**ANLY 501**

**Fall 2018**

**Project Assignment 3**

**DUE: Dec 10 + 24 hours grace period at 11:59 pm ET**

**(It is not possible to submit after Dec 11 for any reason)**

**PROCEDURES AND LATE POLICY REMINDER**

* **Turn-in:** Please turn in your project through Canvas.
* \*\*\* Each member of each Group must submit a full copy of the project to their own Canvas. If you do not do this, the penalty is -15%.
* **Late policy**: It is not possible to submit after Dec 11 at 11:59 pm ET.

**Overview**

This final project assignment asks you to **finalize** your **data science story**. Specifically, your group will conduct one more analyses on your data set that augments your final data science story. You will then add some visualizations, including **two or more interactive plotly visualizations**.

**The idea here is to pull everything together, add a couple more interesting analyses, several excellent visualizations, and then wrap it up.**

**Additional Analysis (20%)**

The analysis conducted in this component will be integrated into the results of your overall story. Only one additional analysis is required - you determine which is most appropriate for your data and your overall “story”.

Below are suggestions/options for different potential analyses. Choose one. Explain why you believe your choice will improve your data science story. Complete the analysis. Add the results to your project “story”. If you feel that the analysis does not add any value to your "story" you can choose to leave it out or to do an analysis that does add to your overall story. This is open-ended and so it up to you.

**Choices (choose one - if you are set on a different option - email me - briefly and clearly explain the other option):**

***Sentiment analysis*** – If you have textual data, you may be interested in determining how positive or negative the data is. You can present your results using text or visualizations. Either way, make sure you explain what you are finding sentiment for and what your sentiment results mean.

***Topic modeling*** – Again, if you have text data, you may be interested in topic modeling. Topic modeling allows you to looks at your different text features and see (1) what topics are prevalent in your data set, (2) the distribution of topics for each “document” and (3) what text is similar to other text based on similar topics. You can present your results using text or visualizations. Either way, make sure you explain what you are finding topics for and what the results mean.

***Network analysis*** – We discussed different network metrics, network clustering, and network information flow. Any/all of those are options are good. At a minimum, you should compute the standard metrics local metrics, betweeness, degree, and clustering coefficient, and some global network metrics, density, triangles, and averages for the centrality metrics. Then you can either do a clustering analysis or an information flow analysis.

**Visualizations (30%)**

For this part of the assignment you will create **at least 5** visualizations, and at least two must be an interactive visualization. **I suggest plotly.**

NOTES: Even though I am giving you a minimum of 5 vis with 2 interactive, the truth is that you should create many visualizations :) While you MUST use plotly/python for the 2 interactive visualizations, in addition to those, you can also use other great methods such as….

* Bokeh<http://bokeh.pydata.org/en/latest/>
* Matlabplot<http://matplotlib.org/>
* Tableau<https://public.tableau.com/s/>
* Google Charts<https://developers.google.com/chart/>

**Final story (40%)**

We have talked about a lot of different data science cases through the course of the semester - such as the Case Studies we have shared and various examples in class. For this task, you will put together **your story**. You should do this as a Word document or a website. Make sure to include your citations.

**Presentation (10%)**

You will be giving a 15 minute presentation of your final project. You can use your website, a Powerpoint presentation, or a combination. Every group member must present.

***A few final notes:***

* You will submit an electronic version of your project through Canvas. This will include the .py code, a Word Document with copies of all of your graphics, a README, the Story Document (or link to the Story on a website - you can use Google Sites if you wish), needed data, and anything else needed to fully run and explore your code, story, analysis, and results.
* Using a website to tell your story will gain you **extra credit. (+3%)**
* You can earn an additional **2%** extra credit on this project assignment. You can perform 2 analyzes instead of 1 and/or can create additional visualizations beyond the requirement.
* Remember to explain the limitations of your findings, e.g. is your data especially noisy, did you have a lot of missing data, etc.
* Be creative and have fun!