Project #2 Talking Points

* Intro - 1
  + Hello everyone! So the name of our project is Real Estate Market Machine Learning & Sentiment Analysis.
  + With the housing market going crazy right now we wanted to take a closer look at current trends and see if there are any similarities to what happened back in 2008 with housing prices at the time of the recession with today.
* Goals - 2
  + We set out with 3 main goals in mind for our analysis.
    - 1) We wanted to find good historical data that gave us enough to work with for the current market and the market pre-2008.
    - 2) We wanted to take a look at public sentiment through media and databases.
    - 3) We wanted to use prediction models to show if the market is trending toward a correction.
* Hypothesis - 2
  + As a group we discussed our opinions on the topic from personal experience and what we have heard in the media.
  + The hypothesis we came up with is:
    - If the current housing market is at unsustainable price levels, then a market correction is inevitable.
  + We feel that prices increasing dramatically and there is no way they can keep going up for an extended period of time, especially with the US moratorium end looming over the country.
* Data Prep - 3/4/5/6
  + We decided that Quandl’s Zillow API data had the best data for what we were trying to accomplish in our models.
  + For sentiment we used Fannie Mae, News API and the New York Times.
  + An interesting challenge we ran into was the size of Data sets we were initially finding.
  + Our first session in class we had another data set picked out from a national database called Freddie Mac which was great, it had a ton of different indicators like square footage and all sorts of stuff but the amount was too massive.
    - My computer physically could not run the train test split due to size so we went with just Zillow API.
    - Even Zillow gave us issues at times with data size. One of our data frames had over 127,000,000 rows of data. We tried running it on Google Colab to see if their service could handle the workload but we ultimately had to trim down our data set.
* I am now going to pass it along to Brian.