Observing The Relationship Between Varying Levels of Social Media Use Among Adults and Mental Health

Jonah Kurian

DS 2001: Health Programming Practicum

Abstract

Social Media, although a quite recent innovation, has become almost instantly integrated into our lives. Social media has proved its merits as an effective way to express one's opinion and educate individuals. How does it affect us mentally? Recent studies almost all highlight the importance in moderation of Social Media for developing individuals (teenagers). The question that remains is: Do adults experience the same mental problems/tribulations due to social media in contrast to their younger couterparts? This study seeks to examine the relationship of social media use in adults across different categories of usage and age groups. Results indicate mixed findings across the different categories of usage and different age groups.

Background and Questions

General Background

An Overwhelming amount of evidence has come to light that exhibits that Social Media has drastic implications on developing individuals (teenagers and younger). One Study cited that "spending more than 3 hours on social media per day puts adolescents at a higher risk for mental health problems (1). Additionally, another study found that 94% of participants reported feeling troubled when they didn't have their phone. 80% were jealous when someone else used their phone, and 70% expected to feel depressed, panicked, and helpless if their phone went missing or they couldn't find it (1). This is especially concerning because this exhibits an almost physical addiction to social media. Social Media can also be extremely distracting , disruptive to one's sleep schedule, and may expose individuals to bullying, can be used as a medium for spreading rumors, and may harbor unrealistic views of other people's lives (2). To gauge an individual's current mental health status it's important to identify indicators of detriment in one’s mental health. The Mayo Clinic compiled a list of actions that are associated with declines in mental health including excessive fears, worries and anxieties and an inability to cope with daily problems and activities(3). Quality of life was also an alluring metric for further use in this study as it seems like it encompasses a variety of other tell-tale signs of decline in mental health such as prolonged depression and Social Withdrawal (3). A metric that is exacerbated as a result of social media detriment is the fear of missing out. The fear of missing out is a phenomenon when an individual has knowledge that others they identify closely with are having favorable experiences without them. This phenomenon is exacerbated by social media. In fact, FOMO has been linked to intensive social media use and is associated with lower mood and life satisfaction. This correlation with social media may have something to do with social media allowing people to actively showcase what they are doing at all times. A British study from 2018 linked social media to physical trauma as well. This physical trauma was described as a decrease, disruption, and a delay in sleep(4). Such Trauma is associated with higher rates of depression and anxiety.

Background To My Project

My project will use a dataset from the Pew Research Center called “August 7-September 16, 2013 - Pervasive Connectivity”. The Data was recorded from August 7th to September 2013 and was collected through a series of phone calls. The Calls were made to landline phones to prevent bias since cell phones are compatible with social media. The data was also collected in survey format: meaning there are a lot of similar values within columns. There were hundreds of columns of data so I decided to fixate on the columns that had the most to do with indicators of deterioration in mental health. I decided that q1, q2c, q2j, and q11d were the best indicators. Q1 was a metric that determined how they would rate their quality of life. Responses ranged from 1-5. 1 being “excellent”, 2 being “very good”, 3 being “good”, 4 being “fair”, and 5 being “poor”. Q2C had to do with levels of how frequent they were nervous/stressed. Responses ranged from 1 to 4. 1 being frequent, 2 being Sometimes, 3 being Hardly ever, and 4 being Never. Q2J had to do with being overwhelmed and Q11d had to do with the degree of which someone fears missing out. Although Q2J had the same scale as Q2C, Q11D’s scale went from Very well (1) , Somewhat(2), Not much(3), to Not at all (4). Each of these metrics correlate to signs of mental deterioration that I mentioned in my preliminary research section. The survey also has some details on the respondent such as their income and marital status, which gives insight to if the sample is representative of the United States' Population.

Questions

This project aims to understand the relationship between adults' use of social media and telltale signs of declines in mental health. It is apparent that as you get older physical and mental deterioration become harder and harder to combat. This leads to a series of questions.

1. As you enter older stages of adulthood do symptoms of social media become worsened?
2. How much does different levels of social media use affect Adults?

Importing Libraries and Reading Data

```
# Load Libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import os

# Changes Directory
os.getcwd()
os.chdir('C:\\Users\\jonah')
```

This chunk imports the necessary libraries for my project. It also changes my working directory because my data was in an external folder.

```
In [44]: # read and preview data
df = pd.read_csv('2013_Data.csv')
df.head()
```

Out[44]:

	psraid	sample	lang	state	cregion	usr	form	sex	q1	marital	...	birth_hisp	race	inc	ql1a	ql2hh	qc1	qc2hh	zipcode	weight	standwt
0	100004	1	1	36	1	S	2	1	2	2	...		1	6.0		2			14624	2.59375	0.847977
1	100019	1	1	12	3	S	2	2	3	3	...		1	3.0		3			34471	2.68750	0.878627
2	100035	1	1	36	1	S	1	2	3	1	...	1	1	3.0	2				13126	5.40625	1.767470
3	100040	1	1	21	3	R	2	2	3	1	...		1	NaN		2			40353	4.06250	1.328157
4	100041	1	1	34	1	S	2	2	3	5	...		1	3.0		3			7026	1.43750	0.469963

5 rows × 134 columns

This chunk reads in my csv file as a dataframe. This makes it easy to use pandas exclusive functions such as head. The function is essential for previewing data frame dimensions and formats.

Data Cleaning and Organizing

In [121]:
```
# Isolates Facebook Data
df_facebook = df[['inc','marital','sex','q1','q2c','q2j','q11d','sns2e','age']]
```

This chunk helps isolate facebook data. In this dataset the column that has to do with facebook usage is denoted by sns2e, thus why sns2e is included in the dataframe

In [122]:
```
# Removes Data with Blanks
IndexesToRemove = []
def clean_data(df, smcol):
    index = 0
    sm_index = 0
    for column in df[smcol]:
        if column == " ":
            IndexesToRemove.append(sm_index)
            sm_index += 1
        else:
            sm_index += 1
    WithoutDuplicates = []
    [WithoutDuplicates.append(x) for x in IndexesToRemove if x not in WithoutDuplicates]
    df = df.drop(df.index[WithoutDuplicates])
    IndexesToRemove.clear()
    return df
```

This chunk utilizes a for loop and a list comprehension to help remove any missing data in the facebook data frame.

In [124]:
```
# Makes Seperate Dataframes According to How Much a User uses Facebook
import collections
def snsCategories(df,val,sns):
    return df.loc[df[sns] == val]
several = snsCategories(df_fb,"1","sns2e")
onceaday = snsCategories(df_fb,"2","sns2e")
threetofive = snsCategories(df_fb,"3","sns2e")
onetotwo = snsCategories(df_fb,"4","sns2e")
everyfewweeks = snsCategories(df_fb,"5","sns2e")
lessoften = snsCategories(df_fb,"6","sns2e")
```

This function helps isolate each sns category. There are 6 calls to the function because there is a range of numbers from 1-6.

1 Several times a day

2 About once a day

3 3 to 5 days a week

4 1 to 2 days a week

5 Every few weeks (OR)

6 Less often

```
# Breaks down Each Usage Category by age demographics
early_several = several.query("age>=18" and "age<=34")
earlymiddle_several = several.query("age>=35" and "age<=44")
middle_several = several.query("age>=45" and "age<=64")
late_several = several.query("age>=65")
severalbyage = [early_several, earlymiddle_several, middle_several, late_several]
severalbyage[0]
```

```
# Breaks down Each Usage Category by age demographics
early_onceaday = onceaday.query("age>=18" and "age<=34")
earlymiddle_onceaday = onceaday.query("age>=35" and "age<=44")
middle_onceaday = onceaday.query("age>=45" and "age<=64")
late_onceaday = onceaday.query("age>=65")
onceadaybyage =[early_onceaday, earlymiddle_onceaday, middle_onceaday, late_onceaday]
onceadaybyage[1].describe()
```

```
# Breaks down Each Usage Category by age demographics
early_threetofive = threetofive.query("age>=18" and "age<=34")
earlymiddle_threetofive = threetofive.query("age>=35" and "age<=44")
middle_threetofive = threetofive.query("age>=45" and "age<=64")
late_threetofive = threetofive.query("age>=65")
threetofivebyage =[early_threetofive, earlymiddle_threetofive, middle_threetofive, late_threetofive]
threetofivebyage[0].describe()
```

```
# Breaks down Each Usage Category by age demographics
early_onetotwo = onetotwo.query("age>=18" and "age<=34")
earlymiddle_onetotwo = onetotwo.query("age>=35" and "age<=44")
middle_onetotwo = onetotwo.query("age>=45" and "age<=64")
late_onetotwo = onetotwo.query("age>=65")
onceadaybyage =[early_onetotwo, earlymiddle_onetotwo, middle_onetotwo, late_onetotwo]
onceadaybyage[0]
```

```
# Breaks down Each Usage Category by age demographics
early_everyfewweeks = everyfewweeks.query("age>=18" and "age<=34")
earlymiddle_everyfewweeks = everyfewweeks.query("age>=35" and "age<=44")
middle_everyfewweeks = everyfewweeks.query("age>=45" and "age<=64")
late_everyfewweeks = everyfewweeks.query("age>=65")
everyfewweeksbyage =[early_everyfewweeks, earlymiddle_everyfewweeks, middle_everyfewweeks, late_everyfewweeks]
everyfewweeksbyage[0]
```

These series of query functions help break down each data frame further into age driven dataframes.

```
# Compiled By Converting each query to csv and copying the mean values
ListOfLessOften = [[2.833333333,2.333333333,3.055555556,3],
                   [2.769230769,2.307692308,3.025641026,3.153846154],
                   [2.728571429,2.214285714,2.914285714,3.042857143],
                   [1.050030525,1.043907845,1.724632997,1.71344607]]
ListOfEveryFewWeeks = [[2.133333333,2.4,3.133333333,3.266666667],
                       [2.565217391,2.347826087,3.043478261,3.086956522],
                       [2.363636364,2.477272727,3.227272727,3.113636364],
                       [2.333333333,2.666666667,3.2,3]]
ListOfOneToTwo = [[2.851851852,1.888888889,2.814814815,2.777777778],
                  [2.865384615,1.846153846,2.634615385,2.865384615],
                  [2.694444444,2.101851852,2.87962963,2.935185185],
                  [2.53125,2.875,3.34375,3]]
ListOfThreeToFive = [[2.606060606,2.090909091,2.757575758,2.848484848],
                     [2.469387755,2.224489796,2.93877551,2.816326531],
                     [2.61627907,2.244186047,3.023255814,2.709302326],
                     [2.2,2.4,3.2,3]]
ListOfOnceADay = [[2.578947368,2.276315789,3.342105263,2.894736842],
                  [2.575221239,2.221238938,3.274336283,2.920353982],
                  [2.597014925,2.323383085,3.213930348,2.895522388],
                  [2.423076923,2.653846154,3.346153846,2.730769231]]
ListOfSeveralADay = [[2.459459459,2.283783784,3.067567568,2.641891892],
                     [2.479262673,2.253456221,3.059907834,2.728110599],
                     [2.542319749,2.250783699,3.059561129,2.796238245],
                     [2.542857143,2.285714286,3.171428571,3]]
```

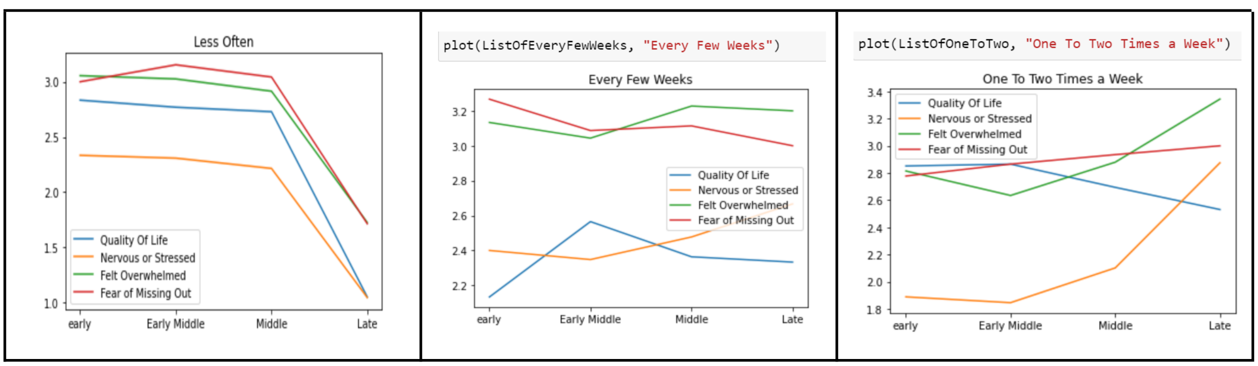
By using .describe() function on every query and converting the queries to csv format I was able to manually compile a mean value for each point of interest across all 4 age progressions for each category.

ListOfLessOften
Row One: Mean QOL in the Less Often Category across all 4 age groups
Row Two: Mean Nervousness/Stress in the Less Often Category across all 4 age groups
Row Three: Mean rate of difficulties/being overwhelmed in the Less Often Category across all 4 age groups
Row Four: Mean value of level of fear of missing out in the Less Often Category across all 4 age groups

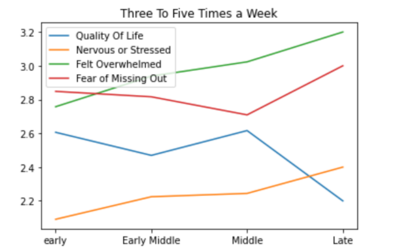
```
# Plots Data
def plot(category,title):
    LOC = ["early", "Early Middle", "Middle", "Late"]
    qol = [row[0] for row in category]
    nervous = [row[1] for row in category]
    difficulties = [row[2] for row in category]
    fomo = [row[3] for row in category]
    plt.plot(LOC,qol,label = "Quality Of Life")
    #qol 1 (excellent) - 5 (poor)
    plt.plot(LOC, nervous, label = "Nervous or Stressed")
    plt.plot(LOC, difficulties, label = "Felt Overwhelmed")
    plt.plot(LOC, fomo, label = "Fear of Missing Out")
    # 1 (frequent) - 5 (never)
    plt.legend()
    plt.title(title)
plot(ListOfLessOften, "Less Often")
```

The function plot takes in one of the 6 lists we compiled from the previous chunk and plots each age group against each point of interest on one plot for every usage category.

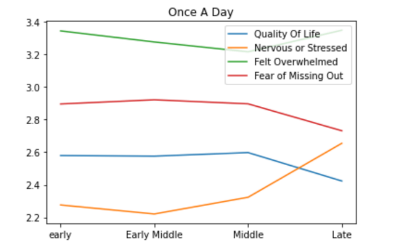
Results/Visualizations



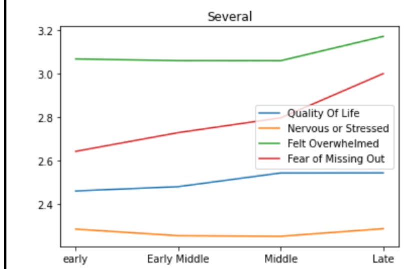
```
plot(ListOfThreeToFive, "Three To Five Times a Week")
```




```
plot(ListOfOnceADay, "Once A Day")
```




```
plot(ListOfSeveralADay, "Several")
```



The results ultimately showed the opposite of what was originally hypothesized. It turns out for almost every metric: individuals almost all felt as if things were getting better as time progressed. Quality of Life improved, Nervousness and Stress subsided, Overwhelming feelings were lessened, and the degree of fear of missing out dwindled. It is also worth acknowledging the “several” category which indicates a person is constantly and continually using social media. The several category had preliminary higher rates of nervousness, higher degrees of fear of missing out, and lower quality of life. Being overwhelmed was the only metric that didn’t align. As for the other categories the results were sometimes mixed/inconclusive.

Limitations

Categorical Data:

I think the survey Data format, especially the social media usage column abstracted the exact values of usage, making it hard to differentiate real usage rates in more broad categories like “several times a week”. This abstraction is huge because hypothetically someone could be checking social media anywhere from twice a day to twenty times a day because of how the categories are structured. Since the “several” category exhibited the most promising and consistent results it’s unfortunate not being able to observe the exact values progressed across each age demographic.

External Factors:

Marital Status

```
# Controlling For external Factors part 1
df_fb = clean_data(df_facebook,'sns2e')
print(df_fb['marital'].value_counts())
sum = df_fb['marital'].value_counts().sum()

# breakdown of Marital Status Within the Dataframe
print(str(round((496/sum), 2)) + " Are Married")
print(str(round((217/sum), 2)) + " Never Been Married")
print(str(round((95/sum), 2)) + " Divorced")
print(str(round((78/sum), 2)) + " Living With Partner")
print(str(round((48/sum), 2)) + " Widowed")
print(str(round((2/sum), 2)) + " NA")
```

```
1    496
6    217
3     95
2     78
5     48
4     24
9      2
Name: marital, dtype: int64
0.52 Are Married
0.23 Never Been Married
0.1 Divorced
0.08 Living With Partner
0.05 Widowed
0.0 NA
```

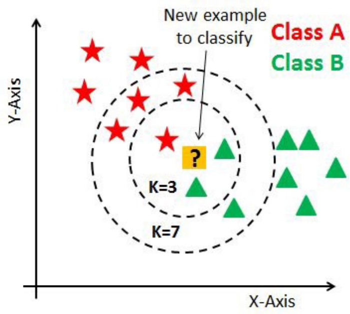
Marital Status is a key factor in one's mental health. Countries where marriage is emphasized, being married is related to lower depression, lower anxiety, lower suicide risk, and lower substance abuse, on average (5). Making sure the sample was representative of the entire population was essential in determining if the data was significant. 52% of the dataset was married at the time of the interview , while about 53% of adults are married in 2019 declining from a high of 67% in 1990. The dataset is more or less representative of the population. A model predicted that about 25% of Americans may never be married, which is flush with the actual dataset (7). I couldn't find percentages for the other categories of marital status, however. The lack of knowledge could be a potential source of error, but would have to be researched further.

Income

Income is another key factor in one's mental health. A 2011 study found that Low levels of household income is associated with several lifetime mental disorders, suicide attempts, and a reduction in household income is associated with increased risk for incident mental disorders (8). The problem was that in my dataset there was a lot of missing income data, approximately 15%. To combat this issue I used a K nearest Neighbors Algorithm.

K Nearest Neighbor Algorithm

A k-nearest neighbor algorithm is a supervised machine learning algorithm. A supervised learning algorithm essentially predicts an output based on input/output pairs. In other words it maps incomplete data based on previous complete data. In a k nearest neighbor algorithm an arbitrary number K is selected. The k closest neighbors to a specific data point will ultimately serve as a reference point for prediction. The K closest neighbors in terms of euclidean distance in turn can help in classification and regression methods. The image below shows how the number of neighbors can affect the classification of a point. When k=3 the point seems to be leaning towards class B while the opposite is true for k=7. The chunk Below shows how I initialized the algorithm I used a K value of two in this instance. When I viewed my updated data frame the missing values had been replaced.




```python
# Replaces Missing Income Values Using A K Nearest Nei
# the average income is to the natinal average.
from sklearn.impute import KNNImputer
def ImputedLessOften(x):
    impute_knn = KNNImputer(n_neighbors = 2)
    lessoften = impute_knn.fit_transform(x)
    return lessoften

def ImputedOnceADay(x):
    impute_knn = KNNImputer(n_neighbors = 2)
    onceaday = impute_knn.fit_transform(x)
    return onceaday

def ImputedOneToTwo(x):
    impute_knn = KNNImputer(n_neighbors = 2)
    onetotwo = impute_knn.fit_transform(x)
    return onetotwo

def ImputedThreeToFive(x):
    impute_knn = KNNImputer(n_neighbors = 2)
    threetofive = impute_knn.fit_transform(x)
    return threetofive

def ImputedSeveral(x):
    impute_knn = KNNImputer(n_neighbors = 2)
    several = impute_knn.fit_transform(x)
```

```python
def getIncome(category):
    income = category[:, 1]
    print(np.sum(income)/len(income))
    return (np.sum(income)/len(income))

# representative of $30000
AverageIncomeCategory = 3
several_income = float(getIncome(s))
onceaday_income = float(getIncome(oad))
onetotwo_income = float(getIncome(ott))
threetofive_income = float(getIncome(ttf))
lessoften_income = float(getIncome(lo))
everyfewweeks_income = float(getIncome(efw))
```

Sums the income of every category

```
def Closeness(AIC, CI):
    return CI / AIC

ClosenessSeveral = Closeness(AverageIncomeCategory,several_income)
ClosenessOAD = Closeness(AverageIncomeCategory,onceaday_income)
ClosenessOTT = Closeness(AverageIncomeCategory,onetotwo_income)
ClosenessTTF = Closeness(AverageIncomeCategory,threetofive_income)
ClosenessLO = Closeness(AverageIncomeCategory,lessoften_income)
ClosenessEFW = Closeness(AverageIncomeCategory, everyfewweeks_income)
print("Income")
print("Several Closeness To Average: " + str(ClosenessSeveral))
print("Once A Day Closeness To Average: " + str(ClosenessOAD))
print("Once To Two Day's Closeness To Average: " + str(ClosenessOTT))
print("Three To Five Day's Closeness To Average: " + str(ClosenessTTF))
print("Less Often's Closeness To Average: " + str(ClosenessLO))
print("Every Few Week's Closeness To Average: " + str(ClosenessEFW))
```

Calculates the closeness of the dataset’s categories to the national average

```
2.8446327683615817
2.6343612334801763
2.65
2.75
2.4578313253012047
2.4237288135593222
Income
Several Closeness To Average: 0.9482109227871939
Once A Day Closeness To Average: 0.8781204111600588
Once To Two Day's Closeness To Average: 0.8833333333333333
Three To Five Day's Closeness To Average: 0.9166666666666666
Less Often's Closeness To Average: 0.8192771084337349
Every Few Week's Closeness To Average: 0.8079096045197741
```

Conclusion

In this project I seeked to understand the relationship between varying levels of social media use among adults and mental health. There is convincing evidence that adult individuals that use social media several times a day have higher preliminary rates of nervousness/stress, have poorer quality of life, have higher rates of fomo, and are more likely to feel overwhelmed. There seems to be no conclusive evidence that an increase in age exacerbates consequences for mental health. The several categories may be telling because it’s the closest category to excessive usage of social media. Mental health consequences due to social media are most commonly observed in individuals that excessively use social media. Throughout the study, we confirmed that the dataset’s income and marital status was more or less a representative sample of the population.

References

1. 40+ Frightening Social Media and Mental Health Statistics — Etactics
2. Teens and social media use: What's the impact? - Mayo Clinic
3. Mental Illness and the Family: Recognizing Warning Signs and How to Cope | Mental Health America (mhanational.org)
4. Here's How Social Media Affects Your Mental Health | McLean Hospital
5. Mental Health and Marital Status - Spiker - - Major Reference Works - Wiley Online Library
6. Rising Share of U.S. Adults Are Living Without a Spouse or Partner | Pew Research Center
7. The Share of Never-Married Americans Has Reached a New High | Institute for Family Studies (ifstudies.org)
8. Relationship between household income and mental disorders: findings from a population-based longitudinal study - PubMed (nih.gov)