan & Gowel, 2022). Social media has become an integral part of the daily lives of Millennials and Gen Z, with these two groups being the most active users of various platforms. As a result, their shopping habits have been heavily impacted by the content they consume on social media.

According to a study by KPMG titled "The Truth About Online Consumers", 60% of consumers aged 18-24 agree that social media platforms influence their purchasing decisions (KPMG, 2018). Also, research conducted by Pew Research Center reveals that 88% of 18-24-year-olds use social media platforms regularly (Smith & Anderson, 2018). Social media platforms have opened new avenues for self-expression, allowing young adults to curate their identities and showcase their preferences and lifestyles. Similar to the concept of window shopping in the past, social media platforms provide a virtual storefront for brands to engage with potential customers. The aspirational nature of social media content, product recommendations, and influencer endorsements play a pivotal role in shaping the purchasing decisions of Millennials and Gen Z members. A substantial 75% of Gen Z consumers globally rely on social media to guide their purchase choices, and 61% have made purchases based on social media advertisements. Additionally, social media influencers have a notable impact on Gen Z's buying behavior, with 68% of them expressing greater trust in influencers over traditional celebrities, as reported by a Morning Consult study. Moreover, apart from social media platforms, the shopping behavior of Millennials and Gen Z is significantly impacted by peer recommendations. Personal connections and social networks play a crucial role in shaping their

attitudes and choices when making purchases. Growing up in an interconnected era, Millennials and Gen Z heavily depend on their friends, family, and online communities for product reviews and recommendations. The combined effect of social media and peer recommendations has transformed the way young consumers navigate the world of shopping, making them more discerning and informed in their choices. As digital natives who have grown up in an interconnected era, their reliance on these influential factors continues to shape the landscape of consumer behavior.

The specific goal of the exploratory data analysis (EDA) is to gain insights and understanding of the dataset focused on studying and recognizing the purchasing behaviors influenced by social media among millennials and Gen Z members. The EDA aims to explore data to identify patterns, trends, and relationships between variables related to social media's impact on shopping habits among the target age groups. The main objective of the exploratory data analysis is to perform an in-depth examination of the dataset, utilizing various statistical and visualization techniques, to achieve the following: (1) Understand the distribution of different segments in the dataset, the EDA will examine the segment type column to identify the various categories or groups of participants and their proportions within the dataset. (2) Investigate opinions about social influence on shopping habits, the EDA will analyze the answers given to survey questions to understand millennial and Gen Z participants' opinions about social media's influence on their shopping habits. It will identify aspects they view positively or negatively, which may include factors influencing their trust in online merchants and purchasing decisions. (3) Uncover trends and correlations between different segments, the EDA will explore relationships between different variables and segments to uncover any interesting trends or correlations. For instance, it may identify whether a particular demographic value or use of certain types of social influence on their shopping habits more than others. (4) Utilize count and percentage columns to provide insights, the EDA will leverage the count column to understand the number of responses obtained for each survey question. Additionally, the percentage column will be used to interpret the count data and gain insights into the generations' social influence on shopping based on accumulated percentages.

Overall, the exploratory data analysis seeks to provide a comprehensive understanding of the dataset and its implications on purchasing behaviors, allowing for data-driven insights to inform further research or decision-making related to social media marketing strategies and its impact on millennials and Gen Z consumers.

II. Results and Discussion

This section analyzes and presents the findings, which conduct an understanding of the obtained dataset of social influence on shopping. The data’s results and discussion explain the procedure, processes, and interpretation of varying variables that affect the dataset.

In examining the data, it is vital to obtain an accurate sample, size, data process, and analysis used that undermines the identification and classification of the data’s variables. The data collection method utilizes survey questionnaires to gain information regarding the purchasing behavior of 300,000 millennials and Gen Z members. The survey initializes to determine the range of insights and findings of the participants. The specified number of individuals is the dataset’s sample size, which comprises 2,676 data points that participated in the survey. In

analyzing the collected data and assessing the data findings, numerical variable analysis is utilized. This analysis includes numerical continuous data features such as the dataset's variables segment type and the accumulated survey answer.

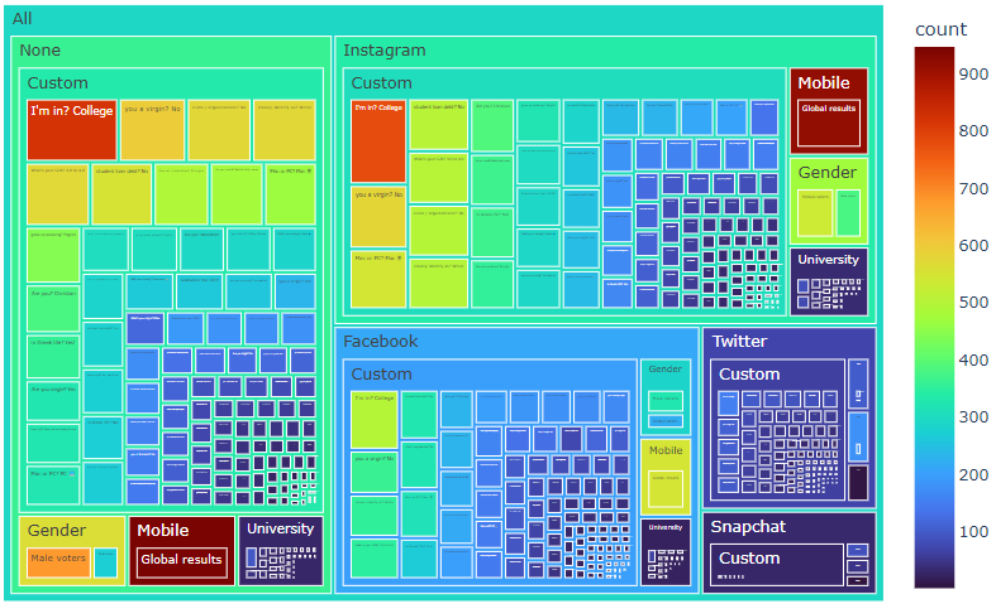


Illustration 1.0 Graphical Representation of All Variables With its Count

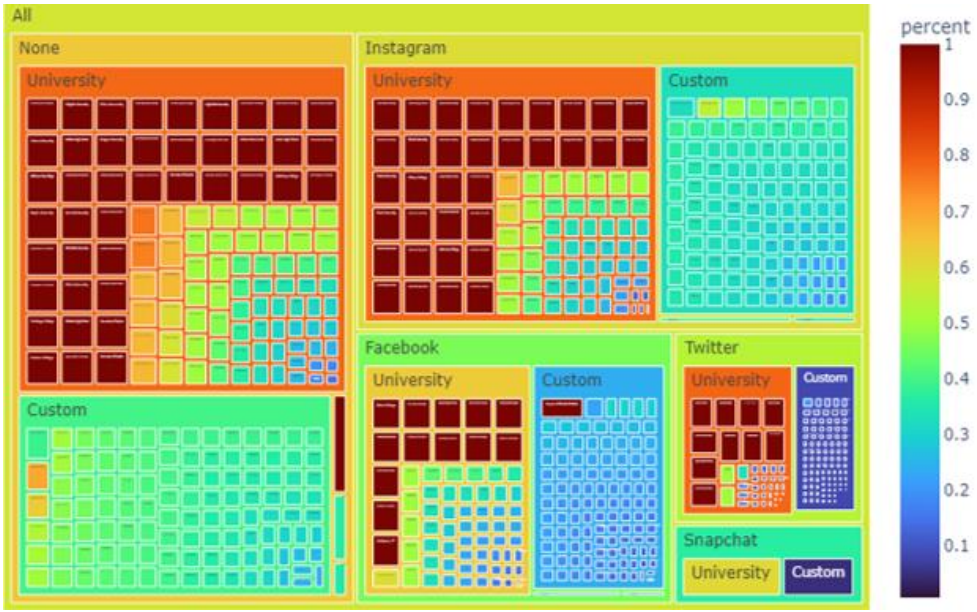
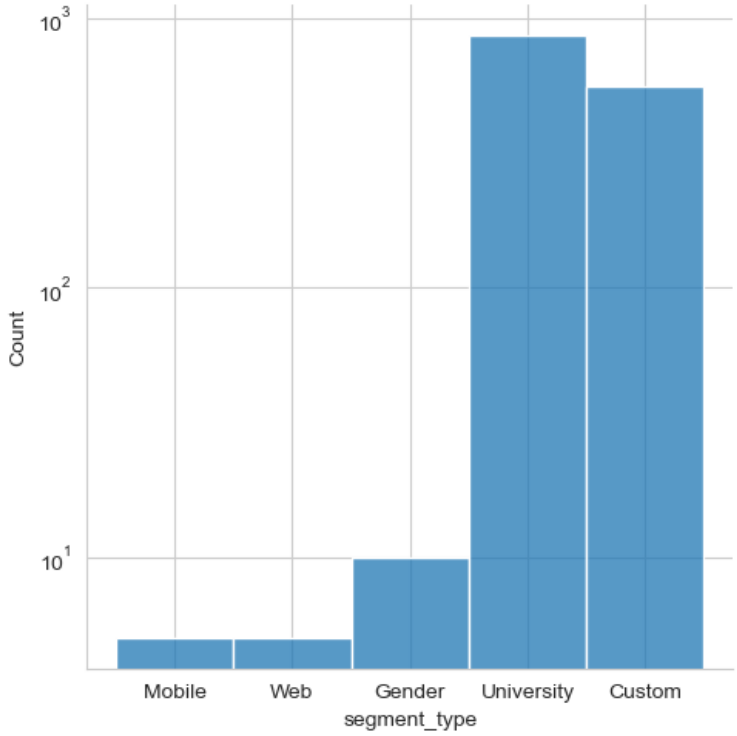


Illustration 1.1 Graphical Representation of All Variables With its Percentage

In Illustration 1.0 and 2.0, Treemap was used to show all the variables — Answer, Segment Type, and Segment Description, and their respective Count and Percentage, respectively.

Analysis of Numeric Variable



Graph 1.0 Graphical Representation of Survey Segment Type

The illustrated graph presents the distribution of each segment type, and its obtained count based on the number of respondents. The y-axis of the graph signifies the number of times a specific range of value in the x-axis appeared. The x-axis is the provided segment type based on the dataset. This implies that each bar of data point on the graph shows the frequency with different transpired values. As observed in the graph, the highest ranked is the university segment type which concludes that it has the most answered count based on the number of respondents accumulating to almost a thousand count among others. While the second highest ranked is the custom segment type that follows the gender segment type of 10 count. The least ranked observed in the graph that shares the same count value is both the mobile and web segment type.

Data Visualization

The data visualization encompasses statistical graphs which in this section are pie charts that represent a category's proportion. The pie chart symbolizes the analysis of the answers provided within the dataset from the code embedded.

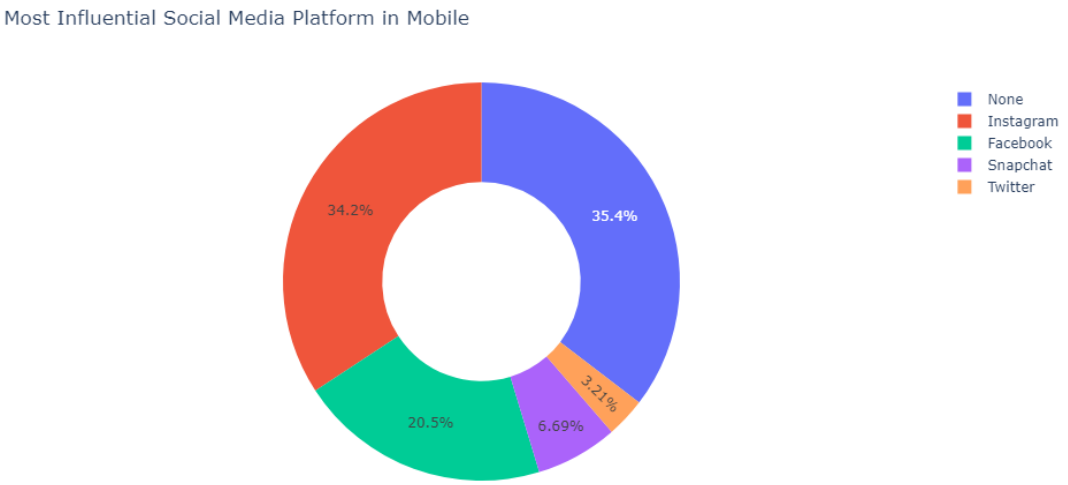


Illustration 2.0 Mobile Visualization Pie Chart

The visualization pie chart of the most influential social media platform in the segment type, mobile gathered its lowest percentage of 3.21% which is the twitter platform. While its highest ranging percentage is 35.4% which is none. Instagram obtained a close percentage result to none with 34.2%. This implies that people are less inclined to the social media platform, twitter as to their decision in social media habits and Instagram as the most influential social media platform in mobile.

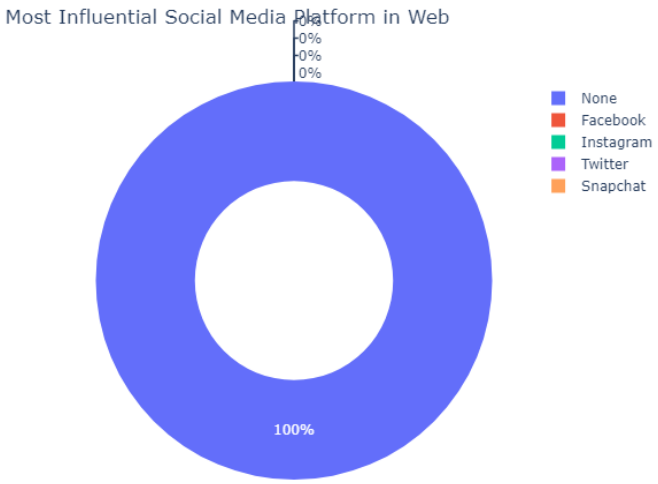


Illustration 2.1 Web Visualization Pie Chart

In this visualization pie chart of Web answers, it is highly evident that it accumulated 0% for social media platforms of Facebook, Instagram, Twitter, and Snapchat. While 100% of answers referred to none. This highlights that the segment type, web has zero accumulated answers.

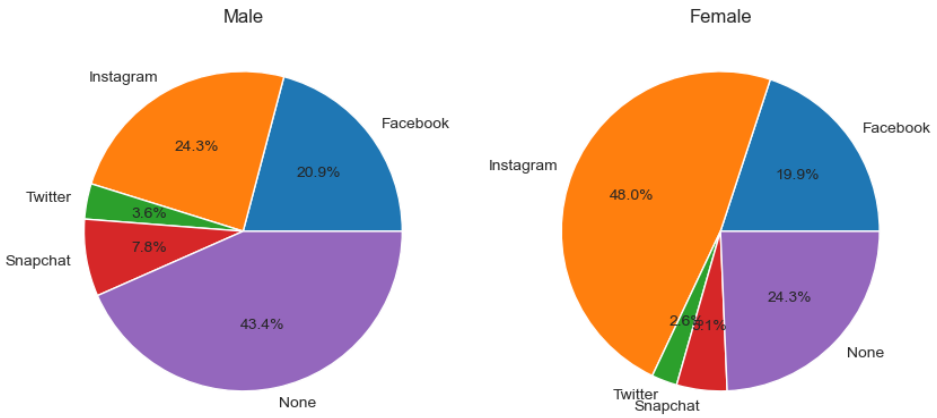


Illustration 2.2.1 Male and Female Visualization Pie Chart

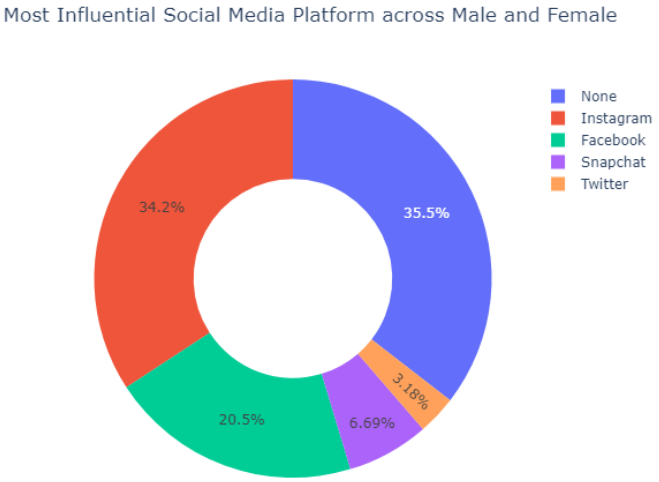


Illustration 2.2.2 Overall Gender Visualization Pie Chart

The visualization pie chart of the most influential social media platform in the segment type of gender obtained its highest percentage which is none with 35.5% and its lowest percentage is Twitter with 3.18%. This implies that the most influential social media platform next to the highest percentage is Instagram having 34.2%. With the given accumulated results, in the female sector, Instagram gained the highest ranked with 45.0%. This means that comparing the female and male genders, the females are more inclined to be influenced on their Instagram social media platform than male, since the male Instagram percentage garnered only 24.3%. Accordingly, the males were

not influence influenced in any social media platform towards their shopping habits as regards to the none answers as the highest ranked of 43.4%

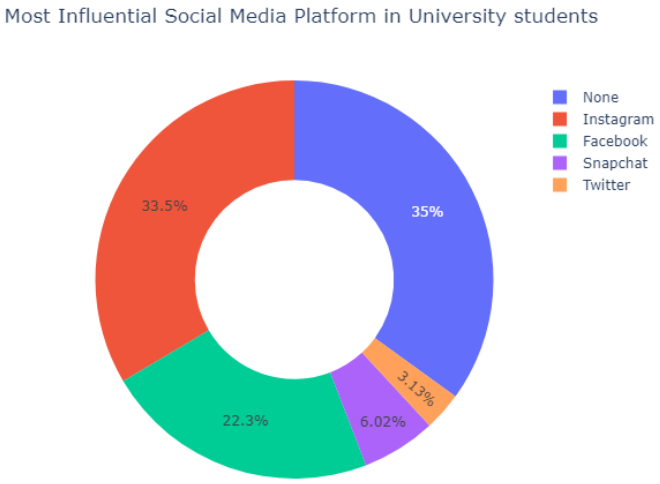


Illustration 2.3 University Visualization Pie Chart

The visualization for the segment type, university has gathered varying answers of different social media platforms that is viewed in the pie chart. The highest percentage gathered is none with 35% and the lowest percentage is Twitter with 3.13%. This implies that in each university the most elevated answer among every social media platform is neither of them. However, as presented in the pie chart, Instagram ranked second to the highest social media platform designated of each university.

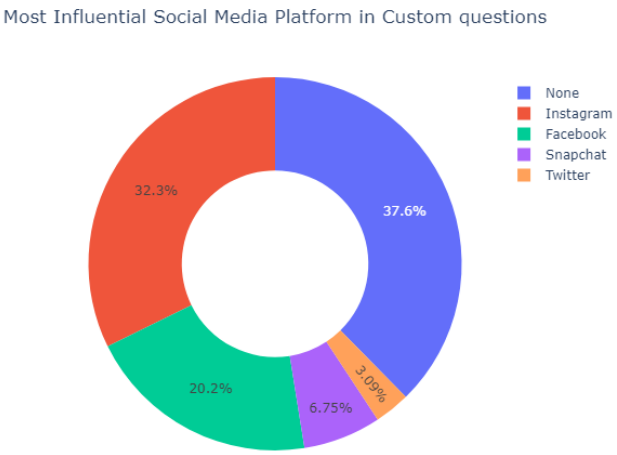


Illustration 2.4 Custom Questions Visualization Pie Chart

This visualization pie chart for custom question obtained 37.6% of none answers as its highest percentage and 3.09% of Twitter as its lowest percentage. This implies that the most influential social media platform is Instagram with 32.3% as the second highest percentage.

Numeric Bivariate Analysis

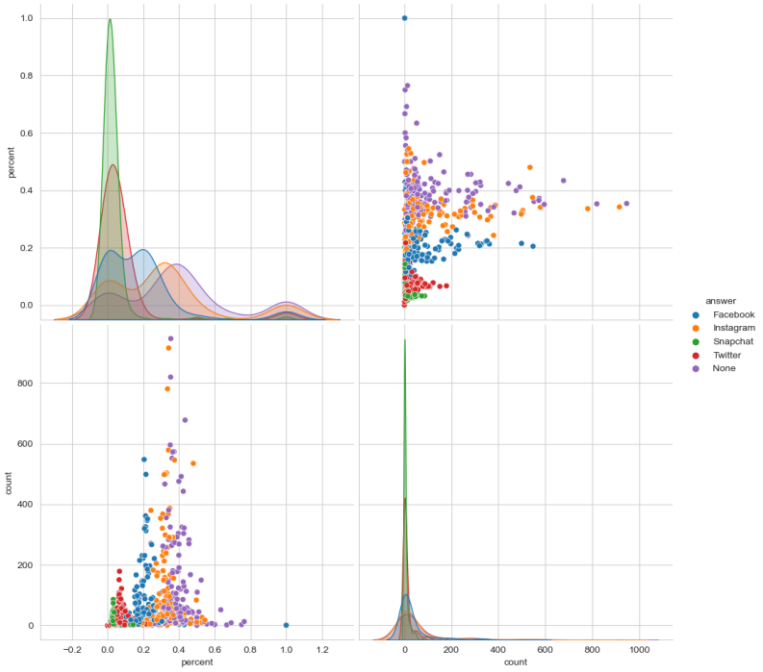


Illustration 3.0 Numeric Bivariate Analysis of Answer, Percentage, and its Count

Pairplots, a compelling data exploration tool, uncovers captivating revelations about the dataset's distributions and relationships. One striking finding is that a significant portion of users (over 500 in count) opted for the "None" option as their answer, accounting for no more than 0.8% of the total. Surprisingly, Instagram garnered a substantial number of votes, also surpassing 500, but with a slightly lower percentage, capping at 0.6%. Facebook follows closely behind, garnering votes within the range of 400 to 550, constituting up to 1.0% of the dataset. In contrast, Twitter and Snapchat registered the lowest count of votes, ranging from 0 to 0.2%, suggesting they were less popular choices among users. This comprehensive visualization provides valuable insights into the users' preferences and highlights the dominant and less favored options among the survey respondents.

III.CONCLUSION

In conclusion, based on the exploratory data analysis (EDA) conducted on the dataset of social influence on shopping behavior among Millennials and Gen Z members, several key findings and conclusions can be drawn: (1) The EDA revealed that the university segment type had the highest count of respondents, indicating that a significant number of participants were university students. The custom segment type followed while the gender segment type had the lowest count of respondents. (2) Among the different segment types, Instagram emerged as the most influential social media platform, with a considerable percentage of respondents across all segments selecting

it as their preferred platform (3) Twitter and Snapchat were the least favored social media platforms in terms of influencing shopping behavior. They obtained the lowest percentages across different segment types, indicating that they played a less significant role in shaping purchasing decisions among Millennials and Gen Z. (4) The "None" option, indicating that social media had no influence on their shopping habits, also obtained a notable percentage in each segment. This suggests that a considerable portion of Millennials and Gen Z members are not heavily influenced by social media when making shopping decisions.

List of References:

Diwan, R. G. M. M. (2022). Influence of social media marketing on buying decision making process of consumers. www.philstat.org. https://doi.org/10.17762/msea.v71i4.1138

Helper, A. (2022, December 5) Social influence on shopping. Kaggle. https://www.kaggle.com/code/toanvdst/data-analysis-for-social-influence-on-shopping

KPMG (2017). The Truth About Consumers Online. 2017 Global Online Consumer Report. assets.kpmg.com/content/dam/kpmg/xx/pdf/2017/01/the-truth-about-online-consumers.pdf

Shaw, J., & Shaw, J. (2023). The role of social media in Gen Z's purchasing decisions. Kadence. https://kadence.com/the-role-of-social-media-in-gen-zs-purchasing-decisions/

Smith, A. (2021, April 7). Social Media Use 2018: Demographics and Statistics | Pew Research Center. Pew Research Center: Internet, Science & Tech. https://www.pewresearch.org/internet/2018/03/01/social-media-use-in-2018/