Michael Reilly

Professor Dugas

CS-581-WS

12/9/20

I pledge my honor that I have abided by the Stevens Honor System.

Purpose: With my chosen analysis I will be able to illustrate what may and may not affect how much someone uses the Internet. I am interested in showing this in order to see what kind of outside factors may actually come into effect for someone to use the Internet on a different basis as someone who has to use it almost constantly. This is important as it can either illustrate how someone uses the Internet or instead showcase how commonplace Internet Usage has become over time.

Input: To execute the program through command line: python3 Reilly_HW9.py

Output: The output includes a variety of things. It starts by making a Plots folder to store all the charts and graphs that will be created. Also there are nine different charts made within the program's execution that are then saved in the created Plots folder as pngs to see as the code executes itself. Then the value counts for all the charts and graphs created are printed out since, particularly the pie charts, do not include these counts. There are also correlation coefficients and short summaries about them in regards to each pair of data compared to Internet Usage which are used for the relationship analysis between variables and printed out onto the screen.

What the program does: During the program's execution the Pew_Survey.csv file is read into the program so it's data can be used. Then the program creates a Plots folder to store the created graphs and charts. Then for each variable analyzed it creates a chart that best represents the data, gives the value counts for every variable except for age due to it not being helpful in that case,

and then finds the correlation coefficient between them which can be used to make an analysis about the relationship between variables.

Results: The first result done was whether people will use the Internet more if they have Social Media versus if they do not. My hypothesis is that if someone has Social Media they are more likely to use the Internet often due to the time sink that Social Media can be.

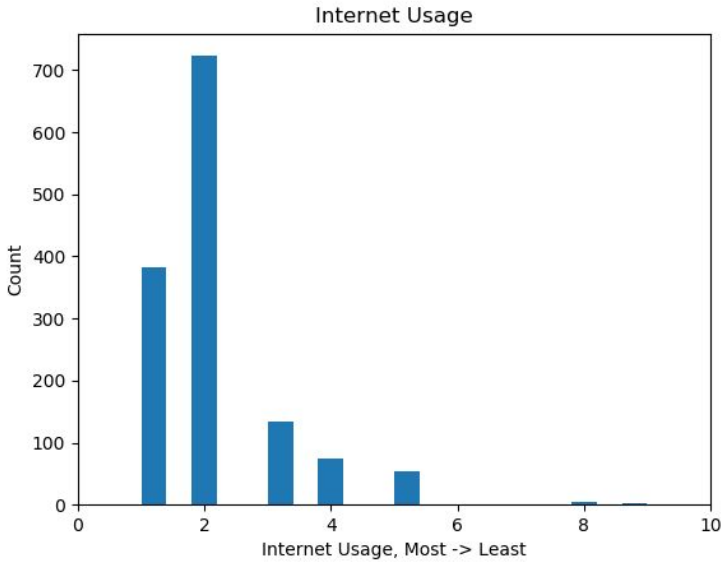

Figure 1: Histogram of Internet Usage

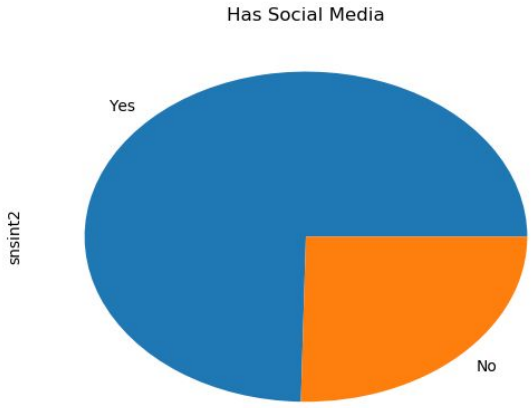

Figure 2: Pie Chart of those Who Do and Don't have Social Media

Based on the charts it can be seen that a large majority of these respondents use the Internet often, and also a majority of the respondents have social media. Then taking the correlation coefficient between them it was found that the correlation between them was about 29.5%, which indicates a slight positive correlation between Internet usage and having social media. Based on that it can be concluded that people are slightly more likely to use the Internet more if they have social media. That conclusion supports my initial hypothesis.

The next variable analyzed was Region of the US. My hypothesis is that people all over the country need the Internet so it is unlikely there is a correlation between them.

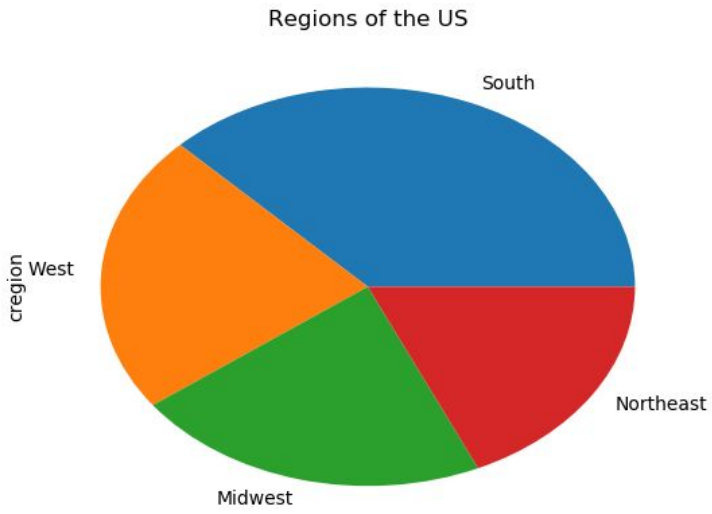

Figure 3: Distribution of Respondents throughout the Regions of the US

The respondents are fairly evenly split between the regions of the US. Then taking the correlation coefficient between Internet Usage and Regions of the US it was found it was about -0.6%, which indicates there is no relationship between Internet usage and the regions of the US people are from, which supports my initial hypothesis.

The next variable analyzed was Gender. My hypothesis is there is no correlation between them as people of all Genders need the Internet.

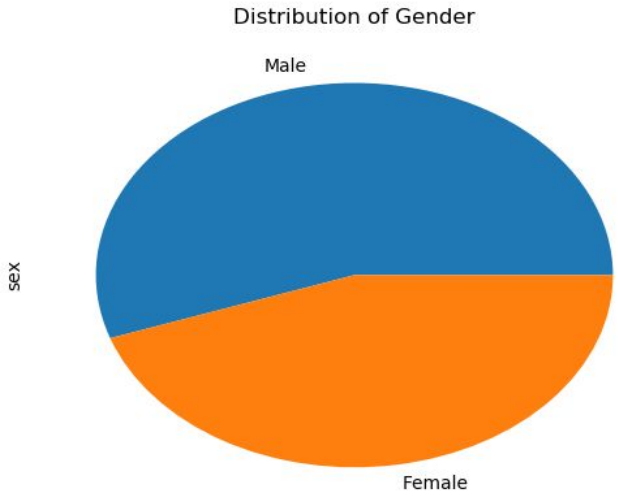

Figure 4: Gender Distribution of Respondents

There are slightly more men than women who responded to the survey, but this will not skew our data analysis. Then taking the correlation coefficient between Internet Usage and Gender it was found it was about 7%, which statistically is not significant enough to indicate a relationship between Internet Usage and Gender, which supports my initial hypothesis.

The next variable analyzed was Age. My hypothesis is that the younger people are the more likely they are to use the Internet often. As previously concluded, people use the Internet more when they have social media, and younger people are more likely to use the Internet than older people so I'd assume they'd also use the Internet more.

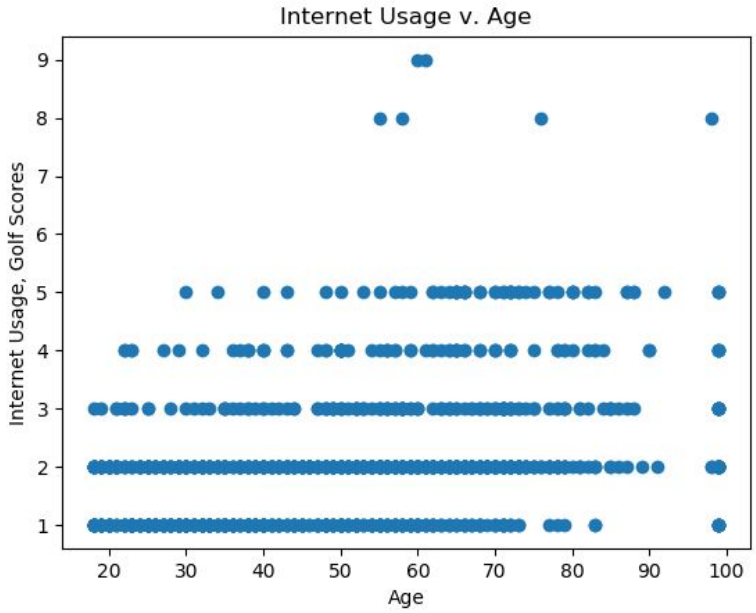

Figure 5: ScatterPlot of Internet Usage v. Age, 1=more internet usage, not less, think golf scores

As can be seen by the scatter plot there is a higher concentration of points in the bottom left than bottom right, indicating more young people use the Internet constantly than older people. As you move up to using the Internet less and less the points start to shift to the right, indicating older people use the Internet less than younger people. To confirm this the correlation coefficient between Internet usage and age was taken and was found to be about 33%, which indicates a slight positive correlation between Internet usage and age. This means that the younger people are, they are slightly more likely to use the Internet constantly, which supports my initial hypothesis.

The next variable analyzed was Marital Status. My hypothesis is that marital status has no effect on how much someone uses the Internet.

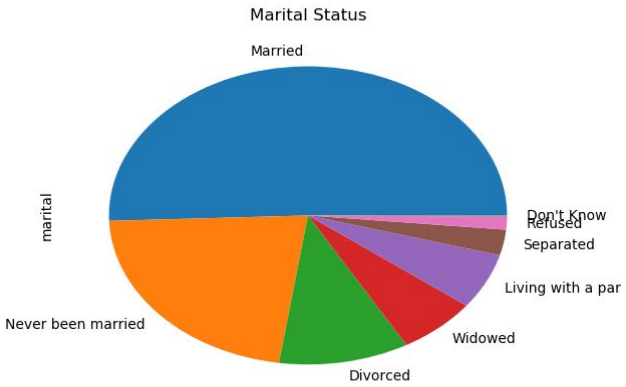

Figure 6: Marital Status Distribution of Respondents

A majority of the respondents are currently married, so likely the age of the respondents is overall a bit older than college students or new college graduates. Taking the correlation coefficient between Internet usage and marital status it was found to be -0.5%, which indicates there is no relationship between the Internet usage of a person and their marital status. That supports my initial hypothesis.

The next variable analyzed was Employment Status. My hypothesis is that employment status has no effect on how much someone uses the Internet.

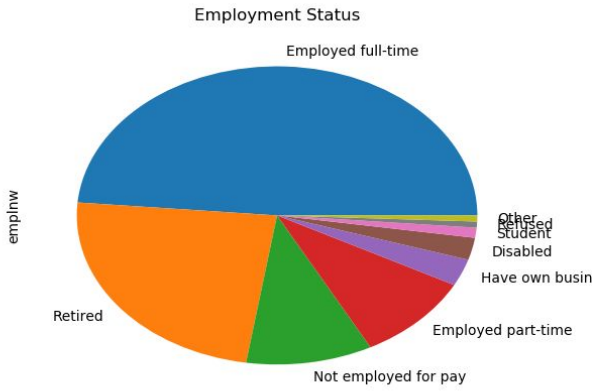

Figure 7: Employment Status Distribution of Respondents

A majority of the respondents were either employed full-time or retired, likely meaning the age distribution is skewed a bit older than it would be if they were mostly students or part-time employees, which may indicate slightly lower Internet usage overall due to the deduction that older people use the Internet less frequently than younger people. This lines up well with the marital statuses of the respondents too in that the overall age distribution is skewed a bit older. After taking the coefficient correlation it was found the correlation between Internet usage and Employment status is about 4%, which statistically is not significant enough to indicate a relationship between Internet usage and employment status. As a result, my initial hypothesis is supported.

The next variable analyzed was Education Level. My hypothesis is that education level has no effect on a person's Internet usage.

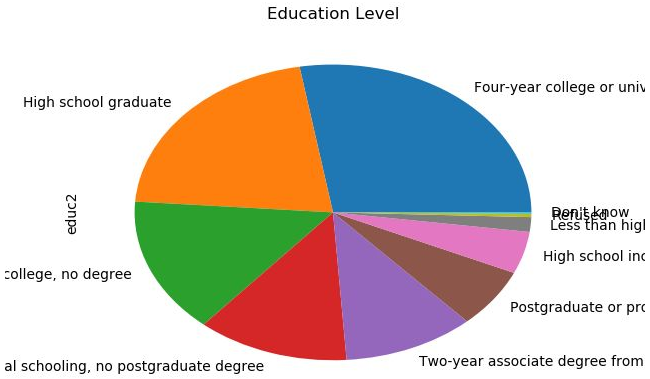

Figure 8: Education Level Distribution of Respondents

A majority of the respondents either have done four years of college and got a bachelor's degree, didn't go to college but finished high school, or have done some college but with no degree from it. Based on my initial hypothesis this won't affect the data at all. Then I took the correlation coefficient between Internet usage and education level which is -3.5% which

statistically means there is no significant relationship between Internet usage and education level. This supports my initial hypothesis.

The final variable analyzed was Political Affiliation. My hypothesis is that political affiliation will not have any effect on Internet Usage.

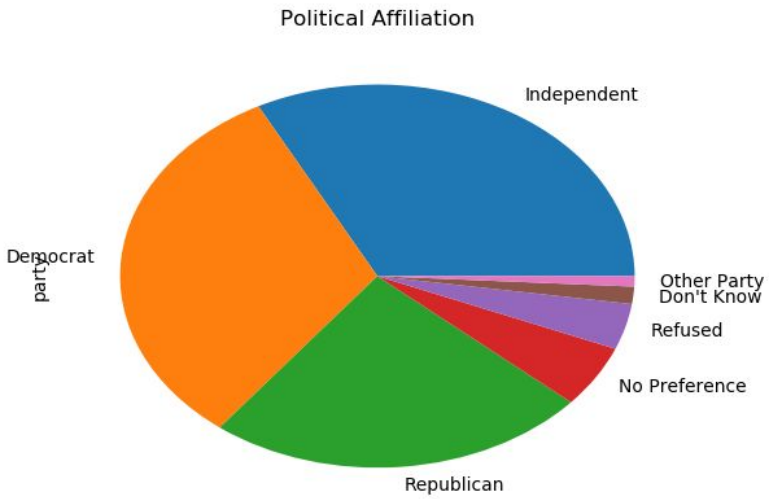

Figure 9: Political Distribution of Respondents

A majority of the respondents are either Independents, Democrats, or Republicans unsurprisingly. This should have no effect on the data based on my initial hypothesis. I then took the correlation coefficient between Internet usage and political affiliation which ended up being 6%, which statistically means there is no significant relationship between Internet usage and political affiliation. This also supports the initial hypothesis given.

In summary, the only variables of those studied that have any correlation with someone's Internet usage is whether they have social media and their age. All other studied variables such as region within the US, gender, marital status, employment status, education level, and political affiliation all have no statistically significant correlation with respect to their Internet Usage.

This leads to the conclusion that consistent Internet usage has become extremely commonplace among people, with the only real exceptions being that social media encourages people to use it more, and younger people who have grown up with Internet use it more frequently than older people who did not have it when they were younger.

Additional Information: This was studied in hopes the data would show something unexpected, such as gender having a moderate effect on someone's Internet usage, but sadly that didn’t happen. Otherwise there is no additional information worth adding.