Seoul Bike Sharing Demand

By Guérou Rayane, Hadji Soudaisse, Khegar Yanis

Ins and outs of the problem

- Prediction of the needs of the clients in bike rentals
- Work on the efficiency of the companies to answer the customers needs without failing in having items

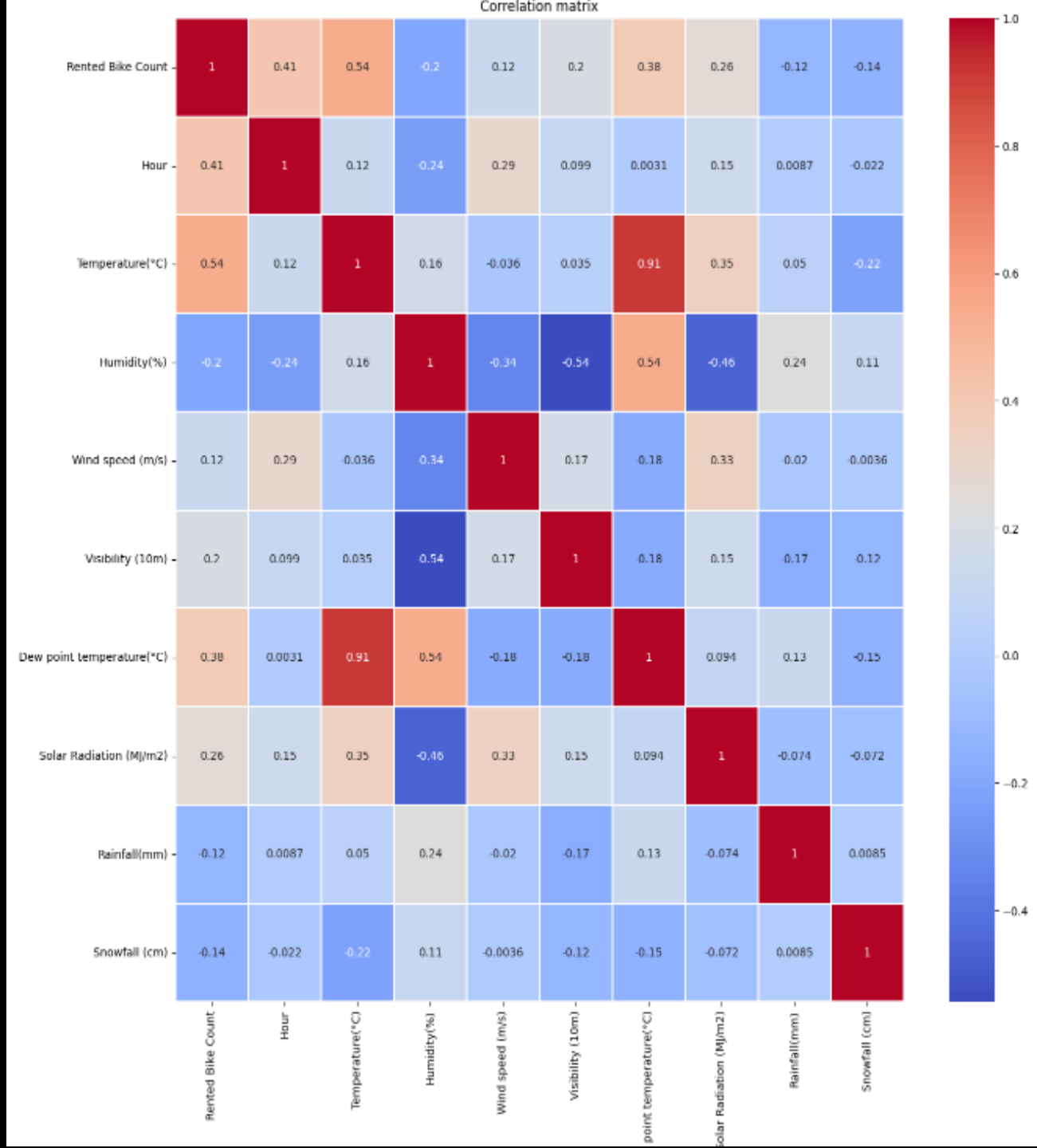
A very documented dataset to work on

	Date	Rented Bike Count	Hour	Temperature(°C)	Humidity(%)	Wind speed (m/s)	Visibility (10m)	Dew point temperature(°C)	Solar Radiation (MJ/m2)	Rainfall(mm)	Snowfall (cm)	Seasons	Holiday	Functioning Day
0	01/12/2017	254	0	-5.2	37	2.2	2000	-17.6	0.0	0.0	0.0	Winter	No Holiday	Yes
1	01/12/2017	204	1	-5.5	38	0.8	2000	-17.6	0.0	0.0	0.0	Winter	No Holiday	Yes
2	01/12/2017	173	2	-6.0	39	1.0	2000	-17.7	0.0	0.0	0.0	Winter	No Holiday	Yes
3	01/12/2017	107	3	-6.2	40	0.9	2000	-17.6	0.0	0.0	0.0	Winter	No Holiday	Yes
4	01/12/2017	78	4	-6.0	36	2.3	2000	-18.6	0.0	0.0	0.0	Winter	No Holiday	Yes
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
8755	30/11/2018	1003	19	4.2	34	2.6	1894	-10.3	0.0	0.0	0.0	Autumn	No Holiday	Yes
8756	30/11/2018	764	20	3.4	37	2.3	2000	-9.9	0.0	0.0	0.0	Autumn	No Holiday	Yes
8757	30/11/2018	694	21	2.6	39	0.3	1968	-9.9	0.0	0.0	0.0	Autumn	No Holiday	Yes
8758	30/11/2018	712	22	2.1	41	1.0	1859	-9.8	0.0	0.0	0.0	Autumn	No Holiday	Yes
8759	30/11/2018	584	23	1.9	43	1.3	1909	-9.3	0.0	0.0	0.0	Autumn	No Holiday	Yes

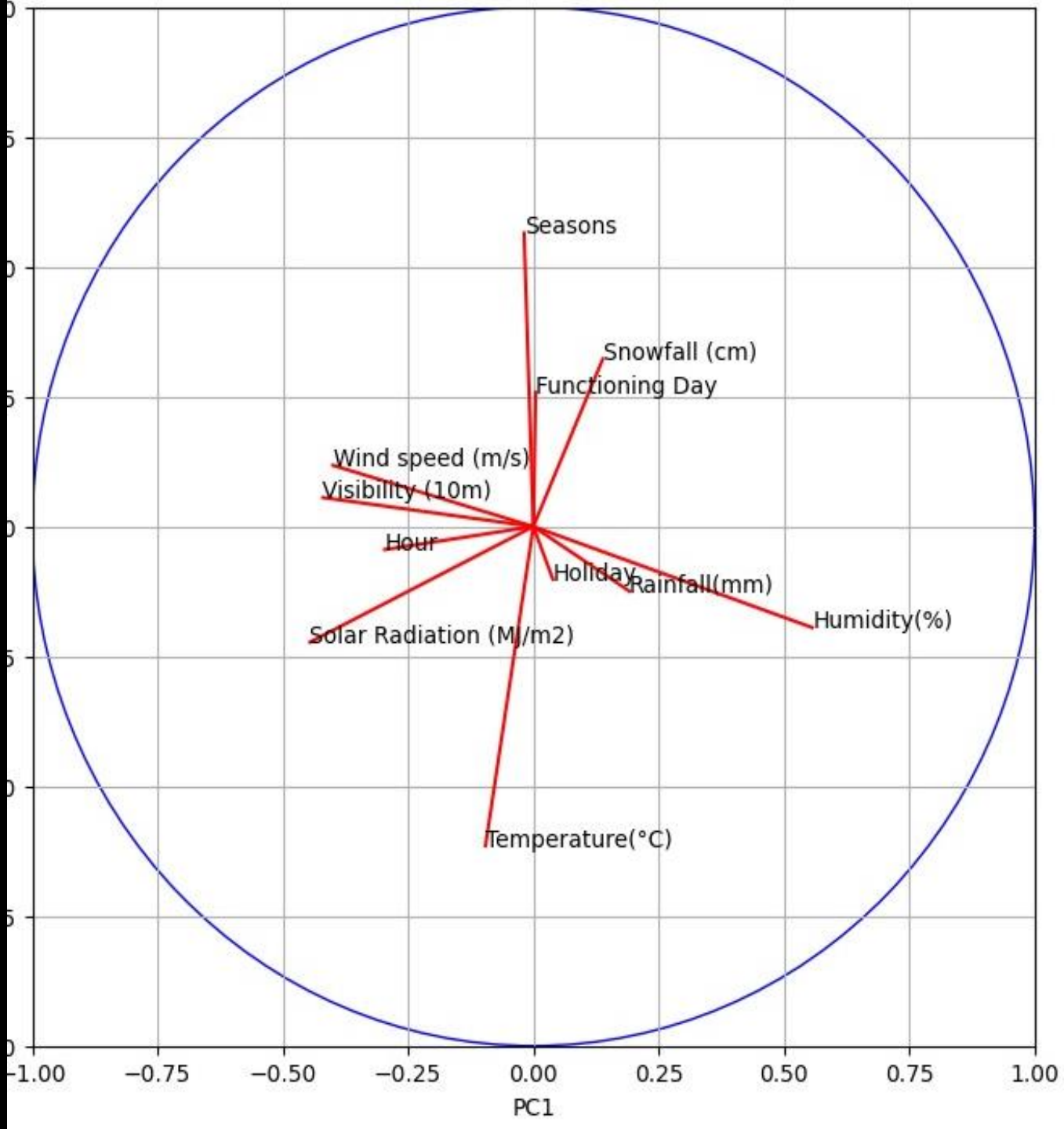
Data preprocessing:

- No missing value
- Normalization of the data set
- Encoding the categorical variables
- Creation of a subset with only numerical values

Correlation Matrix

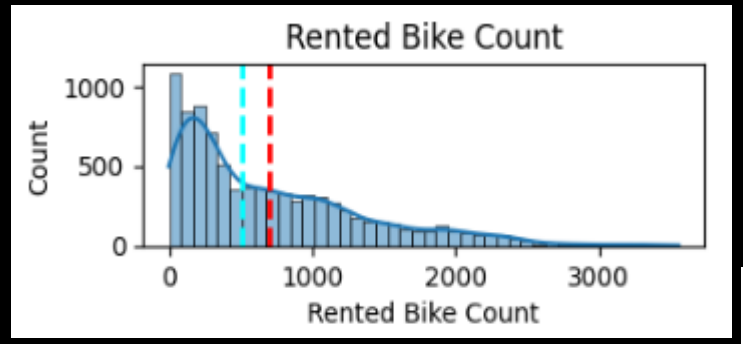


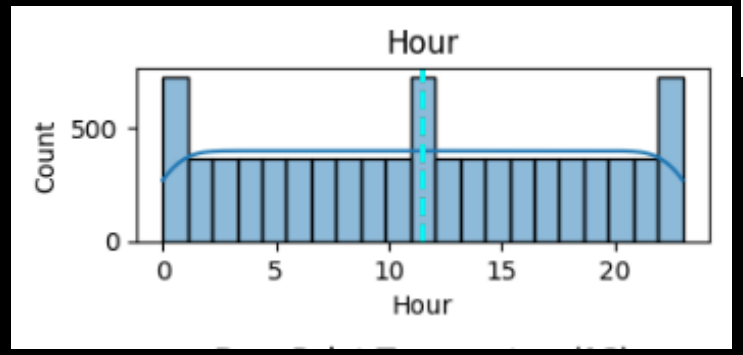
Correlation circle

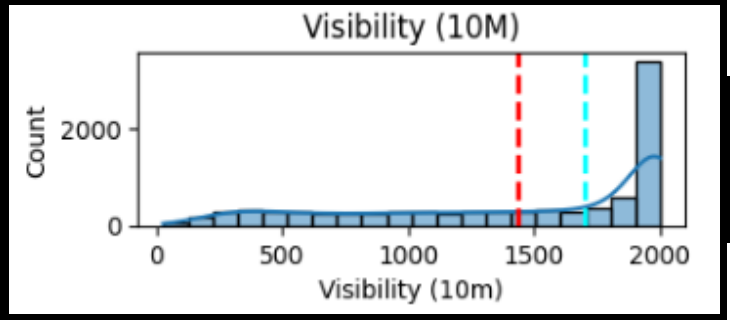


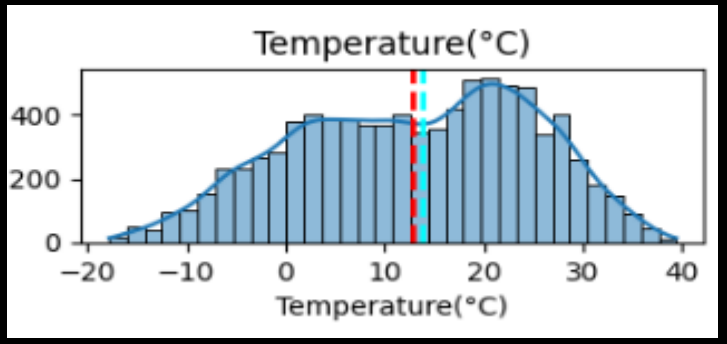
So we wanted to see how our data was distributed on every parameter

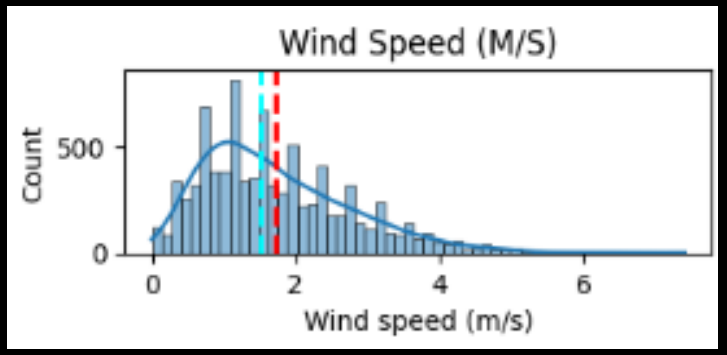
Some distribution graphics we found



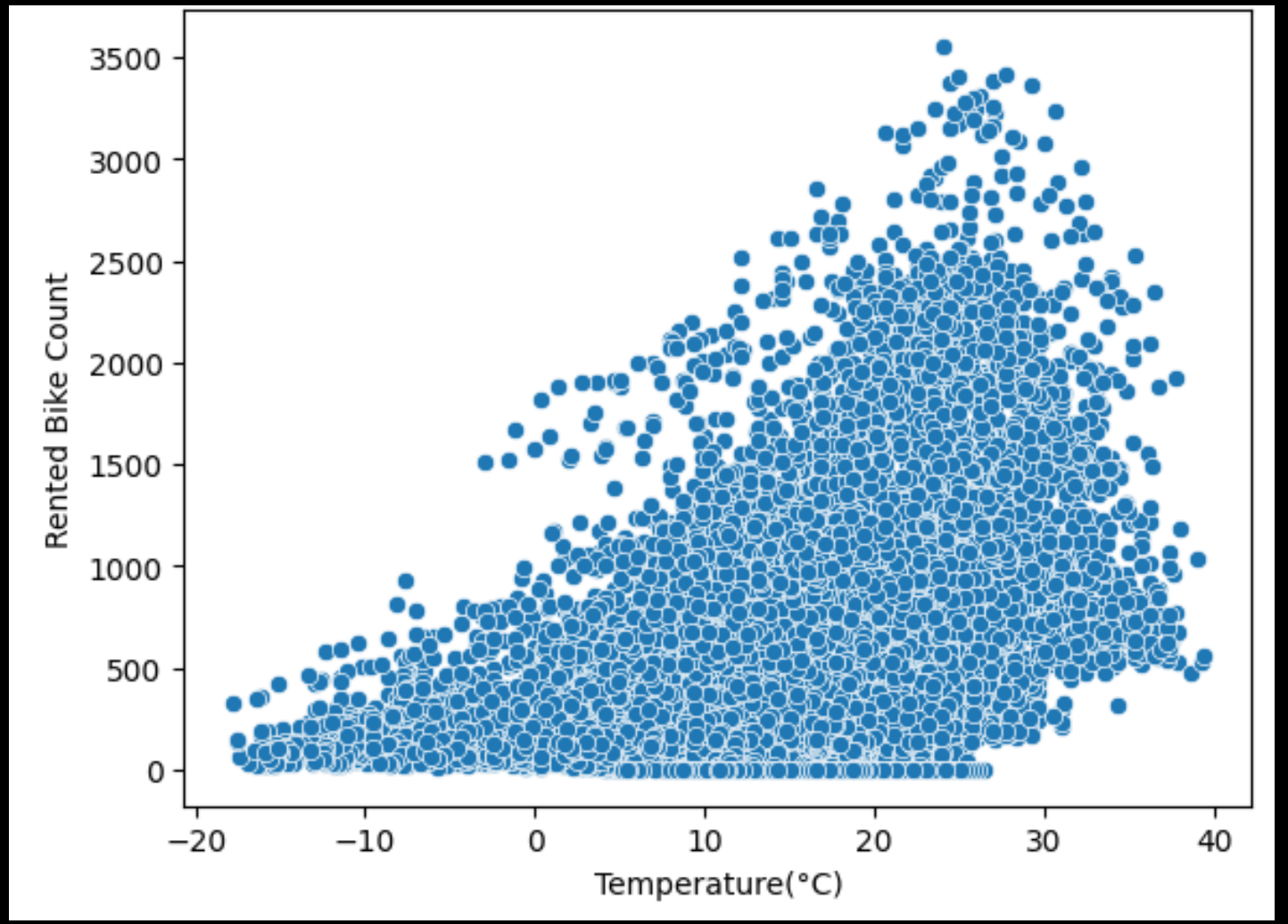




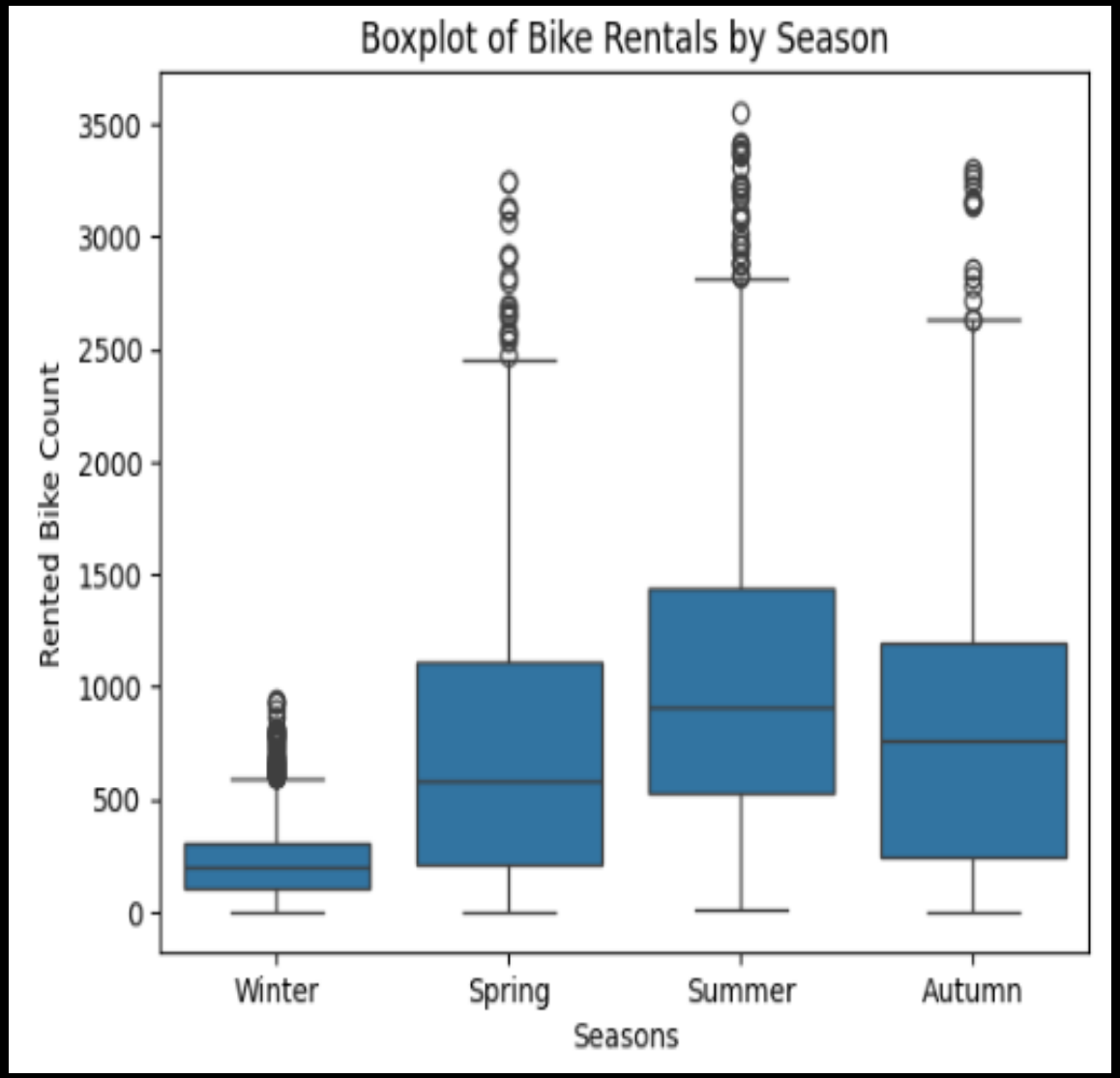




How impactful the temperature is on bike location

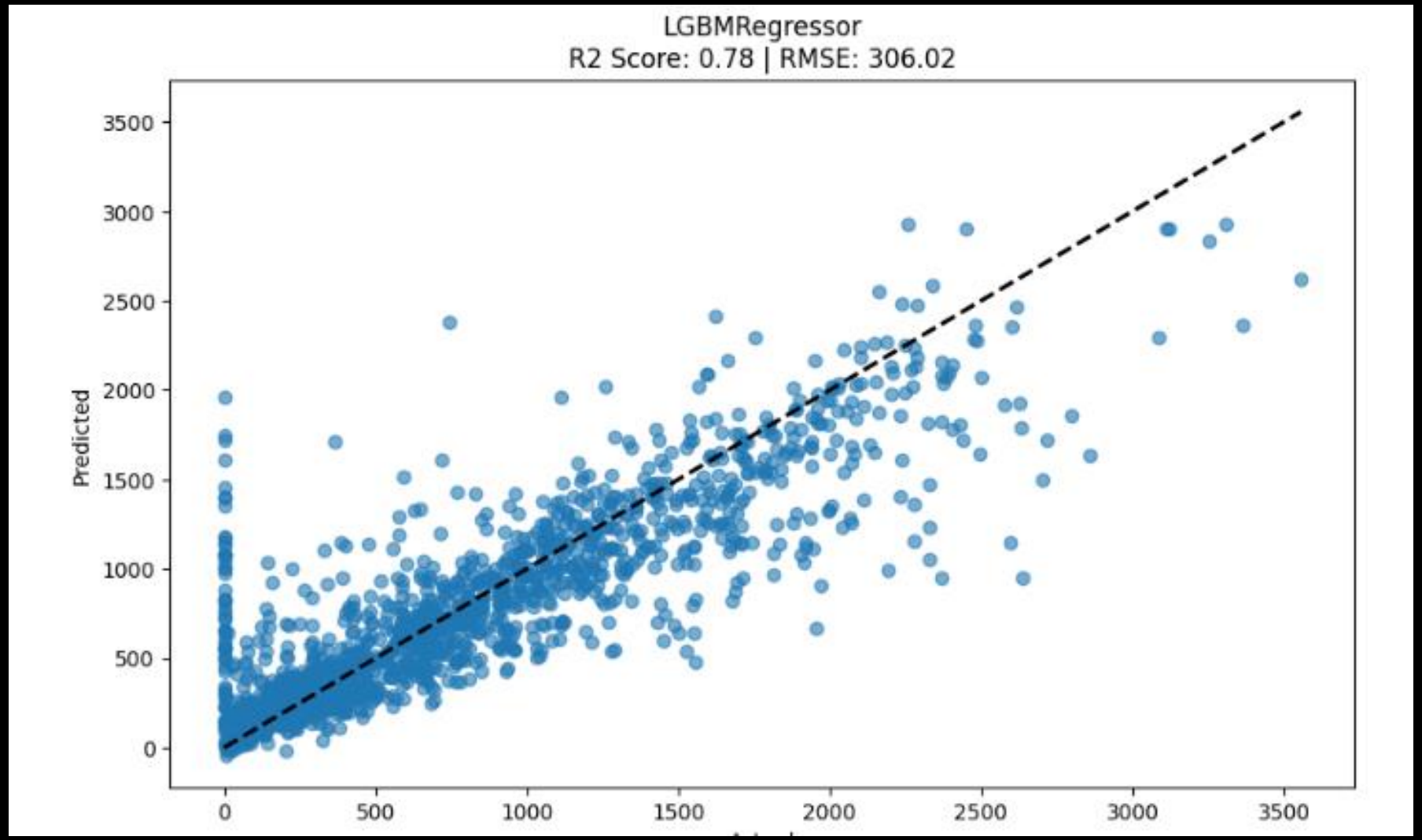


Impact of seasons on bike rentals



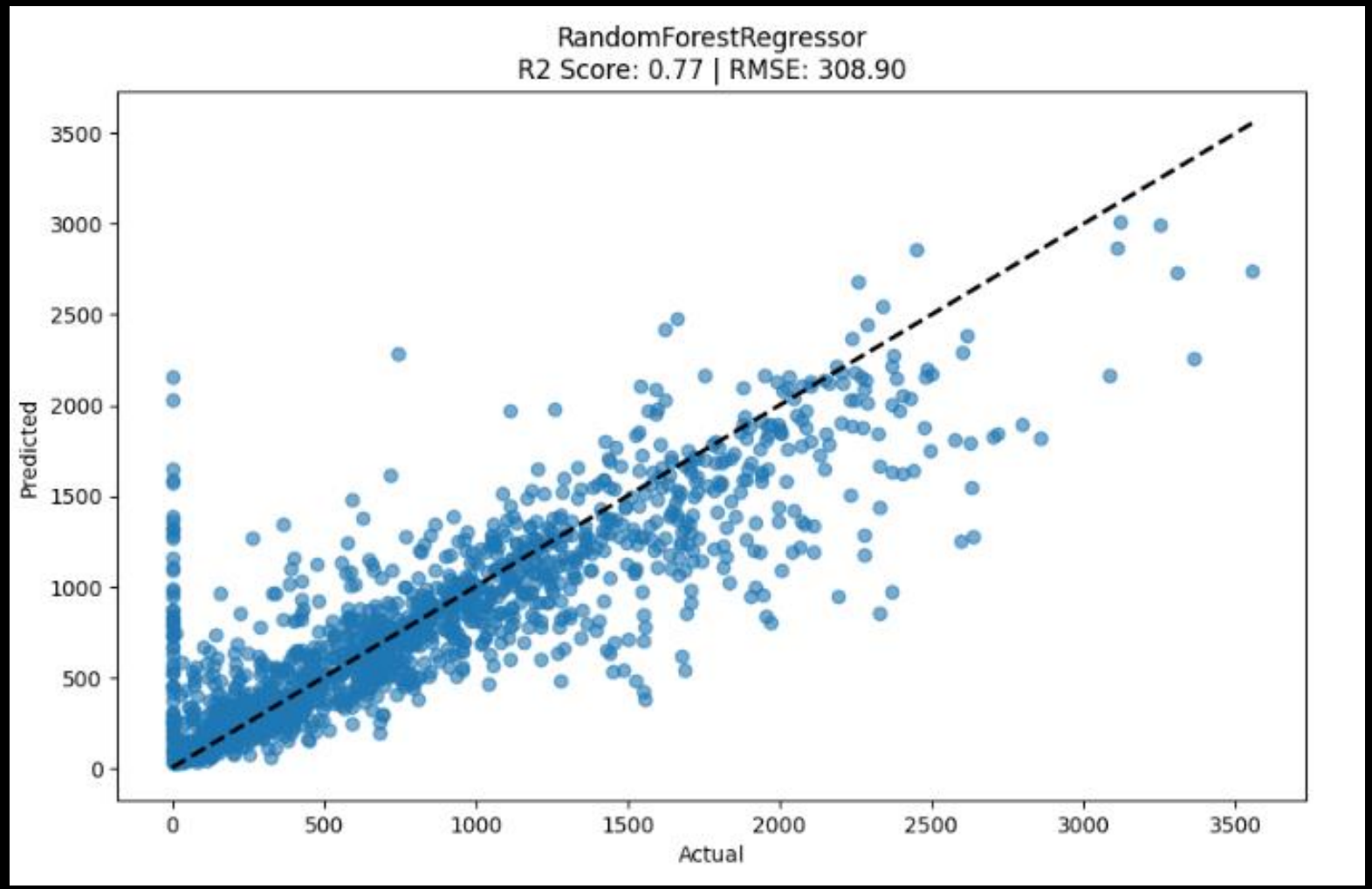
After analyzing how the data was distributed, we started to use different models

The best models



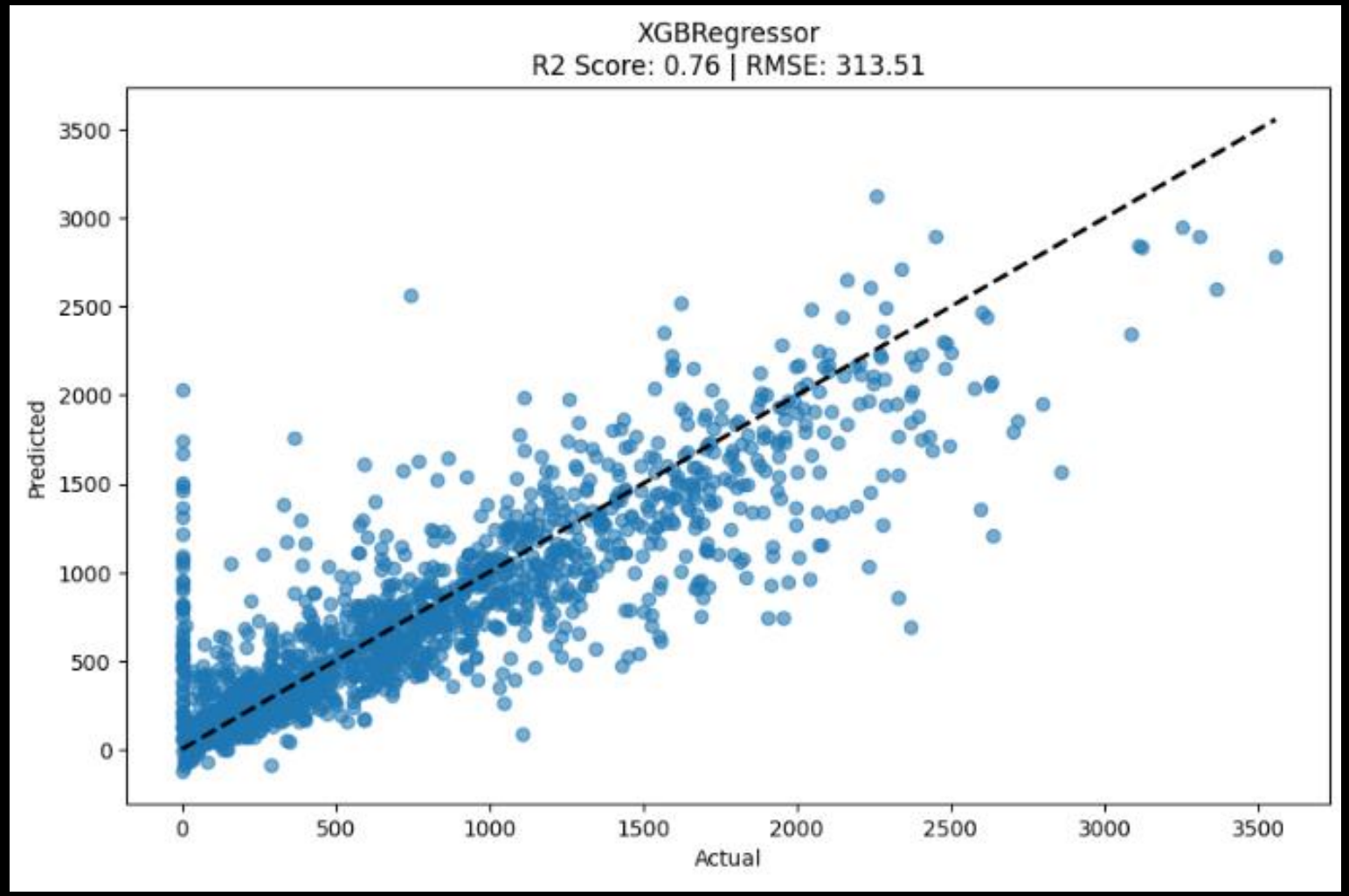
- LGBM Regressor Model, score 0,78

The best models



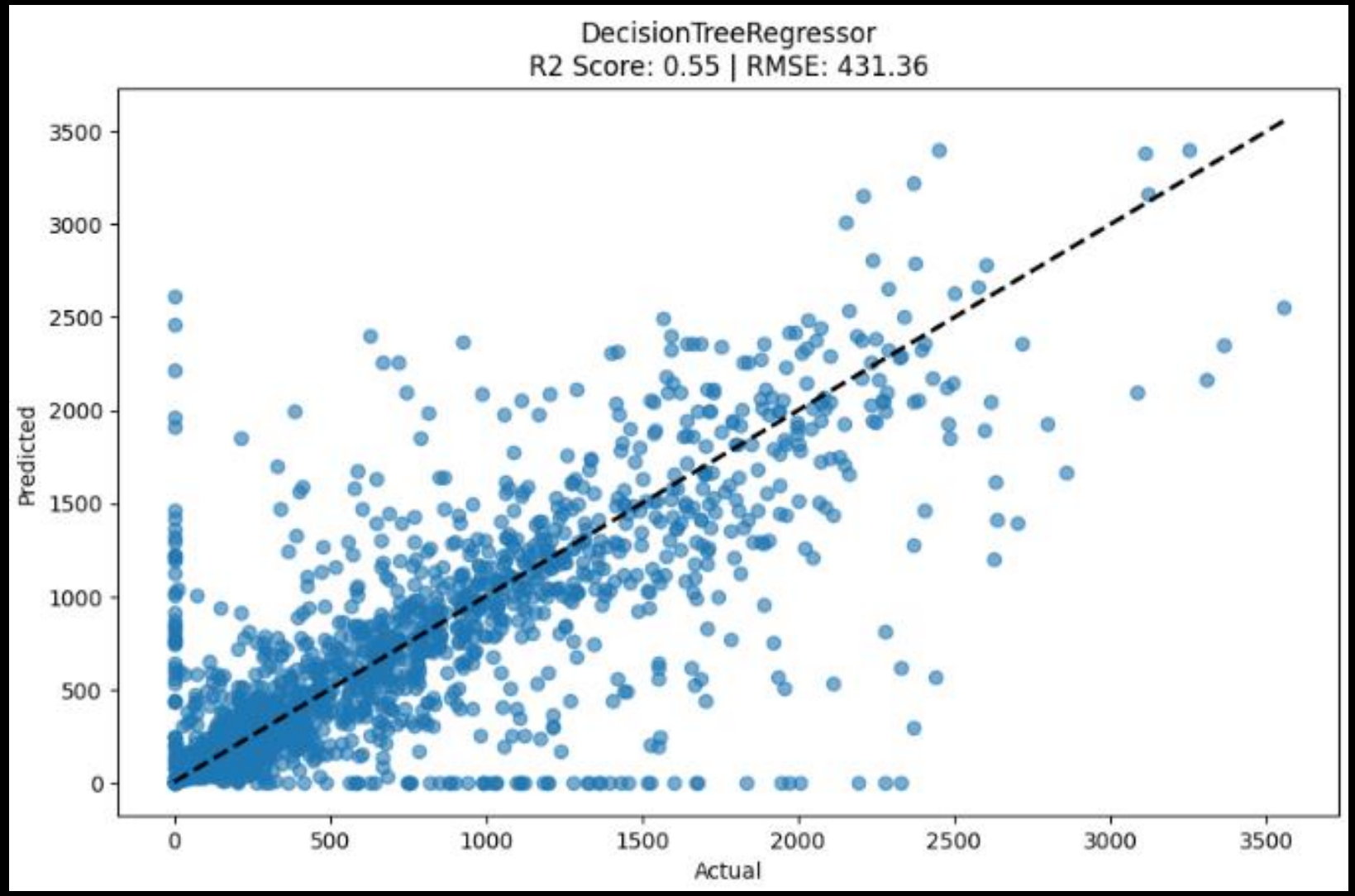
RandomForest Regressor Model, score 0,77

The best models



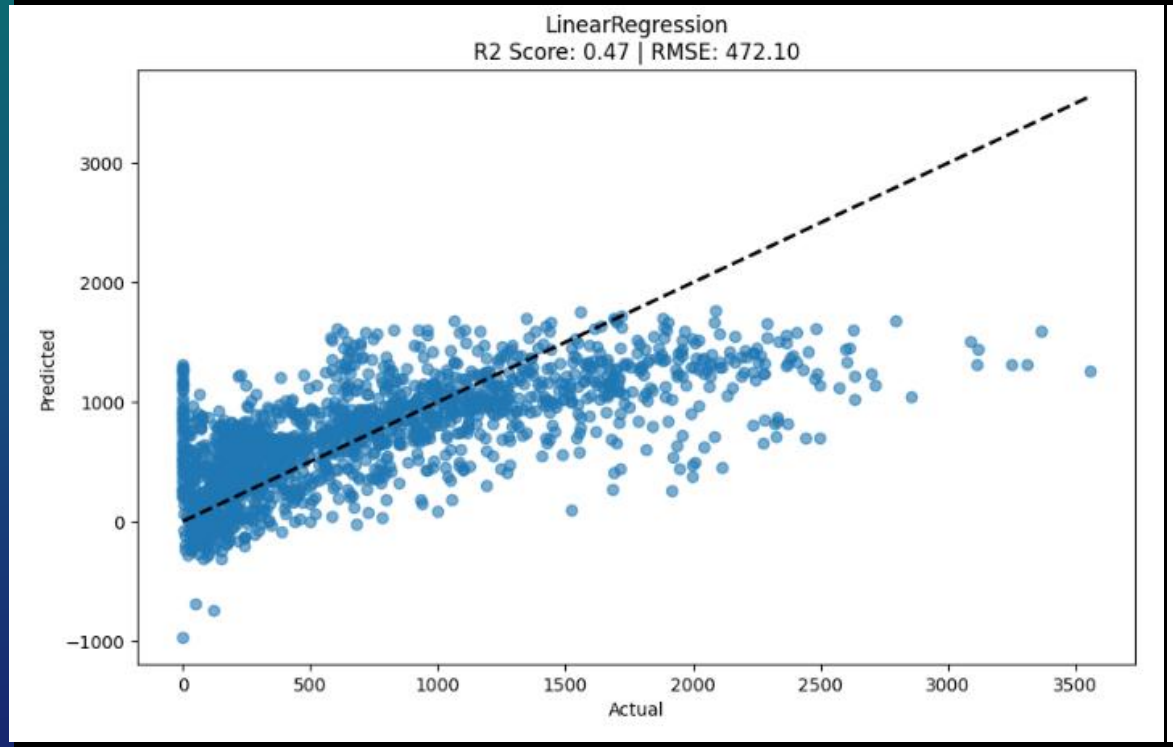
XGB Regressor Model, score 0,76

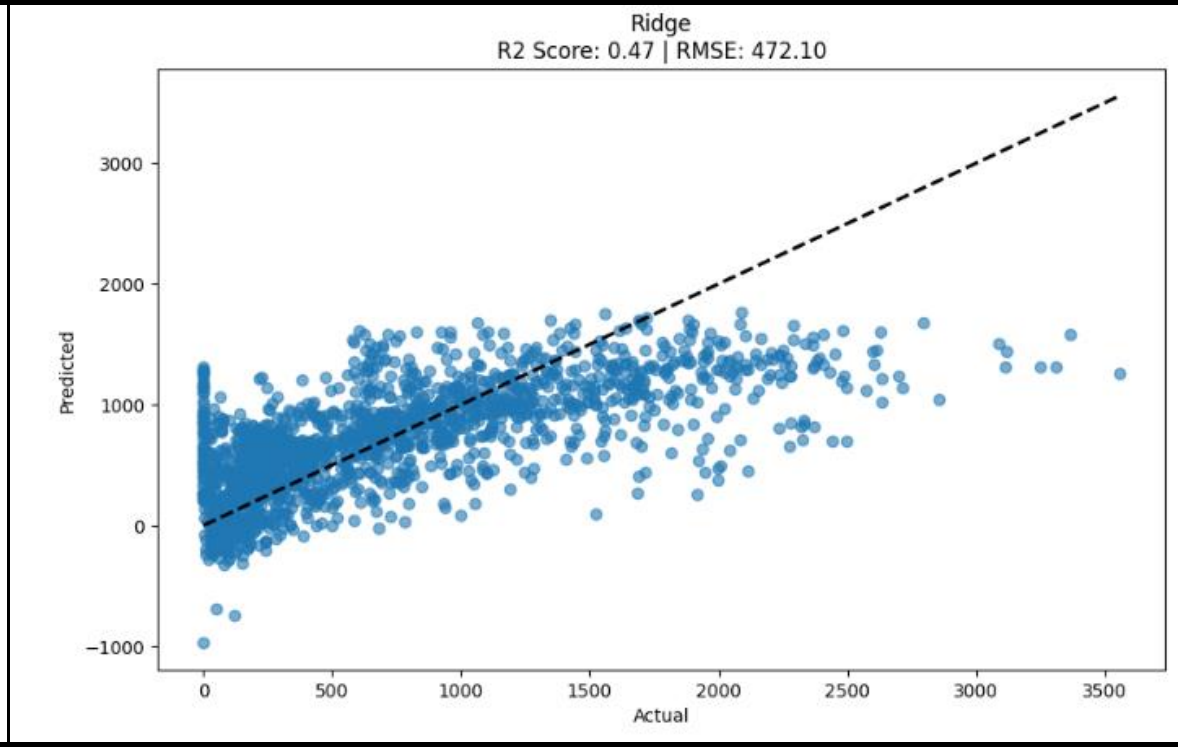
The medium models



Decision Tree Regressor Model, Score 0,55

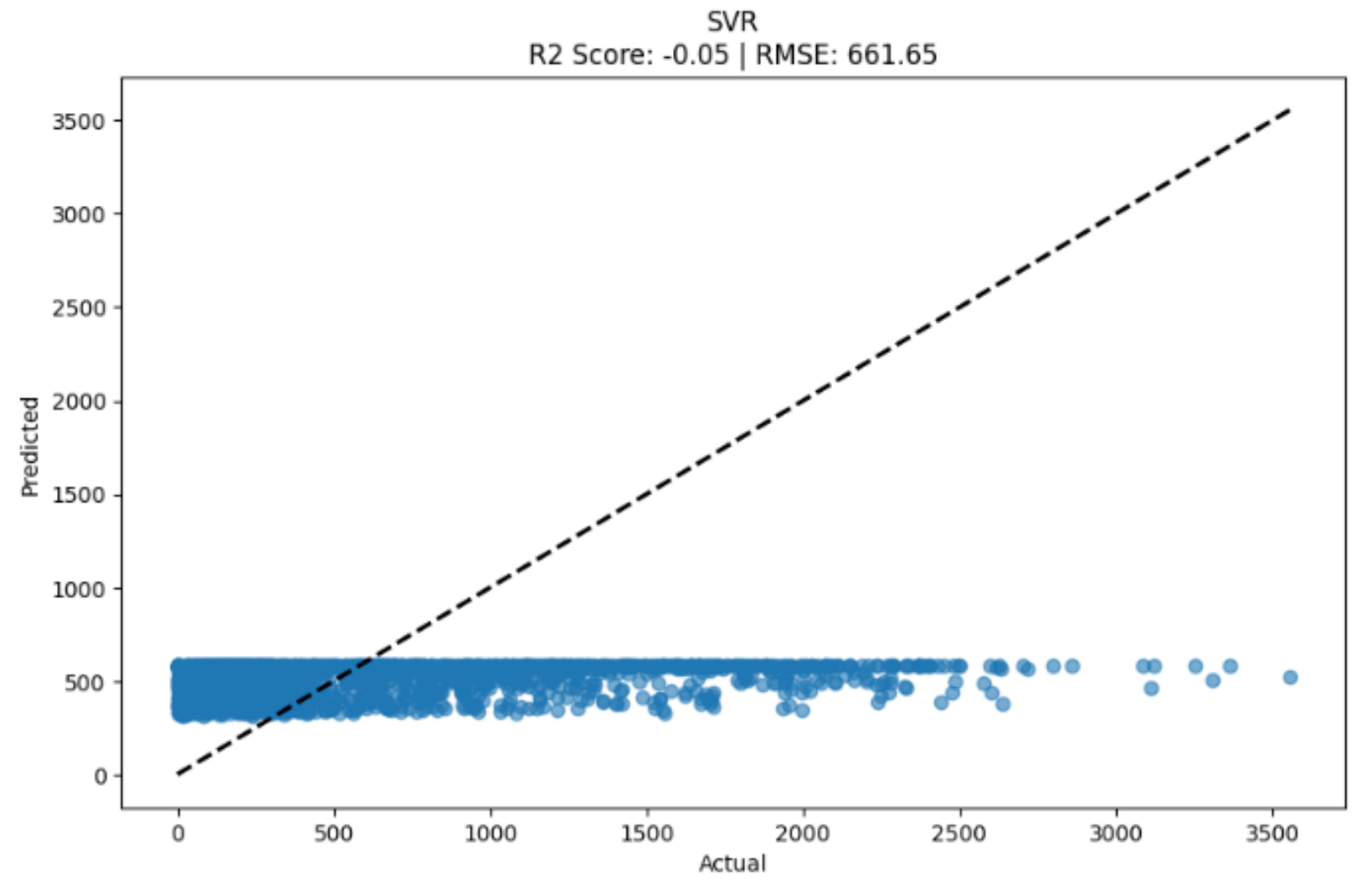
The medium models





Linear Regression and Ridge Models, both score 0,47

The worst Model



- SVR model, score -0,05, the only negative score for all the models we tested

In Conclusion

- accurate Model: R2 0,78
- Predicts well using LGBM Regressor Model
- Helping the company understand and manage their rents and items