The work for the first deliverable was split up fairly well. The dataset regarding upcoming albums was collected by Trystan and preprocessed by Trystan and Ethan. The dataset regarding top 200 songs was collected and preprocessed by Ethan. The dataset regarding song count per artists was put together by David by deconstructing the top 200 songs dataset. The dataset regarding the lifespan of songs on the charts was put together by Trystan by deconstructing the top 200 songs dataset. Exploration into the Spotify API was completed by Ethan and exploration into the Twitter API was completed by Trystan. The main notebook has contributions from all three of us. The introduction and preprocessing summary was written by Ethan.

Our Git repository is organized into four folders. One folder is labeled “Datasets”, which houses cleaned datasets collected so far. Another folder is labeled “Processing”, which houses notebooks, python scripts and bash scripts related to retrieving the raw data from the web as well as preprocessing code. The next folder holds files related to Spike Stories, which is code we may end up use in the future if we decide to use the explored technologies. The final folder is labeled “Deliv\_1”, which houses the notebook that describes our process as well as out presentation information. To see our preprocessing process explore the folder labeled Processing.