**GWU Data Science Datathon Guidelines – January 2022**

1. **DATES:** January 15th 10:00 am – January 17th 11:59 pm;

Optional in person room: (TBD) 10:00 am – 10:00 pm January 15th

1. **DATASET:** The [dataset](https://github.com/fgh-global/2022-GWU-Datathon) is pulled from Twitter’s [rest API](https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/object-model/tweet). Each row of data collates features about a single tweet such as [**user data**](https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/object-model/user) (like name, screen name, follower count, verified, etc); **tweet information** (like tweet text, created at, etc.); **engagement data** (like retweet count, favorite count, reply count, etc); etc. This is a high dimensional dataset and will require manual pre-processing to subset relevant predictor features.
2. **PROBLEM STATEMENT:** Twitter users engage with Tweets using tools such as favorite, retweets, replies, etc. Taking relevant user and tweet features from the dataset, train a suitable model to predict tweet engagement. You can consider creating dummy variables or work with appended fields like [sentiment](https://arxiv.org/pdf/1601.06971.pdf) of a tweet to build a robust model. You have the discretion of selecting a Target Variable from the set of engagement features provided in the dataset or [create](https://eclincher.com/your-tweet-got-ratioed-what-next/) an engagement variable using the available features.
3. **REPORT:** The report (a Jupyter notebook or an R markdown file) should include following sections:

Section 1: Summary/Abstract

Section 2:Project pipeline

* Data Preprocessing
* Exploratory data analysis
* Hypothesis testing (optional)
* Model selection and validation
* Results interpretation

Section 3:Recommendations and Next Steps

1. **VIDEO:** Record a 5-minute-long presentation, focusing on the result interpretation and recommendations described in the report (above).
2. **DELIVERABLE:** Each participant must submit a text file (firstname\_lastname.txt) to [jakelieberfarb@gwmail.gwu.edu](mailto:jakelieberfarb@gwmail.gwu.edu). The submission file should contain participant’s name and email on top, followed by a link to their GitHub repository with their final report and a link to the video presentation.
3. **PRIZE/REWARD:**
4. First Prize - $150
5. Second Prize - $100
6. Third Prize - $50