CENTER FOR
COMPLEXITY &
EMERGING
TECHNOLOGIES

Probationary Lyrids
Tidy Tuesdays - Deliverables

Introduction

Now that you've completed your bootcamp training, to help you hone your skills on data preparation, visualization, and analysis, we have prepared some activities for you to work on during the term.

After completing these activities and projects, you will be evaluated based on your performance. These activities will serve as the basis for promotion to a cohort member at the end of Term 3.

Prerequisite

Kindly create a GitHub repository called *lyrid-training.* This repository will serve as your personal and submission repository for all the individual deliverables.

Tidy Tuesdays

Tidy Tuesdays is a project that is available for the R community to help them practice data cleaning and processing to make meaningful charts. We're using their idea as a way for you to also practice your skills but using Python.

You will be given datasets per week to explore and analyze. Choose only one out of the provided options. After which, you would have to create meaningful visualizations from the data.

As a guide, you can follow this outline for going through the chosen dataset in a Jupyter Notebook. For submissions, kindly commit your Jupyter notebook to your GitHub repository. Make a folder for every week's Tidy Tuesday submission.

Tidy Tuesday Activity Checklist

Data Preparation
- Show the step-by-step process of how you cleaned the dataset (depending on the data given)
  - Removing null values
  - Formatting issues
    - Numerical values
    - Date & Time
    - Varying representation (M -> Male)
    - Duplicate data
- Make sure for the processing choices that you make, provide some supplementary comments or narrative using the markdown cells in the Jupyter notebook

Data Analysis
- Perform descriptive analysis and some exploratory data analysis with the data.
- Make sure to also write out what you observe from the data using the markdown cells (i.e. notice any anomalies, correlations, etc.)
- Make use of basic charts to also show your exploration steps
  - The charts in this stage need not to be well designed but should at least have labels and supporting captions or text based on what you observe.

Data Visualization
- Create a final visualization as a summary to what you learned from the visualization or as a means to communicate the data to others.
- Make sure that this figure (can be a single plot or multiple subplots) is properly labeled with titles and units in the labels.
  - Transform the data visualized to not use the default scientific notation (you can perform divisions!)
  - Choose the right color scheme for your data.
    - Categorical data uses distinct colors.
    - Ordinal (ordered categorical data) can be represented using sequential color schemes.
    - If your data has positive and negative values with a distinct midpoint (i.e. correlation that have positive and negative values), you can use a diverging color scheme for this.

- - Use diverging color schemes when your data goes in opposite directions.
  - Use sequential colors schemes when your data just goes in one direction.
- Also provide a detailed caption of your visualization.

Term 2 Timeline

The table below shows the schedule of release of dataset choices and the deadline for submission of each Tidy Tuesday exercise. Unlike the original Tidy Tuesdays, we will only be doing the exercise every 2 weeks. The release of dataset choices will be every Tuesday and submission will be on the Friday of the following week.

Week 1	N/A	N/A
Week 2	N/A	N/A
Week 3	March 16 (T)	Tidy Tuesdays 1
Week 4	March 26 (F)	Submission number 1
Week 5	BREAK	Holy Week
Week 6	April 6 (T)	Tidy Tuesday 2
Week 7	April 16 (F)	Submission number 2
Week 8	April 20 (T)	Tidy Tuesdays 3
Week 9	BREAK	Independent Learning Week
Week 10	May 7 (F)	Submission number 3
Week 11	May 10 (M)	BREAK
Week 12	N/A	Independent Learning Week
Week 13	N/A	FINALS-TERM WEEK (NO TASKS)

There will be a workshop to be held during the term. The date of the workshop is to be announced on the FB page.