Assignment 3A – Data Science Teams

In this assignment, you will further develop the tasks that were chosen to be included in the experiment last week. This is a crucial part of the project, and will weigh heavily on both the success of your project and your team's overall assessment.

1. **Rate the list of tasks according to your team's preferences**
   To help us assign tasks, we ask you to rank the top 10 tasks you would like to work on. We have added a new tab called "Preferences" in the Wisdom of the Crowd worksheet. **Please submit your preferences by Saturday, 11:59am PST.** We will assign each team 2 tasks by Saturday evening, taking into account your preferences as much as possible.

2. **Write short description for the domains**

   Prior to answering questions in each domain, we will provide users with a short description of what we are about to show them. For example, in the case of matching countries with their flags, we might say:

   "We will now ask you a series of multiple-choice questions. For each question, we will show you a picture of a national flag, and ask you to identify the country it represents."

   Please provide short description for both of the domains to which you are assigned.

3. **Define the corpus of questions for each of the two domains**

   We would like to select the questions for each domain in a principled manner. For example, for the country/flag question domain, we do not simply want to select the first 20 countries that come to mind. Instead, we might start with the list of all countries listed on Wikipedia, and then randomly select 20 of them. Alternatively, we might randomly select countries proportional to their population.

   Formally define your method to generate the questions. This includes determining the *universe* of possible questions, and also the sampling procedure. Identify specific data sources you plan to use, and justify the procedure you propose. We also suggest that you try out your procedure to ensure that it is both possible and reasonable in practice. While this may be clear how to do in some domains, it may be quite difficult in others!

4. **Determine the set of answers [only for the multiple-choice questions]**

   For domains that require multiple-choice questions, we also need a method for generating the answers. Specifically, we require each question to have 5 answers, one of which is the correct answer, and 4 of which are incorrect answers. Describe and justify your methodology for selecting the incorrect answers. In the final experiment, the order of the five possible answers will be displayed to each user in a random order.

An example submission of what we expect is at the bottom of this document. **Please submit your completed assignment on CrowdGrader by Tuesday, 11:59pm PST.**

Assignment 3B – Platform Design Teams

Continue working on the development of the experimental platform. Review the domains in the "Final Tasks" tab, as you will need to ensure that the platform you build is able to support the types of tasks your colleagues are working on. Please also see the schematic outline of the experiment in the GitHub repository (doc/experimental_design.pdf).

Example Submission – Data Science Teams

Task: Identify the Flag of a Country

Detailed Description shown to participants: "We will now ask you a series of multiple-choice questions. For each question, we will show you a picture of a national flag, and ask you to identify the country it represents."

Input Type: Image

Corpus
Wikipedia's list of countries and population:
http://en.wikipedia.org/wiki/List_of_countries_and_dependencies_by_population

Representative tasks methodology
Sample 20 countries from this list weighted by their population.

Answer type: multiple-choice (list of countries)

Multiple choice option selection: 4 incorrect answers will be sampled from the list of countries weighted by population.