

Ciekawe wnioski z danych sprzedażowych filmów

Sebastian Czarnecki

Maszyna do obliczeń

| | |
|----------|--------------------------|
| System | Windows 10 64 bit |
| Procesor | i5 323M 2 x 2,6 GHz |
| Pamięć | 8Gb 1600Mhz DDR3 Samsung |
| Dysk | 128 GB SSD |

Collection Movies

- 5043 documents
- movies across 100 years
- 66 countries
- 2399 unique directors names
- thousand names of actors

Collection fields

| | |
|----------------------|-----------------------------------|
| movie_title | Tytuł filmu |
| movie_facebook_likes | Liczba polubień filmu na facebook |
| director_name | Imie i nazwisko reżysera |
| actor_1_name | Pierwszy aktor |
| actor_2_name | Drugi aktor |
| actor_3_name | Aktor 3 |
| language | Język filmu |

Collection fileds cd.

| | |
|-------------------------|----------------------------|
| gross | Przychody filmu |
| budget | Koszt produkcji |
| title_year | Rok produkcji |
| imdb_score | Ocena filmu z imdb |
| country | Kraj produkujący |
| duration | Czas trwania w min |
| director_facebook_likes | Like reżysera na facebooku |

Import danych

```
mongoimport.exe -d NoSQLExam -c Movies --type csv movie_metadata.csv  
--headerline
```

Duration: 0.3 s

Imported documents: 5043

Czy wielki kryzys z 2007 roku wpłynął na sprzedaż filmów na świecie?

```
var stage2 = {  
  "$bucket": {  
    "groupBy": "$title_year",  
    "boundaries": [2000, 2007, 2014],  
    "default": "Other",  
    "output": {  
      "count" : { "$sum": "$gross" }  
    }  
  }  
}  
  
db.Movies.aggregate({  
  "$facet": {  
    "GrossInYear": [stage2]  
  }  
});
```

```
{
  "GrossInYear" : [
    {
      "_id" : 2000.0,
      "count" : NumberLong(56600834410)
    },
    {
      "_id" : 2007.0,
      "count" : NumberLong(67734330800)
    },
    {
      "_id" : "Other",
      "count" : NumberLong(77244941694)
    }
  ]
}
```


Dane obrobione

| | |
|-------------|------------------------|
| 2000 - 2007 | 566 008 344 10 dolarów |
| 2007 - 2014 | 677 343 308 00 dolarów |

Czy rzeczywiście był kryzys gospodarczy ? Na pewno nie w branży filmów

Krys dotknął głównie USA. Może chociaż tam wpłynął on negatywnie na branżę filmów?

```
var stage1 = {
  "$match": {"country": {"$eq": "USA"}}
};

var stage2 = {
  "$bucket": {
    "groupBy": "$title_year",
    "boundaries": [2000, 2007, 2014],
    "default": "Other",
    "output": {
      "count" : { "$sum": "$gross" }
    }
  }
}

db.Movies.aggregate({
  "$facet": {
    "GrossInYear": [stage1, stage2]
  }
});
```

Jednak nie. I tym razem 7 lat po kryzysie to większy dochód z filmów niż 7 lat przed kryzysem.

```
{
  "GrossInYear" : [
    {
      "_id" : 2000.0,
      "count" : NumberLong(48404032696)
    },
    {
      "_id" : 2007.0,
      "count" : NumberLong(61711998767)
    },
    {
      "_id" : "Other",
      "count" : NumberLong(68503222922)
    }
  ]
}
```

Może chociaż chwilowy spadek? Sprzedaż w latach 2005 - 2010

```
var stage1 = {
  "$match": {"title_year": {
    $gte: 2001,
    $lt: 2008}}
};

var stage2 = { $group: { _id: "$title_year", totalGross: { $sum: "$gross" } } };

var stage3 = { $sort: { _id : 1 } };

db.Movies.aggregate(
  stage1, stage2, stage3);
```

Spadek jest ale bardzo mały.

| | |
|------|-----------------------|
| 2005 | 835 530 608 8 dolarów |
| 2006 | 847 537 688 6 dolarów |
| 2007 | 828 188 268 3 dolarów |
| 2008 | 895 927 538 5 dolarów |
| 2009 | 993 459 964 3 dolarów |
| 2010 | 993 175 687 6 dolarów |

Czy liczba polubień na facebooku jest adekwatna do ogólnej oceny filmu na imDB?

```
var stage0 = {"$match": {"title_year": {"$gte": 2012}}}  
  
var stage1 = {"$project": {"director_name":1, movie_facebook_likes:1, imdb_score:1}}  
  
var stage2 = { $group: { _id: "$director_name",  
    movieLikes: { $avg: "$movie_facebook_likes" } ,avgMoviesScore: { $avg: "$imdb_score" } } };  
  
var stage3 = { $sort: { movieLikes : -1 } };  
  
var stage4 = { $sample: { size: 10 } }  
  
db.Movies.aggregate(  
stage0,stage1,stage2, stage3);
```

Dane posortowane

```
/* 1 */
{
  "_id" : "Christopher Nolan",
  "movieLikes" : 256500.0,
  "avgMoviesScore" : 8.55
}

/* 2 */
{
  "_id" : "George Miller",
  "movieLikes" : 191000.0,
  "avgMoviesScore" : 8.1
}

/* 3 */
{
  "_id" : "Morten Tyldum",
  "movieLikes" : 165000.0,
  "avgMoviesScore" : 8.1
}

/* 4 */
{
  "_id" : "Zack Snyder",
  "movieLikes" : 157500.0,
  "avgMoviesScore" : 7.05
}

/* 5 */
{
  "_id" : "Quentin Tarantino",
  "movieLikes" : 156500.0,
  "avgMoviesScore" : 8.2
}
```

Sample 5 random result

```
/* 1 */
{
  "_id" : "Scott Hicks",
  "movieLikes" : 29000.0,
  "avgMoviesScore" : 6.5
}

/* 2 */
{
  "_id" : "David DeCoteau",
  "movieLikes" : 102.0,
  "avgMoviesScore" : 3.0
}

/* 3 */
{
  "_id" : "Mark Griffiths",
  "movieLikes" : 170.0,
  "avgMoviesScore" : 6.3
}

/* 4 */
{
  "_id" : "Richard Glatzer",
  "movieLikes" : 45000.0,
  "avgMoviesScore" : 7.5
}

/* 5 */
{
  "_id" : "Phil Joanou",
  "movieLikes" : 0.0,
  "avgMoviesScore" : 4.7
}
```