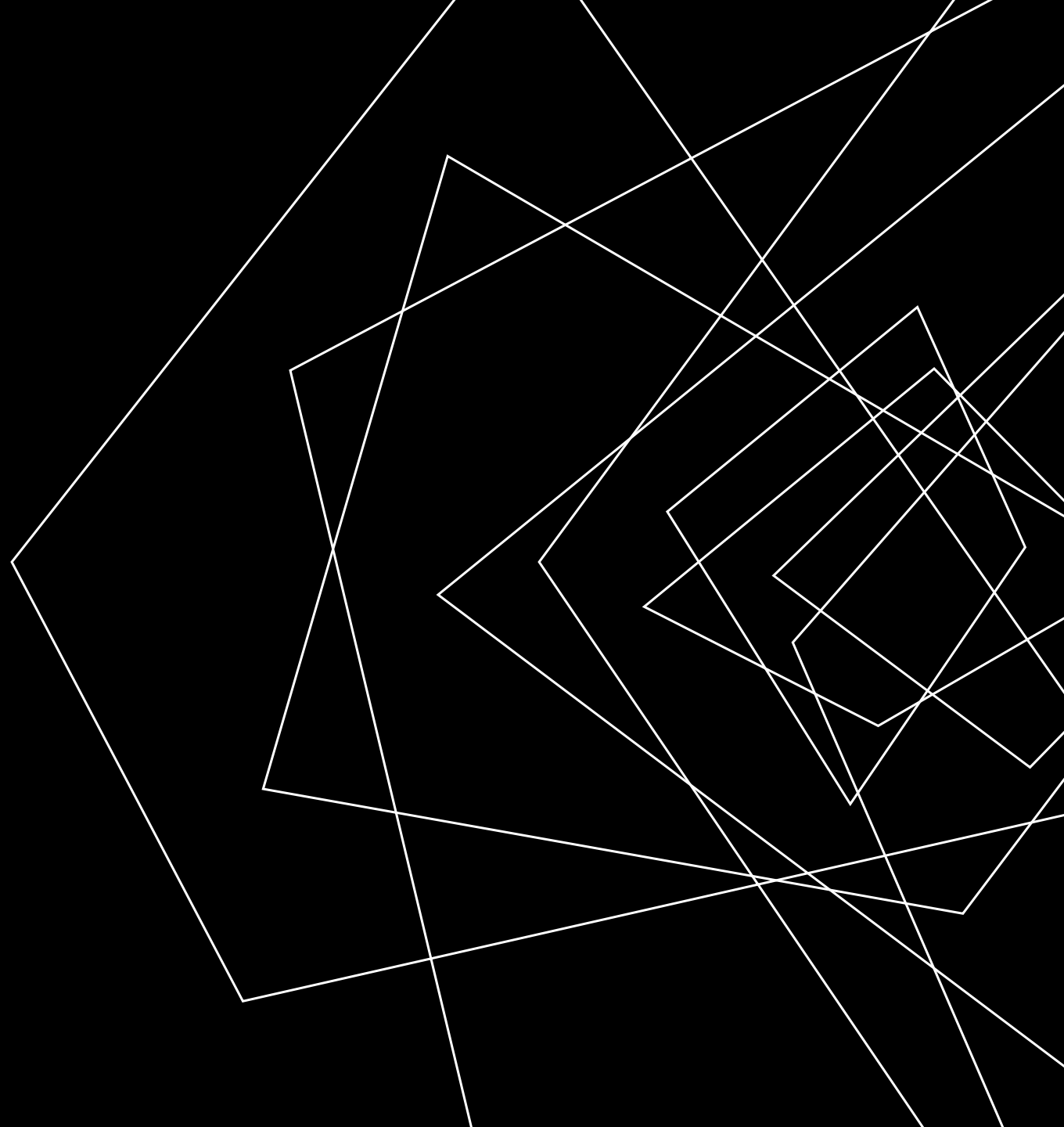
DATA ENGINEERING PROJECT

ANALYSE AND FIND CORRELATION AMONG VARIOUS WEATHER PARAMETERS IN TWO AACHEN WEATHER DATASETS

Md Shaon Hossain Sani

AGENDA

- Introduction
- Process
- Datasets
- Different Visualization Insights
- Finding Correlation among different features
- Conclusion



INTRODUCTION

The main outcome of this project is to analysis two weather datasets that helps to make correlation between temperature vs other parameters. In addition, it helps to find a pattern if there is any environmental relationship between temperature and other parameters exist or not.

PROCESS

- Extract dataset from source url.
- Clean and restructure datasets.
- Visualize different parameters of the datasets.
- Find Correlation among different parameters and specially with temperature.

DATASETS

Aachen Meteostat data source's dataset

	Hour	temp	dwpt	rhum	prcp	wdir	wspd	wpgt	pres
count	744.000000	744.000000	744.000000	744.000000	744.000000	744.000000	744.000000	744.000000	744.000000
mean	11.500000	16.532124	13.147446	81.596774	0.093683	215.172043	12.230780	21.530645	1016.720565
std	6.926843	3.071153	1.497686	12.478948	0.259672	87.954151	5.197554	8.150745	5.406774
min	0.000000	9.800000	8.400000	48.000000	0.000000	1.000000	3.700000	7.400000	1001.500000
25%	5.750000	14.275000	12.200000	73.000000	0.000000	194.000000	7.400000	14.800000	1014.200000
50%	11.500000	16.100000	13.300000	84.000000	0.000000	225.000000	11.100000	20.400000	1017.200000
75%	17.250000	18.400000	14.300000	91.000000	0.000000	254.000000	14.800000	25.900000	1020.000000
max	23.000000	25.700000	16.500000	100.000000	2.200000	360.000000	29.600000	46.300000	1027.300000

DATASETS

Aachen Meteostat data source's dataset after Clean and restructure

	Date	Hour	temp	dew point	relative humidity	precipitation	wind direction	average wind speed	average wind gust	pressure
0	2021-08-01	0	14.1	13.0	93	0.0	216	9.3	14.8	1010.9
1	2021-08-01	1	13.8	12.9	94	0.0	212	11.1	16.7	1010.6
2	2021-08-01	2	13.6	13.0	96	0.0	213	11.1	16.7	1010.3
3	2021-08-01	3	13.4	12.5	94	0.0	215	13.0	18.5	1010.3
4	2021-08-01	4	13.2	12.6	96	0.0	217	13.0	18.5	1010.1

DATASETS

Aachen Mobilthek data source's dataset

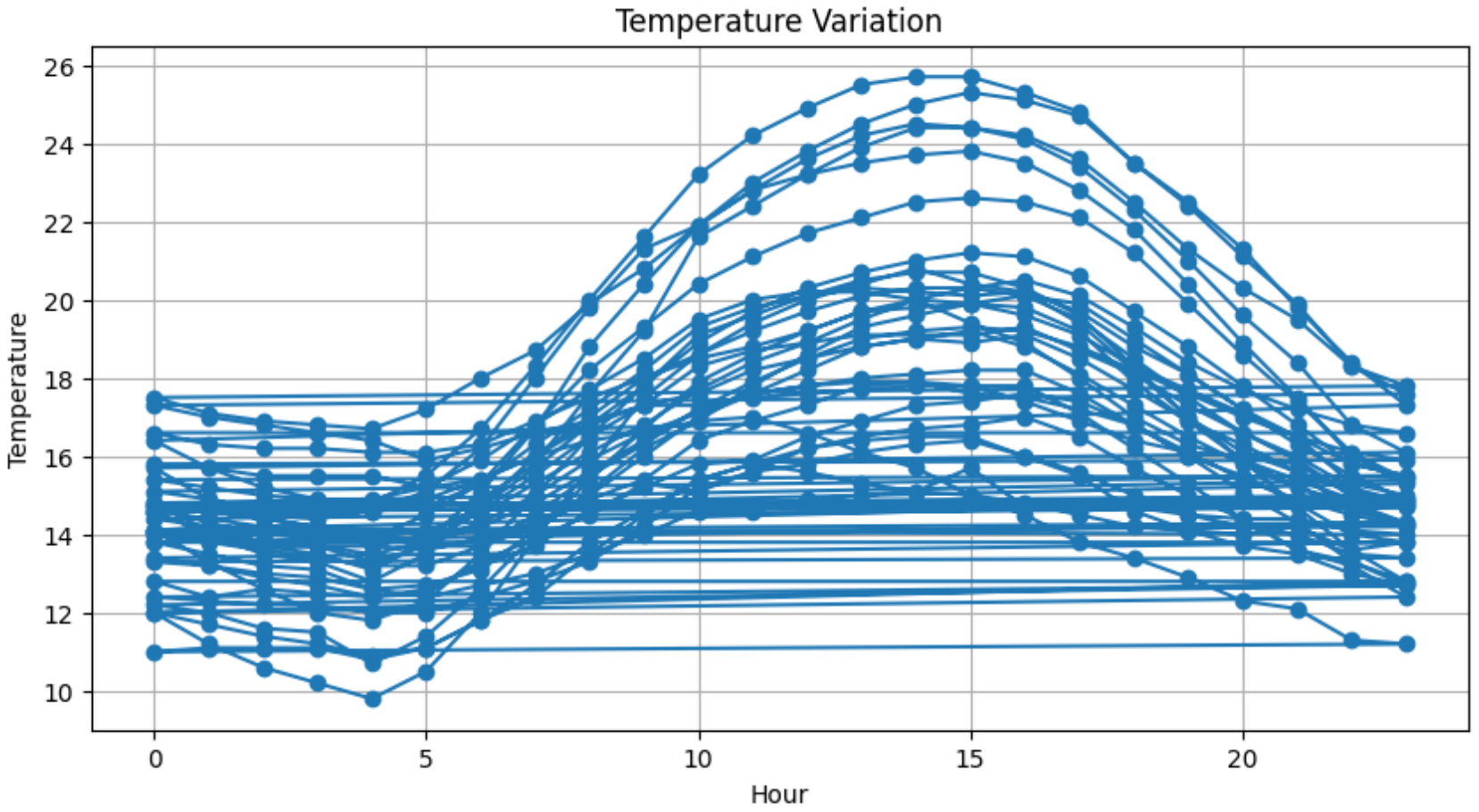
	parameter	value	NumberOfVehicles	Geohash7	Period
0	Humidity	36.18	3	u1h2gky	2021-08-01 00:00:00 UTC
1	Humidity	39.80	3	u1h2gmn	2021-08-01 00:00:00 UTC
2	Humidity	34.33	3	u1h2gmp	2021-08-01 00:00:00 UTC
3	Humidity	38.33	2	u1h2gmq	2021-08-01 00:00:00 UTC
4	Humidity	40.09	2	u1h2gmj	2021-08-01 00:00:00 UTC

DATASETS

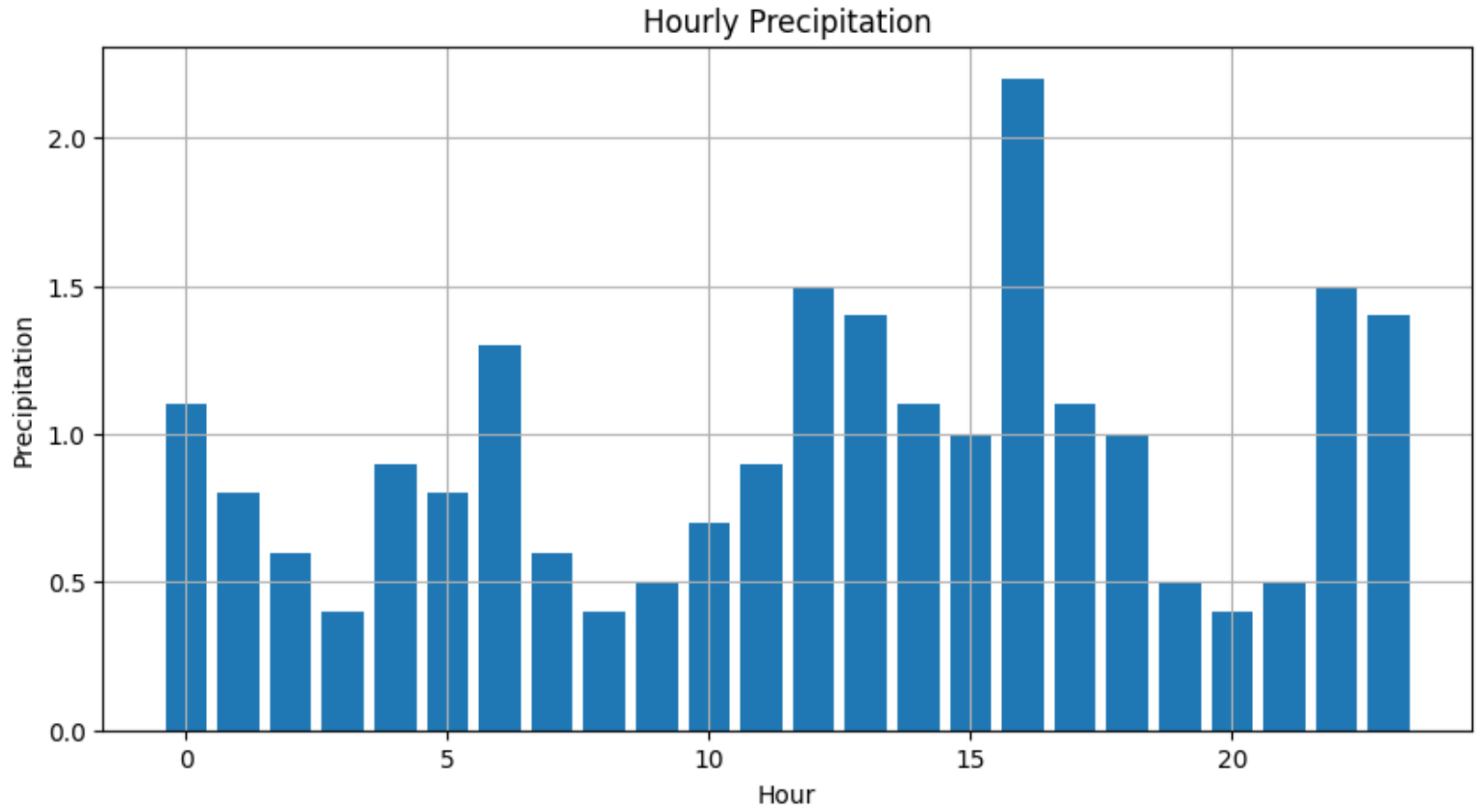
Aachen Mobilithek data source's dataset after Clean and restructure

	Period	Humidity	PM10	PM2.5	Temperature
2021-08-01 00:00:00 UTC	36.746000	5.016000	4.613000	29.075714
2021-08-01 07:00:00 UTC	39.000000	4.200000	4.290000	NaN
2021-08-01 07:30:00 UTC	NaN	3.920000	4.000000	NaN
2021-08-01 19:00:00 UTC	43.060000	4.000000	3.590000	NaN
2021-08-02 00:00:00 UTC	47.356132	6.159545	5.799312	24.938299

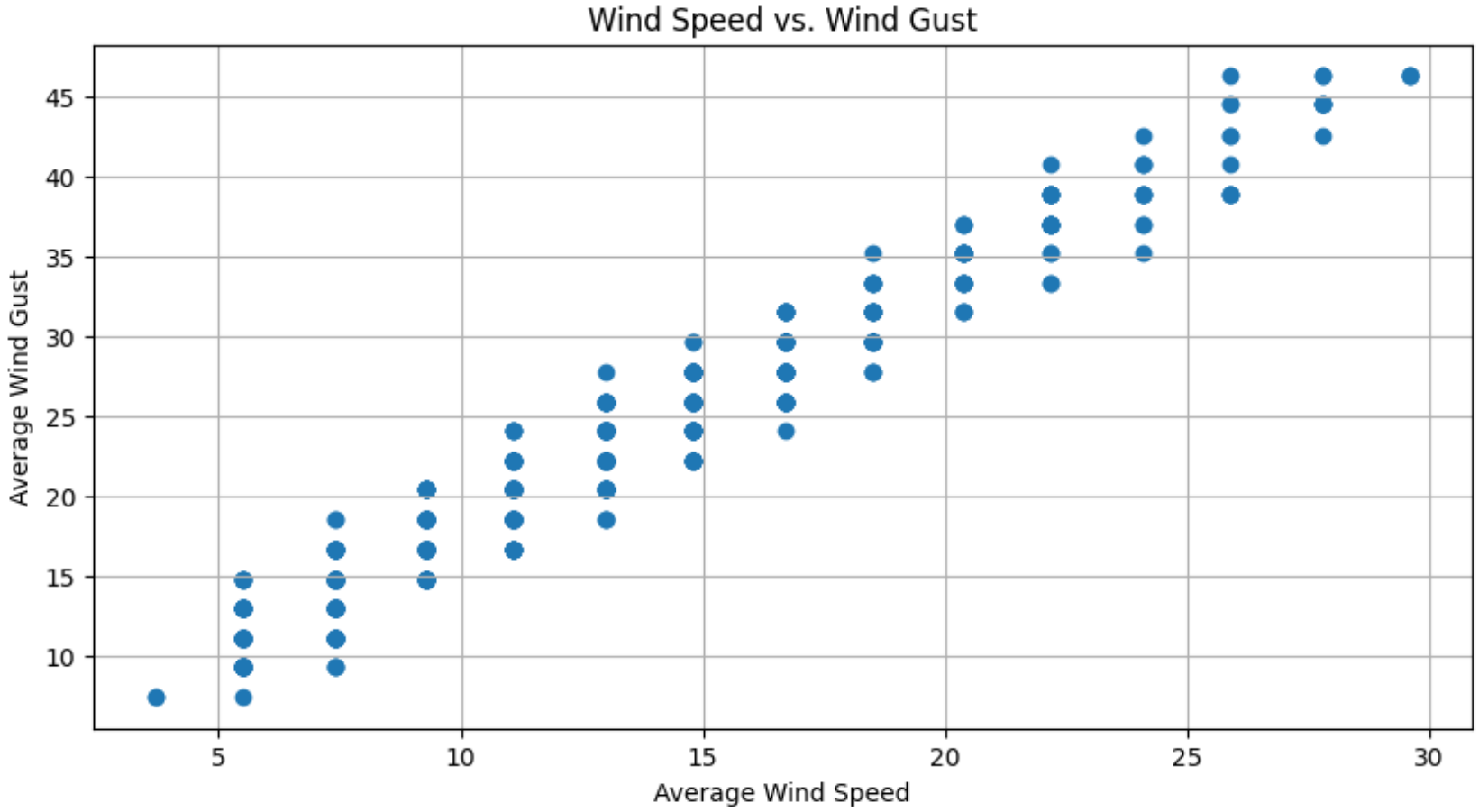
VISUALIZATION FOR **AACHEN METEOSTAT** DATASET



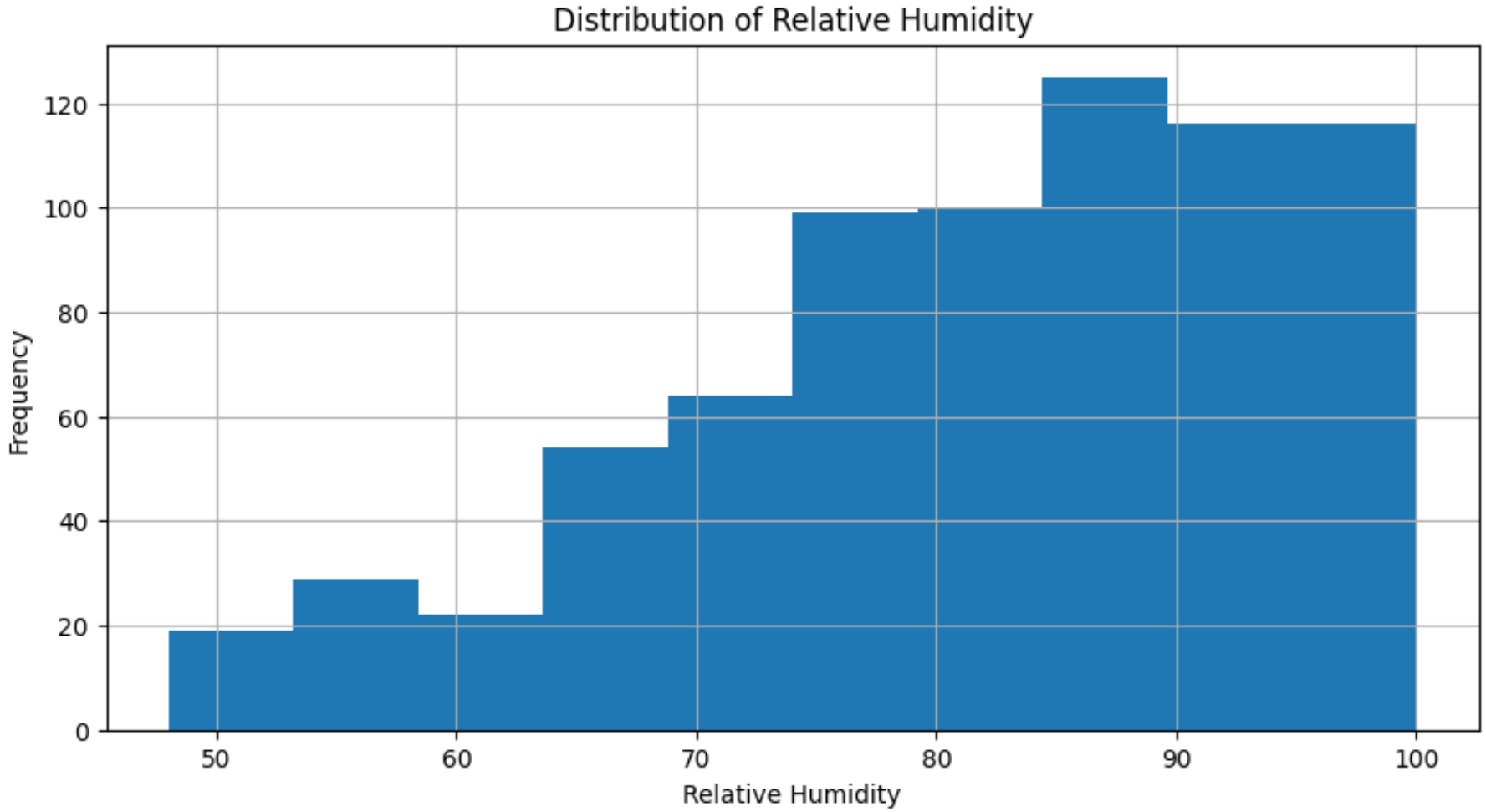
VISUALIZATION FOR **AACHEN METEOSTAT** DATASET



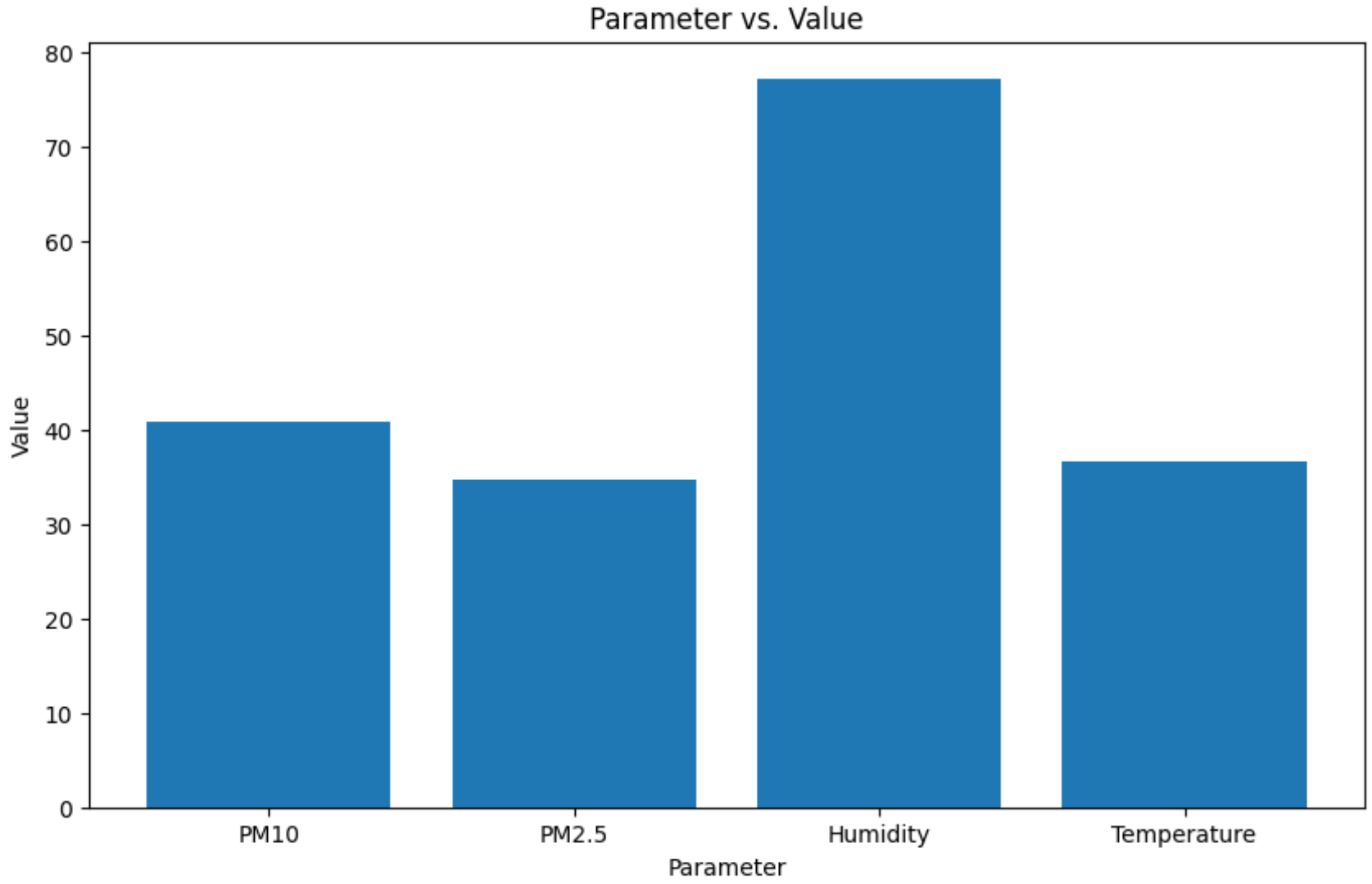
VISUALIZATION FOR **AACHEN METEOSTAT** DATASET



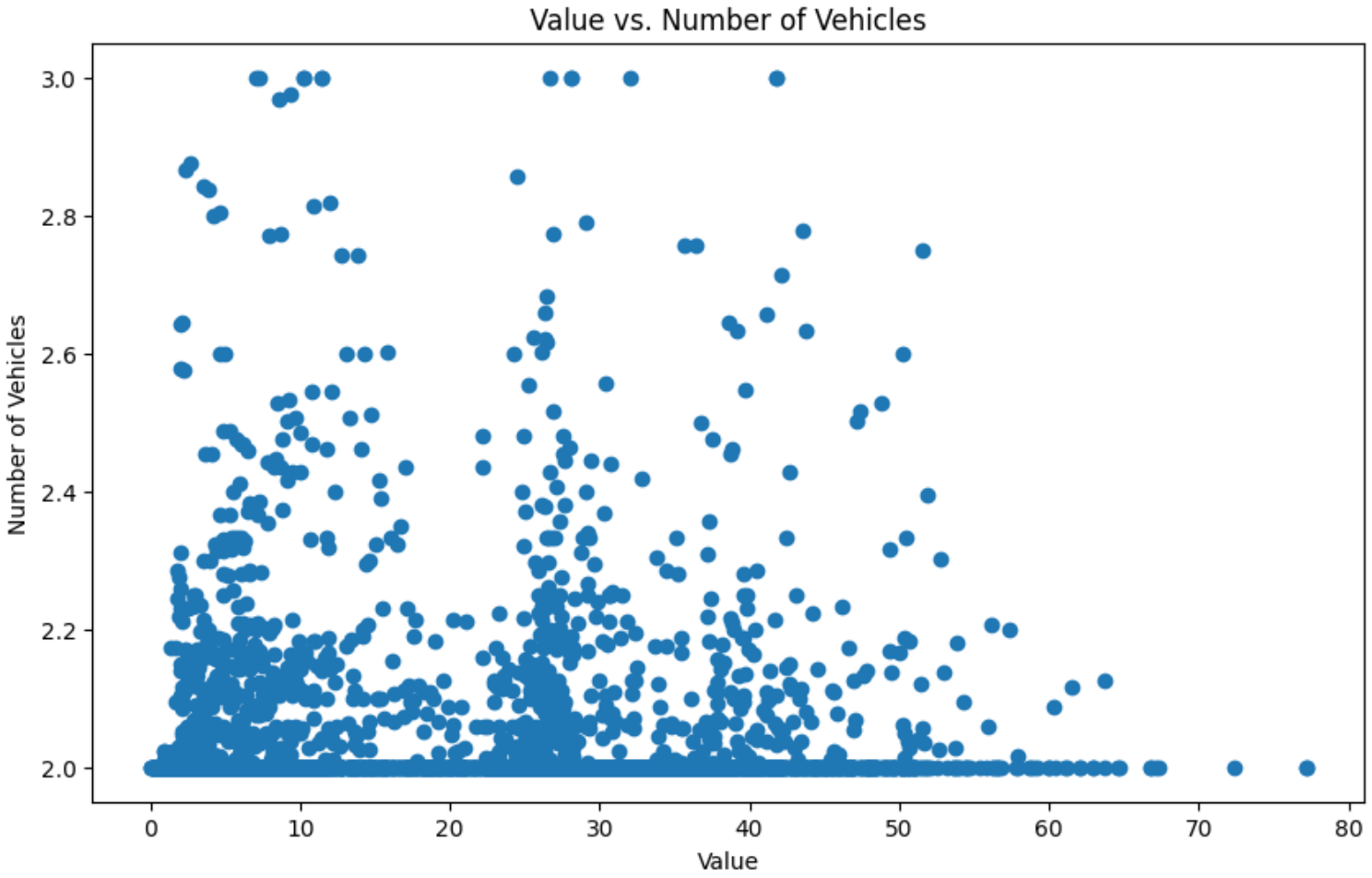
VISUALIZATION FOR **AACHEN METEOSTAT** DATASET



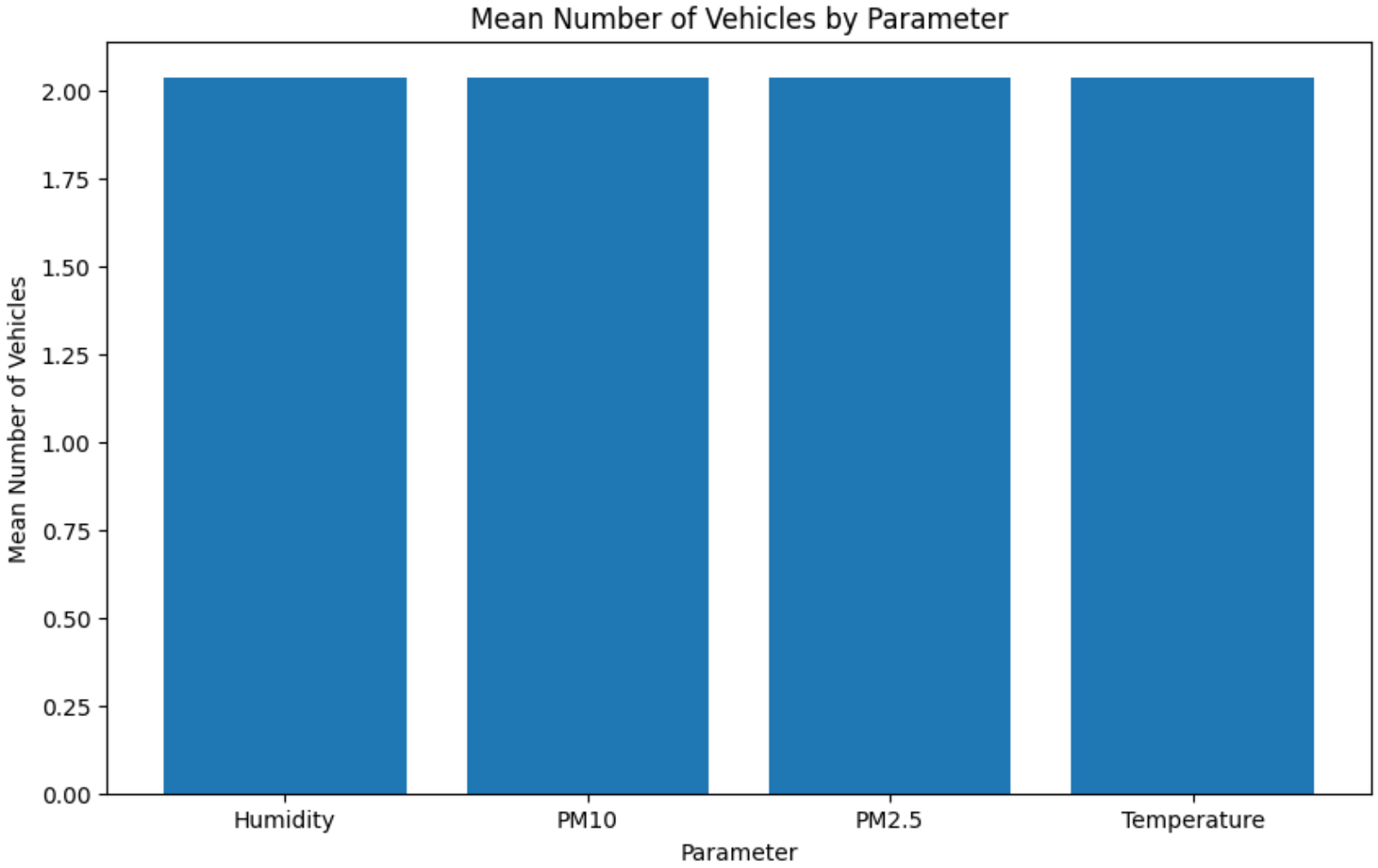
VISUALIZATION FOR **AACHEN MOBILITHEK** DATASET



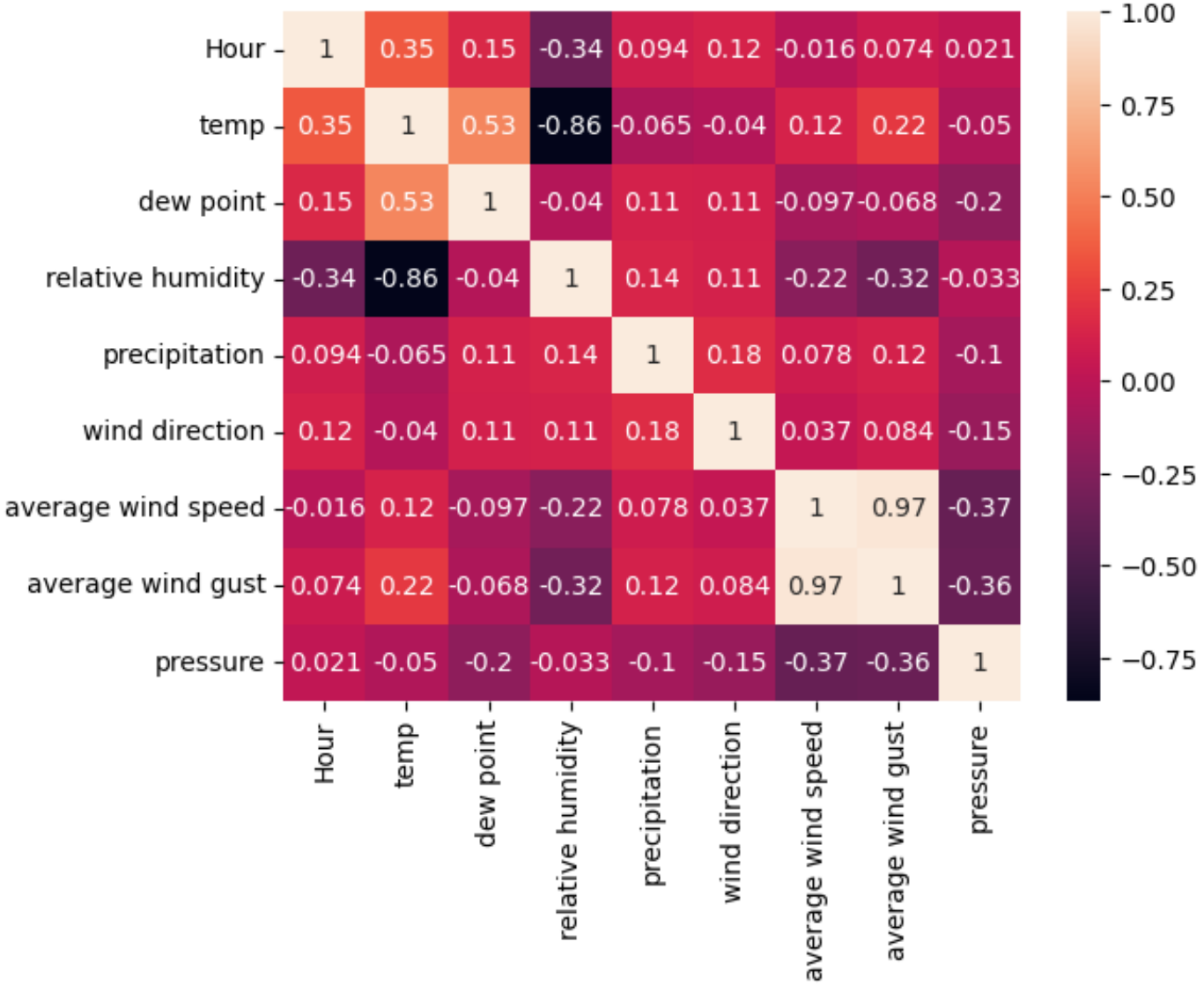
VISUALIZATION FOR **AACHEN MOBILITHEK** DATASET



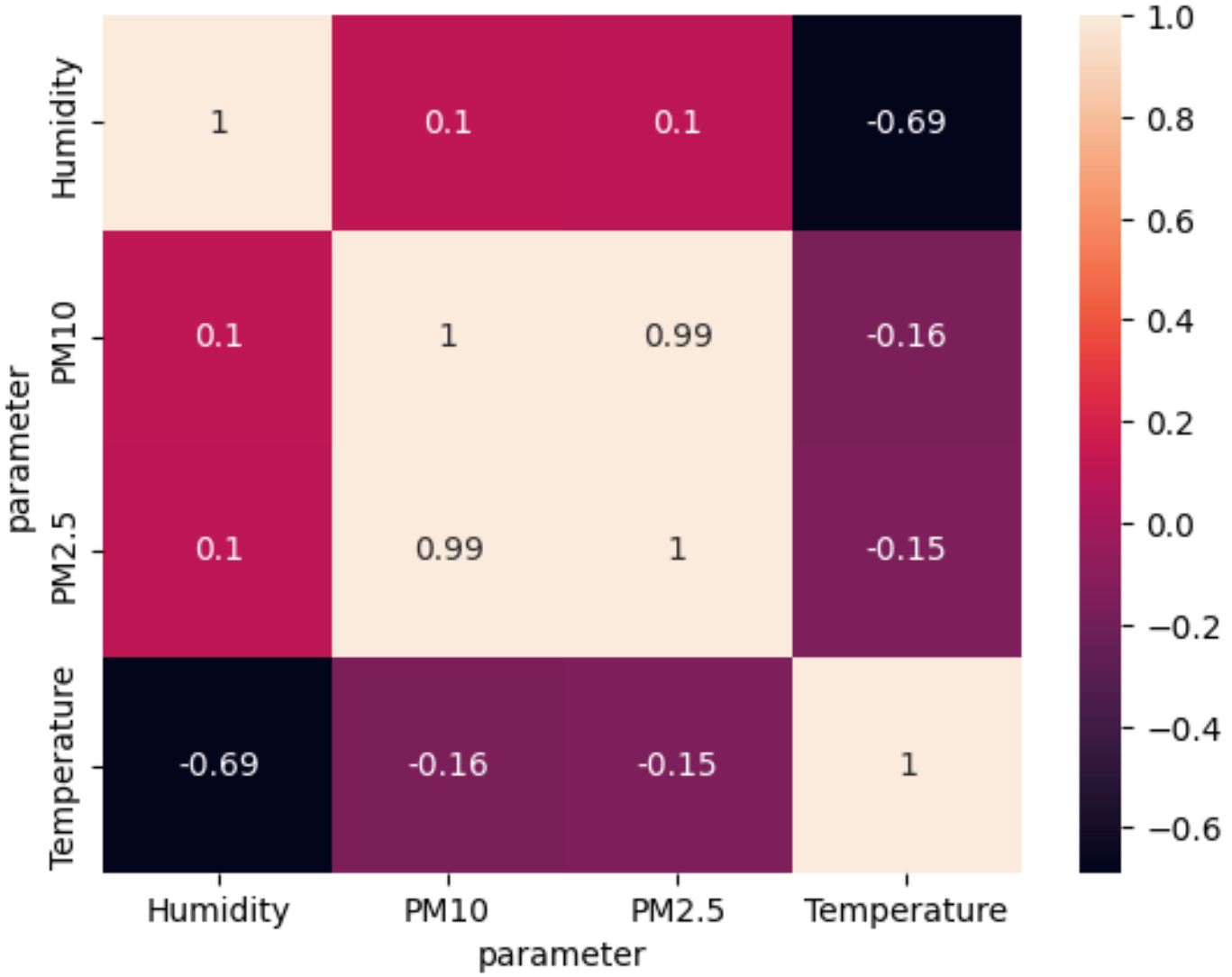
VISUALIZATION FOR **AACHEN MOBILITHEK** DATASET



CORRELATION COEFFICIENTS MATRIX FOR AACHEN METEOSTAT DATASET



CORRELATION COEFFICIENTS MATRIX FOR AACHEN MOBILITHEK DATASET



CONCLUSION

So from that correlation coefficients matrix and theory and our observation from both Aachen datasets, we have seen that, in Mobilithek dataset,the temperture has a negative correlation with other parameters of the dataset's. And on the other hand, in Meteostat we have seen that there are both positive and negative correlation among the parameters. But temperature has considerably strong correlation with humidity in both datasets.

THANK YOU

Any Questions?