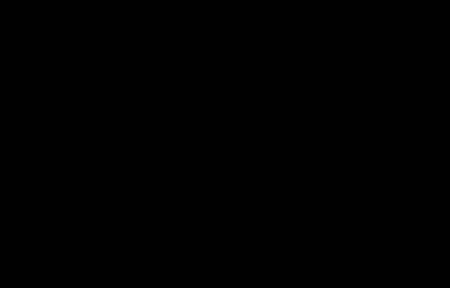


***Iteration Zero Proposal – Formula 1***

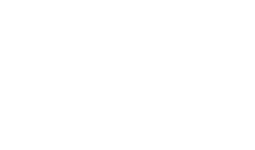
***Kaloyan Rakov***

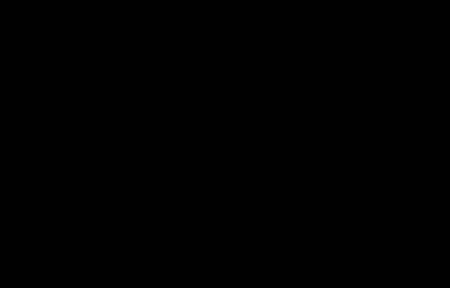


****

**Table of Contents**

1. **Introduction**
   1. **What?**
   2. **Why?**
   3. **Who?**
   4. **When?**
   5. **How?**
2. **Analytic Approach**
   1. **Defining the target variable(s)**
3. **Domain Understanding**
4. **Data requirements**



****

1Intrduction

What is Formula 1?

Formula 1 is a motorsport race, including 10 teams with 2 drivers each. The race is based around 2 titles: the Drivers' Championship (individual for each driver) and the Constructor’s Championship (Team-based).

* What?

I am going to start this project off by designing a machine learning model, that could be used for predicting the potential winners of the races. As the project continues, the precise target variables might change into analyzing team strategies and other factors.

* Why?

Formula 1 is a business that generates loads of money, and many people involved might want to have a more hands-on approach on the data. The sport is also very harmful to our planet, so the data might be used to move into a greener direction.

* Who?

To ensure that the project is going to be useful, I will conduct a stakeholder interview with a person in the field.

Formula 1 includes 10 teams:





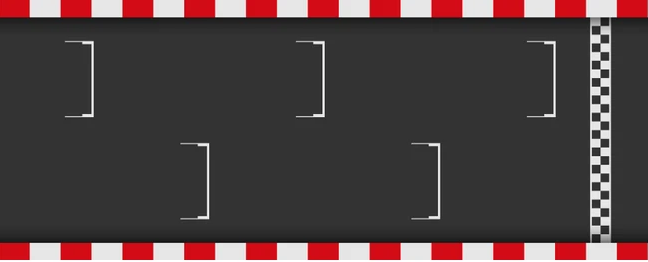
* How?

Many different algorithms can be used for precise prediction. I aim to have a deeper project than simple winner prediction, so I may have to go through many different models until the most precise is clear. As I get deeper into the project, new target variables may appear such as tire management, fuel management, etc.

* When?

After this initial iteration 0, the project is going to go through 4 more, collecting feedback and updating it accordingly

**Week 6 Week 7 Week 8 Week 9 Week 10**



**Iteration 0 Iteration 2 Iteration 5**

**Iteration 1 Iteration 4**

2 Analytic Approach

* Defining the target variable(s):

The target variables for now would be the racers times their positions at the end of the race. This could be used for team performance analysis and further on I want to dive deeper into strategies and environmental factors waste-wise and maybe what the results of the driver’s performances would mean for the team’s finances.

**3 Domain Understanding**

In order to ensure broad domain understanding, a research must be conducted. It will be centered around a stakeholder interview, as well as a number of research questions; e.g:

* What factors influence a driver’s performance?
* What goes into a team’s strategy?
* How do different teams perform on different locations?
* How do teams analyze data to optimize race strategies?
* What role does fuel efficiency play in race strategy?

Research Methods:

* Literature study
* Official F1 Team Report Analysis
* Watching F1 (With strictly educational purposes!)

Analysis of historical data is helpful, but in order to make sure that the research is rooted in real life scenarios and not only in theory, a stakeholder interview is going to be conducted with a person in the field.



4 Data requirements

* Define Objectives:

The goal of this project is to dive deeper into Formula 1: The factors that influence players performances, team performances, strategies, economical and societal factors.

* Data Requirements:

I am going to base my analysis on Temperatures, Amounts of Fuel, Tire Types, Locations, etc.

* Data Sources:

I am going to use the following dataset:

<https://www.kaggle.com/datasets/rohanrao/formula-1-world-championship-1950-2020>

* Data Legality and Ethics: To make sure that I am not overstepping my legal boundaries, I will only be using publicly available data.
* Data Diversity:

To ensure wider data diversity the dataset given may be combined with others, speciating in other areas,

* Version Control: To keep track of the different version of my project, and to prevent harmful losses, I am going to use Git.
* Iterative Process: The project is going go through 4 more versions, collecting feedback and updating it accordingly