

# Homework 1

Dhairav Chhatbar, Mael Illien, Salma Elshahawy

08/31/2020

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Statement of the Problem</b>	<b>2</b>
<b>3</b>	<b>Data Exploration</b>	<b>3</b>
3.1	Imputing Missing Values . . . . .	4
3.2	Correlation Matrix . . . . .	5
<b>4</b>	<b>Data Preparation</b>	<b>6</b>
4.1	Outliers . . . . .	6
4.2	Box Cox Transformation . . . . .	6
<b>5</b>	<b>Models comparasion</b>	<b>7</b>
<b>6</b>	<b>Selected Model</b>	<b>7</b>
<b>7</b>	<b>Prediction on Evaluation Data</b>	<b>7</b>
<b>8</b>	<b>Appendix A</b>	<b>8</b>
8.1	R source code . . . . .	8

Prepared for:

Dr. Nasrin Khansari

City University of New York, School of Professional Studies - Data 621

Prepared by:

Dhairav Chhatbar

Mael Illien

Salma Elshahawy

# 1 Introduction

The ability to analyze and predict performance of a professional baseball team using many dimensions is critical to competitive success for our organization. Therefore, we have analyzed the records of numerous professional baseball team from the years 1871 to 2006. Our hope is that the following report and the resulting predictive models will better inform the organization and assist in making data driven decisions moving forward.

“The goal of a baseball team is to win more games than any other team. Since one team has very little control over the number of games other teams win, the goal is essentially to win as many games as possible. Therefore, it is of interest to measure the player’s contribution to the team’s wins.” Grabiner, B. D. <sup>1</sup> While we do not have the variables at the player’s individual contribution level, we do have the entire teams contributions as an aggregate and will analyze that information.

## 2 Statement of the Problem

The purpose of this report is to determine the batting, baserun, pitching, and fielding effects on a baseball team’s ability to win.

---

<sup>1</sup>(Grabiner, B. D. (n.d.). The Sabermetric Manifesto. Retrieved September 10, 2016 from <http://seanlahman.com/baseball-archive/sabermetrics/sabermetric-manifesto/>)

### **3 Data Exploration**

## 3.1 Imputing Missing Values

## 3.2 Correlation Matrix

## **4 Data Preparation**

### **4.1 Outliers**

### **4.2 Box Cox Transformation**

**5 Models comparasion**

**6 Selected Model**

**7 Prediction on Evaluation Data**

## **8 Appendix A**

### **8.1 R source code**