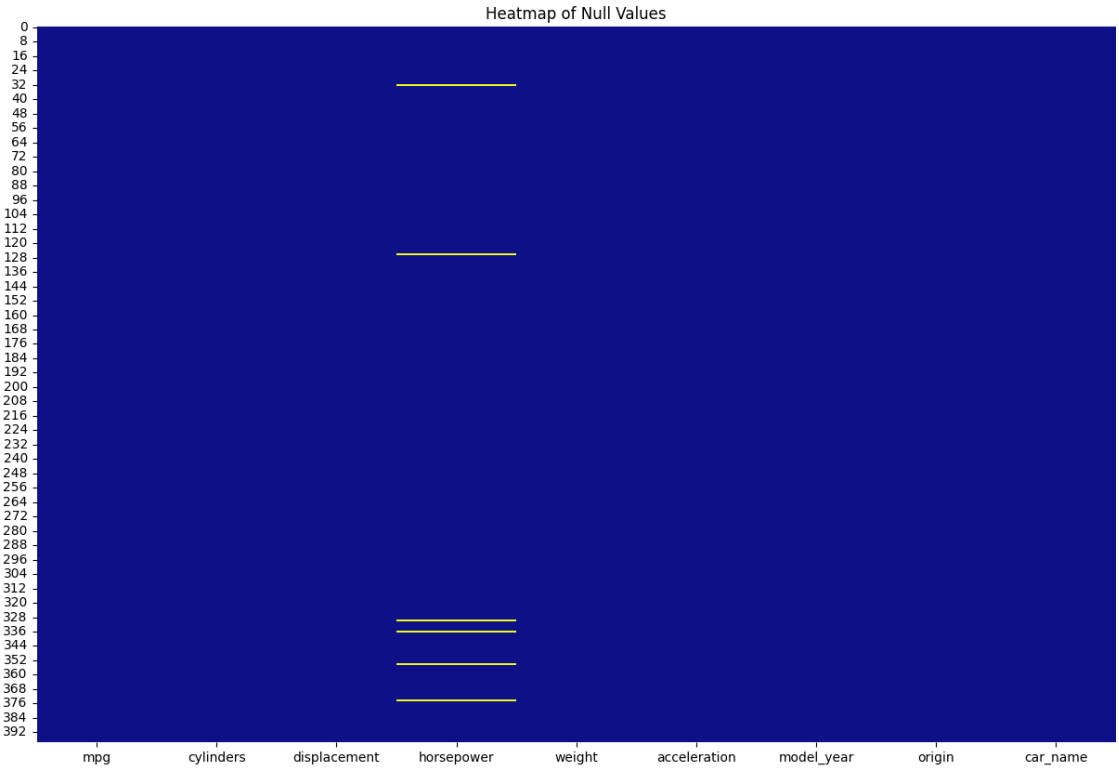
Data Preprocessing And EDA

1. **Auto MPG**

Dataset data types

Column	Data Type
mpg	float64
cylinders	int64
displacement	float64
horsepower	float64
weight	float64
acceleration	float64
model_year	int64
origin	int64
car_name	object

Handling Missing data –



Initial shape with missing values: (398, 9)

Final shape without missing values: (392, 9)

Rows with missing values removed: 6

Mean mpg by brand –

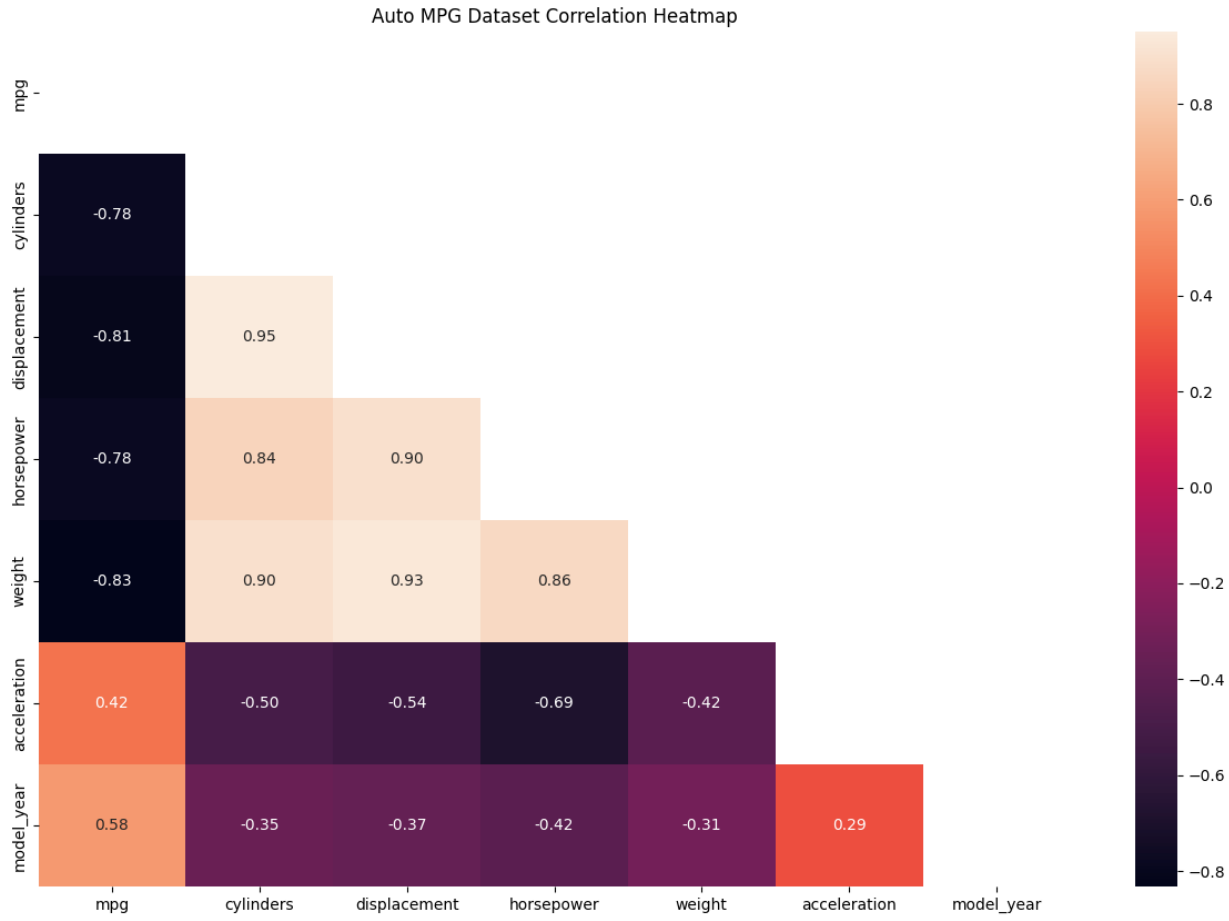
amc	18.070370
audi	26.714286
bmw	23.750000
buick	19.182353
Cadillac	19.750000
Capri	25.000000
Chevrolet	20.219149
Chrysler	17.266667
Datsun	31.113043
Dodge	22.060714
Fiat	28.912500
Ford	19.475000
Honda	33.761538
Mazda	30.058333
mercedes-benz	23.966667
mercury	19.118182
Nissan	36.000000
Oldsmobile	21.100000
Opel	25.750000
Peugeot	23.687500
Plymouth	21.703226
Pontiac	20.012500
Renault	29.666667
Saab	23.900000
Subaru	30.525000
Toyota	28.165385
Triumph	35.000000
Unknown	9.000000
Volkswagen	31.840909
volvo	21.116667

Mean mpg by cylinders:

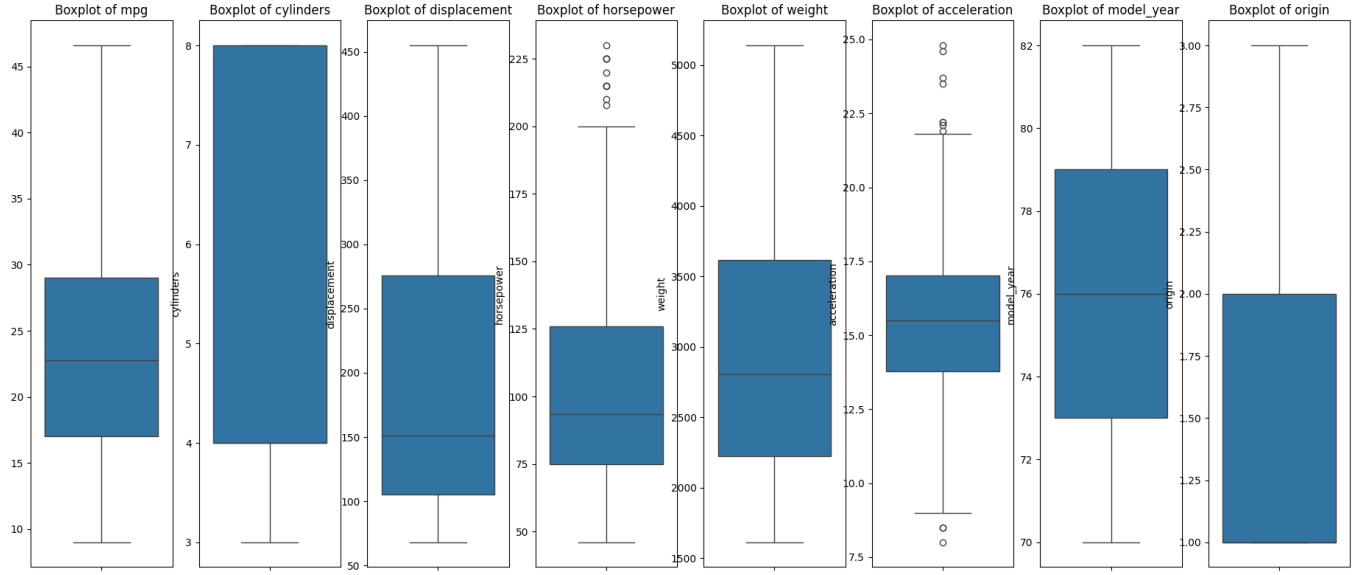
3	20.550000
4	29.283920
5	27.366667
6	19.973494
8	14.963107

Mean mpg by model_year:

70	17.689655
71	21.111111
72	18.714286
73	17.100000
74	22.769231
75	20.266667
76	21.573529
77	23.375000
78	24.061111
79	25.093103
80	33.803704
81	30.185714
82	32.000000



Outlier Analysis



Summary of Statistics:

Feature	Count	Mean	Std Dev	Min	25%	Median
mpg	392	23.45	7.81	9.0	17.00	22.75
cylinders	392	5.47	1.71	3.0	4.00	4.00
displacement	392	194.41	104.64	68.0	105.00	151.00
horsepower	392	104.47	38.49	46.0	75.00	93.50
weight	392	2977.58	849.40	1613	2225.25	2803.5
acceleration	392	15.54	2.76	8.0	13.78	15.50
model_year	392	75.98	3.68	70.0	73.00	76.00

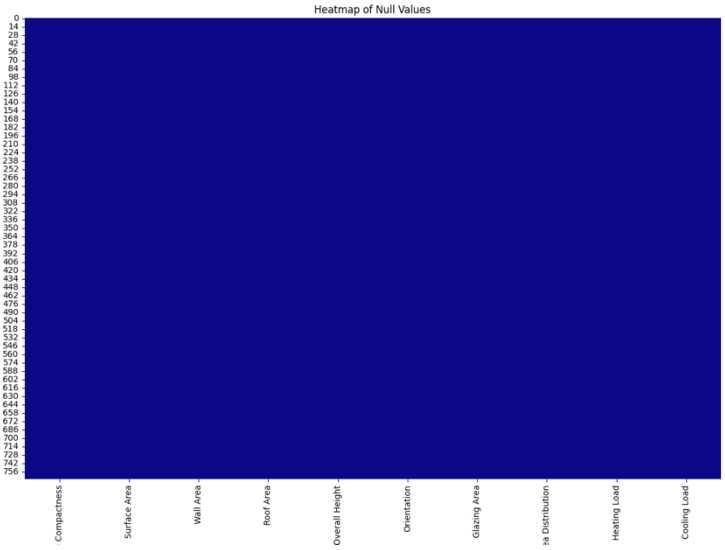
Feature	75%	Max	Kurtosis	Skewness
mpg	29.00	46.6	-0.516	0.457
cylinders	8.00	8.0	-1.398	0.508

displacement	275.75	455.0	-0.778	0.702
horsepower	126.00	230.0	0.697	1.087
weight	3614.75	5140	-0.809	0.520
acceleration	17.03	24.8	0.444	0.292
model_year	79.00	82.0	-1.167	0.020

2. **Energy Efficiency Data**

Feature	**Data Type**
Relative Compactness	float64
Surface Area	float64
Wall Area	float64
Roof Area	float64
Overall Height	float64
Orientation	int64
Glazing Area	float64
Glazing Area Distribution	int64
Heating Load	float64
Cooling Load	float64

Handling Missing Data –



Initial shape with missing values: (768, 10)

Final shape without missing values: (768, 10)

Rows with missing values removed: 0

By Relative Compactness

Relative Compactness	Heating Load	Cooling Load
0.62	14.28	15.24
0.64	16.62	20.23
0.66	12.82	15.87
0.69	12.39	15.24
0.71	12.04	15.04
0.74	11.89	14.81
0.76	35.66	36.41
0.79	38.61	40.24

0.82	25.56	28.03
0.86	28.55	30.91
0.90	31.63	33.82
0.98	27.65	29.22

By Surface Area

Surface Area	Heating Load	Cooling Load
514.5	27.65	29.22
563.5	31.63	33.82
588.0	28.55	30.91
612.5	25.56	28.03
637.0	38.61	40.24
661.5	35.66	36.41
686.0	11.89	14.81
710.5	12.04	15.04
735.0	12.39	15.24
759.5	12.82	15.87
784.0	16.62	20.23
808.5	14.28	15.24

By Wall Area

Wall Area	Heating Load	Cooling Load
245.0	11.89	14.81
269.5	12.04	15.04
294.0	22.86	25.12
318.5	23.33	25.91

343.0	27.61	30.23
367.5	14.28	15.24
416.5	35.66	36.41

By Roof Area

Roof Area	Heating Load	Cooling Load
110.25	27.65	29.22
122.50	33.65	35.11
147.00	30.91	33.06
220.50	13.34	16.07

By Overall Height

Overall Height	Heating Load	Cooling Load
3.5	13.34	16.07
7.0	31.28	33.10

By Orientation

Orientation	Heating Load	Cooling Load
2	22.31	24.60
3	22.38	24.31
4	22.26	24.48
5	22.28	24.95

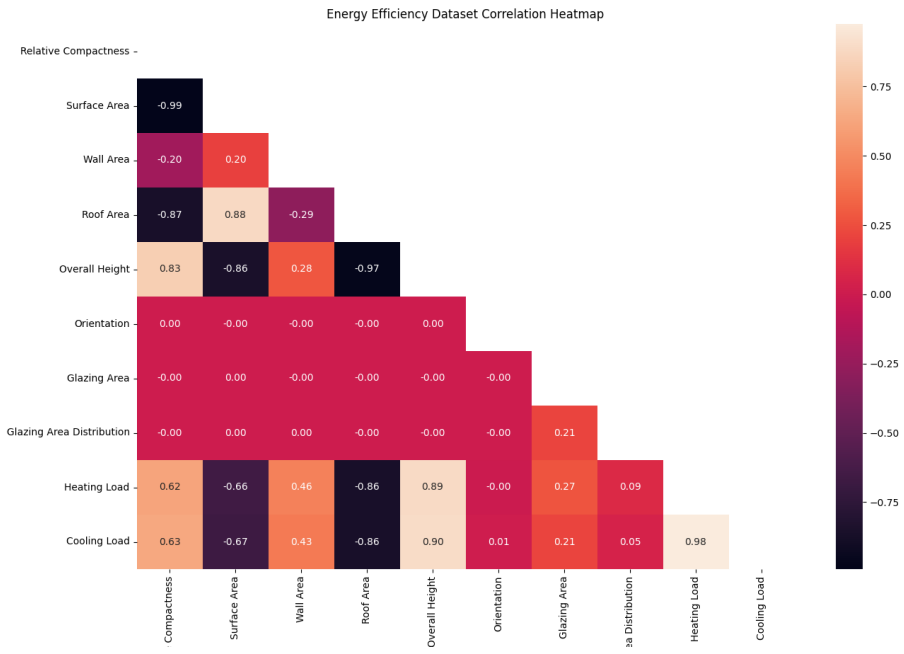
By Glazing Area

Glazing Area	Heating Load	Cooling Load
0.00	14.29	19.71

0.10	20.36	22.94
0.25	22.76	24.89
0.40	25.41	26.91

By Glazing Area Distribution

Glazing Area Distribution	Heating Load	Cooling Load
0	14.29	19.71
1	23.03	25.18
2	22.93	25.00
3	22.68	24.66
4	22.89	25.02
5	22.68	24.72



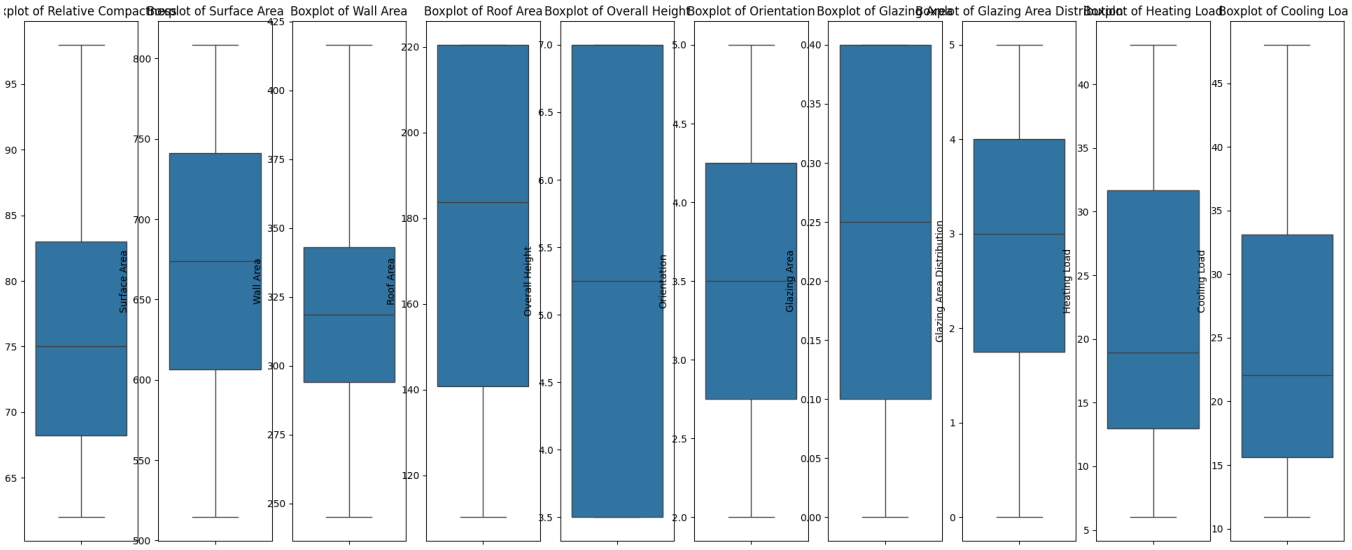


Table: Summary Statistics – Energy Efficiency Dataset

Feature	Count	Mean	Std Dev	Min	25%	50%
Relative Compactness	768	0.764	0.106	0.62	0.683	0.75
Surface Area	768	671.71	88.09	514.5	606.38	673.75
Wall Area	768	318.50	43.63	245.0	294.00	318.50
Roof Area	768	176.60	45.17	110.3	140.88	183.75
Overall Height	768	5.25	1.75	3.50	3.50	5.25
Orientation	768	3.50	1.12	2.00	2.75	3.50
Glazing Area	768	0.234	0.133	0.00	0.10	0.25
Glazing Area Distribution	768	2.81	1.55	0.00	1.75	3.00
Heating Load	768	22.31	10.09	6.01	12.99	18.95
Cooling Load	768	24.59	9.51	10.90	15.62	22.08

Feature	75%	Max	Kurtosis	Skewness
Relative Compactness	0.8300	0.98	-0.707	0.496
Surface Area	741.13	808.50	-1.059	-0.125

Wall Area	343.00	416.50	0.117	0.533
Roof Area	220.50	220.50	-1.777	-0.163
Overall Height	7.00	7.00	-2.005	0.000
Orientation	4.25	5.00	-1.361	0.000
Glazing Area	0.40	0.40	-1.328	-0.060
Glazing Area Distribution	4.00	5.00	-1.149	-0.089
Heating Load	31.67	43.10	-1.246	0.360
Cooling Load	33.13	48.03	-1.147	0.396

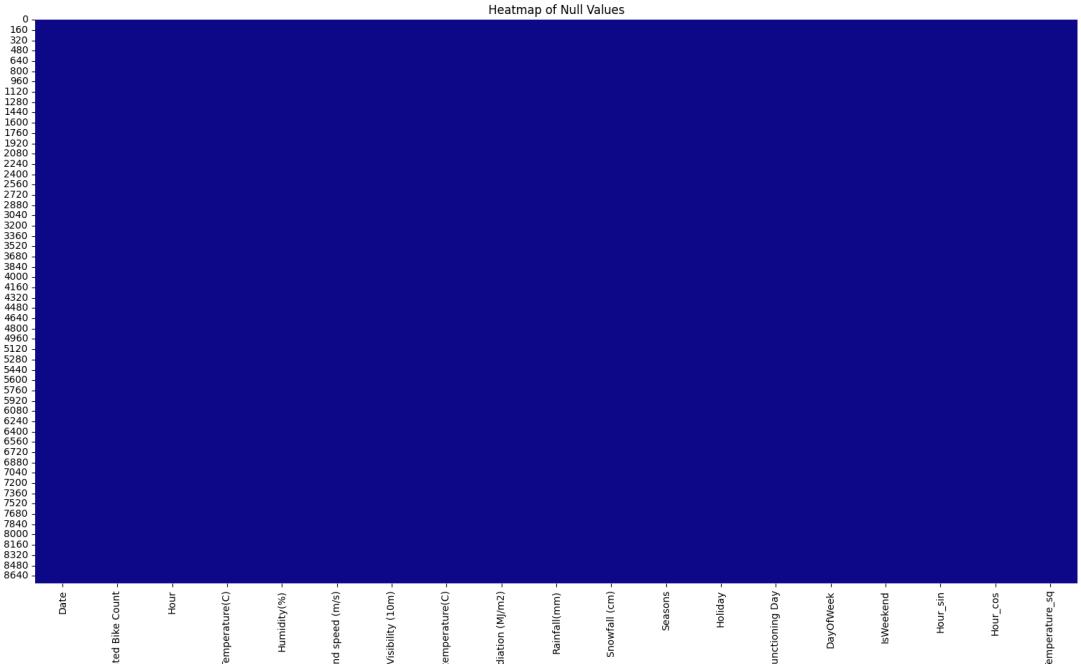
3. Seoul Bike Sharing

Dataset – Data Types

Column	**Data Type**
Date	datetime64[ns]
Rented Bike Count	int64
Hour	int64
Temperature (°C)	float64
Humidity (%)	int64
Wind speed (m/s)	float64
Visibility (10m)	int64
Dew point temperature (°C)	float64
Solar Radiation (MJ/m²)	float64
Rainfall (mm)	float64
Snowfall (cm)	float64
Seasons	object
Holiday	object

Functioning Day	object
DayOfWeek	int32
IsWeekend	int64
Hour_sin	float64
Hour_cos	float64
Temperature_sq	float64

Handling Missing Values



Initial shape with missing values: (8760, 19)

Final shape without missing values: (8760, 19)

Rows with missing values removed: 0

Mean Rented Bike Count by Seasons & Holiday

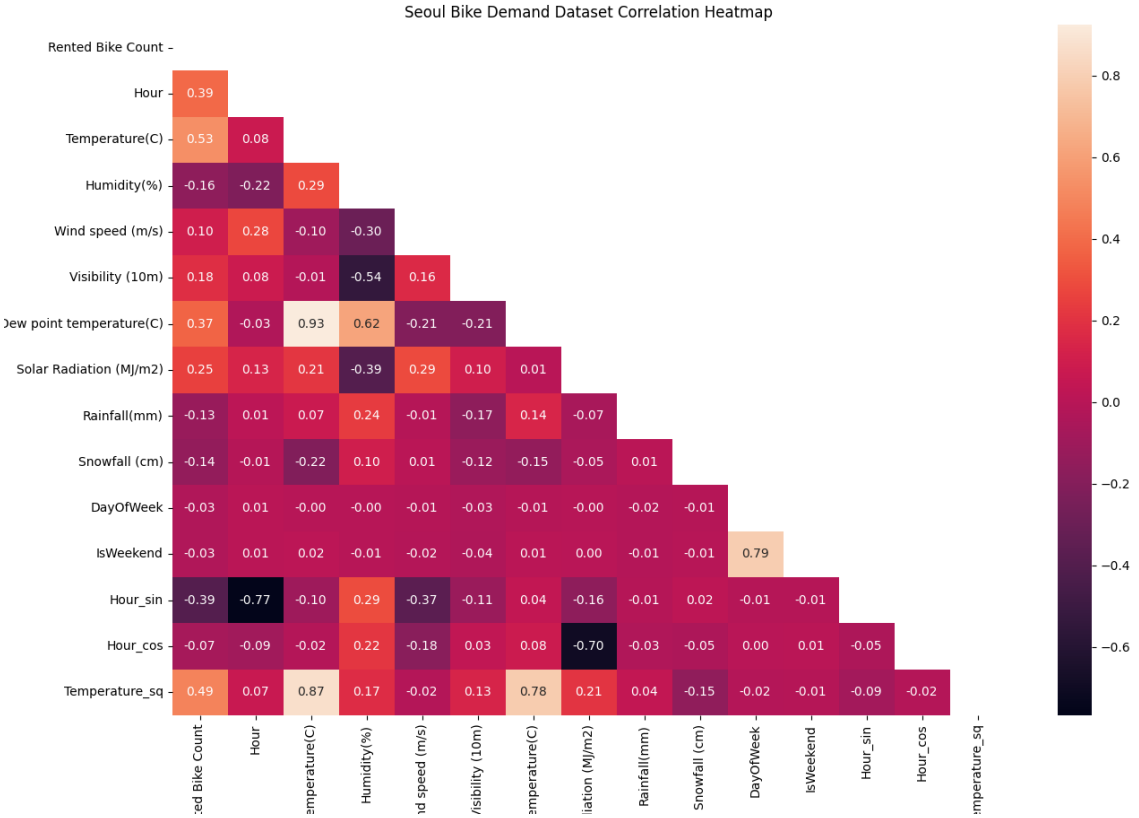
Category	Mean Rented Bike Count
Autumn	767.53
Spring	658.60
Summer	947.62
Winter	224.45

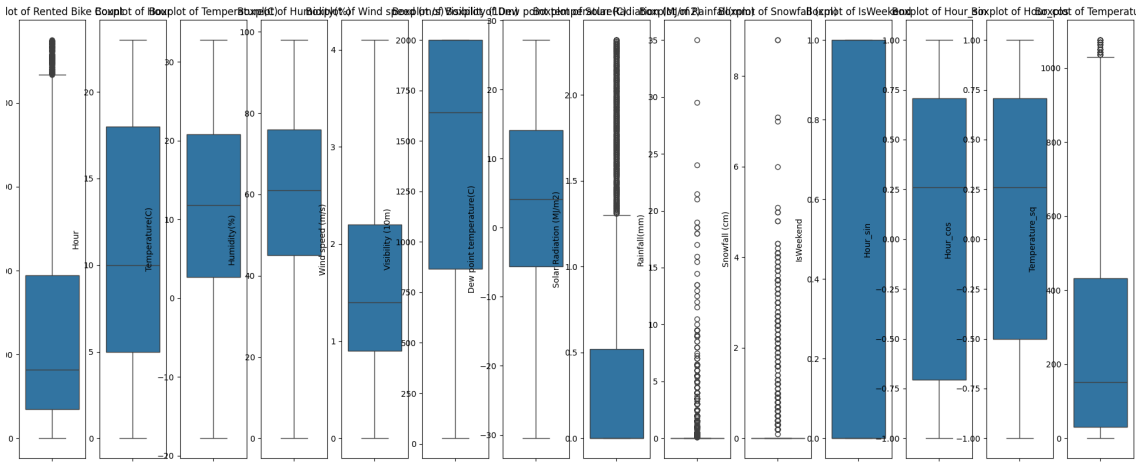
Holiday Status	Mean Rented Bike Count
Holiday	455.80
No Holiday	643.59

Mean Rented Bike Count by Hour of Day

Hour	Mean Count	Hour	Mean Count
0	544.17	12	547.80
1	427.87	13	588.78
2	301.92	14	619.02
3	203.90	15	746.23
4	133.33	16	893.80
5	139.31	17	1125.95
6	288.26	18	1042.29
7	609.71	19	1031.33
8	1001.01	20	1001.57
9	647.76	21	1009.33
10	501.87	22	926.22

11	511.36	23	673.18





Summary Statistics

Feature	Count	Mean	Std Dev	Min	25%	50%
Rented Bike Count	7628	634.24	593.76	0.00	173.00	409.00
Hour	7628	11.08	7.22	0.00	5.00	10.00
Temperature (°C)	7628	11.45	11.36	-17.8	2.70	11.85
Humidity (%)	7628	60.51	20.23	0.00	45.00	61.00
Wind speed (m/s)	7628	1.58	0.92	0.00	0.90	1.40
Visibility (10m)	7628	1405.20	622.81	27.0	866.00	1639.50
Dew point temp (°C)	7628	3.40	13.06	-30.5	-5.60	4.10
Solar Radiation (MJ/m²)	7628	0.36	0.61	0.00	0.00	0.00

Rainfall (mm)	7628	0.17	1.20	0.00	0.00	0.00
Snowfall (cm)	7628	0.08	0.47	0.00	0.00	0.00

Feature	**75%**	**Max**	**Kurtosis**	**Skewness**
Rented Bike Count	971.00	2375.00	0.293	1.095
Hour	18.00	23.00	-1.299	0.135
Temperature (°C)	20.80	32.80	-0.865	-0.195
Humidity (%)	76.00	98.00	-0.767	-0.082
Wind speed (m/s)	2.20	4.10	-0.232	0.669
Visibility (10m)	2000.00	2000.00	-1.110	-0.612
Dew point temp (°C)	14.10	27.20	-0.799	-0.297
Solar Radiation (MJ/m²)	0.52	2.32	1.660	1.679
Rainfall (mm)	0.00	35.00	254.61	13.732
Snowfall (cm)	0.00	8.80	82.683	7.947

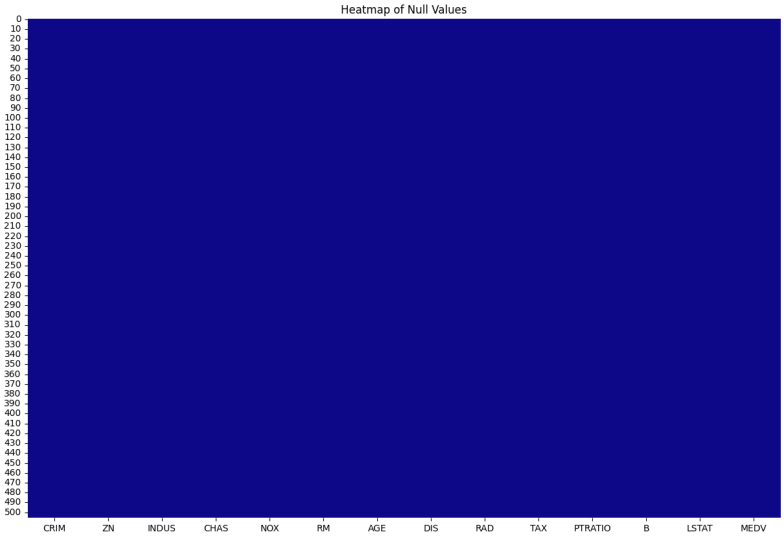
DayOfWeek	5.00	6.00	-1.246	-0.008
IsWeekend	1.00	1.00	-1.115	0.941
Hour_sin	0.71	1.00	-1.513	-0.137
Hour_cos	0.71	1.00	-1.446	-0.166
Temperature_sq	432.64	1075.84	-0.006	1.005

4. Boston Housing Dataset

Dataset – Data Types

Column	Data Type
CRIM	float64
ZN	float64
INDUS	float64
CHAS	int64
NOX	float64
RM	float64
AGE	float64
DIS	float64
RAD	int64
TAX	float64
PTRATIO	float64
B	float64
LSTAT	float64
MEDV	float64

Handling Missing Values :



Initial shape with missing values: (506, 14)

Final shape without missing values: (506, 14)

Rows with missing values removed: 0

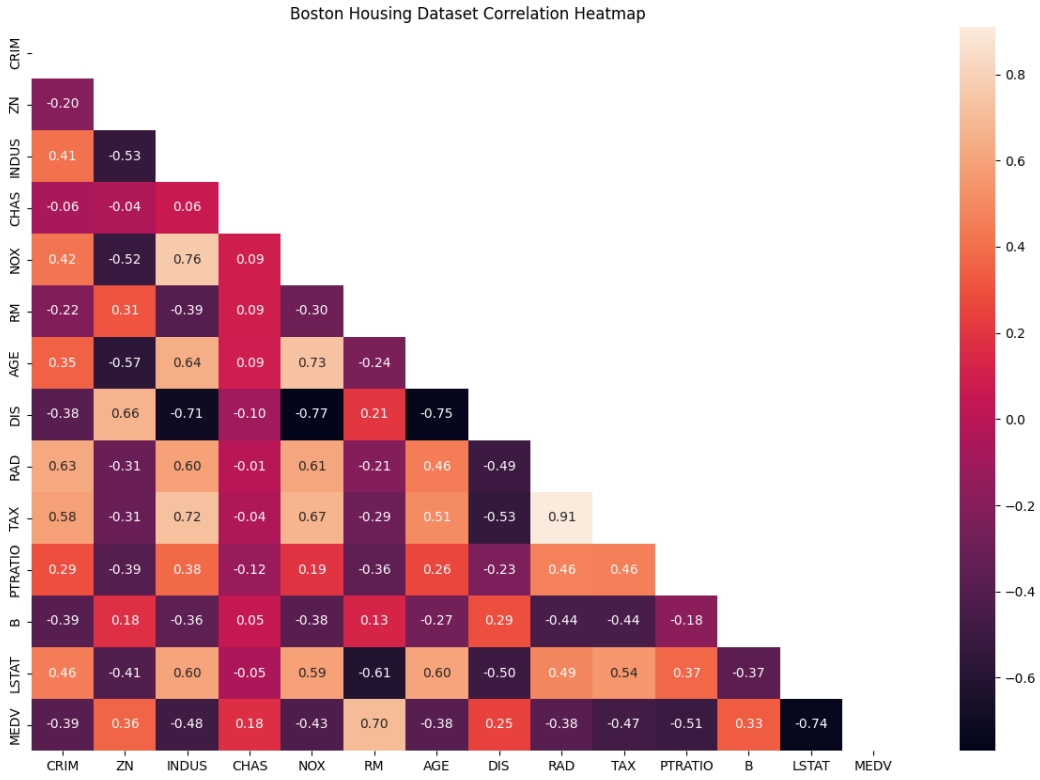
Mean MEDV by CHAS

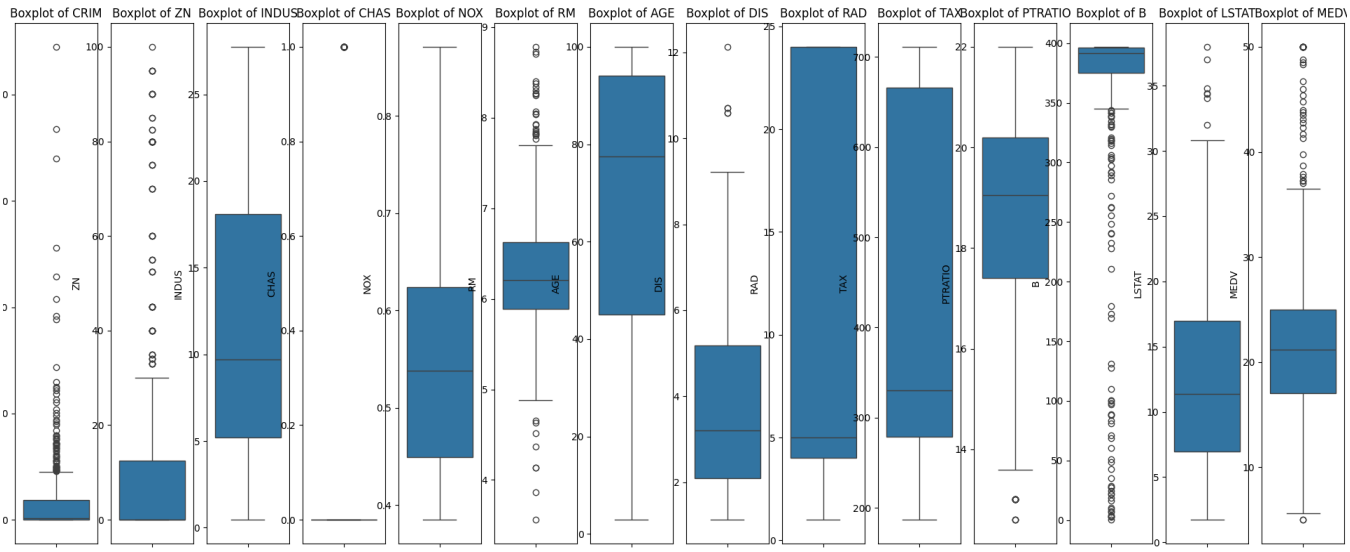
CHAS	Mean MEDV
0	22.09
1	28.44

Mean MEDV by RAD

RAD	Mean MEDV
1	24.37
2	26.83

3	27.93
4	21.39
5	25.71
6	20.98
7	27.11
8	30.36
24	16.40





Summary Statistics

Feature	Count	Mean	Std Dev	Min	25%	50%
CRIM	506	3.61	8.60	0.01	0.08	0.26
ZN	506	11.36	23.32	0.00	0.00	0.00
INDUS	506	11.14	6.86	0.46	5.19	9.69
CHAS	506	0.07	0.25	0.00	0.00	0.00
NOX	506	0.55	0.12	0.39	0.45	0.54
RM	506	6.28	0.70	3.56	5.89	6.21
AGE	506	68.57	28.15	2.90	45.03	77.50
DIS	506	3.80	2.11	1.13	2.10	3.21
RAD	506	9.55	8.71	1.00	4.00	5.00
TAX	506	408.24	168.54	187.0	279.0	330.0
PTRATIO	506	18.46	2.16	12.6	17.4	19.05
B	506	356.67	91.29	0.32	375.38	391.44
LSTAT	506	12.65	7.14	1.73	6.95	11.36
MEDV	506	22.53	9.20	5.00	17.03	21.20

Feature	75%	Max	Kurtosis	Skewness
CRIM	3.68	88.98	37.13	5.22
ZN	12.50	100.0	4.03	2.23
INDUS	18.10	27.74	-1.23	0.30
CHAS	0.00	1.00	9.64	3.41
NOX	0.62	0.87	-0.06	0.73
RM	6.62	8.78	1.89	0.40
AGE	94.08	100.0	-0.97	-0.60
DIS	5.19	12.13	0.49	1.01
RAD	24.00	24.00	-0.87	1.00
TAX	666.0	711.0	-1.14	0.67
PTRATIO	20.20	22.00	-0.29	-0.80
B	396.23	396.9	7.23	-2.89
LSTAT	16.96	37.97	0.49	0.91
MEDV	25.00	50.00	1.50	1.11

Quality of Fit

1. Linear Regression

Dataset	Split	RMSE	MSE	R^2
Auto MPG	Train	3.31350	10.97926	0.82600
Auto MPG	Test	3.27275	10.71086	0.79015
Auto MPG	CV	3.37577	11.49514	0.81009
Energy Efficiency – Heating	Train	1.72897	2.98933	0.97038
Energy Efficiency – Heating	Test	1.94472	3.78194	0.96372
Energy Efficiency – Heating	CV	1.78563	3.20675	0.96831
Energy Efficiency – Cooling	Train	1.91126	3.65293	0.95926
Energy Efficiency – Cooling	Test	2.03965	4.16017	0.95510
Energy Efficiency – Cooling	CV	1.78563	3.20675	0.96831
Seoul Bike Sharing	Train	390.84837	152762.45167	0.57036
Seoul Bike Sharing	Test	368.30763	135650.51203	0.60074
Seoul Bike Sharing	CV	387.33403	150143.39919	0.57387
Boston Housing	Train	4.65203	21.64141	0.75089
Boston Housing	Test	4.92860	24.29112	0.66876
Boston Housing	CV	4.84279	23.48860	0.71522

2. Ridge regression

Dataset	Split	RMSE	MSE	R^2
Auto MPG	Train	3.31351	10.97934	0.82600
Auto MPG	Test	3.27151	10.70278	0.79031

Auto MPG	CV	3.37538	11.49302	0.81013
Energy Efficiency – Heating	Train	2.96290	8.77876	0.91302
Energy Efficiency – Heating	Test	3.10700	9.65346	0.90739
Energy Efficiency – Heating	CV	3.01585	9.11315	0.90957
Energy Efficiency – Cooling	Train	3.27517	10.72675	0.88036
Energy Efficiency – Cooling	Test	3.21641	10.34530	0.88835
Energy Efficiency – Cooling	CV	3.28442	10.81190	0.87954
Seoul Bike Sharing	Train	390.84957	152,763.38810	0.57036
Seoul Bike Sharing	Test	368.26568	135,619.60752	0.60083
Seoul Bike Sharing	CV	387.33181	150,141.89174	0.57388
Boston Housing	Train	4.67163	21.82412	0.74878
Boston Housing	Test	4.94744	24.47719	0.66622
Boston Housing	CV	4.86681	23.71106	0.71222

3. Lasso Regression

Auto MPG

Outliers	**Split**	**RMSE**	**MSE**	**R^2**
Included	Train	3.32287	11.04147	0.82502
Included	Test	3.27933	10.75401	0.78930
Included	CV	3.38156	11.55546	0.80944
Excluded (Z-score)	Train	3.29274	10.84212	0.82165

Excluded (Z-score)	Test	3.19319	10.19649	0.80882
Excluded (Z-score)	CV	3.33598	11.14766	0.80962
Excluded (IQR)	Train	3.28707	10.80482	0.81451
Excluded (IQR)	Test	2.96876	8.81355	0.84535
Excluded (IQR)	CV	3.29364	10.89574	0.80959

Energy Efficiency - Heating

Outliers	**Split**	**RMSE**	**MSE**	**R^2**
Included	Train	3.06604	9.40061	0.90685
Included	Test	3.17633	10.08905	0.90321
Included	CV	3.12162	9.76459	0.90300
Excluded (Z-score)	Train	3.06604	9.40061	0.90685
Excluded (Z-score)	Test	3.17633	10.08905	0.90321
Excluded (Z-score)	CV	3.12162	9.76459	0.90300
Excluded (IQR)	Train	3.06604	9.40061	0.90685
Excluded (IQR)	Test	3.17633	10.08905	0.90321
Excluded (IQR)	CV	3.12162	9.76459	0.90300

Energy Efficiency — Cooling

Outliers	**Split**	**RMSE**	**MSE**	**R^2**
Included	Train	3.37523	11.39216	0.87294
Included	Test	3.32276	11.04076	0.88084
Included	CV	3.38818	11.51477	0.87167
Excluded (Z-score)	Train	3.37523	11.39216	0.87294
Excluded (Z-score)	Test	3.32276	11.04076	0.88084
Excluded (Z-score)	CV	3.38818	11.51477	0.87167

Excluded (IQR)	Train	3.37523	11.39216	0.87294
Excluded (IQR)	Test	3.32276	11.04076	0.88084
Excluded (IQR)	CV	3.38818	11.51477	0.87167

Seoul Bike Sharing

Outliers	**Split**	**RMSE**	**MSE**	**R^2**
Included	Train	390.85081	152,764.35702	0.57035
Included	Test	368.25523	135,611.91605	0.60085
Included	CV	387.33332	150,142.93099	0.57388
Excluded (Z-score)	Train	388.93936	151,273.82748	0.56897
Excluded (Z-score)	Test	381.87741	145,830.35997	0.59258
Excluded (Z-score)	CV	388.67931	151,102.22929	0.57088
Excluded (IQR)	Train	391.83377	153,533.70690	0.55774
Excluded (IQR)	Test	393.34848	154,723.02335	0.55112
Excluded (IQR)	CV	393.25363	154,706.26373	0.55368

Boston Housing

Outliers	**Split**	**RMSE**	**MSE**	**R^2**
Included	Train	4.76863	22.73981	0.73824
Included	Test	5.01554	25.15559	0.65697
Included	CV	4.97119	24.72610	0.69986
Excluded (Z-score)	Train	3.63314	13.19972	0.77065
Excluded (Z-score)	Test	5.87271	34.48867	0.57929
Excluded (Z-score)	CV	4.24526	18.81221	0.70131
Excluded (IQR)	Train	2.75586	7.59474	0.73432

Excluded (IQR)	Test	2.68575	7.21326	0.77935
Excluded (IQR)	CV	2.92209	8.76718	0.68983

4. Bridge Regression

Dataset	**Split**	**MSE**	**RMSE**	**R^2**
Auto MPG	Train	11.93054	3.45406	0.81093
Auto MPG	Test	11.92606	3.45341	0.76634
Auto MPG	5-fold CV (mean)	12.32304	3.49956	0.79547
Energy Efficiency	Train	4.14060	2.03485	0.95897
Energy Efficiency	Test	5.06664	2.25092	0.95139
Energy Efficiency	5-fold CV (mean)	4.35648	2.08308	0.95688
Energy Efficiency	Train	6.03619	2.45687	0.93268
Energy Efficiency	Test	6.55260	2.55980	0.92928
Energy Efficiency	5-fold CV (mean)	6.15749	2.47781	0.93149
Seoul Bike	Train	155072.42405	393.79236	0.56386
Seoul Bike	Test	137268.42961	370.49754	0.59597
Seoul Bike	5-fold CV (mean)	152257.15924	390.05669	0.56791
Boston Housing	Train	25.00473	5.00047	0.71217
Boston Housing	Test	28.59227	5.34717	0.61011
Boston Housing	5-fold CV (mean)	27.02414	5.19069	0.67358

5. Transformed Regression

Auto MPG

Technique	MSE	R2	RMSE
Logarithmic Regression (In-Sample)	8.79549	0.85525	2.96572
Logarithmic Regression (Train)	9.03284	0.85685	3.00547
Logarithmic Regression (Test)	7.69540	0.84923	2.77406
Logarithmic Regression (5X Validation)	0.01323	0.87360	0.11425
Square Root Regression (In-Sample)	9.49663	0.84371	3.08166
Square Root Regression (Train)	9.64170	0.84720	3.10511
Square Root Regression (Test)	8.77465	0.82808	2.96220
Square Root Regression (5X Validation)	0.09506	0.85233	0.30654
Box-Cox Regression (In-Sample)	8.97803	0.85224	2.99634
Box-Cox Regression (Train)	9.17444	0.85460	3.02893
Box-Cox Regression (Test)	7.99014	0.84346	2.82668
Box-Cox Regression (5X Validation)	0.04952	0.86787	0.22114
Yeo-Johnson Regression (In-Sample)	8.99253	0.85201	2.99876
Yeo-Johnson Regression (Train)	9.18992	0.85436	3.03149
Yeo-Johnson Regression (Test)	8.00071	0.84325	2.82855
Yeo-Johnson Regression (5X Validation)	0.03777	0.86768	0.19314

Energy Efficiency — Heating Load

Technique	MSE	R2	RMSE
Logarithmic Regression (In-Sample)	4.33421	0.95737	2.08188
Logarithmic Regression (Train)	4.18235	0.95856	2.04508
Logarithmic Regression (Test)	5.08235	0.95124	2.25441
Logarithmic Regression (5X Validation)	0.00694	0.96619	0.08267
Square Root Regression (In-Sample)	3.17045	0.96882	1.78057
Square Root Regression (Train)	3.01110	0.97016	1.73525
Square Root Regression (Test)	3.85364	0.96303	1.96307
Square Root Regression (5X Validation)	0.03047	0.97355	0.17398
Box-Cox Regression (In-Sample)	3.88411	0.96180	1.97081
Box-Cox Regression (Train)	3.90643	0.96129	1.97647
Box-Cox Regression (Test)	4.80550	0.95390	2.19215
Box-Cox Regression (5X Validation)	0.01807	0.96889	0.13357
Yeo-Johnson Regression (In-Sample)	3.86420	0.96200	1.96576
Yeo-Johnson Regression (Train)	3.88805	0.96148	1.97181
Yeo-Johnson Regression (Test)	4.78202	0.95412	2.18678
Yeo-Johnson Regression (5X Validation)	0.01311	0.96947	0.11381

Energy Efficiency — Cooling Load

Technique	MSE	R2	RMSE
Logarithmic Regression (In-Sample)	4.66804	0.94835	2.16056
Logarithmic Regression (Train)	4.57924	0.94893	2.13992
Logarithmic Regression (Test)	5.11724	0.94477	2.26213

Logarithmic Regression (5X Validation)	0.00582	0.95858	0.07628
Square Root Regression (In-Sample)	4.05452	0.95514	2.01358
Square Root Regression (Train)	3.96652	0.95576	1.99161
Square Root Regression (Test)	4.50296	0.95140	2.12202
Square Root Regression (5X Validation)	0.03467	0.96186	0.18602
Box-Cox Regression (In-Sample)	4.98549	0.94484	2.23282
Box-Cox Regression (Train)	5.08398	0.94330	2.25477
Box-Cox Regression (Test)	5.60525	0.93951	2.36754
Box-Cox Regression (5X Validation)	0.00342	0.95587	0.05847
Yeo-Johnson Regression (In-Sample)	4.95374	0.94519	2.22570
Yeo-Johnson Regression (Train)	5.05019	0.94367	2.24726
Yeo-Johnson Regression (Test)	5.57276	0.93986	2.36067
Yeo-Johnson Regression (5X Validation)	0.00249	0.95631	0.04991

Seoul Bike

Technique	**MSE**	**R2**	**RMSE**
Logarithmic Regression (In-Sample)	156577.35063	0.55581	395.69856
Logarithmic Regression (Train)	159313.84921	0.55193	399.14139
Logarithmic Regression (Test)	142534.87846	0.58047	377.53792
Logarithmic Regression (5X Validation)	0.53065	0.78915	0.72818
Square Root Regression (In-Sample)	133632.33667	0.62091	365.55757
Square Root Regression (Train)	137268.54594	0.61394	370.49770
Square Root Regression (Test)	118028.86160	0.65260	343.55329
Square Root Regression (5X Validation)	49.20838	0.65891	7.01307
Box-Cox Regression (In-Sample)	135724.73632	0.61497	368.40838
Box-Cox Regression (Train)	139318.74642	0.60817	373.25427

Box-Cox Regression (Test)	119042.98436	0.64962	345.02606
Box-Cox Regression (5X Validation)	16.82297	0.71567	4.10046
Yeo-Johnson Regression (In-Sample)	134985.86145	0.61707	367.40422
Yeo-Johnson Regression (Train)	138554.46263	0.61032	372.22905
Yeo-Johnson Regression (Test)	118390.37537	0.65154	344.07902
Yeo-Johnson Regression (5X Validation)	24.81002	0.68661	4.97964

Boston Housing

Technique	**MSE**	**R2**	**RMSE**
Logarithmic Regression (In-Sample)	18.59032	0.77979	4.31165
Logarithmic Regression (Train)	18.32633	0.78905	4.28093
Logarithmic Regression (Test)	17.30027	0.76409	4.15936
Logarithmic Regression (5X Validation)	0.03520	0.75766	0.18642
Square Root Regression (In-Sample)	19.34757	0.77082	4.39859
Square Root Regression (Train)	19.13472	0.77974	4.37433
Square Root Regression (Test)	18.82078	0.74335	4.33829
Square Root Regression (5X Validation)	0.21799	0.75040	0.46187
Box-Cox Regression (In-Sample)	18.77342	0.77762	4.33283
Box-Cox Regression (Train)	18.53569	0.78664	4.30531
Box-Cox Regression (Test)	17.70265	0.75860	4.20745
Box-Cox Regression (5X Validation)	0.14579	0.75761	0.37876
Yeo-Johnson Regression (In-Sample)	18.77313	0.77762	4.33280
Yeo-Johnson Regression (Train)	18.53051	0.78670	4.30471
Yeo-Johnson Regression (Test)	17.69277	0.75874	4.20628
Yeo-Johnson Regression (5X Validation)	0.10066	0.75838	0.31473

6. Symbolic Regression

Auto MPG

Technique	**MSE**	**R^2**	**RMSE**
Symbolic Regression (In-Sample)	21.15553	0.65183	4.59951
Symbolic Regression (Train)	13.93365	0.77918	3.73278
Symbolic Regression (Test)	14.83214	0.70941	3.85125
Symbolic Regression (5X Validation)	19.19235	0.68241	4.34933

Energy Efficiency – Heating Load

Technique	**MSE**	**R^2**	**RMSE**
Symbolic Regression (In-Sample)	4.68666	0.95391	2.16487
Symbolic Regression (Train)	4.56375	0.95478	2.13629
Symbolic Regression (Test)	5.17672	0.95033	2.27524
Symbolic Regression (5X Validation)	4.69599	0.95365	2.16484

Energy Efficiency – Cooling Load

Technique	**MSE**	**R^2**	**RMSE**
Symbolic Regression (In-Sample)	4.68666	0.94815	2.16487
Symbolic Regression (Train)	4.56375	0.94910	2.13629
Symbolic Regression (Test)	5.17672	0.94413	2.27524
Symbolic Regression (5X Validation)	4.58848	0.94912	2.13941

Seoul Bike

Technique	MSE	R^2	RMSE
Symbolic Regression (In-Sample)	139412.16052	0.60451	373.37938
Symbolic Regression (Train)	159667.19851	0.55094	399.58378
Symbolic Regression (Test)	128269.69507	0.62246	358.14759
Symbolic Regression (5X Validation)	161524.92371	0.54081	401.76557

Boston Housing

Technique	MSE	R^2	RMSE
Symbolic Regression (In-Sample)	25.30109	0.70029	5.03002
Symbolic Regression (Train)	32.48102	0.62611	5.69921
Symbolic Regression (Test)	22.89217	0.68784	4.78458
Symbolic Regression (5X Validation)	28.71602	0.65961	5.30243

7. Linear Regression (SCALATION)

Auto MPG

	MSE	**R2**	**RMSE**
Train	10.58177	0.82866	3.25296
Test	12.35949	0.78237	3.51561
5X Cross-Validation	11.12565	0.81653	3.31992

Seoul Bikes

	MSE	R2	RMSE
Train	182212.36226	0.63898	426.8634
Test	180604.26059	0.62452	424.9756
5X Cross Validation	182983.23743	0.63660	427.7654

Boston Housing

	MSE	**R2**	**RMSE**
Train	21.66111	0.74101	4.65415
Test	23.42061	0.73222	4.83948
5X Cross Validation	24.35068	0.72170	4.80633

Energy Heating

	MSE	**R2**	**RMSE**
Train	3.05731	0.97044	1.74852
Test	3.54334	0.96259	1.88238
5X Cross Validation	3.14440	0.96885	1.76928

22.31 heat, 24.59 cool

Energy Cooling

	MSE	R2	RMSE
Train	3.64424	0.95984	1.90899
Test	4.16069	0.95320	2.03978
5X Cross Validation	3.75542	0.95774	1.93305

8. **Ridge Regression (SCALATION)**

Auto MPG

	MSE	R2	RMSE
Train	10.61785	0.82808	3.25850
Test	12.02821	0.78820	3.46817
5X Cross-Validation	11.09213	0.81734	3.31432

Seoul Bikes

	MSE	R2	RMSE
Train	143631.60095	0.63895	378.98760
Test	125407.13616	0.62480	354.12872
5X Cross Validation	135838.9798	0.63674	368.56344

Boston Housing

	MSE	R2	RMSE
Train	21.66556	0.74096	4.65463
Test	23.48550	0.73148	4.84618
5X Cross Validation	23.91439	0.72612	4.76804

Energy Heating

	MSE	R2	RMSE
Train	3.06363	0.97038	1.75032
Test	3.51792	0.96286	1.87561
5X Cross Validation	3.11648	0.96917	1.76033

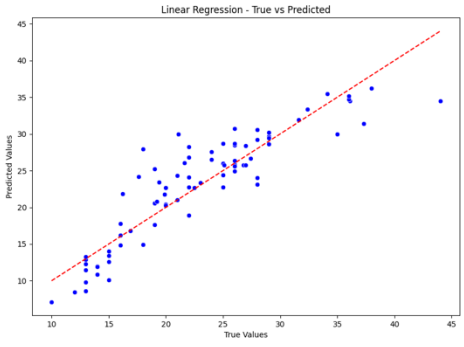
Energy Cooling

	MSE	R2	RMSE
Train	3.65027	0.95978	1.91057
Test	4.17470	0.95304	2.04321
5X Cross Validation	3.74785	0.95783	1.93098

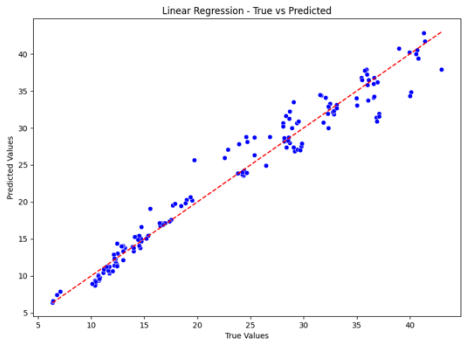
Statistical Summaries and Plots

Linear Regression

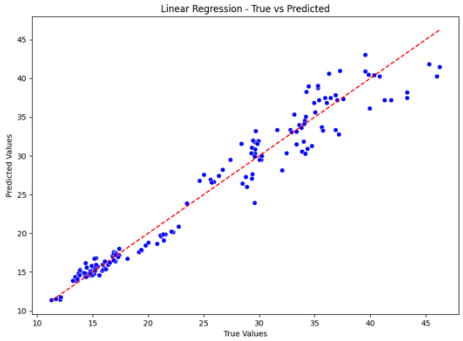
1. Auto MPG



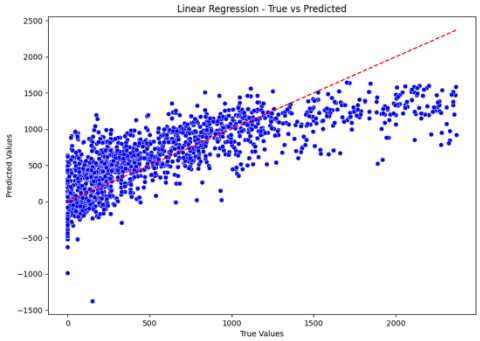
2. Energy Efficiency - Heating



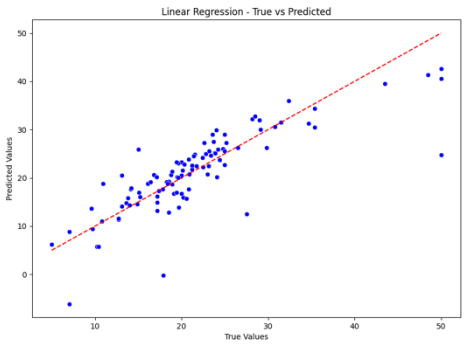
3. Energy Efficiency - Cooling



4. Seoul Bike Sharing Demand

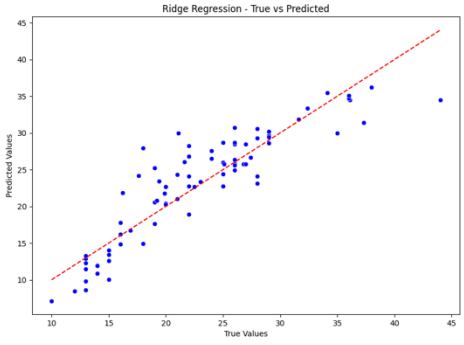


5. Boston Housing

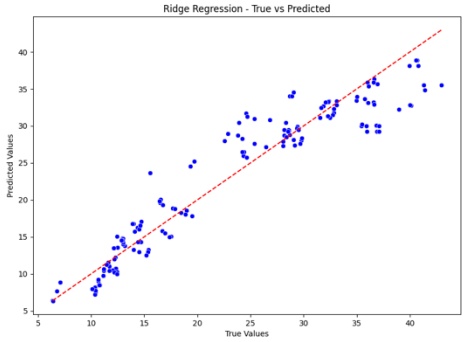


Ridge Regression

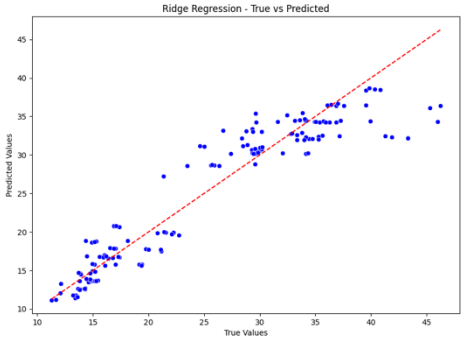
1. Auto MPG



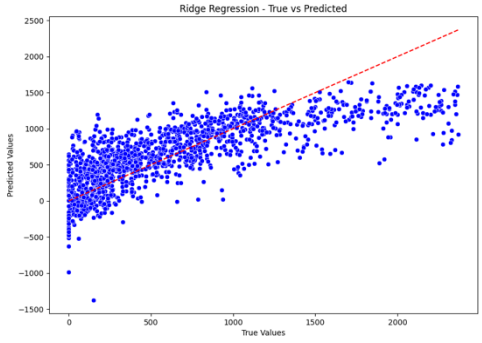
2. Energy Efficiency - Heating



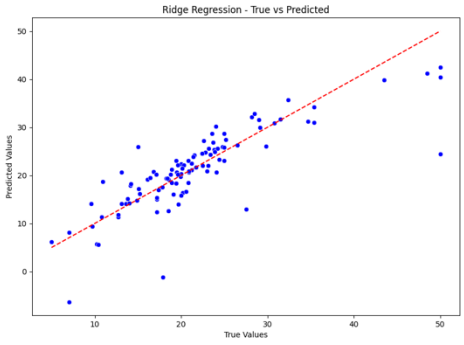
3. Energy Efficiency - Cooling



4. Seoul Bike Sharing Demand

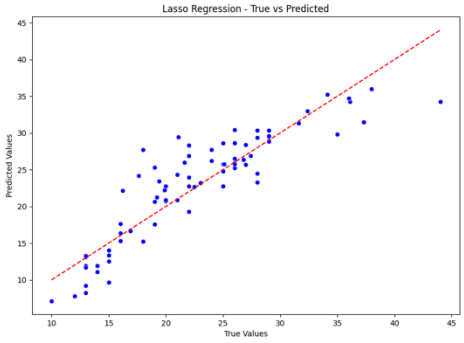


5. Boston Housing

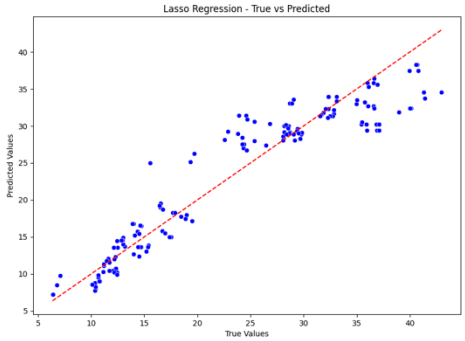


Lasso Regression

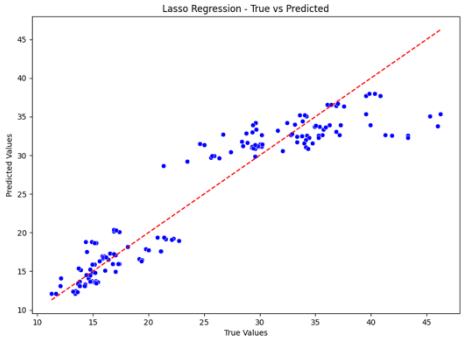
1. Auto MPG



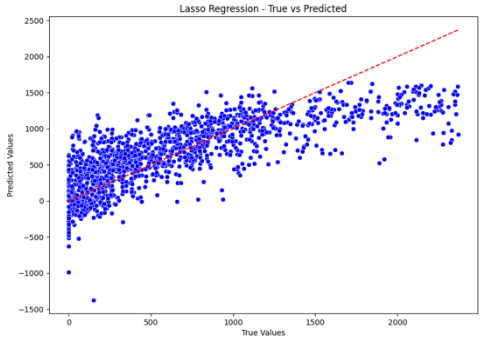
2. Energy Efficiency - Heating



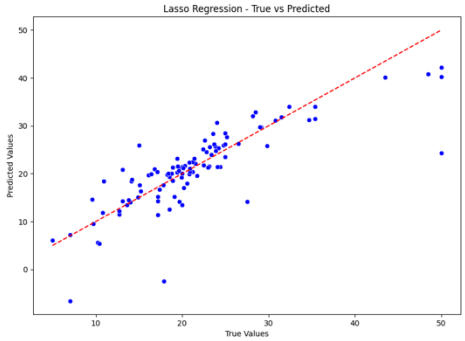
3. Energy Efficiency - Cooling



4. Seoul Bike Sharing Demand

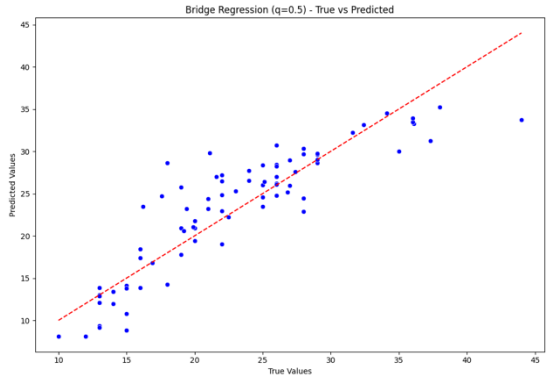


5. Boston Housing

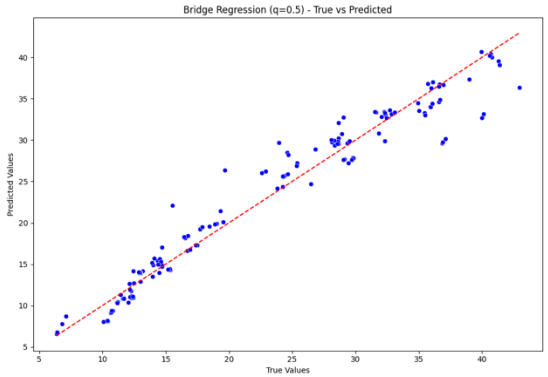


Bridge Regression

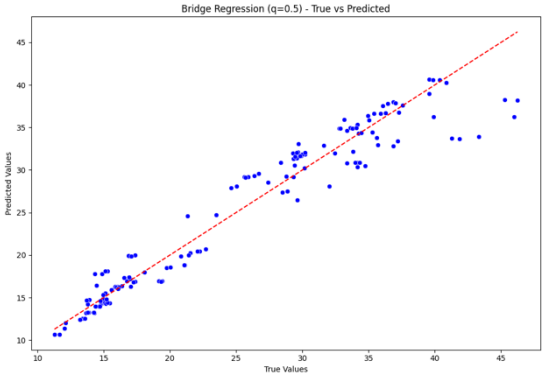
1. Auto MPG



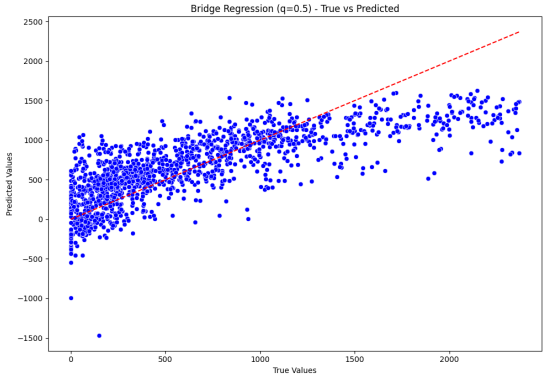
2. Energy Efficiency - Heating



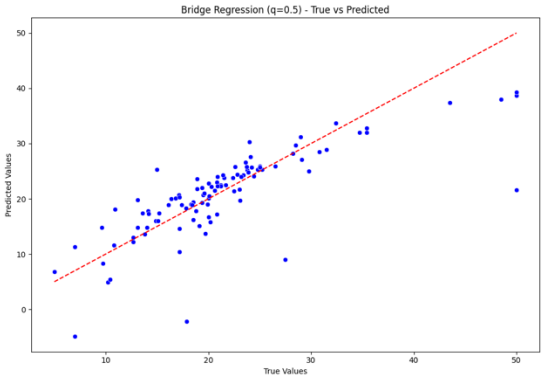
3. Energy Efficiency - Cooling



4. Seoul Bike Sharing Demand

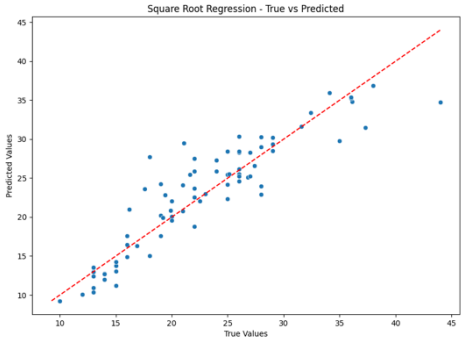


5. Boston Housing

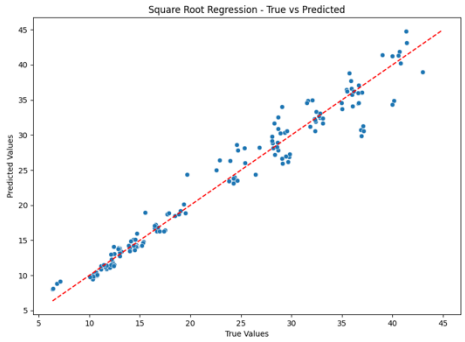


Transformed Regression (Sqrt)

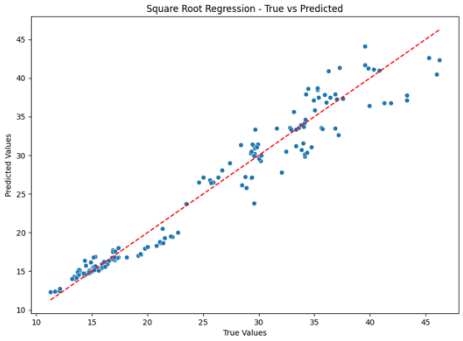
1. Auto MPG



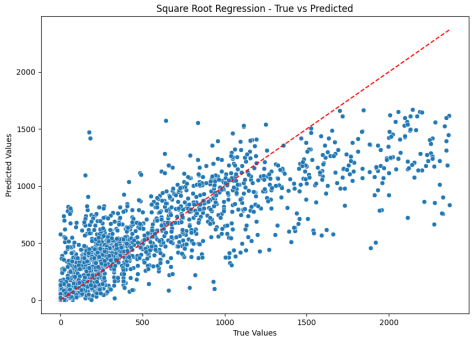
2. Energy Efficiency - Heating



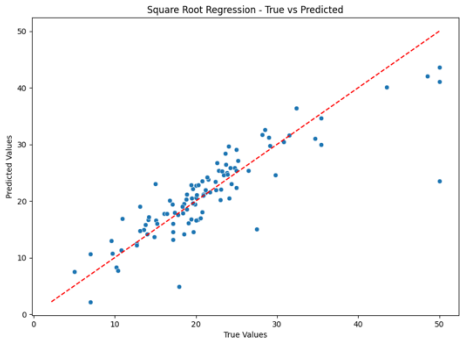
3. Energy Efficiency - Cooling



4. Seoul Bike Sharing Demand

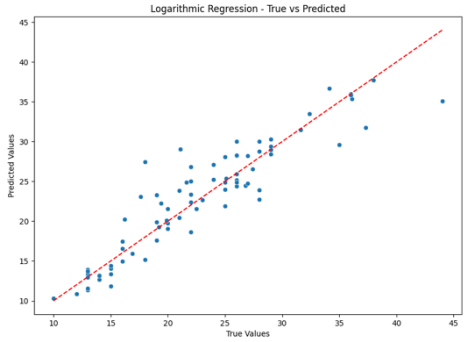


5. Boston Housing

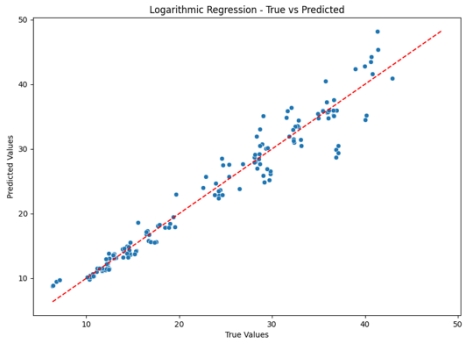


Transformed Regression (Log1p)

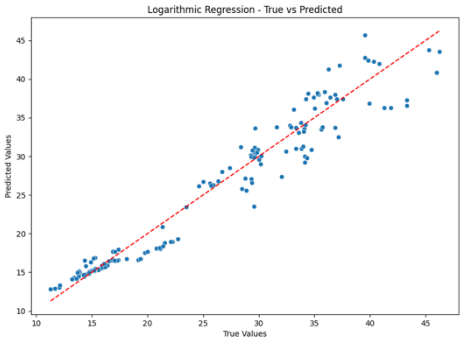
1. Auto MPG



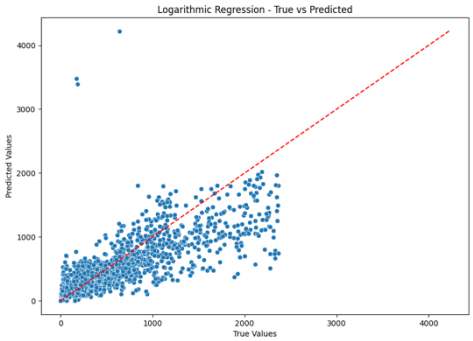
2. Energy Efficiency - Heating



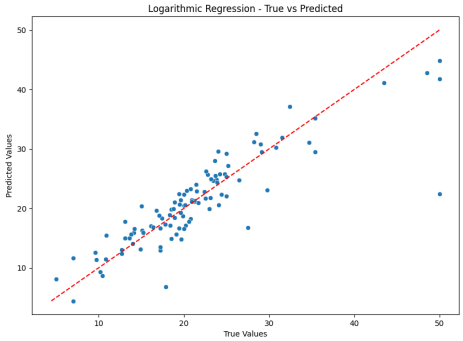
3. Energy Efficiency - Cooling



4. Seoul Bike Sharing Demand

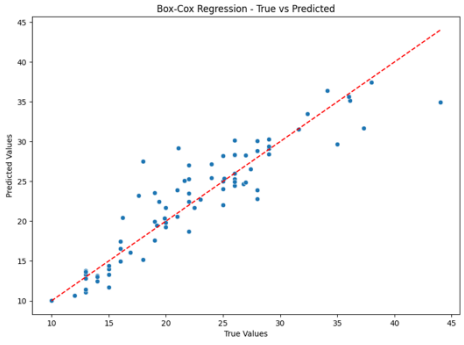


5. Boston Housing

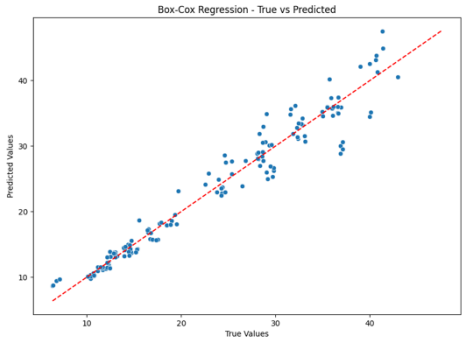


Transformed Regression (Box-Cox,)

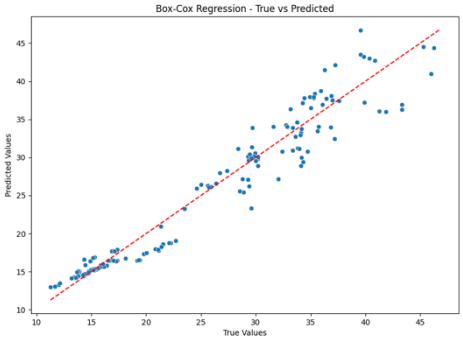
1. Auto MPG



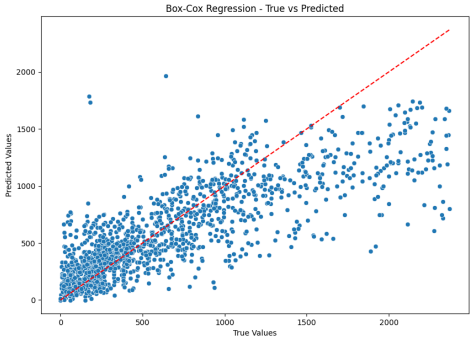
2. Energy Efficiency - Heating



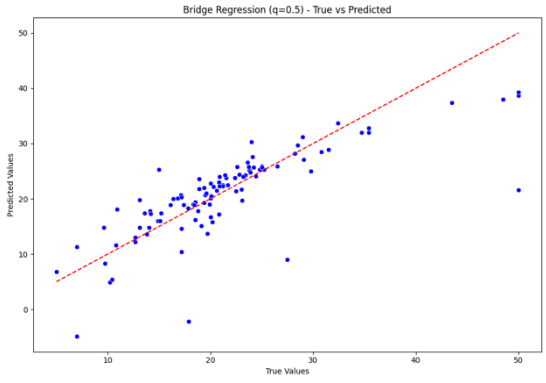
3. Energy Efficiency - Cooling



4. Seoul Bike Sharing Demand

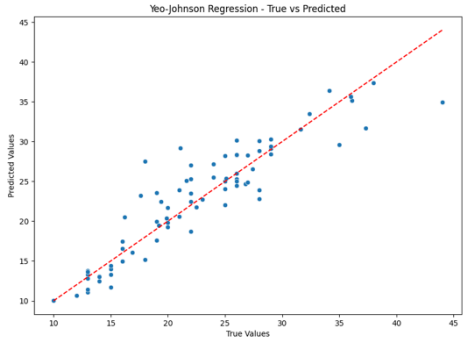


5. Boston Housing

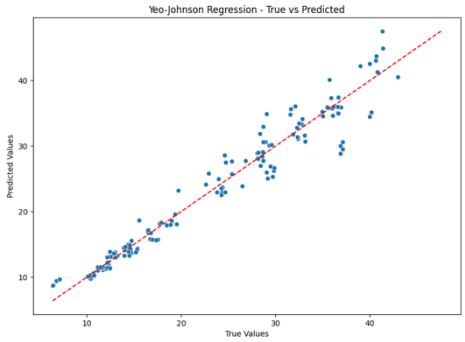


Transformed Regression (Yeo-Johnson)

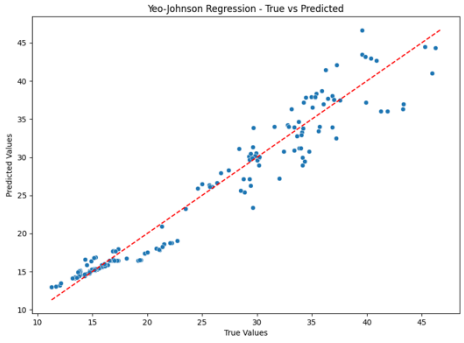
1. Auto MPG



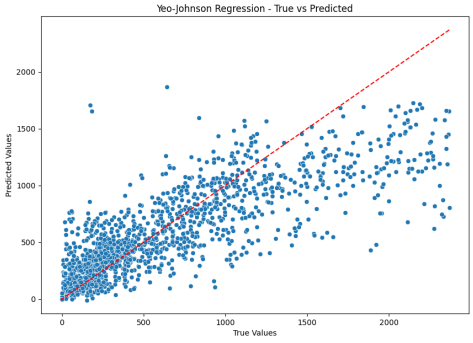
2. Energy Efficiency - Heating



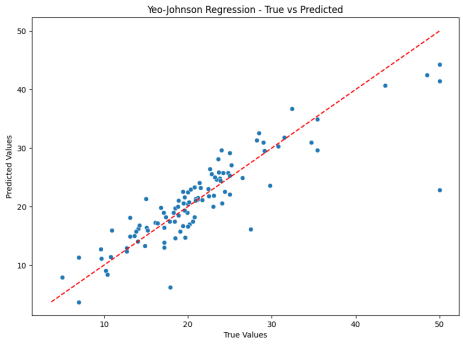
3. Energy Efficiency - Cooling



4. Seoul Bike Sharing Demand

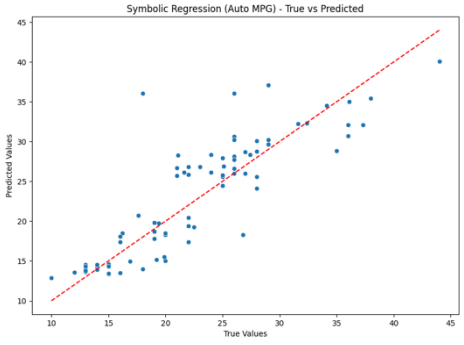


5. Boston Housing

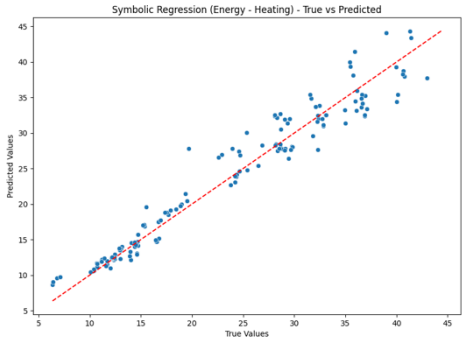


Symbolic Regression

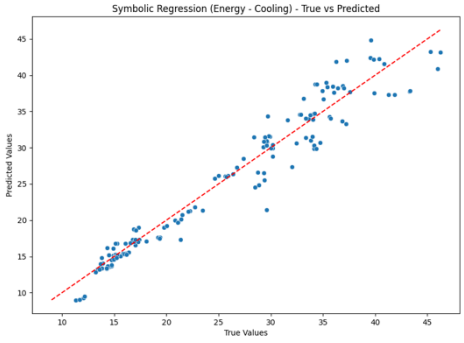
1. Auto MPG



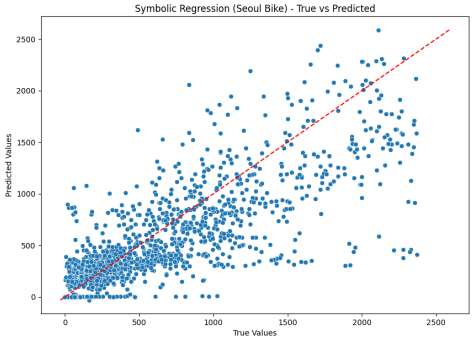
2. Energy Efficiency - Heating



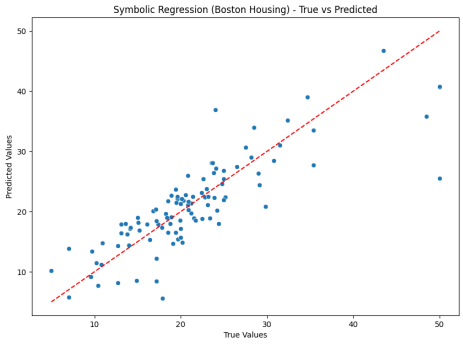
3. Energy Efficiency - Cooling



4. Seoul Bike Sharing Demand



5. Boston Housing



Feature Selection

Auto MPG

Method	R^2	Adj R^2	# Feats	Selected (order)
Forward	0.82148	0.81822	7	weight, model_year, origin, displacement, horsepower, cylinders, acceleration
Backward	0.82148	0.81822	7	cylinders, displacement, horsepower, weight, acceleration, model_year, origin
Stepwise	0.82148	0.81822	7	weight, displacement, horsepower, cylinders, model_year, origin, acceleration

Energy Efficiency — Heating

Method	R^2	Adj R^2	# Feats	Selected (order)
Forward	0.96909	0.96876	8	Cooling Load, Glazing Area, Overall Height, Wall Area, Glazing Area Distribution, Orientation, Surface Area, Relative Compactness
Backward	0.96893	0.96860	8	Relative Compactness, Surface Area, Wall Area, Roof Area, Overall Height, Glazing Area, Glazing Area Distribution, Cooling Load
Stepwise	0.96893	0.96860	8	Cooling Load, Overall Height, Roof Area, Surface Area, Relative Compactness, Wall Area, Glazing Area, Glazing Area Distribution

Energy Efficiency — Cooling

Method	R^2	Adj R^2	# Feats	Selected (order)
Forward	0.95861	0.95817	8	Heating Load, Overall Height, Glazing Area, Roof Area, Glazing Area Distribution, Orientation, Wall Area, Relative Compactness
Backward	0.95787	0.95748	7	Relative Compactness, Surface Area, Wall Area, Roof Area, Overall Height, Glazing Area, Heating Load
Stepwise	0.95787	0.95748	7	Heating Load, Overall Height, Roof Area, Surface Area, Relative Compactness, Wall Area, Glazing Area

Seoul Bike Sharing

Method	R^2	Adj R^2	# Feats	Selected (order)
Forward	0.57664	0.57558	19	Temperature(C), Hour, Functioning Day_Yes, Humidity(%), Seasons_Winter, Rainfall(mm), Seasons_Spring, Hour_sin, IsWeekend, Seasons_Summer, Temperature_sq, Holiday_No Holiday, Dew point temperature(C), Hour_cos, Solar Radiation (MJ/m2), Snowfall (cm), DayOfWeek, Visibility (10m), Wind speed (m/s)
Backward	0.57664	0.57558	19	Hour, Temperature(C), Humidity(%), Wind speed (m/s), Visibility (10m), Dew point temperature(C), Solar Radiation (MJ/m2), Rainfall(mm), Snowfall (cm), DayOfWeek, IsWeekend, Hour_sin, Hour_cos, Temperature_sq, Seasons_Spring, Seasons_Summer, Seasons_Winter, Holiday_No Holiday, Functioning Day_Yes
Stepwise	0.57664	0.57558	19	Temperature(C), Temperature_sq, Seasons_Winter, Hour, Hour_sin, Dew point temperature(C), Seasons_Summer, Solar Radiation (MJ/m2), Functioning Day_Yes, Visibility (10m), Humidity(%), Snowfall (cm), Rainfall(mm), Wind speed (m/s), Hour_cos, Holiday_No Holiday, IsWeekend, DayOfWeek, Seasons_Spring

Boston Housing

Method	R^2	Adj R^2	# Feats	Selected (order)
Forward	0.74064	0.73379	13	LSTAT, RM, PTRATIO, DIS, NOX, CHAS, B, ZN, CRIM, RAD, TAX, INDUS, AGE
Backward	0.74064	0.73379	13	CRIM, ZN, INDUS, CHAS, NOX, RM, AGE, DIS, RAD, TAX, PTRATIO, B, LSTAT
Stepwise	0.74064	0.73379	13	LSTAT, RM, PTRATIO, INDUS, TAX, NOX, CRIM, RAD, AGE, ZN, B, DIS, CHAS

Discussion Of Results

Auto MPG:

- Manufacturers such as Nissan (36 MPG) and Honda (33.76 MPG) consistently delivered vehicles with better mileage, highlighting design emphasis on efficiency.
- Cylinder count remains a critical factor: 4-cylinder cars averaged almost double the mileage of 8-cylinder models (29.28 vs. 14.96 MPG).
- Fuel economy shows a gradual upward trend across model years, aligning with regulatory pushes and technological advances.
- Weight and horsepower remain strong determinants of fuel consumption, and their skewed distributions suggest concentration in certain ranges.
- Model Comparisons:
    - *In-Sample:* Logarithmic Regression produced the highest explanatory power (R^2 = 0.85525).
    - *Train:* Logarithmic Regression again led (R^2 = 0.85685).
    - *Test:* R^2 of 0.84923 confirmed its consistency.
    - *Cross-Validation:* R^2 of 0.8736 reinforced robustness across folds.

Energy Efficiency – Heating Load:

- Building geometry played a dominant role: compact designs reduced heating demand, while taller structures required substantially more.
- Orientation had minimal effect compared to glazing and height.
- Heating demand exhibited clear nonlinearity, favoring transformation-based approaches.
- Model Comparisons:
    - *In-Sample:* Square Root Regression stood out ($R^2 = 0.96882$).
    - *Train:* Square Root Regression achieved R^2 of 0.97016.
    - *Test:* Maintained strong accuracy with $R^2 = 0.96303$.
    - *Cross-Validation:* Square Root Regression confirmed superiority ($R^2 = 0.97355$).

Energy Efficiency – Cooling Load:

- Cooling requirements increased for buildings with lower compactness and larger glazed surfaces, consistent with expected thermal behavior.
- Roof and wall dimensions influenced cooling more strongly than orientation.
- Model Comparisons:
    - *In-Sample:* Square Root Regression provided the best fit ($R^2 = 0.95514$).
    - *Train:* Slightly higher $R^2 = 0.95576$.
    - *Test:* R^2 of 0.95140 confirmed generalizability.
    - *Cross-Validation:* R^2 of 0.96186 demonstrated the model's resilience across splits.

Seoul Bike Sharing Demand:

- Seasonal fluctuations were pronounced: summer demand peaked near 950 rentals/day, while winter dropped below 250.
- Strong commuting patterns appeared at 8 AM and between 5–7 PM.
- Rentals declined during holidays, consistent with lower work-related travel.
- Model Comparisons:
    - *In-Sample:* Square Root Regression produced the best R^2 (0.62091).
    - *Train:* Similar trend with $R^2 = 0.61394$.

- *Test:* Stronger performance at $R^2 = 0.65260$.
- *Cross-Validation:* Interestingly, Logarithmic Regression performed best ($R^2 = 0.78915$), suggesting higher stability under repeated resampling.

Boston Housing:

- Properties near the Charles River (CHAS=1) commanded higher average prices, showing location premiums.
- Access to radial highways (e.g., RAD=8) correlated with higher values, though extreme accessibility (RAD=24) corresponded with lower prices, hinting at noise and congestion effects.
- Socioeconomic indicators, especially LSTAT, strongly predicted housing values, confirming its importance.
- Model Comparisons:
  - *In-Sample:* Logarithmic Regression gave the best R^2 (0.77979).
  - *Train:* Slightly higher R^2 of 0.78905.
  - *Test:* Best generalization at $R^2 = 0.76409$.
  - *Cross-Validation:* R^2 of 0.75766 reaffirmed the reliability of log-transformation.

Overall Observations:

- The **Energy Efficiency dataset** consistently produced the highest accuracies ($R^2 > 0.96$), reflecting its strong linear and non-linear relationships.
- **Auto MPG** and **Boston Housing** achieved moderate to high fits, showing well-captured but noisier patterns.
- **Seoul Bike** presented greater unpredictability (R^2 ~0.6), likely due to behavioral and external weather factors.
- Across datasets, **Square Root Regression** excelled in Energy datasets, while **Logarithmic Regression** dominated Auto MPG and Boston Housing.
- Cross-validation outcomes emphasized model stability, with transformation-based methods offering the best generalization.