**Instructions:** This is a team activity. Each team receives a data analytics request. The team shall discuss the request and answer a number of data acquisition questions.

**Data Analytics Request 1**

Cal Poly wants to understand how parking is utilized on campus, in order to reorganize the parking lot designations, and to allow both students and faculty to spend less time looking for parking upon arrival on campus. The campus is planning to build a "lead me to a parking spot" mobile application that would indicate to people coming to campus where the empty parking spots are.

You are engaged for a preliminary study whose goal is to discover the parking usage patterns. The campus has information about parking permit holders, all students, faculty and staff. The campus, however, does not have any information regarding the use of parking that is NOT anecdotal.

**Instructions:** This is a team activity. Each team receives a data analytics request. The team shall discuss the request and answer a number of data acquisition questions.

**Data Analytics Request 2**

A national chain of organic grocery stores is looking to open ten new locations. The chain has information about the performance of its current 50 stores located in a number of different states and in a number of different areas (urban, suburban, ex-urban, small town, rural). The goal of the analytical task is to  analyze the performance of existing stores, determine what factors significantly contribute to the success of the store (defined largely as "profit!"), and find locations in the USA where the demographic (and other) factors predict high performance.

**Instructions:** This is a team activity. Each team receives a data analytics request. The team shall discuss the request and answer a number of data acquisition questions.

**Data Analytics Request 3**

A small electronic gadget manufacturer is trying to establish its name in the market for some specialty electronics equipment. The manufacturer is very sensitive to the opinions of its products. The products are largely sold through on-line retail stores. The products get reviews in on the on-line retail store sites, as well as in a number of other places. The manufacturer identifies a handful of web sites devoted to product reviews, and a number of forum-style web sites that cater to the niche occupations for which the manufacturer produces the gadgets.

In addition, a wide range of technology sites and other news organizations (in the broad sense of the word) may from time to time either reference the manufacturer's products in their articles, or might write reviews (sometimes without the manufacturer knowing ahead of time).

The manufacturer would like to receive summaries of customer's (and others') opinions on their products. Where opinions run negative, the manufacturer would like to understand the specific reasons of negativity, and the level of prevalence of negative opinions about a specific product.

**DATA 401: Data  Science**                                              **October 10**
**Alex Dekhtyar**                                              **Data Acquisition Activity**
**Dennis Sun**

**Instructions:** This is a team activity. Each team receives a data analytics request. The team shall discuss the request and answer a number of data acquisition questions.


**Data Analytics Request 4**

A new web site is trying to tie together individual's tastes in music, literature, film, food, and drinks.  The long-term plan for the web site is to collect certain information about their users' preferences in these four areas and use model-based collaborative filtering techniques to produce recommendations (e.g., based on understanding what music and wine a person likes, suggest a book they should read).

Before launch, the web site needs to collect data on the five domains it studies, in order to build some feature-based similarity models for individual items (e.g., determining that two movies are similar, or that two varieties of beer as similar). Your data science team has been hired to build these models from scratch.  The web site does not have any datasets available (but is willing to go to necessary expense to acquire access to any data needed for this exercise).

**Instructions:** This is a team activity. Each team receives a data analytics request. The team shall discuss the request and answer a number of data acquisition questions.


**Data Analytics Request 5**


A social science research team is studying the psychology of winning and losing. The research team uses sporting events (e.g., NCAA football and basketball games) and is interested in analyzing the sentiment of the fans as the game progresses.  The research team is interested in the "gut reactions" of the fans to the events in the games as the games unfold.

The data science team is engaged to build a platform that allows for collection of the information about the instantaneous fan reactions, as well as about the actual events taking place in the games.

**TEAM:_____**


**STUDENTS: _____**

### Questions Worksheet


Please answer the following questions about the specific data analytical request your team received.

**Question** **1.**  Identify the main datasets you will need to collect. For each dataset, specify whether it exists, or whether the data science team will have to create the data from scratch.

**Question 2.** Identify the owners/origins of each of the main datasets

| Dataset | Owners/origins |
|---------|----------------|
|         |                |
|         |                |
|         |                |
|         |                |
|         |                |
|         |                |
|         |                |
|         |                |

**Question 3.** For each dataset, specify its velocity (static, batch, periodic, near real-time, real-time)

| Dataset | Type |
|---------|------|
|         |      |
|         |      |
|         |      |
|         |      |
|         |      |
|         |      |
|         |      |
|         |      |

**Question 4.** Identify the biggest challenge your team will face when acquiring the data, and how your team would attempt to meet this challenge.