



HOGESCHOOL VAN AMSTERDAM

CAPSTONE BIG-DATA IN URBAN TECHNOLOGY

# Batting

*Klaas Berbee*

*Thierry Ligeon*

supervised by

Dr. J.R. HELMUS

November 2, 2017

# Contents

<b>1</b>	<b>Capstone Batting</b>	<b>3</b>
1.1	Data importeren via MySQL . . . . .	3

*this page is intentionally left blank*

# 1 Capstone Batting

*Klaas Berbee & Thierry Ligeon*

Voor de capstone van de minor Big-Data in urban technology aan de Hogeschool van Amsterdam moeten wij de Batting dataset gaan bestuderen.

## 1.1 Data importeren via mySQL

Verbinding maken met de mySQL database.

Vervolgens maken we verbinding met de locale mySQL server. Vervolgens drukken we alle tabellen van de database af.

```
my_db <- dbConnect(MySQL(),
  user='klaasberbee',
  password='1190KVU',
  dbname='Batting',
  host='localhost')

print(dbListTables(my_db))

## [1] "AllstarFull"      "Appearances"      "AwardsManagers"
## [4] "AwardsPlayers"   "AwardsShareManagers" "AwardsSharePlayers"
## [7] "Batting"         "BattingPost"      "CollegePlaying"
## [10] "Fielding"        "FieldingOF"       "FieldingOFsplit"
## [13] "FieldingPost"    "HallOfFame"       "HomeGames"
## [16] "Managers"        "ManagersHalf"     "Master"
## [19] "Parks"           "Pitching"         "PitchingPost"
## [22] "Salaries"        "Schools"          "SeriesPost"
## [25] "Teams"           "TeamsFranchises"  "TeamsHalf"
```

Hier kunnen we zien dat de Batting dataset uit 27 verschillende tabellen bestaat. Uit de ERD analyse was gebleken dat ..... de belangrijkste tabellen zijn.

```
dbTables <- dbListTables(my_db)

allTables <-
  lapply(dbTables, function(table) {
    dbGetQuery(my_db, paste("select * from", table))
  })
names(allTables) <- paste("db", dbTables, sep = "_")
list2env(allTables, envir = .GlobalEnv)

## <environment: R_GlobalEnv>
```

Uit deze datasets kan alle data gevonden worden over de de gehele teams, maar ook over de individuele spelers en managers, zoals hun salaris en hun stats. Verder wordt er weergegeven hoe een game verloopt en wat voor activiteiten er in voor komen en door wie deze zijn uitgevoerd. Als laatst kan er ook uit de datasets opgemaakt worden hoeveel een speler verdient.

Het doel is om deze dataset om te zetten naar informatie en dit dan vervolgens om te zetten naar kennis. Er kan namelijk met al deze data kennis worden opgebouwt zoals welke factoren zijn belangrijk voor een sterk team of wat maakt een bepaalde speler zo goed.

belanghebbende voor deze dataset: gokkers: Wanneer gokkers een voorspelmodel kunnen maken van deze dataset met wie er wint, dan kunnen zij er veel geld mee verdienen op goksites.

De coaches van de team: De coach kan uit deze data set halen wat de sterke en de zwake punten zijn van zijn spelers, hierdoor weet hij welke skills hij bij bepaalde spelers moet verbeteren. Ook is deze dataset handig voor hem omdat hij de zwake en de sterke punten van de tegenstander weet. Op deze manier kan hij de inzet van zijn spelers inzetten dat hij er een voordeel uit kan halen.

Spelers die op zoek zijn naar een opleidings school: De spelers die op zoek zijn naar een school waar ze hun baseball skills willen verbeteren kunnen uit deze dataset opmaken waar welke speler is opgeleid. Uit de dataset kan misschien ook wel opgemaakt worden welke skills je bij welke school goed leert. Dit zou dan kunnen helpen bij het maken van zijn beslissing.

de raad van een stadion: Uit deze dataset kan opgemaakt worden hoeveel kijkers er komen kijken bij een wedstrijd. Bij de ene wedstrijd komen er namelijk meer dan bij een ander. Dit wetende kan de raad van een stadion bijvoorbeeld zijn hoeveelheid personeel hierop aanpassen of de mate van beveiliging.