**Please use the following template for your project proposals (maximum 2 pages):**

1. Name, Contact info (e.g. email/phone).

Lindsay Pritikin  
[lpritikin@regis.edu](mailto:lpritikin@regis.edu)  
(630) 205-2573

1. Title of the project: Pagedip
2. High level description of the project: what question or problem are you addressing?

Pagedip enables users to create data-driven, interactive documents that prevent readers from having a fragmented experience that moves them away from the site. The data Pagedip gathers can help their clients make better decision around which content is working and which isn’t. The problem we are solving for is enabling Pagedip to provide the most meaningful engagement analytics to their clients. This will include but is not limited to:

* Data Validation
  + Look at existing pagedips and do clean tests of interactions to make sure the data is as pure as they think it is. Test clicking on things and validating that it came through
* Engagement index
  + Measure how well users are interacting with the pagedips. Interactions are measured through views, time on page, total interactions, engagement factor (how much interaction was achieved vs. how much is possible). Look to determine a proper scale through segmentation
* Corrections/accountability for idiosyncrasies of widgets
  + i.e. Tab-its that always have the first tab open. Look for techniques to account for these views more accurately

1. What type of data science task is it?
   * Data study to validate data
   * Segmentation to develop the engagement index scale
   * Data modeling to account for widget idiosyncrasies
2. Data: Brief description of data. How big do you expect the data will be? Is amount of your data too big or too small? If you're web-scraping or collecting data, how long do you expect to collect the data?
   * Data is provided by the client. The data looks small and has many missing values
3. How will you analyze the data? What machine learning methods do you plan to use, and/or what business intelligence aspect do you plan on incorporating?
   * I plan to do some type of segmentation to develop the engagement index. I also plan to use some type of logic to model fuzzy widget data. For data validation, I will do a data study with tests to ensure the data is coming through accurately.
4. Describe any anticipated difficulties and problems. Discuss how you may overcome the problems.
   * Data is limited and there are many missing values. I will ask Sherisse if there is any additional data she can provide
   * Without access to the pagedip platform, I am not sure how I will perform a data validation study. Perhaps I will have to take note of the datetime of all of my actions, and have Sherisse pull a report after the fact. IP address is not collected, so the only way to know that it is me is the datetime and location; hopefully given the limited data this is unique enough
   * I am unsure at this time how we will perform segmentation for the engagement index. There is not much data on individual widgets, so I envision this more as scoring by pagedip handles. These pages can have statistics such as engagement ratios (actual engagement/potential engagement), time spent, views, and so on. We can then use this data to build our model, find the optimal # of clusters, and cluster the data into our score scale.
   * I am not sure how we will calculate potential engagement automatically if it is not in the data; can they start tracking this or have it in a separate table that we can reference?
5. Suggest a timeline for the project.  This should be a weekly breakdown of what you plan on doing each week.
   * Week 2:
     1. Familiarize myself with the data provided and gather a list of questions
     2. Confirm if it is possible to calculate potential engagement with data
     3. Import the data into Jupyter
     4. Start on EDA and summary statistics to understand the range of data, # of missing values, average time spent, which widgets are driving the most engagement, etc.
   * Week 3:
     1. If there is any additional data I will add that into Jupyter
     2. Continue EDA
     3. Perform data validation study
     4. Determine which widgets need data adjustments and gather ideas
   * Week 4:
     1. Review data validation study results
     2. Regroup with team on widget adjustment ideas
     3. Test out fuzzy widget adjustments
   * Week 5:
     1. Regroup with team on widget adjustments to choose the best fit
     2. Test engagement index segmentation
   * Week 6:
     1. Review engagement index results with the team
     2. Optimize segmentation index
   * Week 7:
     1. Data visualizations to present results
     2. Start on deck
   * Week 8:
     1. Finalize deck
     2. Presentations
6. Create GitHub repository for your Practicum project. Add this proposal, begin a ReadMe document, and begin adding your data to your repository. Add a link to your GitHub repository to this document. <https://github.com/lpritikin/Regis-MSDS-Practicum.git>

Don't worry if you can't answer all of these in week 1. But by the end of week 2, we need to have a clear outline of your project.