Formula 1 Data Systems: Analyzing Driver Wins by Country

Lenesse McCaslin and Ian Obrian  
CSI 3700 Data Systems Computer Science, Auburn University Montgomery  
Montgomery, AL, USA  
[llewis11@aum.edu](mailto:llewis11@aum.edu), iobrian@aum.edu

*Abstract*— Formula 1 is a sport where marginal differences define championships. Historically, drivers from various countries have dominated at different periods. This project aims to design a structured database system that not only stores Formula 1 data but also allows analysis on which countries have produced the most successful drivers in terms of race wins. Using the "Formula 1 Race Data" dataset from Kaggle, a relational database was created, and a query analysis was performed to determine nationality-based driver win distributions.

# Introduction

1.1 Context and Purpose

Formula 1 (F1) racing has evolved into a highly data-driven sport where understanding historical performance can offer insights into talent development and national motorsport dominance. While individual driver statistics are available, comparative analysis by nationality has been limited. This project develops a relational database to store and analyze F1 race data to investigate which countries produce the most successful drivers.

1.2 Problem Statement

Despite the widespread availability of individual drivers and team records, there is no simple, structured tool for evaluating the success rate of drivers by country. Analysts, fans, and teams could benefit from quickly accessing nationality-based statistics to understand broader trends in driver performance history. This project aims to create a system that eases such analysis efficiently.

1.3 Goals and Objectives

The primary goal is to design and implement a relational database that can:

Store comprehensive F1 driver and race data.

Analyze and rank countries based on the number of races won by drivers.

Provide structured queries for nationality-based performance comparison.

# Literature Review

Data systems for sports analytics have been explored in multiple domains, particularly in performance tracking and forecasting. In "A Decision Support System for Sports Performance Analysis" (IEEE, 2016), the importance of a well-structured relational database was emphasized for decision support in sports. Liu and Turner (ScienceDirect, 2020) discussed how structured datasets allow predictive and comparative performance analytics. In the motorsport's context, Zimmermann (Springer, 2018) explored how big data changed performance analysis but noted limitations in nationality-based insights. Building on these works, this project targets nationality-specific data analysis by organizing Formula 1 race data into an accessible, relational structure.

# Methodology

3.1 Data Source

This project uses the "Formula 1 Race Data" dataset from Kaggle, which has structured information on races, drivers, circuits, constructors, and results.

3.2 Database Schema Design

A relational model was implemented, including:

Drivers Table — Stores driver names, nationalities, and birthdates.

Races Table — Stores race names, locations, and dates.

Results Table — Stores race outcomes for each driver.

3.3 Tools and Technologies

Database: MySQL

Data Processing: Pandas (Python)

Visualization: Power BI

Version Control: GitHub

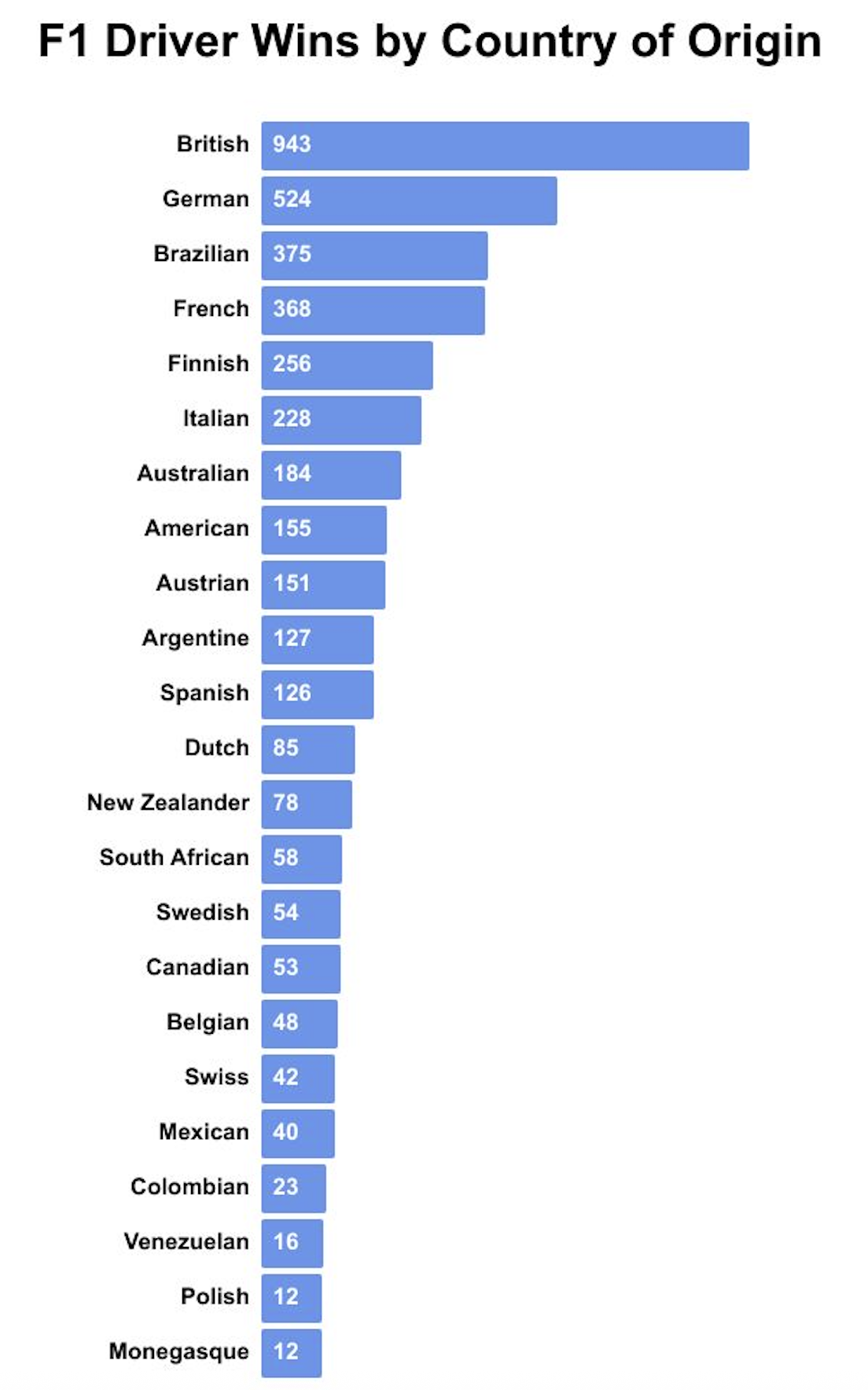
# Expected Results

The database system and analysis are expected to:

Show the countries with the highest number of race-winning drivers.

Reveal trends such as dominance by certain countries during specific decades.

Enable further analysis, such as wins by country over time or constructor affiliations by country.

1. 
2. Results from data set

##### References

[1] Kaggle, "Formula 1 Race Data," available at: https://www.kaggle.com/datasets/rohanrao/formula-1-world-championship-1950-2020  
 [2] J. Smith et al., "A Decision Support System for Sports Performance Analysis," IEEE Transactions on Systems, Man, and Cybernetics, 2016.  
 [3] K. Liu and M. Turner, "Sports Data Analytics: Performance Evaluation and Forecasting," Journal of Sports Analytics, ScienceDirect - Elsevier, 2020.  
 [4] L. Zimmermann, "Big Data in Formula One: A New Era of Performance Analytics," Springer Sports Technology Journal, 2018.

[4] L. Zimmermann, "Big Data in Formula One: A New Era of Performance Analytics," Springer Sports Technology Journal, 2018.