# About the Dataset

This data was pulled from a CSV file on Kaggle.com. The data shows the finishing time of each participant, their team, rank and year they participated in Tour De France.

URL: https://www.kaggle.com/datasets/sujaykapadnis/le-tour-de-france-database-1903-2023

Tour De France Background

Tour De France is an annual bicycle race that takes place in France that covers about 2,235 miles. Riders complete this course in 21 stages within 23 days. The race has about 20 professional teams and about 9 riders for each team. The winner is determined based on fastest average speed and placement within each stage. Tour De France started in 1902 and included male riders only up until the first Tour De France Femmes in 1955.

Dataset Limitations

* Riders average speed are slightly inaccurate as the field was calculated using the distance formula on the finishing time and distance for each participant. This may be different than the official method used to calculate the cyclist average speed.
* Although the Tour De France Femmes started in 1955 this dataset only has records for 2022 and 2023. This impacts the accuracy in comparison of average speeds between male and female participants.

# Data Analysis

## Data Import and Cleaning

* Import Tour De France CSV file and Tour De France Femmes file

## A screenshot of a computer Description automatically generated

## A screenshot of a computer Description automatically generated

* Calculate average speed and add a gender column to identify the gender of each participant

A screenshot of a computer program

Description automatically generated

## Data Visualizations and Findings

* Average speed comparison on gender

A screenshot of a computer

Description automatically generated

* Average Speed Line Graph 1902 – 2023

A graph on a computer screen

Description automatically generated

Result: Cyclist average speed has an overall positive trend since the start of the race. The trend levels out around 1939 then increases again around 1947. This is because Tour De France was cancelled between 1940 – 1946 due to Frace’s involvement in World War 2.

* Are the changes in speed due to a change in the distance of the race?

A screenshot of a computer

Description automatically generated

* Were cyclist achieving higher speeds than the years before?

A screenshot of a computer

Description automatically generated

Result: The data show that there is about a 3 kph difference between the top cyclist in 1926 in comparison to 1927. Also, the cyclist in positions 2-5 achieved a higher overall average speed than the winner in 1926. A similar trend occurs between 1938 and 1939. This shows that overall the cyclist were achieving higher average speeds.

* Scatter plot representing the top 10 cyclist from 1902 – 2023

A screen shot of a graph

Description automatically generated

* What’s the record for the most Tour De France wins?

A screenshot of a computer

Description automatically generated

Result: The maximum number of tour the France wins is 5 and was achieved by 4 cyclist. Miguel Indurian is the only cyclist with 5 consecutive Tour De France wins.

* Which team holds the record for the most Tour De France wins?

A screenshot of a computer

Description automatically generated