1. **Size**

There are 109085 rows (horses in matches) in the data. There are 8672 matches in the data.

There are 61 columns.

Reference: https://special.hkjc.com/racing/info/en/betting/guide.asp

1. **Variables**
2. index 編號，無關變量
3. Rdate: Race date 比賽日期，可篩選走過早以前的比賽
4. Rid: Race ID 比賽編號，無關
5. Hid: Horse ID 馬匹編號，無關
6. Venue: Race venue (HV, ST) 不同馬場不同場地條件
7. Track: Race track (TURF, AWT) 草地跑道與泥地跑道
8. Going: Race track condition 賽道當日狀況，the track surface of a horse racing track prior to a horse race or race meet. The going is determined by the amount of moisture in the ground and is assessed by an official steward on the day of the race.
9. Course: Race course (AWT will have no specific race course description) 跑道資料，其中AWT 為泥地跑道，應與其他樣本分離。詳情可參考 https://racing.hkjc.com/racing/chinese/racing-info/racing\_course.asp
10. Class: Race class of the match 賽事等級，數值越小代表水平越高
11. Distance 比賽距離
12. Rfinishm: Race finish time in centi second (1/100 second) 一場比賽的結束時間
13. Rm1: 1st section finish time of the race in centi second
14. Rm2: 2nd section finish time of the race in centi second
15. Rm3: 3rd section finish time of the race in centi second
16. Rm4: 4th section finish time of the race in centi second
17. Rm5: 5th section finish time of the race in centi second
18. Rm6: 6th section finish time of the race in centi second
19. Horsenum: the horse ID of the horse 馬匹數據庫主鍵，無關
20. Jname: Jockey 騎手
21. Tname: Trainer 訓練師
22. Exweight: the handicapped weight carried by the horse 配重賽配重，A better horse will carry a heavier weight, to give it a disadvantage when racing against slower horses.
23. Bardraw: draw 出發閘口抽籤結果，通常內道更有優勢，但實際上要考慮：horse’s favorite track, distance, running style.
24. Gear: Gear putting on the horse 馬匹的額外裝備，有900多種組合，比較難進行數值化處理
25. Rating: Horse Rating The higher the rating, the better the horse. Each rating point equates to one pound of weight. So a horse rated at 115 should carry 7 pounds more than a horse with an OR of 108. It is an essential component of handicapping because it enables official handicappers to decide how much weight a horse should carry in each race.
26. Ratechg: Rating change of the horse from previous race rating 變化量，可用於衡量狀態？
27. Horseweight: Horse’s weight 馬匹重量
28. Horseweightchg: The Horse’s weight change from previous race 馬匹重量變化，可用於衡量狀態？
29. Besttime: The best finishing time of the horse on the race with same venue, distance and track (minute. second. centi second) 過往最佳成績
30. Age: age of horse 馬匹年齡
31. Priority: Priority of the horse for race given by the trainer. Trump Card Policy，非常難理解，暫時可認為騎師對priority 高的馬匹更有信心
32. Lastsix: the rank from previous 6 races 儲存格式為 '7/6/8/2/11/7'，建議轉化為平均ranking
33. Rank: rank in current match groundtruth，可對每一匹馬進行分類
34. Runpos: The rank in each section of the horse in the race. 馬匹在不同小節的排名，反映馬匹比賽策略
35. P1: The rank in 1st section of the horse
36. P2: The rank in 2nd section of the horse
37. P3: The rank in 3rd section of the horse
38. P4: The rank in 4th section of the horse
39. P5: The rank in 5th section of the horse
40. P6: The rank in 6th section of the horse
41. M1: The finish time of the 1st section of the horse in centi second
42. M2: The finish time of the 2nd section of the horse in centi second
43. M3: The finish time of the 3rd section of the horse in centi second
44. M4: The finish time of the 4th section of the horse in centi second
45. M5: The finish time of the 5th section of the horse in centi second
46. M6: The finish time of the 6th section of the horse in centi second
47. Finishm: The race finish time of the horse in centi second groundtruth，可用於預測馬匹完成比賽的時間
48. D1: The distance from the rank 1 horse in the 1st section (0.25 meant the distance is within 1 horse distance)
49. D2: The distance from the rank 1 horse in the 2nd section (0.25 meant the distance is within 1 horse distance)
50. D3: The distance from the rank 1 horse in the 3rd section (0.25 meant the distance is within 1 horse distance)
51. D4: The distance from the rank 1 horse in the 4th section (0.25 meant the distance is within 1 horse distance)
52. D5: The distance from the rank 1 horse in the 5th section (0.25 meant the distance is within 1 horse distance)
53. D6: The distance from the rank 1 horse in the 6th section (0.25 meant the distance is within 1 horse distance)
54. Datediff: the date difference between the previous match and current match of the horse 馬匹比賽間隔，即備戰時間，可衡量狀態？
55. pricemoney: prize money of the race 獎池獎金
56. Windist: The distance from the Rank1 horse. 可用作 groundtruth，但可能過於複雜
57. Win\_t5: win odds of the horse 5 minutes before the race 冠軍賠率
58. Win: final win odds of the horse 無用
59. Place\_t5: place odds of the horse 5 minutes before the race 前三名賠率
60. Place: final place odds of the horse 無用
61. Ind\_win: indicator of winner of the race (1 for winner, 0 otherwise) groundtruth
62. Ind\_pla: indicator of top 3 places in the race (1 for top 3, 0 otherwise) groundtruth