# Project - Phase IV

SEAL Team 18

Created By:
Doanh Pham
Jake Sikorski
Nate Cherrier
Harsh Tagotra

## Initial Proposal:

How are factors such as age, education, and social media use related to a person's opinions of social media?

## Expected Outcome:

We think that we will find certain demographics of people have measurably different opinions on social media and the internet. The features of individuals used to segregate opinions were namely Age, Gender, Education, Income, Party (Democrat or Republic) and Income.
It is the reasoning that the level of **education** has a correlation with the opinions of social media. Those who are highly educated are more likely to be skeptical of news articles circulated on social media and more likely to fact check. This would give them a negative view of social media as it promotes the proliferation of fake news.
**Age** would also be another factor that influences the opinion of social media by the person. Older people may find social media more effective, allowing them to connect with old childhood and high school friends as social media helps facilitate this. Younger people see more of the profound effects social media can have in the proliferation of fake news and how it harms society, so they may be a tad bit more cautious but young people are still the bulk of social media users and some of the earlier adopters, so not much parity between both ends of the spectrum is expected.
**Social media use** may also influence sentiment. Those who use social media more would tend to have a more positive sentiment, as they see value in social media and believe it provides a benefit to them and society. Otherwise, why would you use a product that much if you don't believe it is a good impact on you or society? However, low parity is predicted here between

both ends, as one that uses social media all the time could think overall it is a negative effect on society.

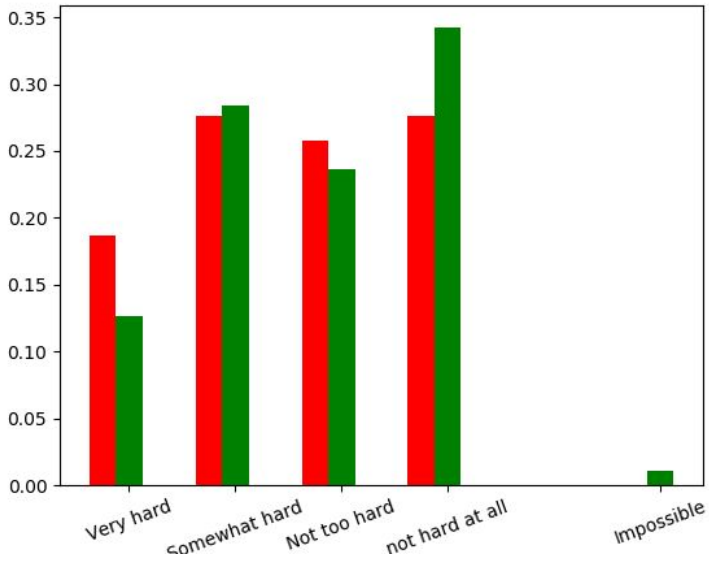
## Observed Outcomes:

The opinions in the dataset were based on the following (most informative) questions asked in the survey:

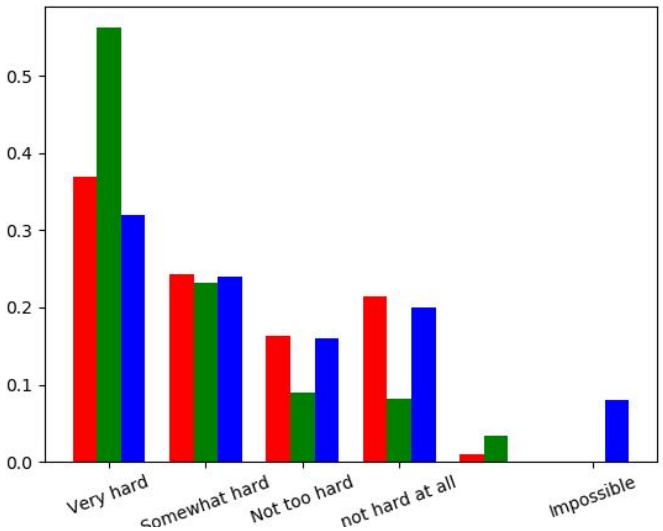
1) Question 5C: How difficult would it be to give up the internet from your life?
2) Question 5D: How difficult would it be to give up social media from your life?
3) Question 11: Is the internet objectively good or bad?

The following features ultimately were responsible for the most conflict of opinions:

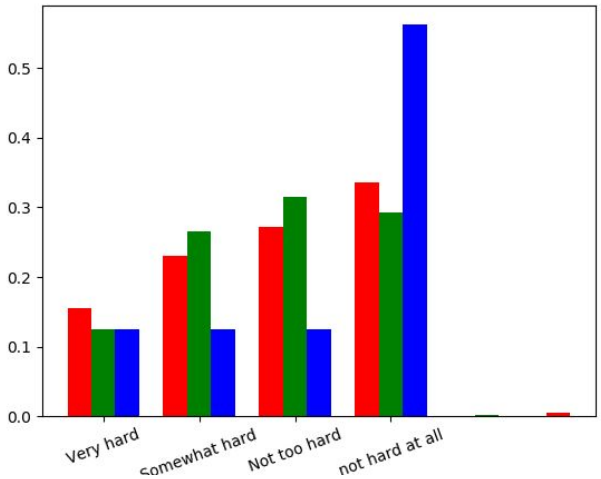
- Age [**Young**, **Old**]
  5C: Trends were more or less the same
  5D: A greater percentage of young people thought it was harder to give up social media from their lives




  11: Trends were the same
- Education: [**High School**, **College**, **Post Graduate**]
  5C: A greater percentage of individuals with some sort of college education thought it would be harder to give up the internet

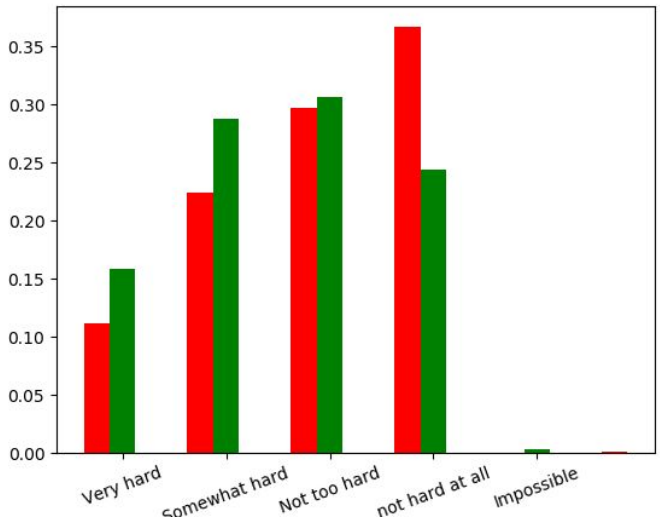



5D: A greater percentage thought that it would be not so hard to give up social media

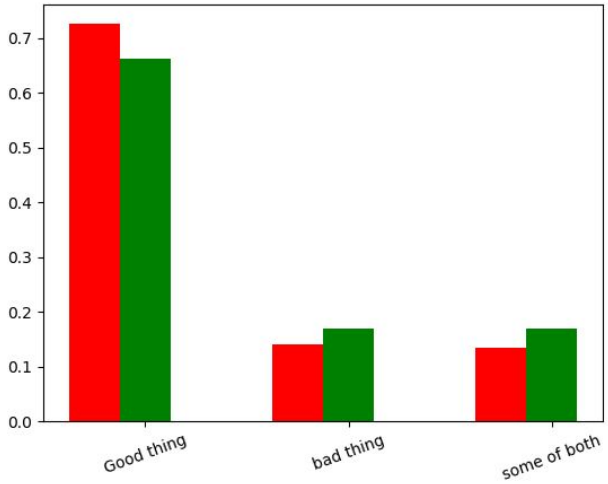



11: Trends were more or less the same

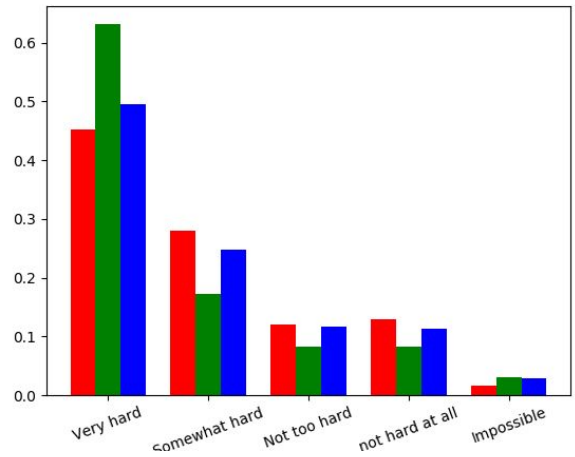
- Gender: [**Male**, **Female**]
  5C: Trends were more or less the same
  5D: A greater percentage of females believed i that giving up social media would be harder



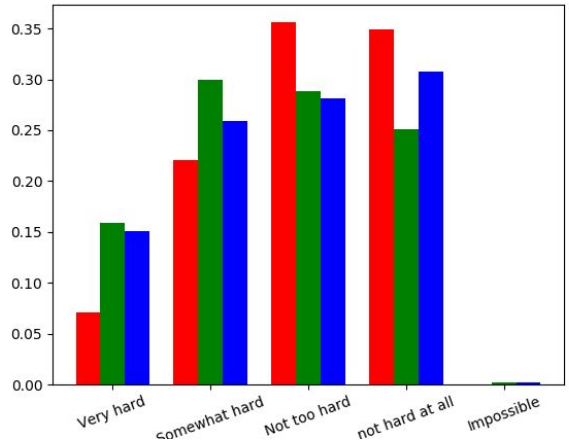

11: A greater percentage of males thought that the internet was objectively a good thing




- Party [**Republic**, **Democrat**, **Independent**]
  5C: A greater percentage of individuals believed that it would be very hard to give up the internet




5D: It was found to be way harder for Democrats and Independent individuals to give up social media




11: A lot more democrats thought the internet was objectively a bad thing

# Overall Project Process:

## Phase 1

Selecting the database as simple as we all agreed that performing an analysis on opinions would give us a lot of information due to the rather polarizing opinions that people have about internet usage. The fact that all our data was numerical also helped since performing calculations on numerical data is generally easier.

### Phase 2

Coming up with table representations of a database with more than 40 features was a little confusing at first. After reading the documentation of our data, we realized that a majority of the features were questions with consistent numeric entries representing the responses. It made sense to include all opinions in just one table with a reference to all persons in its own respective table. To facilitate segregation, separate tables containing information about the demographics, social media usage and device usage were created, all with a similar reference to persons in the Person table.

### Phase 3

Once we had a concrete understanding of our data, referring to our ER diagram and building up the tables in our database was no hassle. No major changes were made to our table definitions.

### Phase 4

Extensive use of the matplotlib, numpy and pandas libraries were used to perform qualitative analysis to figure out the difference in opinions based on demographics. This included slicing data and printing out the percentages of each opinion. Opinion proportions for all demographic categories were printed and the one that separated the data the best was picked out. We also had to add a text representation of all the responses in the tables.

## Team Contributions:

### Doanh Pham

- Planning for all phases
- Creating the skeleton of the project
- Loading person, devices, social media, and person table
- Creating the poster

### Nate Cherrier

- Creating key tables to represent the text meanings of each numerical response
- Planning database structure
- Reloading devices table
- Planning for all phases

### Jake Sikorski

- Planning for all phases

- Loading the demographics table
- Responsible for most of the ER diagram, expected outcome, and project proposal
- Finding the dataset

## Harsh Tagotra

- Planning for all phases
- Loading Opinions and Questions table
- Categorical analysis using python
- Report outcomes and project process