



## CFG'S DATA & SQL MOOC SPRINT & CHALLENGE

### THE CHALLENGE

**Outline:** To consolidate the skills developed throughout the four lessons, students will have **7 days** to complete the challenge set out in this document. The skills required to complete this challenge will have been learnt throughout the four lessons attended by students, and will be based upon a 'real world' scenario. This challenge will allow students to demonstrate their skills in SQL, which can in-turn be used as evidence to current, prospective and future employers.

**Rules:** This challenge is open book, and will allow students to complete the challenge using study aids and online materials. However, plagiarism is not allowed. Students will be expected to submit their project no later than **Wednesday 5th October at 11:59pm**. Anything received after this timeframe will be treated on a case-by-case basis, and this will depend on any extenuating circumstances.

**Purpose:** Through contextualising theoretical skills and applying this to a real-world problem, students will learn how data professionals use SQL to solve issues, and provide actionable insights to organisations.

**Output:** Students will submit the following:

- An **EER diagram** of the tables within their database, noting the relationships and fields included;
- Evidence of the code used to support analysis, including **DDL & DML commands** properly assigned keys and aggregate functions.
- A presentation or **report** answering the key questions set out within the scenario.

## The Challenge

**Scenario:** CFGRacing has provided you with following Formula 1 World Championship datasets recorded from 1950 to 2022:

- circuits.csv: *records of the racing tracks & locations*
- constructors.csv: *records of the car constructors*
- drivers.csv: *records of the drivers' details*
- lap\_times.csv: *records of the drivers' individual lap times for every race*
- pit\_stops.csv: *records of pit stops taken by drivers for every race*
- qualifying.csv: *records of the results of qualifying with the positions of the drivers*
- races.csv: *records of all races in F1 history*
- results.csv: *records of all results of races + fastest lap time*
- sprint\_results.csv: *records of all Friday sprints*
- status.csv: *records of all possible racing statuses*

### **Task:**

Criteria	Definition	Weighting
Innovation	Does the response take a fresh approach to the Challenge?	25%
Analytical Skills	Have the team demonstrated understanding of the DDL and DML concepts and all 5 aggregate functions covered in the MOOC Sprint?	25%
Presentation	How well have the code files, database design and Challenge response been presented?	25%
Adherence to Challenge theme	Did the team adhere to the Challenge?	25%

**Tips & Tricks:**

- As you are working with a very large dataset, MySQL will need some time to import your data properly. It is normal for it to take up more time than usual.
- Time tracking in the datasets: on import, you can keep "time" as text and have milliseconds set as bigint so you can use it for your calculations. It is also recommended to convert durations to milliseconds if you wish to work with them.

**F1 Dictionary:**

If you're unfamiliar with or new to F1, you can watch [this video](#)

If you want to know the rules of F1, you can watch [this video](#)

It's important to understand the structure of the F1 Weekend to understand the difference between practice, qualifying & racing as well as sprints which have started in 2021, you can read the section about Race Weekend [here](#) and/or watch [this video](#) for explanation.

*Good luck!*