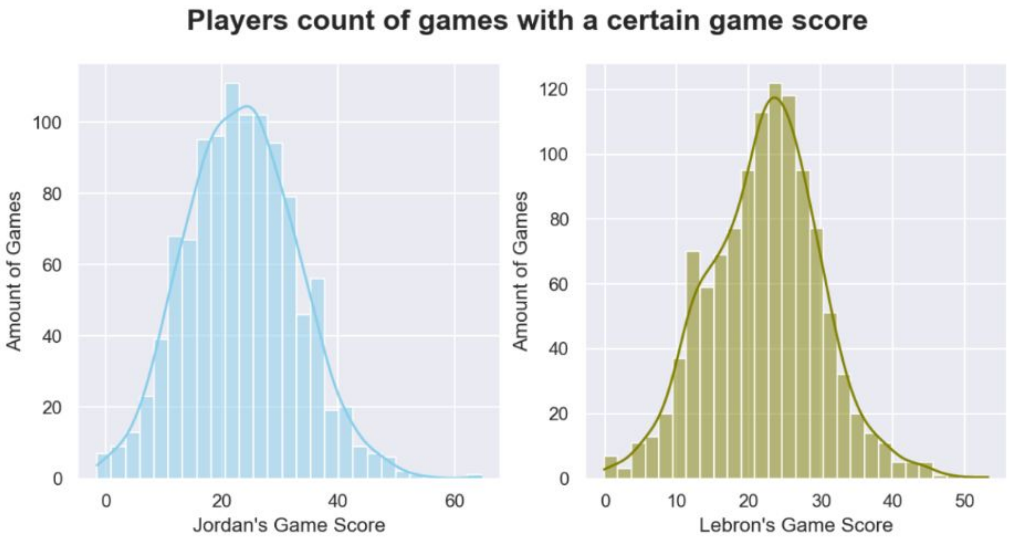
Lebron vs Jordan

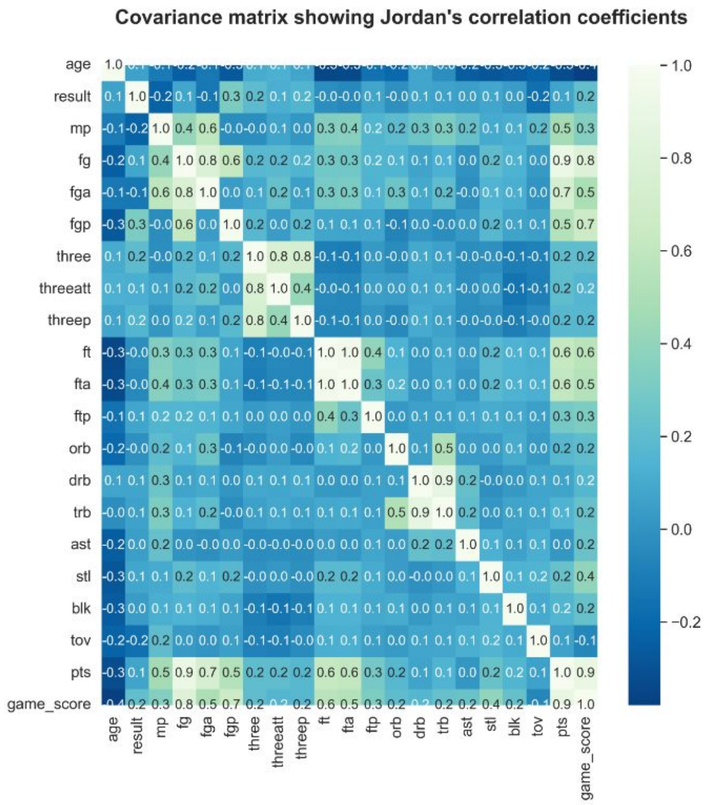
Marc Martín Ortega (1564274)
Marina Vázquez Guerrero (1563735)
Mireia Fernández Fernández (1562636)

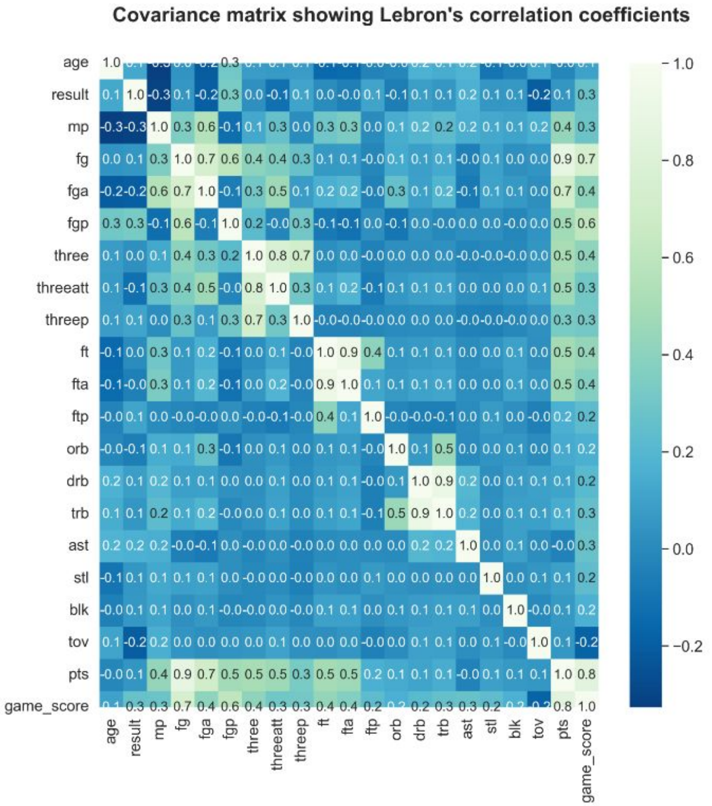
Data visualization

- Michael Jordan showcases a larger range with a more defined Gaussian distribution.
- Jordan got a 64.9 in his best game versus the 53.2 of LeBron’s.
- LeBron who’s counts go over 100 matches with a similar game score and has a slimmer curve.



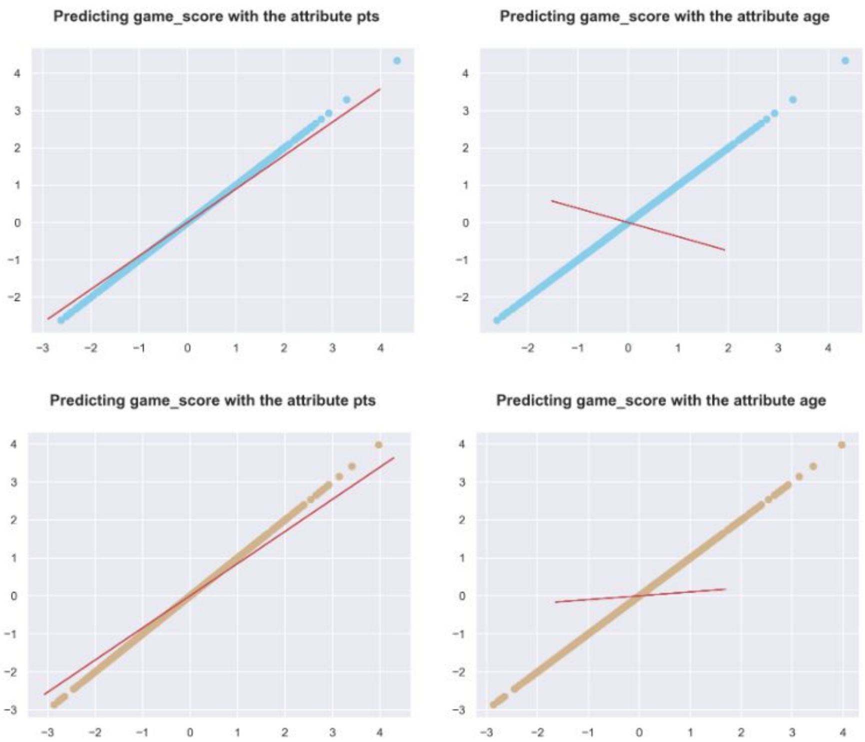
Covariance matrices





The data exhibits that the most important attributes in order to achieve a good prediction, which are the same for both players, will be fg, fgp, ft, fta and pts.

Linear Regressions



- Graphics of Jordan’s game score best and worst linear predictions using just one attribute.

- Graphics of LeBron’s game score best and worst linear predictions using just one attribute

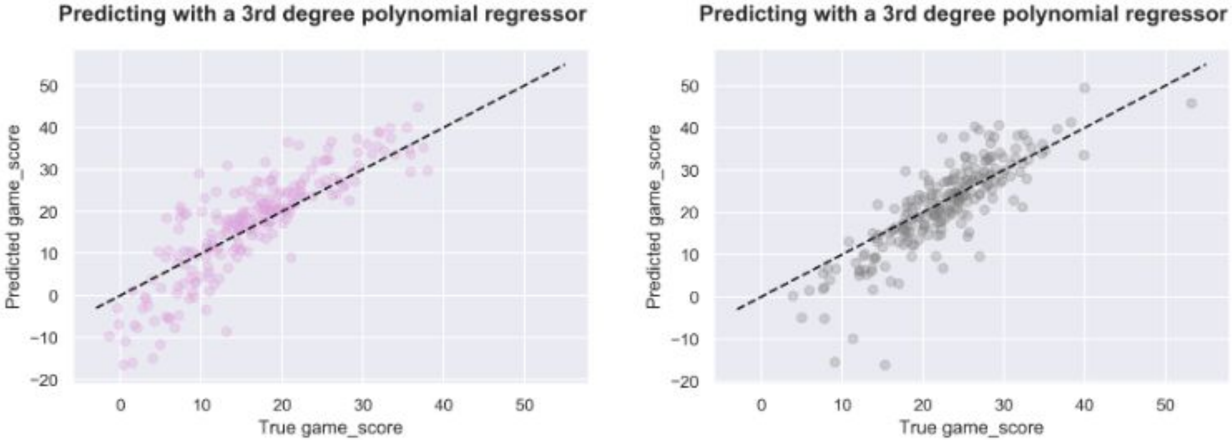
Table of comparison between multivariate regressors

- R^2 score

		Training	Cross Validation	Test
Non Standardized Data	L. James	0.9962	0.9949	0.9954
	M. Jordan	0.9968	0.9951	0.9963
Standardized Data	L. James	0.9962	0.9949	0.9954
	M. Jordan	0.9968	0.9951	0.9963

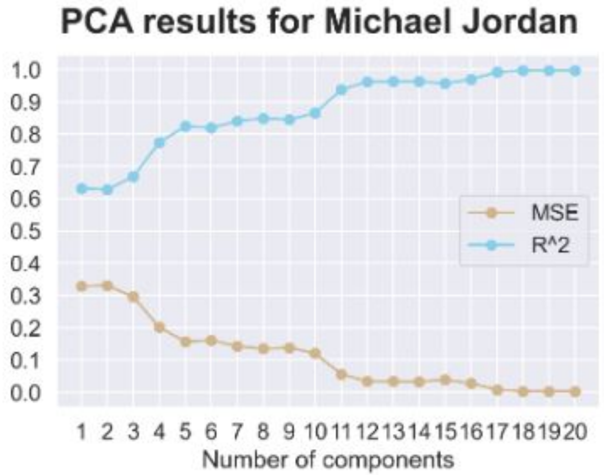
The normalization of the datasets does not influence in the predictions made.

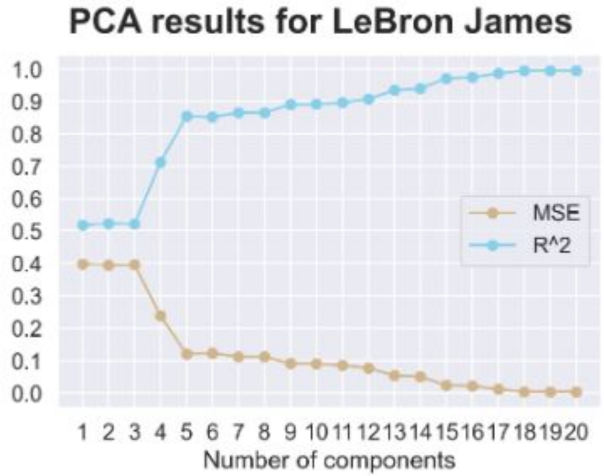
Third degree polynomial fit



The predictions for the cross validation samples returned R2 scores of 0.229665 for LeBron James and 0.365974 for Michael Jordan.

Principal Component Analysis

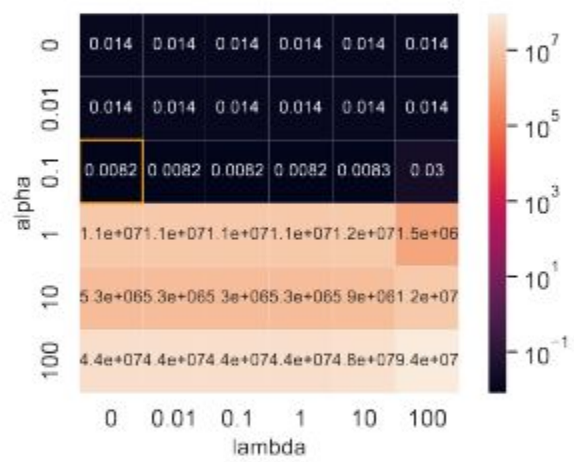




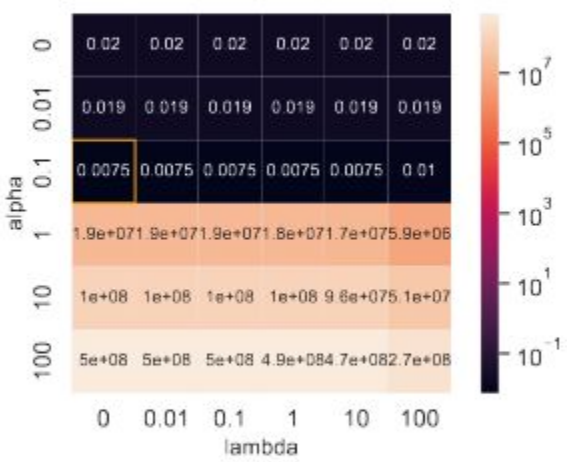
- Jordan’s PCA: it is observable that at four, eleven and seventeen as dimensions, the scores change more notably and then keep similar values.
- James’ PCA: ts R2 increment is much more uniform; with the only big growths in the dimension reductions four and five, and then subtler growths in the dimension reductions nine, thirteen and fifteen

Gradient Descent

MSEs for both players given a set of hyperparameters



Results for the Jordan’s dataset



Results for the LeBron’s dataset

Conclusions

- The models implemented for Michael Jordan would always return better fits, in terms of the R2 and the MSE scores, from a hundredth of the result to increasing LeBron’s score.
- Almost all of the implemented regressors achieved a high R2 score and resulted in decent models; from the linear regressors using only one attribute to the polynomial fits, in exception of the third degree.
- The PCA procedure and then the Gradient descent where consistent but in this case not are very useful.
- Half of the coefficients of the multivariate regressor, returned similar values. This and the outstanding results for the linear regressors gave us the intuition that there was some sort of formula to obtain a player’s game score.