CLASSIFICATION – HOUSE GRADE PREDICTION

PUNEETHKRISHN B

DSFT8

INTRODUCTION:

➢ THE REAL ESTATE COMPANIES PREFER CATEGORIZING VARIOUS HOUSES INTO DIFFERENT GRADES BASED ON VARIOUS PARAMETERS.

➢ THIS WOULD HELP AGENTS IN IDENTIFYING THE TYPE OF HOUSE A CUSTOMER IS LOOKING FOR.

➢ THIS WAY, THE SEARCH FOR THE HOUSE CAN BE NARROWED DOWN BY FOCUSING ONLY ON THE 'CONDITION OF THE HOUSE'.

OBJECTIVE:

➢ TO HELP THE REAL ESTATE COMPANIES BY GROUPING THE HOUSES INTO DIFFERENT CATEGORIES BASED ON VARIOUS PARAMETERS

FEATURE	DATA TYPE	DESCRIPTION
ID	Numeric	Unique record identification number
Totalarea	Numeric	Total area of the house
Trooms	Numeric	Number of rooms in the house
Nbedrooms	Numeric	Number of bedrooms in the house
Nbwashrooms	Numeric	Number of attached washrooms with bedrooms
Twashrooms	Numeric	Total number of washrooms
Roof	Categorical	Does the house have roof?
Roofarea	Numeric	Area of the roof
Lawnarea	Numeric	Area of the lawn
Nflorrs	Numeric	Total floors
API	Numeric	Air pollution index level
ANB	Numeric	Average number belonging
Expected price	Numeri	Approximate cost price of the property
Grade	Categorical	The grade of the house/property

➢DATA PRE-PROCESSING

➢ THERE IS NO ANY MISSING VALUES OR ANY SPECIAL CHARACTERS PRESENT IN THE DATASET
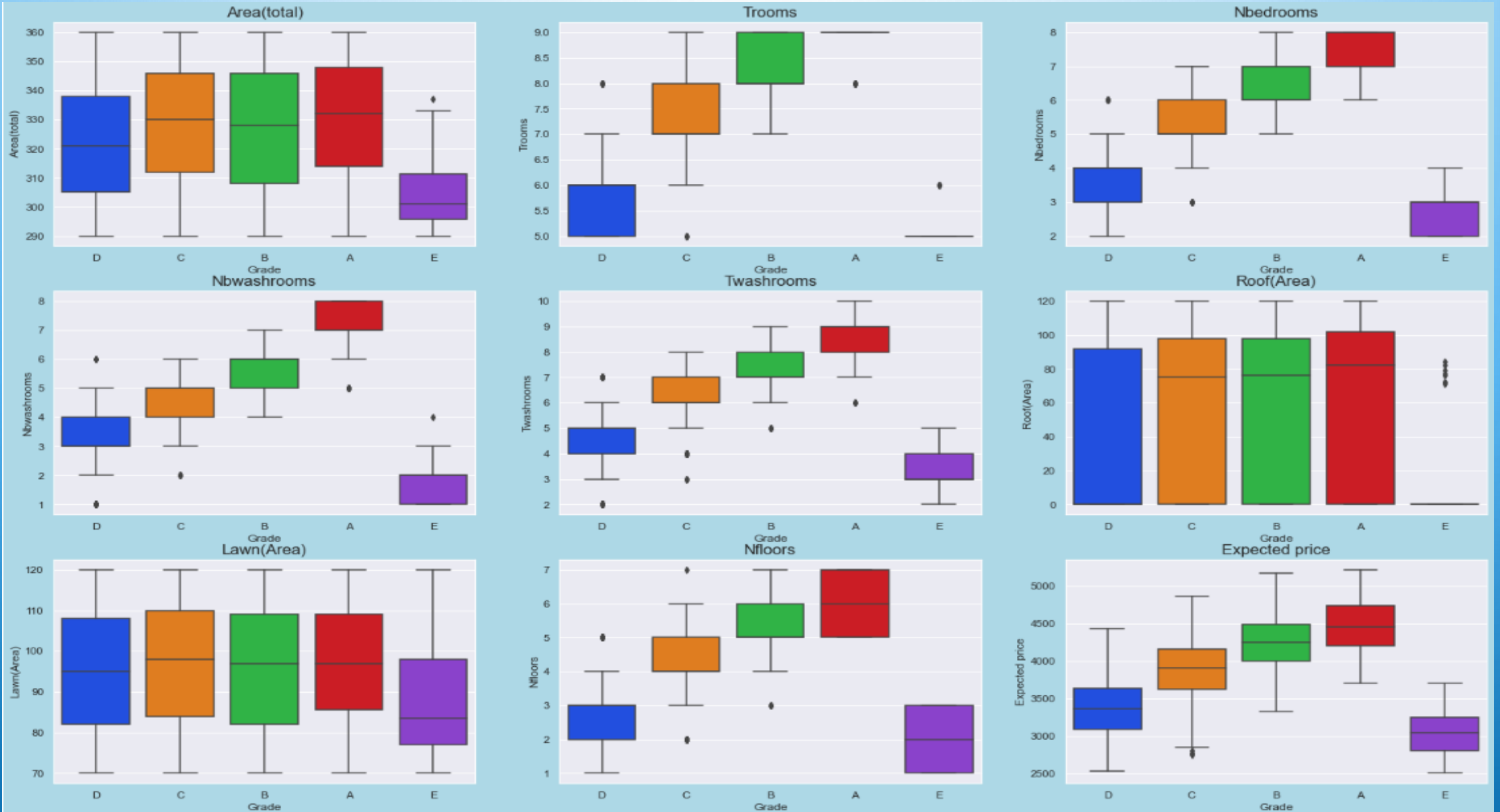
➢ **UNIVARIANT ANALYSIS**
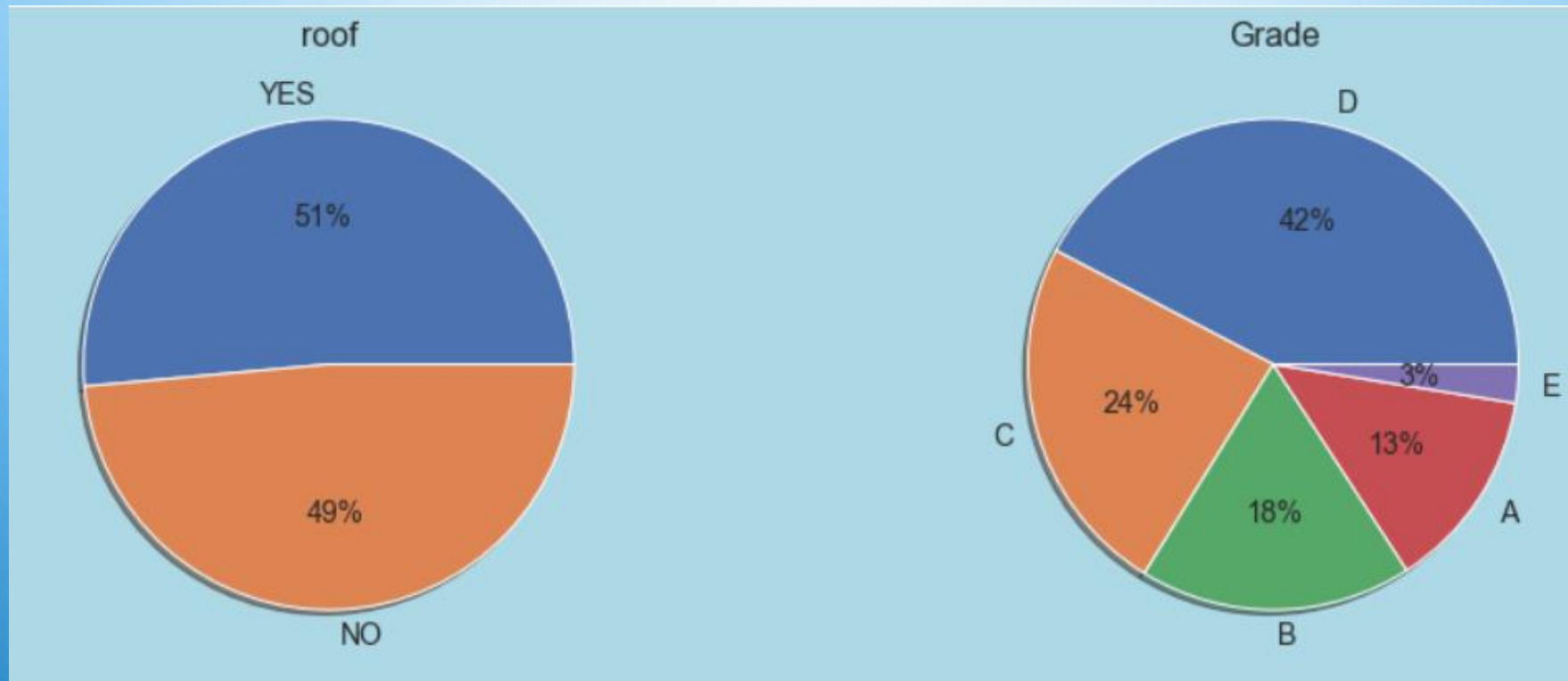
➢ BELOW PLOTS SHOWS THE DISTRIBUTION OF NUMERICAL FEATURES IN THE DATASET AND THERE IS NO OUTLIERS IN ANY FEATURES
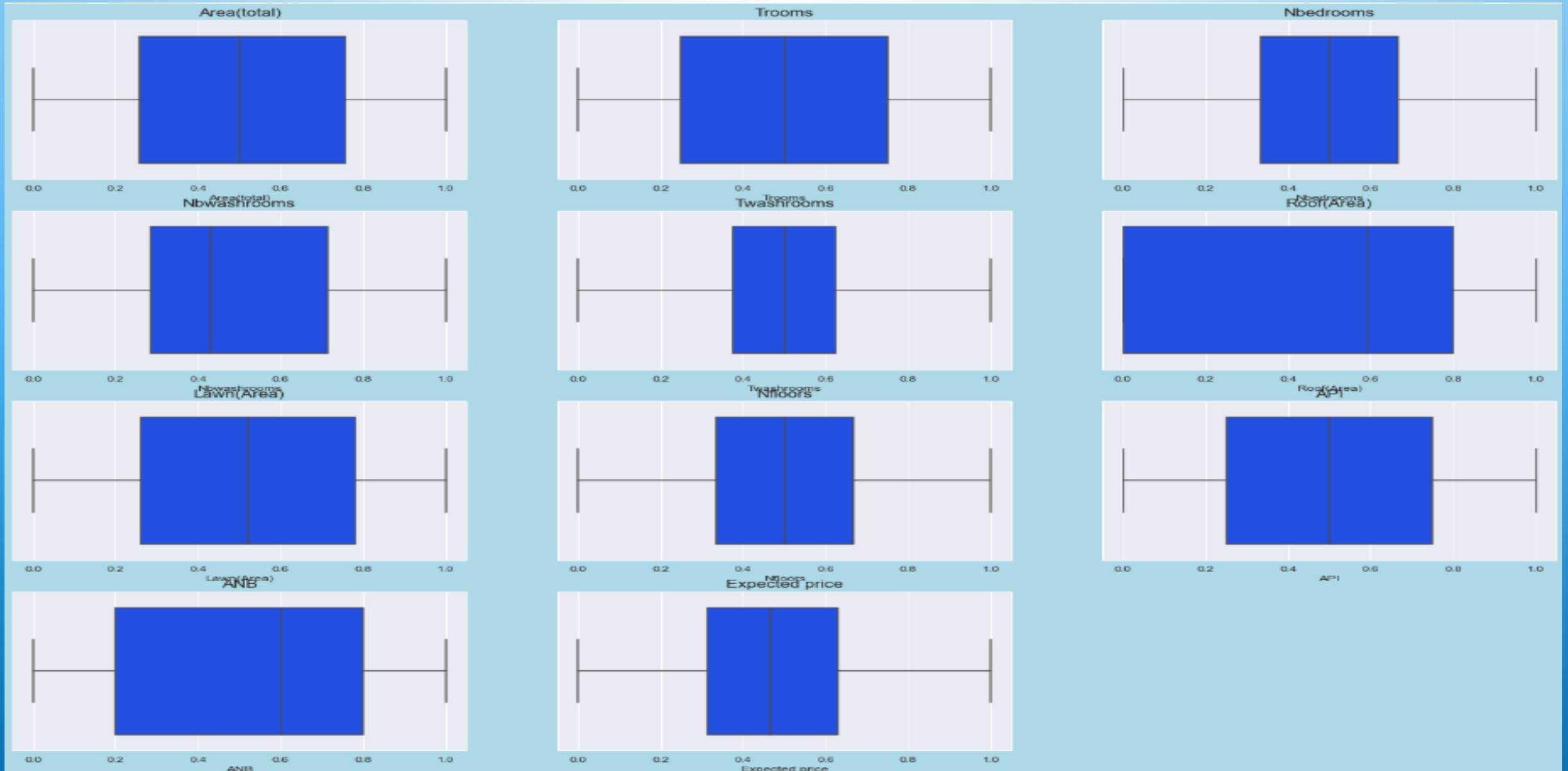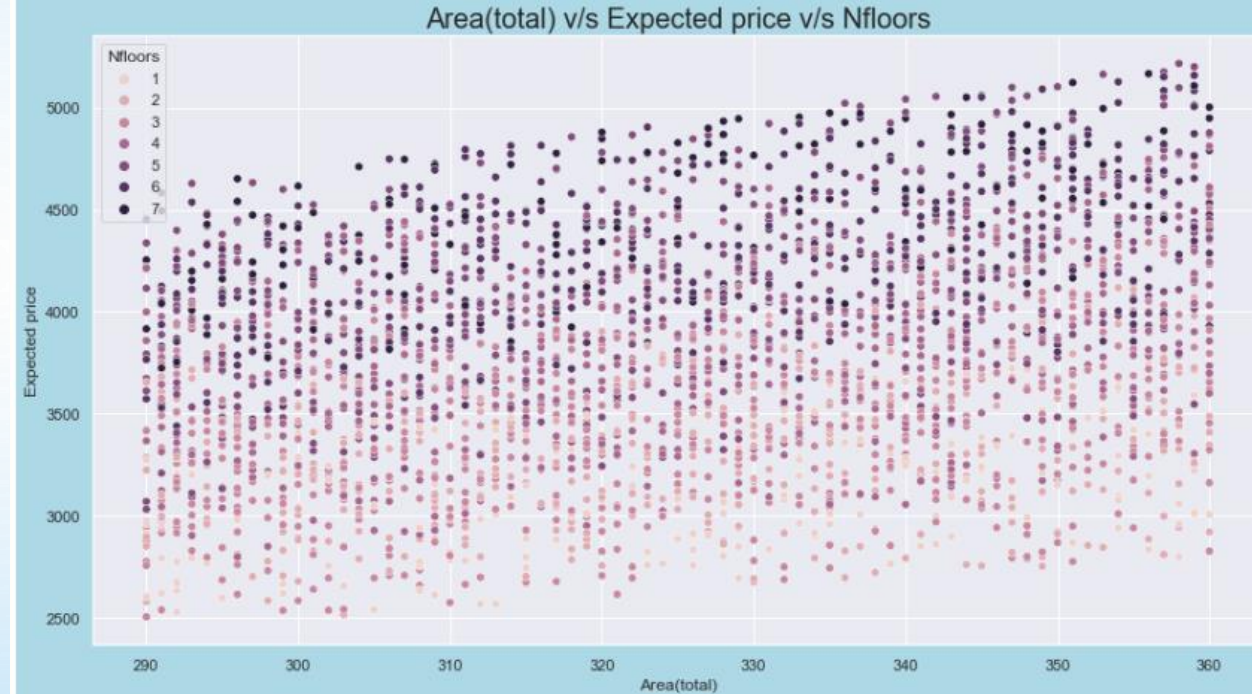
BIVARIANT ANALYSIS

FROM THE PLOT WE CAN SEE THAT HOW THE DISTRIBUTION OF ALL THE FEATURES WITH RESPECT TO HOUSE GRADES

➢ HERE WE CAN SEE HOW TEHE AREA AND EXPECTED PRICE RELATED WITH RESPECT TO NUMBER OF FLOORS


Area(total) v/s Expected price v/s Nfloors

Area(total) v/s Expected price v/s Grade

➢ HERE WE CAN SEE HOW TEHE AREA AND EXPECTED PRICE RELATED WITH RESPECT TO GRADE OF THE HOUSE

Area(total) v/s Trooms v/s Grade

> HERE WE CAN SEE THE DISTRIBUTION OF HOUSE GRADES WITH RESPECT TO TOTAL NUMBER OF ROOMS AND THEIR RESPECTIVE EXPECTED PRICE

Area(total) v/s Nbedrooms v/s Grade
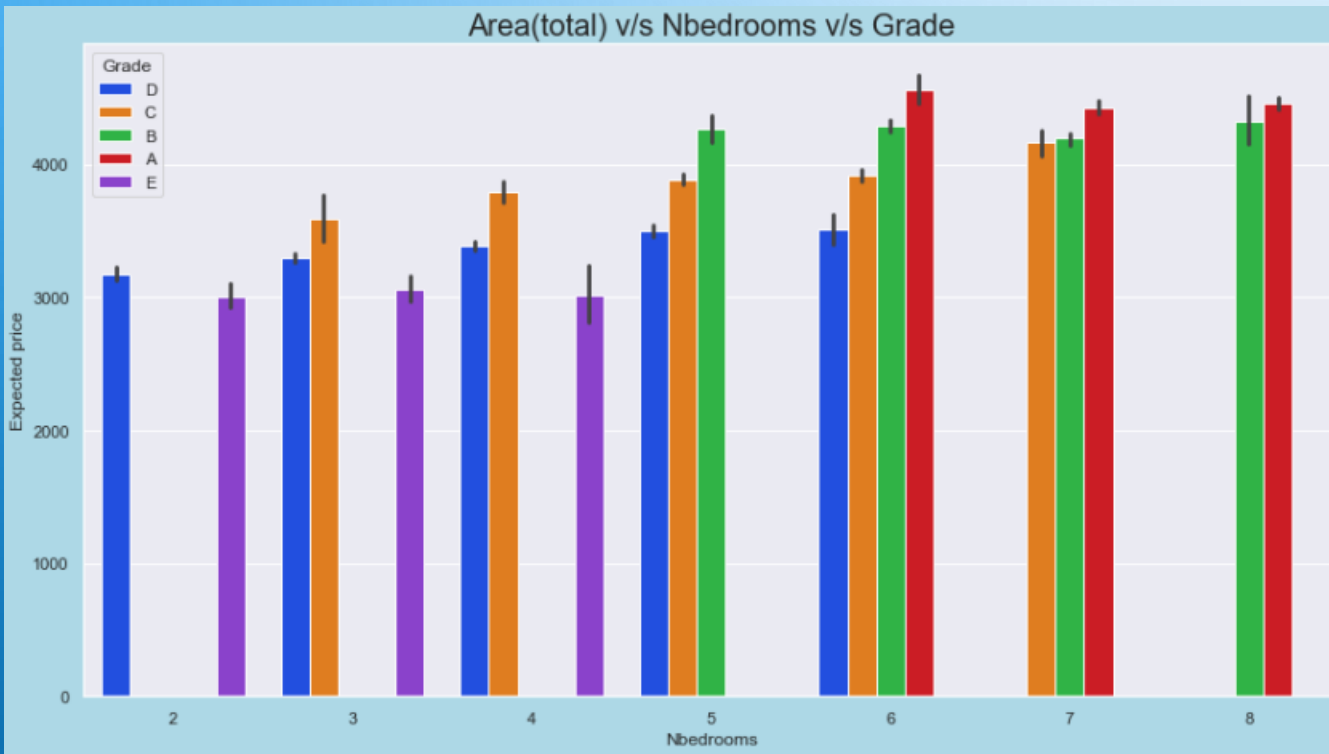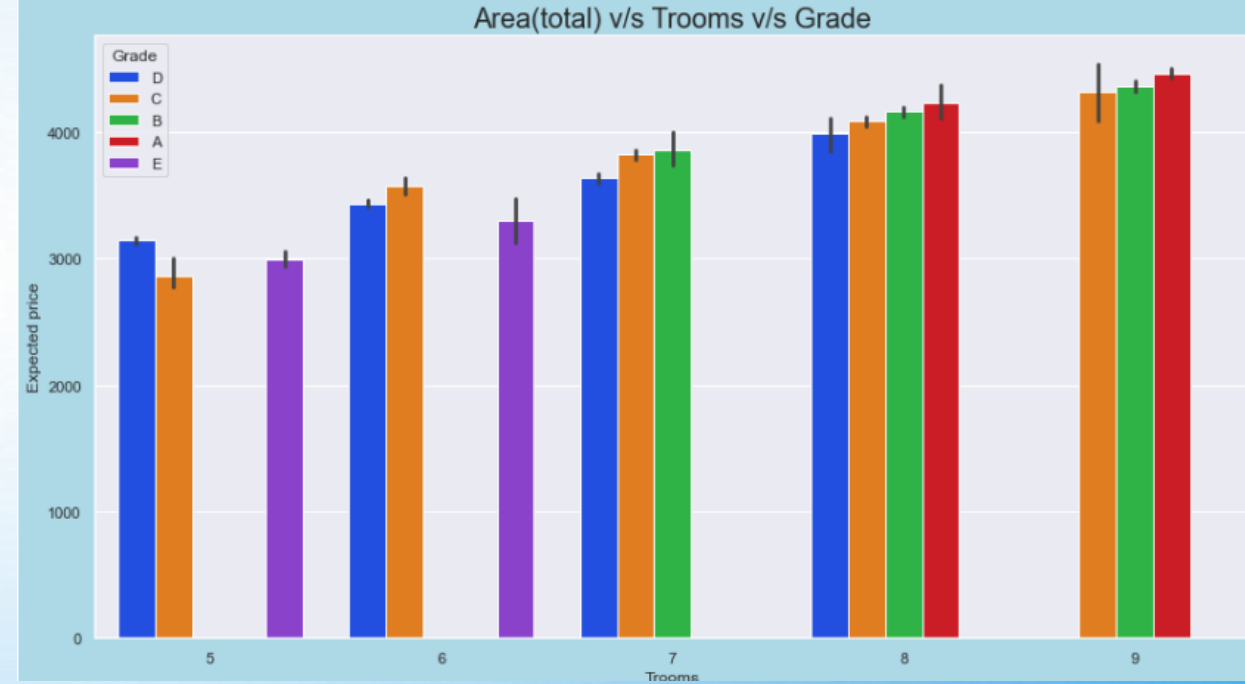
> HERE WE CAN SEE THE DISTRIBUTION OF HOUSE GRADES WITH RESPECT TO TOTAL NUMBER OF BEDROOMS AND THEIR RESPECTIVE EXPECTED PRICE

➢ HERE WE CAN SEE THE DISTRIBUTION OF HOUSE GRADES WITH RESPECT TO NUMBER OF ATTACHED WASHROOMS AND THEIR RESPECTIVE EXPECTED PRICE
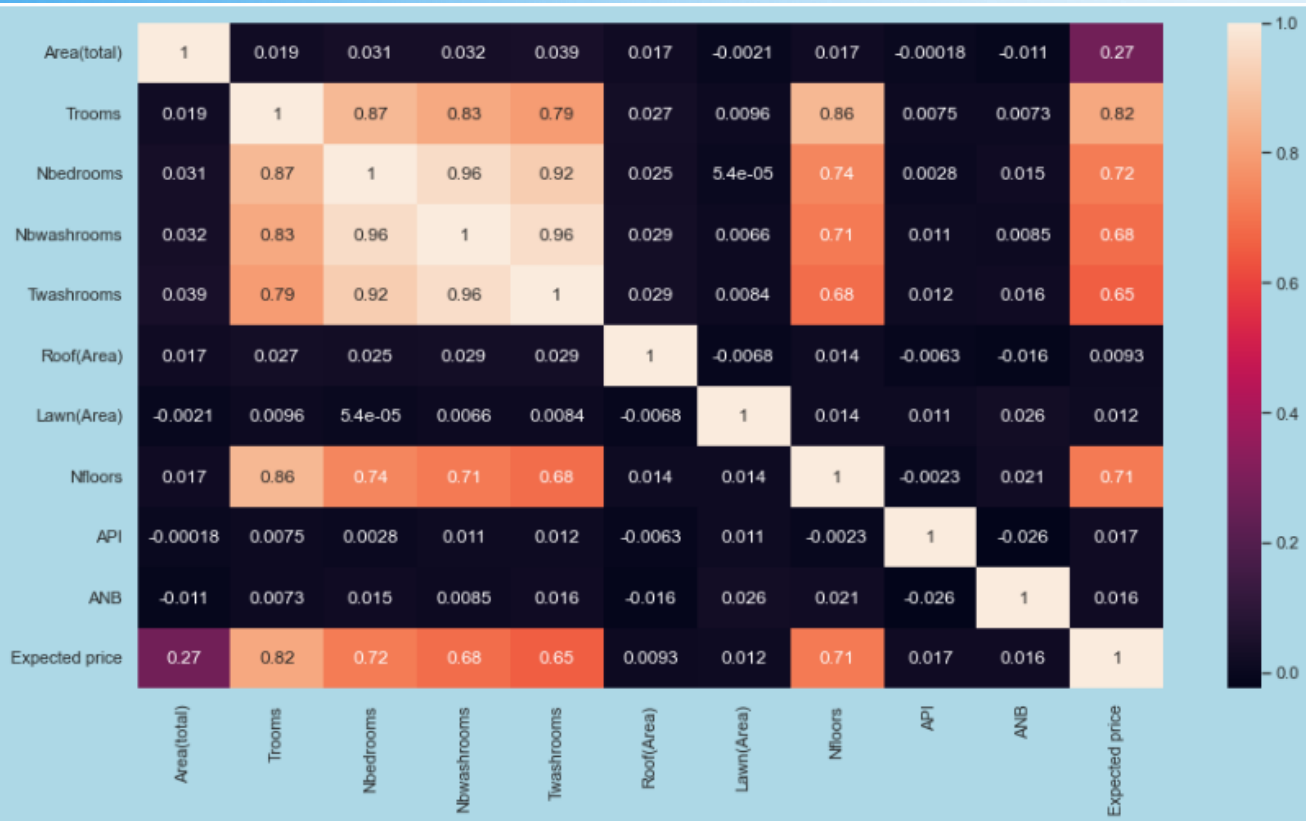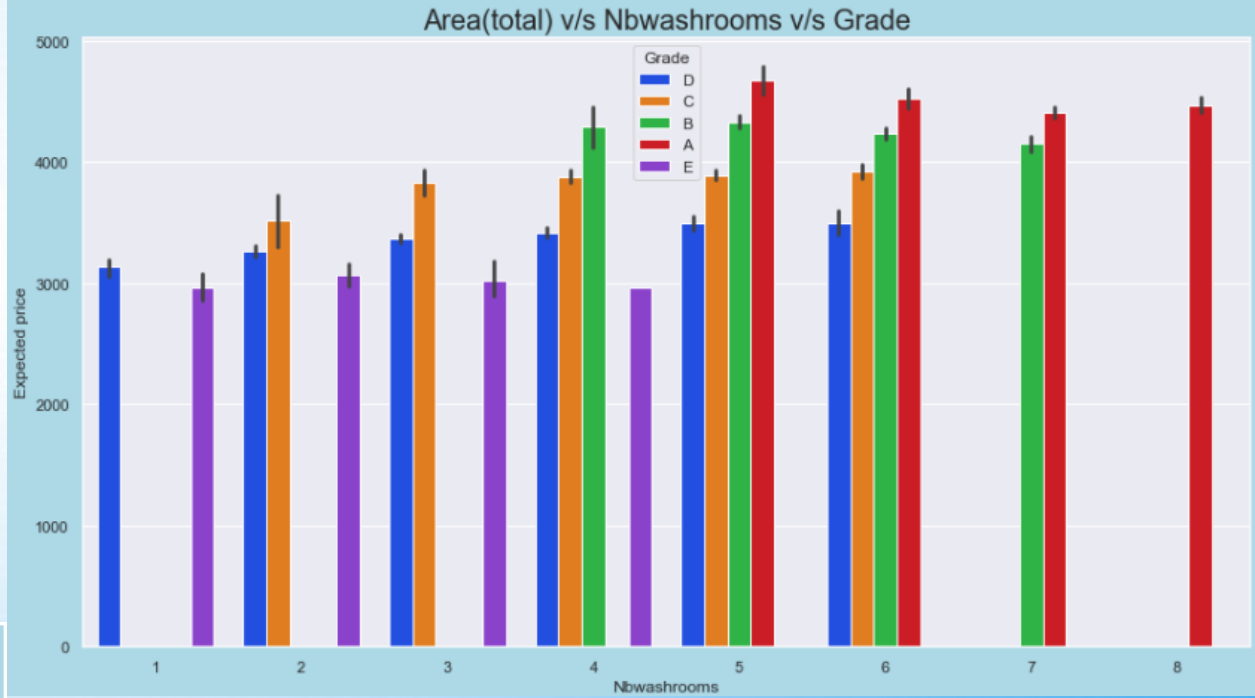
➢ WE CAN SEE HOW THE CORRELATION EXISTS BETWEEN ALL THE FEATURES

FEATURE ENGINEERING

FEATURE TRANFORMATION

- TO TRANSFORM ALL THE FEATURES INTO NUMERICAL DATATYPE
- LABEL ENCODING TECHNIQUE IS USED FOR FEATURE TRANSFORMATION

FEATURE SCALING

- TO GET ALL THE FEATURE INTO SIMILAR RANGE
- SCALING IS DONE THROUGH NORMALIZATION BECAUSE OUTLIERS ARE NOT PRESENT IN THE FEATURES AND THE NORMALIZATION IS IS NOT GOOD WITH HANDLING THE OUTLIERS

FEATURE REDUCTION

- WE CAN SEE THAT THERE IS MORE RELATION EXISTS BETWEEN INDEPENDENT FEATURES IN ORDER TO INCREASE THE PERFORMANCE WE NEED DO PERFORM LDA.

SPLITTING TECHNIQUE

- STRATIFIEDKFOLD TECHNIQUE IS USED IN THIS PROJECT

MODEL BUILDING

EDA OBSERVATIONS

- THE DEPENDENT COLUMN IS DISCREATE.

- THE OUTLIERS ARE NOT PRESENT IN THE INDEPENDENT FEATURES AS WELL AS DEPENDENT FEATURE

- THE BIAS IS MORE IN THE DEPENDENT COLUMN

ALGORITHM SELECTION

- THE ALGORITHM SHOULD BE CLASSIFIER

- AS THE OUTLIERS ARE NOT PRESENT IN THE FEATURES SO ANY ALGORITHM CAN BE SELECTED

- THE BIAS IS MORE SO WE NEED TO SELECT THE ALGORITHM WHICH LOWER THE BIAS

- WE CAN TAKE BOOSTING OR STACKING WHICH HAS ALL THE CAPABILITIES TO GET BETTER PERFORMANCE BY LOWERING THE BIAS BUT TO COMPARE WITH OTHER ALGORITHMS KNN, DECISION TREE, BOOSTING AND STACKING ALGORITHMS ARE SELECTED

MODEL PERFORMANCE AND EVALUATION RESULTS

➢ FROM THE TABLE WE CAN CONCLUDE THAT STACKING WITH FEATURE REDUCTION IS SHOWING GOOD PERFORMANCE COMPARED TO ALL OTHER ALGORITHMS

	Algorithms	Train	Test	Accuracy	Precision	Recall	F1_score
0	KNN	0.89	0.87	0.87	0.88	0.87	0.87
1	Decission Tree classifier	0.77	0.77	0.77	0.76	0.77	0.76
2	Boosting	0.76	0.75	0.75	0.74	0.75	0.74
3	Stacking without feature reduction	0.95	0.94	0.94	0.94	0.94	0.94
4	Stacking with feature reduction	0.96	0.95	0.95	0.95	0.95	0.95

➢ FROM THE PLOT WE CAN SEE THAT THE GRADE B IS MORE WRONGLY PREDICTED COMPARED TO ALL OTHER GRADES
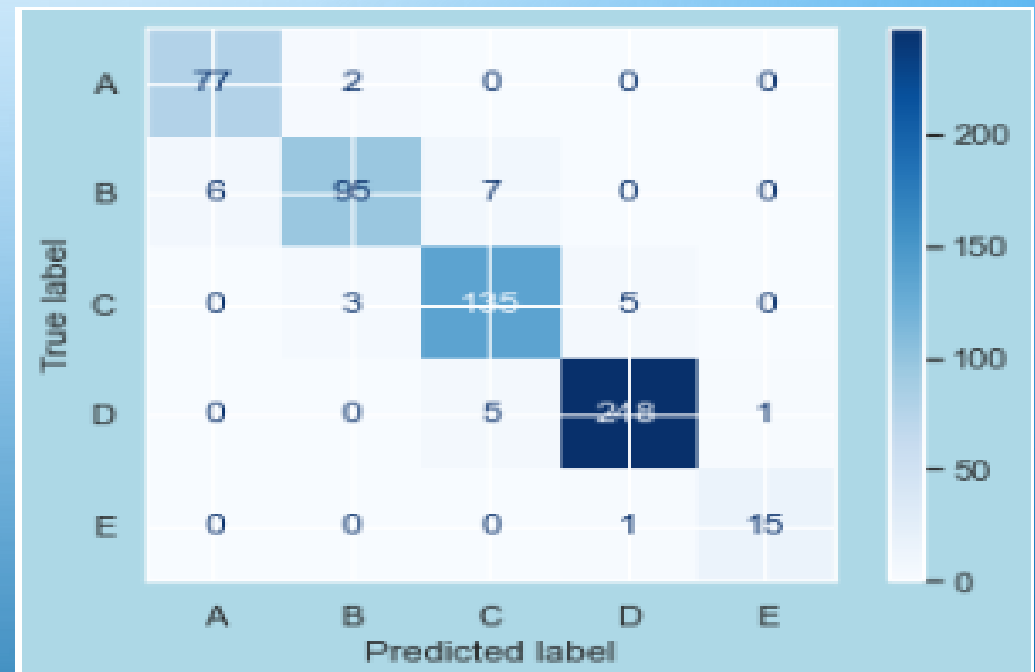
CONCLUSION

- FOR MORE SPACE, ROOMS, FLOORS AND FOR MORE BENEFITS A GRADE IS GOOD AND THE COST IS ALSO HIGH

- FOR BIG FAMILY WITH LESS COST C GRADE HOUSE IS BEST

- LESS RENT COST WITH HIGHEST NUMBER OF WASHROOMS ATTACHED WITH BEDROOMS B GRADE HOUSES IS BEST

- LESS RENT COST WITH MEDIUM SIZE FAMILY CAN CHOOSE D GRADE HOUSES

- FOR BACHELORS OR MARRIED COUPLES WITH LESS MEMBERS CAN CHOOSE E GRADE HOUSES

- THE STACKING IS GIVEN BETTER PERFORMANCE BECAUSE THE BIAS WAS THERE IN THE DATASET SO THE PERFORMANCE OF THE STACKING MODEL IS GOOD

- THE PREDICTIONS ARE HIGHLY ACCURATE

THANK YOU