

Title Screen	Executive Summary	Economic Activity and Sale Price Analysis	As a Business Strategy; Remodelation	The Influence of Space	House Age and Last Construction Age	Overall Quality	Neighborhoods	Houses in the Current Market	Regression Analysis & Predictive Model	Target Houses & Conclusion	A b. .
--------------	-------------------	---	--------------------------------------	------------------------	-------------------------------------	-----------------	---------------	------------------------------	--	----------------------------	--------

7/17/2020

# Business Intelligence and Modeling Analysis for House Price Predictions

Presented By: Huseyin Korkmaz

How can  
**BlueStone Realty**  
increase its revenue  
by at least 16% until the end of 2010?



Source : Ames, Iowa Housing Data (Kaggle)  
Note : According to our business scenario, we used the train data as the data of sold houses, and used the test data as the data of houses in the current market.

Title Screen	Executive Summary	Economic Activity and Sale Price Analysis	As a Business Strategy; Remodelation	The Influence of Space	House Age and Last Construction Age	Overall Quality	Neighborhoods	Houses in the Current Market	Regression Analysis & Predictive Model	Target Houses & Conclusion	A b. .
--------------	-------------------	---	--------------------------------------	------------------------	-------------------------------------	-----------------	---------------	------------------------------	--	----------------------------	--------

BlueStone Realty noticed that the yearly actual revenue is \$4M less than the forecast revenue as of July 2010. It is necessary to increase the revenue by at least 16% with respect to the budget of \$25M until the end of 2010. Actions should be taken to have more revenue opportunities by determining the top houses which are expected to provide the most revenue opportunities in the case of t..

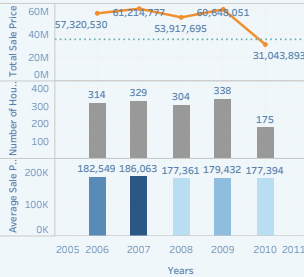
As of July 2010, BlueStone's actual revenue is \$4M behind budget revenue. In order to catch the revenue of 2009, BlueStone should reach the target of \$29M revenue in the remaining 5 months of the year by increasing its \$25M budget revenue by at least 16%.

Our analysis shows that the quality of the house is one of the factors that mostly impacts Sale Price. According to our analysis, it seems house remodeling makes overall quality increased by 1 point. Our main aim is to identify the revenue opportunities that will be occurred in the case of remodeling houses that causes increasing overall quality. By using our regression analysis, we can identify the top houses and calculate that if we sell these target houses by remodeling them, how much more revenue we will gain.

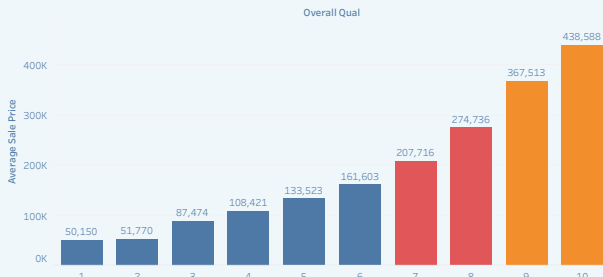
With our predictive model, we identified the top 189 houses which are expected to provide the most revenue opportunities in the case of the newly-remodeling. This strategy can provide \$6.7M more revenue opportunities and increase revenue by ~27%.



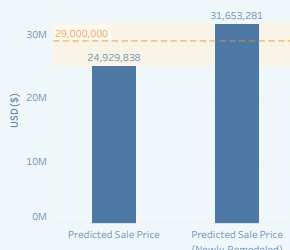
Sales per Year



Average Sale Price w.r.t. Overall Quality of Houses



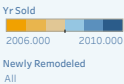
Predicted Total Sale Prices (Revenue Opportunity)



Capstone 2 - Presentation

Title Screen	Executive Summary	Economic Activity and Sale Price Analysis	As a Business Strategy; Remodelation	The Influence of Space	House Age and Last Construction Age	Overall Quality	Neighborhoods	Houses in the Current Market	Regression Analysis & Predictive Model	Target Houses & Conclusion	A b .
--------------	-------------------	---	--------------------------------------	------------------------	-------------------------------------	-----------------	---------------	------------------------------	--	----------------------------	-------

While normally June and July are the months in which most houses are sold and most revenues are gained, as of June 2010, BlueStone Realty’s revenue started severely decreasing. Despite the running total sales were expected \$35M in July, it is around \$31M. To keep the balance, BlueStone should reach at least \$29M revenue in the remaining 5 months of the year.

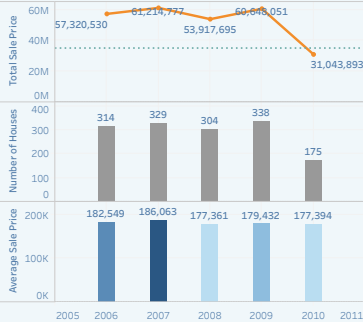


Less houses were sold during the year of 2010 compared to the other years. The average house price was at its peak in 2007 (\$186K) and it was at its lowest point during the year of 2008 and 2010 (\$177K) a difference of \$11K.

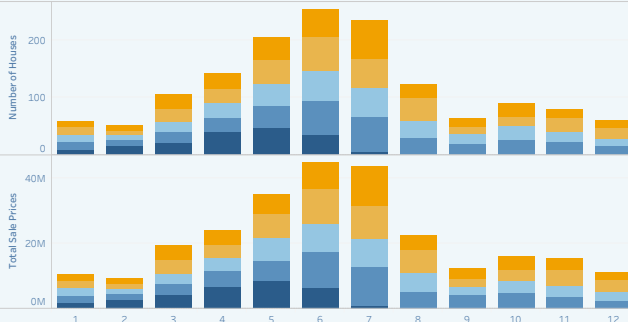
While June and July are the months in which most houses are sold, this period of 2010 is “the worst June and July ever” for BlueStone. Despite the monthly sale price is \$10.8M in June 2009, it is \$6.5M in June 2010.

The distribution of house prices is right skewed.

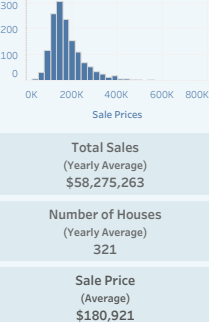
Sales per Year



Sales per Month



Distribution of Sales Prices



Title Screen	Executive Summary	Economic Activity and Sale Price Analysis	As a Business Strategy; Remodelation	The Influence of Space	House Age and Last Construction Age	Overall Quality	Neighborhoods	Houses in the Current Market	Regression Analysis & Predictive Model	Target Houses & Conclusion	A b. .
--------------	-------------------	---	--------------------------------------	------------------------	-------------------------------------	-----------------	---------------	------------------------------	--	----------------------------	--------

Our analysis on the sold houses shows that the sooner the remodeling the higher the value of the house. As a business strategy, remodeling houses before selling them can be an effective solution to receive more revenue opportunities.

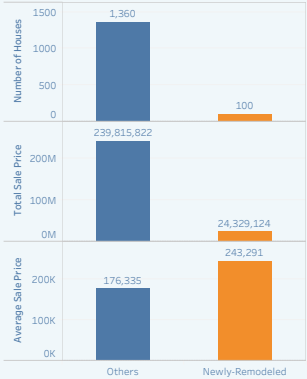
Newly Remodeled  
Others  
Newly-Remodeled

**What does it mean Newly-Remodeled House?**

If a house was remodeled one year or less time before it was sold, it shows this house was remodeled for sale. We will call these houses *Newly-Remodeled*.

We will call other houses as *Others* or *Non-Newly-Remodeled* which weren't remodeled or were remodeled more t..

Total and Average Sale Prices

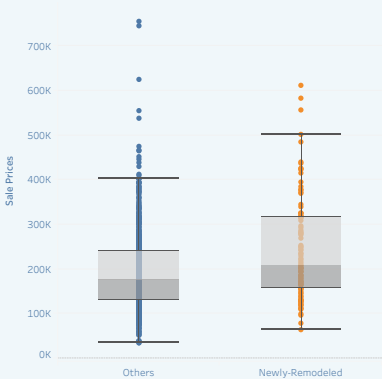


There are 100 newly-remodeled and 1360 non-newly-remodeled houses in our dataset.

The average sale price of newly-remodeled houses is 243K\$. The average sale price of non-newly-remodeled houses is 176K\$. It looks 67K\$ difference between the average sale price of newly-remodeled houses and others'.

The median of sale price of newly-remodeled houses is 208K\$. The median of sale price of non-newly-remodeled houses is 175K\$.

Sale Price Ranges

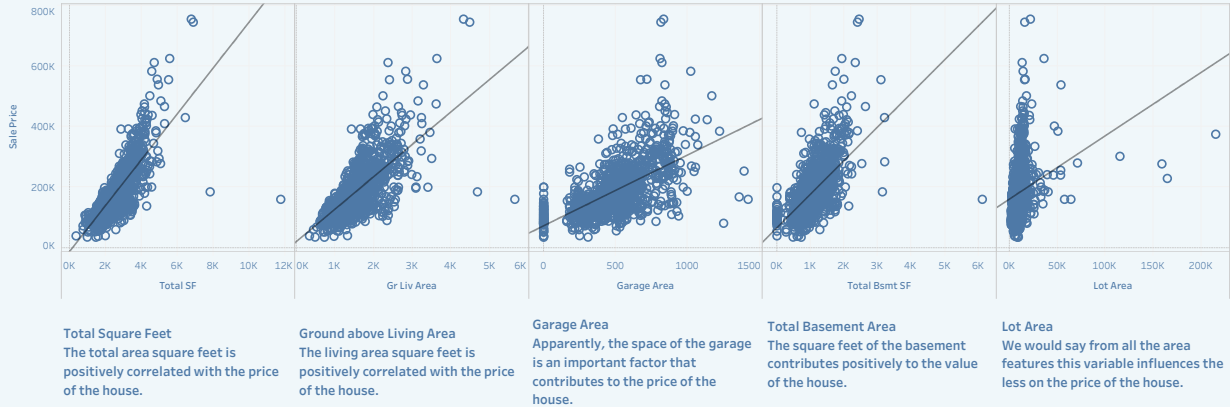


Title Screen	Executive Summary	Economic Activity and Sale Price Analysis	As a Business Strategy; Remodelation	The Influence of Space	House Age and Last Construction Age	Overall Quality	Neighborhoods	Houses in the Current Market	Regression Analysis & Predictive Model	Target Houses & Conclusion	A b. .
--------------	-------------------	---	--------------------------------------	------------------------	-------------------------------------	-----------------	---------------	------------------------------	--	----------------------------	--------

Newly Remodeled  
All

The top variables with the most predictive strength are Overall Quality, Total Area, Ground above Living Area, Garage Area, Total Basement Area, and Lot Size which are further reinforced by the scatter plots below.

Significant Numeric Values vs Sale Prices

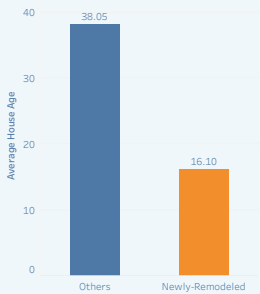


Title Screen	Executive Summary	Economic Activity and Sale Price Analysis	As a Business Strategy; Remodelation	The Influence of Space	House Age and Last Construction Age	Overall Quality	Neighborhoods	Houses in the Current Market	Regression Analysis & Predictive Model	Target Houses & Conclusion	A b.
--------------	-------------------	---	--------------------------------------	------------------------	-------------------------------------	-----------------	---------------	------------------------------	--	----------------------------	------

House Age and Last Construction Age have negative linear relationships with Sale Price. Houses in the high price range were remodeled sooner. The sooner the remodeling the higher the value of the house.

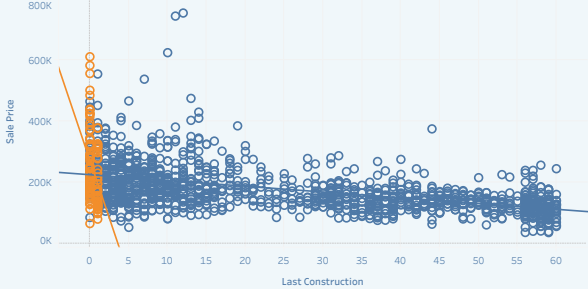
Newly Remodeled  
Newly-Remodeled  
Others

Avarege House Ages



It's not necessary for a house to be old for remodeling. While the average age of all houses is 36.5, Newly-Remodeled houses' average age is 16.

Last Construction Age



This scatter plot shows that there is a negative linear relationship between Last Construction Age and Sale Price. It seems remodeling houses before selling them can increase sale prices. This insight supports our hypothesis of the effect of newly-remodeling on the sale price.

Titl.	Executive Summary	Economic Activity and Sale Price Analysis	As a Business Strategy; Remodelation	The Influence of Space	House Age and Last Construction Age	Overall Quality	Neighborhoods	Houses in the Current Market	Regression Analysis & Predictive Model	Target Houses & Conclusion	About the Project
-------	-------------------	---	--------------------------------------	------------------------	-------------------------------------	-----------------	---------------	------------------------------	--	----------------------------	-------------------

Our analysis shows that the quality of the house is one of the factors that mostly impacts Sale Price. It seems house remodeling makes overall quality increased by 1 point. Moving the quality of a house to the next level by remodeling it can increase its Sale Price.

Clusters (wrt Avg Sale Pri...  
Cluster 1  
Cluster 2  
Cluster 3

OverallQual:  
Rates the overall material and finish of the house

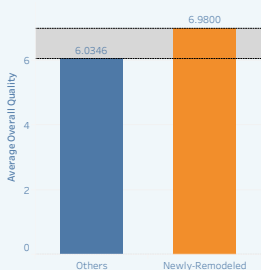
10	Very Excellent
9	Excellent
8	Very Good
7	Good
6	Above Average
5	Average
4	Below Average
3	Fair
2	Poor
1	Very Poor

Average Sale Price w.r.t. Overall Quality of Houses



It seems that the overall material that is used for construction and the finish of the house has a great impact on SalePrice.

Average Overall Quality



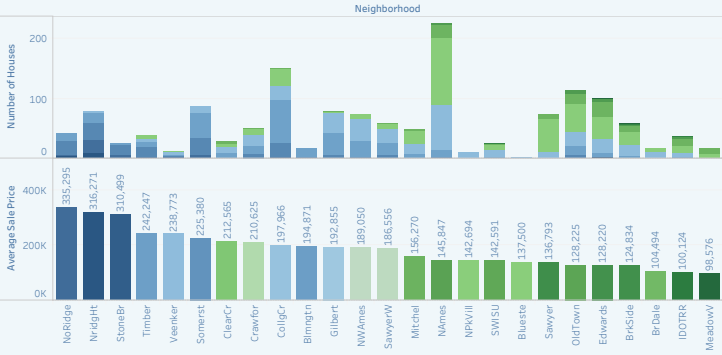
As the average overall quality of non-newly-remodeled houses is 6, Newly-Remodeled houses' is around 7. It is understood that house remodeling makes overall quality increased by 1 point.

Capstone 2 - Presentation

Titl	Executive Summary	Economic Activity and Sale Price Analysis	As a Business Strategy; Remodelation	The Influence of Space	House Age and Last Construction Age	Overall Quality	Neighborhoods	Houses in the Current Market	Regression Analysis & Predictive Model	Target Houses & Conclusion	About the Project
------	-------------------	---	--------------------------------------	------------------------	-------------------------------------	-----------------	---------------	------------------------------	--	----------------------------	-------------------

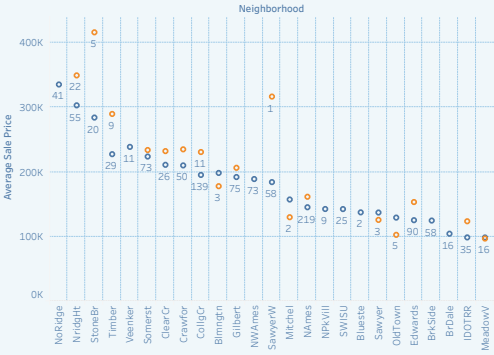
It seems that the overall quality of the house is related to the exclusiveness of neighborhoods. More expensive areas have more high qualified houses. It seems that house remodeling is more common in the areas that have more expensive houses.

Avg. Sale Price and Overall Quality w.r.t. Neighborhoods



The most exclusive Neighborhoods are Northridge, Northridge Heights, and Stone Brook. The average house price is more than 300K in these neighborhoods.

Number of Sold Houses w.r.t. Neighborhood



We can see that newly-remodeled houses generally have a higher average sale price in each neighborhood.



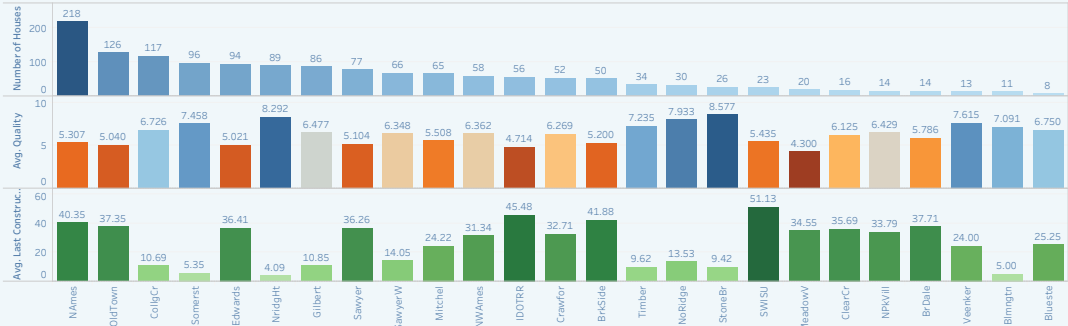
Capstone 2 - Presentation

Titl.	Executive Summary	Economic Activity and Sale Price Analysis	As a Business Strategy; Remodelation	The Influence of Space	House Age and Last Construction Age	Overall Quality	Neighborhoods	Houses in the Current Market	Regression Analysis & Predictive Model	Target Houses & Conclusion	About the Project
-------	-------------------	---	--------------------------------------	------------------------	-------------------------------------	-----------------	---------------	------------------------------	--	----------------------------	-------------------



There are 1459 houses in the current market. 218 of them with 5.3 average quality are based in North Ames. Stone Brook has highest overall quality rate with 8.5.

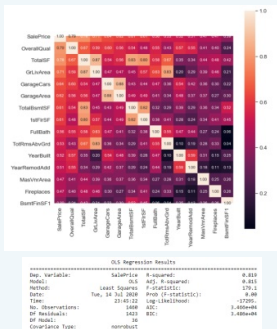
Houses in the Current Market



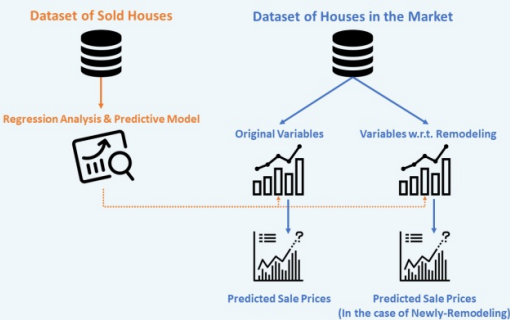
Capstone 2 - Presentation

Titl.	Executive Summary	Economic Activity and Sale Price Analysis	As a Business Strategy: Remodelation	The Influence of Space	House Age and Last Construction Age	Overall Quality	Neighborhoods	Houses in the Current Market	Regression Analysis & Predictive Model	Target Houses & Conclusion	About the Project
-------	-------------------	---	--------------------------------------	------------------------	-------------------------------------	-----------------	---------------	------------------------------	--	----------------------------	-------------------

By using regression analysis and predictive model, it is possible to identify the top houses in the market and calculate that if we sell these target houses by remodeling them, how much more revenue we will gain. We implemented regression analysis and predictive model on the data of sold houses, then applied the same model on the data of houses in the current market to see the predicted sale prices.



We chose numerical features to see if they are strong predictors of housing prices. There are a relatively positive linear relationship between Sale Price vs 'OverallQual', 'TotalSF', 'GrLivArea', 'GarageCars', 'GarageArea', 'TotalBsmSF', '1stFlrSF', 'FullBath', 'TotRmsAbvGrd', and 'YearBuilt'. With an average accuracy of 81%, the improved pricing predictive model can be utilized to estimate the sale prices of houses in the current market.



To see the sale price difference in the case of remodeling, we copied dataset of houses in the market, then modified one of them by assuming the houses were newly remodeled in 2010 and their overall quality moved to the next level. Thus we were able to compare the predicted sale prices of newly-remodeled houses and non-newly-remodeled houses.

Capstone 2 - Presentation

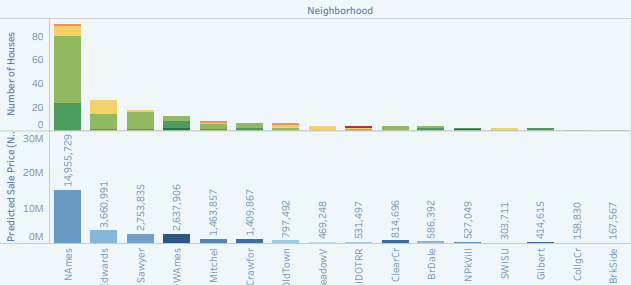
Table of Contents	Executive Summary	Economic Activity and Sale Price Analysis	As a Business Strategy: Remodelation	The Influence of Space	House Age and Last Construction Age	Overall Quality	Neighborhoods	Houses in the Current Market	Regression Analysis & Predictive Model	Target Houses & Conclusion	About the Project
-------------------	-------------------	---	--------------------------------------	------------------------	-------------------------------------	-----------------	---------------	------------------------------	--	----------------------------	-------------------

With our predictive model, we identified the top 189 houses which are expected to provide the most revenue opportunities in the case of the newly-remodeling. This strategy can provide \$6.7M more revenue opportunities and increase revenue by ~27%.



According to the budget of \$25M, we identified the top 189 houses. 89 of them with 6.1 average quality are based in North Ames where there are 218 houses with 5.3 average overall quality.

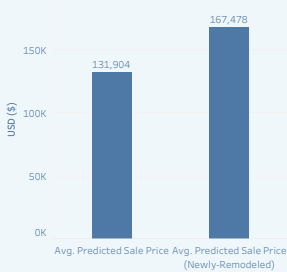
Target Houses



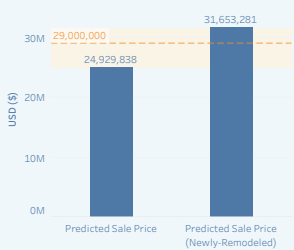
With the effects of newly-remodeling and increasing overall quality by 1 point, these target houses' average predicted sale price increased from \$131.9K to \$167.4K.

To sell target houses after remodeling them gives \$6.7M extra revenue opportunities. With this opportunity, in the rest of 2010, the revenue is expected to increase to \$31.6M.

(Predicted) Average Sale Prices



Predicted Total Sale Prices (Revenue Opportunity)



CONCLUSION

According to our business problem, BlueStone Realty wanted to have a strategy to increase revenues by at least 16% until the end of the year. After our market analysis, house remodeling strategy, regression analysis, house price predictions, and business intelligence analysis, we achieved a strategic solution that can provide increased revenue by ~27%.

Capstone 2 - Presentation

Titl.	Executive Summary	Economic Activity and Sale Price Analysis	As a Business Strategy; Remodelation	The Influence of Space	House Age and Last Construction Age	Overall Quality	Neighborhoods	Houses in the Current Market	Regression Analysis & Predictive Model	Target Houses & Conclusion	About the Project
-------	-------------------	---	--------------------------------------	------------------------	-------------------------------------	-----------------	---------------	------------------------------	--	----------------------------	-------------------

About the Project

1. How was the project picked?

- \* We chose this project because we wanted to **apply business intelligence analysis**.
- \* Instead of the classic prediction analysis approach on a housing data, we **made up a business scenario** to apply the business intelligence analysis approaches.
- \* The most important side of this project, we **liked and enjoyed it** when we analyzing this data set with a real-world business problem.

2. Acquiring the data

- \* We obtained this dataset from **Kaggle**. <https://www.kaggle.com/c/house-prices-advanced-regression-techniques/overview>
- \* We used the original training data set as the dataset of sold houses and the original test data set as the dataset of houses in the current market.



3. Data cleaning

- \* In this dataset, all NaNs are not missing values. So, we filled in these null values by "**None**" instead of dropping them.
- \* We transformed **missing values** by the median of that feature.
- \* For the more accurate analysis, we **added some features** such as 'Total Sq Footage', 'House Age', 'Remodeled', and 'Newly-Remodeled'.
- \* We **dropped a feature** that has only one unique value.



4. Data analysis

- \* We applied **descriptive and inferential analysis** to have a better understanding of the features involved in our data and especially focused on the features that have the **highest correlation towards Sale Price**.
- \* We implemented a **regression analysis** and predictive model on the data of sold houses with an average **accuracy of 81%**, then applied the same model on the data of houses in the current market to see the predicted sale prices.

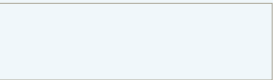


5. What's the impact?

- \* (According to our business scenario) Our main aim was to help BlueStone Realty to make more data-driven decisions by **identifying more revenue opportunities**.
- \* As a result, we were able to **determine the top houses** which are expected to provide the most revenue opportunities in the case of the newly-remodeling.
- \* While the target was 16%, we achieved a strategic solution that **can provide increased revenue by 27%**.
- \* If we find some related data that shows the cost of remodeling houses in Ames, we want to do more analysis on **profitability**.

6. Presentation format

- We preferred to use **Tableau** instead of PowerPoint for the presentation. Because we wanted;
- \* To apply what we have learned from Springboard Data Analytics Bootcamp.
- \* To use **visual advantages** on the plots during the presentation.
- \* To publish our presentation via **Tableau Public**.



7. Conclusion

- \* According to our business problem, BlueStone Realty wanted to have a strategy to increase its revenues by at least 16% until the end of the year.
- \* With our business intelligence analysis, we recommended the strategy of **remodeling houses** before the sale.
- \* Our analysis shows that this **strategic solution can provide increased revenue by ~27%**.

THANK YOU VERY MUCH FOR YOUR TIME!