# **Project 03 – Tour De France Case Study**

1. **Input Data**

* Use the **Input Data** tool to load the following CSV files:
  + tdf\_winners.csv
  + tdf\_tours.csv
  + tdf\_finishers.csv
  + tdf\_stages.csv

## **Actions to be Done**

1. For numeric types, replace nulls with 0.
2. For non-numeric types, replace nulls with empty strings.
3. Join tdf\_winners.csv with tdf\_tours.csv using a common key (e.g., Tour\_ID).
4. Join tdf\_finishers.csv with tdf\_stages.csv using a common key (e.g., Stage\_ID).
5. Append any additional data from tdf\_tours.csv to the tdf\_winners to enrich the dataset.
6. If there are any date fields in tdf\_winners.csv or tdf\_tours.csv, convert them to a standard format (e.g., yyyy-MM-dd).
   1. Extract Year, Month, and Day of the week from the relevant date fields.
7. Split any delimited text fields (e.g., multiple tour names) into separate columns.
8. Use tdf\_finishers data to show the total number of finishers per stage across different tours as columns.
9. From the tdf\_winners sheet, calculate:
   * 1. Total number of winners per tour.
     2. Average finishing time for winners.
     3. Total number of finishers per stage.
10. Create a Full Winner Description field by combining relevant details from tdf\_winners.csv.
11. Calculate the total number of participants in each tour.
12. Filter winners based on specific criteria (e.g., only winners from a certain year).
13. Filter finishers that are above a certain performance threshold.
14. Find unique tours from the tdf\_tours.csv.
15. Extract a list of unique winners from the tdf\_winners.csv.
16. Add a Record\_ID to the tdf\_winners sheet to simulate serial numbers.
17. Get the first 10 winners from the tdf\_winners.csv.
18. Randomly sample 5 records from the tdf\_finishers.csv for quality checks.
19. Select records between the 11th and 20th rows from the tdf\_stages.csv.
20. Sort the tdf\_winners sheet by finishing time from fastest to slowest.
21. Sort tdf\_finishers data to find the best performers based on their finishing times.
22. Use the tdf\_tours.csv to replace tour codes/names in the tdf\_winners.csv using Tour\_ID.
23. Save the final processed data to a new CSV file for further analysis or reporting.