Ironman

Analyzing feature importances and predicting finish times

Imke Binderszewsky
Majbritt Eckert

The Ironman

01 Swimming

3,862 km

02 Cycling

180,236 km

03 Running

42,195 km

01

Get the Data

Web Scraping with Beautiful Soup

Challenges we faced

/ 01

Official Ironman webpage lately deleted all the results

/ 02

We find an unofficial page with data on all (?) ironmans since 2015

/ 03

unofficial site does not use a uniform scheme for gender and age group

/ 04

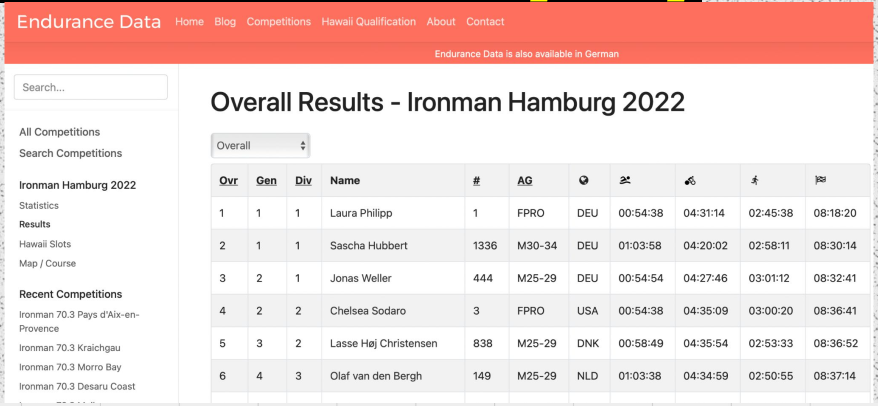
It turns out to be difficult to get past weather data without payment

Challenge Data

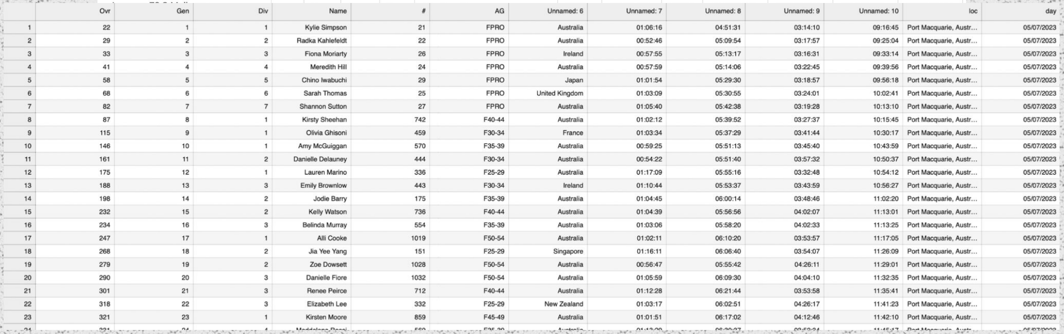
www.endurance-data.com/en/

Participants Data from 283 competitions:

- Finish time and position
- Times for each discipline
- Gender and Age Group
- Country of origin



Ovr	Gen	Div	Name	#	AG					
1	1	1	Laura Philipp	1	FPRO	DEU	00:54:38	04:31:14	02:45:38	08:18:20
2	1	1	Sascha Hubbert	1336	M30-34	DEU	01:03:58	04:20:02	02:58:11	08:30:14
3	2	1	Jonas Weller	444	M25-29	DEU	00:54:54	04:27:46	03:01:12	08:32:41
4	2	2	Chelsea Sodaro	3	FPRO	USA	00:54:38	04:35:09	03:00:20	08:36:41
5	3	2	Lasse Høj Christensen	838	M25-29	DNK	00:58:49	04:35:54	02:53:33	08:36:52
6	4	3	Olaf van den Bergh	149	M25-29	NLD	01:03:38	04:34:59	02:50:55	08:37:14



	Ovr	Gen	Div	Name	#	AG	Unnamed: 6	Unnamed: 7	Unnamed: 8	Unnamed: 9	Unnamed: 10	loc	day
1	22	1	1	Kylie Simpson	21	FPRO	Australia	01:06:16	04:51:31	03:14:10	09:16:45	Port Macquarie, Austr...	05/07/2023
2	29	2	2	Radka Kahlefeldt	22	FPRO	Australia	00:52:46	05:09:54	03:17:57	09:25:04	Port Macquarie, Austr...	05/07/2023
3	33	3	3	Fiona Moriarty	26	FPRO	Ireland	00:57:55	05:13:17	03:16:31	09:33:14	Port Macquarie, Austr...	05/07/2023
4	41	4	4	Meredith Hill	24	FPRO	Australia	00:57:59	05:14:06	03:22:45	09:39:56	Port Macquarie, Austr...	05/07/2023
5	58	5	5	Chino Iwabuchi	29	FPRO	Japan	01:01:54	05:29:30	03:18:57	09:56:18	Port Macquarie, Austr...	05/07/2023
6	68	6	6	Sarah Thomas	25	FPRO	United Kingdom	01:03:09	05:30:55	03:24:01	10:02:41	Port Macquarie, Austr...	05/07/2023
7	82	7	7	Shannon Sutton	27	FPRO	Australia	01:05:40	05:42:38	03:19:28	10:13:10	Port Macquarie, Austr...	05/07/2023
8	87	8	1	Kirsty Sheehan	742	F40-44	Australia	01:02:12	05:39:52	03:27:37	10:15:45	Port Macquarie, Austr...	05/07/2023
9	115	9	1	Olivia Ghisoni	459	F30-34	France	01:03:34	05:37:29	03:41:44	10:30:17	Port Macquarie, Austr...	05/07/2023
10	146	10	1	Amy McGuiggan	570	F35-39	Australia	00:59:25	05:51:13	03:45:40	10:43:59	Port Macquarie, Austr...	05/07/2023
11	161	11	2	Danielle Delauney	444	F30-34	Australia	00:54:22	05:51:40	03:57:32	10:50:37	Port Macquarie, Austr...	05/07/2023
12	175	12	1	Lauren Marino	336	F25-29	Australia	01:17:09	05:55:16	03:32:48	10:54:12	Port Macquarie, Austr...	05/07/2023
13	188	13	3	Emily Brownlow	443	F30-34	Ireland	01:10:44	05:53:37	03:43:59	10:56:27	Port Macquarie, Austr...	05/07/2023
14	198	14	2	Jodie Barry	175	F35-39	Australia	01:04:45	06:00:14	03:48:46	11:02:20	Port Macquarie, Austr...	05/07/2023
15	232	15	2	Kelly Watson	736	F40-44	Australia	01:04:39	05:56:56	04:02:07	11:13:01	Port Macquarie, Austr...	05/07/2023
16	234	16	3	Belinda Murray	554	F35-39	Australia	01:03:06	05:58:20	04:02:33	11:13:25	Port Macquarie, Austr...	05/07/2023
17	247	17	1	Alli Cooke	1019	F50-54	Australia	01:02:11	06:10:20	03:53:57	11:17:05	Port Macquarie, Austr...	05/07/2023
18	268	18	2	Jia Yee Yang	151	F25-29	Singapore	01:16:11	06:06:40	03:54:07	11:26:09	Port Macquarie, Austr...	05/07/2023
19	279	19	2	Zoe Dowsett	1028	F50-54	Australia	00:56:47	05:55:42	04:26:11	11:29:01	Port Macquarie, Austr...	05/07/2023
20	290	20	3	Danielle Fiore	1032	F50-54	Australia	01:05:59	06:09:30	04:04:10	11:32:35	Port Macquarie, Austr...	05/07/2023
21	301	21	3	Renee Peirce	712	F40-44	Australia	01:12:28	06:21:44	03:53:58	11:35:41	Port Macquarie, Austr...	05/07/2023
22	318	22	3	Elizabeth Lee	332	F25-29	New Zealand	01:03:17	06:02:51	04:26:17	11:41:23	Port Macquarie, Austr...	05/07/2023
23	321	23	1	Kirsten Moore	859	F45-49	Australia	01:01:51	06:17:02	04:12:46	11:42:10	Port Macquarie, Austr...	05/07/2023

Weather Data

www.visualcrossing.com/weather/weather-data-services#/

We chose:

- Maximum and average Temperature
- Humidity and Precipitation
- Windspeed
- Cloudiness

	name	datetime	tempmax	temp	humidity	precip	windspeed	cloudcover
1	Port Macquarie, NSW...	05/07/2023	24	15.2	59.5	0	26.6	37.9
2	The Woodlands, TX, ...	04/22/2023	25.8	19.9	57.8	0	22.8	16.6
3	Gqeberha, Eastern C...	03/05/2023	25	21.9	90.3	10.223	48.2	82.1
4	Taupo, Waikato, New ...	03/04/2023	22.7	18.7	81.8	0	20.5	44.1
5	Taupo, Waikato, New ...	12/10/2022	23.2	18.1	91.8	31.4	22.3	95.5
6	Mar del Plata, Bueno...	12/04/2022	31.1	21.2	69.1	0	29.5	36
7	Busselton, WA 6280, ...	12/03/2022	25.8	19	70.6	0	38.4	60.6
8	Port Macquarie, NSW...	05/07/2023	24	15.2	59.5	0	26.6	37.9
9	The Woodlands, TX, ...	04/22/2023	25.8	19.9	57.8	0	22.8	16.6
10	Gqeberha, Eastern C...	03/05/2023	25	21.9	90.3	10.223	48.2	82.1
11	Taupo, Waikato, New ...	03/04/2023	22.7	18.7	81.8	0	20.5	44.1
12	Taupo, Waikato, New ...	12/10/2022	23.2	18.1	91.8	31.4	22.3	95.5
13	Mar del Plata, Bueno...	12/04/2022	31.1	21.2	69.1	0	29.5	36
14	Busselton, WA 6280, ...	12/03/2022	25.8	19	70.6	0	38.4	60.6
15	Port Macquarie, NSW...	05/07/2023	24	15.2	59.5	0	26.6	37.9
16	The Woodlands, TX, ...	04/22/2023	25.8	19.9	57.8	0	22.8	16.6
17	Gqeberha, Eastern C...	03/05/2023	25	21.9	90.3	10.223	48.2	82.1
18	Taupo, Waikato, New ...	03/04/2023	22.7	18.7	81.8	0	20.5	44.1
19	Taupo, Waikato, New ...	12/10/2022	23.2	18.1	91.8	31.4	22.3	95.5
20	Mar del Plata, Bueno...	12/04/2022	31.1	21.2	69.1	0	29.5	36
21	Busselton, WA 6280, ...	12/03/2022	25.8	19	70.6	0	38.4	60.6
22	Port Macquarie, NSW...	05/07/2023	24	15.2	59.5	0	26.6	37.9
23	The Woodlands, TX, ...	04/22/2023	25.8	19.9	57.8	0	22.8	16.6
24	Gqeberha, Eastern C...	03/05/2023	25	21.9	90.3	10.223	48.2	82.1

PREPROCESSING

/ COUNTRY

- Change country codes to number
- Add mean Temperature

/ WEATHER

- Add weather on challenge day

/ AGEGROUPS

- Translating age groups
- PRO as own group

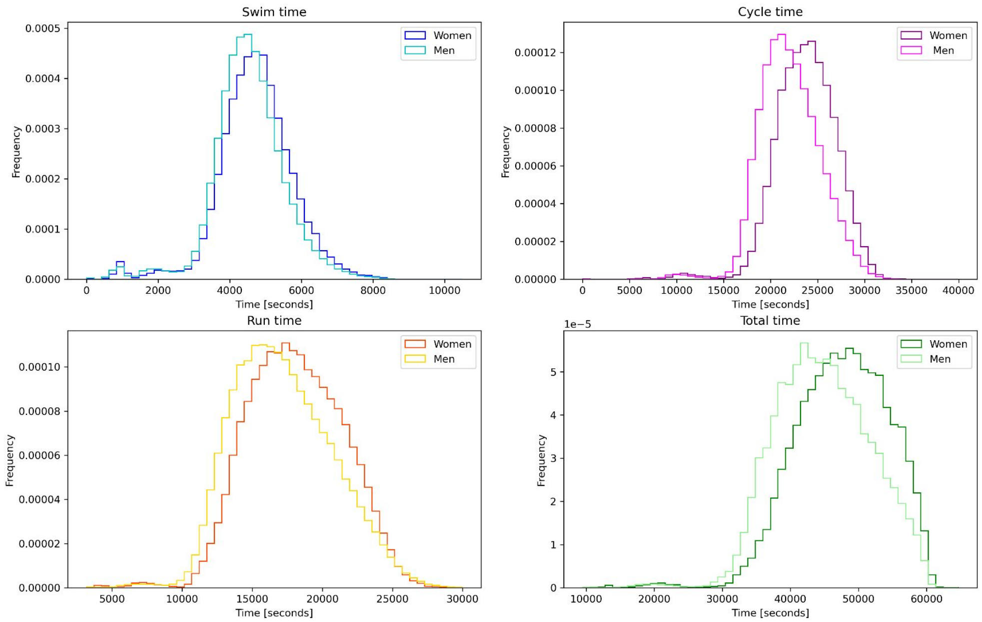
/ DATA CLEANSING

- Exclude unreadable countries and incomplete points
- Exclude outliers to avoid overtraining

FINAL DATA SET

The final data… Rank',
'Name', 'Star… Run Time'',
'Total Time'', … ure',
'Humidity', 'F…

Our preproce…



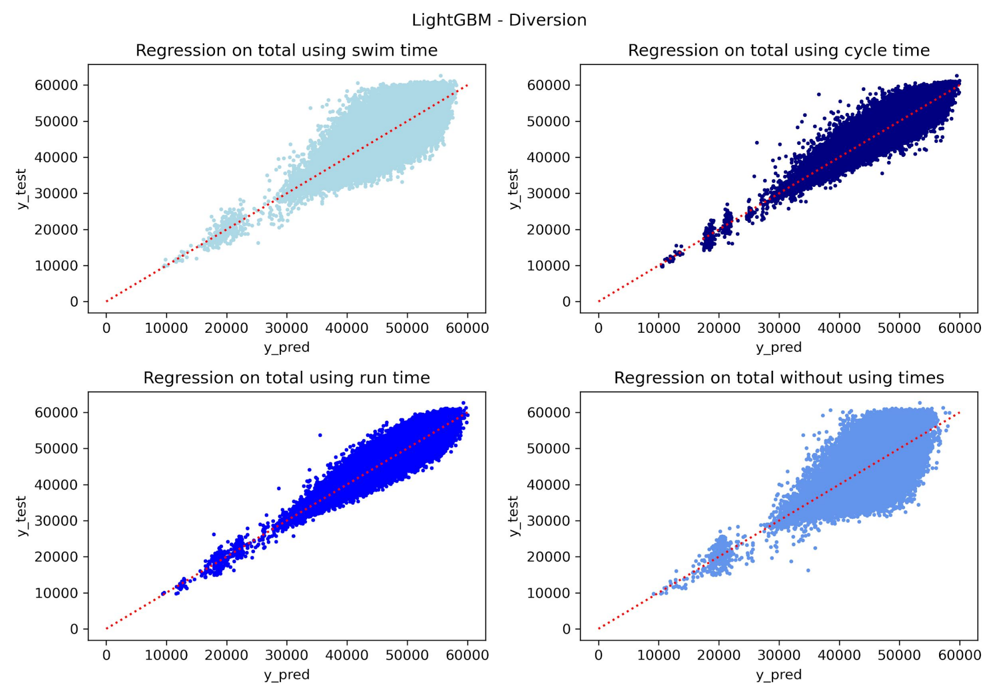
	Overall	Gender	Agegroup															dity	Precip	Wind	Clouds	Country
1	22	1	1	Kylie Sim														9.5	0.0	26.6	37.9	100
2	29	2	2	Radka Kahl														9.5	0.0	26.6	37.9	100
3	33	3	3	Fiona Mc														9.5	0.0	26.6	37.9	152
4	41	4	4	Meredi														9.5	0.0	26.6	37.9	100
5	58	5	5	Chino Iwa														9.5	0.0	26.6	37.9	140
6	68	6	6	Sarah Th														9.5	0.0	26.6	37.9	159
7	82	7	7	Shannon S														9.5	0.0	26.6	37.9	100
8	87	8	1	Kirsty She														9.5	0.0	26.6	37.9	100
9	115	9	1	Olivia Gl														9.5	0.0	26.6	37.9	143
10	146	10	1	Amy McGu														9.5	0.0	26.6	37.9	100
11	161	11	2	Danielle Dela														9.5	0.0	26.6	37.9	100
12	175	12	1	Lauren Marino	336	2	21.65	4629	21316	12768	39252	13	05/07/2023	1	24.0	15.2		59.5	0.0	26.6	37.9	100
13	188	13	3	Emily Brownlow	443	3	9.3	4244	21217	13439	39387	13	05/07/2023	1	24.0	15.2		59.5	0.0	26.6	37.9	152
14	198	14	2	Jodie Barry	175	4	21.65	3885	21614	13726	39740	13	05/07/2023	1	24.0	15.2		59.5	0.0	26.6	37.9	100

02

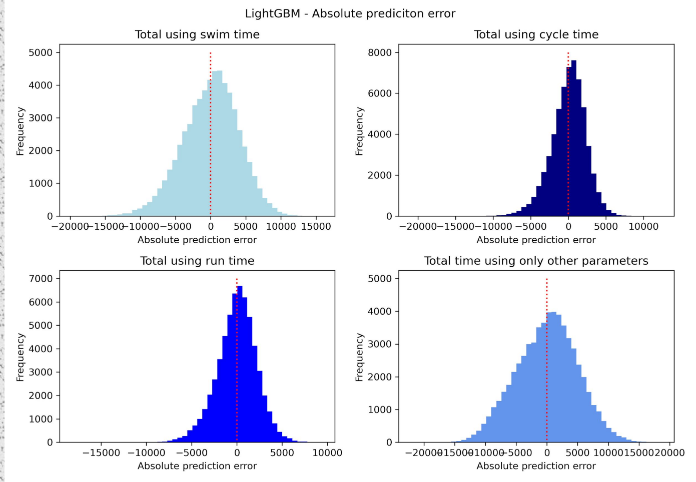
Perform ML

LIGHTGBM – REGRESSION ON 'TOTAL'

['Number', 'Age', 'Climate', 'Swim', 'Cycle', 'Run', 'Total', 'City', 'MF', 'Tempmax', 'Temperature', 'Humidity', 'Precip', 'Wind', 'Clouds', 'Country']

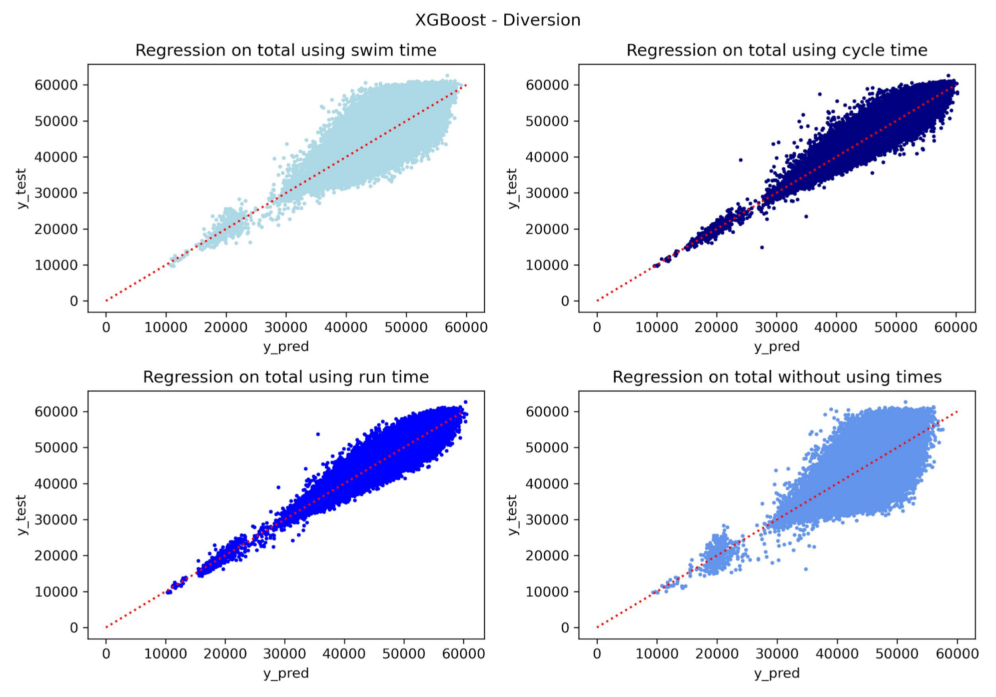


- Regression on 'Total' using one 'other time' at most
- Hyperparameters optimized using Bayesian optimization in optuna

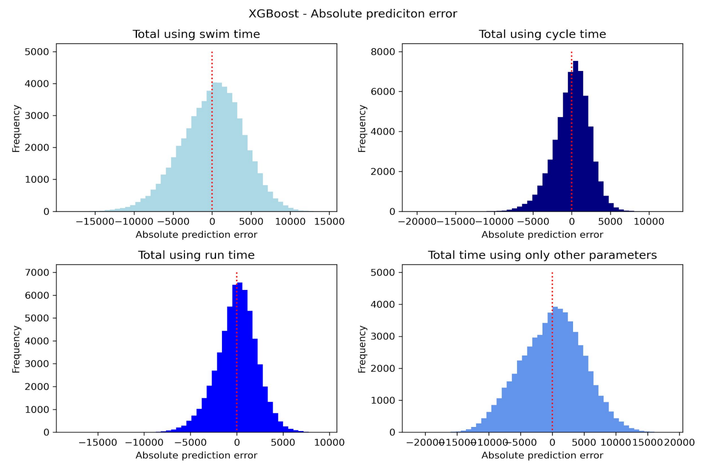


XGBOOST – REGRESSION ON 'TOTAL'

['Number', 'Age', 'Climate', 'Swim', 'Cycle', 'Run', 'Total', 'City', 'MF', 'Tempmax', 'Temperature', 'Humidity', 'Precip', 'Wind', 'Clouds', 'Country']

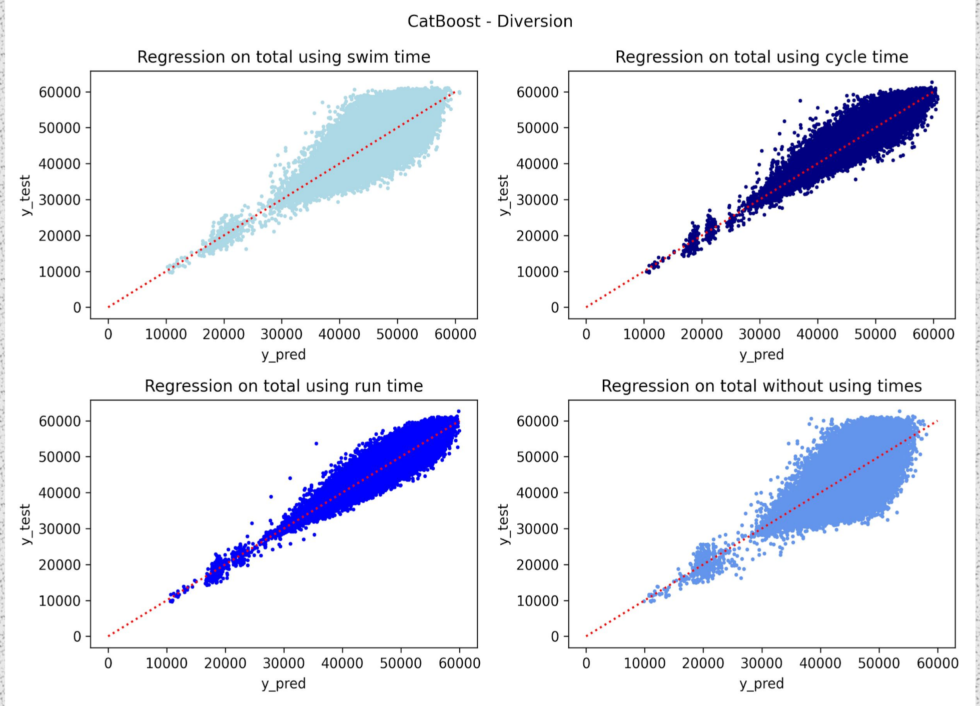


- Regression on 'Total' using one 'other time' at most
- Hyperparameters optimized using Bayesian optimization in optuna

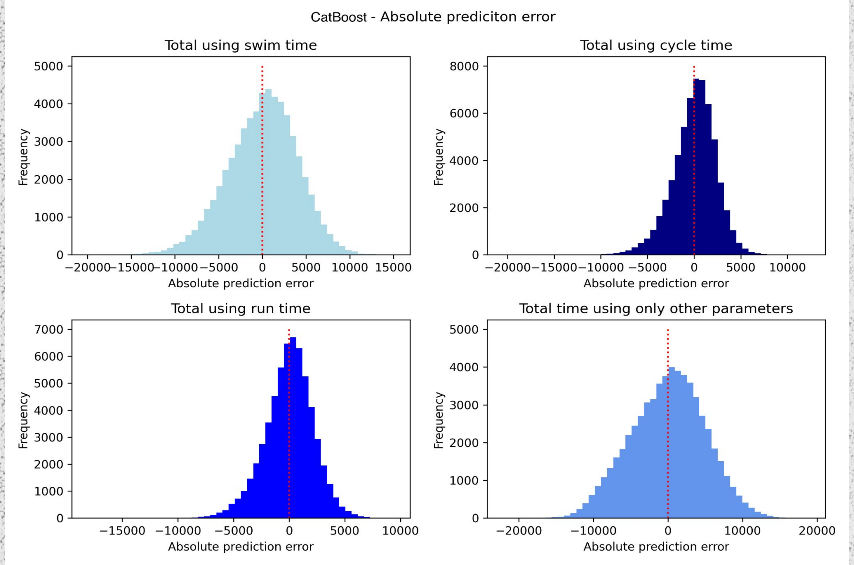


CatBOOST– REGRESSION ON 'TOTAL'

['Number', 'Age', 'Climate', 'Swim', 'Cycle', 'Run', 'Total', 'City', 'MF', 'Tempmax', 'Temperature', 'Humidity', 'Precip', 'Wind', 'Clouds', 'Country']

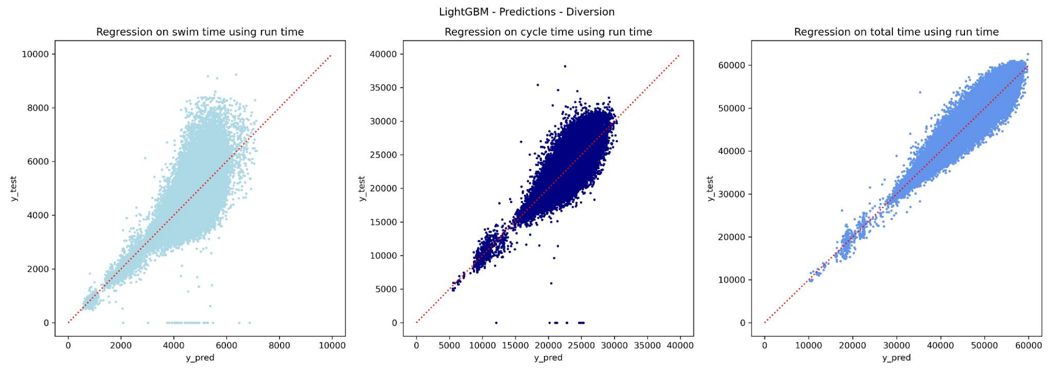


- Regression on 'Total' using one 'other time' at most

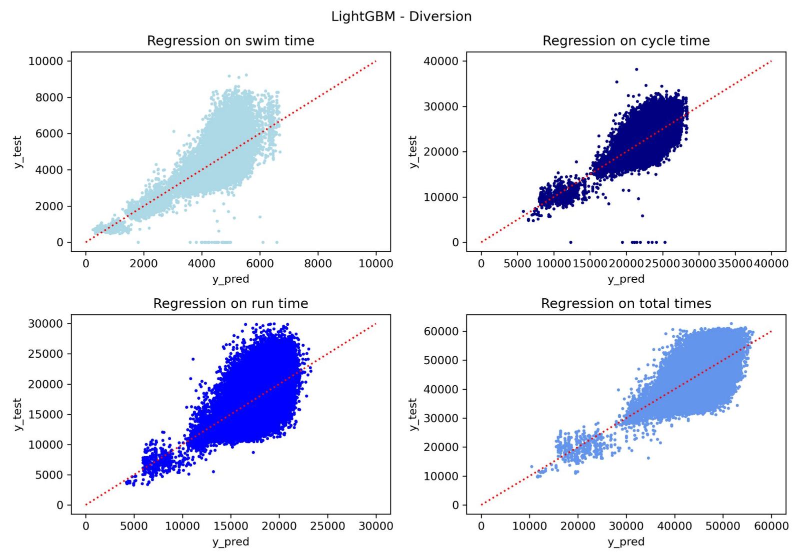


LIGHTGBM – PREDICTIONS

['Number', 'Age', 'Climate', 'Swim', 'Cycle', 'Run', 'Total', 'City', 'MF', 'Tempmax', 'Temperature', 'Humidity', 'Precip', 'Wind', 'Clouds', 'Country']



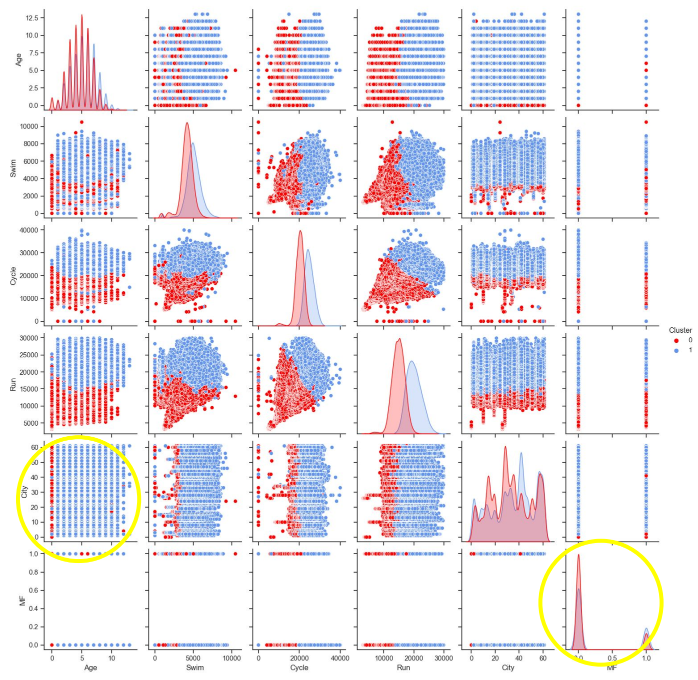
- Regression on 'Swim', 'Cycle' , 'Run', 'Total' using only other parameters

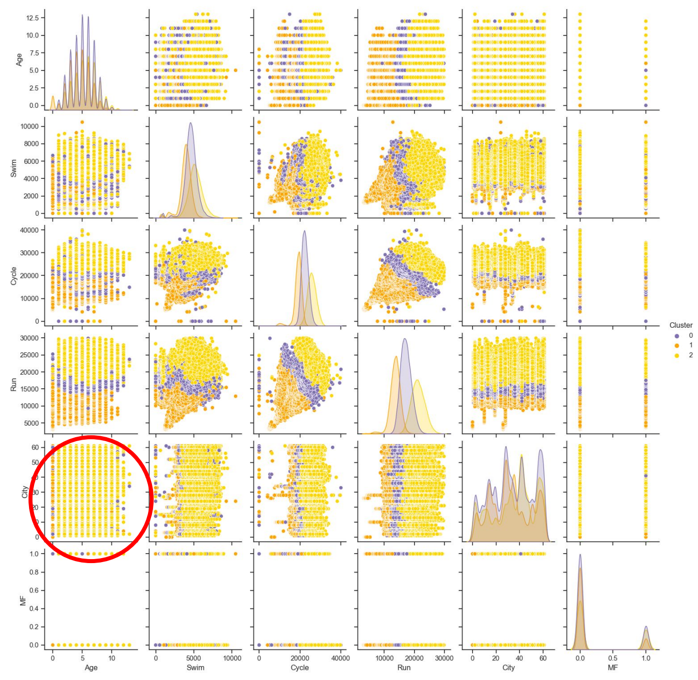


- Regression on 'Swim', 'Cycle' , 'Total' using one 'Run' and other parameters
- Hyperparameters optimized using Bayesian optimization in optuna

KMEANS – CLUSTERING

- Number of clusters with elbow method
- Features used for pairplot: Age, Swim, Cycle, Run, City and MF





- k=2 can kind of see a difference between men and women
- k=2 and k=3 can distinguish between PRO/better and slower athletes
- Tried out PCA and TSNE

03

Evaluation

PERFORMANCE

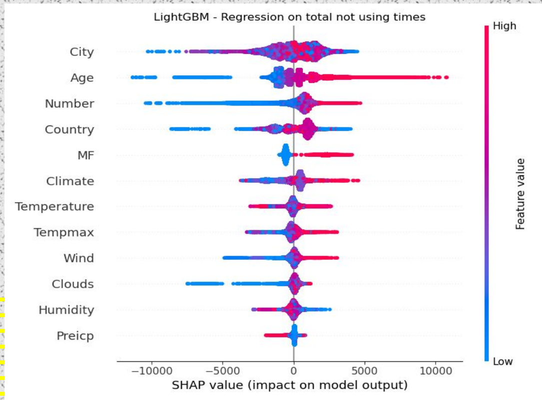
LightGBM - Hyperparameters optimized				
[s]	Predicting total time using swim time	Predicting total time cycle time	Predicting total time using run time	Predicting total time using only other parameters
MAE RAE RMSE	3183.448 0.070 4004.663	1804.379 0.034 2338.322	1701.279 0.038 2186.367	3982.385 0.088 4952.040
XGBoost - Hyperparameters optimized				
[s]	Predicting total time using swim time	Predicting total time cycle time	Predicting total time using run time	Predicting total time using only other parameters
MAE RAE RMSE	3186.601 0.070 4011.422	1801.978 0.040 2334.952	1700.808 0.038 2191.638	3991.294 0.088 4962.010
CatBoost				
[s]	Predicting total time using swim time	Predicting total time cycle time	Predicting total time using run time	Predicting total time using only other parameters
MAE RAE RMSE	3169.863 0.070 3984.920	1802.426 0.040 2334.845	1692.227 0.037 2182.061	4003.017 0.088 4965.796
Mean Absolute Error (MAE), Relative Absolute Error (RAE), Root Mean Squared Error (RMSE)				

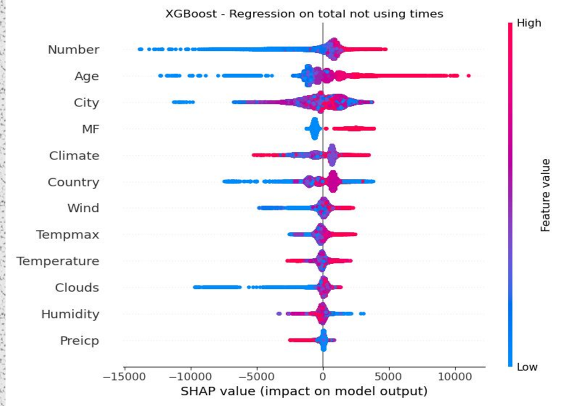
Overall Feature Importances

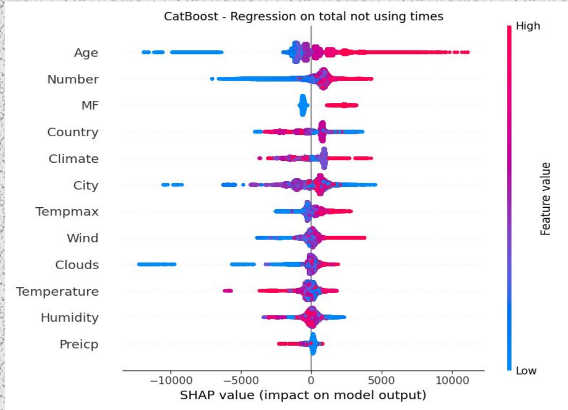
1. Startnumber
2. Age
3. City
4. Origin Country
5. Gender

Paper - M. Thuany et al.: A Machine Learning Approach to Finding the Fastest Race Course for Professional Athletes Competing in Ironman® 70.3 Races between 2004 and 2020

- Data from ironman.com
- Parameters: Gender_ID, EventLocation_ID and Country_ID
- Algorithms: Random Forest, XGBoost, CatBoost, Single Decision Tree
- We also get City and Country as top variables







LIGHTGBM – Predictions

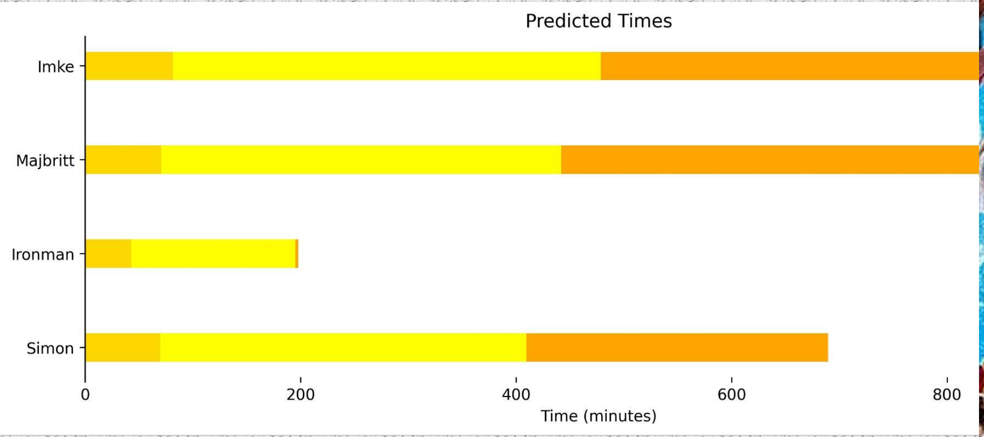
LightGBM - Hyperparameters optimized			
	Predicting swim time using run time	Predicting cycle time using run time	Predicting total time using run time
MAE	478.230	1313.654	1698.317
RAE	0.105	0.059	0.038
RMSE	642.850	1732.432	2190.447
Mean Absolute Error (MAE), Relative Absolute Error (RAE), Root Mean Squared Error (RMSE)			
Top Features (SHAP):	Run, City, **Wind, Tempmax**	Run, City, Gender, Startnumber	Run, City, Gender, Startnumber

LightGBM				
	Predicting swim time	Predicting cycle time	Predicting run time	Predicting total time
MAE	529.140	1781.025	2237.361	4109.109
RAE	0.116	0.080	0.130	0.091
RMSE	699.541	2254.336	2785.358	5072.489
Mean Absolute Error (MAE), Relative Absolute Error (RAE), Root Mean Squared Error (RMSE)				
Top Features (SHAP):	City, Age, **Wind**	Gender, City, Startnumber	Age, Country, Climate	Age, Startnumber, City

04

Results

PREDICTIONS FOR FRIEN



Predictions on friends				
	Known/Guessed Marathon time	Predicting swim time using run time	Pred using run time	using run time
Imke	06:04:00	01:21:33	06:37:25	14:37:18
Majbritt	07:46:40	01:10:41	06:11:21	15:35:39
Simon	04:40:00	01:09:43	05:40:03	11:49:01
Ironman	00:02:44	00:42:55	02:32:49	04:38:26
[hh:mm:ss]				

FINAL Results

Results:

- The best predictions are made by a model that uses the Run time as a parameter
- Important features are the **City** where the event is located, **Country and Climate** where the runner is from and **Age**
- The weather data doesn’t have a big influence, **Wind** and Cycling time are correlated and Wind has importance in predicting Swimming time

Further improvement:

- Train on data from athletes training
- Use more detailed data including break times
- Unsupervised learning

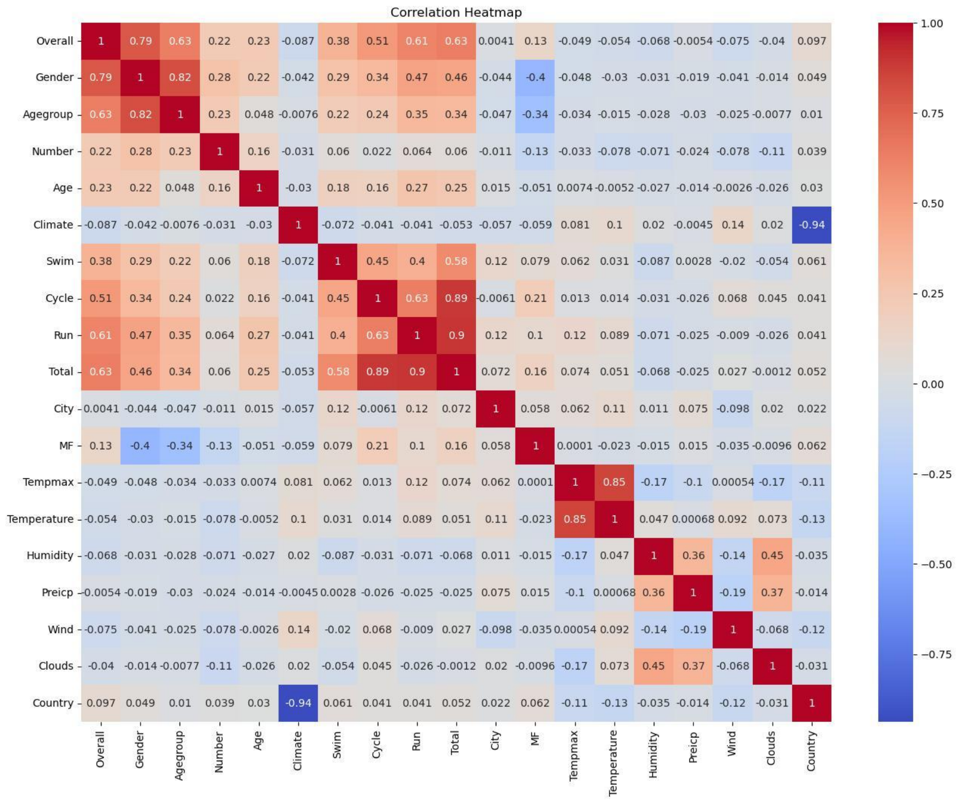
THANKS!

Detailed Data Preprocessing

- Merging race data and weather data based on the location and date of the race (removing lines without matches)
- Replacing country codes with real names (some unknown by hand, some unknown removed)
- Replacing country with average temperature and country index in another column (remove lines with unknown countries)
- Make male/female and age group column from joint string
- Convert times in hh:mm:ss format to seconds (remove runs with missing total times)
- Make list of all run locations and replace location by index in that list
- Removing outliers based on looking at histograms
- Split dataset into male/female

	Overall	Gender	Agegroup	Name	Number	Age	Climate	Swim	Cycle	Run	Total	City	Date	MF	Tempmax	Temperature	Humidity	Precip	Wind	Clouds	Country
1	22	1	1	Kylie Simpson	21	0	21.65	3976	17491	11650	33405	13	05/07/2023	1	24.0	15.2	59.5	0.0	26.6	37.9	100
2	29	2	2	Radka Kahlefeldt	22	0	21.65	3166	18594	11877	33904	13	05/07/2023	1	24.0	15.2	59.5	0.0	26.6	37.9	100
3	33	3	3	Fiona Moriarty	26	0	9.3	3475	18797	11791	34394	13	05/07/2023	1	24.0	15.2	59.5	0.0	26.6	37.9	152
4	41	4	4	Meredith Hill	24	0	21.65	3479	18846	12165	34796	13	05/07/2023	1	24.0	15.2	59.5	0.0	26.6	37.9	100
5	58	5	5	Chino Iwabuchi	29	0	11.15	3714	19770	11937	35778	13	05/07/2023	1	24.0	15.2	59.5	0.0	26.6	37.9	140
6	68	6	6	Sarah Thomas	25	0	8.45	3789	19855	12241	36161	13	05/07/2023	1	24.0	15.2	59.5	0.0	26.6	37.9	159
7	82	7	7	Shannon Sutton	27	0	21.65	3940	20558	11968	36790	13	05/07/2023	1	24.0	15.2	59.5	0.0	26.6	37.9	100
8	87	8	1	Kirsty Sheehan	742	5	21.65	3732	20392	12457	36945	13	05/07/2023	1	24.0	15.2	59.5	0.0	26.6	37.9	100
9	115	9	1	Olivia Ghisoni	459	3	10.7	3814	20249	13304	37817	13	05/07/2023	1	24.0	15.2	59.5	0.0	26.6	37.9	143
10	146	10	1	Amy McGuiggan	570	4	21.65	3565	21073	13540	38639	13	05/07/2023	1	24.0	15.2	59.5	0.0	26.6	37.9	100
11	161	11	2	Danielle Delauney	444	3	21.65	3262	21100	14252	39037	13	05/07/2023	1	24.0	15.2	59.5	0.0	26.6	37.9	100
12	175	12	1	Lauren Marino	336	2	21.65	4629	21316	12768	39252	13	05/07/2023	1	24.0	15.2	59.5	0.0	26.6	37.9	100
13	188	13	3	Emily Brownlow	443	3	9.3	4244	21217	13439	39387	13	05/07/2023	1	24.0	15.2	59.5	0.0	26.6	37.9	152
14	198	14	2	Jodie Barry	175	4	21.65	3885	21614	13726	39740	13	05/07/2023	1	24.0	15.2	59.5	0.0	26.6	37.9	100
15	232	15	2	Kelly Watson	736	5	21.65	3879	21416	14527	40381	13	05/07/2023	1	24.0	15.2	59.5	0.0	26.6	37.9	100
16	234	16	3	Belinda Murray	554	4	21.65	3786	21500	14553	40405	13	05/07/2023	1	24.0	15.2	59.5	0.0	26.6	37.9	100
17	247	17	1	Alli Cooke	1019	7	21.65	3731	22220	14037	40625	13	05/07/2023	1	24.0	15.2	59.5	0.0	26.6	37.9	100
18	268	18	2	Jia Yee Yang	151	2	26.45	4571	22000	14047	41169	13	05/07/2023	1	24.0	15.2	59.5	0.0	26.6	37.9	30
19	279	19	2	Zoe Dowsett	1028	7	21.65	3407	21342	15971	41341	13	05/07/2023	1	24.0	15.2	59.5	0.0	26.6	37.9	100
20	290	20	3	Danielle Fiore	1032	7	21.65	3959	22170	14650	41555	13	05/07/2023	1	24.0	15.2	59.5	0.0	26.6	37.9	100
21	301	21	3	Renee Peirce	712	5	21.65	4348	22904	14038	41741	13	05/07/2023	1	24.0	15.2	59.5	0.0	26.6	37.9	100
22	318	22	3	Elizabeth Lee	332	2	10.55	3797	21771	15977	42083	13	05/07/2023	1	24.0	15.2	59.5	0.0	26.6	37.9	146
23	321	23	1	Kirsten Moore	859	6	21.65	3711	22622	15166	42130	13	05/07/2023	1	24.0	15.2	59.5	0.0	26.6	37.9	100
24	331	24	4	Maddalena Rossi	569	4	21.65	4409	23437	13954	42317	13	05/07/2023	1	24.0	15.2	59.5	0.0	26.6	37.9	100

Correlation Heatmap



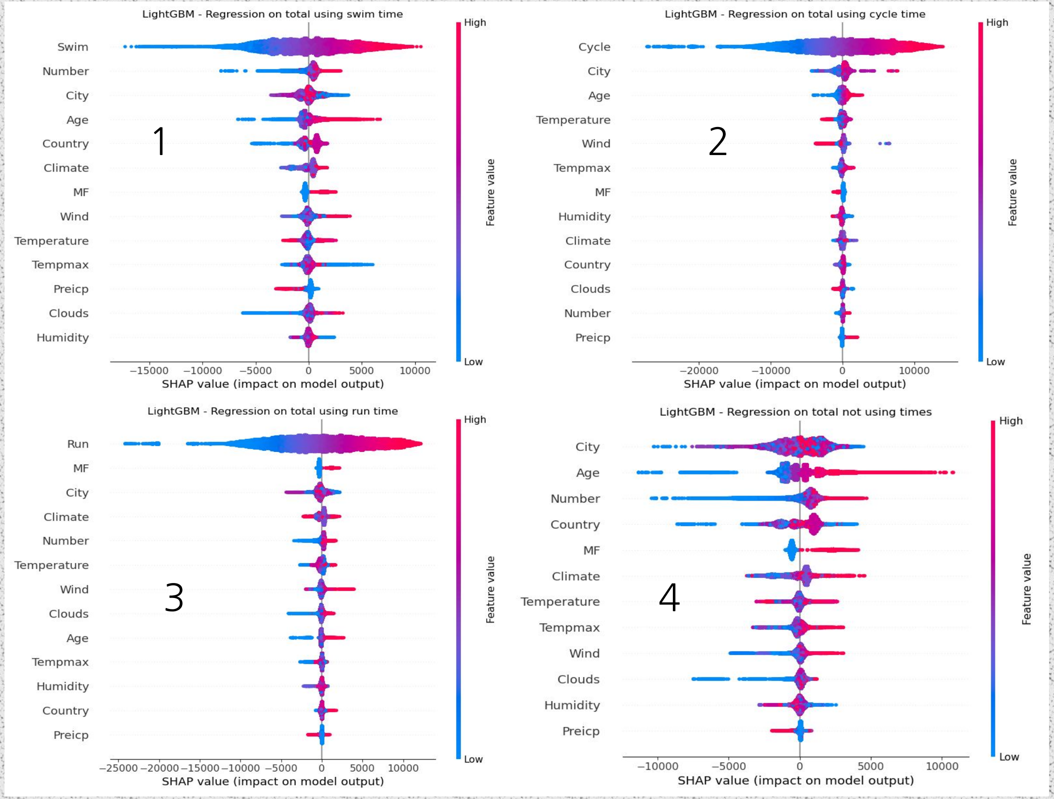
Hyperparameters

- LIGHTGBM:
total: Best Hyperparameters: {'learning_rate': 0.15597416449990634, 'max_depth': 10, 'num_leaves': 198, 'feature_fraction': 0.6296160540883923}
Best RMSE: 4952.039943876661
swim:Best Hyperparameters: {'learning_rate': 0.12816280823430914, 'max_depth': 10, 'num_leaves': 183, 'feature_fraction': 0.8610512261893823}
Best RMSE: 4004.663002332225
cycle:Best Hyperparameters: {'learning_rate': 0.09800265450015587, 'max_depth': 10, 'num_leaves': 181, 'feature_fraction': 0.9844702634058375}
Best RMSE: 2338.3216955607645
run: Best Hyperparameters: {'learning_rate': 0.112136813996687, 'max_depth': 10, 'num_leaves': 200, 'feature_fraction': 0.5376521049190062}
Best RMSE: 2186.367473825411

- XGBOOST:
total:Best Hyperparameters: {'learning_rate': 0.12357139253345416, 'max_depth': 10, 'subsample': 0.5484018958498692, 'colsample_bytree': 0.7452035307775882, 'lambda': 5.823705133797285, 'alpha': 2.0232275922730846e-05}
Best RMSE: 4962.010127368788
swim:Best Hyperparameters: {'learning_rate': 0.12057455458140323, 'max_depth': 9, 'subsample': 0.850744607798772, 'colsample_bytree': 0.8075744401715309, 'lambda': 1.0142267389641621e-05, 'alpha': 0.00020427694930341 72}
Best RMSE: 4011.4218682072856
cycle:Best Hyperparameters: {'learning_rate': 0.11017677248900799, 'max_depth': 10, 'subsample': 0.9359769673858653, 'colsample_bytree': 0.8455505177419685, 'lambda': 7.107806609480754, 'alpha': 0.000993751502189804}
Best RMSE: 2334.951790090717
run:Best Hyperparameters: {'learning_rate': 0.15578388056334727, 'max_depth': 8, 'subsample': 0.7572926891290177, 'colsample_bytree': 0.9466981758239578, 'lambda': 0.00019582984010771745, 'alpha': 4.42274725667293e-05}
Best RMSE: 2191.637916694315

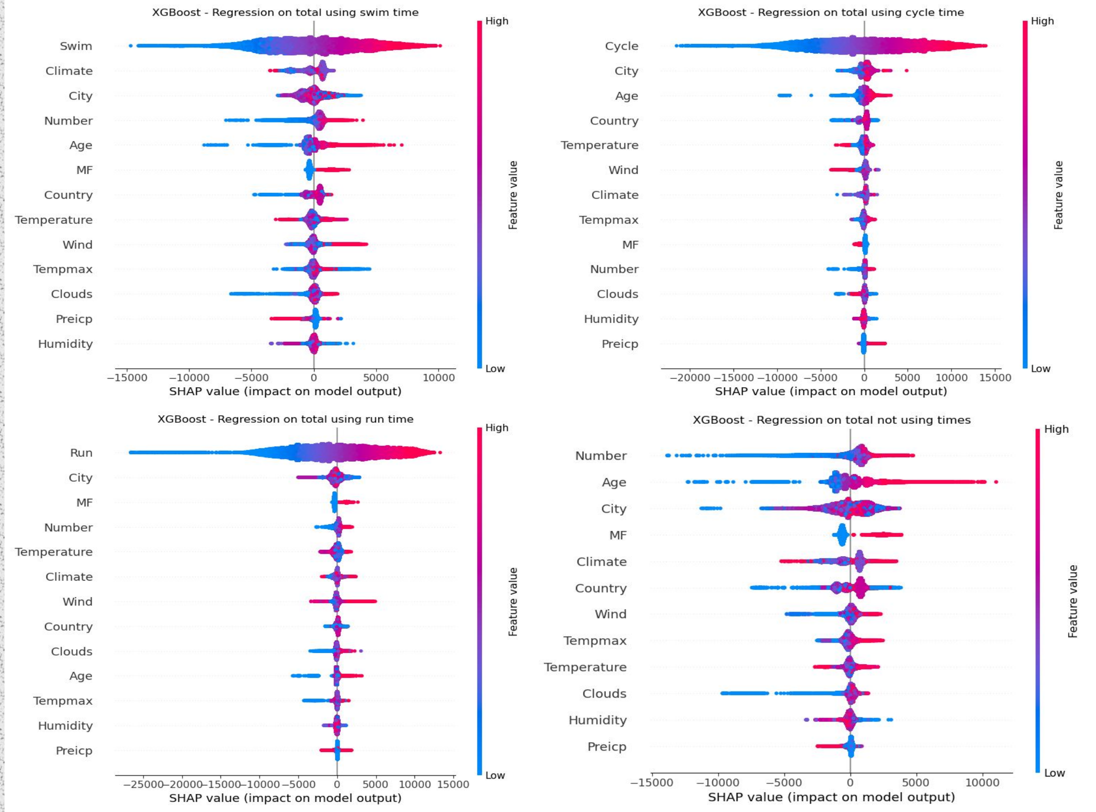
- FRIENDS:
swim:Best Hyperparameters: {'learning_rate': 0.10137359043701283, 'max_depth': 10, 'num_leaves': 165, 'feature_fraction': 0.825008273534942}
Best RMSE: 642.8496741487756
cycle:Best Hyperparameters: {'learning_rate': 0.14434107388313977, 'max_depth': 10, 'num_leaves': 179, 'feature_fraction': 0.8204398355246291}
Best RMSE: 1732.4319819024483
total:Best Hyperparameters: {'learning_rate': 0.15027872303736747, 'max_depth': 10, 'num_leaves': 184, 'feature_fraction': 0.929343948461026}
Best RMSE: 2190.4474294535435

LIGHTGBM – SHAP Values



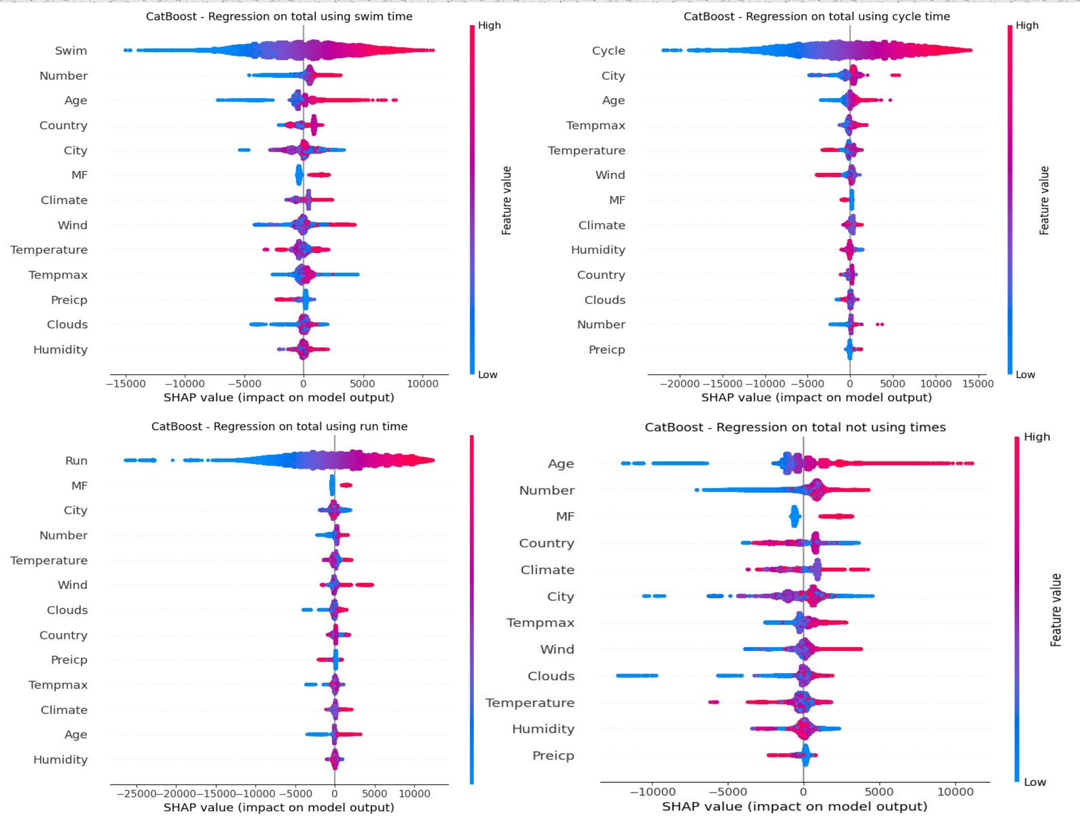
1. Swim, Startnumber, City (Swimming track), Age, Country and Climate where they're from. Weather is not that important.
2. Cycle, City (Cycle track), Age, Temperature and Wind are important (Headwind).
3. Run, Gender, City (Run track), Climate where they're from. Age is less important.
4. City, Age, Startnumber, Country where they're from. Weather is least important.

XGBOOST – SHAP Values



1. Swim, Climate where they're from, City (Swimming track), Startnumber.
   Weather is not that important.
2. Cycle, City (Cycle track), Age, Temperature and Wind are important (Headwind)
3. Run, City (Run track), Gender, Startnumber, Climate where they're from,
   Age is less important
4. Startnumber, Age, City, Gender
   Country where they're from
   Weather is least important

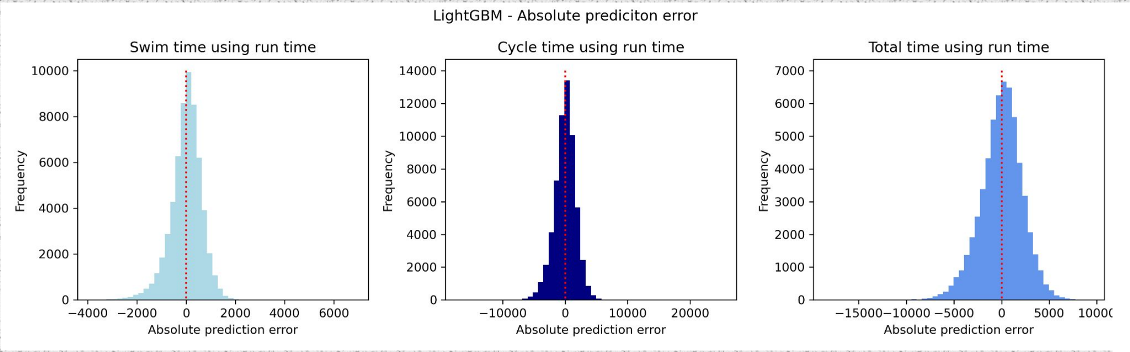
CatBOOST – SHAP Values



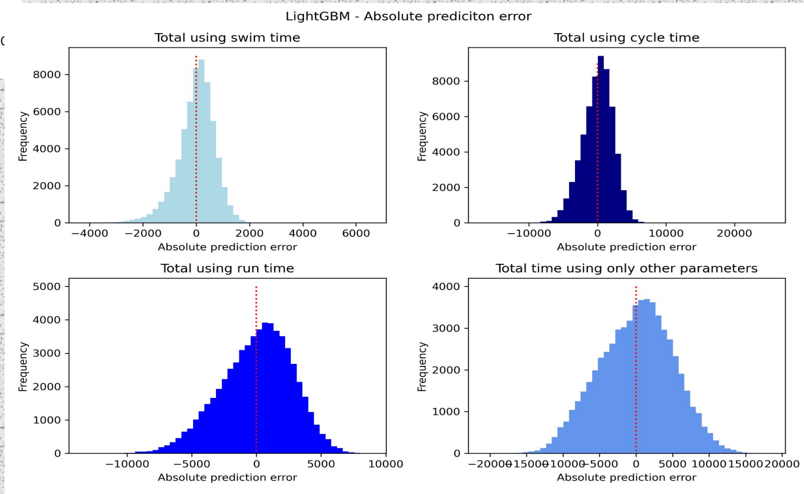
1. Swim, Startnumber, Age, Country where they're from, City (Swimming track).
   Weather is not that important.
2. Cycle, City (Cycle track), Age, Temperature and Wind are important (Headwind)
3. Run, Gender, City (Run track), Startnumber, Climate where they're from,
   Age is less important
4. Startnumber, Age, City, Gender Country where they're from
   Weather is least important

LIGHTGBM – PREDICTIONS

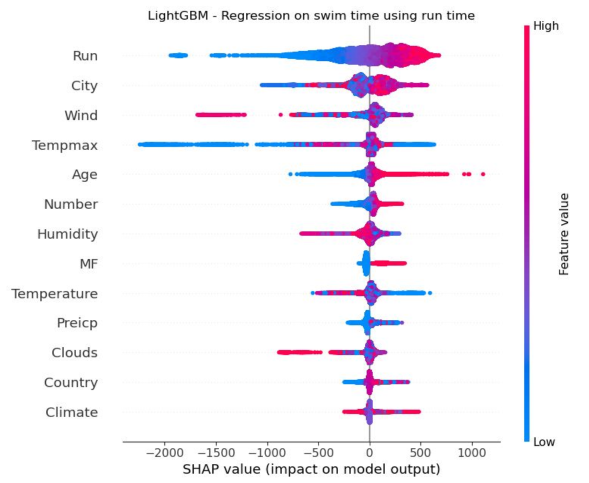
Regression using all variables and Run time



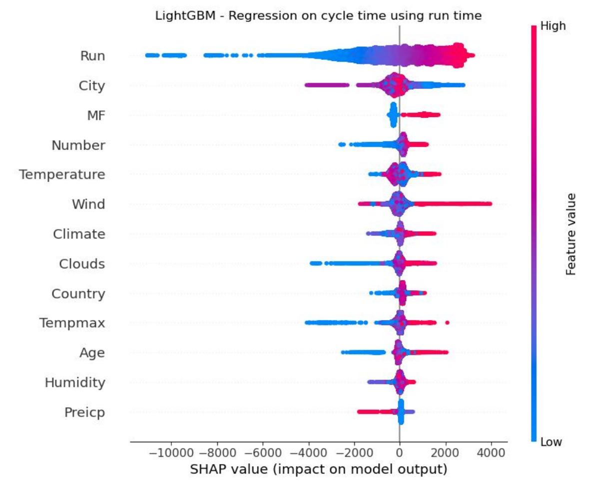
Regression (excluding times)



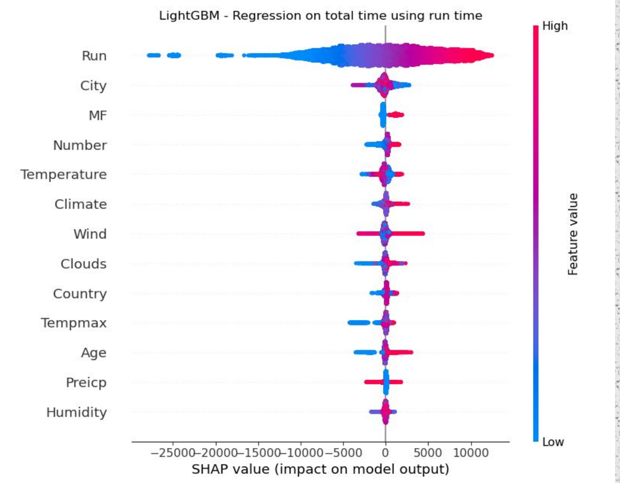
SHAP Values – Predictions



Run, City, Wind, Tempmax, Age

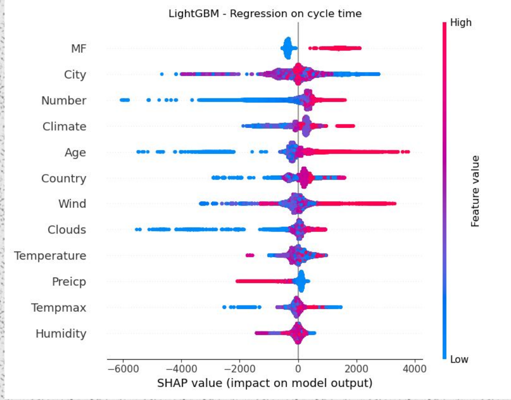


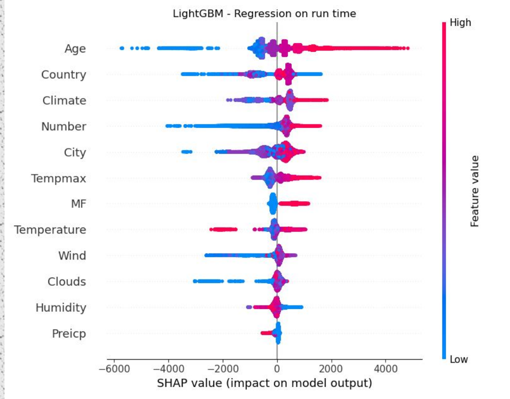
Run, City, Gender, Startnumber

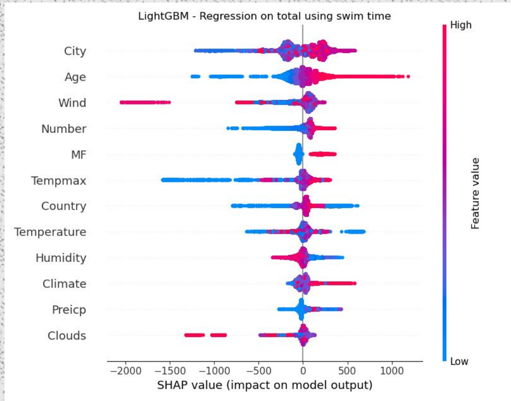


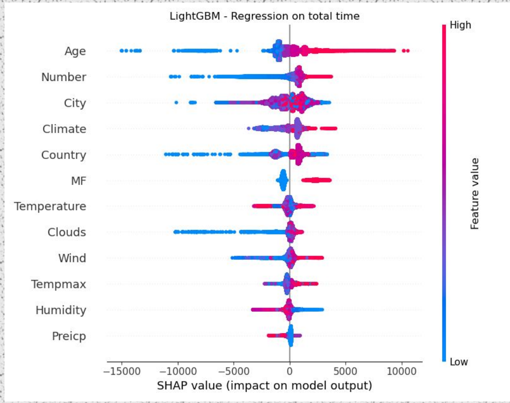
Run, City, Gender, Startnumber

SHAP Values – Predictions

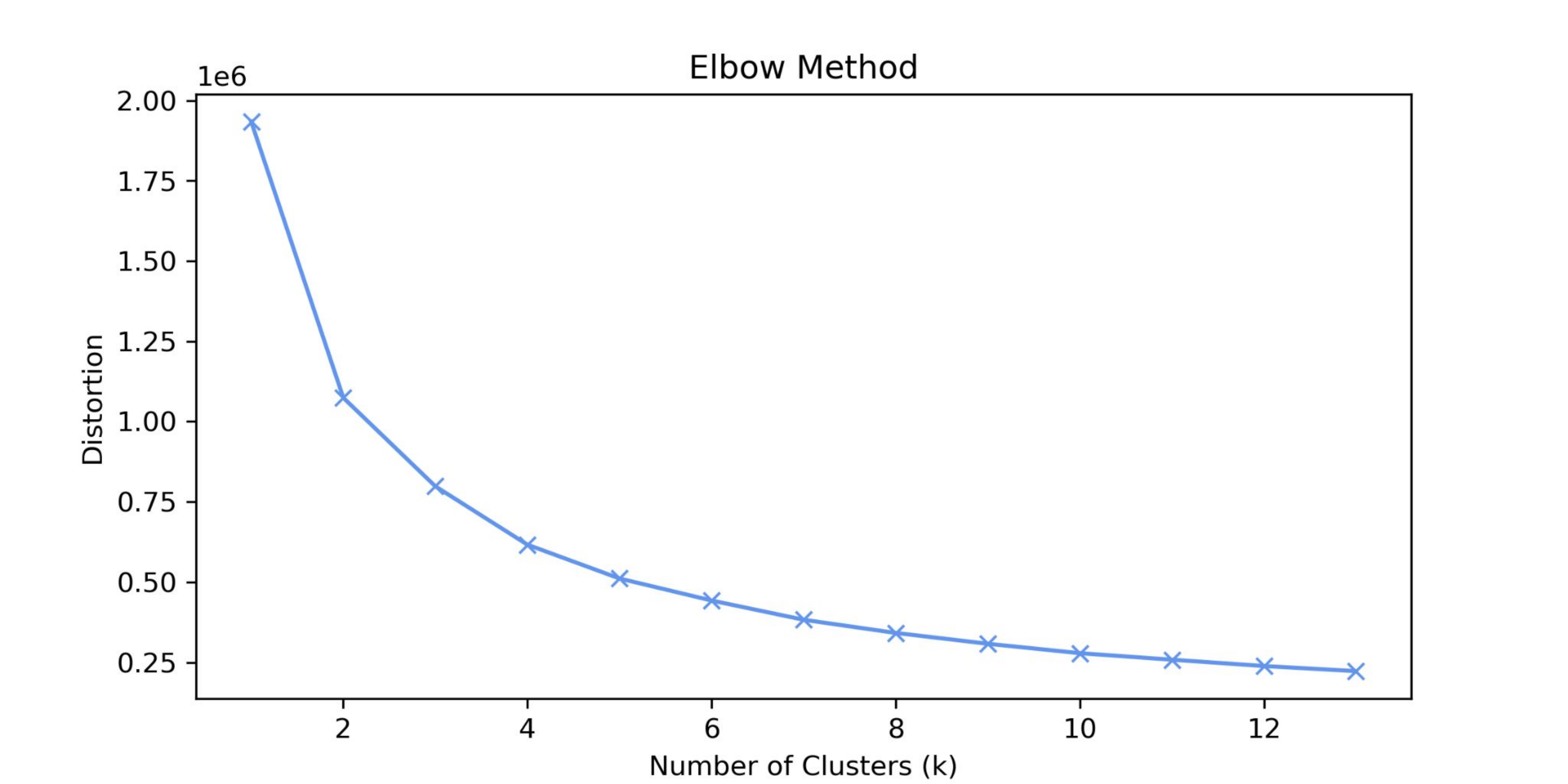






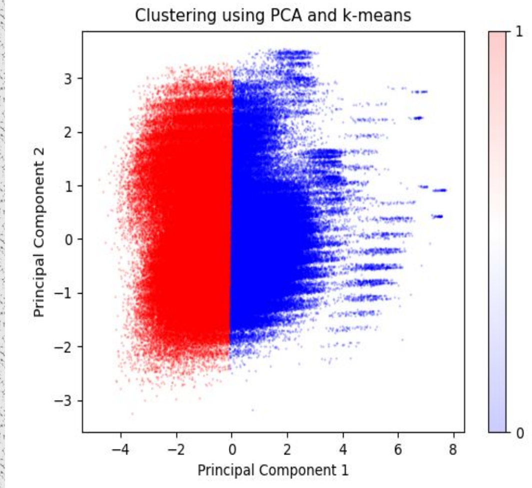


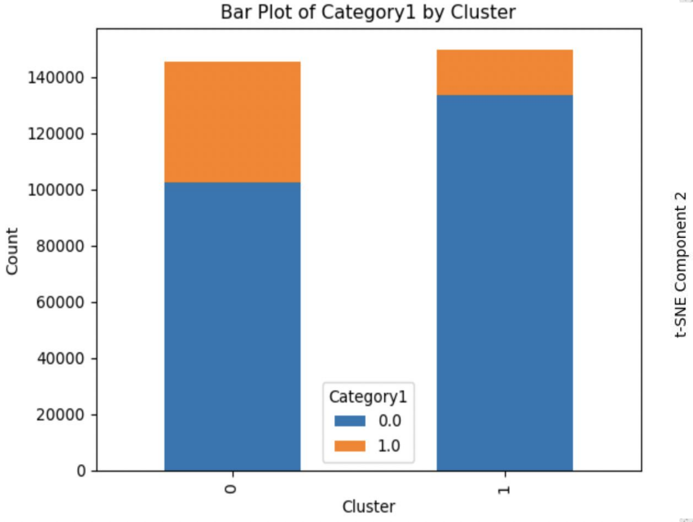
Elbow method

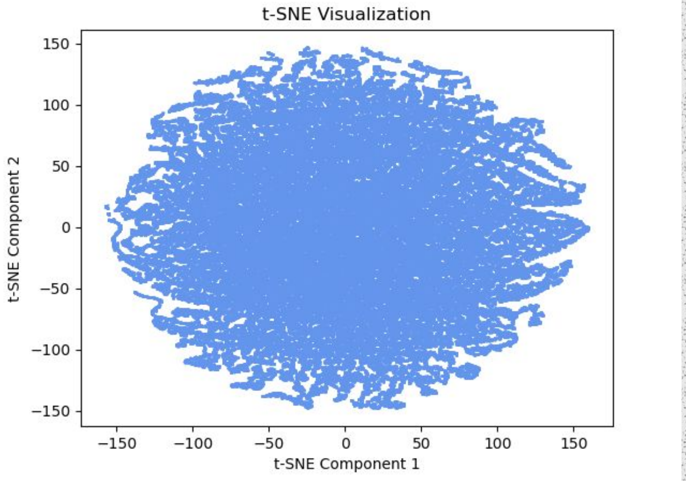


PCA&t-SNE – Clustering

k=2
With dimensionality reduction



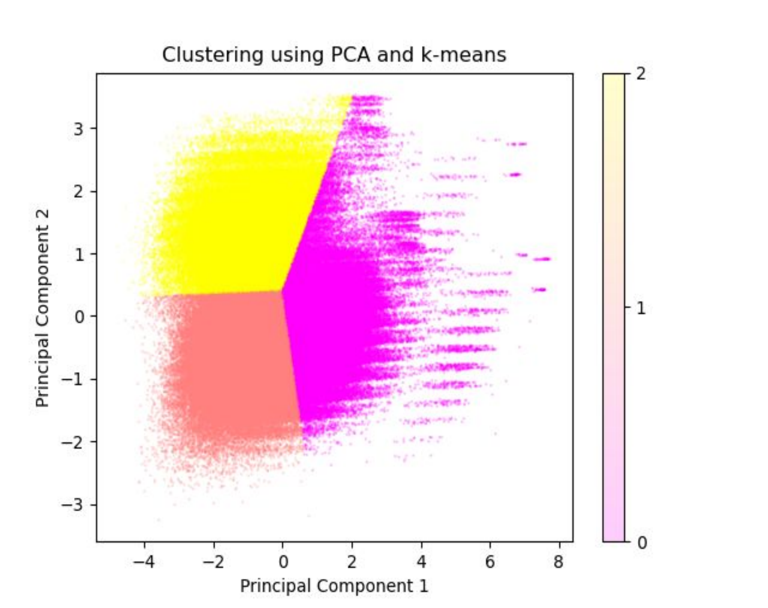


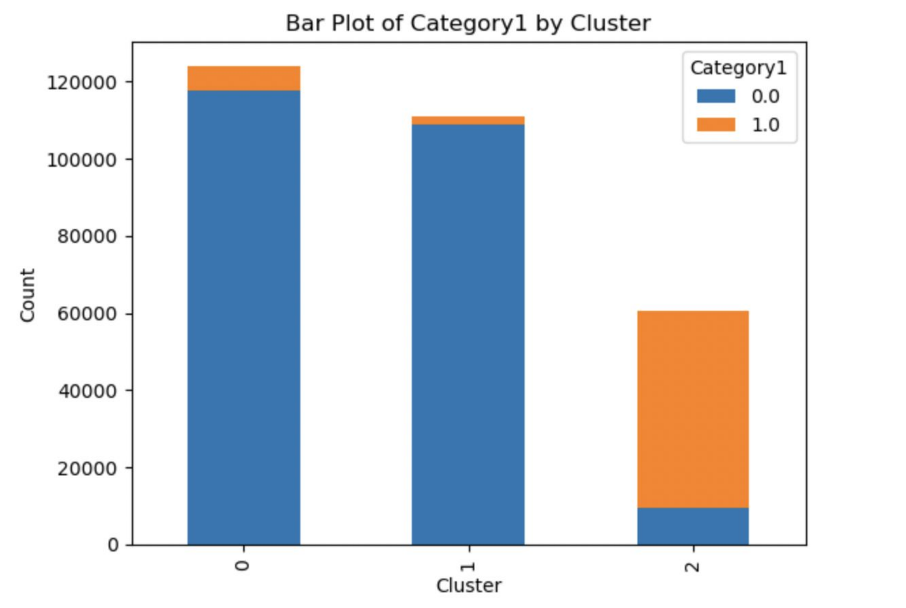


PCA – Clustering

k=3

Splits Men (0,1) and Women (2) reasonably well





RESOURCES

PAPER

- Thuany, M.; Valero, D.; Villiger, E.; Forte, P.; Weiss, K.; Nikolaidis, P.T.; Andrade, M.S.; Cuk, I.; Sousa, C.V.; Knechtle, B. A Machine Learning Approach to Finding the Fastest Race Course for Professional Athletes Competing in Ironman® 70.3 Races between 2004 and 2020. *Int. J. Environ. Res. Public Health***2023**,*20*,3619. https:// doi.org/10.3390/ijerph20043619

IMAGES

- https://stablediffusionweb.com

— All participants contributed evenly —