ETL Project Report

Ryan Dickson and Philip Morlier

March 2020

ETL Project Report

The names 3,380 White House visitors, corporate fat cats, and revolving door lobbyists were correlated with tweets from President Donald Trump to assess the frequency with which the president tweeted about these individuals and to establish a database in which these tweets could be further analyzed.

With respect to materials, ProPublica maintains a dataset of individuals who visited specific officials at the White House during 2017 of the Trump administration (obtained through FOIAs and lawsuits). The log is highly incomplete: only a handful of White House official calendars have been produced, and, of those that have been produced, visitor names are often omitted, redacted, or illegible. In total, our analysis included 3,078 unique individuals who visited the White House in 2017. In addition, LittleSis maintains a list of so-called corporate fat cats—66 individuals who in some way benefited financially from the housing bubble. LittleSis also maintains a list of 242 revolving door lobbyists, which the site does not define. These three lists were stripped of all information except for name, resulting in 3,380 unique individuals. Details regarding the construction of these lists follows.

For the list of White House visitors, a Pandas dataframe containing only attendee names was created. All NaN responses were then dropped. Next, all redacted names—denoted by (b)(6)—were removed. Next, because attendees could contain more than one attendee, separated by a comma, the attendees’ column was exploded into two columns and only the attendee listed first was retained. Finally, all duplicate attendees were dropped. For the fat cat and lobbyist lists, all columns except for name were removed, which was then set as the index. The fat cats list and the lobbyist lists were then appended to the visitors list and the resulting list was checked for duplicates.

TrumpTwitterArchive, a website that contains every tweet from Trump, was scraped.