TLDR: If you just want to have rough year estimates going back 52 eras from the latest round is going back 1 year from now.

The era numbering in the combined train / validation set end always on the saturday 1 week before the latest saturday where the live weekend round started. From this point, going back one era is equivalent to going 1 saturday back.

But keep in mind that this is the date of the round opening, the targets are always 20 Numerai trading days (Tue,Wed,Thu,Fri,Sat) ahead starting from the Friday after the round opening.

Here is an example: Round 424 will be era 1051 and opened on 18th February 2023 (Sat), the trading starts on the 24th February 2023 (Fri), after 20 numerai trading days the round resolves on the 2rd March 2023 (Thu).

If you want to have these dates for previous rounds, you have to go back in a calendar 1 week for each era.

As there are 52 weeks + 1 day (2 days in leap years) per year, and there are at this point 1052 eras available, there are roughly 1052/52 ~ 20 years of data, so the train dataset begins at some point in 2003.

Edit

: Using the python datetime module, you can do the following:

import datetime era_count = 1050 days_per_week = 7 era1051_opening = datetime.date(2023, 2, 18) print(era1051_opening - datetime.timedelta(days_per_week*era_count))

Will result in 2003-01-04

. With this information you can modify your dataframes like this:

import pandas as pd df = pd.read_parquet("train.parquet") df["era"] = df["era"].astype(int) df["opening_date"] = [datetime.date(2003,1,4) + datetime.timedelta(7*e) for e in df["era"]-1]