Using our regression models, we would like to predict before the season starts how many games the 2002 Oakland A's will win.

To do this, we first have to predict how many runs the team will score and how many runs they will allow.

These models use team statistics.

However, when we are predicting for the 2002 Oakland A's before the season has occurred, the team is probably different than it was the year before.

So we don't know the team statistics.

But we can estimate these statistics using past player performance.

This approach assumes that past performance correlates with future performance and that there will be few injuries during the season.

Using this approach, we can estimate the team statistics for 2002 by using the 2001 player statistics.

Let's start by making a prediction for runs scored.

At the beginning of the 2002 season, the Oakland A's had 24 batters on their roster.

Using the 2001 regular season statistics for these players, we can estimate that team on-base percentage will be about 0.339 and team slugging percentage will be about 0.430.

We built the following linear regression equation in R to predict runs scored.

If we plug in 0.339 for on-base percentage and 0.430 for slugging percentage, we predict that the 2002 Oakland A's will score about 805 runs.

Similarly, we can make a prediction for runs allowed.

At the beginning of the 2002 season, the Oakland A's had 17 pitchers on their roster.

Using the 2001 regular season statistics for these players, we can estimate that team opponent on-base percentage will be about 0.307 and team opponent slugging percentage will be about 0.373.

Our regression equation to predict runs allowed was as follows.

By plugging in 0.307 for opponents on-base percentage and 0.373 for opponents slugging percentage, we predict that the 2002 Oakland A's will allow 622 runs.

We can now make a prediction for how many games they will win.

Our regression equation to predict wins is as follows.

We predicted 805 runs scored and 622 runs allowed.

We can plug in the difference between runs scored and runs allowed to predict that the A's will win 100 games in 2002.

Paul DePodesta used a similar approach to make predictions.

It turns out that our predictions and Paul's predictions closely match actual performance.

Our prediction for runs scored was 805 runs.

Paul predicted between 800 and 820 runs.

And it turns out that the 2002 Oakland A's actually scored 800 runs.

For runs allowed, we predicted 622.

Paul DePodesta predicted between 650 and 670.

It turns out that the Oakland A's actually allowed 653 runs.

For wins, we predicted that they would win 100 games.

Paul predicted that they would win between 93 and 97 games.

And they actually 103 games.

These predictions show us that by using publicly available data and simple analytics, we can predict very close to what actually happened before the season even started.

It turns out that the A's won a league record in 2002 by winning 20 games in a row.

And they won 1 more game than the previous year and made it to the playoffs once again.