Project Proposal Title: Predicting Successful F1 Undercuts

Group Members:
Lewis Tu (lewistu@college.harvard.edu)
Evan Jiang (evanjiang@college.harvard.edu)
Paul Jeon (pauljeon@college.harvard.edu)

Background and Motivation:
In Formula 1, an "undercut" is a strategic pit stop where a trailing driver pits earlier to gain time on fresher tyres and potentially overtake the rival once they also pit. Pit stop strategy is a critical but often overlooked aspect of F1. Our goal is to use historical race data to estimate the likelihood that an undercut attempt will succeed and to identify the race conditions under which it is most favorable.

Data:
We will use a public F1 dataset from Kaggle containing tables for lap times, pit stops, race metadata, results, drivers, constructors, and standings. This dataset covers F1 Championship seasons from 1950 to 2024. This source will let us reconstruct race order lap-by-lap and identify windows where one car pits before the car immediately ahead. We anticipate needing to derive new features to improve the analysis.

Link to Data: https://www.kaggle.com/datasets/rohanrao/formula-1-world-championship-1950-2020

Scope:
We will first define an undercut attempt consistently, then create features such as pre-pit time gap, recent pace difference, and stint age. We will then perform exploratory data analysis to identify important features. Afterwards, we will explore decision trees, random forests, and gradient boosting. This project will enable us to practice supervised learning techniques, feature engineering, and model selection on real-world data.