# Introduction

The Tour De France is with its 150 million viewers (<https://www.statista.com/statistics/1254337/tour-de-france-tv-audience/>) per year the biggest and the most prestigious race on the calendar. It consists of 21 individual stages spread across three weeks. Each stage is timed, which is how the race leader and eventually the overall winner is decided.

The TDF stages fall into one of five different categories: flat, mountainous, time trials, team time trials and hill climbs.

At the end of each stage four important jerseys are awarded that the riders will wear during the next stage (https://www.youtube.com/watch?v=p3jlG3XBDjM&ab\_channel=GlobalCyclingNetwork):

* *Maillot Jaune*: that is awarded to the current leader of the general classification
* *Maillot Blanc*: which denotes the best young rider in the race
* *Maillot Vert*: is typically worn by a sprinter, who will accrue points by winning stages and finishing regularly in the top 15 on stages they don’t win. The jersey is worn by the rider with the most points.
* *Maillot à pois*: is typically awarded to the best climber, who will accrue points by winning mountainous stages or by finishing in the top 15. The harder the climb the more points that are on offer.

A team sponsors goal is to maximize name exposure so winning one of the Jerseys is very important. Few teams have Maillot Jaune contenders on their roster so what they’ll typically do is focus on either obtaining the Maillot à Pois or the Maillot Vert as the wearer of one of the jerseys receives significant coverage during the race. Especially the Maillot Vert wearer. For the smallest teams winning a stage is the goal.

In this report we explain how we built a prescriptive model with which a team can decide which riders to select to maximize their chances of obtaining one of the Maillots or to win a stage. A team can either build a team that’s more suited to winning flat stages or it can build a team that focusses on collecting.

# Objective function

We argue that because of the importance of the Tour De France a team’s goal will be to send it’s best possible group of riders to achieve its goal. Riders are ranked based on their overall rating, which is the average of a player’s PCM stats. The best possible team is the team that maximizes the average overall rating.

Max

Constraints:   
-

* Xi = Integer
* Xi ε [0,1]

Where Xi is *“do I select rideri”*and i=1…number of riders in the team.

# Baseline model

We calculated the optimal list for the test sample without using any auxiliary data by calculating the overall rating of each member of the team and then ordering them by the largest overall rating. With the baseline model where we don’t take into account the probabilities to win we assume that the 8 best riders to select are the 8 riders with the largest overall rating. If we take Intermarché-Wanty-Gobert as an example: the baseline objective value is 71.75 (Appendix).

To build the full information decision model for finishing top 5 in one of the flat stages we collected all the teams of the 2021 edition of the TDF that finished top 5 for at least one of the flat stages.   
If we again take Intermarché-Wanty-Gobert as an example: the objective value for the team with which they finished top 5 at least once is 70.48 (Appendix).

Although the baseline model produces an overall better team than the team that finished top 5 in a flat stage, the probability of the baseline model team to win a flat stage is only 56%.

# Data-driven model

## Goals

The goal is to:

* + To predict whether a team is likely to finish top 5 in a particular stage.
  + To select the riders that maximizes a team’s probability to win a stage

Winning by points means a team needs to be well enough to finish top 10 in most races