DS 7330 Term Project Proposal

Donald Anderson, Michael Schulte, Adam Vuinovic, & Austin Webb

**Formula 1 Result Optimization: Team Red Bull**

In our project we want to strengthen our ability to design and implement a database that can meet specific requirements and solve real-world problems. We also want to improve our skills of using a database to analyze data, identify trends, and draw conclusions.

Problem of interest: We are creating a database to track records and performance of drivers, teams, and races over time. The goal is to analyze current and historical data of a particular Formula 1 team/driver to determine what correlating factors could determine finishes.

Find what correlating factors could determine finishes

* Ideal conditions (weather or climate)
* Duration or # of pitstops
* Driver nationality or age
* Team nationality
* Change of positions during race
* Track region
* Qualifying finish vs race finish

Methodology:

1. Data collection

* Find data for Formula 1 racing and determine the type of data and size of data that will need to be stored.

1. Data cleaning/EDA

* Perform exploratory data analysis to get a better understanding of the data that will be worked with. Data cleaning should be done to ensure the data is accurate, consistent, and ready for analysis.

1. Data design

* Create a data design/model that represents the structure of the data, the relationships of the data, the relationships between data entities, and the constraints on the data.

1. Database creation

* Use MySQL to create the physical database, including the tables, indexes, constraints, and other components defined in the database design.

1. Data loading

* Load data into the database.

1. Deployment

* Deploy the database to a production environment, i.e. DataBricks or Google Colab.

1. Analyze the data to solve the problem

* Use database/machine learning methods to solve the problem

1. Presentation

* Share findings