Alpine F1 Data Analysis

Chloe Chan
K M Sabah Khan
Alex MacLeod
Gary Mann
Gurminder Mehat

Guinan Consulting

Table Of Contents

Introduction

Key Question

What areas should Alpine F1 Team focus on to improve performance?
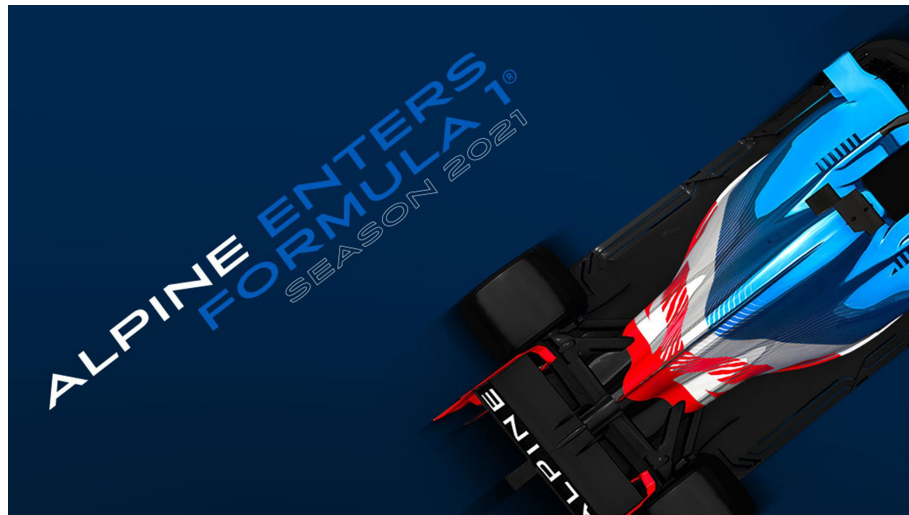
Ultimate Goal

Allow Alpine F1 to go from midfield to top-tier constructor

Analysis Methods

Look at past five years of F1 results, find factors most correlated with podium

Findings Preview

Qualifying is biggest factor, focus on car to achieve better position

ALPINE ENTERS FORMULA 1
SEASON 2021

Data Description

Source

ergast.com

Format

SQL database containing 12 interconnected tables

Data Collection

Data on Formula 1 seasons from 2011

Contents

Race Statistics

Driver Information

Circuit Data

Data Description

Level of Observation

Individual Observations:

- 12 interconnected tables creating data entries ranging from per season to per lap

- Once data was merged, entries were looking at a per lap basis only

Limitations

Basic Data:

- Contains mostly descriptive details such as names, dates, and countries

- Performance data is limited to lap times and positions. No internal team data available such as engine specs

Data Alterations

01

Merge Tables
Imported relevant tables into R, merged into a single table

≈130000 entries due to lap times

Adjust Units and Formats

Many time and speed figures in character format. Durations changed to seconds

02

03

Clean NAs
NA values whenever a driver DNF. Removed NAs to analyze performance when races finished.

Key Variable Definitions: podium

```
==============================================================================================
Statistic              N          Mean       St. Dev.      Min      Pctl(25)   Pctl(75)      Max
----------------------------------------------------------------------------------------------
podium               130,681      0.293       0.455         0          0          1           1
qmean                130,681   72,251.500  24,091.220       0       58,700     91,103.7    117,633
lap_times_seconds    130,681     101.145     83.004      55.404     82.782     107.073    2,126.744
qualifying_position  130,681      7.933       5.562         1          3          12          22
pit_stops_seconds    130,681      99.491     323.518     14.966     22.265     27.136     2,011.266
fastestLapSpeed      130,681     204.768     22.004     149.263    191.013     220.551     255.014
finishing_seconds    130,681   6,155.942   1,217.439   4,526.665  5,489.024  6,369.985  10,943.330
circuitId            130,681      22.892      23.524         1          9          24          76
year                 130,681   2,017.410     1.768       2,015      2,016      2,019       2,020
```

Podium variable created from the finishing position (top 3 podium, rest not)

Units = Podium(1) or no podium(0)

Key Variable Definitions: qmean

```
===============================================================================
Statistic              N        Mean      St. Dev.     Min    Pctl(25)   Pctl(75)    Max
-------------------------------------------------------------------------------
podium               130,681    0.293      0.455        0        0          1          1
qmean                130,681  72,251.500  24,091.220    0      58,700     91,103.7   117,633
lap_times_seconds    130,681   101.145     83.004     55.404   82.782    107.073   2,126.744
qualifying_position  130,681    7.933       5.562        1        3         12         22
pit_stops_seconds    130,681    99.491     323.518    14.966   22.265     27.136   2,011.266
fastestLapSpeed      130,681   204.768      22.004    149.263  191.013    220.551    255.014
finishing_seconds    130,681  6,155.942  1,217.439  4,526.665 5,489.024 6,369.985 10,943.330
circuitId            130,681    22.892      23.524        1        9         24         76
year                 130,681  2,017.410     1.768      2,015    2,016      2,019      2,020
```

Qmean variable created by taking mean of each driver qualifying round times in each race

Units:Milliseconds

Key Variable Definitions: lap_times_seconds

```
==================================================================================
Statistic            N        Mean       St. Dev.    Min     Pctl(25)   Pctl(75)     Max
----------------------------------------------------------------------------------
podium           130,681      0.293        0.455       0          0          1          1
qmean            130,681  72,251.500  24,091.220       0     58,700     91,103.7    117,633
lap_times_seconds 130,681   101.145       83.004    55.404     82.782    107.073   2,126.744
qualifying_position 130,681    7.933        5.562       1          3         12         22
pit_stops_seconds 130,681    99.491      323.518    14.966     22.265     27.136   2,011.266
fastestLapSpeed  130,681    204.768       22.004   149.263    191.013    220.551    255.014
finishing_seconds 130,681 6,155.942   1,217.439 4,526.665  5,489.024  6,369.985 10,943.330
circuitId        130,681     22.892       23.524       1          9         24         76
year             130,681  2,017.410        1.768    2,015      2,016      2,019      2,020
```

Lap times measures the duration of each lap for each driver

Units: Seconds

Key Variable Definitions: qualifying_position

```
================================================================================
Statistic              N        Mean      St. Dev.     Min     Pctl(25)   Pctl(75)      Max
--------------------------------------------------------------------------------
podium             130,681     0.293       0.455         0         0          1           1
qmean              130,681  72,251.500  24,091.220       0       58,700    91,103.7     117,633
lap_times_seconds  130,681   101.145      83.004       55.404    82.782     107.073    2,126.744
qualifying_position 130,681    7.933       5.562         1         3          12          22
pit_stops_seconds  130,681    99.491      323.518      14.966    22.265     27.136     2,011.266
fastestLapSpeed    130,681   204.768      22.004      149.263    191.013    220.551     255.014
finishing_seconds  130,681  6,155.942   1,217.439   4,526.665  5,489.024  6,369.985  10,943.330
circuitId          130,681    22.892      23.524         1         9          24          76
year               130,681  2,017.410     1.768       2,015      2,016      2,019       2,020
```

Qualifying position is position achieved by each driver in qualifying rounds

Units: 1-22 with 1 with assigned to the driver with the first in the rounds

Key Variable Definitions: pit_stops_seconds

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
podium	130,681	0.293	0.455	0	0	1	1
qmean	130,681	72,251.500	24,091.220	0	58,700	91,103.7	117,633
lap_times_seconds	130,681	101.145	83.004	55.404	82.782	107.073	2,126.744
qualifying_position	130,681	7.933	5.562	1	3	12	22
pit_stops_seconds	130,681	99.491	323.518	14.966	22.265	27.136	2,011.266
fastestLapSpeed	130,681	204.768	22.004	149.263	191.013	220.551	255.014
finishing_seconds	130,681	6,155.942	1,217.439	4,526.665	5,489.024	6,369.985	10,943.330
circuitId	130,681	22.892	23.524	1	9	24	76
year	130,681	2,017.410	1.768	2,015	2,016	2,019	2,020

Pit stop seconds is the duration of each pit stop in seconds

Units: Seconds

Key Variable Definitions: fastestLapSpeed

```
==============================================================================
Statistic          N        Mean      St. Dev.     Min    Pctl(25)   Pctl(75)     Max
------------------------------------------------------------------------------
podium            130,681    0.293      0.455        0        0          1           1
qmean             130,681  72,251.500 24,091.220    0      58,700     91,103.7    117,633
lap_times_seconds 130,681   101.145    83.004      55.404   82.782     107.073    2,126.744
qualifying_position 130,681  7.933      5.562        1        3          12          22
pit_stops_seconds 130,681   99.491    323.518     14.966    22.265     27.136    2,011.266
fastestLapSpeed   130,681   204.768    22.004     149.263   191.013    220.551    255.014
finishing_seconds 130,681 6,155.942  1,217.439  4,526.665 5,489.024  6,369.985 10,943.330
circuitId         130,681   22.892     23.524       1        9          24          76
year              130,681 2,017.410    1.768      2,015    2,016      2,019       2,020
```

Fastest lap speed of each driver in each race

Units: Kilometres/Hour

Key Variable Definitions: finishing_seconds

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
podium	130,681	0.293	0.455	0	0	1	1
qmean	130,681	72,251.500	24,091.220	0	58,700	91,103.7	117,633
lap_times_seconds	130,681	101.145	83.004	55.404	82.782	107.073	2,126.744
qualifying_position	130,681	7.933	5.562	1	3	12	22
pit_stops_seconds	130,681	99.491	323.518	14.966	22.265	27.136	2,011.266
fastestLapSpeed	130,681	204.768	22.004	149.263	191.013	220.551	255.014
finishing_seconds	130,681	6,155.942	1,217.439	4,526.665	5,489.024	6,369.985	10,943.330
circuitId	130,681	22.892	23.524	1	9	24	76
year	130,681	2,017.410	1.768	2,015	2,016	2,019	2,020

Finishing seconds is total time the driver took to complete the race

Units: Seconds

Key Variable Definitions: circuitId

```
==================================================================================
Statistic            N        Mean      St. Dev.      Min     Pctl(25)   Pctl(75)     Max
----------------------------------------------------------------------------------
podium             130,681    0.293       0.455         0         0          1          1
qmean              130,681  72,251.500  24,091.220      0      58,700     91,103.7   117,633
lap_times_seconds  130,681   101.145     83.004       55.404   82.782     107.073   2,126.744
qualifying_position 130,681    7.933       5.562         1         3         12         22
pit_stops_seconds  130,681    99.491     323.518      14.966   22.265     27.136    2,011.266
fastestLapSpeed    130,681   204.768      22.004     149.263   191.013    220.551    255.014
finishing_seconds  130,681  6,155.942  1,217.439   4,526.665 5,489.024  6,369.985 10,943.330
circuitId          130,681    22.892      23.524         1         9         24         76
year               130,681  2,017.410    1.768       2,015     2,016      2,019      2,020
```

CircuitId is an unique identifier for each track

Units: Integer

Key Variable Definitions: year

```
============================================================================================
Statistic          N        Mean       St. Dev.      Min     Pctl(25)   Pctl(75)      Max
--------------------------------------------------------------------------------------------
podium           130,681    0.293        0.455         0         0          1           1
qmean            130,681  72,251.500  24,091.220       0       58,700    91,103.7    117,633
lap_times_seconds 130,681   101.145      83.004      55.404    82.782    107.073    2,126.744
qualifying_position 130,681   7.933       5.562         1         3         12          22
pit_stops_seconds 130,681    99.491     323.518      14.966    22.265     27.136    2,011.266
fastestLapSpeed  130,681    204.768      22.004     149.263   191.013    220.551     255.014
finishing_seconds 130,681  6,155.942  1,217.439   4,526.665 5,489.024  6,369.985  10,943.330
circuitId        130,681    22.892       23.524         1         9         24          76
year             130,681  2,017.410      1.768       2,015     2,016      2,019       2,020
============================================================================================
```

Year is the year the race took place in
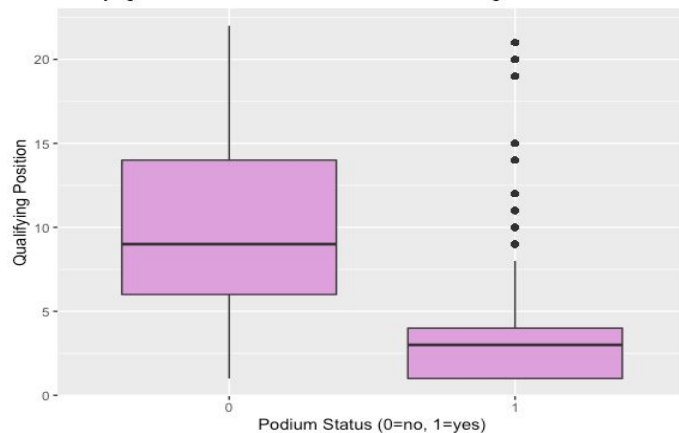
Units: Years from 2015-2020

Key Summary Stats

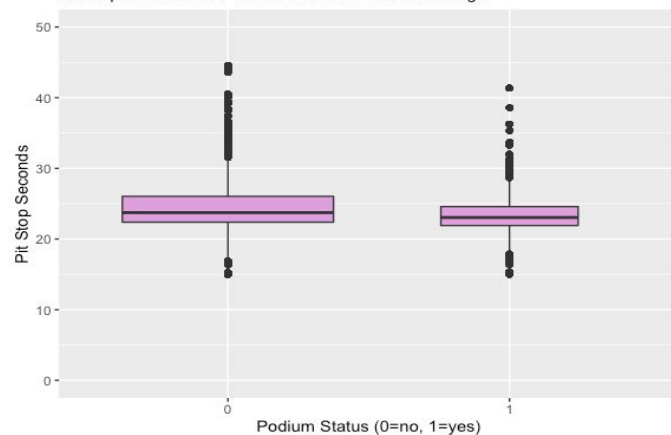Correlogram of Variables Considered for Model
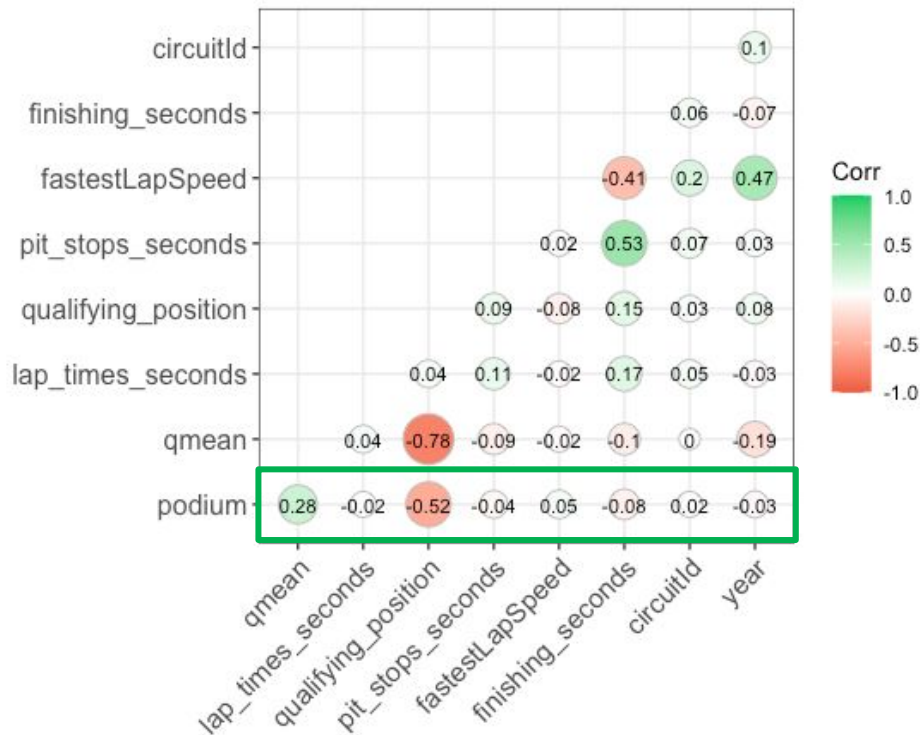
Box Plot
Qualifying Positions for Podium and Non-Podium Placings

Box Plot
Pit Stop Seconds for Podium and Non-Podium Placings

Analysis Methods

Objective

- Identify areas Alpine should focus to improve performance, become top 3 constructor

Model

- LOGIT with podium as binary dependent variable
- IVs: qualifying_position, fastestLapSpeed, pit_stop_seconds, lap_time_seconds
- Controls: year, circuitId

Methods

- Several LOGIT models with podium as DV, different IVs from key variables
- Compared with Alpine data specifically to account for differences in constructors
- Compared to models with points as DV to identify degrees of impact

Analysis Results

LOGIT MODELS

- Podium as DV
- Adding and swapping IV
- Model 3 highest R2, lowest AIC
- McFadden's = 0.318
- AIC = 107848
- Qualifying position stable across models

```
=========================================================================
                                    Dependent variable:
                         ------------------------------------------------
                                           podium
                          (1)        (2)        (3)        (4)        (5)

qualifying_position   -0.413***  -0.413***  -0.413***  -0.413***  -0.413***
                       (0.003)    (0.003)    (0.003)    (0.003)    (0.003)

fastestLapSpeed                   -0.0002    -0.0002    -0.0001    -0.0002
                                  (0.0004)   (0.0004)   (0.0005)   (0.0004)

finishing_seconds                             0.00000
                                             (0.00001)

pit_stops_seconds                           -0.00000   -0.00001   -0.00000
                                            (0.00003)  (0.00003)  (0.00003)

lap_times_seconds                           -0.0001
                                            (0.0001)

year                   0.006      0.007      0.007      0.006      0.007
                      (0.004)    (0.005)    (0.005)    (0.005)    (0.005)

circuitId             0.005***   0.005***   0.005***   0.005***   0.005***
                      (0.0003)   (0.0003)   (0.0003)   (0.0003)   (0.0003)

Constant             -10.151    -12.400    -12.285    -11.589    -12.449
                      (8.812)    (9.591)    (9.600)    (9.758)    (9.598)

-------------------------------------------------------------------------
Observations          130,681    130,681    130,681    130,681    130,681
Log Likelihood      -53,917.640 -53,917.460 -53,917.030 -53,917.330 -53,917.450
Akaike Inf. Crit.   107,843.300 107,844.900 107,848.100 107,848.700 107,846.900
-------------------------------------------------------------------------
Note:                                        *p<0.1; **p<0.05; ***p<0.01
```

Comparison with Alpine Data Only

	Dependent variable:	
	podium	
	(1)	(2)
qualifying_position	-0.413***	-0.143***
	(0.003)	(0.034)
fastestLapSpeed	-0.0002	-0.029***
	(0.0004)	(0.006)
pit_stops_seconds	-0.00000	-0.005*
	(0.00003)	(0.003)
lap_times_seconds	-0.0001	-0.041***
	(0.0001)	(0.005)
year	0.007	18.396
	(0.005)	(317.450)
circuitId	0.005***	-0.015***
	(0.0003)	(0.003)
Constant	-12.285	-37,149.880
	(9.600)	(641,249.700)
Observations	130,681	6,897
Log Likelihood	-53,917.030	-654.364
Akaike Inf. Crit.	107,848.100	1,322.728
Note:	*p<0.1; **p<0.05; ***p<0.01	

Box Plot
Qualifying Positions for Podium and Non-Podium Placings (Alpine)
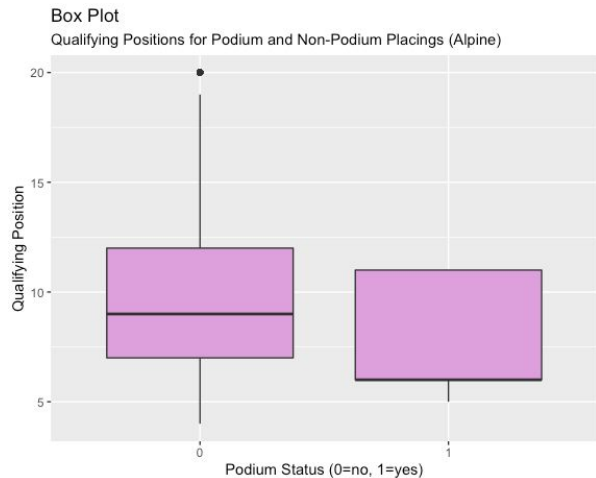
INTERPRETATION

Concern that top teams win qualifying, also win races -> winning qualifying won't help midfield

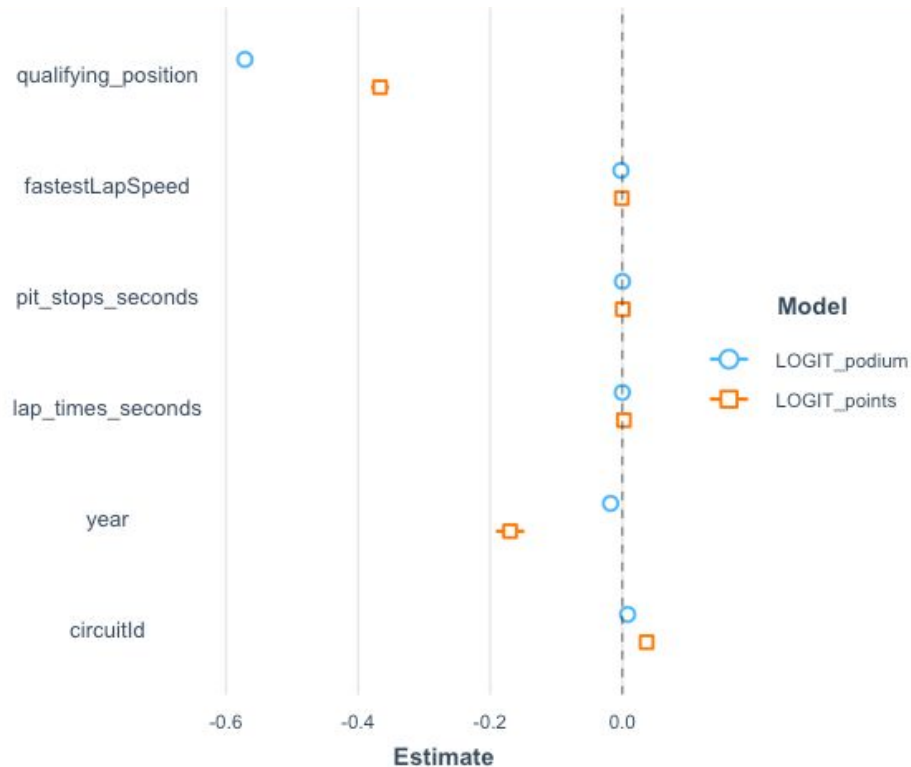Alpine data only shows similar coefficients, distribution of qualifying

Conclusions reached based on all teams applicable to Alpine, but other factors worth considering

Confirming Validity of Relationships

Coefficients of podium model and points model

Confirms qualifying more important to podium

More of an effect than on points, reinforces findings

Interpreting Coefficients

```
================================================
                    Dependent variable:
                ----------------------------
                           podium
------------------------------------------------
qualifying_position        -0.413***
                           (0.003)

fastestLapSpeed            -0.0002
                           (0.0004)

pit_stops_seconds          -0.00000
                           (0.00003)

lap_times_seconds          -0.0001
                           (0.0001)

year                       0.007
                           (0.005)

circuitId                  0.005***
                           (0.0003)

Constant                   -12.285
                           (9.600)

------------------------------------------------
Observations               130,681
Log Likelihood             -53,917.030
Akaike Inf. Crit.          107,848.100
================================================
Note:               *p<0.1; **p<0.05; ***p<0.01
```

INTERPRETATION

For each 1 position better driver qualifies, log odds of podium increase by 0.413

For each 1km/h fastest lap speed increases, log odds of a podium decrease by .0002

For each 1 additional second spent of lap time, log odds of a podium decrease by 0.0001

CONTEXT

Large effect for a position higher, improving qualifying will result in more podiums

Top speed matters, but small coefficient and illogical direction so not basis for recommendations

Useful for modelling based on domain logic, but too small for major recommendations

Conclusion

- Key points
  - Podiums are largely won in the qualifying rounds
  - Race day effects like pit stop efficiency have less of an effect
- Domain Considerations
  - Qualifying purely tests car and driver more than race day
  - Less factors (pits, other cars) play into success

Recommendations

1. Focus budget on car development first and foremost
2. Make slight cuts to pit crew budget to fund car development
3. Conduct further analysis using internal data to determine what car factors determine qualifying

Next Steps

1. Multinomial Regression using qualifying position as DV, include internal car measures if possible
2. Factor DNF status data into analyses to account for car quality issues
3. A/B Tests using two drivers during pre-season testing, replicate qualifying conditions as closely as possible

Thank you.
Any Questions?
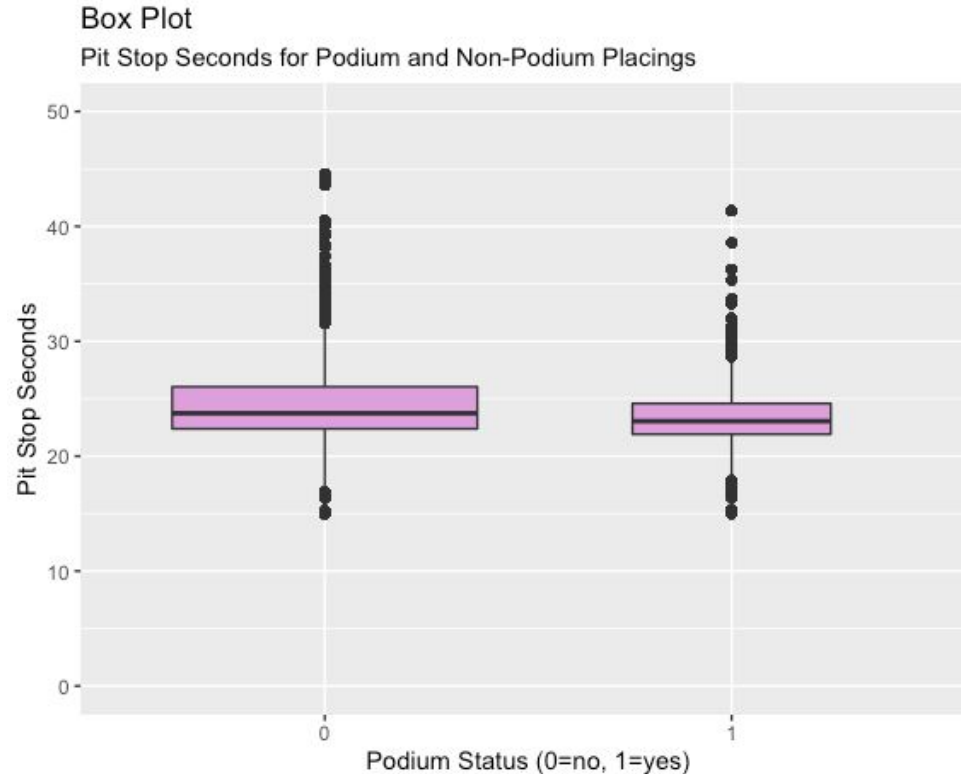
Appendix

A. Box Plots of Pit Stop Seconds

B. Box Pots of Fastest Lap Speed
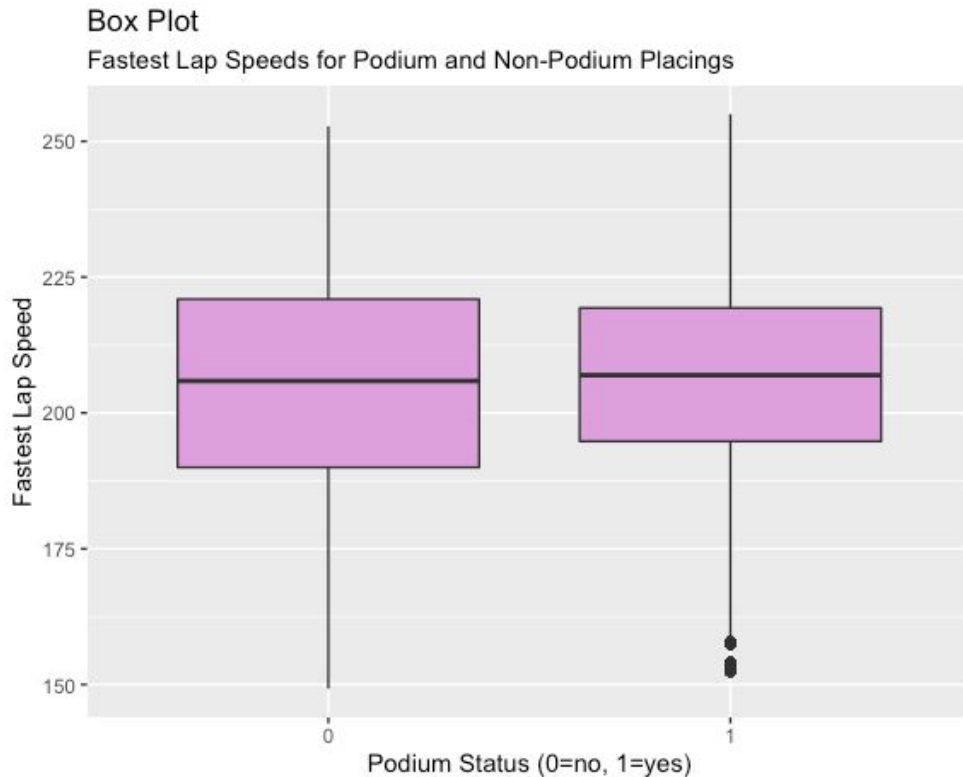
C. Box Plots of Lap Times Seconds

D. Scaled Box Plots of Lap Times
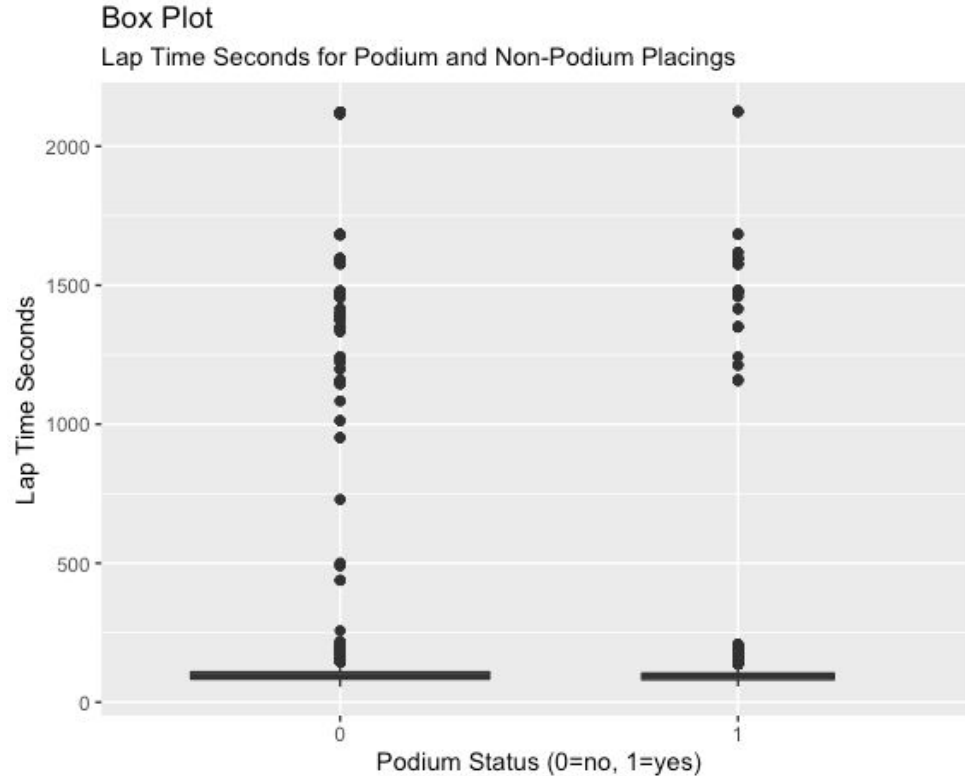
E. Correlation Matrix

Appendix A: Box Plots of Pit Stop Seconds

Box Plot

Pit Stop Seconds for Podium and Non-Podium Placings

Pit Stop Seconds vs. Podium Status (0=no, 1=yes)

Appendix B: Box Plots of Fastest Lap Speed

Box Plot

Fastest Lap Speeds for Podium and Non-Podium Placings

Appendix C: Box Plots of Lap Times Seconds

Box Plot
Lap Time Seconds for Podium and Non-Podium Placings

Appendix D: Scaled Box Plots of Lap Times

Box Plot
Lap Time Seconds for Podium and Non-Podium Placings (Scaled)

Appendix E: Correlation Matrix