Datasets

Gathering the right data is one of the most important task for data engineers. If you are looking for data for your own, self-directed project, or for data to enrich the data for the Udacity provided project, below are some good resources to find interesting datasets.

Analytics

- Google: Dataset Search
- Kaggle Datasets
- Github: Awesome Public Datasets
- Data.gov
- Dataquest: 18 places to find data sets for data science projects
- KDnuggets: Datasets for Data Mining and Data Science
- UCI Machine Learning Repository
- Reddit: r/datasets/

Software Development

- Github: Public APIs
- Last Call: Top 50 Most Popular APIs on RapidAPI (2018)
- Facebook: Graph API

Example

This Kaggle dataset is a good example of someone who found disparate datasets and combined them to provide an even more valuable dataset for others to analyze.