SpectraNeura

INTRODUCE

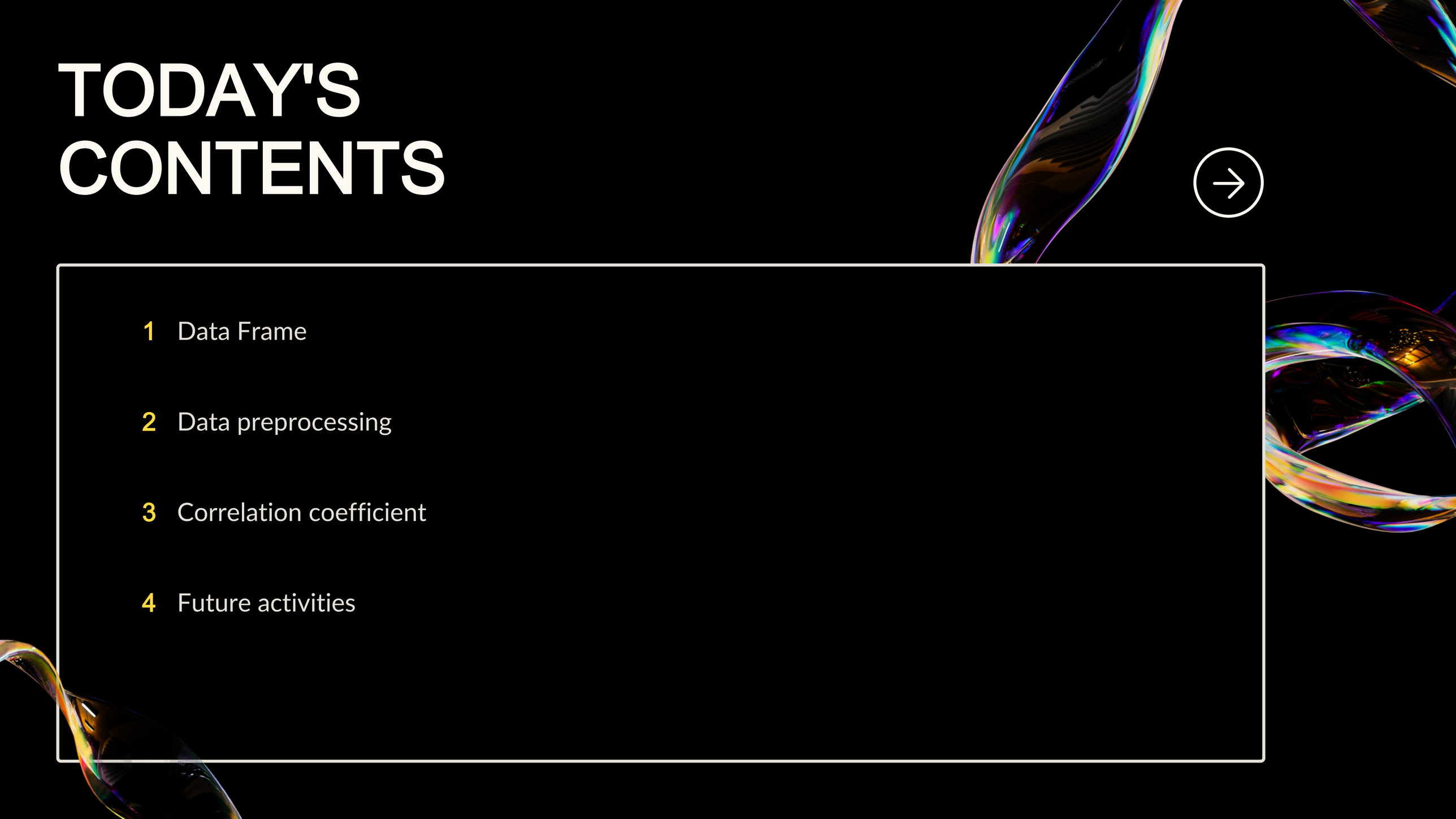
Present by Sunjun Hwang

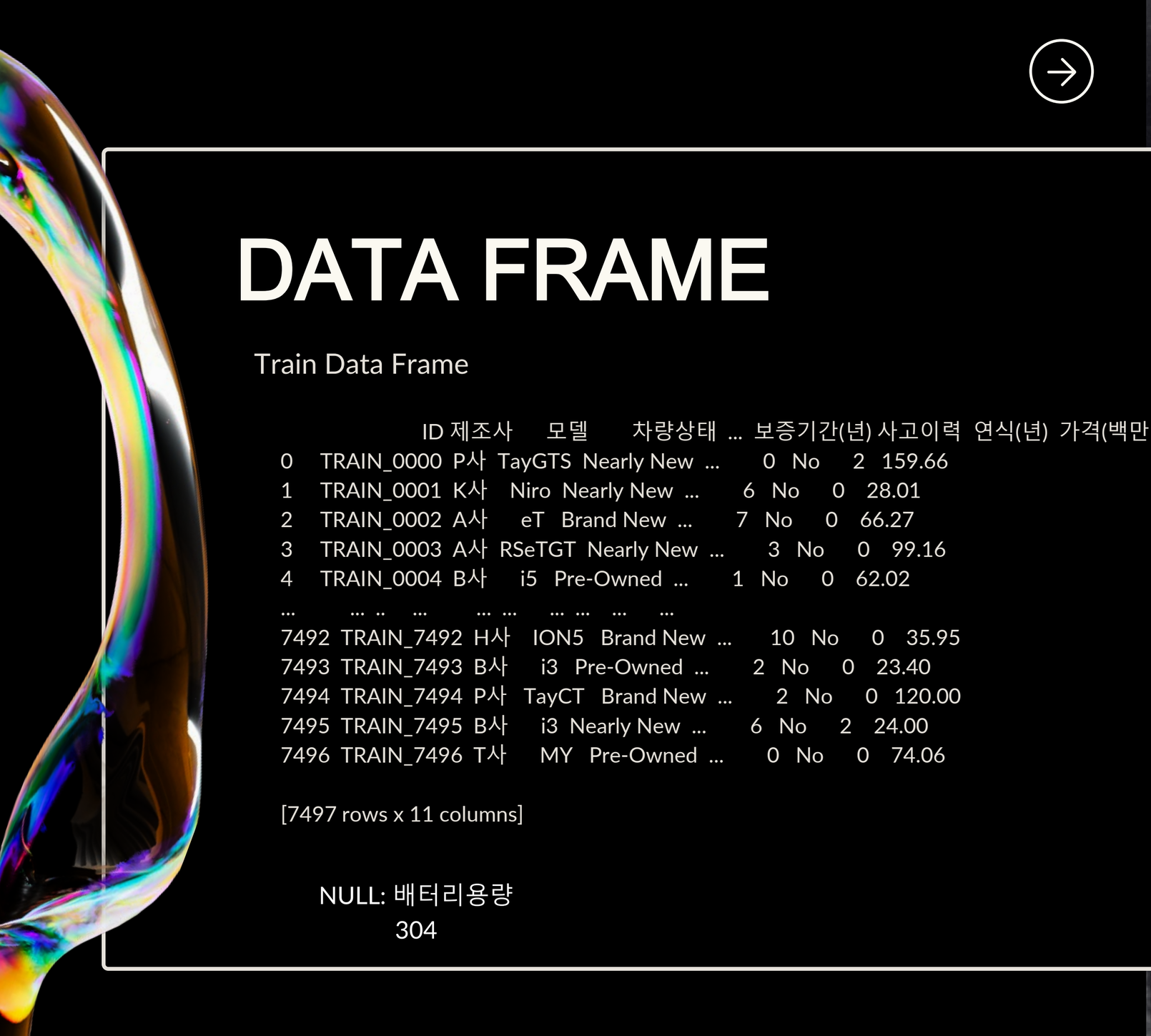
https://sites.google.com/view/seonjunhwang

sunjun7559@gmail.com

seoul, chungdam

TODAY'S CONTENTS





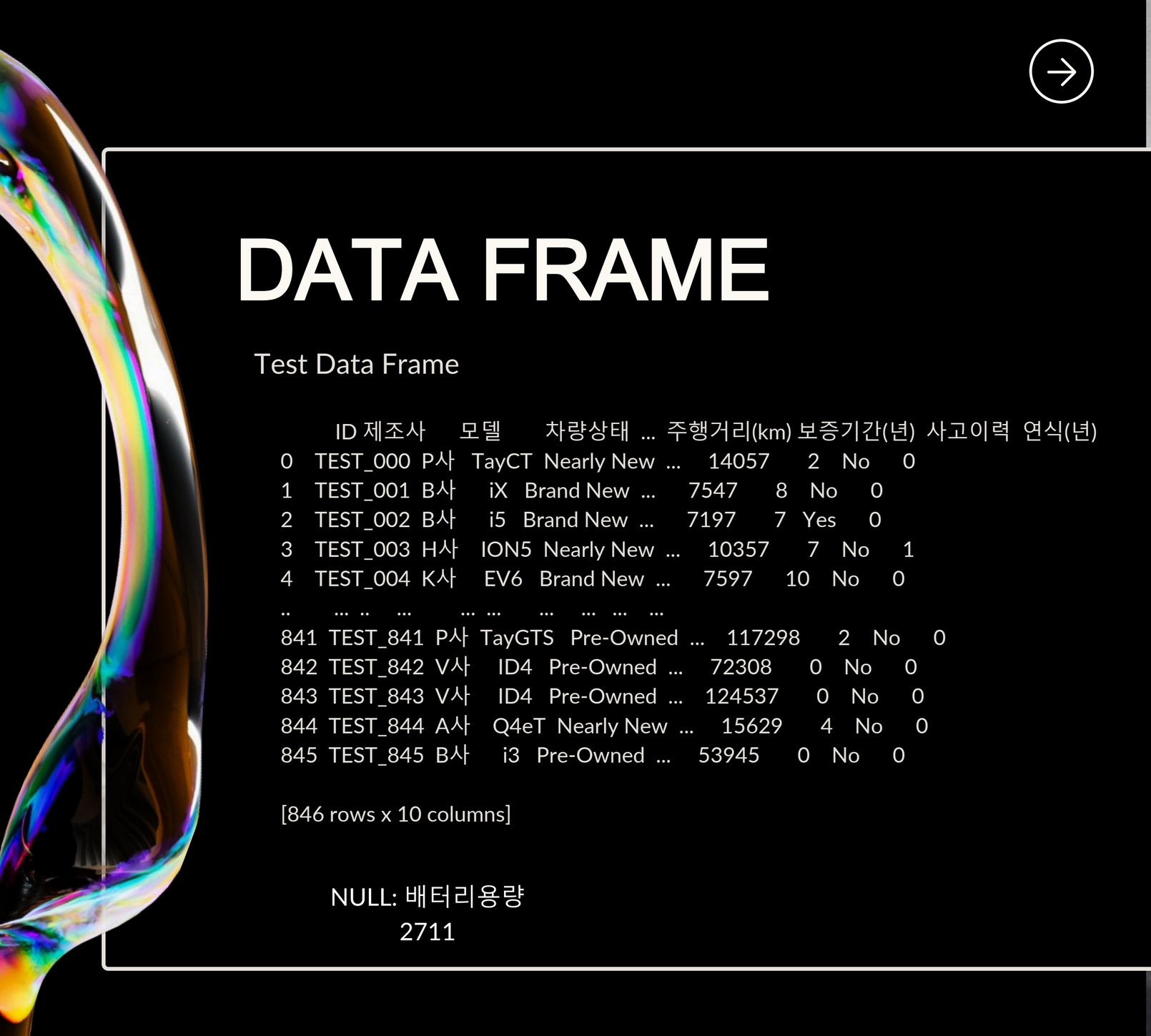
DATA FRAME

Train Data Frame

	ID	제조사	모델	차량상태	...	보증기간(년)	사고이력	연식(년)	가격(백만원)
0	TRAIN_0000	P사	TayGTS	Nearly New	...	0	No	2	159.66
1	TRAIN_0001	K사	Niro	Nearly New	...	6	No	0	28.01
2	TRAIN_0002	A사	eT	Brand New	...	7	No	0	66.27
3	TRAIN_0003	A사	RSeTGT	Nearly New	...	3	No	0	99.16
4	TRAIN_0004	B사	i5	Pre-Owned	...	1	No	0	62.02
...	...	...	...	...	...	...	...	...	...
7492	TRAIN_7492	H사	ION5	Brand New	...	10	No	0	35.95
7493	TRAIN_7493	B사	i3	Pre-Owned	...	2	No	0	23.40
7494	TRAIN_7494	P사	TayCT	Brand New	...	2	No	0	120.00
7495	TRAIN_7495	B사	i3	Nearly New	...	6	No	2	24.00
7496	TRAIN_7496	T사	MY	Pre-Owned	...	0	No	0	74.06

[7497 rows x 11 columns]

NULL: 배터리용량
304



DATA FRAME

Test Data Frame

	ID	제조사	모델	차량상태	...	주행거리(km)	보증기간(년)	사고이력	연식(년)
0	TEST_000	P사	TayCT	Nearly New	...	14057	2	No	0
1	TEST_001	B사	iX	Brand New	...	7547	8	No	0
2	TEST_002	B사	i5	Brand New	...	7197	7	Yes	0
3	TEST_003	H사	ION5	Nearly New	...	10357	7	No	1
4	TEST_004	K사	EV6	Brand New	...	7597	10	No	0
..	...	..	...	...	...	...	...	...	...
841	TEST_841	P사	TayGTS	Pre-Owned	...	117298	2	No	0
842	TEST_842	V사	ID4	Pre-Owned	...	72308	0	No	0
843	TEST_843	V사	ID4	Pre-Owned	...	124537	0	No	0
844	TEST_844	A사	Q4eT	Nearly New	...	15629	4	No	0
845	TEST_845	B사	i3	Pre-Owned	...	53945	0	No	0

[846 rows x 10 columns]

NULL: 배터리용량
2711

DATA PREPROCESSING

Model-based Imputation

결측치가 없는 데이터를 사용해, 회귀 모델을 학습한 후 주행거리, 보증기간, 연식 등의 피처를 활용해서 배터리 용량을 예측. 예측값으로 결측치를 채워 변수간 상관관계를 반영하는 방식이다.

Inconsistent Brand Messaging

전체 데이터의 배터리 용량 평균을 계산하고 결측치에 해당 평균값을 대입. 구현은 간단하지만 변수간 관계를 반영하지 못할 수 있음.

KNN Imputation

결측치가 있는 행에 대해 k개의 유사한 이웃을 찾는다. 이웃들의 배터리 용량 값을 평균하여 결측치를 채운다. 데이터 내 유사성을 반영하기에 현실적인 값을 제공

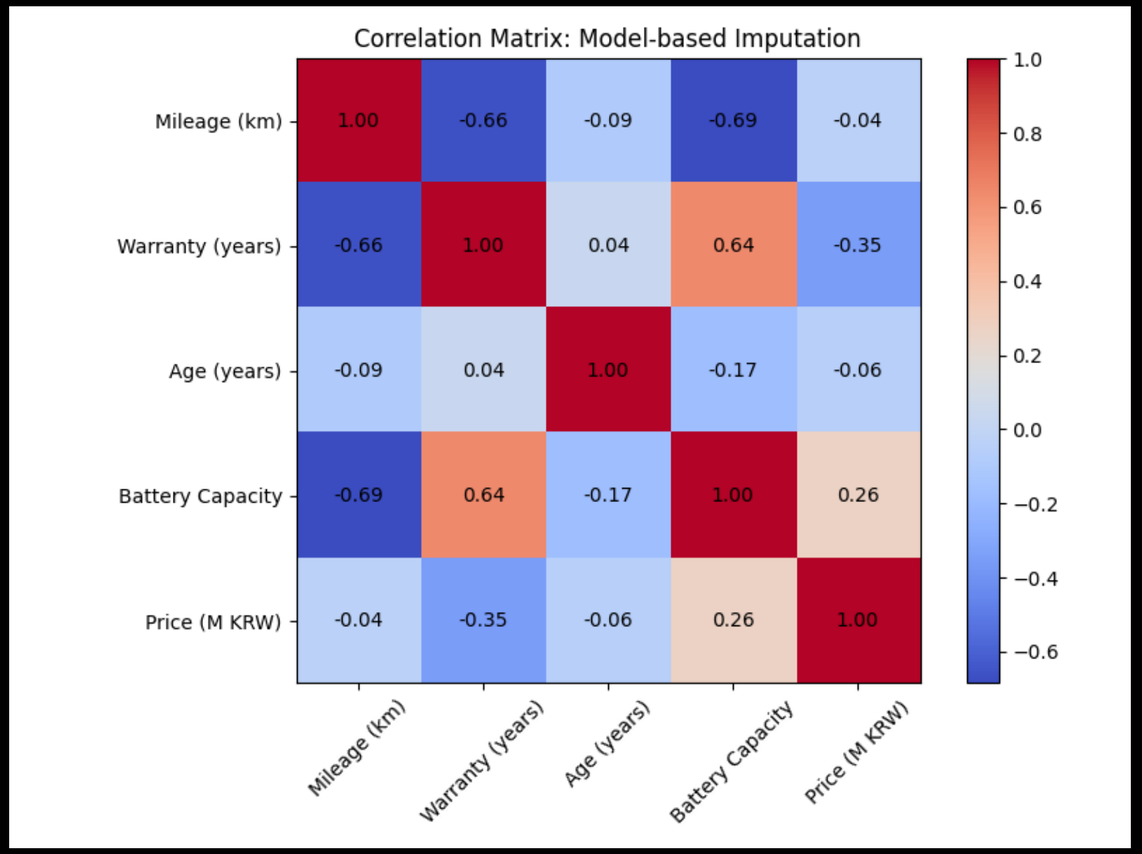
Linear Interpolation

데이터 순서에 따라 인접한 값들을 직선으로 연결하고 선형함수를 활용해 결측치를 보간한다. 순서가 의미 있는 데이터에서 자연스러운 변화로 간주한다.

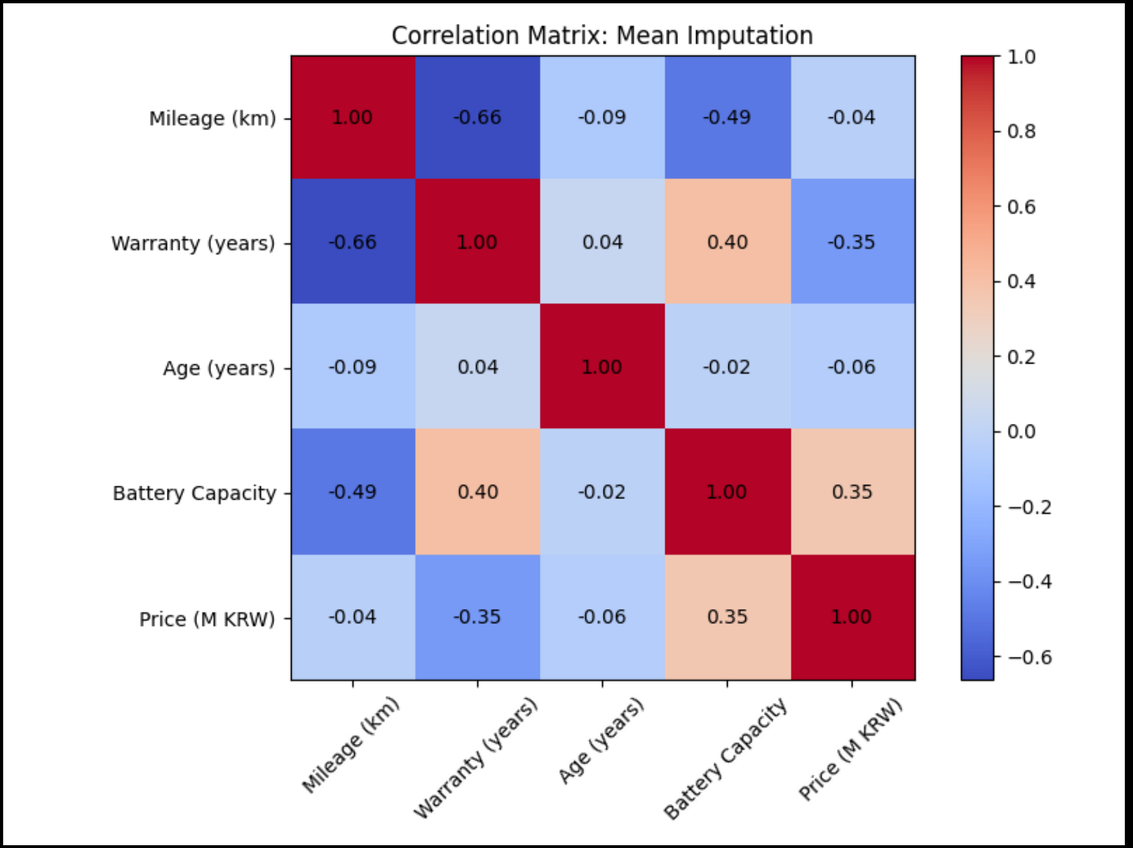
Polynomial Interpolation

선형 보간 대신 다항 함수를 사용하여 결측치를 보간한다. 데이터의 곡선 형태의 변화를 모델링한다. 다항식의 차수에 따라 결과가 달라질 수 있기에 주의가 필요하다.

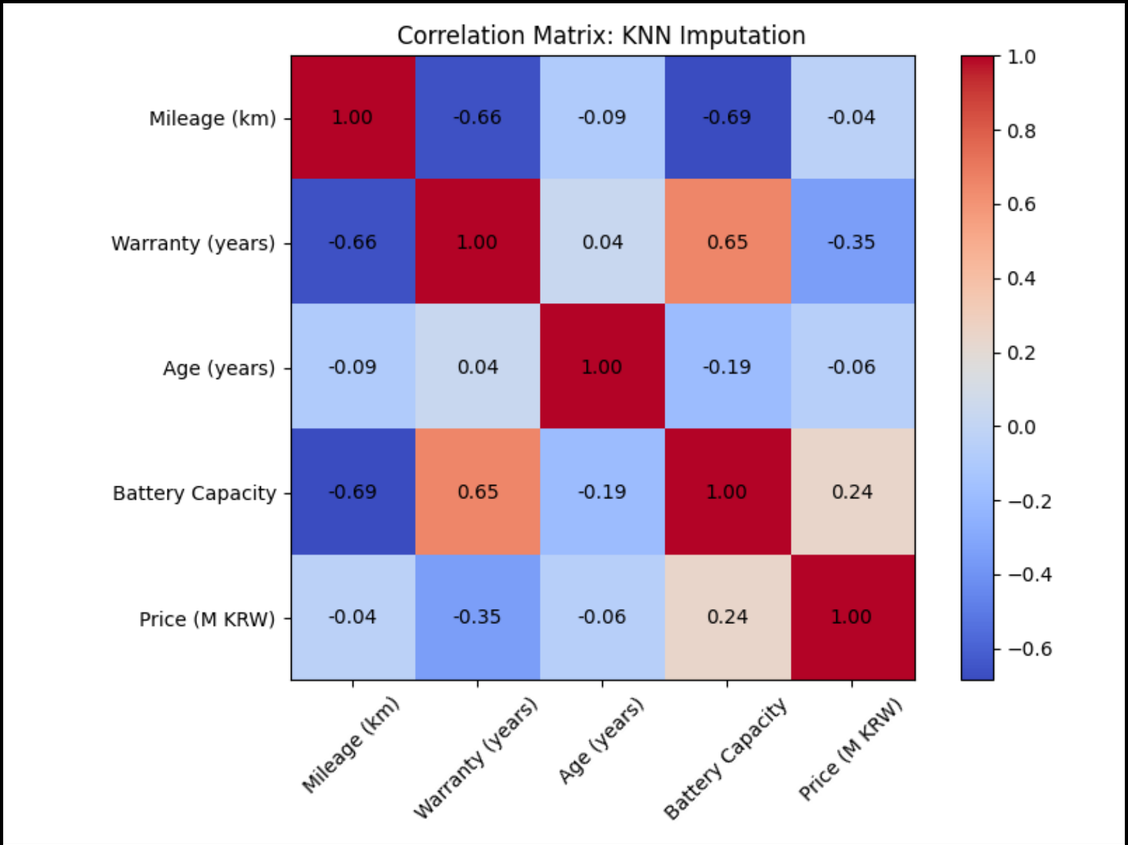
CORRELATION COEFFICIENT



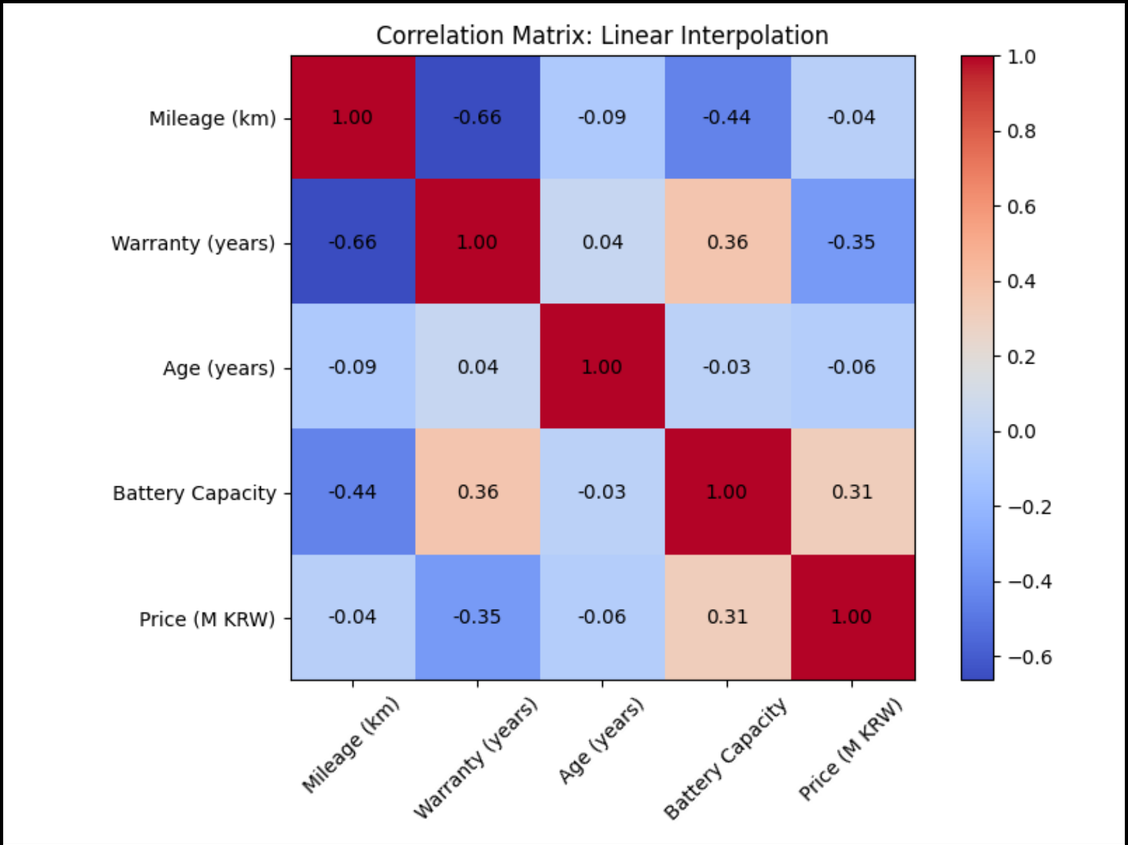
CORRELATION COEFFICIENT



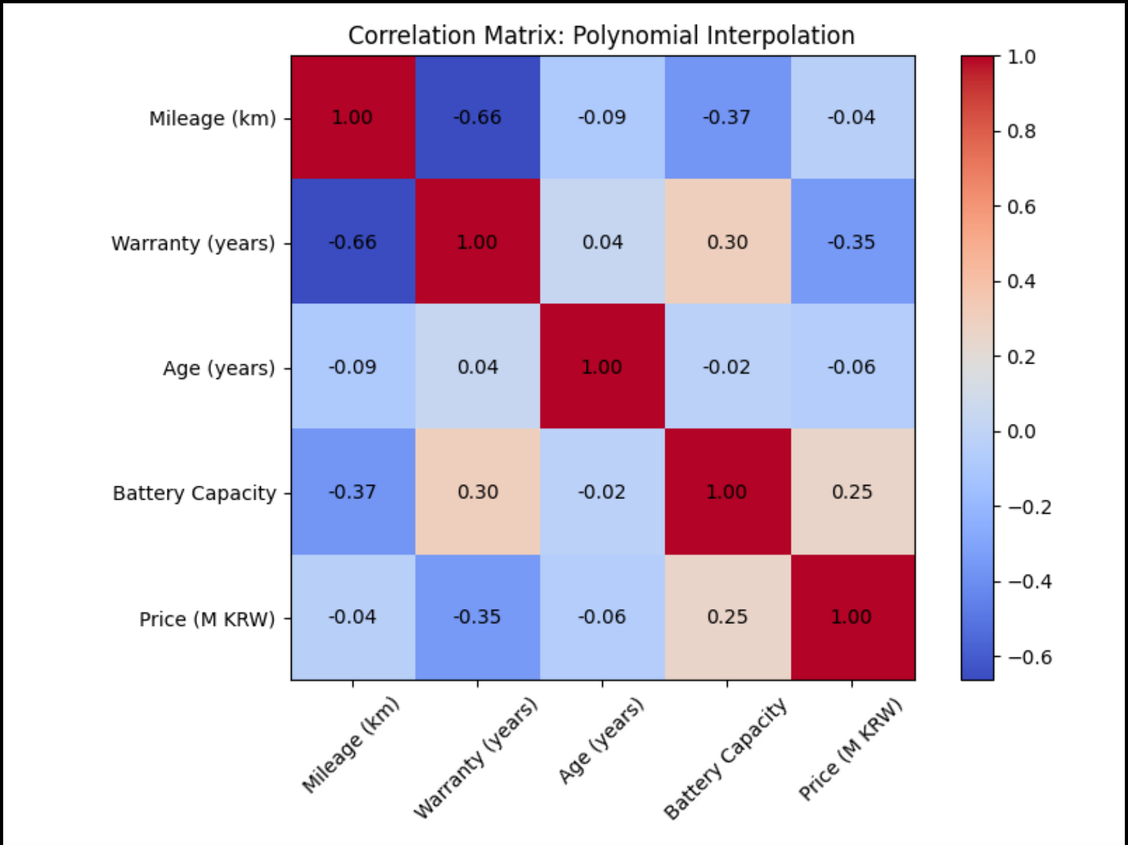
CORRELATION COEFFICIENT



CORRELATION COEFFICIENT



CORRELATION COEFFICIENT



FUTURE PLAN

시계열 데이터를 잘 처리할 수 있는 인공지능 모델을 활용해서 딥러닝을 진행할 예정.

각 모델은 논문을 참고하여 구현해볼 예정입니다.

THANK YOU

for your time and attention

Present by Sj Hwang