

# ***Wydarzenia koszykarskie 2003-2017***

**Piotr Kacprowicz, Damian Centek, Daniel Banasiak**

# ***Informacje o komputerach***

**Piotr Kacprowicz**

System	Win 10 Pro x64
Procesor	Intel Core i7-4710HQ, 4 rdzenie
Pamięć	12 GB, HyperX Savage 2400MHz DDR3
Dysk	Hitachi Travelstar 7K1000 1TB

# ***Informacje o komputerach***

**Damian Centek**

System	Win 10 Pro x64
Procesor	Intel Core i7-4750HQ, 4 rdzenie
Pamięć	8 GB, HyperX Savage 2400MHz DDR3
Dysk	Hitachi Travelstar 7K1000 1TB

# ***Informacje o komputerach***

**Daniel Banasiak**

System	Win 10 Pro x64
Procesor	Intel Core i5-5200U CPU 2.2 GHz 2.19 GHz
Pamięć	16 GB DDR3
Dysk	SSD LITEON LCH-256V2S-11 2.5 7mm 256 GB

# **Plik RegularSeasonDetailedResults.csv**

**Opis:** zawiera wszystkie informacje na temat rozegranych meczy

**Import:**

```
>mongoimport -d basketball -c results --type csv --file E:\MongoDB
```

**Czas importu: 2,744 sekundy**

**Ilość zaimportowanych dokumentów: 76636**

# Wyjaśnienie pól

- "season" - this is the year of the associated entry in seasons.csv (the year in which the final tournament occurs)
- "daynum" - this integer always ranges from 0 to 132, and tells you what day the game was played on. It represents an offset from the "dayzero" date in the "seasons.csv" file. For example, the first game in the file was daynum=20. Combined with the fact from the "season.csv" file that day zero was 10/29/1984, that means the first game was played 20 days later, or 11/18/1984. There are no teams that ever played more than one game on a given date, so you can use this fact if you need a unique key. In order to accomplish this uniqueness, we had to adjust one game's date. In March 2008, the SEC postseason tournament had to reschedule one game (Georgia-Kentucky) to a subsequent day, so Georgia had to actually play two games on the same day. In order to enforce this uniqueness, we moved the game date for the Georgia-Kentucky game back to its original date.
- "wteam" - this identifies the id number of the team that won the game, as listed in the "teams.csv" file. No matter whether the game was won by the home team or visiting team, "wteam" always identifies the winning team.
- "wscore" - this identifies the number of points scored by the winning team.
- "lteam" - this identifies the id number of the team that lost the game.
- "lscore" - this identifies the number of points scored by the losing team.
- "numot" - this indicates the number of overtime periods in the game, an integer 0 or higher.
- "wloc" - this identifies the "location" of the winning team. If the winning team was the home team, this value will be "H". If the winning team was the visiting team, this value will be "A". If it was played on a neutral court, then this value will be "N". Sometimes it is unclear whether the site should be considered neutral, since it is near one team's home court, or even on their court during a tournament, but for this determination we have simply used the Kenneth Massey data in its current state, where the "@" sign is either listed with the winning team, the losing team, or neither team.

# Wyjaśnienie pól c.d.

- wfgm - field goals made
- wfga - field goals attempted
- wfgm3 - three pointers made
- wfga3 - three pointers attempted
- wftm - free throws made
- wfta - free throws attempted
- wor - offensive rebounds
- wdr - defensive rebounds
- wast - assists
- wto - turnovers
- wstl - steals
- wblk - blocks
- wpf - personal fouls

# **Plik Teams.csv**

**Opis: zawiera informacje o nazwach drużyn**

**Import:**

```
mongoimport -d basketball -c teams --type csv --file E:\MongoDB
```

**Czas importu: 0,133 sekundy**

**Ilość zaimportowanych dokumentów: 364**



# Wyjaśnienie pól

## Teams

This file identifies the different college teams present in the dataset. Each team has a 4 digit id number.

# Agregacja 1: Ilość fauli w całym sezonie na przestrzeni 2003-2017

```
db.results.aggregate([
  { $group: {
    _id: { season: "$Season" },
    foul: { $sum: "$Wpf" }
  } },
  { $sort : { "foul": -1 } },
]).pretty()
```

# Agregacja 1: Ilość fauli w całym sezonie na przestrzeni 2003-2017 c.d.

**Wynik:**

```
{ "_id" : { "season" : 2016 }, "foul" : 98603 }  
{ "_id" : { "season" : 2014 }, "foul" : 96787 }  
{ "_id" : { "season" : 2017 }, "foul" : 96525 }  
{ "_id" : { "season" : 2011 }, "foul" : 92232 }  
{ "_id" : { "season" : 2010 }, "foul" : 92021 }  
{ "_id" : { "season" : 2015 }, "foul" : 91983 }  
{ "_id" : { "season" : 2009 }, "foul" : 90794 }  
{ "_id" : { "season" : 2008 }, "foul" : 90095 }  
{ "_id" : { "season" : 2012 }, "foul" : 88873 }  
{ "_id" : { "season" : 2013 }, "foul" : 88398 }  
{ "_id" : { "season" : 2007 }, "foul" : 87969 }  
{ "_id" : { "season" : 2003 }, "foul" : 82537 }  
{ "_id" : { "season" : 2006 }, "foul" : 82076 }  
{ "_id" : { "season" : 2005 }, "foul" : 81363 }  
{ "_id" : { "season" : 2004 }, "foul" : 81355 }
```

**Wniosek: W dzisiejszych czasach koszykówka jest dużo bardziej brutalniejsza**

# Agregacja 2: Wpływ boiska na wygrywanie drużyn

## Etap 1: Wybranie odpowiednich boisk

```
var stepHome1={  
  $match:{  
    "Wloc":{ $eq:"H"}  
  }  
};  
  
var stepAway1={  
  $match:{  
    "Wloc":{ $eq:"A"}  
  }  
};  
  
var stepNeutral1={  
  $match:{  
    "Wloc":{ $eq:"N"}  
  }  
};
```

# Agregacja 2: Wpływ boiska na wygrywanie drużyn c.d.

## Etap 2: Zliczenie wygranych na odpowiednich boiskach

```
var stepHome2={  
    $count:"Home"  
};  
  
var stepAway2={  
    $count:"Away"  
};  
  
var stepNeutral2={  
    $count:"Neutral"  
};
```

# Agregacja 2: Wpływ boiska na wygrywanie drużyn c.d.

**Cała agregacja:**

```
db.results.aggregate([
{
  $facet:{
    "Home": [stepHome1, stepHome2],
    "Away": [stepAway1, stepAway2],
    "Neutral": [stepNeutral1, stepNeutral2]
  }
}
]) .pretty()
```

## Agregacja 2: Wpływ boiska na wygrywanie drużyn c.d.

**Wynik:**

```
{
  "Home" : [
    {
      "Home" : 45512
    }
  ],
  "Away" : [
    {
      "Away" : 23431
    }
  ],
  "Neutral" : [
    {
      "Neutral" : 7693
    }
  ]
}
```

**Wniosek: Najlepiej wygrywa się u siebie**

# Agregacja 3: Najlepsze drużyny w poszczególnych sezonach

## Cała agregacja:

```
db.results.aggregate([
  { $lookup : {from:"team", localField:"Wteam", foreignField: "Team_Id", as:"win_team"}},

  { $group: {
    _id: {season:"$Season",win:"$win_team.Team_Name"},
    score: { $sum: 1 },
  }},
  { $sort : { score: -1} },
  { $project : { "win_team.Team_Name" : 1, score: 1 }}
])
```



## Agregacja 3: Najlepsze drużyny w poszczególnych sezonach c.d.

**Wynik:**

```
{ "_id" : { "season" : 2003, "win" : [ "Kentucky" ] }, "score" : 29 }
{ "_id" : { "season" : 2004, "win" : [ "Stanford" ] }, "score" : 29 }
{ "_id" : { "season" : 2005, "win" : [ "Illinois" ] }, "score" : 32 }
{ "_id" : { "season" : 2006, "win" : [ "Duke" ] }, "score" : 30 }
{ "_id" : { "season" : 2007, "win" : [ "Memphis" ] }, "score" : 30 }
{ "_id" : { "season" : 2008, "win" : [ "Memphis" ] }, "score" : 33 }
{ "_id" : { "season" : 2009, "win" : [ "Memphis" ] }, "score" : 31 }
{ "_id" : { "season" : 2010, "win" : [ "Kentucky" ] }, "score" : 32 }
{ "_id" : { "season" : 2011, "win" : [ "Ohio St" ] }, "score" : 32 }
{ "_id" : { "season" : 2012, "win" : [ "Kentucky" ] }, "score" : 32 }
{ "_id" : { "season" : 2013, "win" : [ "Gonzaga" ] }, "score" : 30 }
{ "_id" : { "season" : 2014, "win" : [ "Wichita St" ] }, "score" : 33 }
{ "_id" : { "season" : 2015, "win" : [ "Kentucky" ] }, "score" : 34 }
{ "_id" : { "season" : 2016, "win" : [ "Villanova" ] }, "score" : 29 }
{ "_id" : { "season" : 2017, "win" : [ "Gonzaga" ] }, "score" : 32 }
```

**Wniosek: Z roku na rok generalnie najlepsza drużyna osiąga coraz lepsze wyniki**

# Agregacja 4: Pięć najskuteczniejszych drużyn na przestrzeni 2012-2017

## Etap 1: Wybranie odpowiednich sezonów i złączenie tabel

```
var step1 = { $match: {"Season": { $gte:2012}}}
var step11 = { $lookup : {from:"team", localField:"Wteam", foreignField: "Team_Id", as:"win_team"}}
```

## Etap 2: Wyliczenie średnich wartości dla wykonanych rzutów oraz celnych rzutów

```
var step2 = { $group :{ _id: {win:"$win_team.Team_Name"}, twoPoints:{ $avg: "$Wfgm"},
  | twoPointT: {$avg: "$Wfga"},threePoints:{ $avg: "$Wfgm3"},threePointsT:{ $avg: "$Wfga3"}}}
```

## **Agregacja 4: Pięć najskuteczniejszych drużyn na przestrzeni 2012-2017 c.d.**

**Cała agregacja:**

```
db.results.aggregate([  
    step1,step11,step2,step3,step4  
]).pretty()
```

## Agregacja 4: Pięć najskuteczniejszych drużyn na przestrzeni 2012-2017 c.d.

**Wynik:**

```
{
  "_id" : {
    "win" : [
      "North Carolina"
    ]
  },
  "twoPoints" : 30.575163398692812,
  "twoPointT" : 63.88235294117647,
  "threePoints" : 5.895424836601308,
  "threePointsT" : 16.26797385620915
}
{
  "_id" : {
    "win" : [
      "Cent Arkansas"
    ]
  },
  "twoPoints" : 30.275,
  "twoPointT" : 62.6,
  "threePoints" : 7.7,
  "threePointsT" : 18.4
}
```

## Agregacja 4: Pięć najskuteczniejszych drużyn na przestrzeni 2012-2017 c.d.

```
{
  "_id" : {
    "win" : [
      "UCLA"
    ]
  },
  "twoPoints" : 30.075187969924812,
  "twoPointT" : 60.10526315789474,
  "threePoints" : 6.917293233082707,
  "threePointsT" : 17.022556390977442
}
{
  "_id" : {
    "win" : [
      "Northwestern LA"
    ]
  },
  "twoPoints" : 29.789473684210527,
  "twoPointT" : 61.3421052631579,
  "threePoints" : 7.592105263157895,
  "threePointsT" : 19.763157894736842
}
{
  "_id" : {
    "win" : [
      "VMI"
    ]
  },
  "twoPoints" : 29.35593220338983,
  "twoPointT" : 63.08474576271186,
  "threePoints" : 10.067796610169491,
  "threePointsT" : 27.627118644067796
}
```

# Agregacja 5: Zsumowana liczba punktów zdobyta przez zwycięskie oraz przegrane zespoły w danym sezonie (2003-2017)

**Etap 1: Wybranie przedziałów czasowych oraz zsumowanie liczby punktów dla zwycięzców oraz przegranych**

```
var step1 = {  
  $bucket:{  
    groupBy:"$Season",  
    boundaries:[2003,2004,2005,2006,2007,2008,2009,2010,2011,2012,2013,2014,2015,2016,2017],  
    default: "Other",  
    output: {  
      "count": { $sum: "$Wscore" }, "count1": { $sum: "$Lscore" },  
    }  
  }  
}
```

## Agregacja 5: Zsumowana liczba punktów zdobyta przez zwycięskie oraz przegrane zespoły w danym sezonie (2003-2017) c.d.

**Cała agregacja:**

```
db.results.aggregate([
  {
    $facet:{
      PointSeason: [step1]
    }
  }
]).pretty()
```

## Agregacja 5: Zsumowana liczba punktów zdobyta przez zwycięskie oraz przegrane zespoły w danym sezonie (2003-2017) c.d.

**Wynik:**

```
"PointSeason" : [  
  {  
    "_id" : 2003,  
    "count" : 349870,  
    "count1" : 294302  
  },  
  {  
    "_id" : 2004,  
    "count" : 342693,  
    "count1" : 287928  
  },  
  {  
    "_id" : 2005,  
    "count" : 349193,  
    "count1" : 292820  
  },  
  {  
    "_id" : 2006,  
    "count" : 354813,  
    "count1" : 298105  
  },  
  {  
    "_id" : 2007,  
    "count" : 377365,  
    "count1" : 316507  
  },  
]
```



## Agregacja 5: Zsumowana liczba punktów zdobyta przez zwycięskie oraz przegrane zespoły w danym sezonie (2003-2017) c.d.

```
{
  "_id" : 2008,
  "count" : 387031,
  "count1" : 324006
},
{
  "_id" : 2009,
  "count" : 388603,
  "count1" : 325335
},
{
  "_id" : 2010,
  "count" : 392896,
  "count1" : 328855
},
{
  "_id" : 2011,
  "count" : 390167,
  "count1" : 327721
},
{
  "_id" : 2012,
  "count" : 385465,
  "count1" : 322618
},
}
```

## Agregacja 5: Zsumowana liczba punktów zdobyta przez zwycięskie oraz przegrane zespoły w danym sezonie (2003-2017) c.d.

```
{
  {
    "_id" : 2013,
    "count" : 387513,
    "count1" : 324031
  },
  {
    "_id" : 2014,
    "count" : 408055,
    "count1" : 345592
  },
  {
    "_id" : 2015,
    "count" : 389133,
    "count1" : 326636
  },
  {
    "_id" : 2016,
    "count" : 420371,
    "count1" : 356110
  },
  {
    "_id" : "Other",
    "count" : 424288,
    "count1" : 359384
  }
}
```

**Wniosek: Z każdym sezonem liczba punktów zdobywana przez zwycięskie, jak i przegrane zespoły rosła**