Molly McNamara
Foundations of Data Science Capstone Project Ideas

1. Build a predictive model for hockey success

Establishing the right balance of player skills and team needs to make a successful run at the playoffs is the challenge of every National Hockey League (NHL) general manager. In the United States, teams see an impact of improved performance and win totals on season ticket sales and sponsorship revenue. I propose to evaluate National Hockey League data from ~2000-2012 to build a model that determines what factors predict which teams will make the playoffs and succeed through multiple rounds. This model would then be tested on 2013-2017 season data to evaluate performance. Model building data is currently in multiple spreadsheets that will need some work to centralize the various attributes of interest. While this dataset actually goes back into the early 20th century, the game has changed substantially in the post expansion era and likely the older data is less useful in predicting modern day success.

2. Evaluation of Avalanche Accident Data

Colorado's outdoor recreation industry generates $994 million in state and local tax revenue (Outdoor Industry Association, 2014). However, outdoor activities are not without risk, as an average of 27 people per year are killed by avalanches in the United States. I propose to utilize nationwide accident data collected by the Colorado Avalanche Information Center to evaluate historical data on avalanche fatalities and determine if temporal, elevation or recreation type trends differ between Colorado and the rest of the country, in order to advise local ski operators, park rangers and tourism organizations on how to best educate and protect visitors.

3. Evaluation of National Park Service Visitation Data

American National Parks in recent years have reached record visitation levels, challenging infrastructure and staff to capacity and park visitors with overcrowding. I propose to evaluate nearly 40 years of National Park Service data to highlight where additional budget and resources are indicated (most utilized facilities, most

visitors per acre, etc) and where visitors can find parks with smaller crowds (both seasonally and overall).