Executive Summary

Background
Pennsylvania Department of Labor and Industry deployed an analytic system developed by CMU students in the previous semesters which included features, such as data integration, time-series prediction and prediction result visualization. In this semester, we continued to work with the PA department to improve the prediction web application and provide insights on employee injury claim mitigation.

Objective
The overall project objective is to provide client insights on injury rate and build tools to predict the future rate. Through this project, we hope to assist the staff in the PA department to do more proactive work that could reduce injury rate.

Methodology
We started by working on data cleaning and preparation. Then we moved to the exploratory data analysis phase, where trend analysis and various granularity data analysis were conducted. To validate and compare with the previous findings, we developed further clustering analysis. Next, we moved to the model stage, where we approached time series models and non time series models in parallel. We studied the existing time series models and added the prediction evaluation function. We also implemented a neural network model to predict the top five injury natures per employee case. Eventually, we reached the final stage to enhance the web application by incorporating new features. We conducted local tests to ensure the application is operating as expected. Based on those output, we finalized our deliverables listed below.

Deliverables
- **Web application** - added neural network prediction model and time-series evaluation, descriptive data analysis
  https://github.com/LansingC001/CapstoneProject_2022Spring
  Model_training zip in package
- **Installation documentation**
  https://github.com/LansingC001/CapstoneProject_2022Spring
- **Video** - Product introduction and commercial
  https://drive.google.com/file/d/1E5YtlaCgejJTOZEHaMzAQPxPHFuOxQZ3/view?usp=sharing
- **Presentation** - Introduce our project output and analytical insights
  https://drive.google.com/file/d/1QmTy58HLoYefBK488vNMm3ccBJbxm5QK/view?usp=sharing
- **Report** - Includes project background, scope, value, data analysis, prediction model, software
  https://docs.google.com/document/d/1vVfYXKRwk5MTBHAWf5QeDlVmbpOOQJik_mJEyXc06dw/edit

(Please contact Lansing at yifanche@andrew.cmu.edu for access permission).

Installation instructions
See Github readme: https://github.com/LansingC001/CapstoneProject_2022Spring
(Please contact Lansing at yifanche@andrew.cmu.edu for access permission).

Recommended next steps
- Augment the PowerBI dashboard
- Consolidate Data Cleaning Logics
- Document data source and FAQs
- Acquire medical cost rankings
- Optimize neural network model