

 master ▾


 3 branches













 0 tags



Go to file

About ▾

Code ▾







	8af6b7f 9 minutes ago	 32 commits
	.ipyn... notebook chan...	8 hours ago
	Imag... Update links to...	2 hours ago
	data first push	9 days ago
	.DS_... update Images...	9 hours ago
	Final... Update graphs...	2 hours ago
	REA... update readm...	1 hour ago
	gith... Add git.pdf pla...	9 hours ago
	note... add notebook...	9 minutes ago
	note... Add notebook....	9 hours ago
	pres... upload presen...	9 hours ago

 README.md 

# Phase-2Project

## Phase 2 Project - Olamide, Dara, Ilene

-  [Readme](#)
-  0 stars
-  1 watching
-  1 fork




### Releases

No releases published  
[Create a new release](#)

### Packages

No packages published  
[Publish your first package](#)

### Contributors 3

-  **olamide-h**
-  **ileneee** Ilene sorto
-  **DaraNadine** Dara Estrada



Authors: Olamide Olayinka, Ilene  
Sorto, Dara Estrada

## Business Problem

---

### Languages

● Jupyter Notebook 100.0%

The cable television channel HGTV is in the beginning stages of researching a pilot episode for a new renovation series. Each episode will feature a different county in popular regions throughout the United States. The first episode is slated to focus on the Pacific Northwest region, specifically within the Seattle metropolitan area of King County, Oregon. Due to the network's aversion to financial risk our data science team was contracted to provide an objective, outside analysis on the best ways to mitigate risk moving forward. Ultimately, the goal of the show is to tell homeowners what improvements to make on an objective and data driven level, for how to increase the value of their house so the owner can maximize value during sale. HGTV will be investing a percentage of renovations and wants to see the factors that would maximize return on their profits. Since we are interested in figuring out the maximizing the price of these houses, we selected that our target. With this as our target, we utilized several variables and analyzed their impact on predicting the price of homes in the county.

## Data

Our data was pulled from the King County House Data: which contains homes built within the county from 1900 until 2014 in Kings County, Seattle. From there we noticed houses with multiple entries in the data set due to renovations. We decided to drop these houses as to add clarity to the dataset. As we want to ensure for our stakeholder, that we are using relevant information in our modeling process.

Futhermore, we noticed that a large amount of houses had no entries on whether or not they were on the waterfront. To ensure of data was more accurate, we had the missing entries filled with 0 value. We deemed the variables with the most importance to our stakeholder to be the following : bedrooms, bathrooms, Sqft(Square Feet) -Living , Sqft Lot, Year built, grade, floors. With grade being a categorical value ranging from Poor to Mansion level housing, we decided to have this one hot encode to fit our modeling. This allowed us to utilize all the variables relevant to our stakeholders ultimate request.

## Methods

We utilized Scikit Learn.

## Models

---

With model 1, we utilized the variables of bedrooms, bathrooms, floors, grade, square feet living, square feet lot, and year built. With this we got an R-squared of .63. In addition, the root mean squared error is 211289. With this being the metric of error in our model, we deemed this as not a huge error dealing with real estate. However we want to explore other models. This model already outperforms the base.

On model 2, we explored bathrooms and bedrooms against target price. This brought our r-squared down to 2.79. Compared to our previous model, this underperformed. So we decided to move on to another model.

With our third model we got an r-squared of .50. This model accounted for squarefoot living, square foot living-15 and squarefoot lot. An improvement from our previous model but not as good as our first model.

## Conclusion/Recommendations

Ultimately we found that our model 1 had the best results for our stakeholder. With this model we factored in bedrooms, bathrooms, sqft- living, sqft-lot, year built, grade and floors to see its impact on the prices of houses within Kings county. With these findings, we advise hgtv to first renovate lower graded homes. We believe that with investing in these homes, there is a great opportunity for growth and maximizing investment. Furthermore, we advise investing in quality amenities and building adequate living space within the homes. As this maximizes sales price. And lastly, we advise utilizing quality materials and contractors to ensure our homes are quality as this has an impact on the grade of the house and the grade has been proven to have an impact on price.

## Navigation

|-----Data Folder - a directory of all data used in project

<https://github.com/olamide-h/Phase-2Project/tree/master/data>

|-----Images Folder - contains images used in presentation

<https://github.com/olamide-h/Phase-2Project/tree/master/Images>

|-----Presentation.pdf - pdf of  
Google Slide presentation  
[https://github.com/olamide-h/Phase-2Project/blob/master/presentation.p  
df](https://github.com/olamide-h/Phase-2Project/blob/master/presentation.pdf)

|-----Github.pdf - pdf of Github  
submission  
<https://github.com/olamide-h/Phase-2Project/blob/master/github.pdf>

|-----Notebook.pdf - Final jupyter  
notebook  
<https://github.com/olamide-h/Phase-2Project/blob/master/notebook.pdf>

|-----README.md -