Assignment 9 – Report

Purpose

The purpose of this assignment was to analyze social network data that was obtained through a Pew Research Survey conducted in January/February 2019. The dataset provided by Pew Research is over 1300 lines long, and contains survey responses for users of the internet and social media. The survey provides a plethora of information that I was able to analyze and present through the use of Python and the Matplotlib library. Going through the data and obtaining insight into how different demographics use technology, the internet, and social media was fascinating, and it was quite interesting to see the story the data told me about different demographics and their relationship with technology. The purpose of the assignment was to study recent survey data on social network use, as well as continue to utilize python and programming to assist us in this endeavor. Additionally, the assignment allowed for us to become familiar with and utilize a new library, Matplotlib, which was incredibly useful for presenting complex data graphically.

Input

The input of information was the Pew Research dataset provided which was obtained through an extensive survey conducted at the beginning of 2019. The dataset format is a csv file, which makes it easy to easily read the data into python, create data structures and perform meaningful analysis. To run the program, just navigate to the folder/path using the terminal window and execute python Marks.py. Ensure that the "Pew_Survey.csv" file is saved in the same path for the python program to access.

Output

The program accesses the csv file, creating a comprehensive list with all of the data from respondents. From this list, various calls are performed to access the data of interest to me in this analysis. Two types of output are provided from the program including output to the terminal window, as well as output of graphical information in the form of pie charts administered by Matplotlib. The first information output to the terminal window are the total number of users of each social network. The second piece of information that I output from the program is the frequency with which the entire dataset use the internet, also in list/tabular format. The next pieces of information output to the terminal window are a breakdown of social media use by generation. I look at each generation, including generation z, millennials, generation x, baby boomers, and the post war generation. I was able to determine the age range of each age group or generation by the age response in the survey, and used Beresford Research as a guide of ages/years for each generation. Next, I thought it would be interesting to look at social media site use by income demographic levels, and determined which network is most used by each income bracket. I also found it interesting to look at the geographic data to see if certain networks were more or less popular in different parts of the country, and output the most used and least used social network for each geographic region. Finally, on the terminal, I also have the program output by age generation, the internet frequency usage, based off the category responses from the survey.

The second type of output of my program is graphical information that I am able to output from my computations with the help of the Matplotlib library. The plots that I output are all pie charts, and first include the overall social media use, and overall internet use frequency of the dataset as a whole by percent. Next, I graphically present a pie graph for each age generation to show their social media use graphically, and present this as percents as well. Finally, for each age generation, I present in a pie chart the internet use frequency, showing the percent of each category in responses.

What the Program Does

First, my program accesses the "Pew_Survey.csv" file, and reads each line/respondent into python as a list. Each of these sublists make up a master list that I utilize throughout the program to access data to perform my analysis.

After creating my master list, I use a for loop to access index 4, 5, 6, 7, 8, 9, 10, 11, and 12, for each respondents' answer to their use of Twitter, Instagram, Facebook, Snapchat, Youtube, Whatsapp, Pinterest, LinkedIn, and Reddit, respectively. These responses are totaled up, and presented both to the terminal and graphically. Next, the program computes totals for frequency of internet use for the population as a whole in a similar manner, and presents to the terminal and graphically. The next computation of the program is using a for loop to look at each age generation and total up use of social networks for each. Conditional statements are used to logically determine which age group each respondent belongs to. Income information is analyzed next to see if there is a correlation of income levels and particular social networking sites, and this again is done through the use of a for loop accessing information from the master list, and totaling up use by using an incrementing counter. Finally, the program accesses the master list of data from respondents to determine the most and least used social networks for each of the four geographical regions, and also computes internet usage for each age generation.

Results

I found these survey results to be fascinating and thoroughly enjoyed analyzing the information from different perspectives to see how different demographics and socioeconomic groups interface with technology, and which social media networks they find most useful and use the most and least. I first thought it would be useful to look at the data from the entire population perspective to see which networks were being used the most overall, and how frequently the population was using the internet. From there, I wanted to compare the behaviors of different groups by age, as their experiences with technology are so different. A couple of things that were surprising to me were that generation z and millennials had just as many respondents that use the internet only several times per day as respondents that used it almost constantly. I would have thought the vast majority of both of these groups would have responded that they were using the internet and social media almost constantly. I also found it interesting that across the board, and at each generation level, Youtube is the most used social media network. However, if you look at the data from income levels, the lower income brackets actually use Facebook the most, while the higher income brackets, starting at $40,000 annual income and up actually use Youtube the most. Also, I was surprised by the geographic information I found. I would have thought there would have been more contrast by social network use by geography, however my findings were that in every region, the most used social network was Youtube, and the least used social network was Reddit.

Additional Information

This project was quite challenging due to the amount of information that respondents provided in the survey and the volume of the file. There were various ways to look at the information, and I found it to be fun trying to hypothesize or come up with expectations on how I thought the information may play out or the story the data would tell, and there were many times, as mentioned/reflected in my results where the information was a little different from what I expected. I enjoyed using the Matplotlib library, and think it will be very beneficial having that knowledge in future projects or courses.