## Section 1: Define the problem

- What is the question your project is attempting to answer?
- Where does your project fit within the broader conversation/controversy surrounding your topic?

The inspiration for our project came from a dissertation by Uyen Nguyen. The paper proposed the question: How has the proliferation of broadband affected US consumer retail behavior from 2004 to 2018?

In response to this Nguyen collected and analyzed data on 7 metrics of consumer behavior, including: (1) trip frequency, (2) total spending, (3) unique chains visited, (4) unique brands purchased, (5) prices, (6) price dispersion, (7) price elasticities of demand.

The broader topic focuses on a prominent discussion at the time: the retail apocalypse.

Our findings add to a growing body of research on the impact of technology on the retail industry, and may help to inform discussions around the future of brick-and-mortar retail in an increasingly digital landscape.

## Section 2: What success would look like

- What are you trying to accomplish?
- What is the outcome you hope to achieve?

- Our goal was to create a R-shiny web app to help interested parties to easily visualize and explore her findings.
- The final app would be interactive, with customizability options for time and percentile for the metrics.
- Relevant parties: policymakers, researchers, and industry stakeholders
- Ultimately, the aim is to inform policies and strategies that can help bridge the digital divide and increase broadband access and adoption, thereby benefiting both consumers and the economy as a whole.

## Section 3: The Data and Data Cleaning

Where does the data come from?

- The broadband data comes from the Federal Communications Commission (FCC) Form 477
- The 7 metrics comes from household and retail scanner data

Possible Bias in the data?

- If a large proportion of the retail scanner data came from stores of similar category. For example, too many stores are all grocery stores like Trader Joes, Whole Foods and Safeway.

Other issues with the dataset (limitations.)?

- There is little distinction among different types of high-speed internet services and the measure between high medium and low is not the most refined
- Providers may not necessarily offer the reported service everywhere in the census block, and the number of providers in the census block does not purport to measure competition.
- Is the data set large enough to support conclusions drawn for the entire US economy and its people?

How did you clean and preprocess the data?

- The data was cleaned and preprocessed by using imputation methods to fill in missing values, harmonizing data from different sources, and aggregating data to different geographical levels for analysis.

The data is organized as follows:

Household scanner data:

- Outcomes:
    - Trip frequency
    - Total spending
    - Unique chains visited
    - Unique brands purchased

Retail scanner data:

- out comes
    - Prices
    - Price dispersion
    - Price elasticity

Both datasets were preprocessed by linking them with the census tract-level broadband data collected by the FCC through Form 477.

Extra

- Nguyen found that contrary to popular media reporting, aggregate retail trends during the rise of broadband were muted and did not support the hypothesis that brick-and-mortar retail was on the verge of collapse. The only significant changes

observed were a decrease in the number of unique brands and an increase in average prices consistent with inflation.

- Our original goal was to have two apps. The first would present visualizations on metrics (1) to (6). This would be the halfway point of the project and provide us with the tools to complete the most important goal which was to create an app for metric (7) price elasticity of demand.
-
- We decided to pivot and make phase 1 the end goal.