

**Project Report** Database Design - ILS Z 511

**F1 Analytics for Fantasy Sports and Fan Engagement**

SubmittedBy:  
 Aditya Prajapati​

Apoorva Ramesh​

Jannat Arora

Prabhruti Chaudhary

Luddy School of Informatics, Computing & Engineering, Indiana University

**FALL- 2023**

​​**Table of contents**

[Introduction](#_lg00ww2qfs2y) 3

Target Community 4

Dataset Information 4

E-R Diagram 6

Relationship Schema 7

Questions Solved 8

SQL Queries - Fantasy Sports 9

SQL Queries - Fan Engagement 11

Ethical & Privacy Concerns 12

Challenges Faced & Resolutions 12

# 

**Introduction**

The world of Formula One, a premier motorsport championship, has witnessed a significant transformation in recent years due to data-driven fan engagement strategies. Since its inception in 1950, Formula One has evolved into a global phenomenon, captivating the imagination of millions of viewers worldwide. Notably, the release of the Netflix documentary series 'Drive to Survive' has played a pivotal role in expanding the sport's audience. This compelling series offers an insider's view of the intense competition, personal stories, and behind-the-scenes drama of Formula One, thereby attracting a new and diverse group of fans. The global viewership of Formula One has surged to an impressive 445 million, reflecting the power of data-driven content in engaging and retaining fans.

Another transformative trend in the world of sports entertainment is the rising popularity of fantasy games. Formula One, with its unique blend of strategy, skill, and cutting-edge technology, has proven to be an ideal candidate for fantasy leagues. Fans now have the opportunity to immerse themselves in the sport by becoming virtual team managers, selecting their favorite drivers and making strategic decisions based on real-time data and performance metrics. This innovative approach to fan engagement not only enhances the connection between the sport and its fans but also leverages data to create a more interactive and immersive experience for the audience.

Furthermore, the growing demand for real-time statistics, historical performance data, and insightful analytics has paved the way for an enhanced spectator experience in Formula One. With advancements in technology and data analysis, fans now have access to a wealth of information that adds depth to their understanding of the sport. Whether it's tracking a driver's lap times, analyzing race strategies, or exploring historical performance trends, Formula One has embraced data-driven insights to cater to the needs and preferences of its dedicated global fanbase. This evolution in fan engagement reflects a broader trend in the sports and entertainment industry, where data-driven approaches are revolutionizing how fans connect with their favorite teams and athletes.

**Target Community**

The primary community targeted by our Formula 1 analytics and fantasy sports project comprises fervent Formula One enthusiasts and passionate fantasy sports aficionados. Formula One, with its rich history dating back to 1950, has evolved into a global phenomenon, captivating the imaginations of millions of viewers worldwide. By identifying and focusing on this dynamic community, our project aims to provide a tailored and enriching experience that aligns with the evolving expectations of Formula 1 fans and fantasy sports enthusiasts worldwide.

**Dataset information**

The Formula 1 World Championship dataset, readily available on Kaggle, offers an extensive and detailed historical record of the championship's events and statistics spanning from its inaugural season in 1950 to 2023. This dataset serves as a valuable resource for researchers, analysts, and enthusiasts interested in delving into the rich history of Formula 1, allowing them to explore a wide range of data, including race results, driver and team performance metrics, circuit information, and much more.

Find it here : [Formula 1 World Championship (1950 - 2023)](https://www.kaggle.com/datasets/rohanrao/formula-1-world-championship-1950-2020/)

A screenshot of a computer

Description automatically generated

This dataset comprises multiple CSV files, each containing specific information related to the history of Formula 1 racing from 1950 to 2023. Here's an overview of the main CSV files included in the dataset

1. circuits.csv: This file contains details about the various racetracks and circuits used in Formula 1 races. It includes information such as circuit ID, name, location, country, and circuit characteristics.

2. constructor\_results.csv: This CSV provides data on the performance of Formula 1 constructors (teams) in each race. It includes constructor results like constructor ID, race ID, points, and other relevant statistics.

3. constructors.csv: This file offers details about Formula 1 constructors, including their constructor ID, name, nationality, and other pertinent information.

4. driver\_standings.csv: It contains information on the standings of drivers in various Formula 1 seasons. This includes driver standings per race, driver ID, race ID, points, and position.

5. drivers.csv: This CSV file provides information about Formula 1 drivers, including driver ID, driver's full name, nationality, date of birth, and other relevant data.

6. lap\_times.csv: This dataset includes lap time information for each driver in different races, allowing for an analysis of lap-by-lap performance.

7. qualifying.csv: It offers details about driver qualifications for races, including driver ID, race ID, grid positions, and more.

8. races.csv: This CSV file provides essential race information such as race ID, name, date, and location, allowing for race-specific analysis.

9. results.csv: This comprehensive file contains race results, including driver and constructor details, race ID, position, points, and other related data.

**Here's how the dataset is related to our target community:**

1. **Historical Insights**: The dataset offers a historical record of Formula One, helping fans explore the sport's evolution over more than seven decades.

2. **Statistical Analysis**: Researchers and analysts can use the dataset for in-depth statistical analysis, allowing them to assess race results, driver/team performance, and other metrics.

3. **Data Visualization**: Enthusiasts can create visual representations of the data to better understand the sport's history, trends, and relationships between various data points.

4. **Fan Engagement**: The dataset enhances engagement by providing data for discussions, fact-checking, and data-driven debates within the Formula One community.

**Entity Relationship Diagram**

A diagram of a company

Description automatically generated

**Relational Schema**

In the realm of database design, the relational schema serves as a fundamental blueprint, delineating the structure and interrelationships among various data entities. In our project, the relational schema served as a fundamental blueprint helping us understand relationships among various data entities, helping us manage the data better. Derived from the Entity-Relationship (ER) diagram, the relational schema translation was designed with careful consideration of primary and foreign keys to uphold data integrity and seamless interactions within the database.

A diagram of a computer

Description automatically generated

**Questions we aim to Solve**

**Our project is divided into 2 domains, one deals with historical data for Fan Engagement and the other one deals with Recent data for Fantasy sports.**

**Questions for Formula 1 fantasy sport using recent data:**For the fantasy Sports part we will be taking the data from the latest regulation change. In this case, the latest regulation change was in 2022 when the car design was updated for all the teams so we will be using that data. This is also because the performance order of all the teams have been changed after the change so including the older data may end up giving unexpected and wrong results.

**1. Which teams have consistently scored points in most of the race?**

Fantasy game participants can acquire useful insights to make well-informed picks by calculating their average points per race and analyzing recent data to find which teams have shown this consistency. With this knowledge, players can select teams that have a history of producing results, giving them the best opportunity to maximize their point total.

**SQL Query:**

**A computer screen shot of a code

Description automatically generated**

**2. Are there certain circuits where specific drivers tend to excel?**

Given the flexibility to change their team composition during the season, players can leverage this data to maximize their team's performance. By assessing driver performances on upcoming tracks, they can strategically adjust their lineup, ensuring that they have drivers who historically excel on those circuits.

**SQL Query:**

**A computer screen shot of a computer code

Description automatically generated**

**3. Based on historical data and current form, which drivers should users consider for their fantasy teams?**

By analyzing historical data, users can pinpoint drivers who have consistently delivered solid performances across various seasons, indicating their reliability for the long haul. Simultaneously, assessing their current form helps ensure that these drivers are still competitive and in peak condition. This question offers a valuable insight into which drivers are not only proven over time but also maintain their competitive edge, enabling fantasy team managers to make more informed and strategic choices, especially if they prefer stability in their lineups throughout a season.

**SQL Query:**

A screenshot of a computer

Description automatically generated

**Questions for Formula 1 fan engagement using historical data :**

For fan management , we consider historical data starting from 1950 as analysis has to be made on the drivers and teams performance over the years.

**1. Who has the highest number of wins/podiums?**

This analysis not only offers fans the understanding of a driver's legacy and their profound impact on Formula 1 but also provides an avenue to celebrate and acknowledge the extraordinary achievements of these motorsport legends. By tracking the ongoing achievements of these drivers, it perpetuates a sense of connection and reverence for their contributions.

**SQL Query:**

A computer screen shot of a code

Description automatically generated

**2. Which drivers have won the championships and in what years?**

**SQL Query:**

**A computer screen shot of a computer code

Description automatically generated**

**3. Which teams excelled in which era or decade and how many titles do they have?**

This analysis allows fans to grasp the context and significance of team achievements over the years, shedding light on the unique contributions made by each dominant team and the lasting impact they've left on the sport. Formula 1's history is replete with teams that enjoyed distinct eras of excellence, and by exploring and acknowledging these historical trends, fans can better understand the rich tapestry of the sport.

**SQL Query:**

A computer screen shot of a code

Description automatically generated

**Ethics or privacy concerns:**

Since the dataset is available as an open-source resource, there are no specific ethical or privacy concerns that need to be addressed. Open-source datasets are typically made available with the consent of the data owners or with anonymized, publicly accessible data, which means there are no privacy violations or data breaches associated with its usage.

However, the mention of adhering to industry standards of data integrity for open-source data is essential. Even though the dataset is open-source, it's crucial to handle it with care and responsibility. This includes ensuring that the data is used for its intended purpose, respecting any licensing or attribution requirements, and avoiding any misuse or unethical practices, such as re-identifying individuals or disclosing sensitive information. Adhering to industry standards helps maintain the trust and ethical use of open-source data, promoting responsible data analysis and research within the data science community.

**Challenges Faced & Resolutions**Challenge 1: Multiple null values wrongly represented.

Resolution: We changed data with '\N' or '0' values to NULL to get the data loaded seamlessly.

Challenge 2: Few of the tables were out of scope(ex: Lap Times, pit stops).

Resolution: Careful consideration had to done to remove data to avoid disjointness and to maintain integrity of the rest of the data.​