

Tasks that we have completed and pending tasks (question 1 and question 2):

1. Web Crawling Setup & Data Extraction (40 hours)
 - a. Identify and list Formula One data sources (like formula1.com, espn.com/f1, etc.).
 - b. Develop scripts using Python, BeautifulSoup, and requests for web scraping.
 - c. Extract relevant data such as driver statistics, team information, race results, etc.
 - d. Ensure data accuracy and handle any web scraping challenges.
2. Website Creation & Hosting (30 hours)
 - a. Design the website layout and user interface.
 - b. Implement the front-end using suitable technologies (HTML, CSS, JavaScript).
 - c. Develop back-end functionalities using Flask.
 - d. Set up a local server for hosting the website.
 - e. Integrate the scraped data into the website.
3. Creating Statistics and Data Visualizations (20 hours)
 - a. Analyze the collected data to find interesting patterns and statistics.
 - b. Create visualizations (graphs, charts) to represent data effectively.
 - c. Implement these visualizations on the website.
 - d. Ensure that data is presented in a user-friendly and accessible manner.
4. Setup Feedback Collection System (10 hours)
 - a. Develop a system to collect user feedback.
 - b. Implement features for users to submit feedback directly through the website.
 - c. Plan for processing and responding to feedback.
 - d. Use feedback for continuous improvement of the website.

Here are all the tasks that we are planning to complete. We separated the project into 5 sections: **Web Crawling Setup & Data Extraction, Website Creation & Hosting, Creating Statistics and Data Visualizations, Setup Feedback Collection System.**

We have just completed the first part of our project, which is the **Web Crawling Setup & Data Extraction**. A comprehensive web crawling system has been established, utilizing tools like Python, BeautifulSoup, and requests. This system has enabled the extraction of a wide range of data essential to our project, including detailed driver statistics, team information, and race results from various authoritative Formula 1 data sources. API setup has been implemented for the data retrieval. We are able to obtain up-to-date information. The APIs have been integrated for continuous data flow. The development focuses on maintaining the integrity and accuracy of the data. The web crawling process has been tailored to handle the complexities of different data structures and formats found on various F1-related websites.

We will need to complete part 2 to part 4 to complete the project

Challenges that we have faced (question 3):

The data extracted from various sources through web crawling and APIs has presented some inconsistencies in format and quality. Ensuring uniformity and accuracy across diverse data sets can be challenging. We have to be ready to handle cases when some of the data is not obtainable, maybe the website is down or the API is being down during maintenance.