Datasets and Sources

- https://www.kaggle.com This is one of the best sources for data science-related datasets. It has a full code notebook and analysis that help prompt you to discover and build new information.
- https://github.com/rfordatascience/tidytuesday A lot of datasets for analysis
- https://opendata.cityofnewyork.us NYC datasets
- https://fred.stlouisfed.org/ – Economic Data regarding the US
- https://www.data.gov/ – Data repo from the US government
- https://datasetsearch.research.google.com/ – Search engines for Datasets
- https://github.com/OpportunityInsights/EconomicTracker – This includes COVID lockdown dates, changes in local policy, unemployment changes, etc. at the state and local levels), employment, consumer spending, education related statistics, and Google/Apple mobility reports.
- https://paperswithcode.com/datasets – Papers with code datasets
- https://datahub.io/collections – This has a lot of data regarding finance
- https://archive.ics.uci.edu/ml/datasets.php – your source for your standard ML benchmark datasets – things like MSINT, Iris, Titanic, among plenty of others
- https://www.earthdata.nasa.gov/learn/find-data – Earth Science Data
- https://apps.who.int/gho/data/node.home – WHO global health data
- https://data.fivethirtyeight.com/ – US politics and sports
- https://github.com/BuzzFeedNews source data from Buzzfeed News.
- https://github.com/awesomedata/awesome-public-datasets – Some public datasets
- https://snap.stanford.edu/data/ – Several social media-related datasets
- https://research.google.com/youtube8m/ – 8 million categorized YouTube videos
- https://research.atspotify.com/datasets/ – lots of music/podcast related data
- https://huggingface.co/docs/datasets/index – Lots of great text datasets