**Additional Final Project Information**

**Please read the instructions for the Final Project carefully. The required content for the MAIN notebook is clearly outlined in the Instructions.**

You will need to **START** the Final Project to access the instructions and grading rubric by clicking the Instructions button (once you start), which opens the side panel. An attempt is only counted once you submit the assignment. The “Resume,” you will see means you are still on your first attempt and can continue working without any issues.

**Reminder:**For the Final Project you will be using the same Spotify song dataset.

songs\_url = ' https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2020/2020-01-21/spotify\_songs.csv'

df = pd.read\_csv( songs\_url )

**Failure to do so will result in a point deduction.**

**Final Project Submission**

For submission, only the following file formats will be accepted:

* .ipynb (Jupyter Notebook)
* .html (HTML export of the notebook)

For every Jupyter Notebook there should be a corresponding HTML file.Word and PDF files will **NOT** be accepted. (i.e., 4 Jupyter notebooks means you should also have 4 html files)

**Jupyter Notebook Formatting**

The ***Grading Rubric*** is associated with the **MAIN** notebook, which will be the primary focus during evaluation. The purpose of the supporting notebook is to verify any confusions or concerns regarding your MAIN notebook.

To ensure accurate grading, please make sure your MAIN notebook follows the required template and guidelines provided in the Final Project Instructions.

**Main Notebook**

* Contains your response to ***Part A*** and all information listed in the ***Instructions*** under ***Reporting.***
* Every section from Part B through Part F contains a ***Reporting*** section
* All discussions and inferences **must be included** in the MAIN notebook to receive full points.
* Explanations of methods used
* Reasoning for the visualizations included in the MAIN notebook
* Example: If KMeans ends up being the visualization that contributes most effectively compared to hierarchical clustering, then only KMeans should be included in the MAIN notebook—along with the required explanation and your reasoning for choosing it over other clustering methods. All other approaches you explored should be included in your supporting notebook.
* The MAIN notebook should include discussions and inferences along with the important, finalized content—such as key EDA visualizations, key cluster visualizations, and major modeling results and visualizations—as outlined in the instructions and grading rubric.
* Please note that some of the instructions are different depending on whether you have a regression or classification problem. Only implement the visualizations that are relevant to your problem.
* Example 1: Both **classification** and **regression** problems are expected to plot model predictions using the visualization grid, including confidence intervals, to assess performance.
* Example 2: ROC is only required for **classification** problems. If you are doing regression and include ROC you will lose points.

**Supporting Notebook**

* Draft notebooks where you explore different approaches while working on the Final Project
* Explanations and discussions are not required for the supporting notebook.