



HOGESCHOOL VAN AMSTERDAM

CAPSTONE BIG-DATA IN URBAN TECHNOLOGY

Batting

Klaas Berbee

Thierry Ligeon

supervised by

Dr. J.R. HELMUS

November 2, 2017

Contents

1	Capstone Batting	3
1.1	Context van de dataset	3
1.1.1	Belanghebbende voor deze dataset	3
1.2	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. Als eerste beginnen we met het globaal inzicht krijgen van de dataset.

1.1 Context van de dataset

De Batting dataset bevat data over o.a.: teams, spelers en managers. Ook bevat de dataset informatie over hun salarissen en skills. Verder bevat deze dataset informatie over hoe de individuele wedstrijden verlopen zijn, welke activiteiten hierin zijn voorgekomen en door wie deze zijn uitgevoerd.

Het doel is om deze dataset om te zetten naar nuttige informatie en dit vervolgens conclusies uit kunnen trekken. Teams kunnen bijvoorbeeld een inzicht krijgen in welke factoren een grote rol spelen in de prestatie van het team.

1.1.1 Belanghebbende voor deze dataset

Kansspel spelers

Er kan een voordeel voor kansspelers ontstaan zodra zij bezitten over voorspelmodellen voor komende wedstrijden.

Teamcoaches

De coach kan uit de dataset de sterke en zwakke punten van de spelers inzien. Hierop kan hij zijn trainingen specificeren op bepaalde aspecten voor zijn spelers. Daarnaast kunnen de sterke en zwakke punten van de competitie bekeken worden om zo beter voor te kunnen bereiden op aankomende wedstrijden.

Junior spelers

Beginnende spelers kunnen hun idolen en vaardigheden vergelijken en bekijken bij welke club zij getraind hebben. Dit zou een aanleiding kunnen zijn om een club te kiezen

Stadions

Met deze dataset kan de hoeveelheid bezoekers voorspeld worden. Hierop kunnen prijzen, hoeveelheid beveiliging en promotie op worden aangepast.

1.2 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>
```