

Data Management and Data Analytics Capstone Topic Approval Form

Student Name: [REDACTED]

Student ID: [REDACTED]

Capstone Project Name: Lap Time Progression Analysis: A Measure of Rookie Viability in Formula 1

Project Topic: Analysis of the progression in lap times of rookie Formula 1 drivers to determine their potential and viability for future success

Research Question: How does the progression of lap times for rookie drivers in Formula 1 during their first season correlate with their potential for long term success in the sport?

Hypothesis: Rookie drivers who demonstrate a consistent improvement in their lap times over the course of their first season are more likely to secure long term contracts, and achieve greater success in Formula 1.

Context: Formula 1 is the pinnacle of motorsport, and a testament to human engineering prowess and a relentless pursuit of speed. The difficulty of being a successful Formula 1 team doesn't end at the car; it also extends to the individuals behind the wheel – the drivers. In this high stakes environment, the selection of these drivers becomes a strategic decision, one that can greatly influence a team's future.

Teams heavily invest in scouting and growing talent, with the understanding that a proficient driver can make all the difference. Given the monumental costs associated with training and development, the early identification of a driver's potential becomes that much more important. Adding in the cost-cap structure each team has to be aware of, and the fact that crashes can result in hundreds of thousands or even millions of dollars being used to repair to the car, quickly identifying the potential of the driver is stressed further.

In the modern era, where data is king, a data driven approach to evaluate rookies offers a more objective and reliable method. Analyzing metrics such as lap time progression can provide insights into a rookie's adaptability, consistency, and potential for growth. Analysis such as this can aid teams in making informed decisions, ensuring that their investments in their rookies are more likely to yield returns in the form of race wins, or higher places in the Constructor's standings, which means larger reward payouts to the team at the end of the season.

Data: Detailed performance metrics of all rookie drivers, historical data of successful rookies, and performance data of current drivers.

The dataset will be retrieved from the Kaggle dataset ([here](#)) titled "Formula 1 World Championship (1950 - 2023)". The dataset comprises of 14 csv's:

1. *circuits.csv*
2. *constructor_results.csv*



QMM1: BSDMDA Capstone Topic Approval and Release Form

3. *constructor_standings.csv*
4. *constructors.csv*
5. *driver_standings.csv*
6. *drivers.csv*
7. *lap_times.csv*
8. *pit_stops.csv*
9. *qualifying.csv*
10. *races.csv*
11. *results.csv*
12. *seasons.csv*
13. *sprint_results.csv*
14. *status.csv*

The dataset on Kaggle is publicly available and can be used for research and analysis purposes.

Data Gathering: Data will be downloaded directly from Kaggle. Additional preprocessing might be needed to extract relevant features for rookies, past successful rookies, and veteran teammates, to gauge potential success.

Data Analytics Tools and Techniques: Supervised machine learning models, including decision trees, random forests, and gradient boosting.

Justification of Tools/Techniques: Decision trees and ensemble methods like random forests and gradient boosting are adept at handling complex datasets with multiple features. These models can provide insights into feature importance, helping to highlight which metrics are most indicative of a rookie's potential success.

Application Type, if applicable (select one):

- ☐ Mobile
- ☐ Web
- ☐ Stand-alone

Programming/Development Language(s), if applicable: Python

Operating System(s)/Platform(s), if applicable: Windows, MacOS, & Linux.

Database Management System, if applicable: N/A

Project Outcomes:

1. A machine learning model that predicts the potential success of a rookie based on performance metrics.
2. Insights into the most significant performance metrics influencing rookie success.
3. A benchmarking tool that allows F1 teams to compare current rookies against historical data and their current teammates.

Projected Project End Date: 10/2023

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



WESTERN GOVERNORS UNIVERSITY

Human Subjects or Proprietary Information

Does your project involve the potential use of human subjects? (Y/N): N

Does your project involve the potential use of proprietary company information? (Y/N): N

STUDENT SIGNATURE

[Redacted Student Signature]

By signing and submitting this form, you acknowledge that any cost associated with the development and execution of your data analytics solution will be your (the student) responsibility.

TO BE FILLED BY A COURSE INSTRUCTOR

The capstone topic is approved by a course instructor.

COURSE INSTRUCTOR SIGNATURE:



Jim Ashe, Ph.D. Mathematics

COURSE INSTRUCTOR APPROVAL DATE:

Tuesday, [Redacted] 2023

Project Compliance with IRB (Y/N): Y

