Formula 1

By Michael Mejia

What is Formula 1

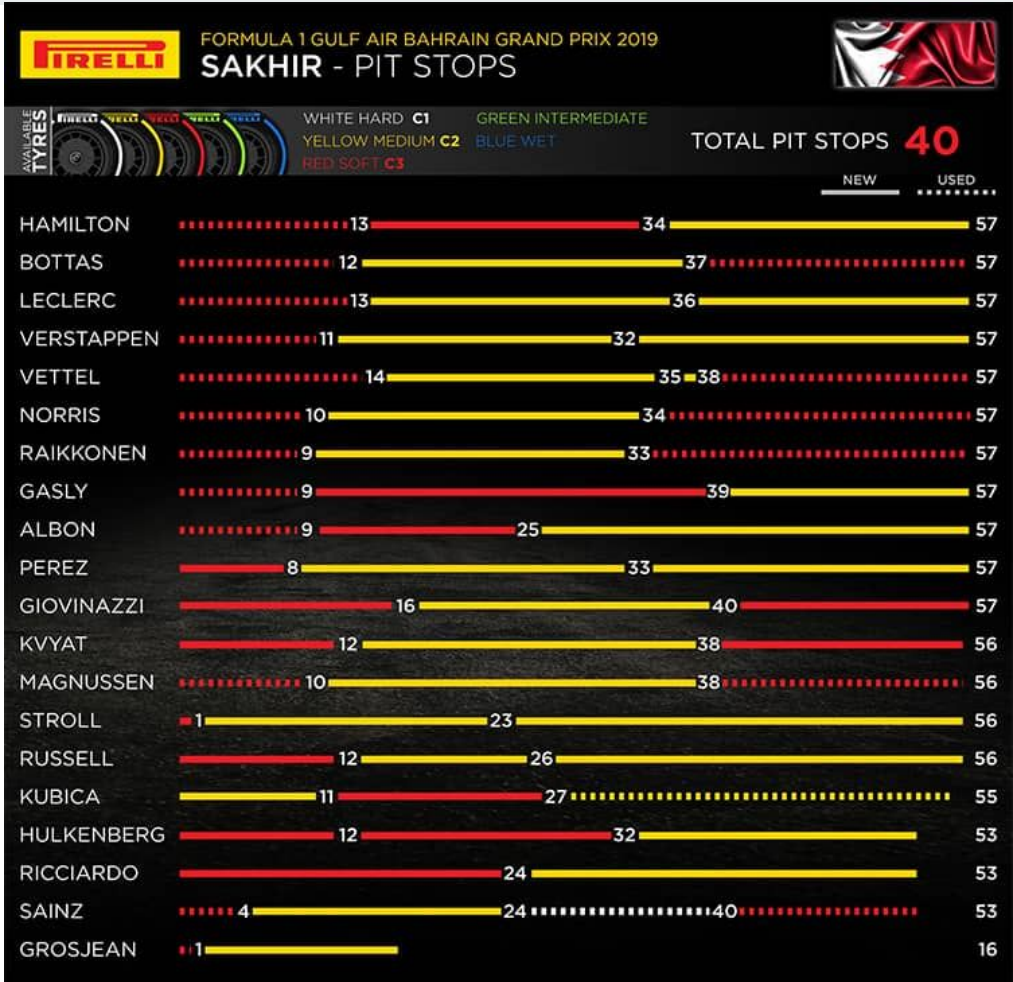
- Racing since 1950
- 13 teams 26 cars
- Combined worth of all teams 4.05 Billion dollars

When to pit and why

Each car has enough fuel for the entire race

Tires wear

Grid placement



Data Gathering

The Ergast Developer API

Kaggle

Racefans.net

JSON Example Response


```
{
  "MRData": {
    "xmlns": "http://ergast.com/mrd/1.0",
    "series": "f1",
    "limit": "30",
    "offset": "0",
    "total": "22",
    "RaceTable": {
      "season": "2008",
      "round": "1",
      "Races": [
        {
          "season": "2008",
          "round": "1",
          "url": "http://en.wikipedia.org/wiki/2008_Australian_Grand_Prix",
          "raceName": "Australian Grand Prix",
          "Circuit": {
            "circuitId": "albert_park",
            "url": "http://en.wikipedia.org/wiki/Melbourne_Grand_Prix_Circuit",
            "circuitName": "Albert Park Grand Prix Circuit",
            "Location": {
              "lat": "-37.8497",
              "long": "144.968",
              "locality": "Melbourne",
```

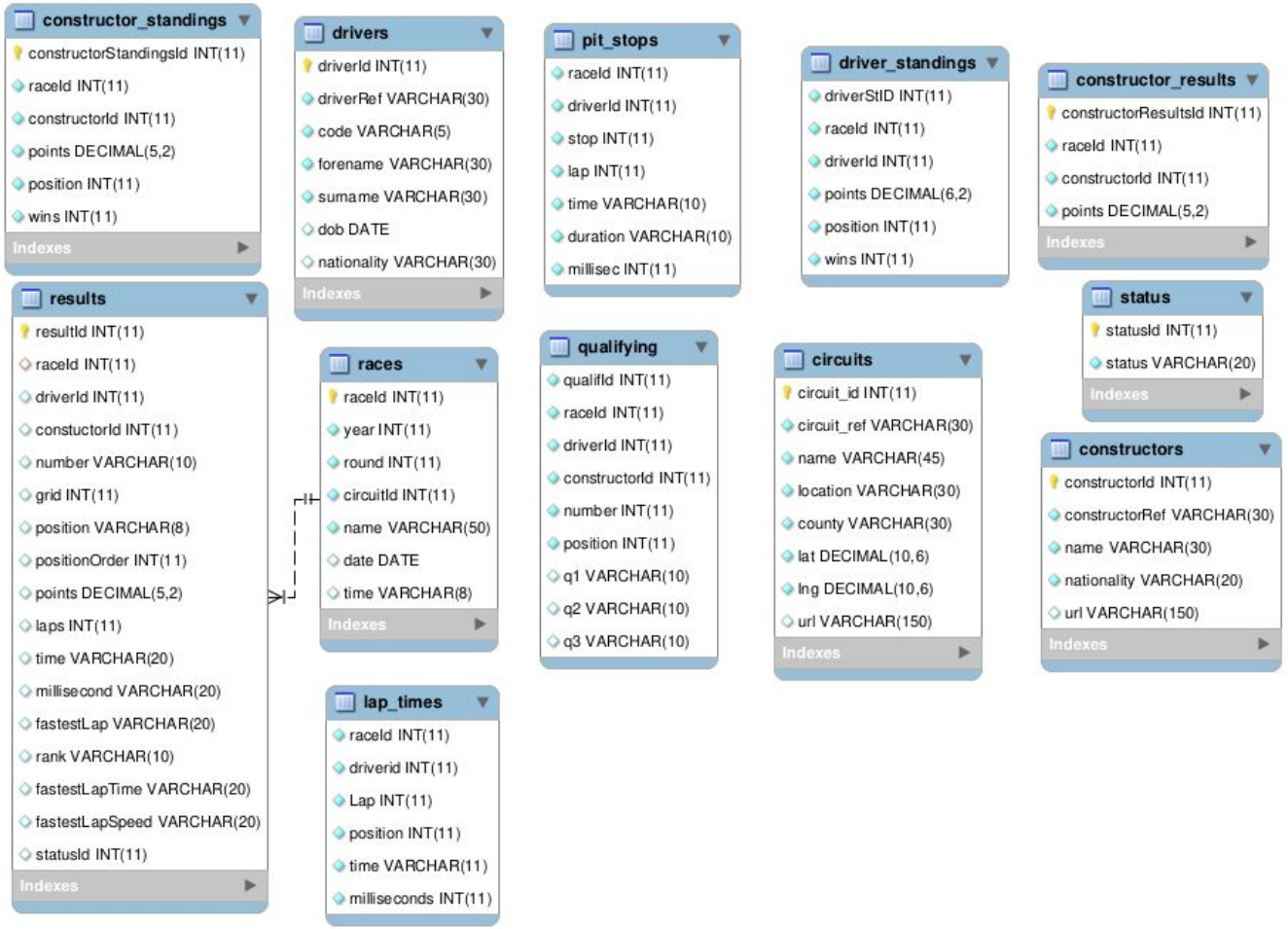
Setup

AWS server

Amazon Relational database server

Local Mysql server

Database


constructor_standings
constructorStandingsId INT(11)
raceId INT(11)
constructorId INT(11)
points DECIMAL(5,2)
position INT(11)
wins INT(11)
Indexes
drivers
driverId INT(11)
driverRef VARCHAR(30)
code VARCHAR(5)
forename VARCHAR(30)
surname VARCHAR(30)
dob DATE
nationality VARCHAR(30)
Indexes
pit_stops
raceId INT(11)
driverId INT(11)
stop INT(11)
lap INT(11)
time VARCHAR(10)
duration VARCHAR(10)
millisec INT(11)
driver_standings
driverStID INT(11)
raceId INT(11)
driverId INT(11)
points DECIMAL(6,2)
position INT(11)
wins INT(11)
constructor_results
constructorResultsId INT(11)
raceId INT(11)
constructorId INT(11)
points DECIMAL(5,2)
Indexes
results
resultId INT(11)
raceId INT(11)
driverId INT(11)
constuctorId INT(11)
number VARCHAR(10)
grid INT(11)
position VARCHAR(8)
positionOrder INT(11)
points DECIMAL(5,2)
laps INT(11)
time VARCHAR(20)
millisecond VARCHAR(20)
fastestLap VARCHAR(20)
rank VARCHAR(10)
fastestLapTime VARCHAR(20)
fastestLapSpeed VARCHAR(20)
statusId INT(11)
Indexes
races
raceId INT(11)
year INT(11)
round INT(11)
circuitId INT(11)
name VARCHAR(50)
date DATE
time VARCHAR(8)
Indexes
qualifying
qualifId INT(11)
raceId INT(11)
driverId INT(11)
constructorId INT(11)
number INT(11)
position INT(11)
q1 VARCHAR(10)
q2 VARCHAR(10)
q3 VARCHAR(10)
circuits
circuit_id INT(11)
circuit_ref VARCHAR(30)
name VARCHAR(45)
location VARCHAR(30)
county VARCHAR(30)
lat DECIMAL(10,6)
lng DECIMAL(10,6)
url VARCHAR(150)
Indexes
status
statusId INT(11)
status VARCHAR(20)
Indexes
constructors
constructorId INT(11)
constructorRef VARCHAR(30)
name VARCHAR(30)
nationality VARCHAR(20)
url VARCHAR(150)
Indexes
lap_times
raceId INT(11)
driverid INT(11)
Lap INT(11)
position INT(11)
time VARCHAR(11)
milliseconds INT(11)

Models

Logistic regression

Random forests

Neural networks

Accuracy too good to be true

98% accuracy for all three models.

- False 0.984263
- True 0.015737

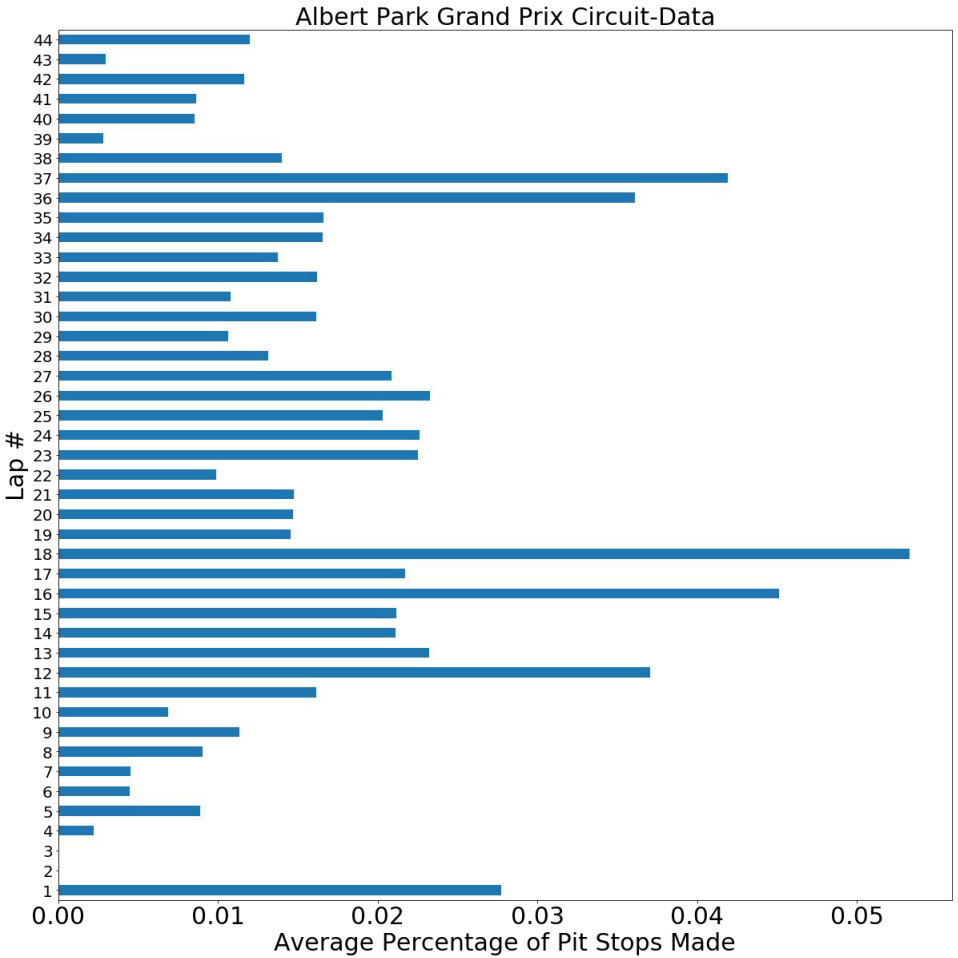
Sensitivity tells a different story

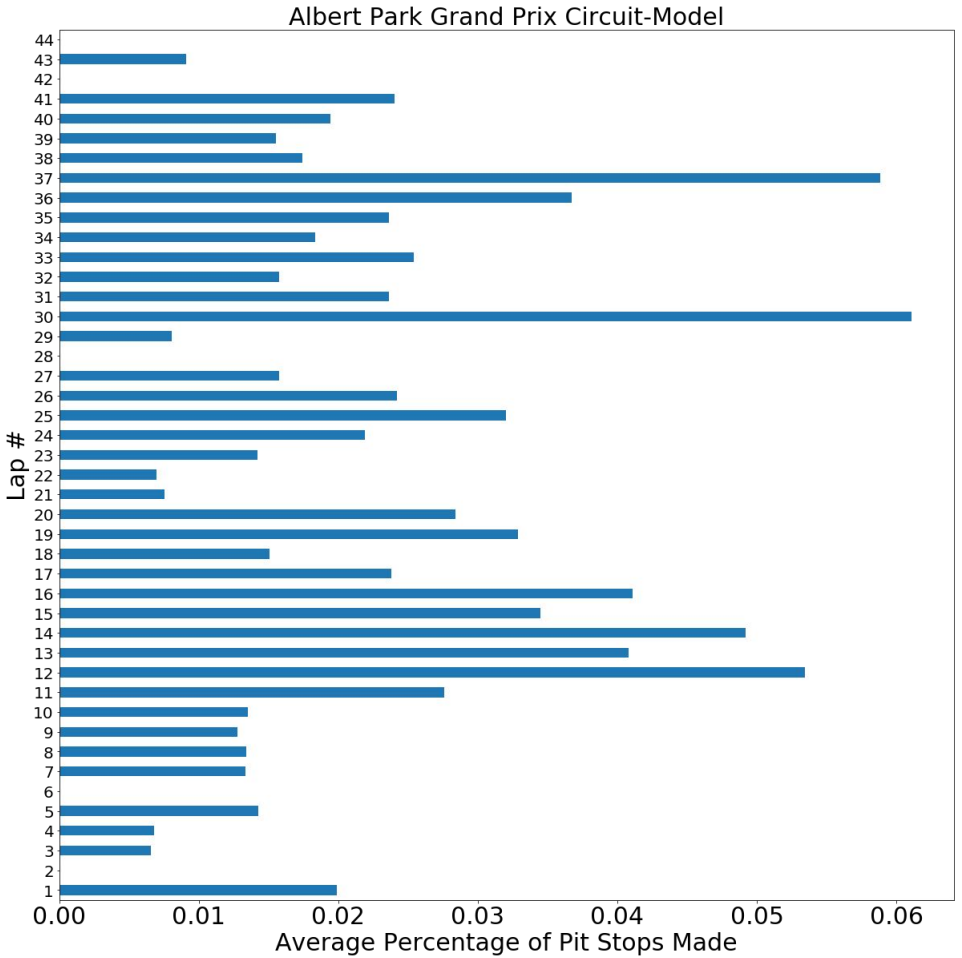
- Logistic regression:
  - Sensitivity: 0.0
    - Specificity: 1.0
- Simple Neural Network:
  - Sensitivity: 0.0
    - Specificity: 1.0
- Random Forest
  - Sensitivity: 0.1084
    - Specificity: 0.9995

Simple Decision trees

Accuracy: Sensitivity: 0.3085

- **Sensitivity: 0.3085**
- **Specificity: 0.9888**


Albert Park Grand Prix Circuit-Data
Lap #
Average Percentage of Pit Stops Made
0.00
0.01
0.02
0.03
0.04
0.05

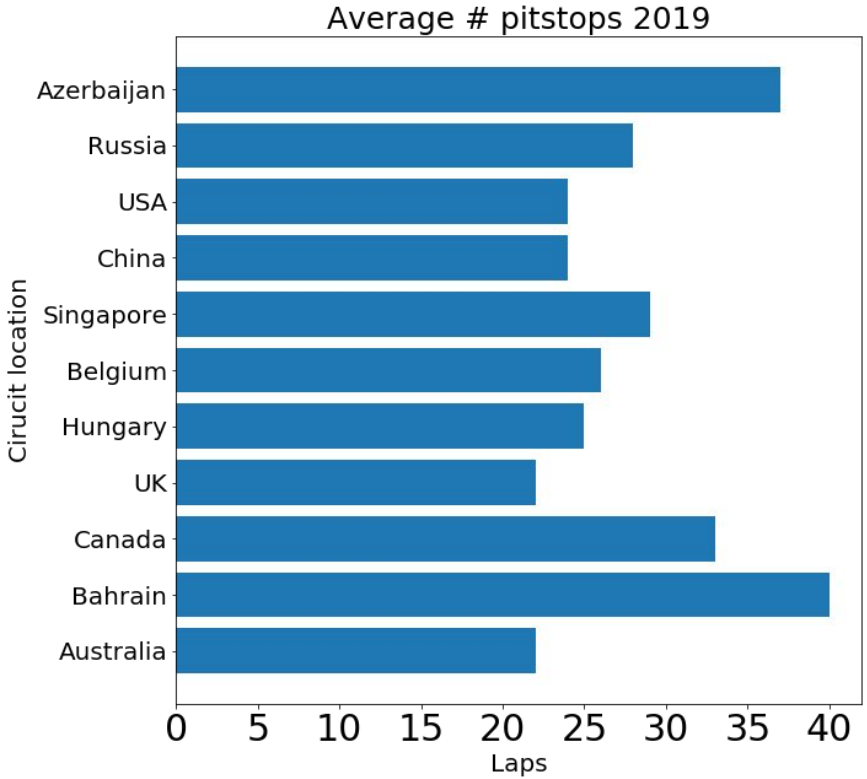

Albert Park Grand Prix Circuit-Model
Lap #
Average Percentage of Pit Stops Made
0.00
0.01
0.02
0.03
0.04
0.05
0.06

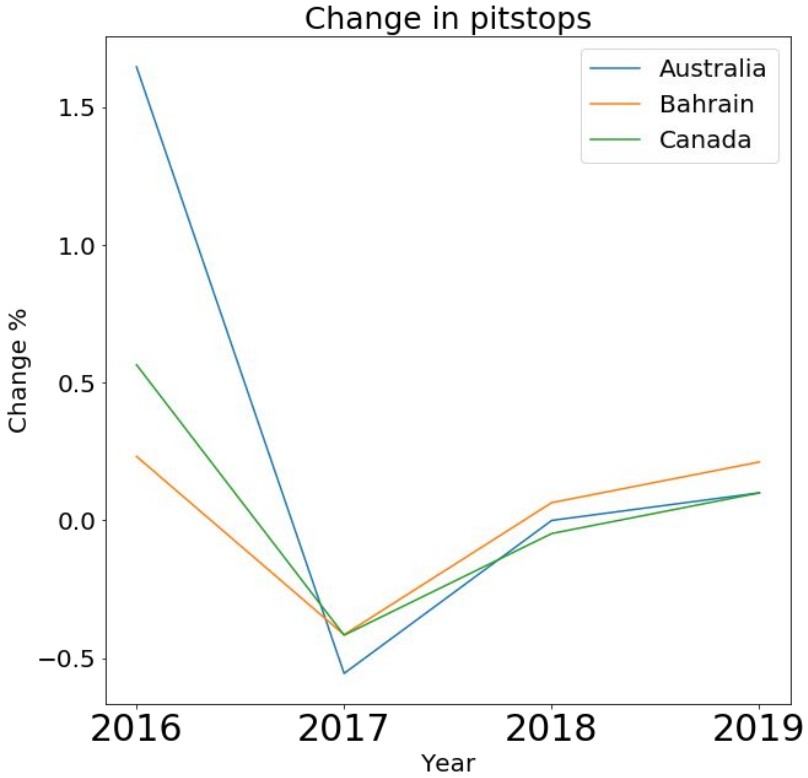
Limitations

A lot of data but still missing a few categories.

Unable to utilize RNN

Yearly rule changes


Average # pitstops 2019
Azerbaijan
Russia
USA
China
Singapore
Belgium
Hungary
UK
Canada
Bahrain
Australia
Cirucit location
0
5
10
15
20
25
30
35
40
Laps


Change in pitstops
Australia
Bahrain
Canada
1.5
1.0
0.5
0.0
−0.5
Change %
2016
2017
2018
2019
Year

Conclusions