**Requirement Phase**

**Team**: Mount Harvard

**Members**: Jenna Beutler, Seth Grossman, Miao Hua, Samuel Statton, Nhi Vo

**Dataset**: <https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge>

**Business** **Case**: Common factors/symptoms that employees/employers should look out for to help prevent spread of COVID-19 in the workplace

**Functional Requirements:**

The following functional requirements need to be in place to achieve the end product as described in the Project Requirement document:

* Data Ingestion
  + Loading the data in AWS via EMR master node
* Coding
  + Collaboration amongst group members via github and AWS to write, review, and deploy code
  + Enable horizontal scaling
* Data Visualization
  + Using Pyspark to perform basic aggregate functions and visualize the data
* Text Analysis
  + Utilizing Pyspark text analysis tools to gain insights into each article Abstract
* Business Applications
  + Translate insights from the data into business applications
* Deliverables
  + April 12: Requirements Document, GitHub Classroom
  + April 25: Design and Development Test
  + April 28: Deployment & Presentation

**Non-Functional Requirements:**

The following non-functional requirements have been considered by the team as factors that will influence how the project is conducted but deemed non-functional:

* Collaboration
  + Using Google Colab to share code and ideas before posting to GitHub
* Communication
  + WhatsApp is our primary text communication app while Slack will be utilized when we have questions for Peter.
  + Zoom will also be used for checkpoints and to share screens for code examples
* Time Analysis
  + Segment data by data (pre-2000 and post-2000)