Distributed Computing Project 2
Initial Planning Document

Group Members
- Joshua Ingram
- Nikolas Santamaria
- Fehmi Neffati
- Paul Hwang
- Aaron Spielman

1) List a few lessons that members of your team learned from Project 1

There were 5 main takeaways from Project 1:
1. Keep the scope of the project in mind
2. Be mindful of and plan for workflow bottlenecks in the project
3. Prioritize getting data collection and data preparation completed early
4. Have multiple group members check the validity of the data
5. Use informative file and variable names for better readability

2) List a few things that you are doing differently in this project than in the previous project in light of the lessons you have learned. For example, you might mention things related to communications strategy, meeting strategy, project management strategy and/or use of technology.
More regular, frequent meetings

Limiting time on planning
        -
Primary methods of communication (ranked):
1. Google Space Group Chat
2. Google Doc Logbook
3. Email

Technologies for Organization:
1. GitHub (https://github.com/FNeffati/DCP1)
2. Google Drive

3) What is your initial high-level plan of attack?

Communication Strategy:
Our team will maintain regular communication via Google Space, a Google Doc logbook, and comments for any commits on GitHub. As much as possible, progress on the project will be through decentralized individual or small team efforts. Progress made will be communicated regularly and when needed for major challenges, the full group meetings will be organized.

Meeting Strategy:

Each meeting that takes place will have an agenda organized by the project manager. Group members will present their individual progress since the last group meeting, their team progress, and will present any problems they have had that need to be addressed in the meeting. At the end of each meeting, a plan of action will be produced with a tentative next meeting date.

Frequency of Meetings:
- Full group meetings will be made on an as-needed basis (estimated 1-2 times per week). These will be held in person.
- Partial group meetings will be up to the discretion of the members of the teams. These can be virtual or in person.

Plan of Action:
1. Topic: Are there any interesting trends in political advertisement on Facebook?
2. Data munging
   a. Extract targeted ad information
   b. Check for missing or bad data
   c. Compare two datasets
   d. Collect data and combine into one large dataset by matching related variables
3. Exploratory data analysis
4. Analysis
   a. Time series analysis of ads
   b. Categorical data analysis of ad posters
   c. Analysis of ad targets
5. Create presentation

4) Initially, who is doing what?
Group Member Roles:
Joshua Ingram - Project manager, time series analysis
Nikolas Santamaria - pyspark, exploratory data analysis
Fehmi Neffati - collecting and formatting data on targets
Aaron Spielman - data munging, exploratory data analysis
Paul Hwang - exploratory data analysis