Flotation plant data analysis

Luís Felipe Souza Magalhães
lupi.mag2001@gmail.com

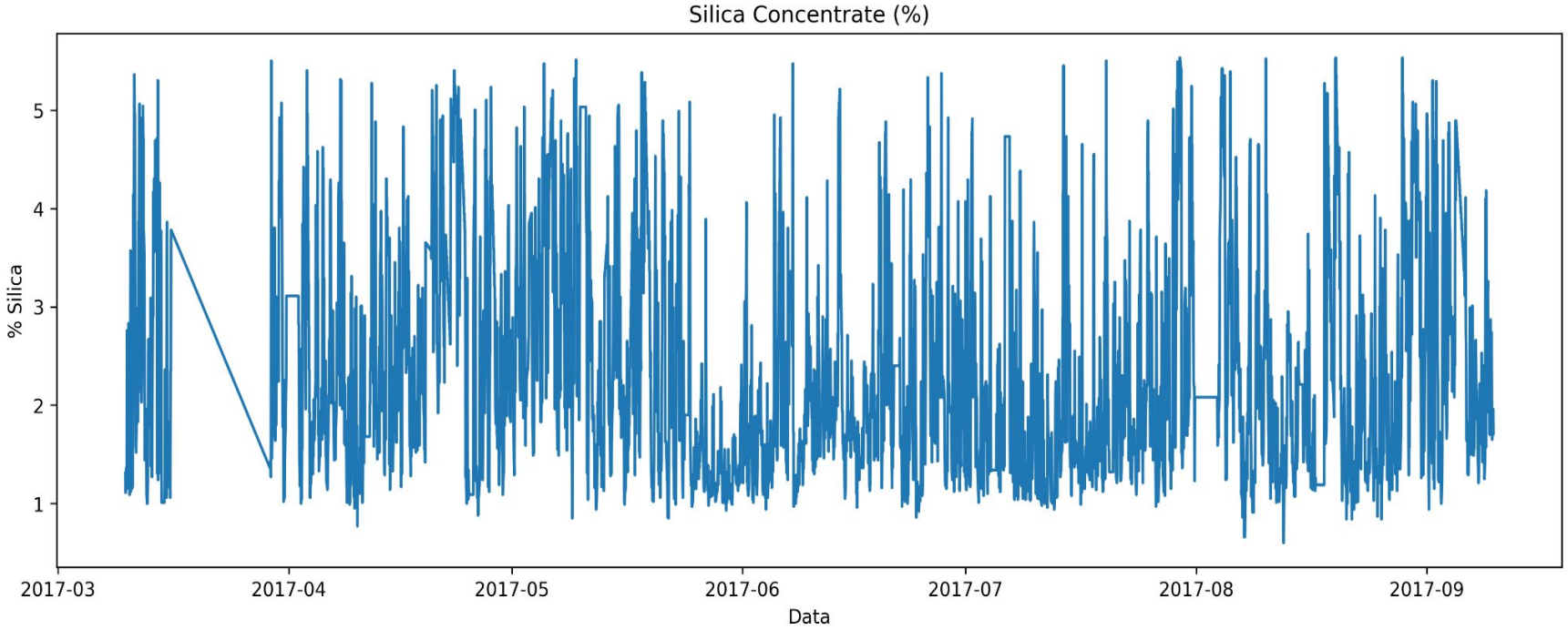
The Data set

- An overview

	date	% Iron Feed	% Silica Feed	Starch Flow	Amina Flow	Ore Pulp Flow	Ore Pulp pH	...	Flotation Column 03 Level	Flotation Column 04 Level	Flotation Column 05 Level	Flotation Column 06 Level	Flotation Column 07 Level	% Iron Concentrate	% Silica Concentrate
0	2017-03-10 01:00:00	55,2	16,98	3019,53	557,434	395,713	10,0664	...	424,954	443,558	502,255	446,37	523,344	66,91	1,31
1	2017-03-10 01:00:00	55,2	16,98	3024,41	563,965	397,383	10,0672	...	432,939	448,086	496,363	445,922	498,075	66,91	1,31
2	2017-03-10 01:00:00	55,2	16,98	3043,46	568,054	399,668	10,068	...	434,61	449,688	484,411	447,826	458,567	66,91	1,31
3	2017-03-10 01:00:00	55,2	16,98	3047,36	568,665	397,939	10,0689	...	442,865	446,21	471,411	437,69	427,669	66,91	1,31
4	2017-03-10 01:00:00	55,2	16,98	3033,69	558,167	400,254	10,0697	...	450,523	453,67	462,598	443,682	425,679	66,91	1,31
5	2017-03-10 01:00:00	55,2	16,98	3079,1	564,697	396,533	10,0705	...	460,449	439,92	451,588	433,539	425,458	66,91	1,31
6	2017-03-10 01:00:00	55,2	16,98	3127,79	566,467	392,9	10,0713	...	452,306	431,328	443,548	444,575	431,251	66,91	1,31
7	2017-03-10 01:00:00	55,2	16,98	3152,93	558,777	397,002	10,0722	...	461,64	442,067	441,73	461,77	449,679	66,91	1,31
8	2017-03-10 01:00:00	55,2	16,98	3147,27	556,03	394,307	10,073	...	459,103	455,074	439,798	457,738	455,915	66,91	1,31
9	2017-03-10 01:00:00	55,2	16,98	3142,58	565,857	393,105	10,0738	...	460,665	457,225	453,236	449,898	455,75	66,91	1,31

Data preparation

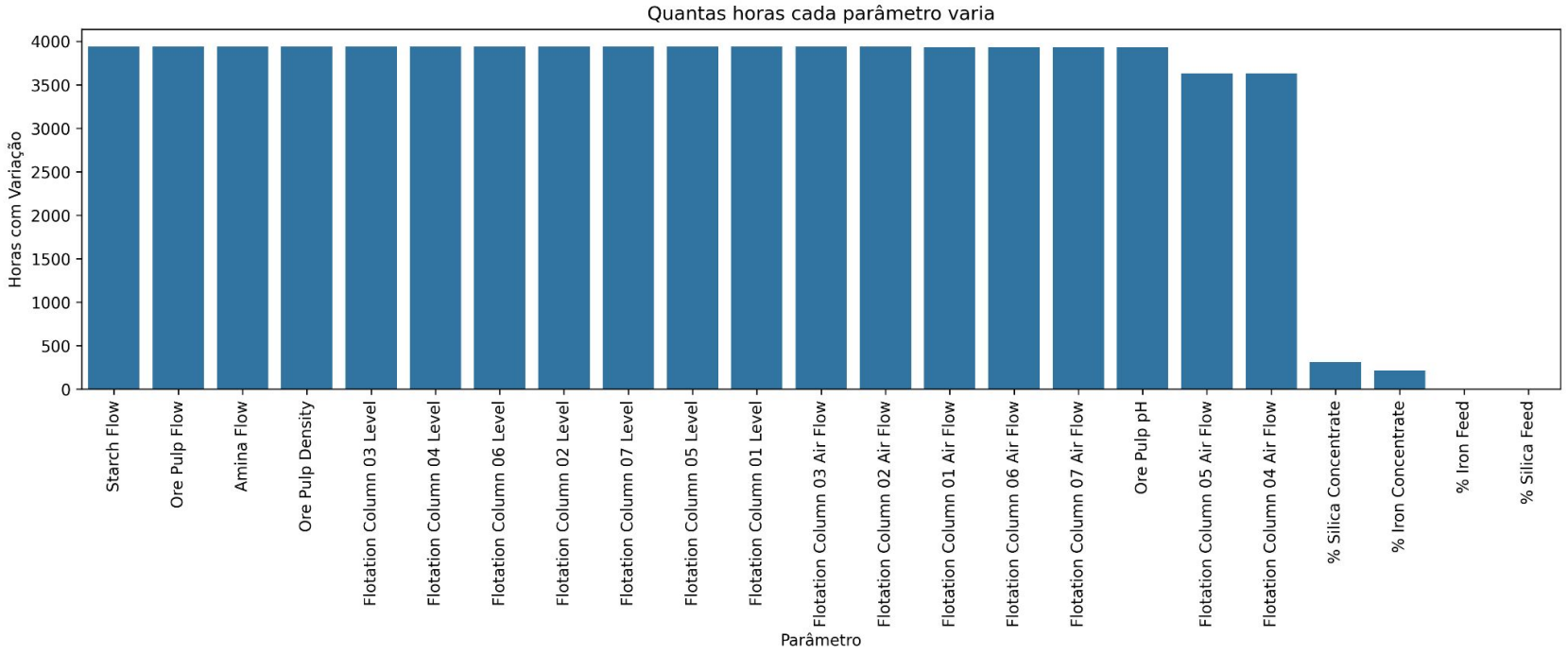
- Plotting the target variable



There is a gap !

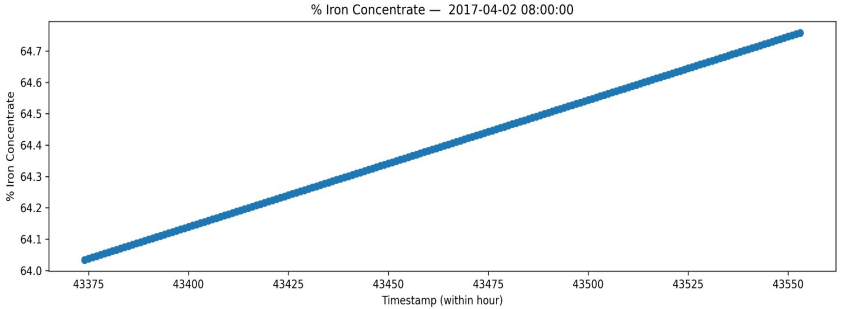
	date	% Iron Feed	% Silica Feed	Starch Flow	Amina Flow	Ore Pulp Flow	Ore Pulp pH	...	Flotation Column 03 Level	Flotation Column 04 Level	Flotation Column 05 Level	Flotation Column 06 Level	Flotation Column 07 Level	% Iron Concentrate	% Silica Concentrate
26809	2017-03-16 05:00:00	57,45	12,95	3066,8	465,393	401,016	9,5156	...	555,06	516,39	498,014	509,734	522,018	64,03	3,78
26810	2017-03-16 05:00:00	57,45	12,95	3540,63	466,156	400,151	9,52077	...	561,454	525,862	499,143	499,934	527,975	64,03	3,78
26811	2017-03-16 05:00:00	57,45	12,95	3593,75	472,565	401,06	9,52595	...	566,984	521,038	498,069	492,85	520,885	64,03	3,78
26812	2017-03-16 05:00:00	57,45	12,95	3612,11	467,926	400,532	9,53113	...	571,083	511,466	493,886	493,102	511,233	64,03	3,78
26813	2017-03-16 05:00:00	57,45	12,95	3634,38	465,698	399,683	9,5363	...	570,412	514,396	494,988	498,45	512,914	64,03	3,78
26814	2017-03-29 12:00:00	60,18	9,34	2044,99	349,911	405,625	9,74604	...	614,9847142857	428,43	413,662	447,887	464,869	66,44	1,36
26815	2017-03-29 12:00:00	60,18	9,34	2020,55	353,064	404,402	9,74169	...	659,5574285714	431,146	422,308	444,326	463,224	66,44	1,36
26816	2017-03-29 12:00:00	60,18	9,34	2012,84	356,217	403,18	9,73733	...	704,1301428571	433,863	430,954	440,765	461,579	66,44	1,36
26817	2017-03-29 12:00:00	60,18	9,34	2007,67	359,371	401,957	9,73298	...	748,7028571429	436,58	439,6	437,204	459,935	66,44	1,36
26818	2017-03-29 12:00:00	60,18	9,34	2003,92	362,524	400,735	9,72862	...	793,2755714286	439,296	448,246	433,643	458,29	66,44	1,36

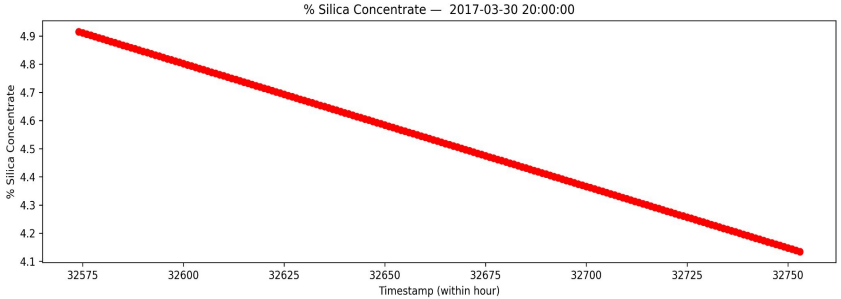
Hourly VS 20 seconds columns



% Silica\Iron concentrate variation

- What happens in the hours these variables change?

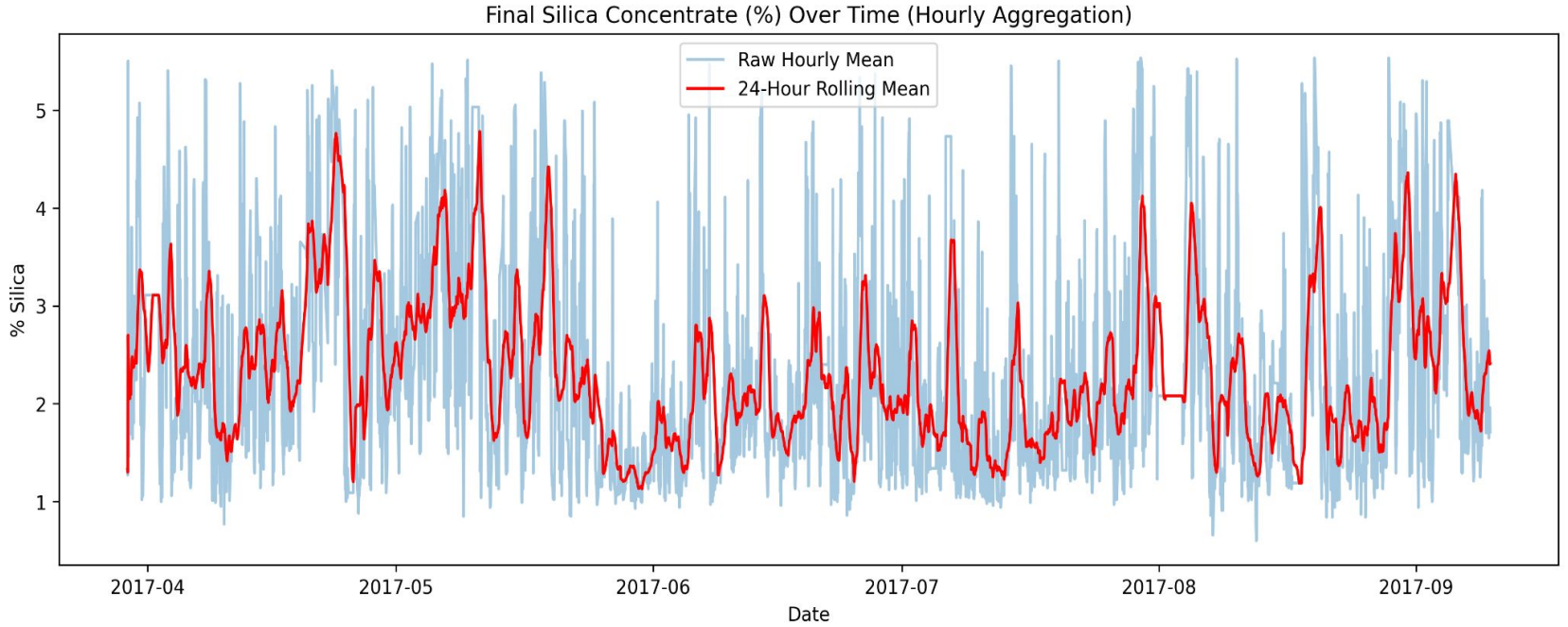




Hourly Dataset

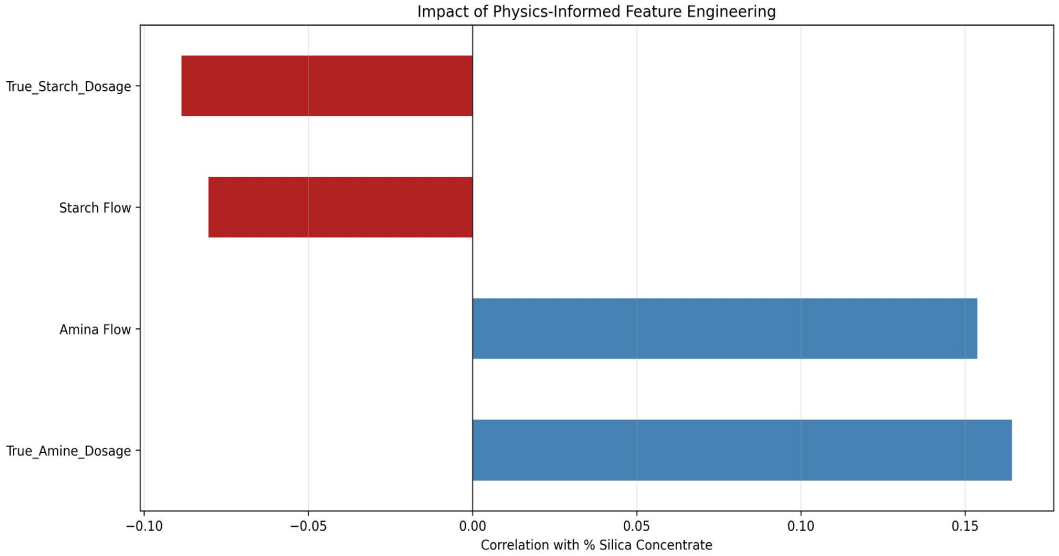
	% Iron Feed	Starch Flow	Amina Flow	Ore Pulp Flow	Ore Pulp pH	Ore Pulp Density	Flotation Column 01 Air Flow	...	Flotation Column 03 Level	Flotation Column 04 Level	Flotation Column 05 Level	Flotation Column 06 Level	Flotation Column 07 Level	% Silica Concentrate	% Silica Concentrate Delta
date															
2017-03-29 12:00:00	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	1.36	NaN
2017-03-29 13:00:00	60.18	1060.859297	379.836203	400.982539	9.527187	1.550374	200.035839	...	860.557818	477.349197	452.512522	478.219647	470.105547	1.43	NaN
2017-03-29 14:00:00	60.18	2034.926525	322.233975	400.468947	9.700679	1.532017	199.990122	...	869.419421	483.816131	469.448172	471.607075	462.665478	1.33	0.07
2017-03-29 15:00:00	60.18	1435.425811	474.656322	399.157083	9.687788	1.645200	200.019311	...	879.704989	456.147231	453.227000	447.708467	453.475667	1.27	-0.10
2017-03-29 16:00:00	60.18	618.460387	396.377733	398.942522	9.851689	1.559929	199.940833	...	882.093896	449.410744	448.581506	450.697578	448.660094	5.50	-0.06
2017-03-29 17:00:00	59.54	1367.502858	317.134072	400.836283	9.936163	1.530984	199.878922	...	884.840698	450.123050	451.774694	451.574422	449.406894	1.64	4.23
2017-03-29 18:00:00	59.54	1729.412710	410.978717	400.036094	9.477018	1.567739	199.906183	...	869.178824	431.977917	452.877969	462.790350	449.586028	1.46	-3.86
2017-03-29 19:00:00	59.54	1639.599185	460.216806	399.255361	9.604897	1.626786	199.901339	...	883.408518	336.023814	449.695350	452.206889	449.973436	1.69	-0.18
2017-03-29 20:00:00	59.54	1913.013778	495.235056	399.948317	9.558475	1.673500	199.866906	...	882.990989	559.462817	556.967506	518.052867	523.231725	1.81	0.23
2017-03-29 21:00:00	59.54	2033.815739	500.493172	400.512483	9.279860	1.682864	199.898189	...	883.013811	589.507347	601.724111	549.201878	548.972244	1.71	0.12

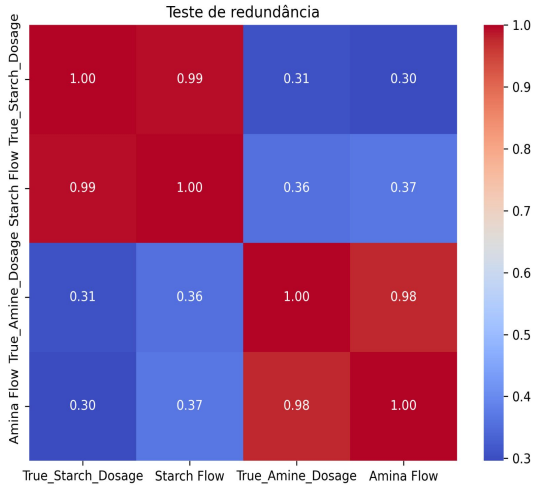
New Plot


Final Silica Concentrate (%) Over Time (Hourly Aggregation)
Raw Hourly Mean
24-Hour Rolling Mean
% Silica
1
2
3
4
5
2017-04
2017-05
2017-06
2017-07
2017-08
2017-09
Date

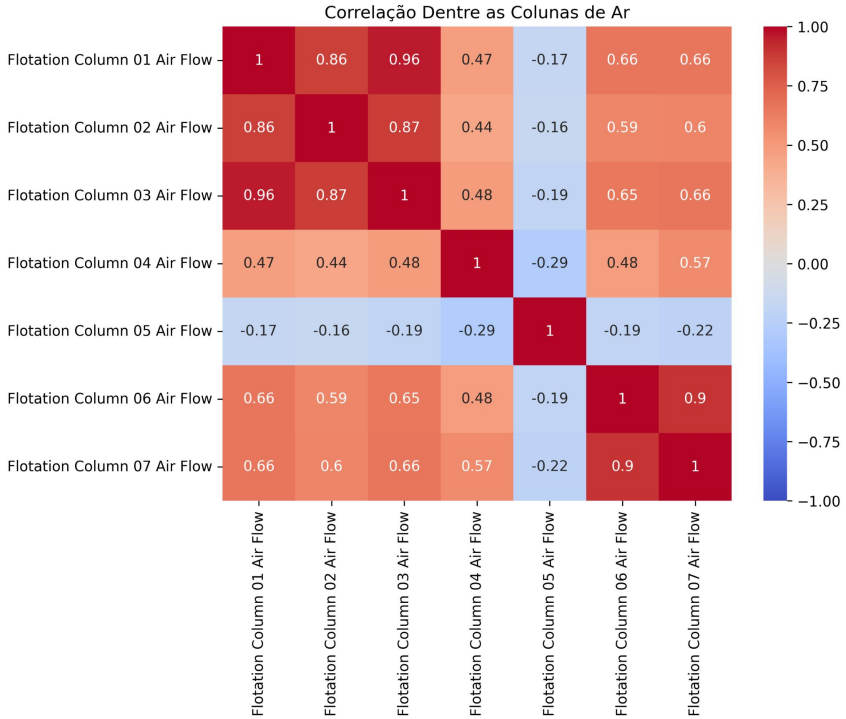
Features

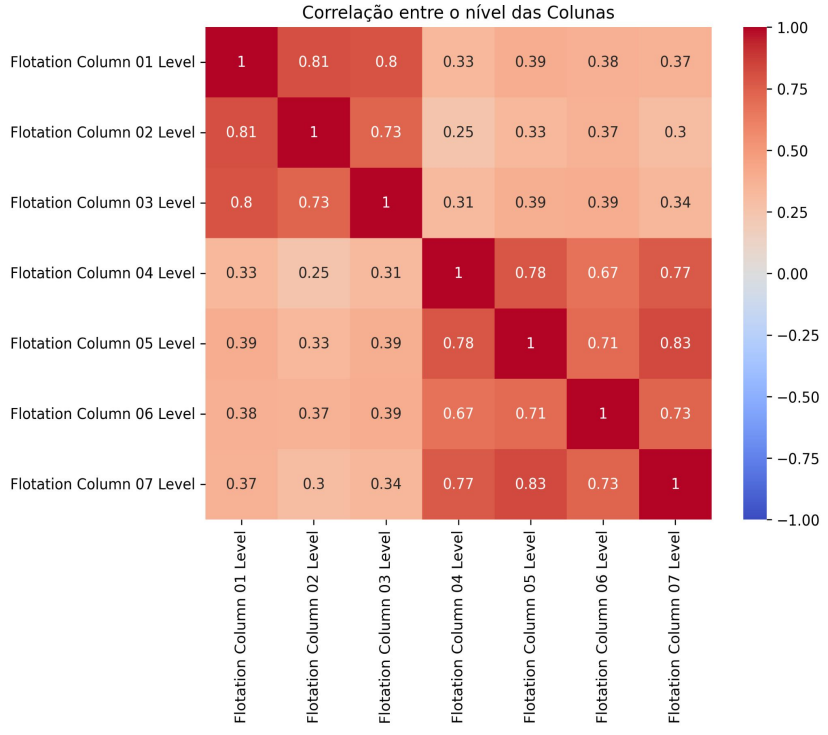
$$\text{Dosage} = \frac{\text{Reagent Flow}}{\text{Pulp Flow} \times \text{Pulp Density}}$$



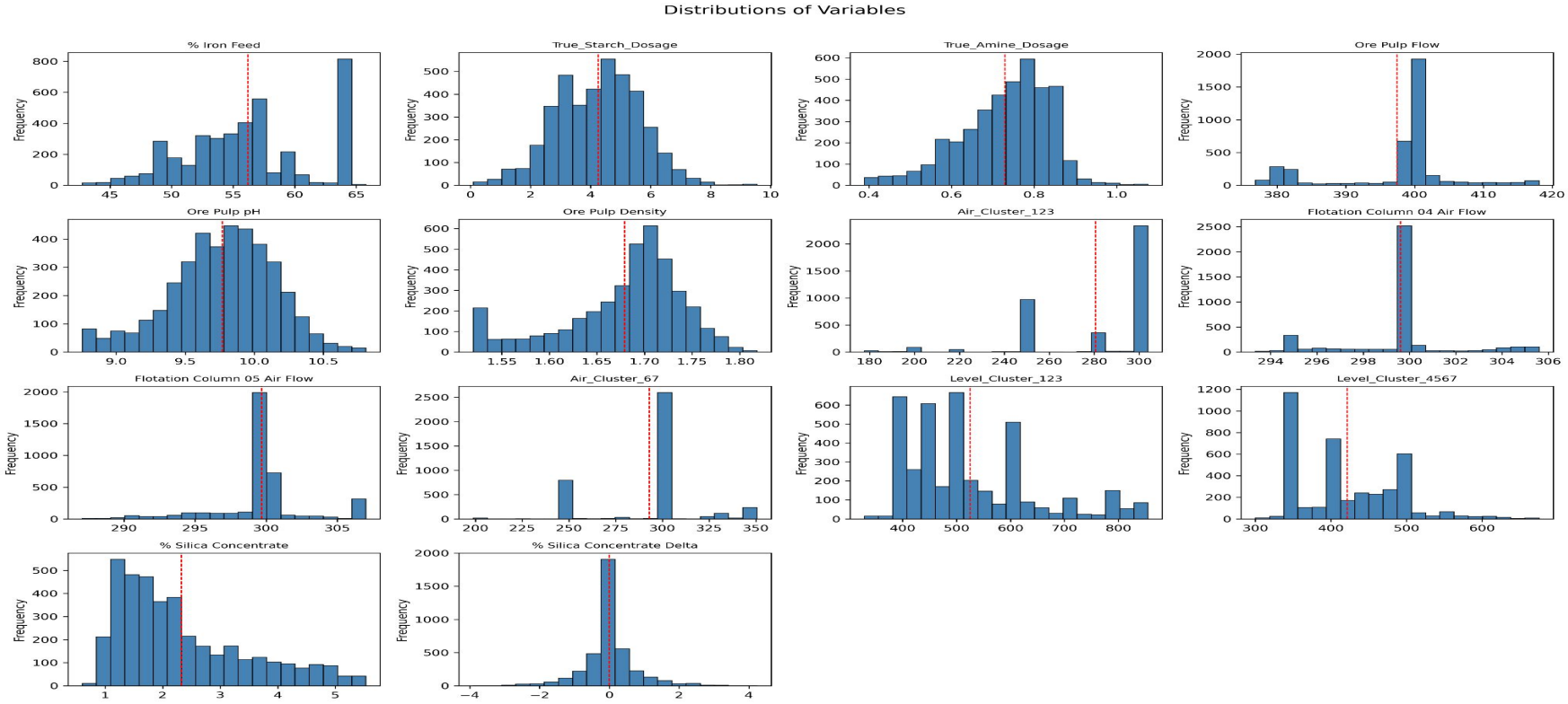


Air\Column level



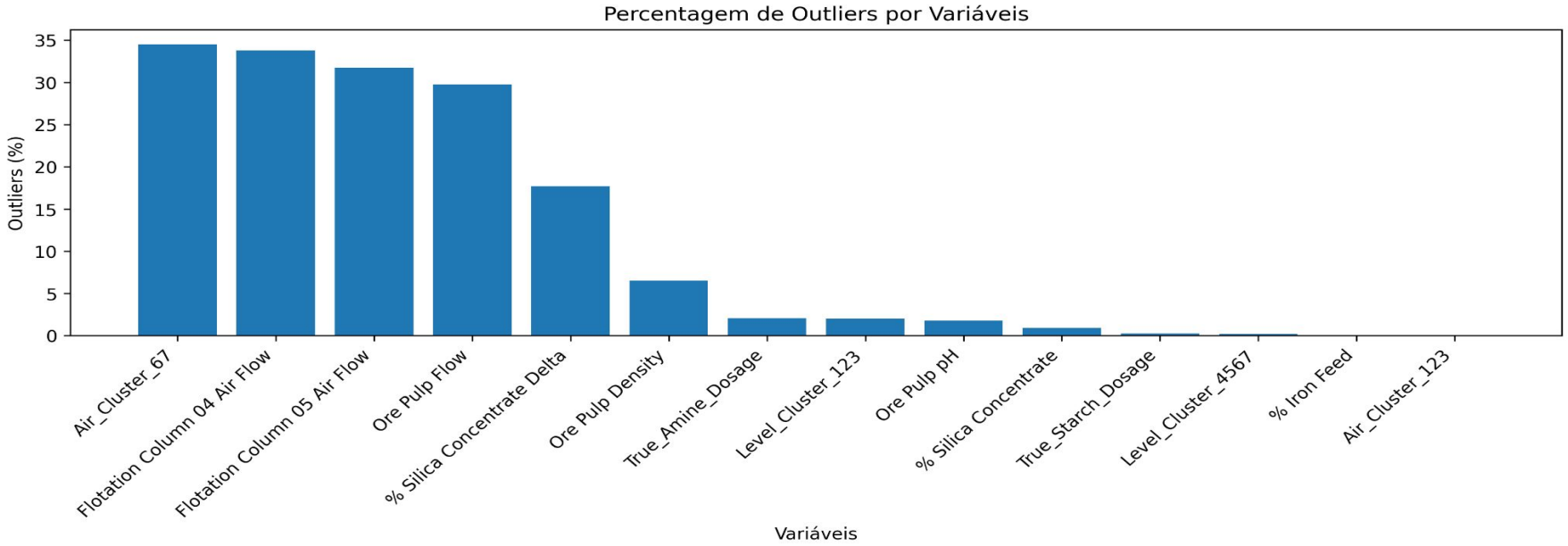


Distributions



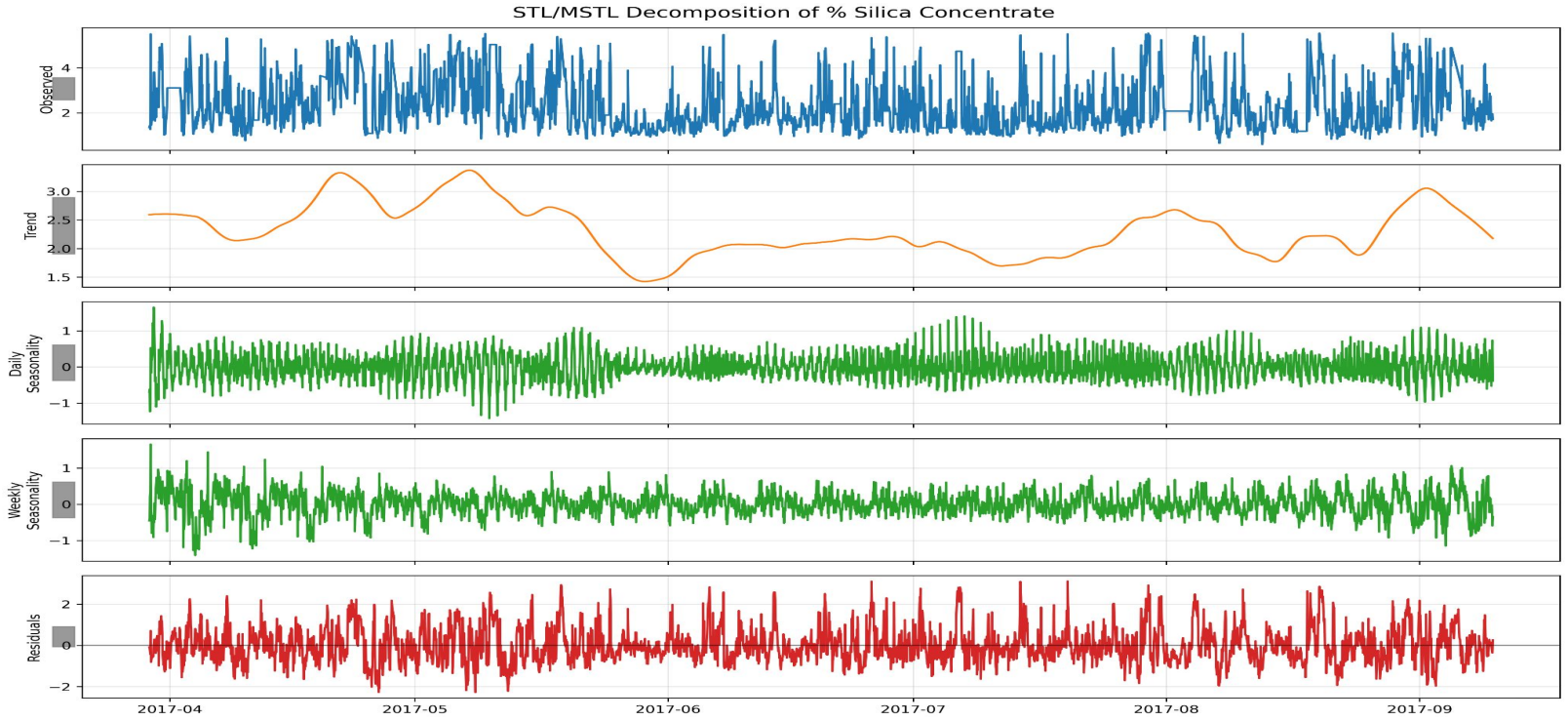
Outliers

$$IQR = Q_3 - Q_1$$
$$LF = Q_1 - (1.5 \times IQR)$$
$$UF = Q_3 + (1.5 \times IQR)$$



Trend \ Seasonality

$$y_t = T_t + S_t + R_t$$



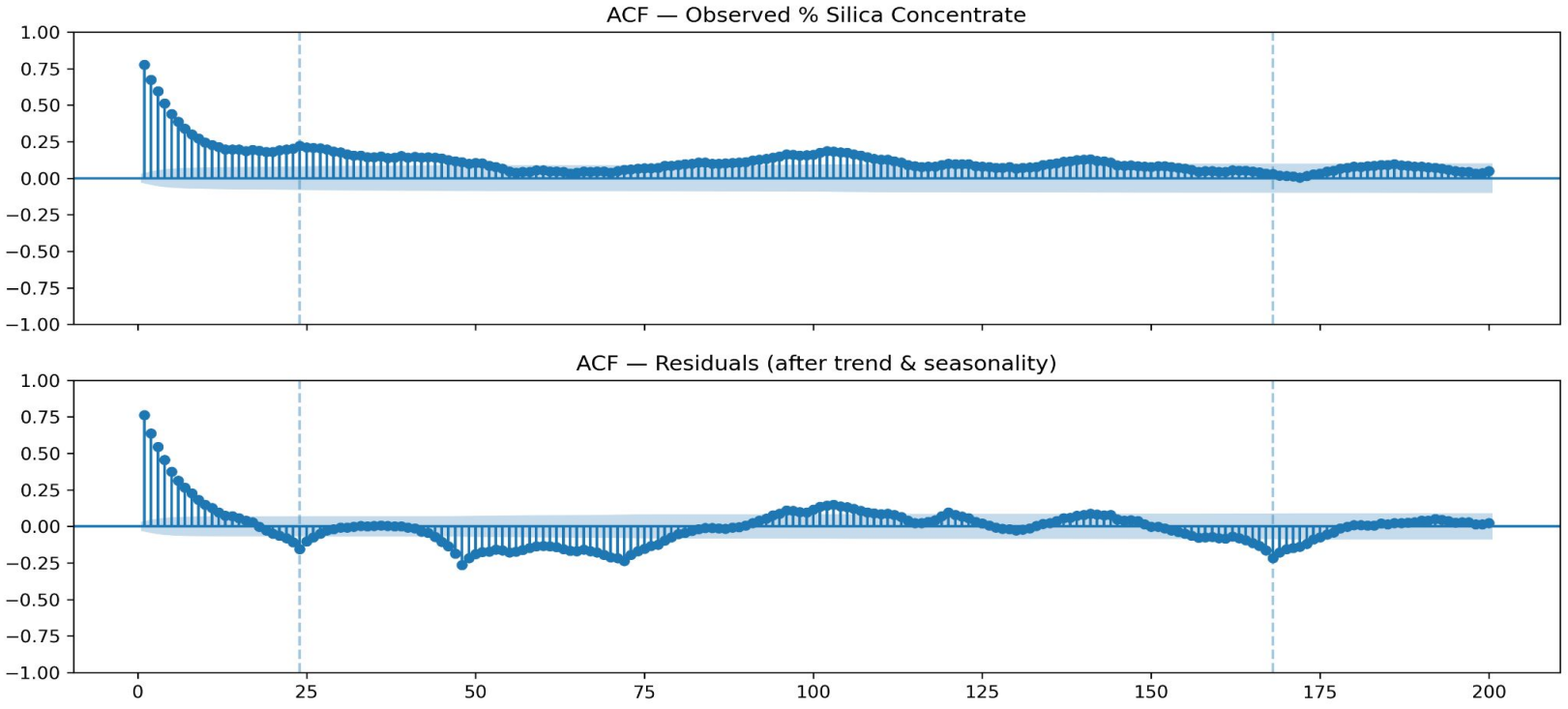
Trend/Seasonality strength

$$F_T = \max\left(0, 1 - \frac{\mathrm{Var}(R_t)}{\mathrm{Var}(T_t + R_t)}\right) = 0.244$$

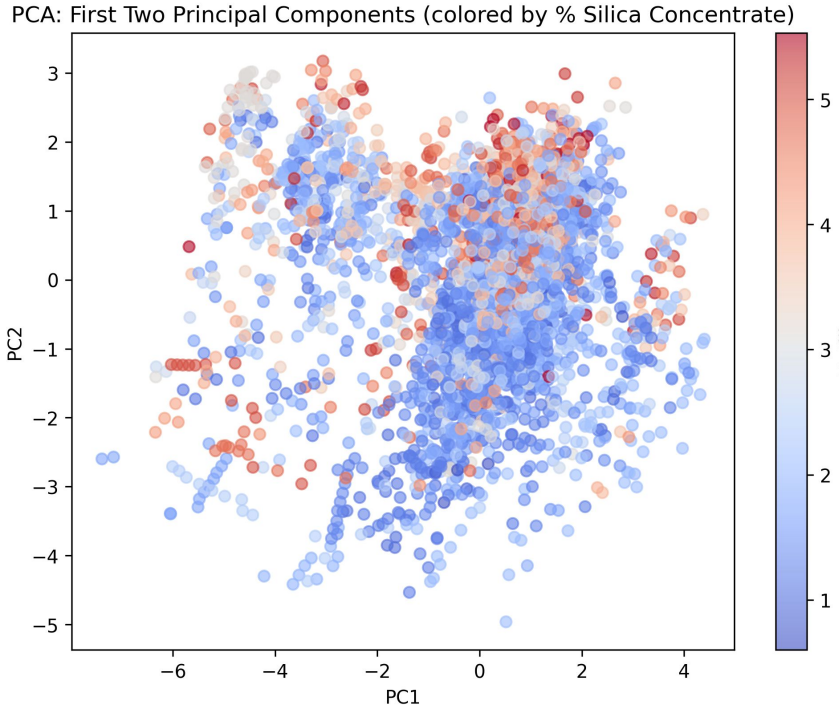
$$F_{Sweekly} = \max\left(0, 1 - \frac{\mathrm{Var}(R_t)}{\mathrm{Var}(S_t + R_t)}\right) = 0.159$$

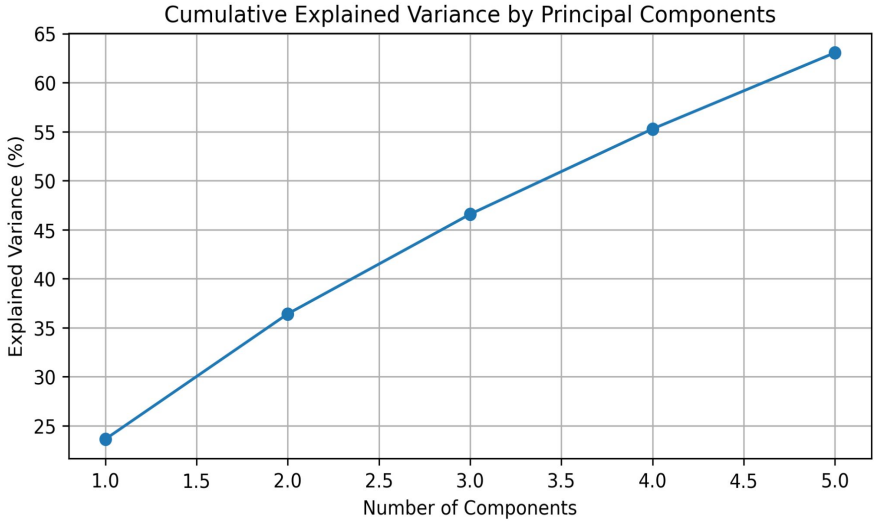
$$F_{Sdaily} = \max\left(0, 1 - \frac{\mathrm{Var}(R_t)}{\mathrm{Var}(S_t + R_t)}\right) = 0.181$$

Autocorrelation function



PCA


PCA: First Two Principal Components (colored by % Silica Concentrate)
PC2
PC1
% Silica


Cumulative Explained Variance by Principal Components
Explained Variance (%)
Number of Components