Revised 2020 FIA Formula One Calendar

1. Introduction

1.1. Background

The FIA Formula One World Championship is the highest class of single-seater racing. The 2020 season was intended to run from March to December and comprise of 22 races, in 22 different countries. However, the COVID-19 crisis caused travel and mass gatherings to be severely restricted and, consequently, the first 10 races of the season to be abandoned. With the sport's commercial rights' holders and, more importantly, the teams themselves getting most of their income from broadcast rights, race fees and sponsorship details, this presented the sport with a series problem.

The governing body, the FIA, and teams believe the best way to recover their income – keep their companies in business and employees employed – would be to run an 18-race season over a condensed period of 6 months.

1.2. Business Problem

Condensing the season introduces practical and logistical problems for the teams, who must ship equipment around the world, and team members, who must do so much travelling in such a short space of time. There are two questions that ned answering:

1. Which 4 of the 22 races should be cut from the calendar?
2. How should the season be arranged to minimise the strain on the teams and travelling team members?

The answers to these questions would be of interest to the FIA and the teams who are currently trying to put a new race calendar together

2. Data

2.1. Data Required

To answers the two questions posed in 1.2, we must first decide which 4 races to remove from the calendar. Data pertaining to the relative importance of each race to the sport would offer an impartial metric in which to choose the races to leave out. For this, a metric of interest in each race will be sort such that the races that fans are least interested in can be cut from the calendar. This will likely keep the sponsors happy as their primary reason for their involvement in the sport is advertising. Keeping the races that have the most fans will mean the best advertising.

Once the races have been decided upon, the order in which they are to run will needs deciding such the pressure on logistics and travelling team members is minimised. For this, clusters of races will be proposed that minimise travel between them and allow a break for team members between clusters.

2.2. Data Sources

The current 2020 calendar will be sourced from scraping data from Wikipedia and parsing with BeautifulSoup. A measure of interest in the races will be found by leveraging FourSquare data. The number of photos and 'likes' at each race circuit will be used as a metric for interest in that particular race. This information will then be used to remove the 4 races with the least interest.

The final list of 18 races will then be clustered by region to provide the proposed final calendar.