Columns

nyra\_start\_table.csv

* track\_id - 3 character id for the track the race took place at. AQU -Aqueduct, BEL - Belmont, SAR - Saratoga.
* race\_date - date the race took place. YYYY-MM-DD.
* race\_number - Number of the race. Passed as 3 characters but can be cast or converted to int for this data set.
* program\_number - Program number of the horse in the race passed as 3 characters. Should remain 3 characters as it isn't limited to just numbers. Is essentially the unique identifier of the horse in the race.
* weight\_carried - An integer of the weight carried by the horse in the race.
* jockey - Name of the jockey on the horse in the race. 50 character max.
* odds - Odds to win the race passed as an integer. Divide by 100 to derive the odds to 1. Example - 1280 would be 12.8-1.
* position\_at\_finish - An integer of the horse's finishing position. (added to the dataset 9/8/22)

nyra\_race\_table.csv

* track\_id - 3 character id for the track the race took place at. AQU -Aqueduct, BEL - Belmont, SAR - Saratoga.
* race\_date - date the race took place. YYYY-MM-DD.
* race\_number - Number of the race. Passed as 3 characters but can be cast or converted to int for this data set.
* distance\_id - Distance of the race in furlongs passed as an integer. Example - 600 would be 6 furlongs.
* course\_type - The course the race was run over passed as one character. M - Hurdle, D - Dirt, O - Outer turf, I - Inner turf, T - turf.
* track\_condition - The condition of the course the race was run on passed as three characters. YL - Yielding, FM - Firm, SY - Sloppy, GD - Good, FT - Fast, MY - Muddy, SF - Soft.
* run\_up\_distance - Distance in feet of the gate to the start of the race passed as an integer.
* race\_type - The classification of the race passed as as five characters. STK - Stakes, WCL - Waiver Claiming, WMC - Waiver Maiden Claiming, SST - Starter Stakes, SHP - Starter Handicap, CLM - Claiming, STR - Starter Allowance, AOC - Allowance Optionl Claimer, SOC - Starter Optional Claimer, MCL - Maiden Claiming, ALW - Allowance, MSW - Maiden Special Weight.
* purse - Purse in US dollars of the race passed as an money with two decimal places.
* post\_time - Time of day the race began passed as 5 character. Example - 01220 would be 12:20.

nyra\_tracking\_table.csv

* track\_id - 3 character id for the track the race took place at. AQU -Aqueduct, BEL - Belmont, SAR - Saratoga.
* race\_date - date the race took place. YYYY-MM-DD.
* race\_number - Number of the race. Passed as 3 characters but can be cast or converted to int for this data set.
* program\_number - Program number of the horse in the race passed as 3 characters. Should remain 3 characters as it isn't limited to just numbers. Is essentially the unique identifier of the horse in the race.
* trakus\_index - The common collection of point of the lat / long of the horse in the race passed as an integer. From what we can tell, it's collected every 0.25 seconds.
* latitude - The latitude of the horse in the race passed as a float.
* longitude - The longitude of the horse in the race passed as a float.

nyra\_2019\_complete.csv - This file is the combined 3 files into one table. The keys to join them trakus with race - track\_id, race\_date, race\_number. To join trakus with start - track\_id, race\_date, race\_number, program\_number.

* track\_id - char(3)
* race\_date - date
* race\_number - char(3)
* program\_number - char(3)
* trakus\_index - int
* latitude - float
* longitude - float
* distance\_id - int
* course\_type - char(1)
* track\_condition - char(3)
* run\_up\_distance - int
* race\_type - char(5)
* post\_time - char(5)
* weight\_carried - int
* jockey - char(50)
* odds - int
* position\_at\_finish - An integer of the horse's finishing position. (added to the dataset 9/8/22)