* What are we building?
  + Problem Statement
    - Users are currently struggling with poor classifications of companies in sectors and industries. There is often a mismatch between what a company actually does and what industry it is classified in. There is a lot of subjective human error in the current manual classification process. This is also very labor intensive and not interpretable. The evolution of a company as well as a sector or industry is not captured in the current data.
  + Goal
    - Create a solution that gives more objectivity, interpretability, flexibility, and automation into the company classification process.
    - Build a solution that classifies companies into maximum of 10 sectors and no more than 100 industries, but also provide flexibility to change these parameters to less or more classifications.
    - The solution will provide detailed metrics and weights behind each of the company classifications and enable the user to have a deeper understanding of why a particular company fits into a sector or industry.
    - Provide the user the ability to both tune/optimize the models but also swap/compare different models more easily.
    - The ability to review parameters for the weighting ofcompanies in each sector and industry.
* Who are we building for? (personas)
  + Consumer
    - Objectives
      * Capture and assess the impact of global, regional or local industry trends on a portfolio
      * Compare and report on industry sector exposures versus peers or benchmarks
      * Pinpoint industry investment opportunities across developed and emerging markets
      * Analyze sector and industry contributions to portfolio performance
      * Construct consistently defined global or regional sector-based and sector rotation strategies
    - Personas
      * Company Executives
      * Economists
      * Investors
      * Regulators
      * Employees
  + Data Scientist
    - Continuing to improve and expand the model that we build here
    - Have a deeper understanding of the methods and logic underlying the classification and grouping of companies in sectors and industries.
* What data is needed?
  + Baseline current NASDAQ classifications
  + 10K
    - EDGAR API
  + Financial Information
    - EDGAR API
  + Company Website
    - Web Scraping -- Keyword and Description <meta> tags
  + Company News
  + Wikipedia
* Define Personas
* Develop User Stories
  + As a product owner, I need to be able to review data and re-collect all the information used for the research.
  + As a product owner, I need to be able to generate a list of 10 sectors and no more than 100 industries classifications group under the sectors with related companies
  + weighting from collected data by running no more than two commands.
  + As a product owner, I need to be able to review classifications using an infinite scroll HTML web page. The web page shall include data visualization, analysis and insight
  + of sectors and industries with recommended weighting of each company based on 10 years of historical data.
  + As a DevSecOps engineer, I need to be able to stand up entire infrastructure using 100% automated scripts with a minimum number of commands.
  + As a data scientist, I need to be able to have an infinity scroll infographic to view the raw data.
  + As a data scientist, I need to be able to understand the classification and grouping of companies in sectors and industries.
  + As a data scientist, I need to be able to review parameters for the weighting of companies in each sector and industry.
* Prioritize
* Planning
* Understanding the distances between clusters

Data Sources

* 10K (Harold)
  + EDGAR API
* Financial Information (Eugene)
  + EDGAR API
* Company Website (Carlos)
  + Web Scraping -- Keyword and Description <meta> tags
* Company News (Zach)
* Wikipedia (Zach)

Data Modelling

* Doc2Vec in R (David Moore)