# Project Design Write-up and Approval Document

Anne Robison

DS\_SM\_15

June 20, 2017

**### Project Problem and Hypothesis**

I think there are many questions that could be addressed by this data. I could use some guidance about which of these would be reasonable for me to do.

Some possibilities are:

1. Classifying respondents as users of specific methods of entertainment consumption based on variables such as
   * demographics,
   * usage levels of other methods of consumption,
   * ownership of hardware
2. Predicting choice of method based on circumstances of the occasion, such as:
   * Group or solo watching (including who is in group)
   * Day of Week/ Time of day
   * Location
   * Other factors
3. Predicting cable cord cutting based on variables such as
   * demographics,
   * usage levels of other methods of consumption,
   * ownership of hardware
4. Predicting usage levels of consumption methods or total consumption based on   
   variables such as
   * demographics,
   * usage levels of other methods of consumption,
   * ownership of hardware
5. Predicting which streaming service is used based on variables such as:
   * demographics,
   * usage levels of other methods of consumption,
   * ownership of hardware

These are classification problems, except for #4.

Having a clearer understanding of what motivates consumers to pick a certain method or to be heavy or light users would inform choices in advertising messages or media.

Conventional wisdom is pretty convinced that demographics are the most important driver of entertainment consumption. I'm hoping to find that something else is important.

**### Datasets**

Survey data collected via online interviews conducted against a general population sample from a panel provider. 2364 interviews among general pop sample of Americans 12-74, with an augment of 340 respondents to fill out key usage segments (e.g., streamers, digital content buyers, etc.)

Excel file with details of the questionnaire (Still working on it)

**### Domain knowledge**

I have 20 years of experience with survey research in the entertainment space, with 6 years specifically working with home entertainment issues. My experience with the category will give me a better sense of when results "make sense" or not. Our prior analysis of this data was very basic. We didn't try to find any relationships within the data.

I know that data science is being used for entertainment but it tends to be done within a company and results are not shared publicly.

**### Project Concerns**

*\* What questions do you have about your project? What are you not sure you quite yet understand?*

I have so many variables that I'm afraid that using all of them to see which are important will take a long time and not checking all of them will mean that I could be missing something important. Also, not sure which models would be most useful to apply.

*\* What are the assumptions and caveats to the problem?*

*\* What data do you not have access to but wish you had?*

*\* What is already implied about the observations in your data set?*

Everyone is tech savvy enough to have Internet and be willing to participate in survey panels. There's some risk of non-response bias.

*\* What are the risks to the project?*

*\* What's the cost of your model being wrong?*

Nothing really. The industry will continue doing everything the way they've done it till now.

*\* What's the benefit of your model being right?*

Better and cheaper targeting and better messaging to consumers in terms of prompting them to use higher margin consumption options.

\* Is any of the data incorrect? Could it be incorrect?

I'm not sure how to answer any of the questions below…

### Outcomes

\* What do you expect the output to look like?

\* What does your target audience expect the output to look like?

\* What gain do you expect from your most important feature on its own?

\* How complicated does your model have to be?

\* How successful does your project have to be in order to be considered a "success"?

\* What will you do if the project is a bust (this happens! but it shouldn't here)?