**1301 Holling Dr, Frederick, MD 21701**

Capstone Project Proposal

Ben Berg

Project summary for capstone project to be completed during SpringBoard Data Science Intensive online course.

08

**Fall**

Table of Contents

Initial Project Ideas 3

What is the problem being solved? 4

Who is the client and why do they care? 4

What data is being used? 4

Approach Outline 5

Project Deliverables 5

# Initial Project Ideas

1. Predict the volume of inventory of job openings that will become open in the near future based on historical job opening data
   1. By industry
   2. By Job Title
   3. By Location
2. Predict the lifespan in days of a job opening being available online
   1. By industry
   2. By Title
   3. By what job board site it is posted on
   4. By Location
3. Predict the likelihood of membership sign up based on inventory available that is relevant to a user
   1. By Industry
   2. By Job Title
   3. By Location

# What is the problem being solved?

As a potential customer of iHire, I only have insight into how many jobs are open for me at the current moment. Being informed on level of inventory of jobs I can expect to become available through iHire services, I am able to make a more informed decision and more clearly understand the value of iHire’s service of providing new job leads from across the internet.

As a product person, I would like to highlight/expose to our current and potential customers the future volume of job openings likely to be available. This will aid us in communicating value of product.

As a marketer, I would like to understand future inventory of job content so I am able to react to changes in inventory prior to it actually happing. Having this information will prevent spending advertising dollars acquiring customers that iHire is less likely to have job inventory for because of a swing in inventory.

# Who is the client and why do they care?

Membership customers of iHire’s premium job seeker service would like the information because it sets their expectations of incoming job openings in the future.

Product Team would like to use information to inform customers of inventory.

Marketing Team would like to save money and optimize ROAS by spending wisely in market segmets that inventory is likely to be available and not in market segments that inventory is not predicted to be adequate.

Content Team would like to understand where to prioritize their time in finding new sources of jobs for weak segments as well as prioritizing time spend adding to classification system that classifies jobs into segments

# What data is being used?

iHire owned historical job posting database will be used. This data contains the date the job became available, the date it became unavailable, and many features that can be used in the prediction model. Multiple years of data is accessable and will be used in hopes of understanding if seasonality has an impact – back to 2013 is available.

# Approach Outline

1. Acquire and clean data for sample and test data
2. Feature identification and prioritization
   1. Hypothesis around these features
3. Data mining
4. Trial and error with Machine learning algorithms for prediction
5. Power point creation and presentation

# Project Deliverables

Algorithms tested and documented in Github repository

Powerpoint presentation telling the story of approaches used, any potentially valuable insights learned in process of mining the job data set