Slide 1   
  
In attempting to out scout the scouts, several sentiment analysis models were applied to the NBA draft data to determine if **negative vs positive reviews provided by analysts**, could be used as predictors of player success.

The bar chart shows that for our sample data, the majority of “bust” players have highly positive sentiment scores as represented by the green color of the last bar.

Bust players have the highest distribution among negative scores, however it is clearly immaterial when compared to the entire dataset, **suggesting severe skewness towards positive** scores in the data.

Next slide - Slide 2

Given the skewness of the data, and poor results obtained from runs of various models and feature combinations, **specifically** when trying to predict positive, negative, and neutral sentiment scores,

WE arrived at a final training data set which focused on only two classes, positive and negative. This in turn produced a model with a great overall accuracy of 92.5%. We also saw that the SVM model with a linear kernel produced consistent results across both classes and cross-sections.

Next Slide - Slide 3

As far as using Sentiment analysis as a predictor of Draft Player success in the NBA:

We can say that based on our sample data and model results, there is potential for this application. The distribution charts of the top 100 positive and top 100 negative scores confirm that there is a higher incidence of bust players among the negative sentiment scores than among the positive ones. That said, significant improvements should be made to the data collection for better model training.

Next Slide – Slide 4

To extract more insights from the NBA Draft Data we applied the Latent Dirichlet Allocation model to identify main themes within the Analysts reviews.

Using a grid search method based on Coherence scores, which measures statistical dependencies between words in each topic, 25 main topics were discovered. – The plot shows the moment when the line drops, indicating the instant when the topics became less coherent, this signals that adding more topics would not improve the quality of the topics – and here we can see the resulting main themes, focusing on players game skills, physical strengths and ball handling abilities among others.

Next slide please

Additional insights were captured using the pyLDAvis visualization: just to highlight a few points

Basically, each bubble is a topic and the distance between bubbles indicates the degree of similarity between topics. – for our data, we can see the topics a quite similar

The frequency and relevance of words within each of the 25 topics is represented by the **bar chart and size of the bubbles.** The larger the bubble or bar the higher the frequency and the more significant the words are within the collection.

Again, the main goal here was to identify **common themes among the reviews**. By harnessing these topics and their prevalence, analysts can build predictive models that consider **their influence** on **players' future success or struggles in the NBA.**

And now Matt Smith will take over