

1. There are four 'Generate' notebooks that use the Wikipedia Pageview API to create the four data files (csv files), one for each of the four classes of articles (historical people, historical institutions/ideas/events, world cities, modern people). The csv files are included in this folder, so no need to run the Generate notebooks.
2. There are four 'Analysis' notebooks that analyze each of the four classes individually, including running the clustering on them. In their current form they load the.pkl files (contained in this folder) for the clustering models created to run the clustering. There are four.pkl files, one for each of the four classes.
3. There is a single combined 'Analysis' notebook that analyzes all of the articles together. No clustering is run on the combined set, but the linear modelling component is run in this notebook.
4. There are four.html files that contain the visualization for the clustering of the four classes.
5. There is the 'Months_Max'.pdf file that contains the output of the articles with the largest relative view spikes in each month analysis, plus the probable reason for this spike.
6. There is the 'HistoryViews Final Report'.pdf file that is the final report for the project.