INTRO
Hi everyone, I'm Nate and I'm a data scientist. I believe I have a wide audience here today. After talking to a few of you, I know some of you are looking for more info before you purchase a house, or are a builder trying to flip houses and want to maximize profits.

PROBLEM STATEMENT
So today we will be diving into the question:
What features increase the value of a house the most?

INFO ABOUT DATA
First, I want to give you a quick summary of the data I used for my analysis.
The data contains houses sold between 2006 and 2010.
The whole set of data is for properties in Ames, Iowa.
Each entry had 81 variables that I looked into. These varied greatly from square footage of a basement, if the house had an alley, the quality of the kitchen, and many more features.

OUTLINE
Next, I want to break down the steps I'll be going through.
Together, we will explore the data.
I will point out areas of significance.
I will explain a model I used to predict the sale price of a house based on certain features.
Last, I will offer my recommendations.

So, let's dig in!

DISTRIBUTION
To start off I want to give you all an idea of the sale price of homes in this dataset.
This is a normal distribution with just a small percentage of houses selling for more than $400,000
The median, or middle value is represented with the vertical line. This value is around $160,000

What I think is the most important thing to consider when purchasing property is:

Location

NEIGHBORHOODS
Let's take a closer look at the neighborhoods. On the left is the number of houses sold in each neighborhood. On the right is the average sale price of houses in each neighborhood. The vertical line on the right graph is once again the median, or middle sale price.

The neighborhoods are all over the place in that some are well above the average price and some are well below.

The neighborhood at the top, North Ames, had the most properties sold during this time span, and the average price was just below average.

Northridge Heights had a decent number of properties sold with a very high average sale price. Maybe a good place to buy something cheap and fix it up.

You can use this information in different ways, depending on your needs. Someone looking to settle down in a nice neighborhood might be looking in a different place compared to someone trying to fix up a property and make some money off the sale.

Now that you have a better idea of the neighborhoods in Ames, let's get into some details about houses and what features have an impact on their value.

ABOVE GRADE LIVING AREA vs SALE PRICE
Here we have a scatter plot comparing the above grade living area and sale price. Above grade basically means everything that's not the basement. What we are seeing here is a positive linear relationship between these two variables. As the area increases, so does the price. Nothing too surprising here. Bigger house = more expensive

TOTAL BASEMENT SF vs SALE PRICE
Next we have the square footage of the basement and the sale price. Once again, there is a positive linear relationship. An interesting thing to note is the grouping of houses at 0. Not having a basement at all puts a cap on the potential value of a house. There are no houses without a basement that sold for over 300k. You would have to have a basement to exceed that mark.

GARAGE AREA vs SALE PRICE
This graph is very similar to the last. This time we are looking at the relationship between garage area and sale price. Like with the basement, there seems to be a price cap if you do not have a garage.

Let's move away from square footage and look at some other features.

FIREPLACES
Here we are looking at how the number of fireplaces in a house affects the sale price. Whether you have 1 fireplace or 4, the value doesnt change much. The biggest jump comes from not having a fireplace, to having one.

FULL BATHROOMS
This next graph shows there are a few unfortunate people who dont have a full bathroom. Most of the houses in this dataset had 1 or 2 full bathrooms. There is a sizable jump going from 1 bath, to 2.

These last graphs will have to do with the quality of certain features.

OVERALL QUALITY
First, we will look at the overall quality of a home. No real surprises here. The better quality homes generally sell for more.

EXTERIOR QUALITY
Next, let's look how the exterior quality affects sale price. There are some big jumps in value as we go up in quality. A quality 5 exterior is more than triple the value compared to a quality 2. This makes a lot of sense because of the importance of curbside appeal.

KITCHEN QUALITY
Everybody likes nice things. Having a nice quality kitchen also has a huge impact on sale price. Just like the exterior, the quality 5's average value is more than triple than that of a quality 2.

BASEMENT QUALITY
The last quality metric we will look at is the basement quality. Earlier we saw that having a basement was beneficial to the value of a house. Now we can see that having a high quality basement is also very valuable. The high value of basements in Iowa may be because of its geographic location. Iowa is located in the tornado alley, where there is a higher frequency and magnitude of tornadoes. So, if you need to find shelter, why not have a comfortable living space?

With all this exploring of data, I was able to piece together some features that I thought would increase home value the most. I was able to test these features with a model to predict the sale price. I'll spare you the details on how it works and instead show you how it did.

BASELINE
First, we need something to compare to. A baseline model is what I'll be using. Simply put, this model guessed the average sale price every time.
The scoring metric I'll be using to compare is called the Root Mean Squared Error or RMSE for short. To put it simply again, the lower the number, the better. The baseline RMSE I will be comparing my model to is 76,414.

RESULTS
Each model I tested I ran 50 times and then took the average of those scores. The two regression models I used were Lasso and Ridge. Lasso came in at just under 25,000, and Ridge came in around 24,200. Remember, the lower the score the better. Both models did significantly better than the baseline in predicting the sale price of a house based on the features I fed the model. The Ridge Regression model slightly outperformed Lasso for me.

So, I spent a lot of time manipulating the models. Fine tuning parameters and deciding which features I wanted the model to train on.

FEATURES
I've chosen 7 features that I believe you all should focus on if you want to increase the value of your home the most.

- Above grade living area and overall quality are pretty straight forward. Big, nice, houses are worth more.
- By not having a garage, your ceiling for house value is limited.
- The same goes for the basement. In addition to having a basement, having one that is of higher quality will give a significant boost in value.
- Exterior quality is important for the curbside appeal.
- A high quality kitchen goes a long way.
- Last, I want to stress the importance of location again. You can change almost everything you want about a house, so choose a location that fits you best.

Thank you all for your time! I'd be happy to answer any questions.