The folder contain following files:

**Twitter-Data-Extraction.ipynb:** Data Extraction code for January, February and March (2021)

(*Note*: Sample code for the month of February. Similar approach across granularity on days used to extract data for January and March)

**Data Cleaning.ipynb:** Data Cleaning (Stopwords & Punctuations removal, Lemmatization, Tokenization) on the extracted data

(*Note*: The code is worked only on tweets from January 2021. The code remains same for Feb and March with different input files)

**Individuals vs Businesses (Part1).ipynb:** Code for filtering out organizational tweets from personal tweets using bag-of-words on 2.74 million tweets

**Individuals vs Businesses (Part2).ipynb:** This code was used to addgranularity to organizational tweets separation and data cleaning

**Sentiment analysis\_Final.ipynb:** VADER sentiment analysis to classify tweets as positive, negative, neutral, overly-positive, overly-negative by running this vader sentiment analysis code

**Unsupervised\_LDA.ipynb:** This code helped us determine 18 different topics across the 12 week data and their corresponding keywords

**Semi-supervised\_Corex.ipynb:** Code for CorEx topic-modeling used to determine NTC (Normalized Topic Correlation) score for each observed theme across the 12 weeks