MSA 8010 - Data Programming

Predicting Major League Baseball World Series Winner

# Introduction

* Matan
* What baseball is
* How the playoffs work
* Analytics in baseball
* Project goal – predict how far each team will go

# Data

* Michael
* Data gathering
* Graphs
* Description of statistics
* Data cleaning
* Python packages

# Variable Selection

* Deepti
* Recursive Feature elimination
* ExtraTrees Classifier
* Random Forest Regressor
* Compare output of each method

# Data Models

* Rohan
* KNN
* Logistic regression

# Conclusions

* Michael
* Model accuracy
* Shortcomings
* Future work