Rafal Stapinski mongo mini-project

1. Find all reviews for a hotel 'Desert Rose Hotel'

```
db.reviews.findOne(
    {
        "HotelInfo.HotelID": db.reviews.findOne(
            {
                "HotelInfo.Name": "Desert Rose Resort"
            }
        ).HotelInfo.HotelID
    }
).Reviews
```

```
[
    {
        "Ratings" : {
            "Service" : 5,
            "Cleanliness" : "4",
            "Overall" : 5,
            "Value" : 5,
            "Location" : "4",
            "Rooms" : "4"
        },
        "AuthorLocation" : "",
        "Title" : ""“Best Deal for large families.”"",
        "Author" : "pb1bicdic",
        "ReviewID" : "UR127021251",
        "Content" : "So I found out about the hotel from searching this
    site. I found the reviews to be accurate and honest. When we check in, not
    five minutes into the hotel concierge gave us great info on free shows on
    the strip and the best way to maneuver from strip to downtown. If I could
    change anything it would be the hours of contenital breakfast. That time
    went so fast and the line so long that I only wanted to grab juice and
    leave. However, the staff was very accommodating and friendly that more
    than made up for it.",
        "Date" : ISODate("2012-03-31T04:00:00Z")
    }
    // ... more not included
]
```

2. Number of ratings for each hotel. Sort the results.

```
db.reviews.aggregate(
    [
```

```
        {
            $project: {
                HotelID: "$HotelInfo.HotelID",
                numberOfReviews: {
                    $size: "$Reviews"
                },
                _id: 0
            }
        },
        {
            $sort: {
                numberOfReviews: -1
            }
        }
    ]
)
```

```
{ "HotelID" : "2515499", "numberOfReviews" : 4936 }
{ "HotelID" : "91703", "numberOfReviews" : 4778 }
{ "HotelID" : "2515740", "numberOfReviews" : 4510 }
{ "HotelID" : "91891", "numberOfReviews" : 4373 }
{ "HotelID" : "2515443", "numberOfReviews" : 3738 }
{ "HotelID" : "97704", "numberOfReviews" : 3680 }
{ "HotelID" : "2515676", "numberOfReviews" : 3120 }
{ "HotelID" : "530483", "numberOfReviews" : 3065 }
{ "HotelID" : "2515633", "numberOfReviews" : 3030 }
{ "HotelID" : "91687", "numberOfReviews" : 3024 }
{ "HotelID" : "2515773", "numberOfReviews" : 3008 }
{ "HotelID" : "91967", "numberOfReviews" : 2932 }
{ "HotelID" : "260444", "numberOfReviews" : 2920 }
{ "HotelID" : "97737", "numberOfReviews" : 2911 }
{ "HotelID" : "2515540", "numberOfReviews" : 2842 }
{ "HotelID" : "2515796", "numberOfReviews" : 2814 }
{ "HotelID" : "2515577", "numberOfReviews" : 2808 }
{ "HotelID" : "503598", "numberOfReviews" : 2738 }
{ "HotelID" : "601419", "numberOfReviews" : 2728 }
{ "HotelID" : "97786", "numberOfReviews" : 2718 }
// more not included
```

3. Average overall ratings for each hotel. Sort the results

```
db.reviews.aggregate(
    [
        {
            $unwind: "$Reviews"
        },
        {
            $group: {
                _id: "$HotelInfo.HotelID",
```

```
                avgOverall: {
                    $avg: "$Reviews.Ratings.Overall"
                }
            }
        },
        {
            $sort: {
                avgOverall: -1
            }
        },
        {
            $project: {
                HotelID: "$_id",
                avgOverall: "$avgOverall",
                _id: 0
            }
        }
    ]
)
```

```
{ "HotelID" : "949915", "avgOverall" : 5 }
{ "HotelID" : "877948", "avgOverall" : 5 }
{ "HotelID" : "805712", "avgOverall" : 5 }
{ "HotelID" : "758193", "avgOverall" : 5 }
{ "HotelID" : "675736", "avgOverall" : 5 }
{ "HotelID" : "675236", "avgOverall" : 5 }
{ "HotelID" : "674318", "avgOverall" : 5 }
{ "HotelID" : "672625", "avgOverall" : 5 }
{ "HotelID" : "671259", "avgOverall" : 5 }
{ "HotelID" : "666557", "avgOverall" : 5 }
{ "HotelID" : "659829", "avgOverall" : 5 }
{ "HotelID" : "666555", "avgOverall" : 5 }
{ "HotelID" : "642935", "avgOverall" : 5 }
{ "HotelID" : "642541", "avgOverall" : 5 }
{ "HotelID" : "639165", "avgOverall" : 5 }
{ "HotelID" : "638980", "avgOverall" : 5 }
{ "HotelID" : "619780", "avgOverall" : 5 }
{ "HotelID" : "614386", "avgOverall" : 5 }
{ "HotelID" : "648897", "avgOverall" : 5 }
{ "HotelID" : "651617", "avgOverall" : 5 }
// more not included
```

4. Show hotels with number of 5.0 overall ratings that they received.

```
db.reviews.aggregate(
    [
        {
            $unwind: "$Reviews"
        },
```

```
                        {
                            $project: {
                                HotelID: "$HotelInfo.HotelID",
                                fiveOCount: {
                                    $cond: [
                                        {
                                            $eq: [
                                                "$Reviews.Ratings.Overall", 5.0
                                            ]
                                        },
                                        1,
                                        0
                                    ]
                                }
                            },
                        },
                        {
                            $group: {
                                _id: "$HotelID",
                                fiveOCount: {
                                    $sum: "$fiveOCount"
                                }
                            }
                        },
                        {
                            $project: {
                                HotelID: "$_id",
                                fiveOCount: 1,
                                _id: 0
                            }
                        }
                    ]
                )
```

```
{ "fiveOCount" : 307, "HotelID" : "99766" }
{ "fiveOCount" : 192, "HotelID" : "99762" }
{ "fiveOCount" : 39, "HotelID" : "99443" }
{ "fiveOCount" : 169, "HotelID" : "99371" }
{ "fiveOCount" : 68, "HotelID" : "99332" }
{ "fiveOCount" : 29, "HotelID" : "99365" }
{ "fiveOCount" : 95, "HotelID" : "98731" }
{ "fiveOCount" : 0, "HotelID" : "99292" }
{ "fiveOCount" : 60, "HotelID" : "98733" }
{ "fiveOCount" : 56, "HotelID" : "98728" }
{ "fiveOCount" : 30, "HotelID" : "98717" }
{ "fiveOCount" : 5, "HotelID" : "98725" }
{ "fiveOCount" : 0, "HotelID" : "98713" }
{ "fiveOCount" : 2, "HotelID" : "98704" }
{ "fiveOCount" : 1, "HotelID" : "98702" }
{ "fiveOCount" : 2, "HotelID" : "98693" }
{ "fiveOCount" : 3, "HotelID" : "98687" }
```

```
{ "fiveOCount" : 4, "HotelID" : "98685" }
{ "fiveOCount" : 0, "HotelID" : "98678" }
{ "fiveOCount" : 0, "HotelID" : "98652" }
// more not included
```

5. Number of ratings given out per month/day of week

```
db.reviews.aggregate(
    [
        {
            $unwind: "$Reviews"
        },
        {
            $group: {
                _id: {
                    $dayOfWeek: "$Reviews.Date"
                },
                reviewsOnDay: {
                    $sum: 1
                }
            }
        },
        {
            $project: {
                _id: 0,
                dayOfWeek: "$_id",
                count: 1
            }
        }
    ]
)
```

```
{ "count" : 217323, "dayOfWeek" : 1 }
{ "count" : 178605, "dayOfWeek" : 7 }
{ "count" : 263769, "dayOfWeek" : 3 }
{ "count" : 208368, "dayOfWeek" : 6 }
{ "count" : 236146, "dayOfWeek" : 4 }
{ "count" : 283292, "dayOfWeek" : 2 }
{ "count" : 234453, "dayOfWeek" : 5 }
```

6. Number of reviews per author. (Sorted)

```
db.reviews.aggregate(
    [
        {
            $unwind: "$Reviews"
```

```
        },
        {
            $group: {
                _id: "$Reviews.Author",
                reviewsByAuthor: {
                    $sum: 1
                }
            }
        },
        {
            $project: {
                _id: 0,
                author: "$_id",
                reviewsByAuthor: 1
            }
        },
        {
            $sort: {
                reviewsByAuthor: -1
            }
        }
    ],
    // {
    //      allowDiskUse: true
    // }
)
```

```
{ "reviewsByAuthor" : 75503, "author" : "A TripAdvisor Member" }
{ "reviewsByAuthor" : 17432, "author" : "lass=" }
{ "reviewsByAuthor" : 4656, "author" : "Posted by an Accorhotels.com
traveler" }
{ "reviewsByAuthor" : 262, "author" : "Posted by an Easytobook.com
traveler" }
{ "reviewsByAuthor" : 178, "author" : "Posted by a hotelsgrandparis.com
traveler" }
{ "reviewsByAuthor" : 97, "author" : "David S" }
{ "reviewsByAuthor" : 90, "author" : "David B" }
{ "reviewsByAuthor" : 89, "author" : "David M" }
{ "reviewsByAuthor" : 85, "author" : "John C" }
{ "reviewsByAuthor" : 84, "author" : "John S" }
{ "reviewsByAuthor" : 80, "author" : "David H" }
{ "reviewsByAuthor" : 74, "author" : "John B" }
{ "reviewsByAuthor" : 69, "author" : "ITA_And_RE_a" }
{ "reviewsByAuthor" : 68, "author" : "John M" }
{ "reviewsByAuthor" : 66, "author" : "John R" }
{ "reviewsByAuthor" : 66, "author" : "David C" }
{ "reviewsByAuthor" : 65, "author" : "Pawel_EPWR" }
{ "reviewsByAuthor" : 64, "author" : "Paul B" }
{ "reviewsByAuthor" : 63, "author" : "Chris B" }
{ "reviewsByAuthor" : 60, "author" : "Paul M" }
// more not included
```

Export to relational format

```
db.reviews.aggregate(
    [
        {
            $unwind: "$Reviews"
        },
        {
            $project: {
                _id: "$Reviews.Ratings.ReviewID",
                HotelName: {
                    $ifNull: [
                        "$HotelInfo.Name", null
                    ]
                },
                HotelID: "$HotelInfo.HotelID",
                Author: "$Reviews.Author",
                Overall: "$Reviews.Ratings.Overall",
                Date: "$Reviews.Date"
            }
        },
        {
            $out: "results"
        }
    ]
)
```

1. Find all reviews for a hotel 'Desert Rose Hotel'

```sql
SELECT * FROM reviews r1
WHERE r1."HotelID" IN
    (SELECT r2."HotelID"
    FROM reviews r2
    WHERE r2."HotelName" = 'Desert Rose Resort'
    LIMIT 1);
```

```
// shortened output

HotelName           HotelID Author          Overall Date
Desert Rose Resort      217598  pb1bicdic           5       2012-03-31
04:00:00
Desert Rose Resort      217598  Nancy I         2       2012-03-30
04:00:00
Desert Rose Resort      217598  John E          5       2012-03-29
04:00:00
Desert Rose Resort      217598  staycation79    4       2012-03-29
```

```
04:00:00
Desert Rose Resort      217598  Wanda M          5           2012-03-29
04:00:00
Desert Rose Resort      217598  xexec            5           2012-03-28
04:00:00
Desert Rose Resort      217598  AJ115            5           2012-03-28
04:00:00
Desert Rose Resort      217598  isabehr          5           2012-03-28
04:00:00
Desert Rose Resort      217598  grammiewags          5       2012-03-27
04:00:00
```

2. Number of ratings for each hotel. Sort the results.

```
SELECT r."HotelID", COUNT(*) as c
FROM reviews r
GROUP BY r."HotelID"
ORDER BY c DESC;
```

```
// shortened output

HotelID c
2515499 4936
91703   4778
2515740 4510
91891   4373
2515443 3738
97704   3680
2515676 3120
530483  3065
2515633 3030
91687   3024
2515773 3008
```

3. Average overall ratings for each hotel. Sort the results

```
SELECT r."HotelID", AVG(r."Overall") as a
FROM reviews r
GROUP BY r."HotelID"
ORDER BY a DESC;
```

```
// shortened output

HotelID,    a
```

```
1095040      5
666555       5
619780       5
253833       5
255395       5
1131703      5
1048931      5
654419       5
1204564      5
484842       5
652557       5
656305       5
1213098      5
```

4. Show hotels with number of 5.0 overall ratings that they received.

This query is ordered unlike the mongo equivalent above

```sql
SELECT r."HotelID", COUNT(*) as c
FROM reviews r
WHERE r."Overall" = 5
GROUP BY r."HotelID"
ORDER BY c DESC;
```

```
// shortened output

HotelID c
2515499 2791
91703   2695
530483  2319
2515443 1974
97704   1942
2515577 1896
1007582 1863
601419  1846
218524  1725
2515540 1724
503598  1646
774804  1530
578371  1413
2515740 1358
```

5. Number of ratings given out per month/day of week

```sql
SELECT
    date_part('month', r."Date") as m,
```

```
      date_part('dow', r."Date") as d,
      COUNT(*) as c
FROM reviews r
GROUP BY m, d;
```

```
// shortened output

m          d          c
1          0          20831
1          1          24494
1          2          22554
1          3          19442
1          4          20506
1          5          17212
1          6          16007
2          0          16783
2          1          22173
2          2          19886
2          3          19017
2          4          16428
2          5          15911
2          6          14033
```

6. Number of reviews per author. (Sorted)

```
SELECT r."Author", COUNT(*) as c
FROM reviews r
GROUP BY r."Author"
ORDER BY c DESC;
```

```
// shortened output

Author                                             c
A TripAdvisor Member                               75503
lass=                                              17432
Posted by an Accorhotels.com traveler              4656
Posted by an Easytobook.com traveler               262
Posted by a hotelsgrandparis.com traveler          178
David S                                            97
David B                                            90
David M                                            89
John C                                             85
John S                                             84
David H                                            80
```