IMT 573: Formulating Data Science Questions

Alexander Davis

Oct 15 2020

Collaborators:

Objectives

In the following lab activity we will practice going through steps of the data science process. Careful planning and defining project steps, as well as adapting to arising challenges, is vital for successful data science projects.

With a small team (2-3 students), you will think about some example data science cases. For each case, consider the steps you would take to plan a project. Think about defining **quantifiable goals, the data needed, and possible approaches to analysis**. Your team should write a short summary of your considerations and thoughts for each.

Before you get into the stage of discussing with your team, you should first think and come up with individual responses. Hence, going through the questions and thinking of responses before class is advisable. If you had to work alone for whatever reason beyond your control (e.g. working in a different timezone, etc.), you can simply upload your individual responses to this lab.

Note: You need to attempt both questions. So you should at least respond to one case from each question.

Question 1

For the following cases, think about what data you might use to answer the research question described.

1. **Team Formation:** Suppose your boss comes to you and says "We need to better understand team performance at our company. I want you to explore the characteristics of teams that perform optimally at solving problems." How might you go about exploring this research question?

   Response:

2. **Population Displacement after Disasters:** Suppose you work for the American Red Cross. Your job is to help determine where to send relief supplies, such as food, medicine and water, after a large-scale disaster. However, you know that when crises occur people move to other, less damaged areas, perhaps to live temporarily with family or friends. How might you understand migration patterns following disasters?

   Response: I would take a look at area code population data and join that with the areas affected by the disaster. I would begin with the path of the hurricaine and compare that with the most densly populated areas that were not struck. If the downtown portion of a city was unaffected then I could imagine that would be a spot where many people would go to. I believe stadiums are often used as areas to house many people as long as they are not damaged.

3. **Predicting Patterns of Mobile Application Usage:** Suppose you are tasked with improving the user experience of a particular mobile phone operating system. Your team comes up with the idea to try

to decrease the time it takes for applications on the phone to start. How might you understand typical application usage patterns? What additional data might you use to help predict what applications are going to be used next by a particular user?

Response:

Question 2

For the following cases, think about research questions you could formulate and explore with the data described.

1. **Random Acts of Pizza:** Consider a dataset with 5,671 textual requests for pizza from the Reddit community Random Acts of Pizza together with their outcome (successful/unsuccessful) and meta-data. All requests ask for the same thing - a free pizza - but each haver their own unique characteristics. The outcome of each request - whether its author received a pizza or not - is known. Meta-data includes information such as: time of the request, activity of the requester, community-age of the requester, up votes and down votes of the request, number of comments for the request, text of the request, and the title of the request.

   Response:

2. **Bike Sharing Systems:** Consider a dataset that contains the hourly and daily count of rental bikes between years 2011 and 2012 in the Capital One bikeshare system with the corresponding weather and seasonal information. This data includes date, season, holiday, day of the week, weather (clear, cloudy, mist, show, rain, thunderstorms, etc.), temperature, humidity, windspeed, and count or both casual and registered users of the bike share system.

   Response: