Finding Replacement Players for Oakland A's

Mohsin Asif

May 18, 2018

Introduction

In this project I will be providing a solution to the actual problem a baseball team Oakland A's faced in 2001 when three of their key players left the team. To find the replacement players,the general manager of Oakland A's took sports analytics to the next level by inventing some new KPIs for finding undervalued but brilliant players . This strategy was so successful that they went on the win 20 consecutive games. This strategy was brough to the world in a book called 'Moneyball: The Art of Winning an Unfair Game' by Michel Lewis. Later, this book was turned into a movie called Moneyball.

In this project we will analyze the KPIs that better account for the offence's success such as Slugging and On Base Percentage. While finding replacement players we have following three constraints:

1. Combined salaries should not exceed 15 million USD
2. Combined At Bat(AB) should be more than the combined AB of lost players
3. Mean On Base Percentage(OBP) should be more than the mean OBP of lost players

Loading the datasets:

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
#loading batting table
batting<-read.csv("C:/Users/Mohsin Asif/Box Sync/MS IS/Moneyball
Project/batting.csv")

#loading salary table
salary<-read.csv("C:/Users/Mohsin Asif/Box Sync/MS IS/Moneyball
Project/salaries.csv")
```

Analyzing the structure of datasets:

```
#analyzing the structure and summarizing batting table
str(batting)
```

```
## 'data.frame':    97889 obs. of  24 variables:
##  $ playerID : Factor w/ 18107 levels "aardsda01","aaronha01",..: 1 1 1 1 1
1 1 2 2 2 ...
##  $ yearID   : int  2004 2006 2007 2008 2009 2010 2012 1954 1955 1956 ...
##  $ stint    : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ teamID   : Factor w/ 149 levels "ALT","ANA","ARI",..: 117 35 33 16 116
116 93 80 80 80 ...
##  $ lgID     : Factor w/ 6 levels "AA","AL","FL",..: 4 4 2 2 2 2 2 4 4 4
...
##  $ G        : int  11 45 25 47 73 53 1 122 153 153 ...
##  $ G_batting: int  11 43 2 5 3 4 NA 122 153 153 ...
##  $ AB       : int  0 2 0 1 0 0 NA 468 602 609 ...
##  $ R        : int  0 0 0 0 0 0 NA 58 105 106 ...
##  $ H        : int  0 0 0 0 0 0 NA 131 189 200 ...
##  $ X2B      : int  0 0 0 0 0 0 NA 27 37 34 ...
##  $ X3B      : int  0 0 0 0 0 0 NA 6 9 14 ...
##  $ HR       : int  0 0 0 0 0 0 NA 13 27 26 ...
##  $ RBI      : int  0 0 0 0 0 0 NA 69 106 92 ...
##  $ SB       : int  0 0 0 0 0 0 NA 2 3 2 ...
##  $ CS       : int  0 0 0 0 0 0 NA 2 1 4 ...
##  $ BB       : int  0 0 0 0 0 0 NA 28 49 37 ...
##  $ SO       : int  0 0 0 1 0 0 NA 39 61 54 ...
##  $ IBB      : int  0 0 0 0 0 0 NA NA 5 6 ...
##  $ HBP      : int  0 0 0 0 0 0 NA 3 3 2 ...
##  $ SH       : int  0 1 0 0 0 0 NA 6 7 5 ...
##  $ SF       : int  0 0 0 0 0 0 NA 4 4 7 ...
##  $ GIDP     : int  0 0 0 0 0 0 NA 13 20 21 ...
##  $ G_old    : int  11 45 2 5 NA NA NA 122 153 153 ...
```

```
summary(batting)
```

```
##       playerID         yearID         stint            teamID
##  mcguide01:   31   Min.   :1871   Min.   :1.000   CHN     : 4720
##  henderi01:   29   1st Qu.:1931   1st Qu.:1.000   PHI     : 4621
##  newsobo01:   29   Median :1970   Median :1.000   PIT     : 4575
##  johnto01 :   28   Mean   :1962   Mean   :1.077   SLN     : 4535
##  kaatji01 :   28   3rd Qu.:1995   3rd Qu.:1.000   CIN     : 4393
##  ansonca01:   27   Max.   :2013   Max.   :5.000   CLE     : 4318
##  (Other)  :97717                                  (Other):70727
##    lgID             G            G_batting            AB
##  AA  : 1890   Min.   :  1.00   Min.   :  0.00   Min.   :  0.0
##  AL  :44369   1st Qu.: 13.00   1st Qu.:  7.00   1st Qu.:  9.0
##  FL  :  470   Median : 35.00   Median : 32.00   Median : 61.0
##  NL  :49944   Mean   : 51.65   Mean   : 49.13   Mean   :154.1
##  PL  :  147   3rd Qu.: 81.00   3rd Qu.: 81.00   3rd Qu.:260.0
##  UA  :  332   Max.   :165.00   Max.   :165.00   Max.   :716.0
```

```
##  NA's:  737                    NA's   :1406     NA's   :6413
##        R                H                X2B             X3B
##  Min.   :  0.00   Min.   :  0.00   Min.   : 0.0   Min.   : 0.000
##  1st Qu.:  0.00   1st Qu.:  1.00   1st Qu.: 0.0   1st Qu.: 0.000
##  Median :  5.00   Median : 12.00   Median : 2.0   Median : 0.000
##  Mean   : 20.47   Mean   : 40.37   Mean   : 6.8   Mean   : 1.424
##  3rd Qu.: 31.00   3rd Qu.: 66.00   3rd Qu.:10.0   3rd Qu.: 2.000
##  Max.   :192.00   Max.   :262.00   Max.   :67.0   Max.   :36.000
##  NA's   :6413     NA's   :6413     NA's   :6413   NA's   :6413
##        HR              RBI               SB                CS
##  Min.   : 0.000   Min.   :  0.00   Min.   :  0.000   Min.   : 0.000
##  1st Qu.: 0.000   1st Qu.:  0.00   1st Qu.:  0.000   1st Qu.: 0.000
##  Median : 0.000   Median :  5.00   Median :  0.000   Median : 0.000
##  Mean   : 3.002   Mean   : 18.47   Mean   :  3.265   Mean   : 1.385
##  3rd Qu.: 3.000   3rd Qu.: 28.00   3rd Qu.:  2.000   3rd Qu.: 1.000
##  Max.   :73.000   Max.   :191.00   Max.   :138.000   Max.   :42.000
##  NA's   :6413     NA's   :6837     NA's   :7713      NA's   :29867
##        BB               SO              IBB              HBP
##  Min.   :  0.00   Min.   :  0.00   Min.   :  0.00   Min.   : 0.000
##  1st Qu.:  0.00   1st Qu.:  2.00   1st Qu.:  0.00   1st Qu.: 0.000
##  Median :  4.00   Median : 11.00   Median :  0.00   Median : 0.000
##  Mean   : 14.21   Mean   : 21.95   Mean   :  1.28   Mean   : 1.136
##  3rd Qu.: 21.00   3rd Qu.: 31.00   3rd Qu.:  1.00   3rd Qu.: 1.000
##  Max.   :232.00   Max.   :223.00   Max.   :120.00   Max.   :51.000
##  NA's   :6413     NA's   :14251    NA's   :42977    NA's   :9233
##        SH               SF             GIDP           G_old
##  Min.   : 0.000   Min.   : 0.0    Min.   : 0.00   Min.   :  0.00
##  1st Qu.: 0.000   1st Qu.: 0.0    1st Qu.: 0.00   1st Qu.: 11.00
##  Median : 1.000   Median : 0.0    Median : 1.00   Median : 34.00
##  Mean   : 2.564   Mean   : 1.2    Mean   : 3.33   Mean   : 50.99
##  3rd Qu.: 3.000   3rd Qu.: 2.0    3rd Qu.: 5.00   3rd Qu.: 82.00
##  Max.   :67.000   Max.   :19.0    Max.   :36.00   Max.   :165.00
##  NA's   :12751    NA's   :42446   NA's   :32521   NA's   :5189
```

```
#analyzing the structure and summarizing batting table
str(salary)
```

```
## 'data.frame':    23956 obs. of  5 variables:
##  $ yearID  : int  1985 1985 1985 1985 1985 1985 1985 1985 1985 1985 ...
##  $ teamID  : Factor w/ 35 levels "ANA","ARI","ATL",..: 4 4 4 4 4 4 4 4 4 4
...
##  $ lgID    : Factor w/ 2 levels "AL","NL": 1 1 1 1 1 1 1 1 1 1 ...
##  $ playerID: Factor w/ 4668 levels "aardsda01","aasedo01",..: 2961 2517
3566 2304 1332 394 4061 2604 3634 2722 ...
##  $ salary  : int  1472819 1090000 800000 725000 641667 625000 581250
560000 558333 547143 ...
```

```
summary(batting)
```

```
##       playerID         yearID         stint           teamID
##  mcguide01:   31   Min.   :1871   Min.   :1.000   CHN    : 4720
```

```
##  henderi01:   29   1st Qu.:1931   1st Qu.:1.000   PHI    : 4621
##  newsobo01:   29   Median :1970   Median :1.000   PIT    : 4575
##  johnto01 :   28   Mean   :1962   Mean   :1.077   SLN    : 4535
##  kaatji01 :   28   3rd Qu.:1995   3rd Qu.:1.000   CIN    : 4393
##  ansonca01:   27   Max.   :2013   Max.   :5.000   CLE    : 4318
##  (Other)  :97717                                  (Other):70727
##    lgID             G            G_batting            AB
##  AA  : 1890   Min.   :  1.00   Min.   :  0.00   Min.   :  0.0
##  AL  :44369   1st Qu.: 13.00   1st Qu.:  7.00   1st Qu.:  9.0
##  FL  :  470   Median : 35.00   Median : 32.00   Median : 61.0
##  NL  :49944   Mean   : 51.65   Mean   : 49.13   Mean   :154.1
##  PL  :  147   3rd Qu.: 81.00   3rd Qu.: 81.00   3rd Qu.:260.0
##  UA  :  332   Max.   :165.00   Max.   :165.00   Max.   :716.0
##  NA's:  737                    NA's   :1406     NA's   :6413
##         R                H               X2B             X3B
##  Min.   :  0.00   Min.   :  0.00   Min.   : 0.0   Min.   : 0.000
##  1st Qu.:  0.00   1st Qu.:  1.00   1st Qu.: 0.0   1st Qu.: 0.000
##  Median :  5.00   Median : 12.00   Median : 2.0   Median : 0.000
##  Mean   : 20.47   Mean   : 40.37   Mean   : 6.8   Mean   : 1.424
##  3rd Qu.: 31.00   3rd Qu.: 66.00   3rd Qu.:10.0   3rd Qu.: 2.000
##  Max.   :192.00   Max.   :262.00   Max.   :67.0   Max.   :36.000
##  NA's   :6413     NA's   :6413     NA's   :6413   NA's   :6413
##        HR              RBI               SB                CS
##  Min.   : 0.000   Min.   :  0.00   Min.   :  0.000   Min.   : 0.000
##  1st Qu.: 0.000   1st Qu.:  0.00   1st Qu.:  0.000   1st Qu.: 0.000
##  Median : 0.000   Median :  5.00   Median :  0.000   Median : 0.000
##  Mean   : 3.002   Mean   : 18.47   Mean   :  3.265   Mean   : 1.385
##  3rd Qu.: 3.000   3rd Qu.: 28.00   3rd Qu.:  2.000   3rd Qu.: 1.000
##  Max.   :73.000   Max.   :191.00   Max.   :138.000   Max.   :42.000
##  NA's   :6413     NA's   :6837     NA's   :7713      NA's   :29867
##        BB               SO              IBB              HBP
##  Min.   :  0.00   Min.   :  0.00   Min.   :  0.00   Min.   : 0.000
##  1st Qu.:  0.00   1st Qu.:  2.00   1st Qu.:  0.00   1st Qu.: 0.000
##  Median :  4.00   Median : 11.00   Median :  0.00   Median : 0.000
##  Mean   : 14.21   Mean   : 21.95   Mean   :  1.28   Mean   : 1.136
##  3rd Qu.: 21.00   3rd Qu.: 31.00   3rd Qu.:  1.00   3rd Qu.: 1.000
##  Max.   :232.00   Max.   :223.00   Max.   :120.00   Max.   :51.000
##  NA's   :6413     NA's   :14251    NA's   :42977    NA's   :9233
##        SH               SF              GIDP            G_old
##  Min.   : 0.000   Min.   : 0.0    Min.   : 0.00   Min.   :  0.00
##  1st Qu.: 0.000   1st Qu.: 0.0    1st Qu.: 0.00   1st Qu.: 11.00
##  Median : 1.000   Median : 0.0    Median : 1.00   Median : 34.00
##  Mean   : 2.564   Mean   : 1.2    Mean   : 3.33   Mean   : 50.99
##  3rd Qu.: 3.000   3rd Qu.: 2.0    3rd Qu.: 5.00   3rd Qu.: 82.00
##  Max.   :67.000   Max.   :19.0    Max.   :36.00   Max.   :165.00
##  NA's   :12751    NA's   :42446   NA's   :32521   NA's   :5189
```

Adding new statistical measures to the dataset:

```
#Adding a new variable BA for calculating Batting Average:
batting$BA<-batting$H/batting$AB
```

```r
#Adding a new variable OBP for calculating On Base Percentage:
batting$OBP<-
(batting$H+batting$BB+batting$HBP)/(batting$AB+batting$BB+batting$HBP+batting
$SF)

#Adding a new variable SLG for calculating Slugging:
#Creating a new variable 1xB to calculate Singles as it is not already in the
batting table
batting$x1B<-batting$H-(batting$X2B+batting$X3B+batting$HR)

#Creating new column for slugging using following formula
batting$SLG<-
(batting$x1B+(2*batting$X2B)+(3*batting$X3B)+(4*batting$HR))/batting$AB

#Quick look at the final batting table:
head(batting, 25)
```

```
##     playerID yearID stint teamID lgID   G G_batting  AB   R   H X2B X3B HR
## 1  aardsda01   2004     1    SFN   NL  11        11   0   0   0   0   0  0
## 2  aardsda01   2006     1    CHN   NL  45        43   2   0   0   0   0  0
## 3  aardsda01   2007     1    CHA   AL  25         2   0   0   0   0   0  0
## 4  aardsda01   2008     1    BOS   AL  47         5   1   0   0   0   0  0
## 5  aardsda01   2009     1    SEA   AL  73         3   0   0   0   0   0  0
## 6  aardsda01   2010     1    SEA   AL  53         4   0   0   0   0   0  0
## 7  aardsda01   2012     1    NYA   AL   1        NA  NA  NA  NA  NA  NA NA
## 8  aaronha01   1954     1    ML1   NL 122       122 468  58 131  27   6 13
## 9  aaronha01   1955     1    ML1   NL 153       153 602 105 189  37   9 27
## 10 aaronha01   1956     1    ML1   NL 153       153 609 106 200  34  14 26
## 11 aaronha01   1957     1    ML1   NL 151       151 615 118 198  27   6 44
## 12 aaronha01   1958     1    ML1   NL 153       153 601 109 196  34   4 30
## 13 aaronha01   1959     1    ML1   NL 154       154 629 116 223  46   7 39
## 14 aaronha01   1960     1    ML1   NL 153       153 590 102 172  20  11 40
## 15 aaronha01   1961     1    ML1   NL 155       155 603 115 197  39  10 34
## 16 aaronha01   1962     1    ML1   NL 156       156 592 127 191  28   6 45
## 17 aaronha01   1963     1    ML1   NL 161       161 631 121 201  29   4 44
## 18 aaronha01   1964     1    ML1   NL 145       145 570 103 187  30   2 24
## 19 aaronha01   1965     1    ML1   NL 150       150 570 109 181  40   1 32
## 20 aaronha01   1966     1    ATL   NL 158       158 603 117 168  23   1 44
## 21 aaronha01   1967     1    ATL   NL 155       155 600 113 184  37   3 39
## 22 aaronha01   1968     1    ATL   NL 160       160 606  84 174  33   4 29
## 23 aaronha01   1969     1    ATL   NL 147       147 547 100 164  30   3 44
## 24 aaronha01   1970     1    ATL   NL 150       150 516 103 154  26   1 38
## 25 aaronha01   1971     1    ATL   NL 139       139 495  95 162  22   3 47
##    RBI SB CS BB SO IBB HBP SH SF GIDP G_old        BA       OBP x1B
## 1    0  0  0  0  0   0   0  0  0    0    11       NaN       NaN   0
## 2    0  0  0  0  0   0   0  1  0    0    45 0.0000000 0.0000000   0
## 3    0  0  0  0  0   0   0  0  0    0     2       NaN       NaN   0
## 4    0  0  0  0  1   0   0  0  0    0     5 0.0000000 0.0000000   0
## 5    0  0  0  0  0   0   0  0  0    0    NA       NaN       NaN   0
```

```
## 6    0  0  0  0  0   0   0  0  0    0    NA       NaN       NaN   0
## 7   NA NA NA NA NA  NA  NA NA NA   NA    NA        NA        NA  NA
## 8   69  2  2 28 39  NA   3  6  4   13   122 0.2799145 0.3220676  85
## 9  106  3  1 49 61   5   3  7  4   20   153 0.3139535 0.3662614 116
## 10  92  2  4 37 54   6   2  5  7   21   153 0.3284072 0.3648855 126
## 11 132  1  1 57 58  15   0  0  3   13   151 0.3219512 0.3777778 121
## 12  95  4  1 59 49  16   1  0  3   21   153 0.3261231 0.3855422 128
## 13 123  8  0 51 54  17   4  0  9   19   154 0.3545310 0.4011544 131
## 14 126 16  7 60 63  13   2  0 12    8   153 0.2915254 0.3524096 101
## 15 120 21  9 56 64  20   2  1  9   16   155 0.3266998 0.3805970 114
## 16 128 15  7 66 73  14   3  0  6   14   156 0.3226351 0.3898051 112
## 17 130 31  5 78 94  18   0  0  5   11   161 0.3185420 0.3907563 124
## 18  95 22  4 62 46   9   0  0  2   22   145 0.3280702 0.3927445 131
## 19  89 24  4 60 81  10   1  0  8   15   150 0.3175439 0.3787167 108
## 20 127 21  3 76 96  15   1  0  8   14   158 0.2786070 0.3561047 100
## 21 109 17  6 63 97  19   0  0  6   11   155 0.3066667 0.3692078 105
## 22  86 28  5 64 62  23   1  0  5   21   160 0.2871287 0.3535503 108
## 23  97  9 10 87 47  19   2  0  3   14   147 0.2998172 0.3959311  87
## 24 118  9  0 74 63  15   2  0  6   13   150 0.2984496 0.3846154  89
## 25 118  1  1 71 58  21   2  0  5    9   139 0.3272727 0.4101222  90
##          SLG
## 1        NaN
## 2  0.0000000
## 3        NaN
## 4  0.0000000
## 5        NaN
## 6        NaN
## 7         NA
## 8  0.4465812
## 9  0.5398671
## 10 0.5582923
## 11 0.6000000
## 12 0.5457571
## 13 0.6359300
## 14 0.5661017
## 15 0.5936982
## 16 0.6182432
## 17 0.5863708
## 18 0.5140351
## 19 0.5596491
## 20 0.5389718
## 21 0.5733333
## 22 0.4983498
## 23 0.6069470
## 24 0.5736434
## 25 0.6686869
```

Cleaning and merging datasets:

```
#Since our Salary data starts from year 1985, we will only select data after
1984 from batting table
```

```r
batting<-subset(batting, batting$yearID>1984,)

#Checking if we applied subset function correctly and only have data after
1984
summary(batting$yearID)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1985    1993    2000    2000    2007    2013
```

```r
#Merging batting and salary data for further analysis on selecting
replacement players
batsal<-merge(batting, salary, by=c('playerID','yearID'))
head(batsal)
```

```
##    playerID yearID stint teamID.x lgID.x  G G_batting AB  R  H X2B X3B HR
## 1 aardsda01   2004     1      SFN     NL 11        11  0  0  0   0   0  0
## 2 aardsda01   2007     1      CHA     AL 25         2  0  0  0   0   0  0
## 3 aardsda01   2008     1      BOS     AL 47         5  1  0  0   0   0  0
## 4 aardsda01   2009     1      SEA     AL 73         3  0  0  0   0   0  0
## 5 aardsda01   2010     1      SEA     AL 53         4  0  0  0   0   0  0
## 6 aardsda01   2012     1      NYA     AL  1        NA NA NA NA  NA  NA NA
##   RBI SB CS BB SO IBB HBP SH SF GIDP G_old  BA OBP x1B SLG teamID.y lgID.y
## 1   0  0  0  0  0   0   0  0  0    0    11 NaN NaN   0 NaN      SFN     NL
## 2   0  0  0  0  0   0   0  0  0    0     2 NaN NaN   0 NaN      CHA     AL
## 3   0  0  0  0  1   0   0  0  0    0     5   0   0   0   0      BOS     AL
## 4   0  0  0  0  0   0   0  0  0    0    NA NaN NaN   0 NaN      SEA     AL
## 5   0  0  0  0  0   0   0  0  0    0    NA NaN NaN   0 NaN      SEA     AL
## 6  NA NA NA NA NA  NA  NA NA NA   NA    NA  NA  NA  NA  NA      NYA     AL
##    salary
## 1  300000
## 2  387500
## 3  403250
## 4  419000
## 5 2750000
## 6  500000
```

```r
#Removing null values
batsal_clean<-na.omit(batsal)
head(batsal_clean)
```

```
##     playerID yearID stint teamID.x lgID.x  G G_batting  AB  R  H X2B X3B
## 3  aardsda01   2008     1      BOS     AL 47         5   1  0  0   0   0
## 10  aasedo01   1989     1      NYN     NL 49        49   5  0  0   0   0
## 11  abadan01   2006     1      CIN     NL  5         5   3  0  0   0   0
## 14 abbotje01   1998     1      CHA     AL 89        89 244 33 68  14   1
## 15 abbotje01   1999     1      CHA     AL 17        17  57  5  9   0   0
## 16 abbotje01   2000     1      CHA     AL 80        80 215 31 59  15   1
##    HR RBI SB CS BB SO IBB HBP SH SF GIDP G_old        BA       OBP x1B
## 3   0   0  0  0  0  1   0   0  0  0    0     5 0.0000000 0.0000000   0
## 10  0   0  0  0  0  3   0   0  0  0    0    49 0.0000000 0.0000000   0
## 11  0   0  0  0  2  0   0   0  0  0    0     5 0.0000000 0.4000000   0
```

```
## 14 12  41  3  3  9 28   1   0  2  5    2    89 0.2786885 0.2984496  41
## 15  2   6  1  1  5 12   0   0  1  1    4    17 0.1578947 0.2222222   7
## 16  3  29  2  1 21 38   1   2  2  1    2    80 0.2744186 0.3430962  40
##          SLG teamID.y lgID.y salary
## 3  0.0000000      BOS     AL 403250
## 10 0.0000000      NYN     NL 400000
## 11 0.0000000      CIN     NL 327000
## 14 0.4918033      CHA     AL 175000
## 15 0.2631579      CHA     AL 255000
## 16 0.3953488      CHA     AL 255000
```

Having a look at lost players statistics:

```
#Subselecting data of players lost in 2001
lost_players<-subset(batsal_clean, yearID==2001 &
                     batsal_clean$playerID %in%
c('giambja01','damonjo01','saenzol01'),

select=c('playerID','H','X2B','X3B','HR','OBP','SLG','BA','AB'))
#Statistics of Lost players in 2001
mean(lost_players$OBP)

## [1] 0.3638687

sum(lost_players$AB)

## [1] 1469
```

Excluding lost players from the dataset:

```
#Excluding lost players from the analysis
remaining_players<-subset(batsal_clean, !(batsal_clean$playerID %in%
c('giambja01','damonjo01','saenzol01')))
```

Selecting the replacement players:

```
#Selecting replacement players according to the defined criteria
replacement_players<-subset(remaining_players,AB>300 & yearID==2001 &
OBP>0.37 & salary<=5000000)

#Arraning for highest OBP
replacement_players<-arrange(replacement_players,-OBP)

## Warning: package 'bindrcpp' was built under R version 3.4.4

#Selecting top 3 players base on OBP
top3<-head(replacement_players, 3)
top3

##    playerID yearID stint teamID.x lgID.x   G G_batting  AB   R   H X2B X3B
## 1 heltoto01   2001     1      COL     NL 159       159 587 132 197  54   2
## 2 berkmla01   2001     1      HOU     NL 156       156 577 110 191  55   5
## 3 gonzalu01   2001     1      ARI     NL 162       162 609 128 198  36   7
```

```
##   HR RBI SB CS  BB  SO IBB HBP SH SF GIDP G_old        BA       OBP x1B
## 1 49 146  7  5  98 104  15   5  1  5   14   159 0.3356048 0.4316547  92
## 2 34 126  7  9  92 121   5  13  0  6    8   156 0.3310225 0.4302326  97
## 3 57 142  1  1 100  83  24  14  0  5   14   162 0.3251232 0.4285714  98
##         SLG teamID.y lgID.y  salary
## 1 0.6848382      COL     NL 4950000
## 2 0.6204506      HOU     NL  305000
## 3 0.6880131      ARI     NL 4833333
```

```
#Statistics of replacement players
mean(top3$OBP)
```

```
## [1] 0.4301529
```

```
sum(top3$AB)
```

```
## [1] 1773
```

```
sum(top3$salary)
```

```
## [1] 10088333
```

Conclusion:

In the end we have successfully chosen three palyers who meet our original criteria we set in the introduction. These replacement players are some of the great players who are highly undervalued in the market. We have unearthed them thanks to the innovative features.

Further details about the players is provided in the links with their names below.

1.Todd Helton

https://www.baseball-reference.com/players/h/heltoto01.shtml

2.Lance Berkman

https://www.baseball-reference.com/players/b/berkmla01.shtml

3.Luis Gonzalez

https://www.baseball-reference.com/players/g/gonzalu01.shtml