

Formula 1: Fastest Lap Prediction

...

Ann Marie Abraham



Formula 1: Background

- Saturday: 3 qualifying sessions
 - Drivers with fastest laps progress to next session
 - 5 drivers “eliminated” after each qualifying session
 - Determines grid position
- Sunday: Race day
 - If you have the fastest lap then you and your team are awarded 1 point



Research question:

Can a driver's fastest lap be predicted from qualifying times, weather variables, race track, and team?

Motive:

People bet on who will get the fastest lap and influences who wins the F1 Championship and Constructors Cup

Data Collection

Data Collection

- Kaggle datasets - data comes from egrast API
- Fast F1 API - qualifying and race weather data
- 2021 Season

qualifyId	raceId	driverId	constructorId	number_car_number	position	q1	q2	...	track_temp_qualifying	wind_direction_qualifying
8740.0	1052.0	817.0	1.0	3.0	6.0	1:30.795	1:30.222	...	33.84359	174.666667
8736.0	1052.0	1.0	131.0	44.0	2.0	1:30.617	1:30.085	...	33.84359	174.666667
8737.0	1052.0	822.0	131.0	77.0	3.0	1:31.200	1:30.186	...	33.84359	174.666667
8743.0	1052.0	4.0	214.0	14.0	9.0	1:30.863	1:30.595	...	33.84359	174.666667
8742.0	1052.0	832.0	6.0	55.0	8.0	1:31.653	1:30.009	...	33.84359	174.666667
...
9159.0	1073.0	815.0	9.0	11.0	4.0	1:23.350	1:23.135	...	29.17284	297.024691
9156.0	1073.0	830.0	9.0	33.0	1.0	1:23.322	1:22.800	...	29.17284	297.024691
9163.0	1073.0	852.0	213.0	22.0	8.0	1:23.428	1:23.404	...	29.17284	297.024691
9157.0	1073.0	1.0	131.0	44.0	2.0	1:22.845	1:23.145	...	29.17284	297.024691
9164.0	1073.0	839.0	214.0	31.0	9.0	1:23.764	1:23.420	...	29.17284	297.024691

Data Cleaning

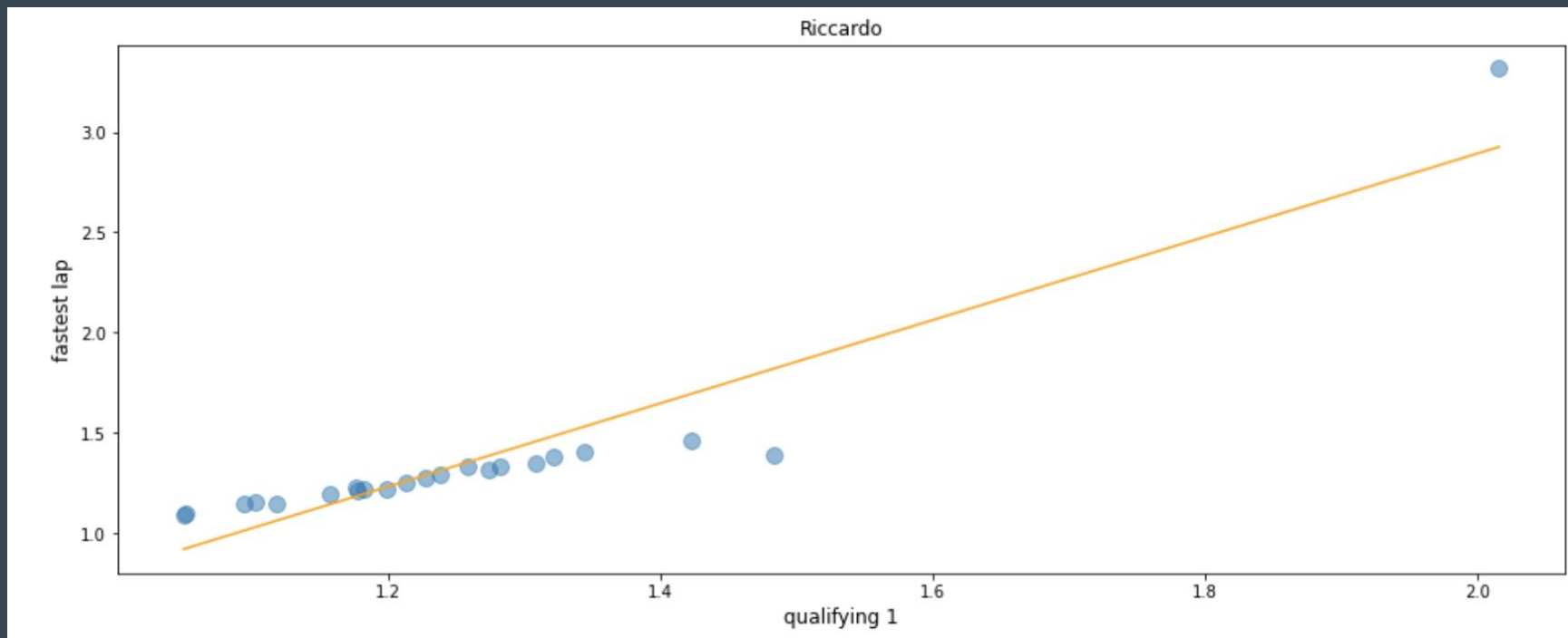
- Finding each driver's fastest qualifying and race laps
- Lap times - strings to floats
 - 1:34.468 -> 1.34468
- Driver reference
 - Lewis Hamilton, HAM, or 44
- Drivers don't race: NaN

fastest_lap	stop	fastest_lap_float	q1_float	q2_float	q3_float
1:34.932	2.0	1.34932	1.30795	1.30222	1.29927
1:34.015	2.0	1.34015	1.30617	1.30085	1.29385
1:32.090	3.0	1.32090	1.31200	1.30186	1.29586
1:36.063	2.0	1.36063	1.30863	1.30595	1.30249
1:34.509	2.0	1.34509	1.31653	1.30009	1.30215
...
1:26.419	3.0	1.26419	1.23350	1.23135	1.22947
1:26.103	3.0	1.26103	1.23322	1.22800	1.22109
1:27.496	2.0	1.27496	1.23428	1.23404	1.23220
1:26.615	1.0	1.26615	1.22845	1.23145	1.22480
1:28.249	1.0	1.28249	1.23764	1.23420	1.23389

Data Exploration

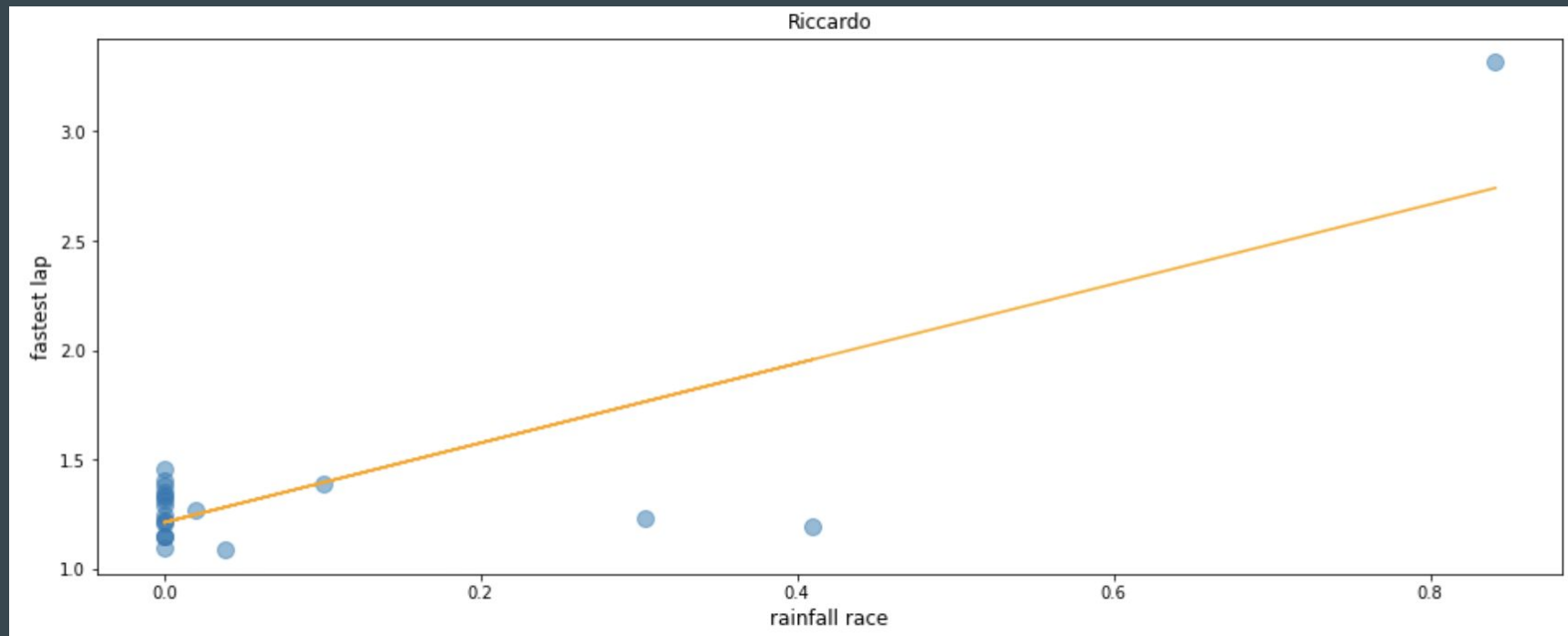
Qualifying 1 vs Fastest Lap

$$R^2 = 0.872$$



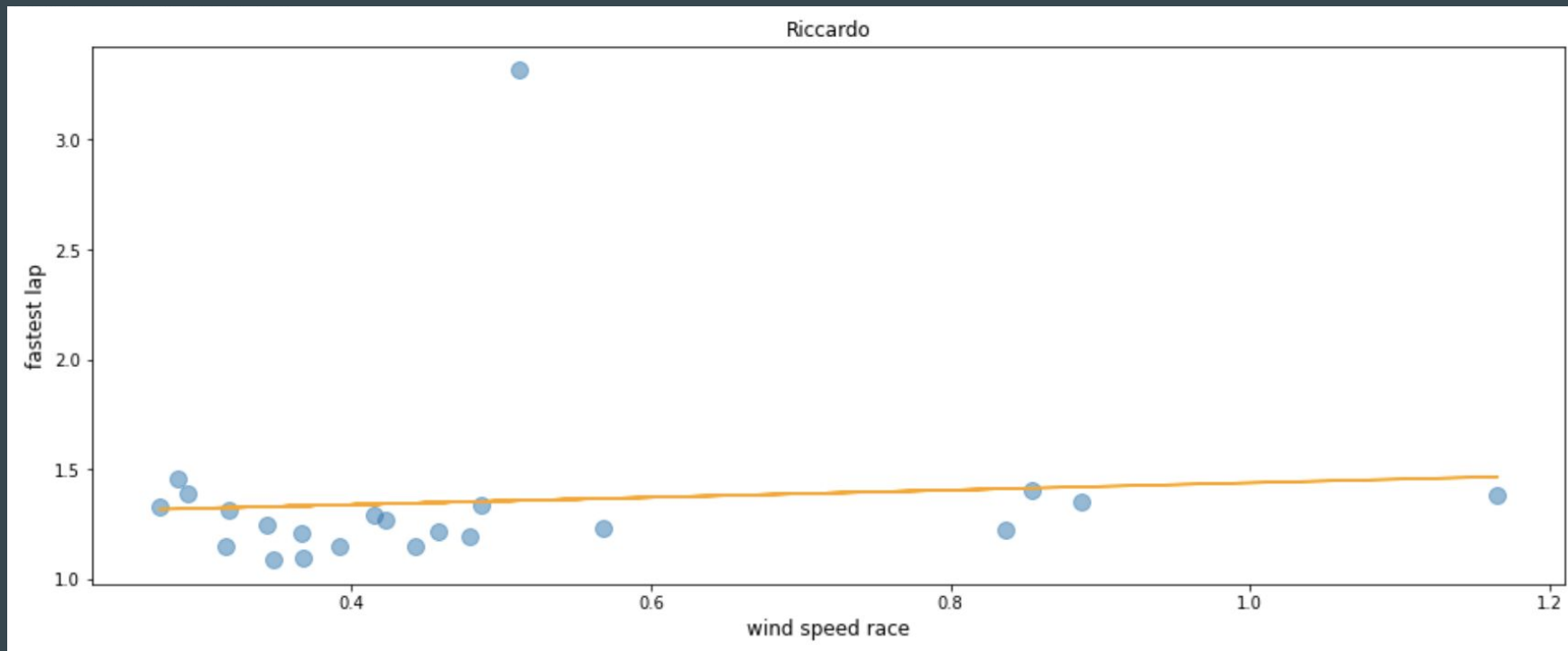
Rainfall vs Fastest Lap

$$R^2 = 0.678$$

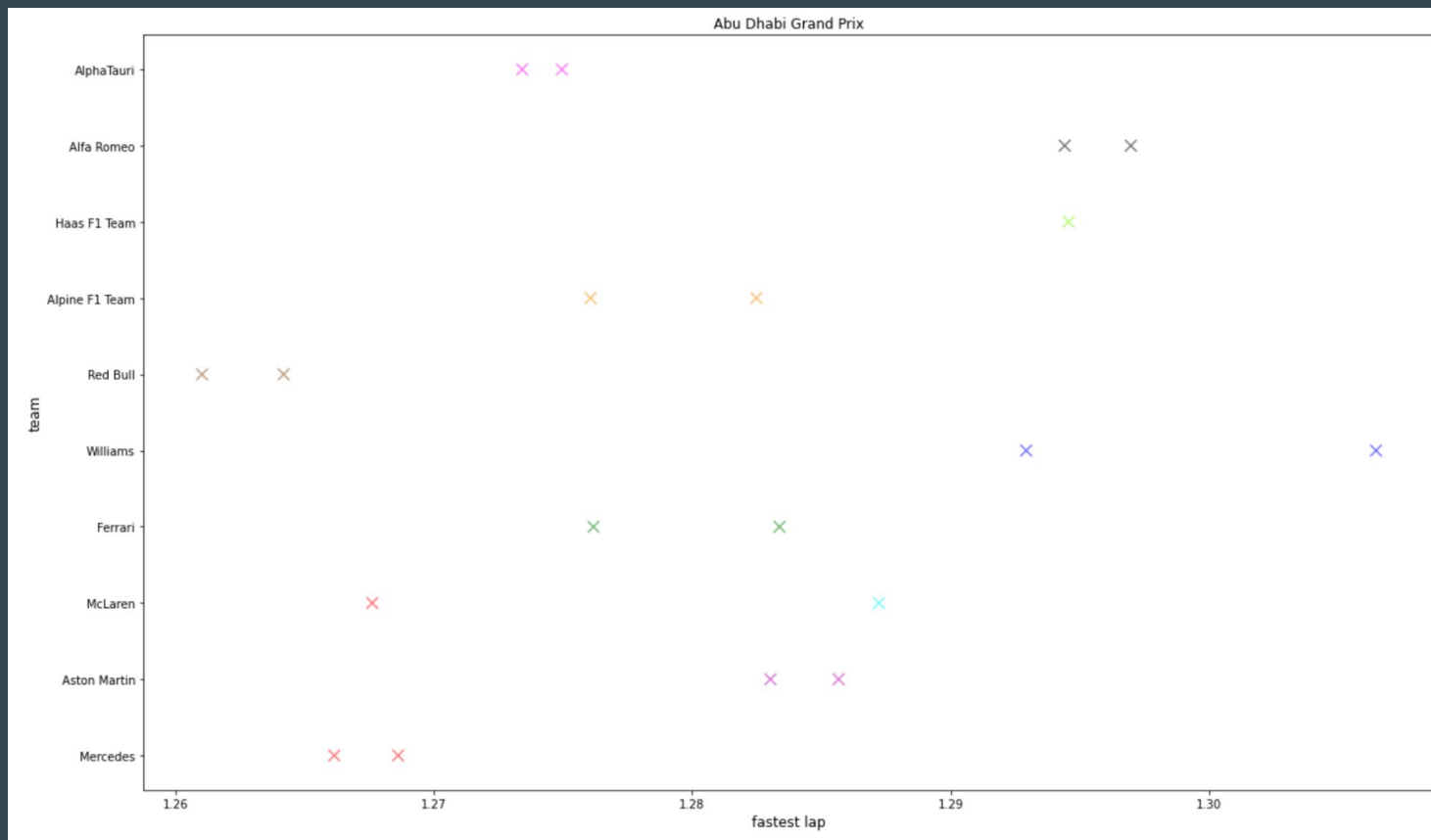


Wind Speed vs Fastest Lap

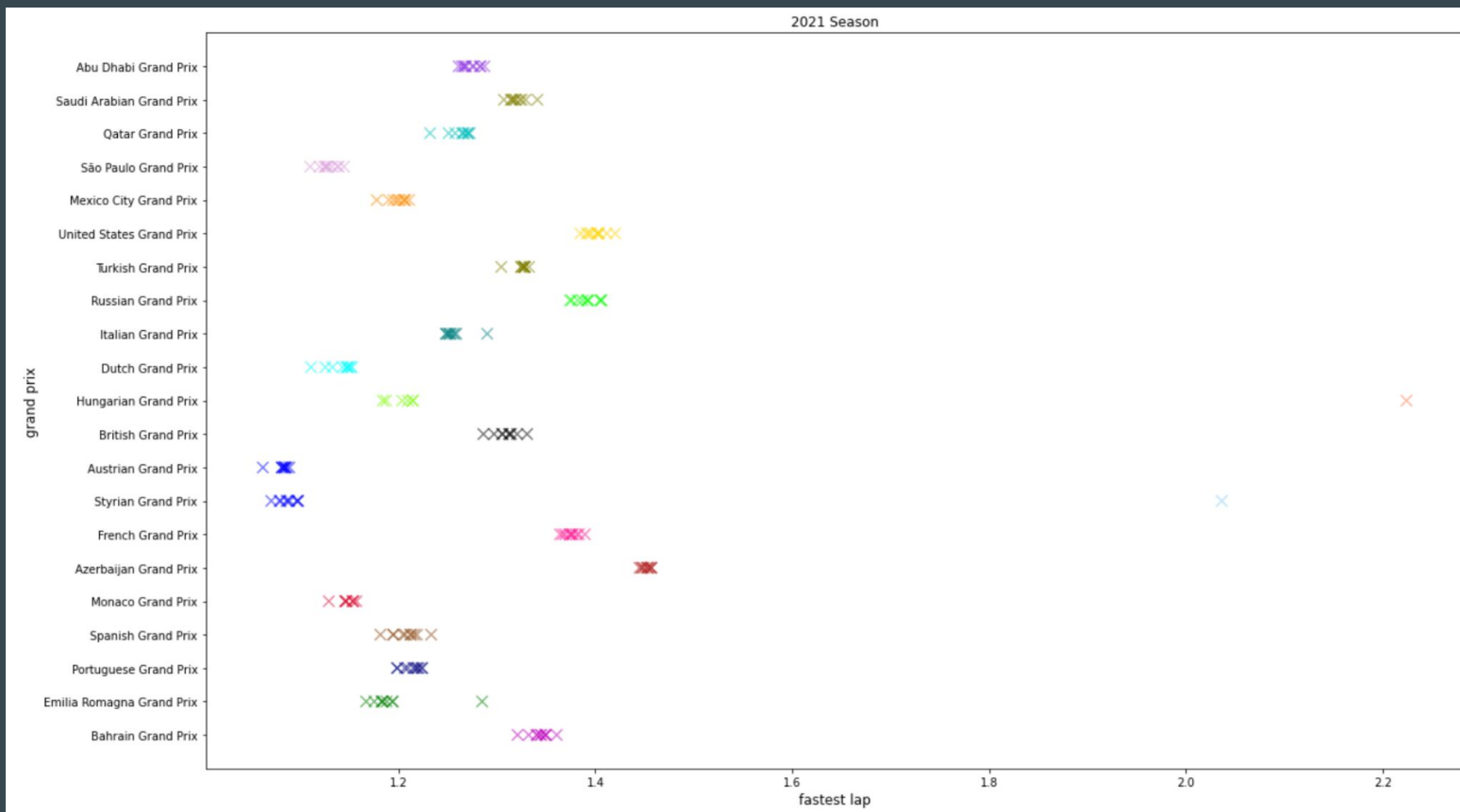
$$R^2 = 0.007$$



Agglomerative Clustering (complete): Abu Dhabi Grand Prix



Agglomerative Clustering (complete): 2021 Season

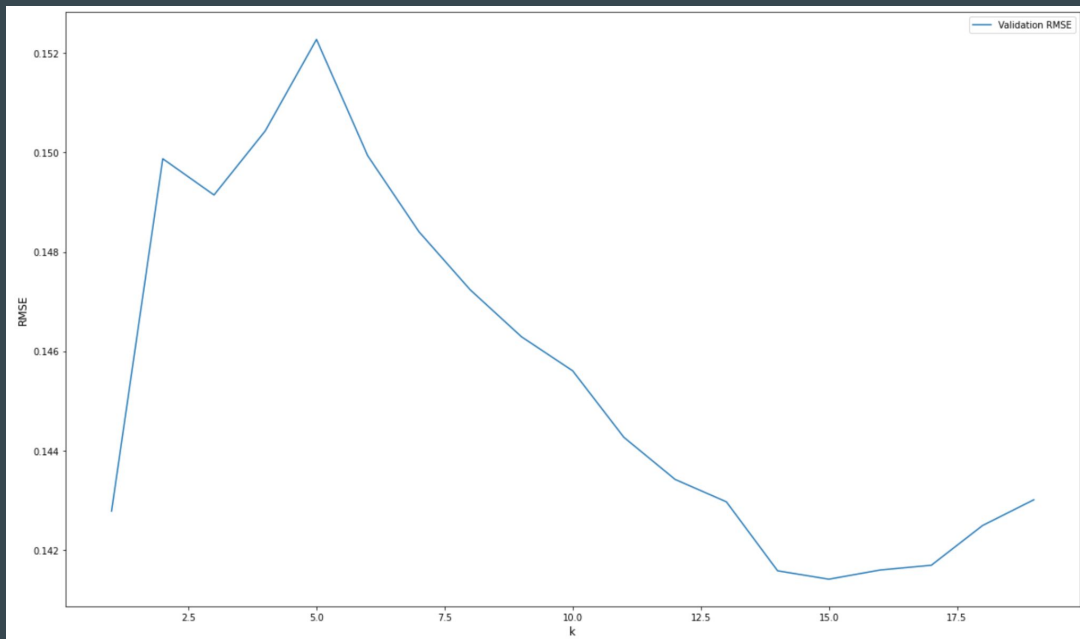


Machine Learning

Machine Learning Models

Variables: qualifying 1, qualifying 2, qualifying 3, race humidity, qualifying humidity, race rainfall, and qualifying rainfall

1. K-Nearest-Neighbors
 - a. $K = 15$
2. Linear Regression
3. Ensemble Models
 - a. Voting
 - b. Stacking



Results

- K-Nearest Neighbors had the best validation RMSE and test RMSE

Model	Validation RMSE	Test RMSE
K-Nearest Neighbors	0.020183	0.03708
Linear Regression	0.021864	0.088008
Voting Model	0.020622	0.054867
Stacking Model	0.021011	0.046329

True Fastest Lap	Predicted Fastest Lap
1:34.932	1:35.829
1:34.015	1:35.674
1:26.419	1:24.827
1:26.615	1:24.827

Thank you!