

Formula 1 Regression



Braden Taack



Agenda

- Introduction
- Methodology
- Results / Conclusions
- Future Work
- Appendix

Introduction

- Motivation: What does it take to be a winner?
- Objective: Model Formula 1 race results against practice data and race day choices
- Goals: Determine key factors for race day performance



Methodology



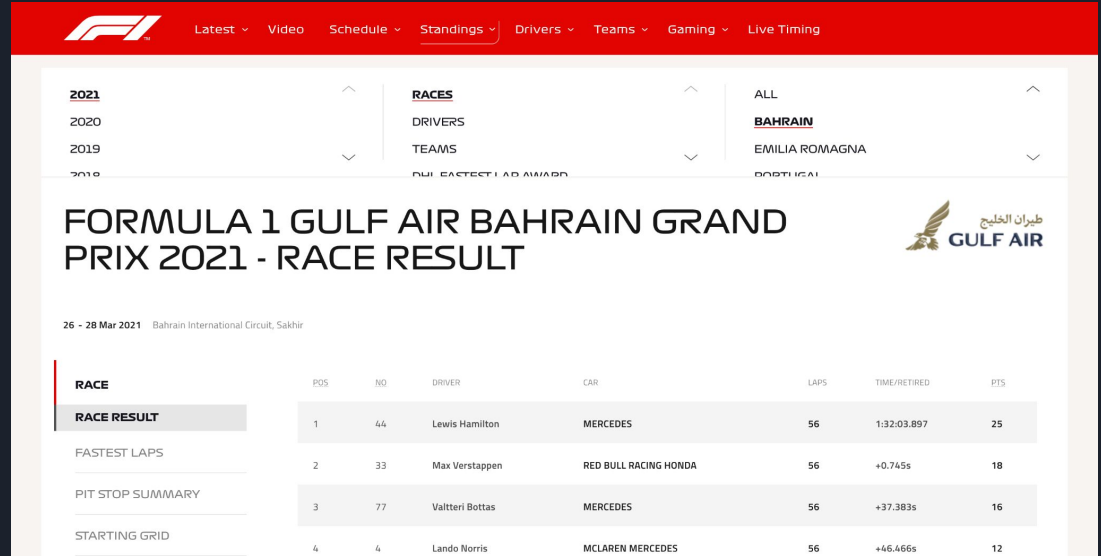


Methodology - Model

- Outcome
 - Predict driver finish position based on pre-race trials and in-race decisions
- Features
 - Fastest_laps_Pos
 - Fastest_laps_Lap
 - SG_Pos_Q_Laps
 - FL_avg_p_laps
 - Fastest_laps_Avg_Speed
 - Pit_stop_summary_Stops
 - Pit_stop_summary_Total
 - Starting_grid_Pos
 - Qualifying_Laps
 - Practice_3_Pos and Laps
 - Practice_2_Pos and Laps
 - Practice_1_Pos and Laps
 - P3_P2_delta
 - P2_P1_delta
 - Q2_Q1_delta

Methodology - Data

- Web-scraped from [Formula1.com](https://www.formula1.com)
- 2006-2021
- ~4200 unique driver finishes
- Initially 26 features



The screenshot shows the Formula 1 website's race results page for the 2021 Gulf Air Bahrain Grand Prix. The page features a red navigation bar at the top with links for Latest, Video, Schedule, Standings, Drivers, Teams, Gaming, and Live Timing. Below the navigation bar, there are three dropdown menus for selecting the year (2021, 2020, 2019, 2018), the race (RACES, DRIVERS, TEAMS), and the location (ALL, SAUDI ARABIA, AUSTRIA, HUNGARY, ITALY, SPAIN, MONACO, AZERBAIJAN, UZBEKISTAN, RUSSIA, BELGIUM, NETHERLANDS, GERMANY, CANADA, UNITED STATES, MEXICO, BRAZIL, ARGENTINA, CHINA, AUSTRALIA, SAUDI ARABIA, AUSTRIA, HUNGARY, ITALY, SPAIN, MONACO, AZERBAIJAN, UZBEKISTAN, RUSSIA, BELGIUM, NETHERLANDS, GERMANY, CANADA, UNITED STATES, MEXICO, BRAZIL, ARGENTINA, CHINA, AUSTRALIA, SAUDI ARABIA). The main heading is "FORMULA 1 GULF AIR BAHRAIN GRAND PRIX 2021 - RACE RESULT". The date and location are "26 - 28 Mar 2021 Bahrain International Circuit, Sakhir". The table below shows the race results with columns for Position (POS), Number (NO), Driver, Car, Laps, Time/Retired, and Points (PTS).

POS	NO	DRIVER	CAR	LAPS	TIME/RETIRED	PTS
1	44	Lewis Hamilton	MERCEDES	56	1:32:03.897	25
2	33	Max Verstappen	RED BULL RACING HONDA	56	+0.745s	18
3	77	Valtteri Bottas	MERCEDES	56	+37.383s	16
4	4	Lando Norris	MCLAREN MERCEDES	56	+46.466s	12

Results



Results

R2: 0.684

MAE: 2.187

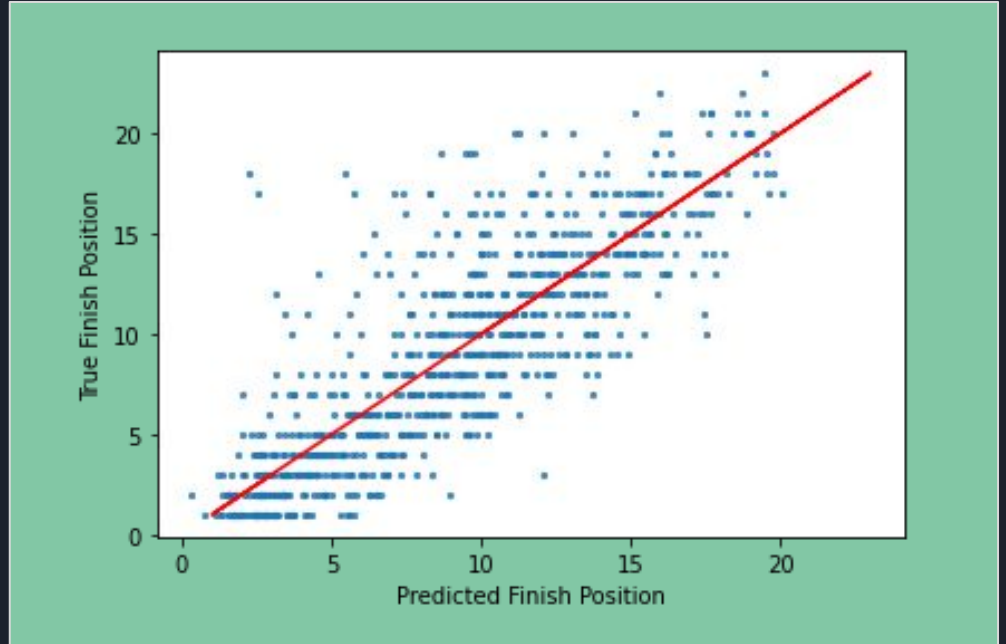
Note: Output is due to discrete nature of target var

Key Features:

- # Pit Stops
- Fastest Lap Pos
- Starting Grid Pos

Honorable Mention :

- # Practice /Qualifying laps





Results

Intercept: -9.764

Coefficients:

- ('Fastest_laps_Pos', 0.308),
- ('Fastest_laps_Lap', 0.003),
- ('SG_Pos_Q_Laps', 0.005),
- ('FL_avg_p_laps', -0.0),
- ('Fastest_laps_Avg_Speed', 0.039),
- ('Pit_stop_summary_Stops', 0.74),
- ('Pit_stop_summary_Total', -0.001),
- ('Starting_grid_Pos', 0.272),
- ('Qualifying_Laps', 0.058),
- ('Practice_3_Pos', 0.077),
- ('Practice_3_Laps', 0.075),
- ('Practice_2_Pos', 0.063),
- ('Practice_2_Laps', 0.083),
- ('Practice_1_Pos', 0.08),
- ('Practice_1_Laps', 0.069),
- ('P3_P2_delta', 0.006),
- ('P2_P1_delta', 0.016),
- ('Q2_Q1_delta', 0.01)

Prediction Examples

- $\hat{Y} = 0.773 \rightarrow y_{\text{true}} = \text{1st}$
- $\hat{Y} = 19.49 \rightarrow y_{\text{true}} = \text{21st}$



Future Work

1. Incorporate more data dating before 2006 when race weekend format was changed
2. Incorporate more features such as money spent, previous success of driver, presence of advanced technology on car, etc
3. Use non-discrete continuous measure of driver success (perhaps fastest average speed)
4. Try more regression methods to deal with discrete data better

Thank You!

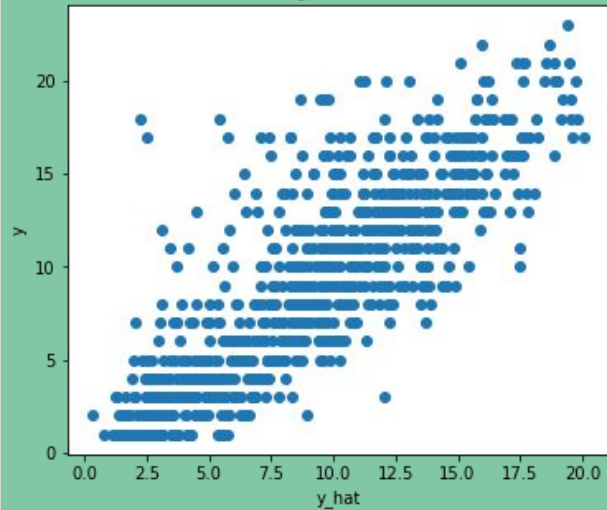


Appendix

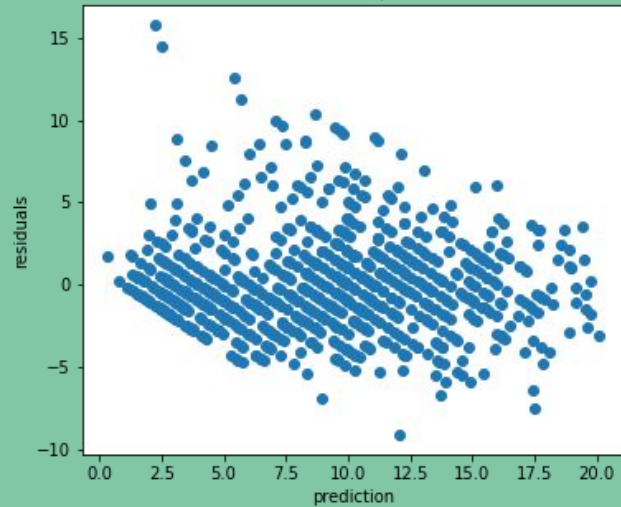


Appendix

Regression fit



Residual plot



Normal Q-Q plot

