A black background with blue text

Description automatically generated

**Course Number: MMA 860**

**Course Name: Acquisition and Management of Data**

**Assignment Name: Project**

**Due Date: July 12, 2024 12pm**

**Team Name: Team Gordon**

|  |  |
| --- | --- |
| **Student Name** | **Student Number** |
| Alisha Sahota | 20497348 |
| Anthony Ramelo | 20499391 |
| Chris Wu | 10182394 |
| Elizabeth Zhang | 20161231 |
| Emily Zhao | 10096273 |
| Sam Hossain | 20466500 |

## Order of files:

|  |  |  |
| --- | --- | --- |
| **Filename** | **Pages** | **Comments and/or Instructions** |
| MMA 860 Project Outline\_Team Gordon | 4 |  |

**Additional Comments:**

|  |
| --- |
|  |

Formula 1: Factors Impacting Race Outcome

### Summary

Formula 1 is the pinnacle of motorsports and one the world’s most prestigious motor racing competition. There are 10 teams and 20 drivers racing, and 24 races that take place all over the world. Driver’s go as fast as 370km/h and decisions are made in a blink of an eye.

Formula 1 is a data driven sport where each millisecond makes a difference to the race outcome. Over the last decade, teams have collected large amounts data to better understand where they can improve race performance. While the driver’s skill and experience are a large part of the race outcome there are other factors such as, number of pitstops, starting position in the race, lap times that influence the race. In this analysis, we are going to be looking at each of these factors to determine what teams should focus on to get the best race outcome.

### Methodology

1. **Data source: Ergast Formula 1 API**
2. **Data Cleaning: Files 14 different, these files will need to be merged together to one dataframe. Treat for any nulls, and no duplicates.**
3. **Descriptive Statistics: Examine each of the columns for the correlation in next step**
4. **Correlation Analysis: Using corr() function, find the correlation race wins and the other columns in dataframe**
5. **Linear Regression Analysis: Look at R2 and determine if the columns in the Correlation analysis are significant to race wins.**
6. **Create a conclusion from the analysis**

### Expected Outcomes

The project is expected to deliver a detailed analysis and model of factors that influence the race outcome. Using linear regression and various techniques learned from this course.