AirBnB Listings

“The data behind the Inside Airbnb site is sourced from publicly available information from the Airbnb site. The data has been analyzed, cleansed and aggregated where appropriate to facilitate public discussion.”
http://insideairbnb.com/get-the-data.html
Pick a city of your choosing!
Choose “listings.csv” for your city of choice: “Summary information and metrics for listings in <CITY>”.
If you really want to get into the weeds, use “listings.csv.gz”. Which has “Detailed Listings data for <CITY>”. (Has a lot more in depth stats that are summarized in listings.csv)

Spotify Dataset

Note: Lots of options for what you can do here. Predicting track metrics based on artist (e.g. loudness, “energy”). Predicting genre (multi-class) based on metrics/artists. Very fat tailed when it comes to author popularity, as expected. Also, who doesn’t like music?
https://www.kaggle.com/mrmorj/dataset-of-songs-in-spotify

Car Accidents

https://www.kaggle.com/sobhanmoosavi/us-accidents
“This is a countrywide car accident dataset, which covers 49 states of the USA. The accident data are collected from February 2016 to Dec 2020, using multiple APIs that provide streaming traffic incident (or event) data. These APIs broadcast traffic data captured by a variety of entities, such as the US and state departments of transportation, law enforcement agencies, traffic cameras, and traffic sensors within the road-networks. Currently, there are about 3 million accident records in this dataset”

Airlines

“This dataset is all about flights in the United states, including information about the number, length, and type of delays. The data is reported for individual months at every major airport for every carrier.”
https://corgis-edu.github.io/corgis/csv/airlines/