Datasets

Below is a non-exhaustive list of some data sources people may find interesting for their SLP. In general, Kaggle is a great resource to find interesting datasets. My recommendation is that people think about an industry or business function they are interested in and google “datasets for machine learning and analysis in {XX} industry”.

Also remember that students need to have a specific target/dependent variable in mind that they are interested in doing analysis against.

Dataset Lists

<https://github.com/justmarkham/DAT3/blob/master/public_data.md>

<https://learn.g2.com/open-data-sources>

<https://medium.com/analytics-vidhya/top-100-open-source-datasets-for-data-science-cd5a8d67cc3d>

Economic and Finance

<https://libguides.lib.cwu.edu/c.php?g=358343&p=2419792>

Banking

<https://www.kaggle.com/datasets/prakharrathi25/banking-dataset-marketing-targets>

Marketing

<https://www.kaggle.com/datasets/jackdaoud/marketing-data>

https://www.kaggle.com/datasets/rodsaldanha/arketing-campaign

<https://www.kaggle.com/datasets/henriqueyamahata/bank-marketing>

Tech Industry

<https://www.kaggle.com/datasets/arnavsmayan/netflix-userbase-dataset>

<https://www.kaggle.com/datasets/netflix-inc/netflix-prize-data>

Politics

[State of the Union Speeches](https://www.presidency.ucsb.edu/documents/presidential-documents-archive-guidebook/annual-messages-congress-the-state-the-union) – Requires some light web scraping

<https://data.world/datasets/election>

<https://electionlab.mit.edu/data>