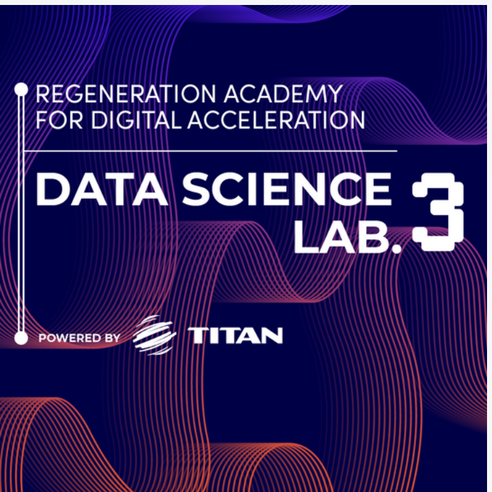
Code.Hub

REGENERATION®

TITAN

TEAM 6

FINAL PROJECT



CONTENT

PREPROCESSING

01

FIX DATES

Different date formats

2019-05-10 09:30:30.000
12/5/2019 6:02

Transformed dates
Dropped duplicates

02

ERROR VALUES FLOAT

We found 9528 ERROR values in the first rows of May.csv

We took action to drop them

03

NAN VALUES

We dropped rows that had more than 25% missing values

We took action to drop them

04

OUTLIERS

Handled using IQR method

Search and rename columns with typos

05

CONCAT FILES

Merge files on common columns

Search and rename columns with typos

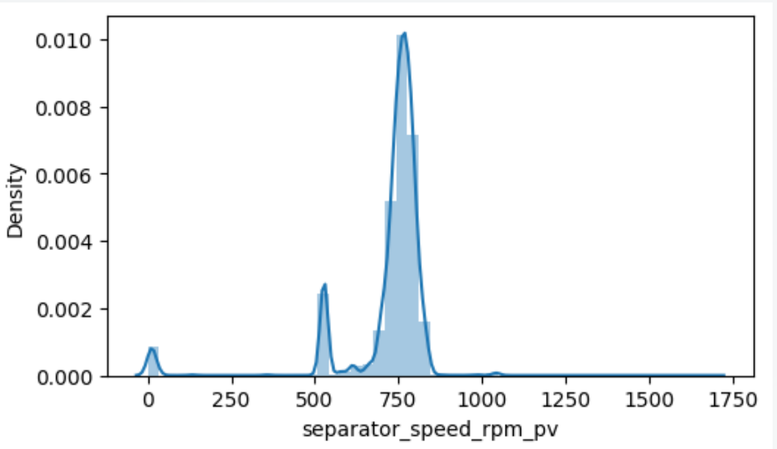
COMPARING CLEAN DATA

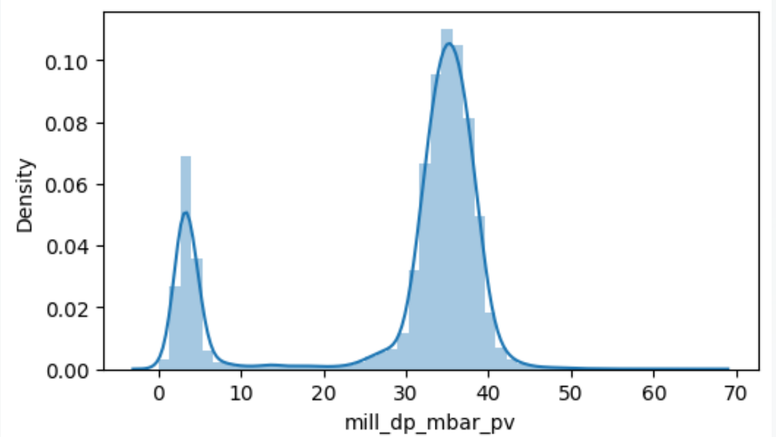
	INITIAL DATA	CLEAN DATA
May	66711	57017
June	77149	76312
July	73170	61224
August	85904	85700
September	77959	77948
October	83405	83404
Total Rows	**464277**	**441605**

MAIN STATISTICAL CHARACTERISTICS

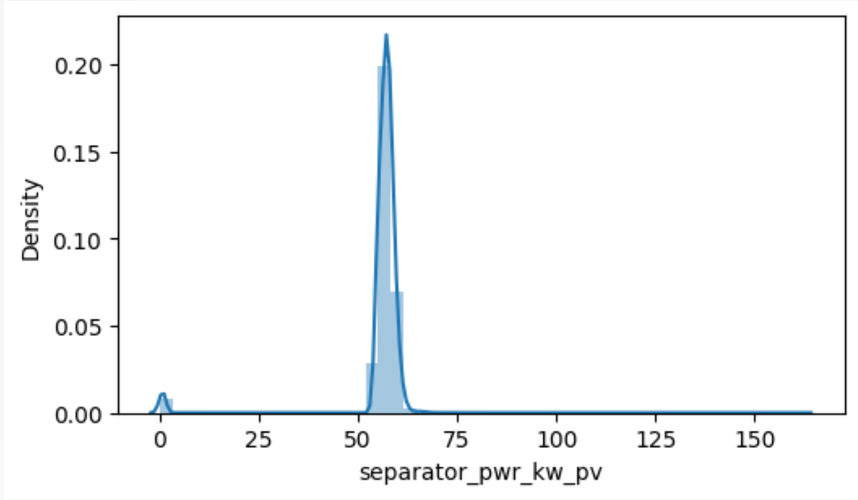
The resulting plot shows the distribution of the data, with the histogram providing a visual representation of the frequency of values in different bins, and the KDE plot providing an estimate of the data's underlying probability density function

Input

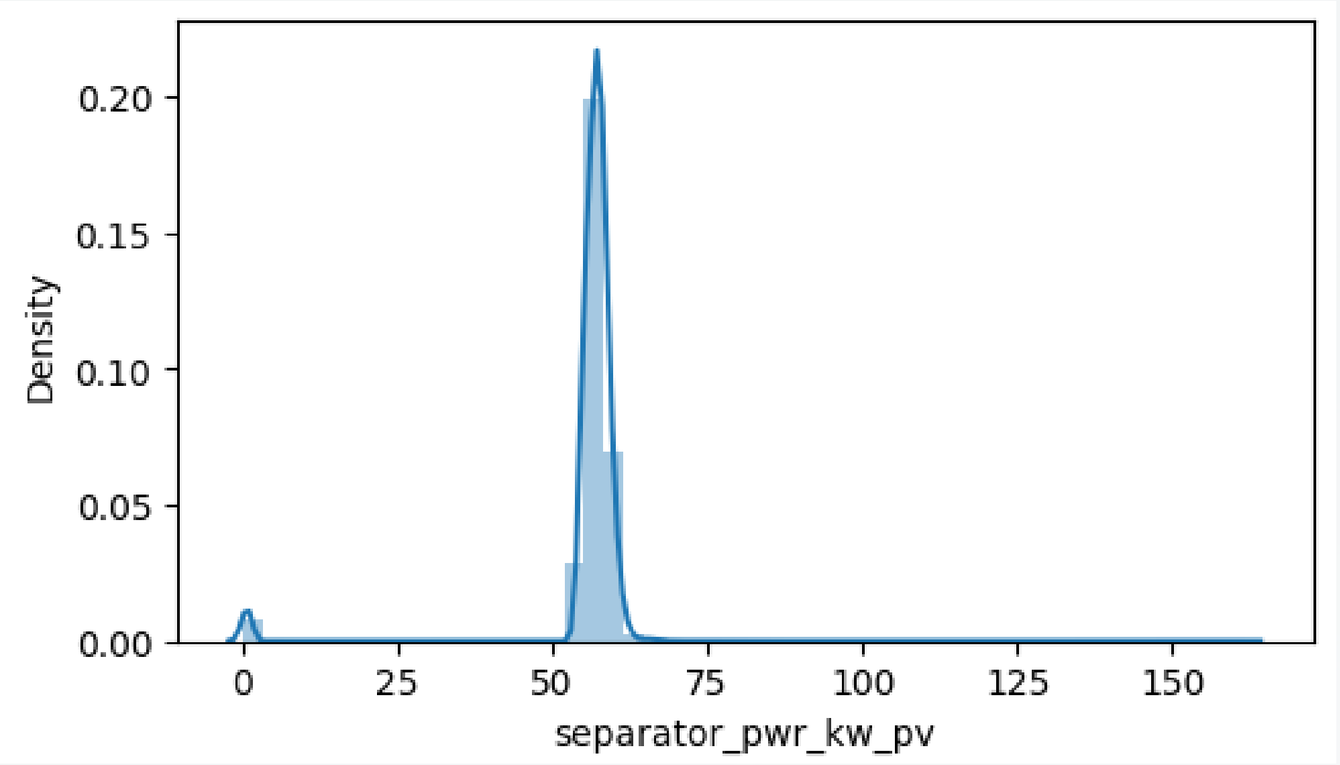




Output



What conclusions can we draw ?



	separator_pwr_kw_pv
count	302433.000000
mean	55.873773
std	9.273690
min	0.000000
25%	55.950000
50%	57.150000
75%	58.240000
max	162.000000

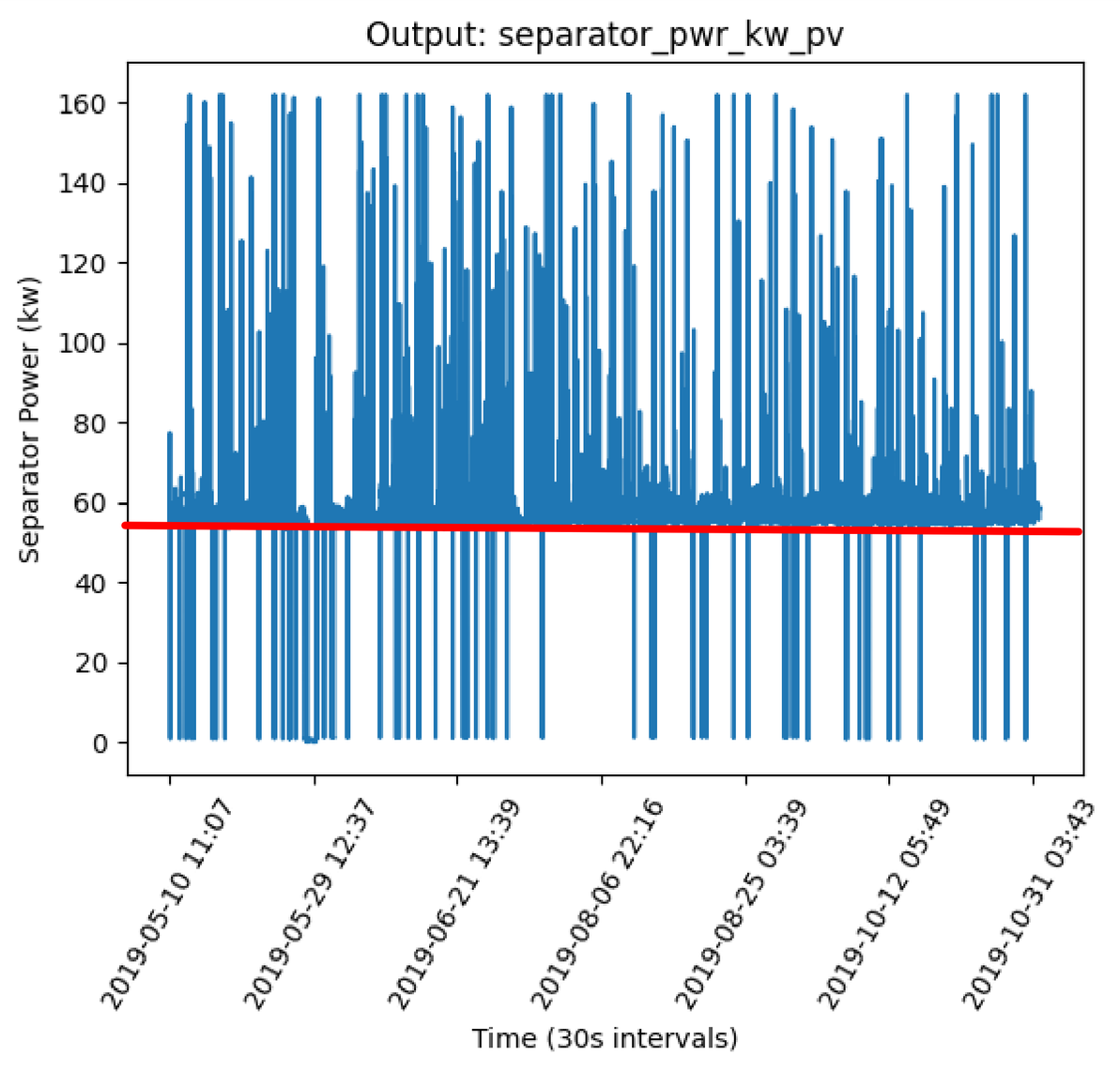
The separator power values within the range of 0 to 162, with a mean value of 55.87

1st Conclusion

The highest peaks in the density within the specific range of 50 to 60 may indicate a normal operation of the mill

2nd Conclusion

After defining normal operation we still have zero values

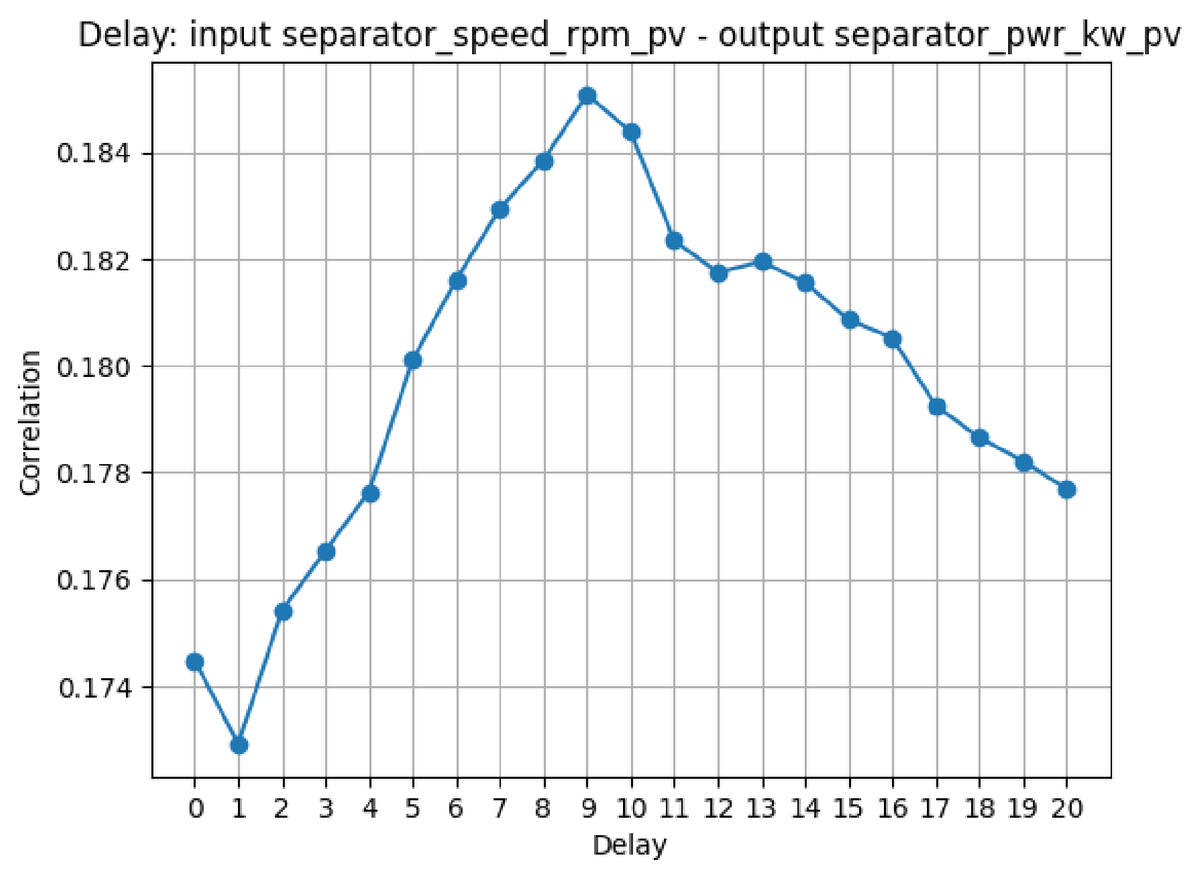


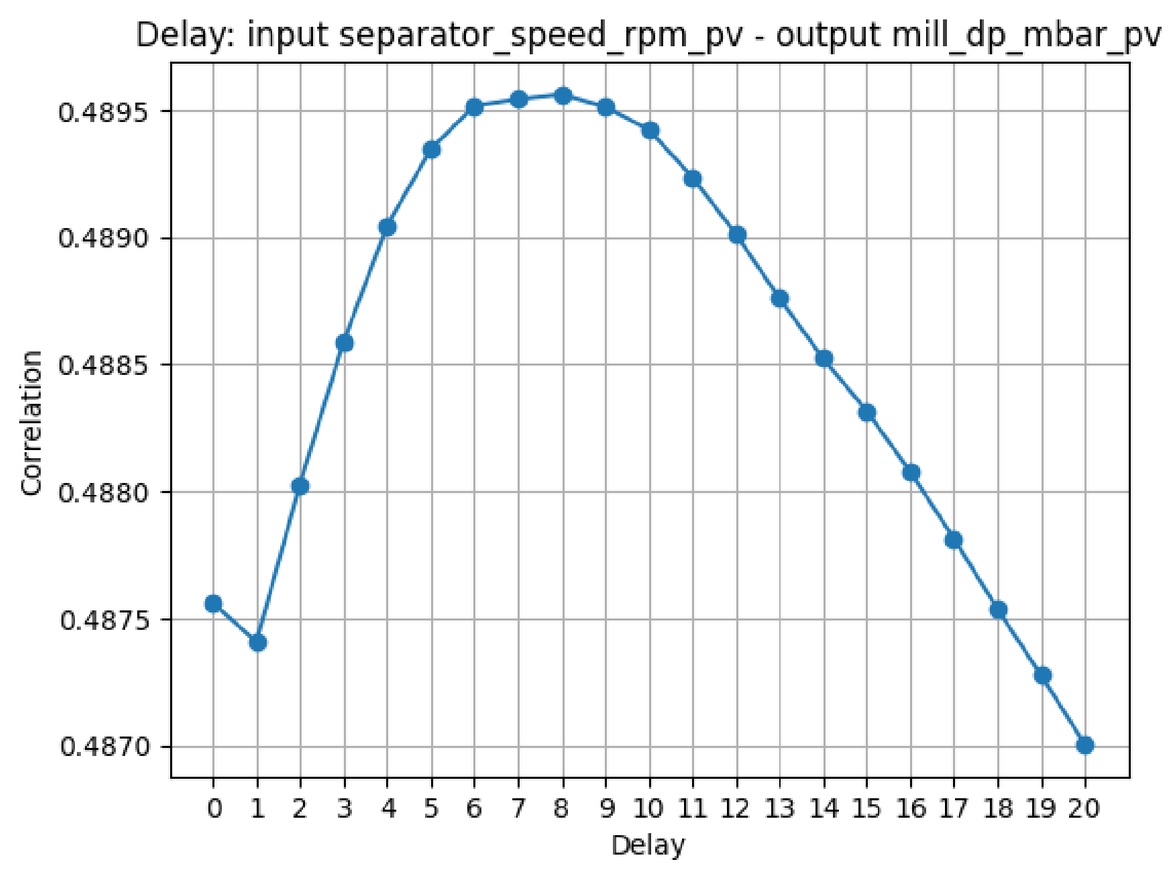
We applied a **limit** to filter output values that are close to zero (no operation of the seperator)

We implemented a new **function** that returns if a time period is **continuous or not**

DELAY ANALYSIS

Separator Speed delay analysis with Separator Speed and Mill DP





DELAY ANALYSIS

We created a function with two inputs and a max delay parameter. Returns the correlations for each

INPUTS	Delay
mill_dp_mbar_pv	4.5m
separator_speed_rpm_pv	2m

BLOCK 1

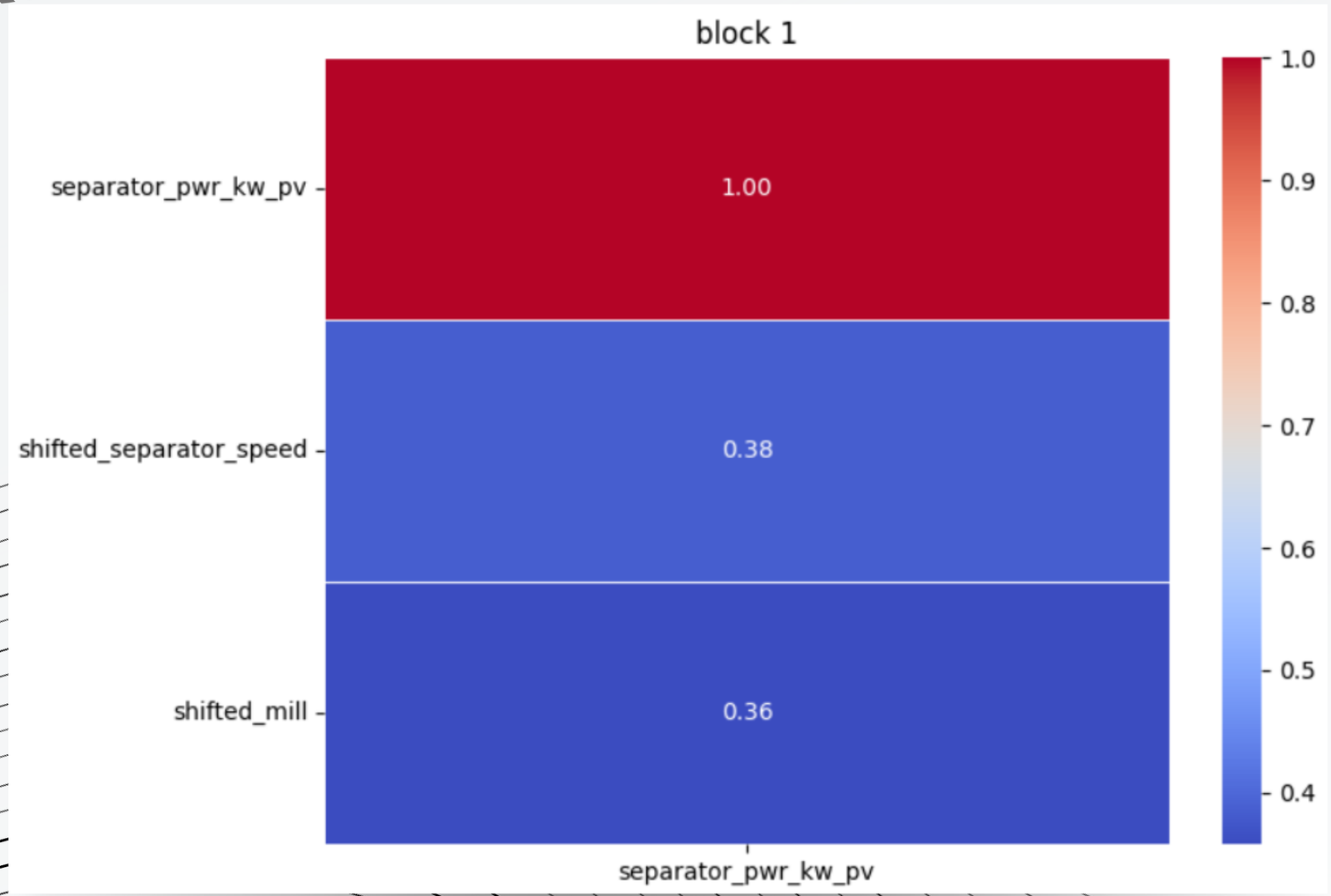
INPUTS	Delay
separator_speed_rpm_pv	4m
mill_injection_water_m3/h_pv	
fly_ash_1_ton/h_pv	
fly_ash_2_ton/h_pv	
grinding_pressure_bar_pv	
mill_out_pres_mbar_pv	

BLOCK 2

We used shifted each input up in the DataFrame to prepare the data for the model.
We ignored 0m delays, due to time restriction

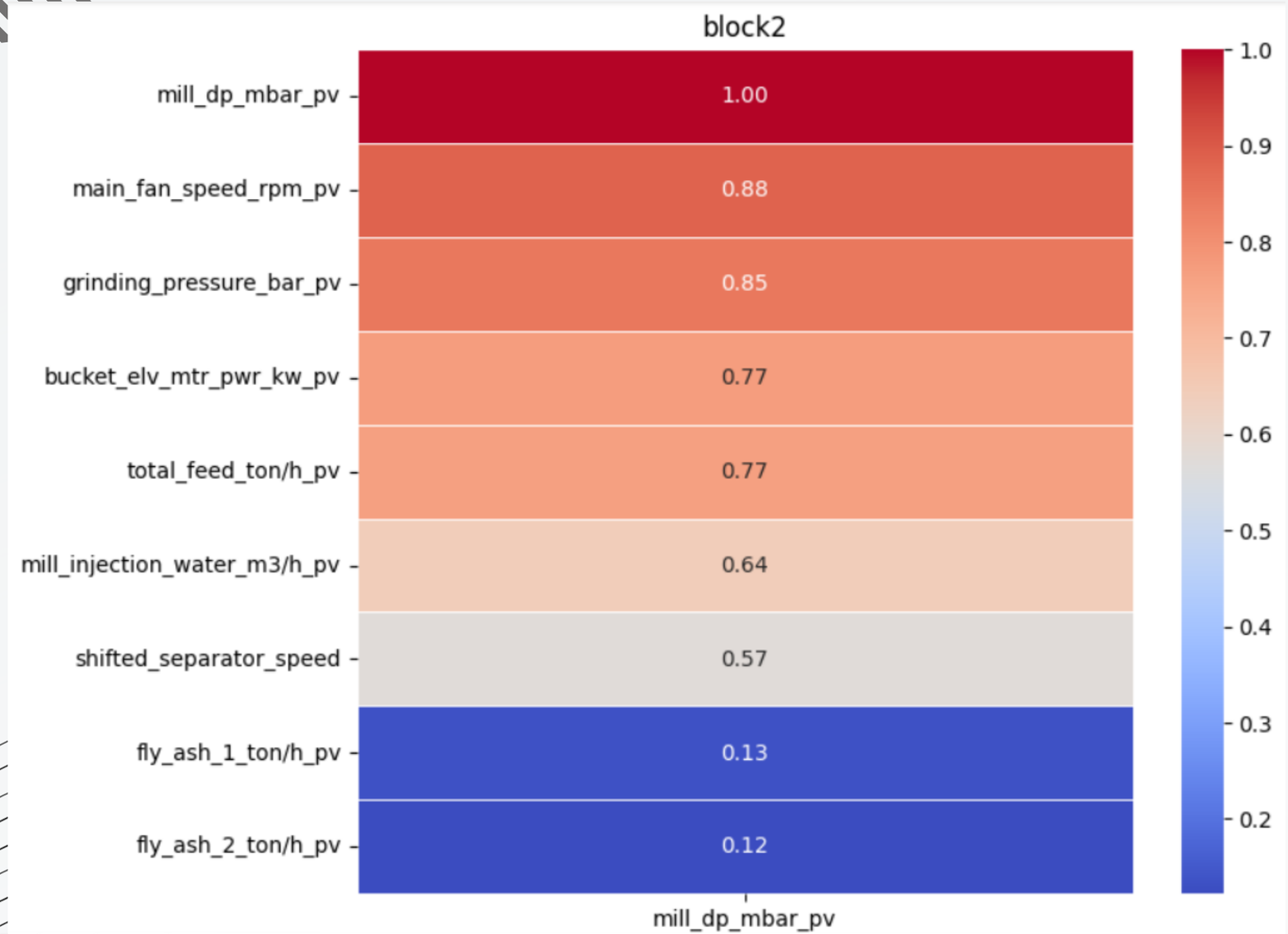
NEXT STEPS

CORRELATION



- Based on the normal operation of the mill
- The corresponding delay has been applied to each input variable
- The correlation of the separator pwr with the shifted separator speed is 0.38
- The correlation of the separator pwr with the shifted mill is 0.36

CORRELATION



- Based on the normal operation of the mill
- The corresponding delay has been applied to each input variable
- The main fan speed and the grinding pressure have the biggest correlation with the output
- The fly ash 1 and the fly ash 2 have the weakest correlation with the output

TWO TYPES OF CEMENT

BLOCK-1

CPII		
	MSE	MAPE
BayesianRidge	4.664	0.0187
XGBRegressor	4.655	0.0178

CPIV		
	MSE	MAPE
BayesianRidge	3.8	0.014
KNeighborsRegressor	4.52	0.015

BLOCK-2

CPII		
	MSE	MAPE
Random Forest	13.003	0.0105
XGBRegressor	18.979	0.0118

CPIV		
	MSE	MAPE
KNeigbors Regressor	8.293	0.014
XGBRegressor	6.497	0.01

REGRESSORS

We used 9 regression algorithms
Two of them stood out for both problems

We performed hyperparameters analysis and tested different values for parameters like n_estimators, learning_rate (XGB), min_samples_split, min_samples_leaf (Random Forest)

BLOCK 1

01 RANDOM FOREST

02 XGB

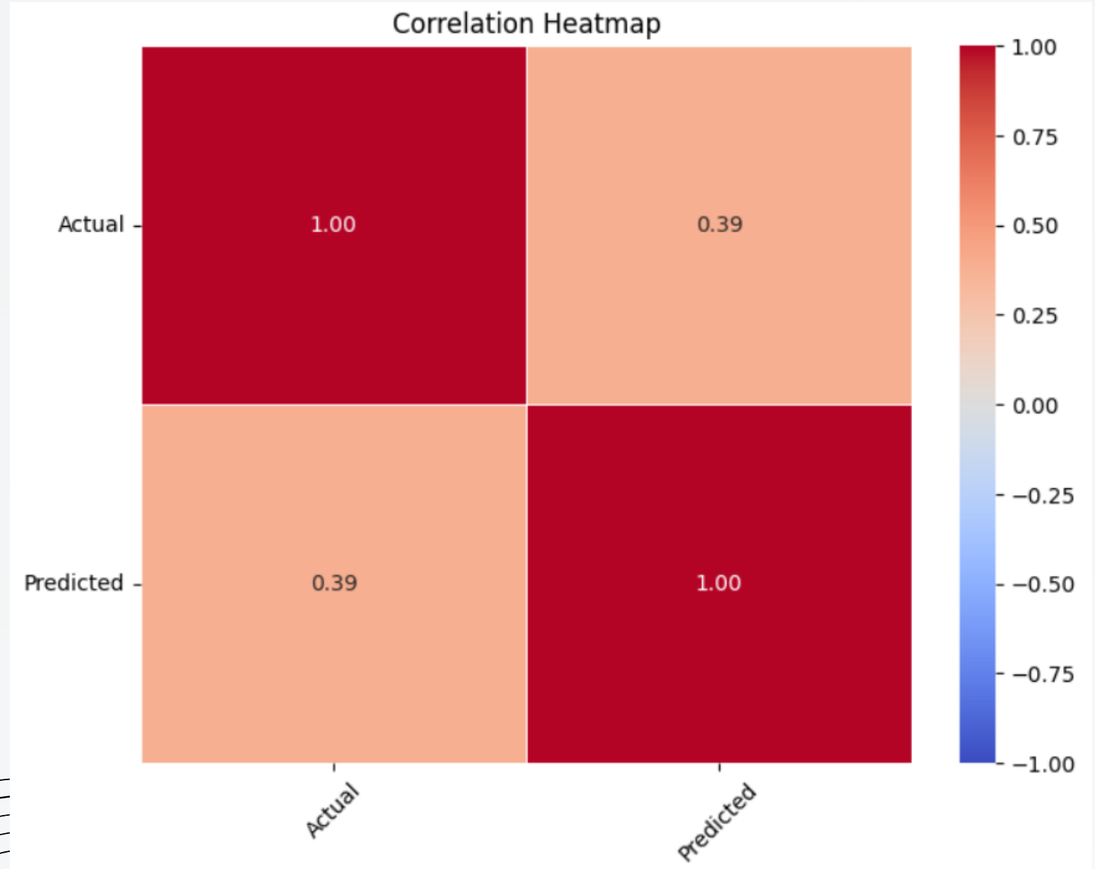
BLOCK 2

01 RANDOM FOREST

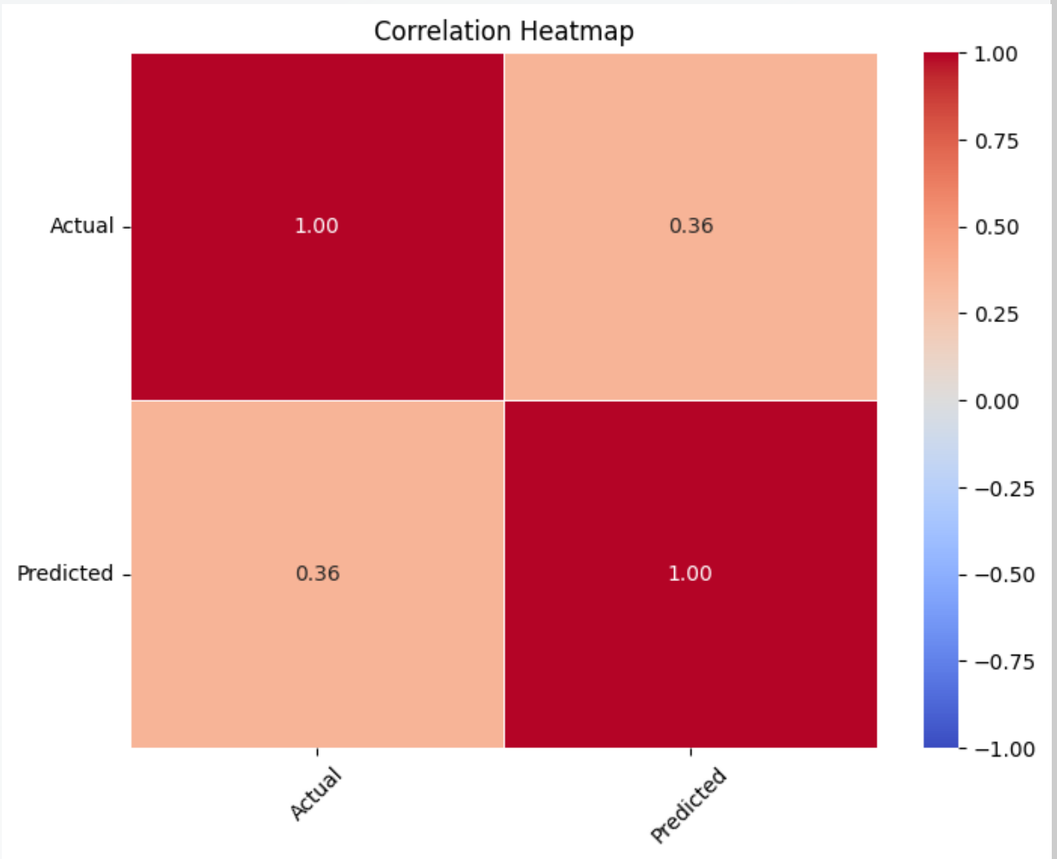
02 XGB

ML MODELS - BLOCK 1

XGBRegressor



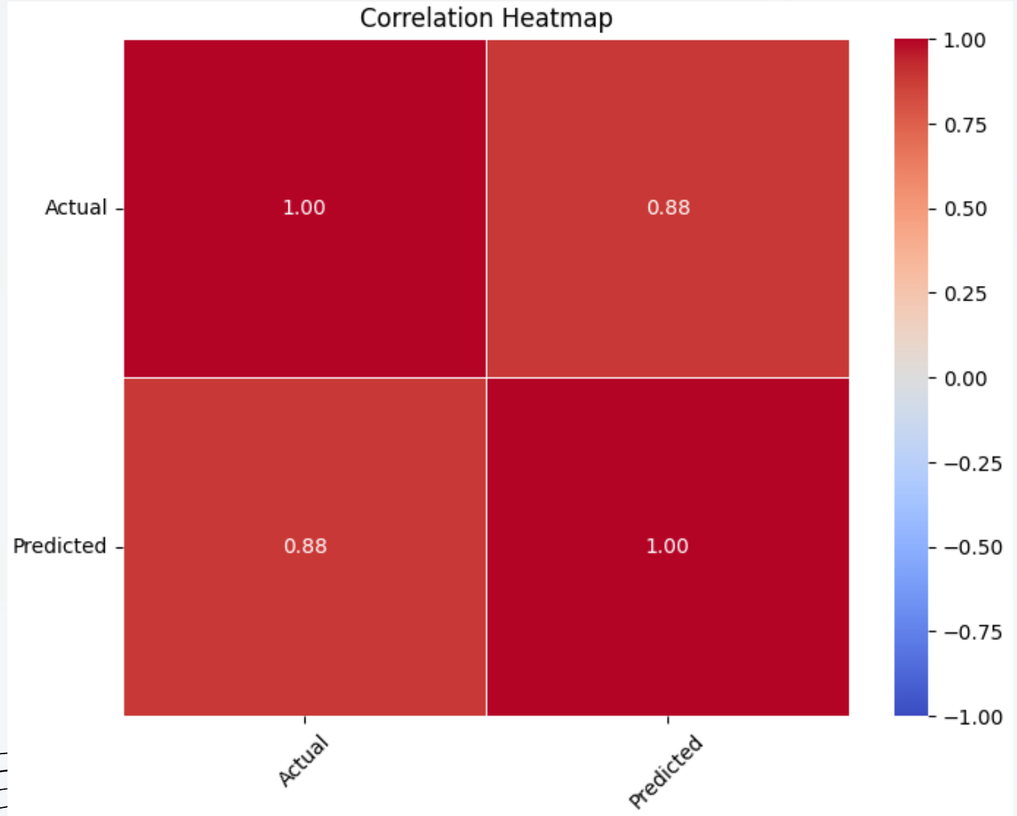
BayesianRidge



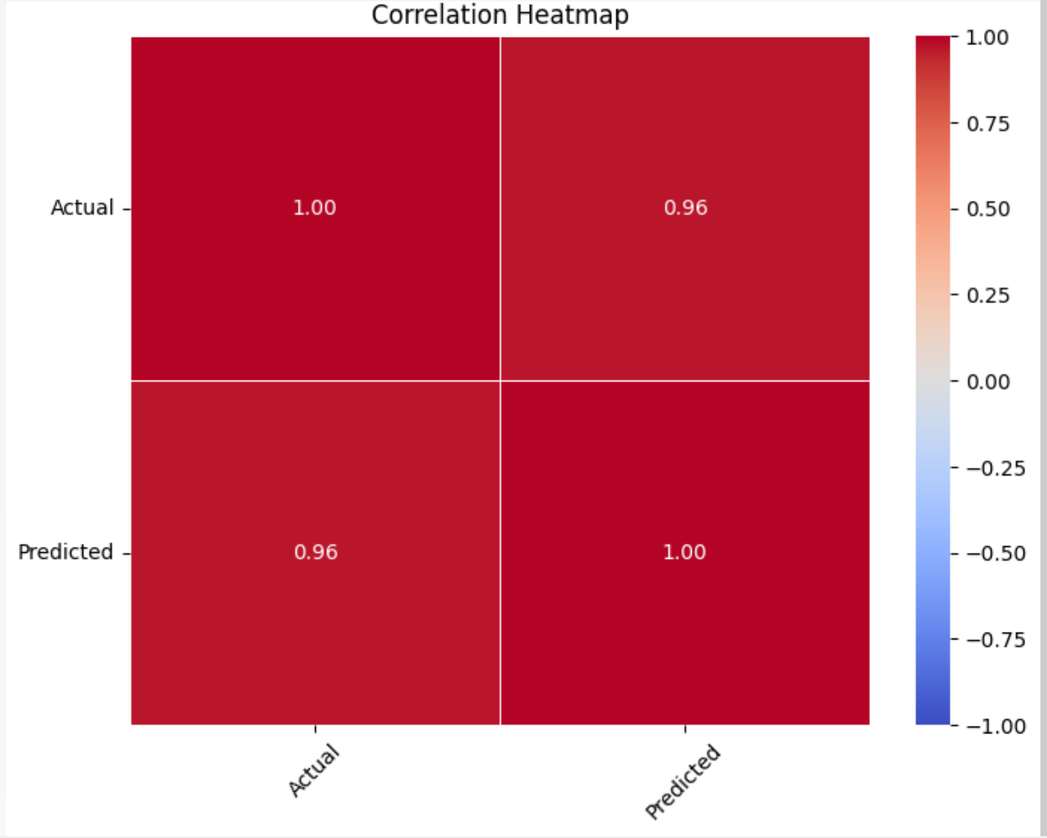
For Block 1 our models were **insufficient**

ML MODELS - BLOCK 2

Random Forest



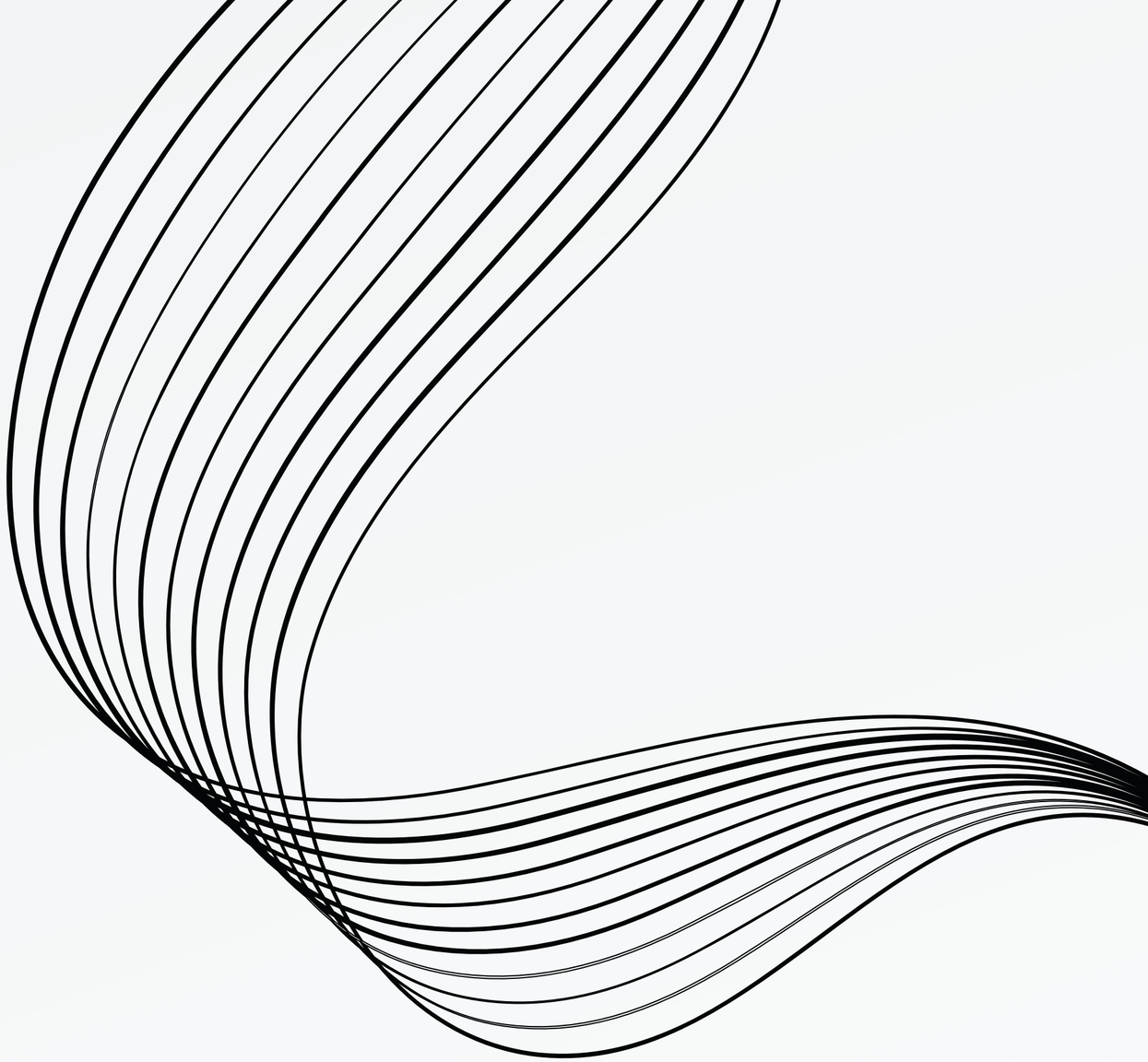
XGBRegressor



After running hyperparameters for both algorithms we concluded that our model using **XGB** is sufficiently **accurate**

FINAL THOUGHTS

This is the end... we still don't know how to code... but at least we met each other and had fun



THANK YOU FOR WATCHING