# Problem and Motivation

*<* *Describe the selected questions you have addressed. Explain the significance and your reasons for seeking a solution to the questions.>*

Our project focuses on the Client Services aspect of the business questions. First, as a response to Question 1, we would like to investigate the relationship between the amount of time spent working with individual clients and how quickly they are employed. This relationship will help the Hire Heroes USA team (HHUSA) find the optimal amount of efforts that they should make to work with each client.

# Approach

*<Give an overview about the major steps you have done in your project. If applicable, provide more detailed information about your ideation approach.>*

For Question 1, we fitted regression models in which the response is the number of days that it took each client to get employed, and the predictors include various indicators of efforts in working with each client. We preprocessed the dataset to obtain the response and predictor variables. In particular, we observe that besides the total number of logged activities, the frequency of logged activities provides additional insights as a predictor, because at a constant activity rate, the total number of logged activities is expected to increase with the time that it took the client to get employed, whereas contact frequency is a better indicator of how closely the HHUSA worked with the client. We chose this overall approach because the slope of the regression contains information about the contribution of each predictor to the job-seeking time, and the p-value indicates the significance of the relationship.

# Datasets

*<For the case that you have added other data (e.g. open data) to the data provided, describe the additional datasets on a high level and indicate their origin (company, your advisor, open data, self generated, etc.), format (structured, text, etc.) and size.*

*Otherwise leave this section empty.>*

(Empty for business question 1.)

# Tools and Analytics

*<Describe which tools you have used for which purposes. Provide details about your analytics approach (methods, techniques, algorithms, etc.). >*

To answer Question 1, we applied a linear regression. The response is the job-seeking time, and the predictors include various measures of activities between the HHUSA and the client, such as the number of logged activities with the client and whether the client requested to have their federal resume reviewed, was told about HHUSA resume tailoring assistance, and had a federal resume reviewed by HHUSA team. To find the optimal model, we tried various data transformations such as the Box Cox Transformation, among others. We selected the best model according to diagnostic plots and Rˆ2.

# Results

*<Describe your results in detail. Which artifacts (data models, analysis models, reports, visualizations, etc.) have you generated in the project? How may the results of your work contribute to the future success of the non-profit organization).**>*

For Question 1, we conclude that the clients who were contacted more frequently tended to get employed significantly faster, since according to our linear regression model, the slope of activity frequency is negative with a p-value less than 2e-16. Additionally, our results showed that the clients who had been told about resume tailoring assistance tended to get employed more slowly, since the slope is positive with a p-value less than 2e-16. Besides, the clients who had a federal resume reviewed by the HHUSA needed longer time to get employed, since the slope is positive with a p-value 0.00164.

From these results, we obtain the following suggestions to HHUSA. First, working with the clients more frequently helps them find jobs faster. Moreover, it would be helpful to develop customized services for the clients who are classified as requiring resume tailoring assistance or requested their federal resume to be reviewed by HHUSA, because these characteristics potentially imply that they need additional help from HHUSA.