To replicate the analyses from our project, run the following notebooks in order.

For files created by the notebook that are greater than 25MB, you can view them through the Google drive links provided in this README file. For files less than 25MB, you can view them in the data folder of the GitHub repository.

1. **requirements.txt**
    - This file contains all the Python libraries and packages that should be imported in order to successfully run the analysis files

2. **influencer_preprocessing.ipynb**
    - This notebook consolidates the zip files from the open-source Influencer and Brand dataset into Brand, Influencer, and Post CSVs
    - Files created by this notebook:
        - Brand_data.csv - original Brands dataset
        - Influencer_data.csv - original Influencers dataset
        - Joined_post_json_1.csv - 1 of 6 Posts datasets
        - Joined_post_json_2.csv - 2 of 6 Posts datasets
        - Joined_post_json_3.csv - 3 of 6 Posts datasets
        - Joined_post_json_4.csv - 4 of 6 Posts datasets
        - Joined_post_json_5.csv - 5 of 6 Posts datasets
        - Joined_post_json_6.csv - 6 of 6 Posts datasts

3. **brand_metrics.ipynb**
    - This notebook calculates brand related metrics which are used for the rest of the analyses
    - Files created by this notebook:
        - influencers_sponsored_mentioning_brand_merge.csv - updated Brand dataset with calculated Brand metrics

4. **influencer_metrics.ipynb**
    - This notebook calculates influencer related metrics which are used for the rest of the analyses
    - Files created by this notebook:
        - sponsored_brand_mentions_by_influencer_merge.csv - updated Influencer dataset with calculated Influencer metrics

5. **twitterdata2.ipynb**
    - This notebook collects Tweets for a specified time period using twint and consolidates the results into a Tweets CSV
    - Files created by this notebook:
        - alltweets.csv - Tweets collected for the during-post period (October 1, 2018 to January 1, 2019)
        - alltweets2.csv - Tweets collected for the pre-post period (July 1, 2018 to September 30, 2019)
        - influencers_sponsored_mentioning_brand_merge_alltweets_pretweets.csv - updated Brand dataset with number of Tweets from pre-post and during-post time periods

6. **exploratory_analysis_relationships.ipynb**
    - This notebook conducts an exploratory analysis of the relationships between the variables in the Influencer dataset
    - Files created by this notebook:
        - Charts and visualizations for the blog post

7. **exploratory_analysis_distributions.ipynb**
    - This notebook conducts an exploratory analysis of the distributions of variables in the Influencer datasets
    - Files created by this notebook:
        - Charts and visualizations for the blog post

8. **Influencer_clustering.ipynb**
    - This notebook clusters influencers based on similarities using the K-Means clustering algorithm
    - Files created by this notebook:
        - Charts and visualizations for the blog post

9. **network_analysis_preprocessing.ipynb**
    - This notebook conducts preprocessing of the datasets to be used as input for the network analysis
    - Files created by this notebook:
        - Influencer_brand_category.csv - influencer-brand relationships dataset

10. **network_analysis.ipynb**
    - This notebook conducts network analysis to understand connections and loyalty between influencers and brands
    - Files created by this notebook:
        - Influencer_brand.gpickle - network of influencer-brand connections
        - Influencer_category.gpickle - network of influencer-category connections

11. post_captions.ipynb

- This notebook extracts post captions from the 6 Posts datafiles and conducts an NLP analysis on them
- Files created by this notebook:
  - Influencers_posts_emoji_hashtag.csv - post captions with associated hashtags and emojis

12. model_preprocessing.ipynb

- This notebook conducts preprocessing of the datasets to be used as input for the models
- Files created by this notebook:
  - model_input.csv - preprocessed variables for the model

13. models.ipynb

- This notebook creates various models to predict brand popularity growth given a set of influencer metric inputs
- Files created by this notebook:
  - Models for the blog post