AMES DEVELOPMENT PROJECT

RAFI RAHMAN

PROBLEM STATEMENT & RESEARCH

- Developing Ames the right way
- Understanding initial cost of development, by calculating land cost
- Fairly compensating home owners

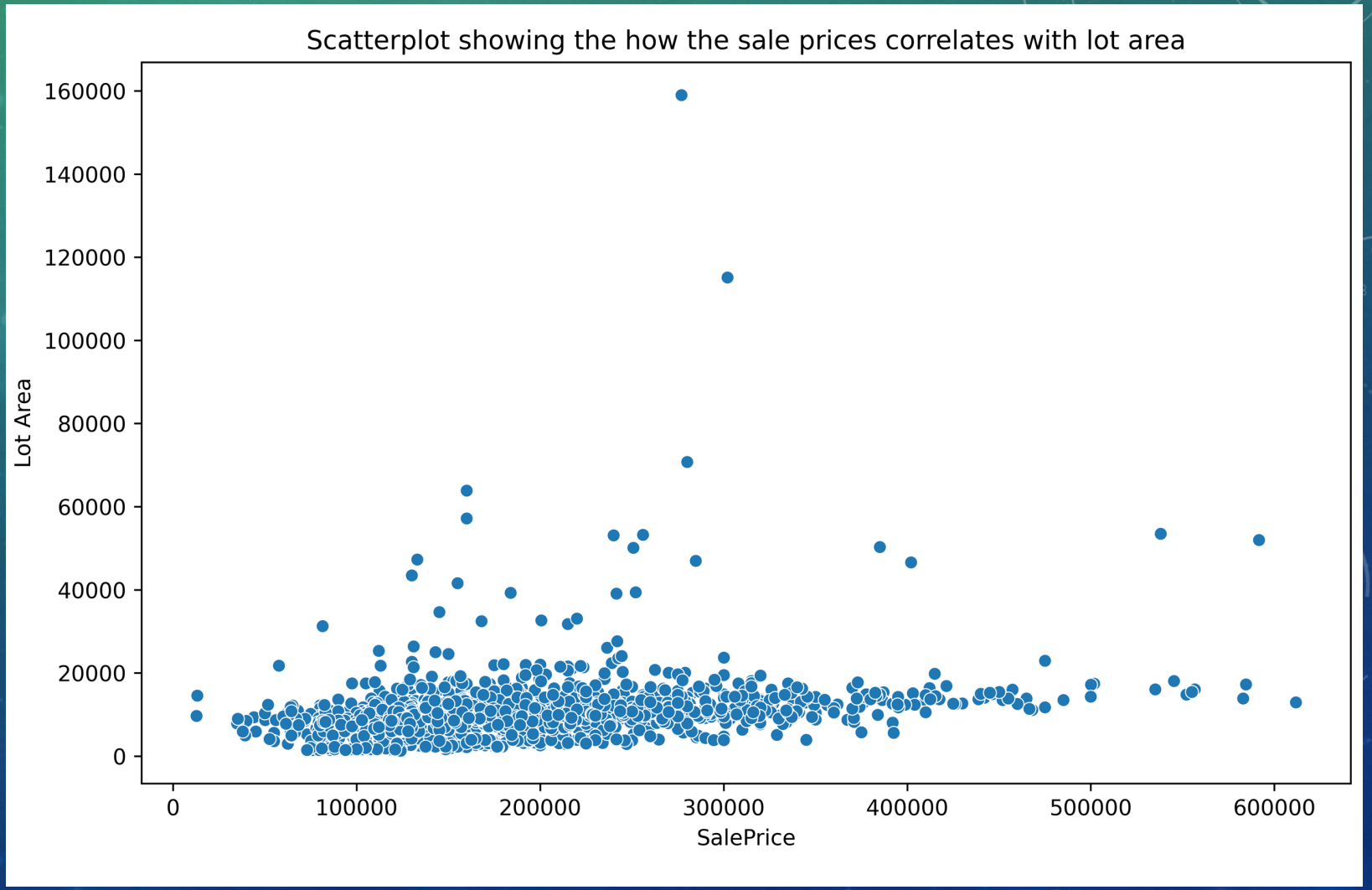
- “Neom, Saudi Arabia - US$500 billion”
- “California High-Speed Rail, US – US$113 billion”
- 6 months to 6 years in preplanning before ground is broken

METHODOLOGY

- Identify most relevant parameter/house features
- Focusing on square footage and area
- Create baseline model
- Introduce Preprocessing and Feature Engineering techniques to improve models performance

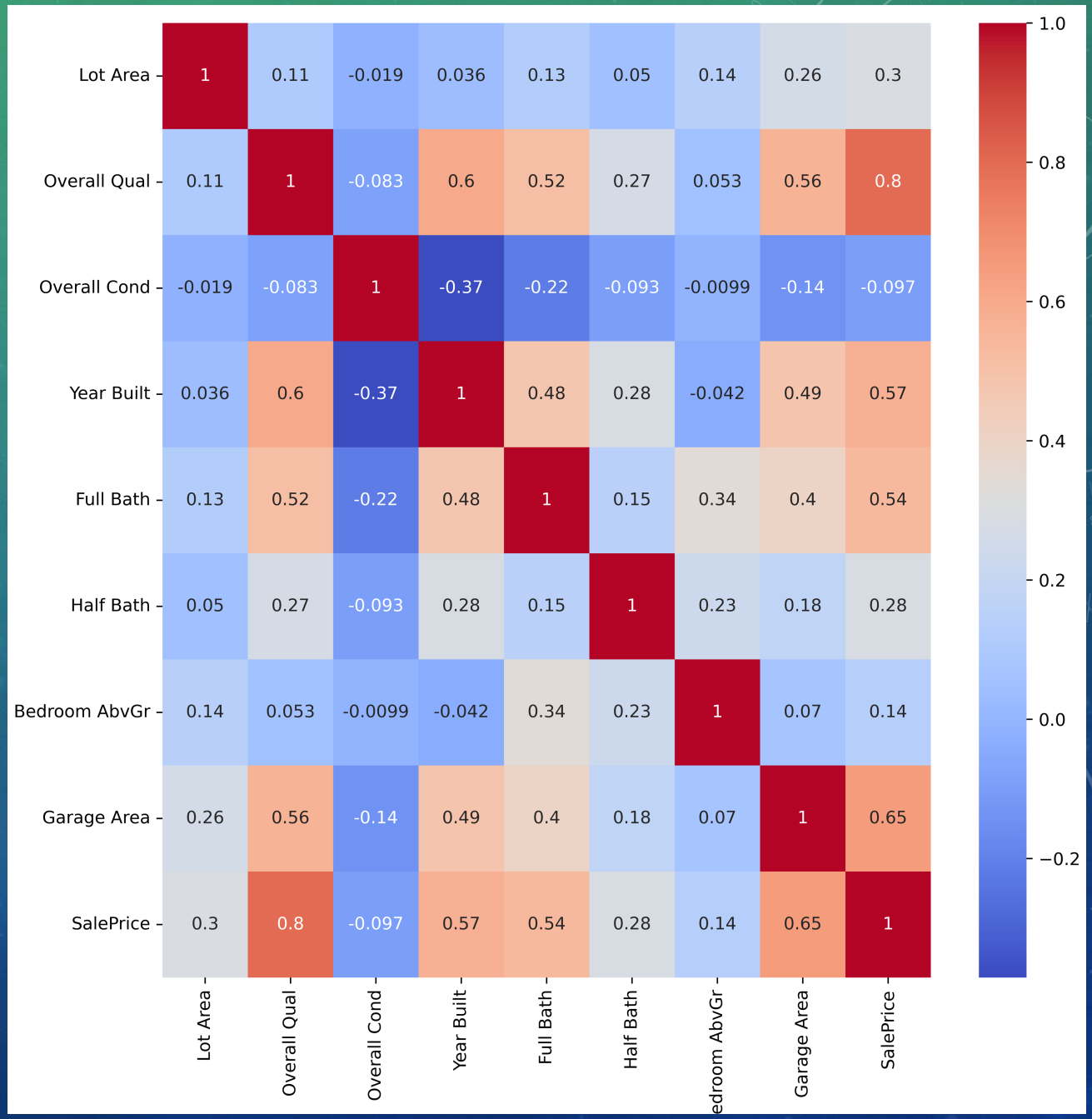
EXPLORATORY DATA ANALYSIS

- No obverse correlation between sales price and lot area



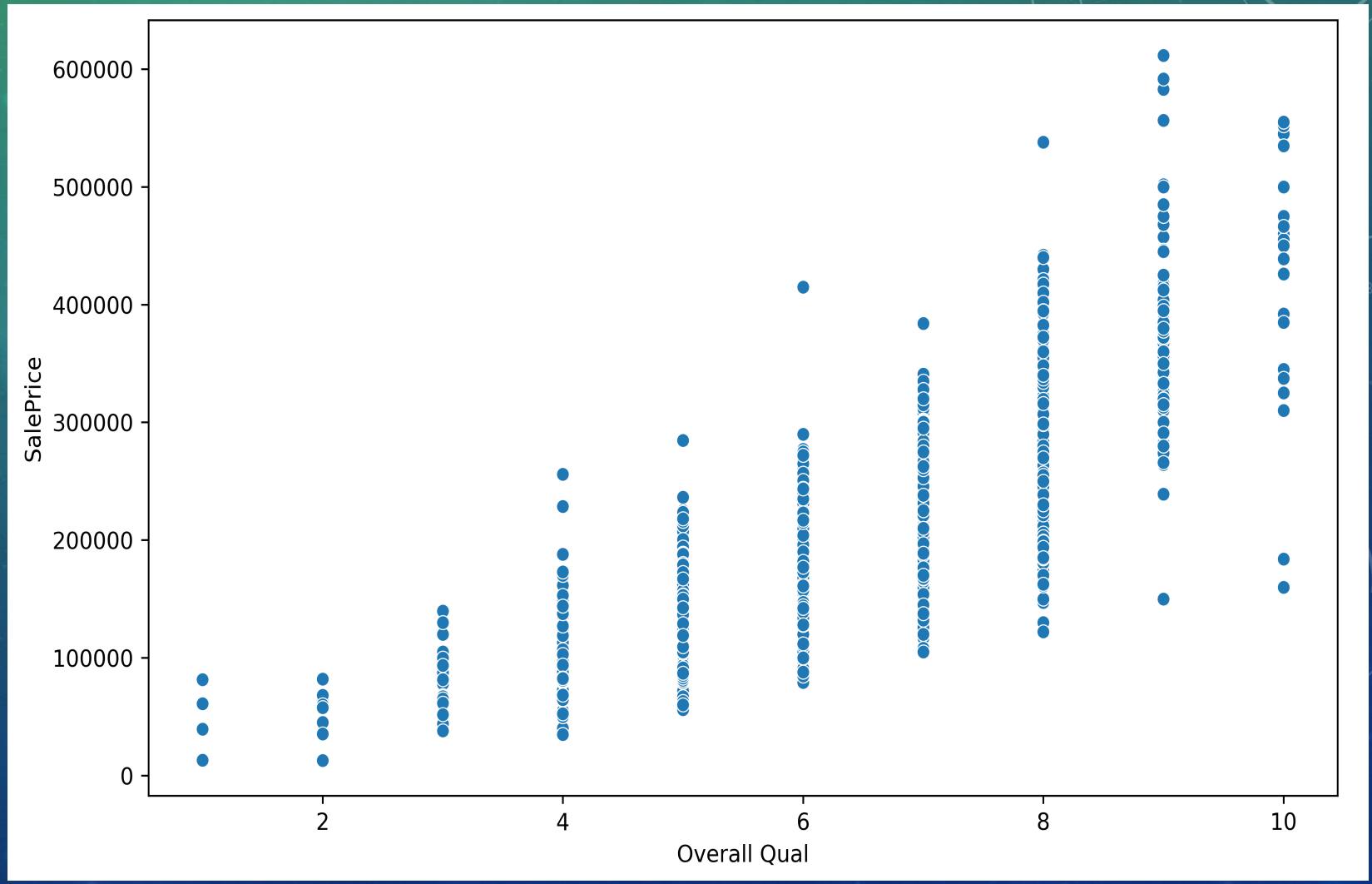
EXPLORATORY DATA ANALYSIS

- No obverse correlation between sales price and lot area
- Overall Quality, Garage Area, Year Built & Full Bath showing higher correlation to the Sale Price.



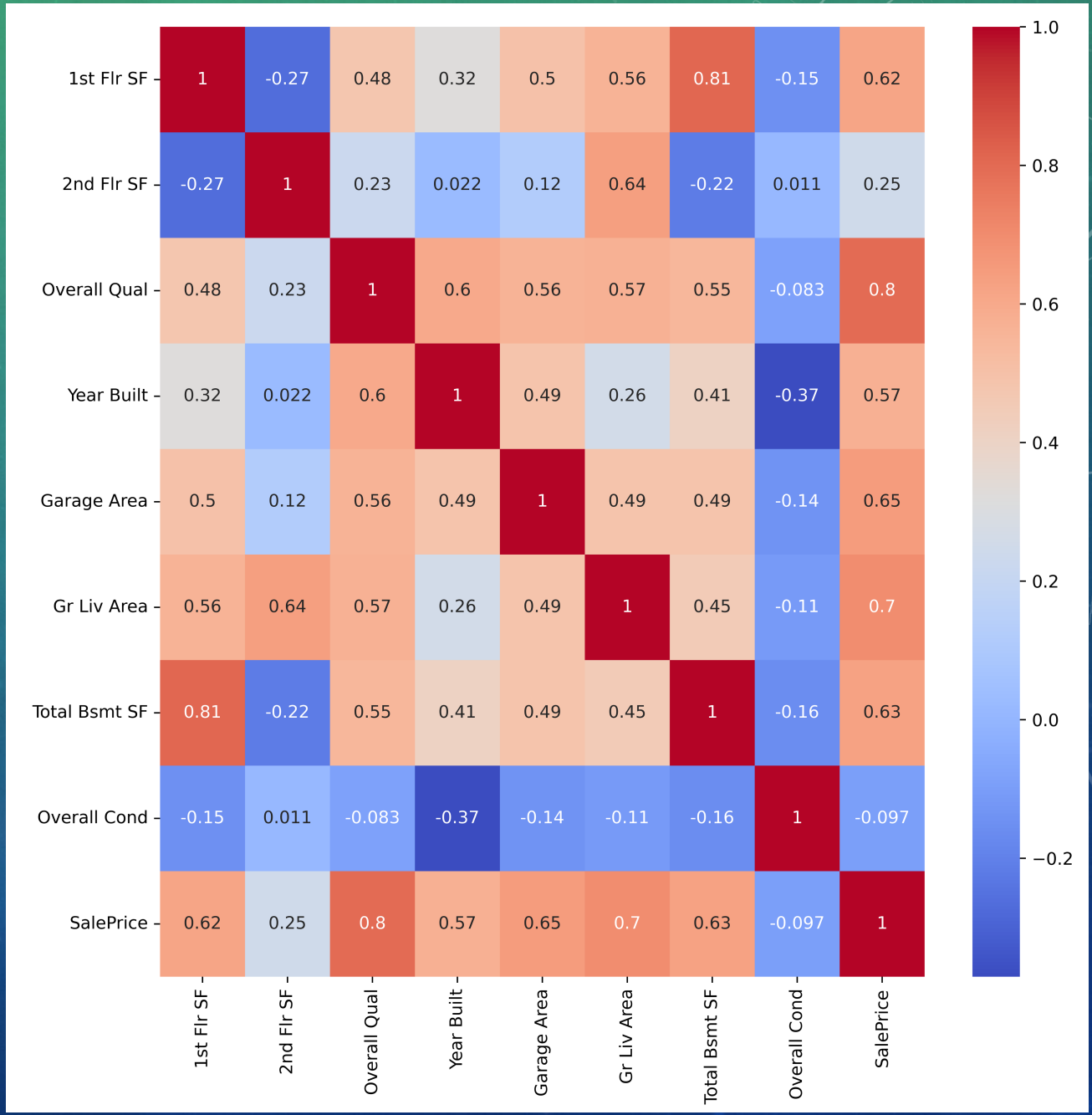
EXPLORATORY DATA ANALYSIS

- No obverse correlation between sales price and lot area
- Overall Quality, Garage Area, Year Built & Full Bath showing higher correlation to the Sale Price.
- We can see how the Sales price increases as the overall quality increases.



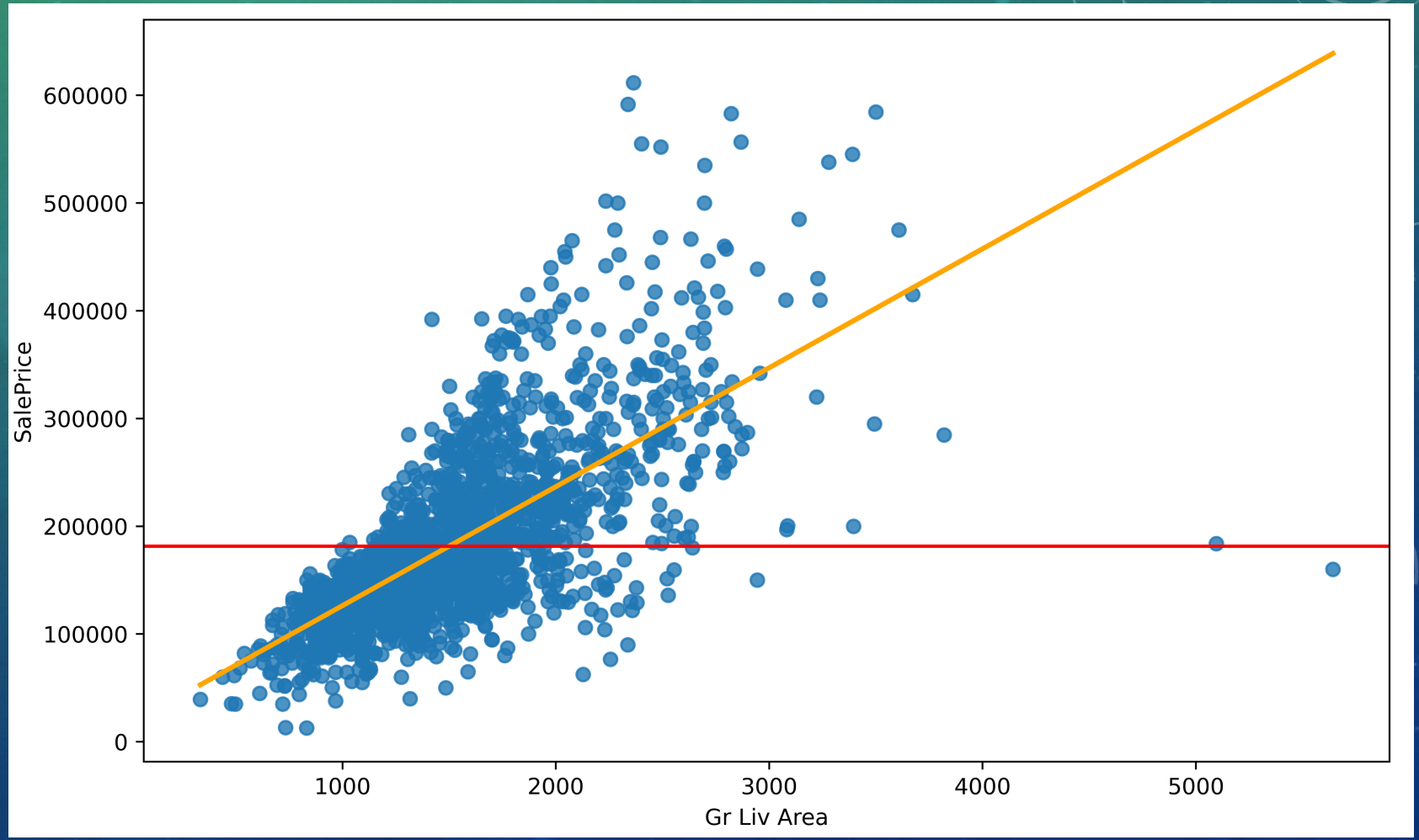
EXPLORATORY DATA ANALYSIS

- No obverse correlation between sales price and lot area
- Overall Quality, Garage Area, Year Built & Full Bath showing higher correlation to the Sale Price.
- We can see how the Sales price increases as the overall quality increases.
- Total Basement SF, 1st Floor SF and Ground living area



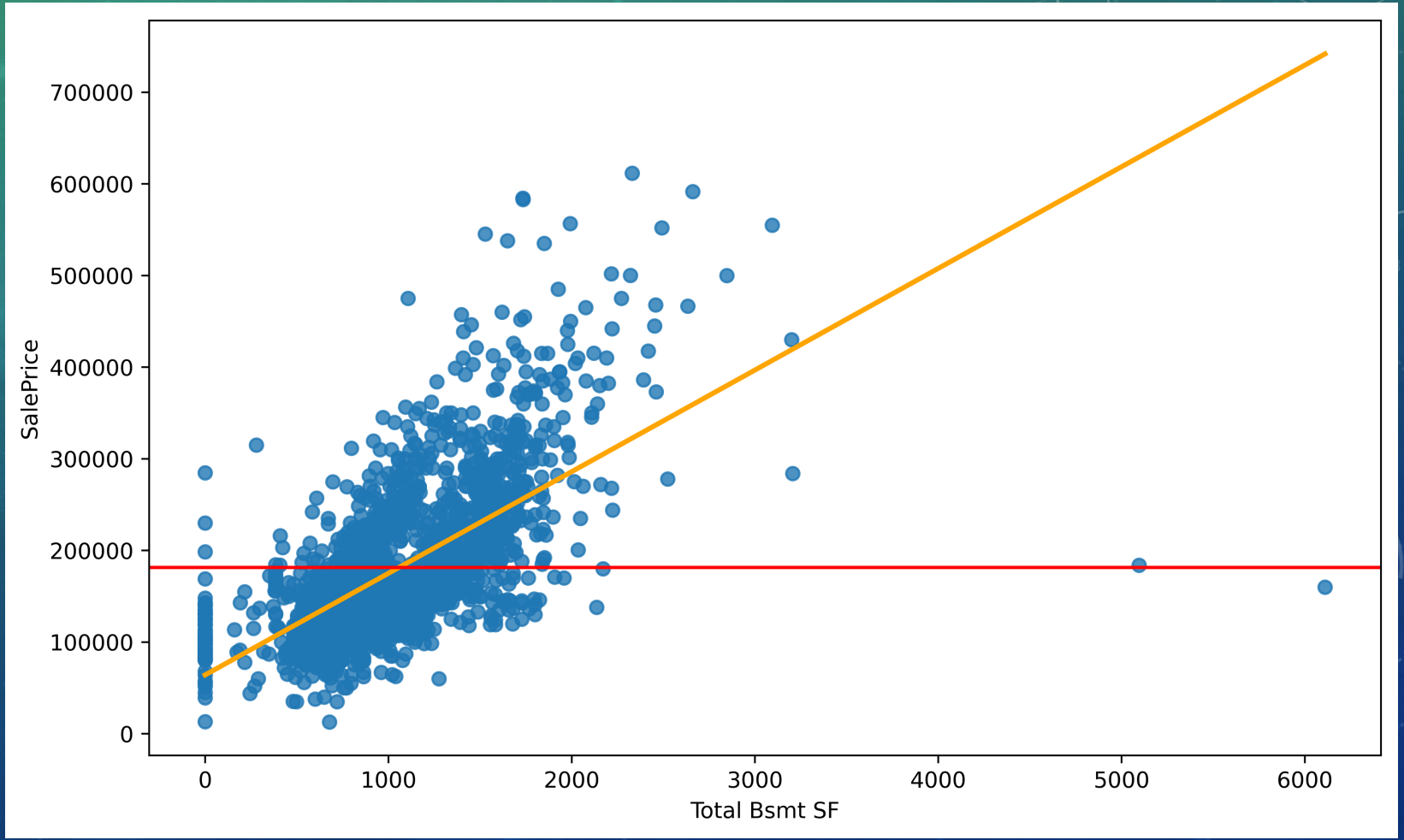
EXPLORATORY DATA ANALYSIS

- No obverse correlation between sales price and lot area
- Overall Quality, Garage Area, Year Built & Full Bath showing higher correlation to the Sale Price.
- We can see how the Sales price increases as the overall quality increases.
- Total Basement SF, 1st Floor SF and Ground living area
- Good positive correlation with a few anomalies



EXPLORATORY DATA ANALYSIS

- No obverse correlation between sales price and lot area
- Overall Quality, Garage Area, Year Built & Full Bath showing higher correlation to the Sale Price.
- We can see how the Sales price increases as the overall quality increases.
- Total Basement SF, 1st Floor SF and Ground living area
- Good positive correlation with a few anomalies
- Helps show how other areas of the property affects Sales price



BASELINE

- Let's have a look at the data that we will be using.

	Id	Overall Qual	Year Built	Full Bath	Half Bath	Garage Area	Total Bsmt SF	Gr Liv Area	1st Flr SF	2nd Flr SF	SalePrice
0	109	6	1976	2	1	475.0	725.0	1479	725	754	130500
1	544	7	1996	2	1	559.0	913.0	2122	913	1209	220000
2	153	5	1953	1	0	246.0	1057.0	1057	1057	0	109000
3	318	5	2006	2	1	400.0	384.0	1444	744	700	174000
4	255	6	1900	2	0	484.0	676.0	1445	831	614	138500

BASELINE

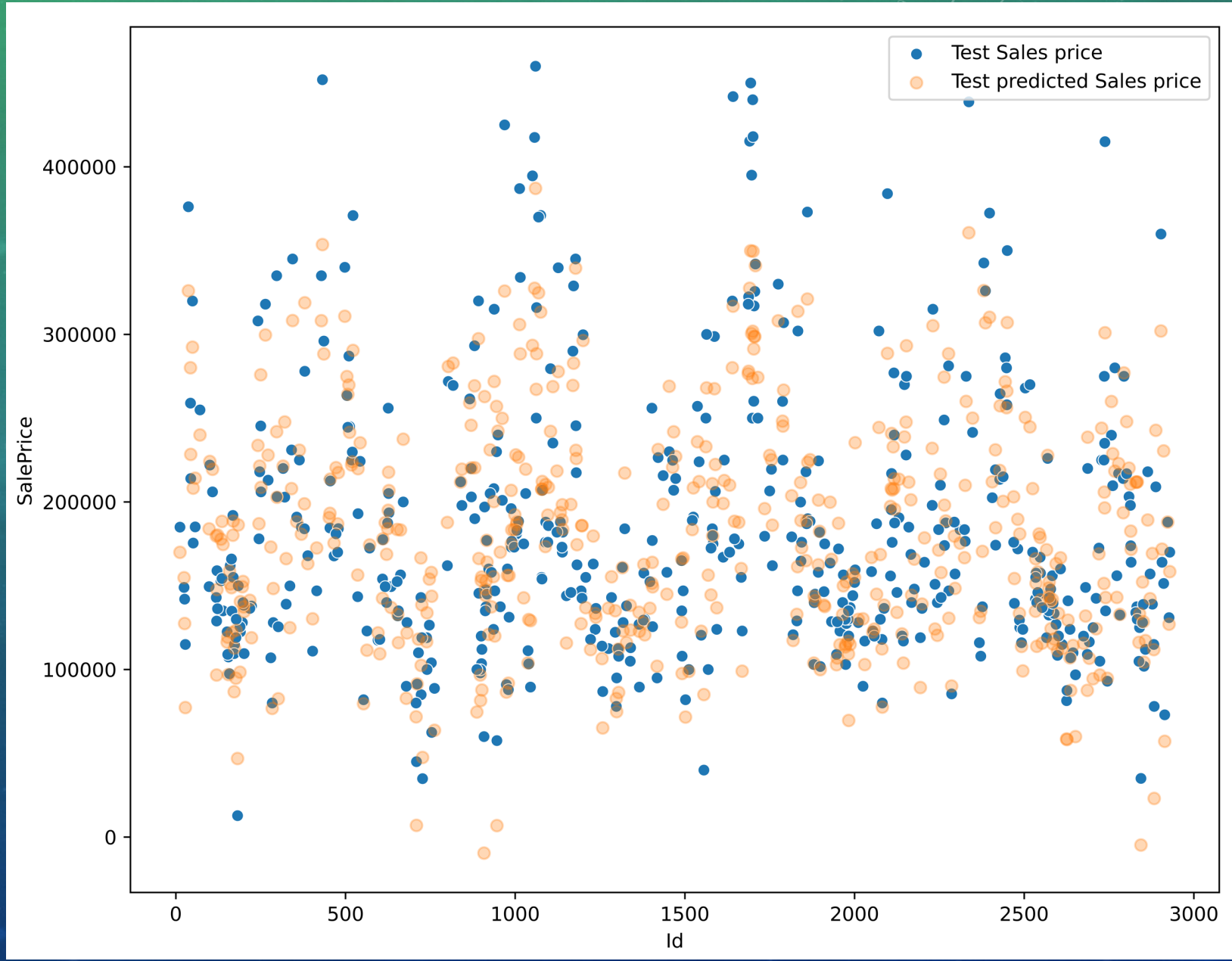
- Let's have a look at the data that we will be using.
- Metrics Metrics Metrics!!!!

	Baseline
R^2 Train score	0.771023
R^2 Validation score	0.835513
MAE	23307.513959
RMSE	31779.748064

BASELINE

- Let's have a look at the data that we will be using.
- Metrics Metrics Metrics!!!!
- Comparing the given provided Sales prices with the predicted

	Baseline
R^2 Train score	0.771023
R^2 Val score	0.835513
MAE	23307.513959
RMSE	31779.748064



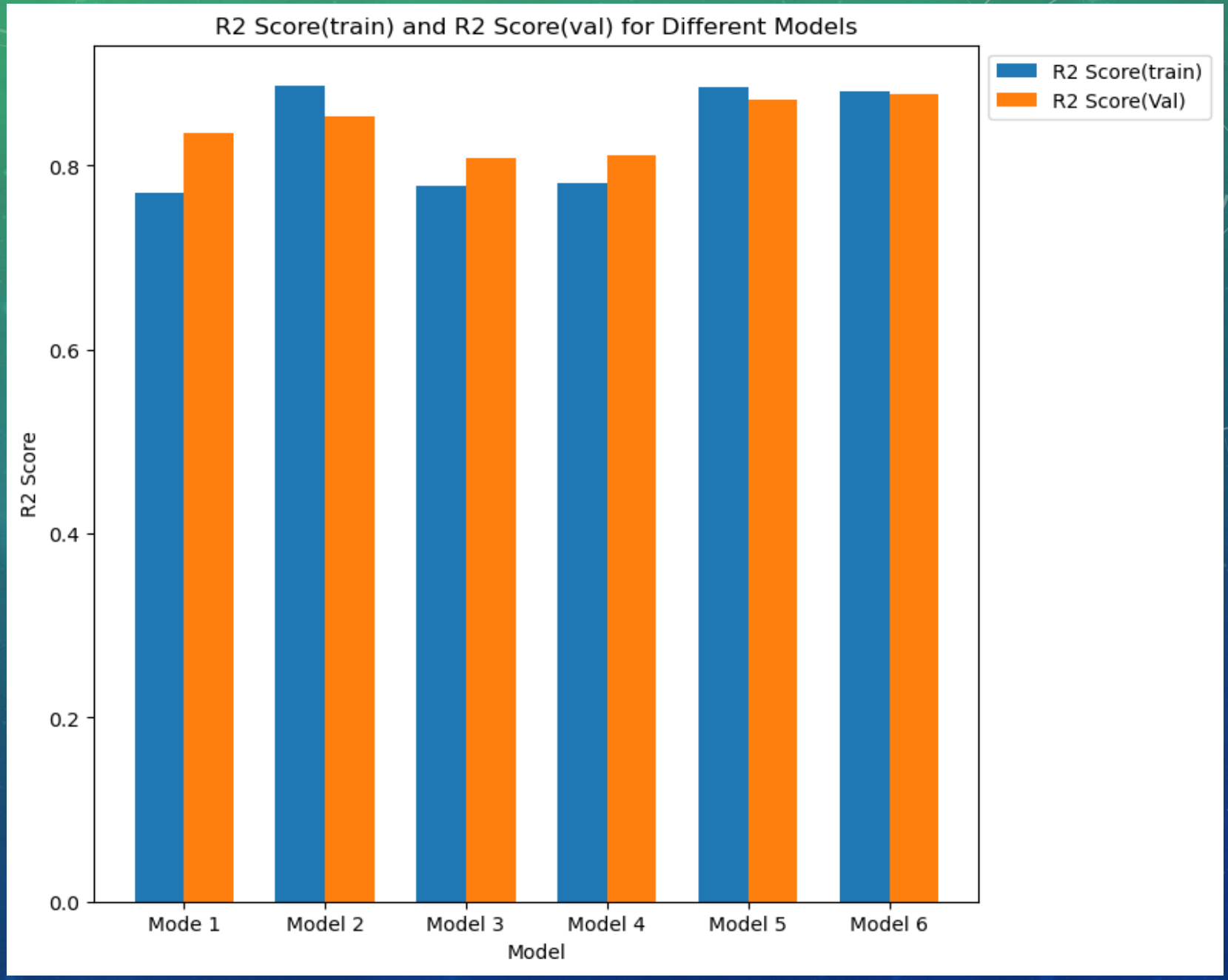
MODEL TUNING

- Polynomial Features
- Standard Scaling
- Both
- ≈$4000k drop in MAE
- Linear Regression vs Ridge CV
  - Regularization
  - Handling Multicollinearity

	Model	R2 Score(train)	R2 Score(Val)	MAE	RMSE	Description
0	Mode 1	0.771023	0.835513	23307.513959	31779.748064	Baseline
1	Model 2	0.886713	0.852980	21157.810887	31094.158241	Poly
2	Model 3	0.777809	0.807667	25021.056681	35564.524098	Dropped Full Bath and Half Bath
3	Model 4	0.780299	0.811278	24684.392809	35229.123817	StandardScaled
4	Model 5	0.885289	0.871826	19836.248501	28053.375859	Poly and SC
5	Model 6	0.880983	0.877115	19523.526618	27468.513989	RidgeCv

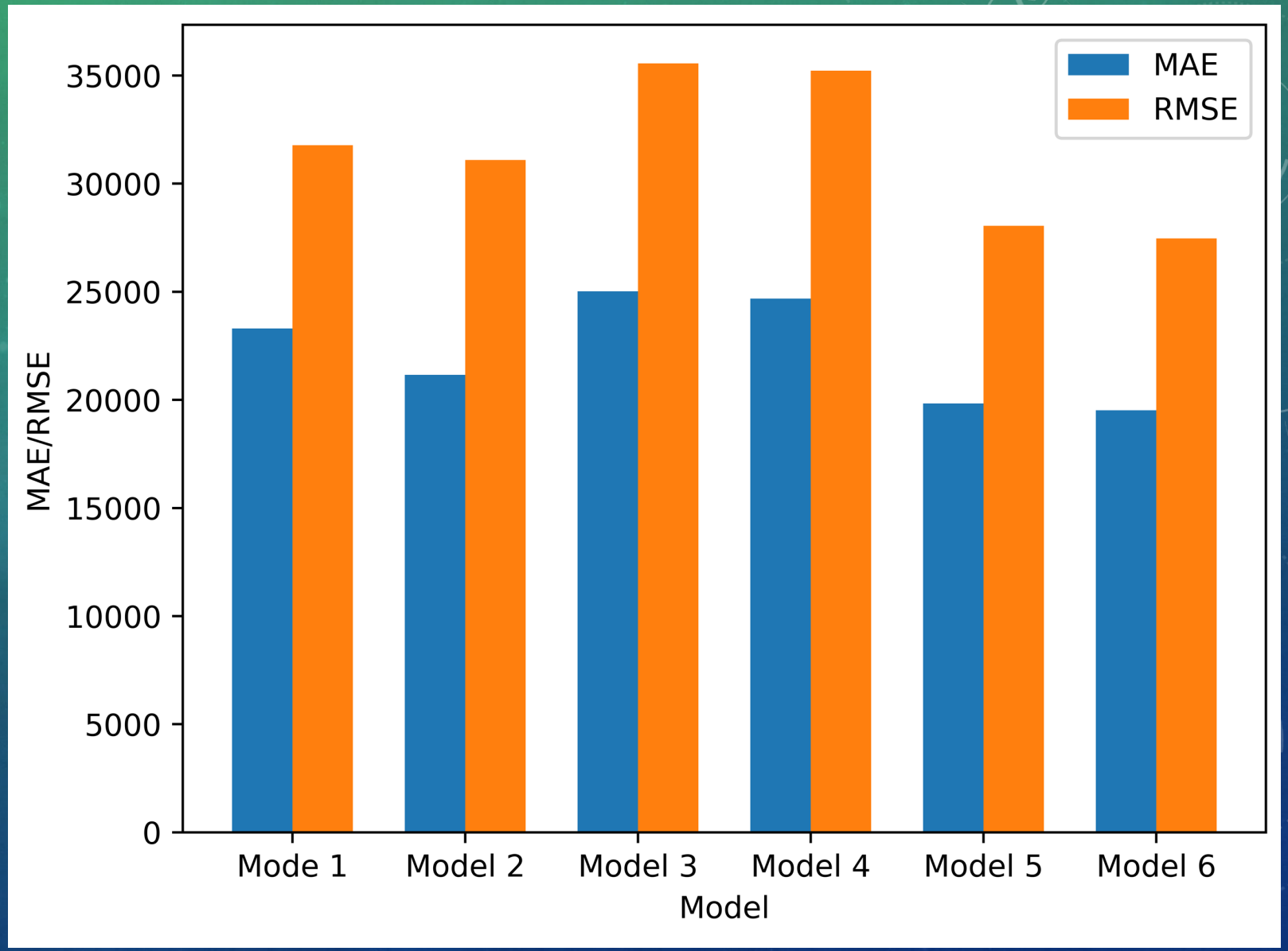
MODEL TUNING

- Polynomial Features
- Standard Scaling
- Both
- ≈$4000k drop in MAE
- Linear Regression vs Ridge CV
  - Regularization
  - Handling Multicollinearity



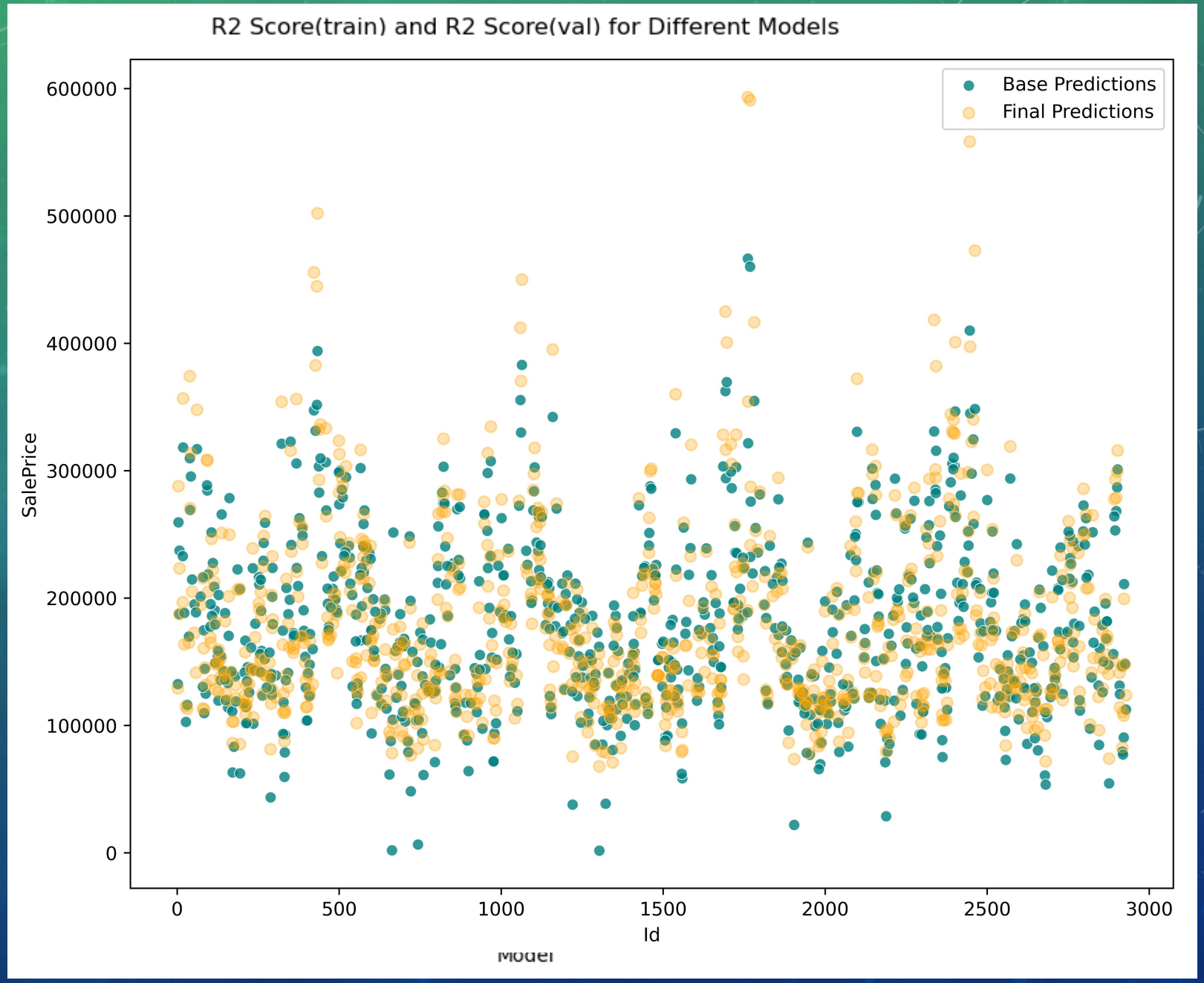
MODEL TUNING

- Polynomial Features
- Standard Scaling
- Both
- ≈$4000k drop in MAE
- Linear Regression vs Ridge CV
  - Regularization
  - Handling Multicollinearity



MODEL TUNING

- Polynomial Features
- Standard Scaling
- Both
- ≈$4000k drop in MAE
- Linear Regression vs Ridge CV
  - Regularization
  - Handling Multicollinearity



CONCLUSION & RECOMMENDATIONS

- Predicted total costing of buying all the properties = $157,135,272.02
- Variation of = $19523.526618
- ≈ $17,141,656.37
- Recommendations
  - Introducing more variables
  - Deeper dive into Neighborhood and House style
  - Investigate if Year of remodel and additional constructions changes the model

THANK YOU FOR LISTENING

ANY QUESTIONS?

CITATIONS AND THANKS

Thanks to Rowan and Tim for answering all my endless questions.

Background picture: https://www.aimircg.com/3d-architectural-rendering-for-a-real-estate-development-in-montana-usa-case-study/

7 Most Expensive Construction Projects: https://www.international-construction.com/news/7-of-the-world-s-most-expensive-construction-projects/8026711.article

Real Estate Development Timeline: https://lev.co/blog/assets/real-estate-development-timeline/