Abstract: (Do not cite references in abstract; write this part last so we know how to end it)

According to Statista, half a billion people watch Formula 1 races. With such a large viewership, we can assume that betting on the outcome of the races is a popular event. We wanted to identify insights to determine if there is a way to predict winners of each race. The goal is to analyze current and historical data of a particular Formula 1 team/driver to determine what correlating factors could determine finishes.

We have created a database to track records and performance of drivers, teams, and races over time. We gathered our data from <http://ergast.com/mrd/> and explored its contents on a deep level. We looked into correlating factors (weather conditions, nationality, track region, etc.) as well as used various statistical methods to create prediction models. The results show that winners can be predicted with LOW ACCURACY due to the randomness of unknown variables. It is hard to predict the winner of the races with the data we had, but as advancements in technology and Formula 1 occur, predictions should increase as well.

<https://www.statista.com/statistics/480129/cable-or-broadcast-tv-networks-formula-one-f1-racing-watched-within-the-last-12-months-usa/>

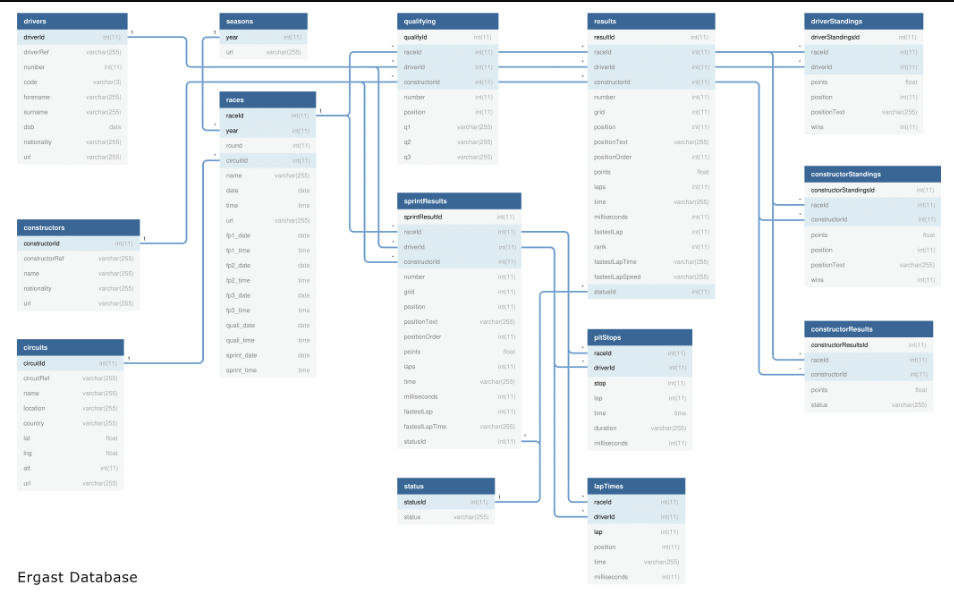
1. Introduction (I do not like this paragraph lol)

In 2021, almost 4.3 billion dollars were generated through the use of sports betting. Due to this absurd amount of money, we wanted to identify a way to give a lower percentage of risk to those who bet on sports – specifically Formula 1 racing. We made a goal of building and implementing a prediction model to identify race finishes. The way that we went about doing this was to first collect current and historical data regarding F1 races. The data we collected included various features that could potentially impact the finishes. With the large amount of data collected, we created a database to store and organize the data. Next, we created the actual prediction model. Finally, we analyzed the results of our model. A successful model will predict the finishes with higher accuracy than current models.

<https://www.legalsportsbetting.com/how-much-money-do-americans-bet-on-sports/#:~:text=How%20Much%20Money%20Is%20Bet%20On%20Sports%20In%20Other%20States,excluding%20tribal%2Drun%20sportsbooks>).

1. Data Collection

In order to create a model to predict the finishes of F1 races, we needed data. We found a great resource to help with the data mining: the Ergast data repository. This repository contained a lot of data and a lot of information. “The Ergast Developer API provides a historical record of motor racing data” from 1950, when the world championship began.

The Ergast database structure can be viewed in the following Entity Relationship Diagram:  


We modeled our database off of this one. By creating a database that easily interconnects with all of our data, we will be able to more easily create our prediction models.