readme-dataset

This is a dataset of British steam trains in the time period of 1921-1968.

IMPORTANT: This is not a totally complete set, and only has some kinds of data, as some have been left out. Also not every train is in the dataset as some classes have been cut due to not being relevant with having all trains of the classes if only a couple of them manage to become part of BR or the Big four, like with the WD austerity 2-8-0.

Sources
This dataset is made out of multiple digital sources. The primary sources are the BRdatabase: https://web.archive.org/web/20221017134057/https://www.brdatabase.info/index.php and Southern Railways E-mail group: https://web.archive.org/web/20220521100132/https://sremg.org.uk/photoind.shtml
Other sources include:
-railuk.info (For getting info on missing data like tonnage) https://web.archive.org/web/20221130223237/https://railuk.info/steam/
-The London & North Eastern Railway Encyclopedia (for information on LNER trains) https://web.archive.org/web/20221024223432/https://www.lner.info/locos/locos.php
-steamlocomotive.com (information on old LMS and GWR trains) https://www.steamlocomotive.com/
-hattons.co.uk (help with finding trains roles) https://web.archive.org/web/20221130072636/https://www.hattons.co.uk/
-Wikipedia Articles on GWR, LMS and BR locomotives, along with articles on Narrow gauge and Minimum railway
-https://web.archive.org/web/20220530013709/https://en.wikipedia.org/wiki/Locomotives_of_the_London,_Midland_and_Scottish_Railway
-https://web.archive.org/web/20221128230317/https://en.wikipedia.org/wiki/Locomotives_of_the_Great_Western_Railway
-https://web.archive.org/web/20221014200320/https://en.wikipedia.org/wiki/Locomotives_of_the_Southern_Railway
-https://web.archive.org/web/20220627225515/https://en.wikipedia.org/wiki/British_narrow-gauge_railways
-https://web.archive.org/web/20210309114010/https://en.wikipedia.org/wiki/List_of_British_private_narrow-gauge_railways
-https://web.archive.org/web/20220804211928/https://en.wikipedia.org/wiki/British_industrial_narrow-gauge_railways

Variable
The dataset has 10 variables(technical 11) to 35034 objects (objects are tied to the variable name in the table). Those 10 variables are Whyte notation, Class, Gauge, weight, Roles, Company, builddato, scrapdato, buildplace and designer.
-Names: Names Consist of a number and company abbreviation, and if there is a name for a train, then that is as well. I have been going with using companies numbers system if it covers the whole class. With companies like LNER and their renumbering, I went with their original number system. With some classes, I took the BR number system due to it being more organised.

-Whyte notation: a classification method based on wheel arrangement. The basic form counts the number of leading wheels, then the number of driving wheels, and finally the number of trailing wheels, numbers being separated by dashes. A train with a tender would use a normal Whyte notation, but for those without one, letters would be put behind the number of trailing wheels.
T is for Side tank locomotive.
ST is for Saddle tank locomotive.
WT is for Well tank locomotive.
PT is for Pannier tank locomotive.
T+T is for tank locomotives with a tender
VB is for Vertical boilered locomotive.
Articulated locomotives have two wheelsets, so it would be a pair of whyte notations with + between them.
-Class: A design that trains are built to, aka to ship classes.
-Gauge: The gauge is the distance between the two rails of a railway track. In this dataset, there are 3 types of gauge; Standard gauge (1,435 mm) the most common gauge. Narrow gauges are track spaced significantly narrower than Standard. Minimum-gauge are the gauge of under 2 feet (610 mm)
-weight: The tonnage of the train, which is in British Longton. For all tender trains, only the weight of the train itself is used.
-Roles: Roles or functions describe the main use of the train or the designed use. This dataset used 11 functions to put the trains in.
  -express Passenger (a passenger train that with great speed travel over long distance no stop between major destinations. They are among the fasted and biggest trains)
  -Passenger (a generic term for trains on passenger duty not of suburban, branch line or an express nature)
  -Freight (a generic term for goods and freight train.)
  -Bank engine (trains that assist other trains that requires additional power or traction to climb a gradient)
  -mixed-traffic () trains that are built to do both freight and passenger duty)
  -suburban services (trains that run passenger duty in a metropolitan area and that need high amounts of speed and acceleration)
  -shunting (Train build to sorting items of rolling stock into complete trains or the reverse)
  -branch services (train build for use on Branch line)
  -mineral traffic (Heavy freight trains used for coal and other mineral goods)
  -Unknown (For trains, where no data on their function can be found)
  -Crane engine (trains with a crane on it)
-Company: The variable of which companies have used the trains. All values are companies abbreviation. Here the list is from BRdatabase. Minimum and Narrow gauge railways are not on this list:
  -LNER: London and North Eastern Railway
  -LMS: London, Midland & Scottish Railway
  -SR: Southern Railway
  -GWR: Great Western Railway
  -BR: British Railways
  -WD: war department
  -GCR: Great Central Railway
  -GER: Great Eastern Railway

-GNR: Great Northern Railway
-GNSR: the Great North of Scotland Railway
-HBR: Hull & Barnsley Railway
-MGNR: the Midland & Great Northern Railway
-NBR: North British Railway
-NER: North Eastern Railway
-LSWR: London and South Western Railway
-LBSCR: London, Brighton and South Coast Railway
-SER: the South Eastern and Chatham Railway
-CR: Caledonian Railway
-FR: Furness Railway
-GSWR: Glasgow and South Western Railway
-HR: Highland Railway
-LNWR: London and North Western Railway
-LTSR: London, Tilbury & Southend Railway
-LYR: Lancashire and Yorkshire Railway
-MCR: Maryport & Carlisle Railway
-MR: Midland Railway
-NSR: North Staffordshire Railway
-SDJR: Somerset & Dorset Joint Railway
-BaR: Barry Railway
-BMJR: Brecon and Merthyr Junction Railway
-BPGVR: Burry Port and Gwendraeth Valley Railway
-CambR: Cambrian Railway
-CarR: Cardiff Railway
-MSWJR: Midland and South Western Junction Railway
-PM: Powlesland and Mason
-PTR: Port Talbot Railway
-RR: Rhymney Railway
-SHT: Swansea Harbour Trust
-TVR: Taff Vale Railway
-VofR: Vale of Rheidol

-builddato: The date of build/intro to service. Both this variable and scrapdato are in years only due to varying amounts of data.
-scrapdato: The date of withdrawal as for most cases, trains get scrapped within the same year they are withdrawn. Trains that survive today have the value of Preserved instead of the year of withdrawal.
-buildplace: This variable is what place built the train, may it be a company or a locomotive work. Based on the primary sources and Wikipedia.
-designer: The designer is who is credited with creating the design of the train. This is based on the two primary sources. Some train designs may not have a personal tie, so it is the building company that will get the credit for it in the dataset. There are a few unique cases, where I have to use NA/Unknown.

rebuild
Major rebuilds that change a train from one class to another are shown with that train having values from both classes that are separate by “ ; “ This affects class, ton, roles, and sometimes designer.

Use and goal
This dataset was created due to the goal of having a united spreadsheet for my project, but also to give the railway community a dataset to work with and to build further on. The hope is this would serve as a framework for future datasets. And the hope that some people would build more on it or use it for their own projects.