Determining the Factors That Most Influence Formula 1 Race Results in the Hybrid Era

Rody Bertolini
CSPB 4502

Questions

1. What race-specific factor has the strongest influence on finishing position?
2. Is the belief that F1 races are determined by the constructors justified?

Motivation

F1 is often thought to be dictated by who has the faster car (which constructor is superior) yet that hasn't been the case this season. While McLaren have the fastest car, Max Verstappen has been able to successfully challenge the McLaren team in his inferior Red Bull. Despite Max's teammates frequently finishing in the bottom 5, Max has won several races and podiumed at most others. The complexities in the chase for this years championship made me want to better understand how various factors affect a race.

Dataset

Formula 1 Championship Dataset

-9 tables used

-Races, Results, Drivers, Constructors, Circuits, Qualifying, Pit Stops, Lap Times, Status

- Over 4,600 entries (driver-race)

-Multiple engineered features (pit stop number, avg lap times)

-Restricted from 2014 to Present

Data Prep Work

Preprocessing:

- Standardized column names across tables
- Converted the place holders to NaN
- Merged relational tables into single dataset
- Engineered multiple features
- Dropped missing/inconsistent rows
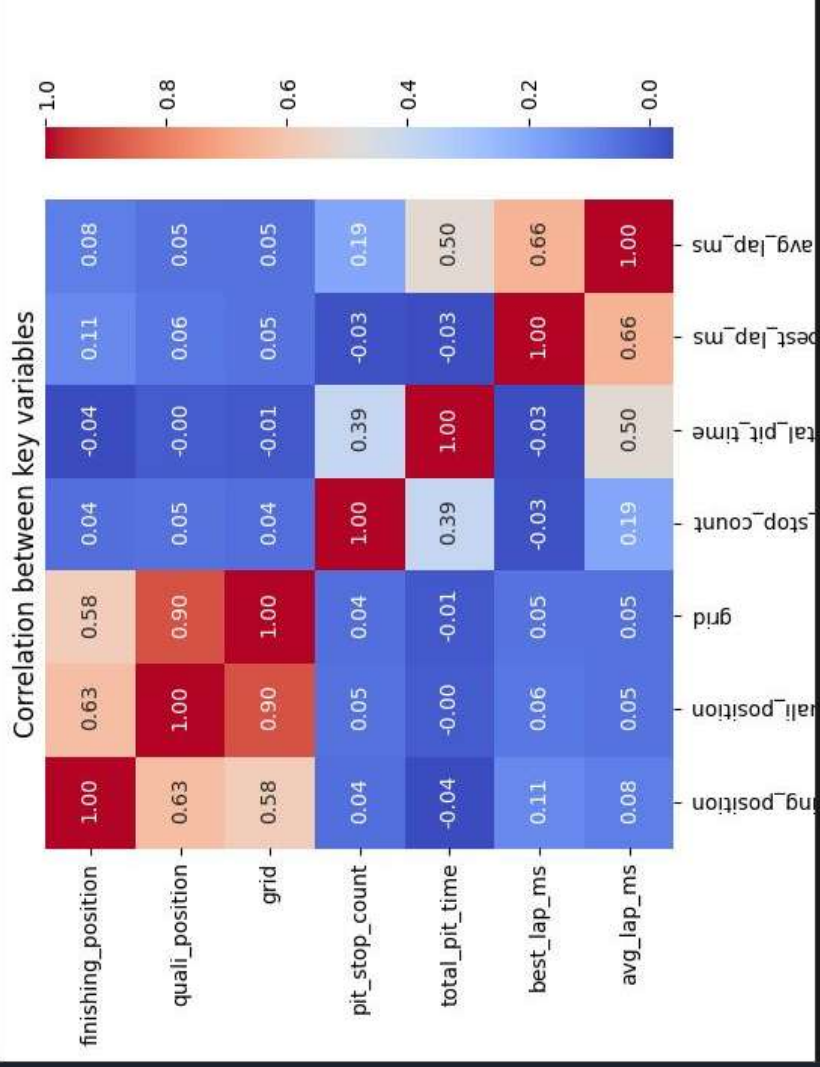- Ensure the numeric typing for modeling

Tools

- Python
- Jupyter Notebooks
- Pandas
- NumPy
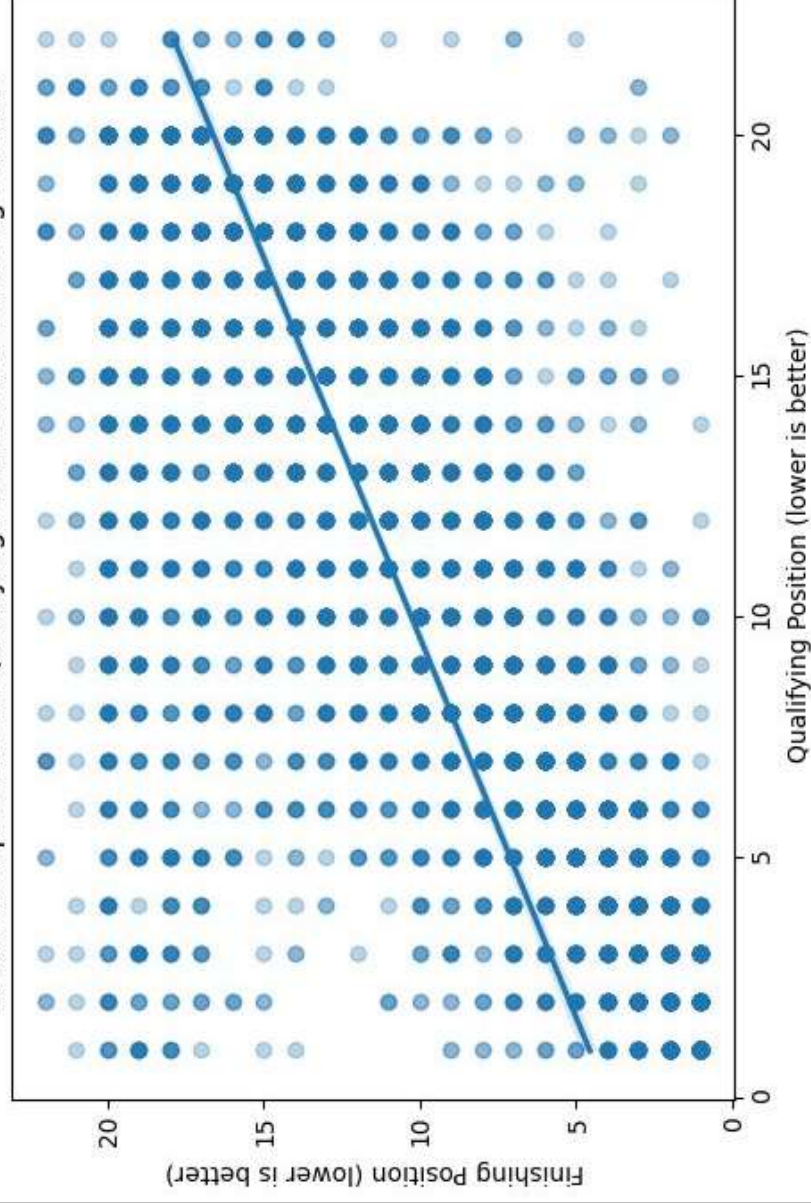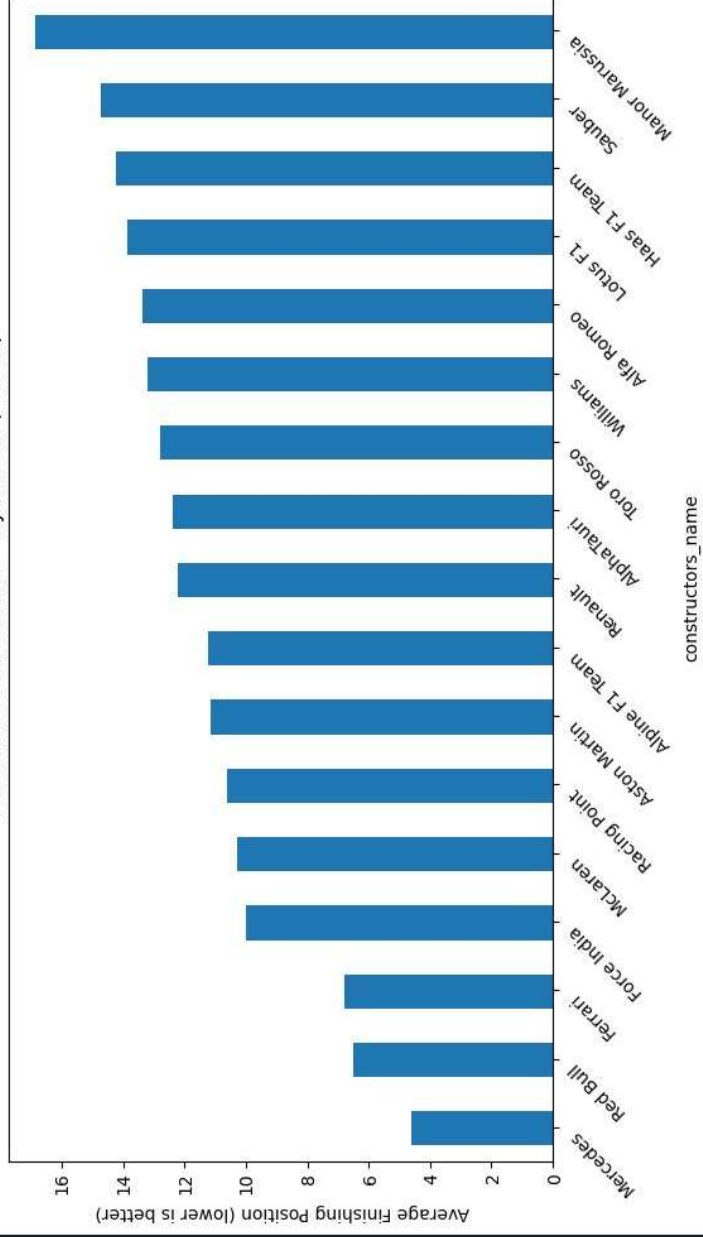- Seaborn + Matplotlib
- scikit-learn

EDA

- Correlation matrices

- Scatterplots with regression lines

- Grouped mean finishing positions

- Distribution visualization

- Sanity check on engineered features

Correlation between key variables

Relationship Between Qualifying Position and Finishing Position

Constructor Performance – Hybrid Era (2014+)

Modeling

Features:

- Quali_position, grid, pit_stop_count, total_pit_time, constructor

R = 0.47 baseline, 0.51 with constructor

- Model explains ~50% variance (due to a high degree of randomness)

-Qualifying coefficient was largest

-Constructor coefficients were smaller offsets

Results

-Qualifying was the most influential.

-Constructors was also meaningful.

-Pit stop and lap times were not effective predictors.

-Formual 1 in the Hybrid era is overwhelmingly determined by track position.