Aggregation and Grouping

1 Count the Number of Movies Per Year (Descending Order)

(Identify the most prolific years for movie releases)

```
db.movies.aggregate([
  { $group: { _id: "$year", count: { $sum: 1 } } },
  { $sort: { count: -1 } }
])
```

```
> db.movies.aggregate([
    { $group: { _id: "$year", count: { $sum: 1 } } },
    { $sort: { count: -1 } }
  ])
< {
    _id: 2013,
    count: 1105
  }
  {
    _id: 2014,
    count: 1073
  }
  {
    _id: 2012,
    count: 955
  }
  {
    _id: 2009,
    count: 917
  }
  {
    _id: 2011,
    count: 893
  }
  {
    _id: 2008,
    count: 886
  }
}
```

```
}
{
  _id: 2004,
  count: 678
}
{
  _id: 2002,
  count: 622
}
{
  _id: 2001,
  count: 612
}
{
  _id: 2003,
  count: 603
}
{
  _id: 2000,
  count: 581
}
{
  _id: 1999,
  count: 515
}
{
```

```
}
{
  _id: 2015,
  count: 480
}
{
  _id: 1997,
  count: 439
}
{
  _id: 1996,
  count: 407
}
Type "it" for more
Atlas atlas-h4o7hc-shard-0 [primary] sample_mflix>
```

---

2 Find the Average IMDB Rating by Genre

(Analyze how different genres are rated on average)

```
db.movies.aggregate([
  { $unwind: "$genres" },
  { $group: { _id: "$genres", avgRating: { $avg: "$imdb.rating" } } },
  { $sort: { avgRating: -1 } }
])
```

```
> db.movies.aggregate([
    { $unwind: "$genres" },
    { $group: { _id: "$genres", avgRating: { $avg: "$imdb.rating" } } },
    { $sort: { avgRating: -1 } }
])
< {
    _id: 'Film-Noir',
    avgRating: 7.397402597402598
}
{
    _id: 'Short',
    avgRating: 7.377574370709382
}
{
    _id: 'Documentary',
    avgRating: 7.365679824561403
}
{
    _id: 'News',
    avgRating: 7.252272727272728
}
```

```
{
    _id: 'History',
    avgRating: 7.1696100917431185
}
{
    _id: 'War',
    avgRating: 7.128591954022989
}
{
    _id: 'Biography',
    avgRating: 7.087984189723319
}
{
    _id: 'Talk-Show',
    avgRating: 7
}
{
    _id: 'Animation',
    avgRating: 6.89669603524229
}
{
    _id: 'Music',
    avgRating: 6.883333333333334
}
```

```
{
  _id: 'Western',
  avgRating: 6.823553719008264
}
{
  _id: 'Drama',
  avgRating: 6.803377338624768
}
{
  _id: 'Sport',
  avgRating: 6.749041095890411
}
{
  _id: 'Crime',
  avgRating: 6.688585405625764
}
{
  _id: 'Musical',
  avgRating: 6.665831435079727
}
{
  _id: 'Romance',
  avgRating: 6.6564272782136396
}
{
```

```
{
  _id: 'Mystery',
  avgRating: 6.527425044091711
}
{
  _id: 'Adventure',
  avgRating: 6.493680884676145
}
{
  _id: 'Comedy',
  avgRating: 6.450214658080344
}
{
  _id: 'Fantasy',
  avgRating: 6.3829847908745245
}
Type "it" for more
```

---

3 Count Movies by Director and Sort by Count

(Find the most prolific directors)

```
db.movies.aggregate([
  { $unwind: "$directors" },
  { $group: { _id: "$directors", movieCount: { $sum: 1 } } },
  { $sort: { movieCount: -1 } },
  { $limit: 10 }
])
```

```
> db.movies.aggregate([
    { $unwind: "$directors" },
    { $group: { _id: "$directors", movieCount: { $sum: 1 } } },
    { $sort: { movieCount: -1 } },
    { $limit: 10 }
  ])
< {
    _id: 'Woody Allen',
    movieCount: 40
  }
  {
    _id: 'Martin Scorsese',
    movieCount: 32
  }
  {
    _id: 'Takashi Miike',
    movieCount: 31
  }
  {
    _id: 'Sidney Lumet',
    movieCount: 29
  }
```

```
    _id: 'Steven Spielberg',
    movieCount: 29
  }
  {
    _id: 'John Ford',
    movieCount: 29
  }
  {
    _id: 'Robert Altman',
    movieCount: 27
  }
  {
    _id: 'Clint Eastwood',
    movieCount: 27
  }
  {
    _id: 'Michael Apted',
    movieCount: 27
  }
  {
    _id: 'Spike Lee',
    movieCount: 27
  }
Atlas atlas-h4o7hc-shard-0 [primary] sample_mflix >
```

---

4 Calculate the Average Number of Cast Members per Movie by Year

(Track how cast sizes have evolved over time)

```
db.movies.aggregate([
  { $project: { year: 1, castSize: { $size: { $ifNull: ["$cast", []] }
} } },
  { $group: { _id: "$year", avgCastSize: { $avg: "$castSize" } } },
  { $sort: { _id: -1 } }
])
```

```
> db.movies.aggregate([
    { $project: { year: 1, castSize: { $size: { $ifNull: ["$cast", []] } } } },
    { $group: { _id: "$year", avgCastSize: { $avg: "$castSize" } } },
    { $sort: { _id: -1 } }
  ])
< {
    _id: '2014è',
    avgCastSize: 4
  }
  {
    _id: '2012è',
    avgCastSize: 4
  }
  {
    _id: '2011è',
    avgCastSize: 4
  }
  {
    _id: '2010è',
    avgCastSize: 2.25
  }
  {
    _id: '2009è',
    avgCastSize: 2.5
  }
  {
    _id: '2007è',
    avgCastSize: 4
  }
```

```
  {
    _id: '2006è2007',
    avgCastSize: 4
  }
  {
    _id: '2006è',
    avgCastSize: 4
  }
  {
    _id: '2005è',
    avgCastSize: 2.5
  }
  {
    _id: '2003è',
    avgCastSize: 4
  }
  {
    _id: '2002è',
    avgCastSize: 4
  }
  {
    _id: '2000è',
    avgCastSize: 4
  }
  {
    _id: '1999è',
    avgCastSize: 1
  }
```

```
db.movies.aggregate([
    { $project: { year: 1, castSize: { $size: { $ifNull: ["$cast", []] } } } },
    { $group: { _id: "$year", avgCastSize: { $avg: "$castSize" } } },
```

```
{
  _id: '1995è',
  avgCastSize: 1
}
{
  _id: '1994è1998',
  avgCastSize: 4
}
{
  _id: '1988è',
  avgCastSize: 4
}
{
  _id: '1987è',
  avgCastSize: 4
}
Type "it" for more
```

---

⑤ Find the Highest Rated Movie Per Director

(Discover a director's top-rated film)

```
db.movies.aggregate([
   { $unwind: "$directors" },
   { $sort: { "imdb.rating": -1 } },
   { $group: { _id: "$directors", topMovie: { $first: "$title" },
rating: { $first: "$imdb.rating" } } },
   { $sort: { rating: -1 } }
])
```

```
> db.movies.aggregate([
    { $unwind: "$directors" },
    { $sort: { "imdb.rating": -1 } },
    { $group: { _id: "$directors", topMovie: { $first: "$title" }, rating: { $first: "$imdb.rating" } } },
    { $sort: { rating: -1 } }
])
< {
    _id: 'Jennifer B. White',
    topMovie: 'Mary Loss of Soul',
    rating: ''
  }
  {
    _id: 'Vanessa Hope',
    topMovie: 'All Eyes and Ears',
    rating: ''
  }
  {
    _id: 'Shalini Kantayya',
    topMovie: 'Catching the Sun',
    rating: ''
  }
  {
    _id: 'Eric Weber',
    topMovie: 'Outliving Emily',
    rating: ''
```

```
{
  _id: 'Elena Hazanov',
  topMovie: 'Sindrom Petrushki',
  rating: ''
}
{
  _id: 'Duke Johnson',
  topMovie: 'Anomalisa',
  rating: ''
}
{
  _id: 'Charlie Kaufman',
  topMovie: 'Anomalisa',
  rating: ''
}
{
  _id: 'Gan Bi',
  topMovie: 'Kaili Blues',
  rating: ''
}
{
  _id: 'Warren P. Sonoda',
  topMovie: 'The Masked Saint',
  rating: ''
}
{
  _id: 'Ilya Chizhikov',
```

```
{
  _id: 'Ilya Chizhikov',
  topMovie: 'Paren s nashego kladbishcha',
  rating: ''
}
{
  _id: 'Simon Rouby',
  topMovie: 'Adama',
  rating: ''
}
{
  _id: 'Gooc-jin Ahn',
  topMovie: 'Alice in Earnestland',
  rating: ''
}
{
  _id: 'Tuija-Maija Niskanen',
  topMovie: 'Landet som icke èr',
  rating: ''
}
{
  _id: 'Anna Kadykova',
  topMovie: 'Krot na more',
  rating: ''
}
{
```

```
  _id: 'Keri Collins',
  topMovie: 'Convenience',
  rating: ''
}
{
  _id: 'Vanessa Roth',
  topMovie: 'No Tomorrow',
  rating: ''
}
{
  _id: 'Brady Corbet',
  topMovie: 'The Childhood of a Leader',
  rating: ''
}
{
  _id: 'Eitan Reuven',
  topMovie: 'Another World',
  rating: ''
}
{
  _id: 'Vasiliy Sigarev',
  topMovie: 'Strana Oz',
  rating: ''
}
Type "it" for more
Atlas atlas-h4o7hc-shard-0 [primary] sample_mflix>
```

6 Count the Number of Movies by Rating Category

(Evaluate how many movies fall into different rating categories)

```
db.movies.aggregate([
  { $group: { _id: "$rated", count: { $sum: 1 } } },
  { $sort: { count: -1 } }
])
```

```
> db.movies.aggregate([
    { $group: { _id: "$rated", count: { $sum: 1 } } },
    { $sort: { count: -1 } }
  ])
< {
    _id: null,
    count: 9895
  }
  {
    _id: 'R',
    count: 5537
  }
  {
    _id: 'PG-13',
    count: 2322
  }
  {
    _id: 'PG',
    count: 1852
  }
  {
    _id: 'APPROVED',
    count: 709
  }
  {
    _id: 'G',
    count: 477
```

```
  }
  {
    _id: 'PASSED',
    count: 181
  }
  {
    _id: 'TV-14',
    count: 89
  }
  {
    _id: 'TV-PG',
    count: 76
  }
  {
    _id: 'TV-MA',
    count: 60
  }
  {
    _id: 'TV-G',
    count: 59
  }
  {
    _id: 'GP',
    count: 44
  }
```

```
  {
    _id: 'M',
    count: 37
  }
  {
    _id: 'Approved',
    count: 5
  }
  {
    _id: 'AO',
    count: 3
  }
  {
    _id: 'TV-Y7',
    count: 3
  }
  {
    _id: 'OPEN',
    count: 1
  }
  {
    _id: 'Not Rated',
    count: 1
  }
Atlas atlas-h4o7hc-shard-0 [primary] sample_mflix>
```

7 Calculate the Average IMDB Rating by Release Year (Last 30 Years)

(Analyze how ratings have trended over recent decades)

```
db.movies.aggregate([
  { $match: { year: { $gte: 1994 } } },
  { $group: { _id: "$year", avgRating: { $avg: "$imdb.rating" } } },
  { $sort: { _id: 1 } }
])
```

```
> db.movies.aggregate([
    { $match: { year: { $gte: 1994 } } },
    { $group: { _id: "$year", avgRating: { $avg: "$imdb.rating" } } },
    { $sort: { _id: 1 } }
  ])
< {
    _id: 1994,
    avgRating: 6.62327868852459
  }
  {
    _id: 1995,
    avgRating: 6.600806451612903
  }
  {
    _id: 1996,
    avgRating: 6.569041769041768
  }
  {
    _id: 1997,
    avgRating: 6.6029612756264235
  }
  {
    _id: 1998,
    avgRating: 6.638791423001949
  }
  {
    _id: 1999,
    avgRating: 6.620582524271844
  }
```

```
  {
    _id: 2000,
    avgRating: 6.518588640275388
  }
  {
    _id: 2001,
    avgRating: 6.5663398692810455
  }
  {
    _id: 2002,
    avgRating: 6.547749196141479
  }
  {
    _id: 2003,
    avgRating: 6.5318407960199005
  }
  {
    _id: 2004,
    avgRating: 6.656342182890856
  }
  {
    _id: 2005,
    avgRating: 6.582748948106592
  }
```

```
{
  _id: 2006,
  avgRating: 6.597286821705427
}
{
  _id: 2007,
  avgRating: 6.585555555555556
}
{
  _id: 2008,
  avgRating: 6.552994350282486
}
{
  _id: 2009,
  avgRating: 6.506543075245365
}
{
  _id: 2010,
  avgRating: 6.515473441108544
}
{
  _id: 2011,
  avgRating: 6.453415453527435
}
{
```

```
{
  _id: 2012,
  avgRating: 6.484486373165618
}
{
  _id: 2013,
  avgRating: 6.491477787851315
}
Type "it" for more
```

8️⃣ Find Popular Actors Based on Number of Appearances

(Identify the most frequently cast actors)

```
db.movies.aggregate([
  { $unwind: "$cast" },
  { $group: { _id: "$cast", appearances: { $sum: 1 } } },
  { $sort: { appearances: -1 } },
  { $limit: 10 }
])
```

```
> db.movies.aggregate([
    { $unwind: "$cast" },
    { $group: { _id: "$cast", appearances: { $sum: 1 } } },
    { $sort: { appearances: -1 } },
    { $limit: 10 }
  ])
< {
    _id: 'Gèrard Depardieu',
    appearances: 67
  }
  {
    _id: 'Robert De Niro',
    appearances: 58
  }
  {
    _id: 'Michael Caine',
    appearances: 51
  }
  {
    _id: 'Bruce Willis',
    appearances: 49
  }
```

```
  {
    _id: 'Morgan Freeman',
    appearances: 48
  }
  {
    _id: 'Samuel L. Jackson',
    appearances: 48
  }
  {
    _id: 'Christopher Plummer',
    appearances: 47
  }
  {
    _id: 'Gene Hackman',
    appearances: 46
  }
  {
    _id: 'Max von Sydow',
    appearances: 45
  }
  {
    _id: 'Nicolas Cage',
    appearances: 45
  }
Atlas atlas-h4o7hc-shard-0 [primary] sample_mflix>
```

9️⃣ Group Movies by Language and Find the Average IMDB Rating

(Compare the performance of different languages in cinema)

```
db.movies.aggregate([
  { $unwind: "$languages" },
  { $group: { _id: "$languages", avgRating: { $avg: "$imdb.rating" },
count: { $sum: 1 } } },
```

```
    { $match: { count: { $gt: 5 } } },  // Only consider languages with
more than 5 movies
    { $sort: { avgRating: -1 } }
])
```

```
> db.movies.aggregate([
    { $unwind: "$languages" },
    { $group: { _id: "$languages", avgRating: { $avg: "$imdb.rating" }, count: { $sum: 1 } } },
    { $match: { count: { $gt: 5 } } },  // Only consider languages with more than 5 movies
    { $sort: { avgRating: -1 } }
  ])
<{
    _id: 'Georgian',
    avgRating: 7.8,
    count: 24
  }
  {
    _id: 'Malayalam',
    avgRating: 7.729411764705882,
    count: 17
  }
  {
    _id: 'Bengali',
    avgRating: 7.663888888888888,
    count: 36
  }
  {
    _id: 'Romany',
    avgRating: 7.457894736842104,
    count: 19
```

```
  {
    _id: 'Tamil',
    avgRating: 7.425301204819276,
    count: 83
  }
  {
    _id: 'Azerbaijani',
    avgRating: 7.414285714285714,
    count: 7
  }
  {
    _id: 'Shanghainese',
    avgRating: 7.4125,
    count: 16
  }
  {
    _id: 'Marathi',
    avgRating: 7.390909090909091,
    count: 22
  }
  {
    _id: 'Khmer',
    avgRating: 7.390000000000001,
    count: 10
```

```
{
    _id: 'Maya',
    avgRating: 7.316666666666666,
    count: 6
}
{
    _id: 'Bulgarian',
    avgRating: 7.2875000000000005,
    count: 24
}
{
    _id: 'Sicilian',
    avgRating: 7.2875,
    count: 8
}
{
    _id: 'Urdu',
    avgRating: 7.28,
    count: 60
}
{
    _id: 'Somali',
    avgRating: 7.271428571428571,
    count: 7
```

```
{
    _id: 'Pushto',
    avgRating: 7.264705882352941,
    count: 17
}
{
    _id: 'Turkish',
    avgRating: 7.250993377483445,
    count: 151
}
{
    _id: 'Hakka',
    avgRating: 7.25,
    count: 6
}
{
    _id: 'Persian',
    avgRating: 7.247692307692308,
    count: 130
}
{
    _id: 'Kurdish',
    avgRating: 7.234782608695652,
    count: 23
```

```
}
{
    _id: 'Min Nan',
    avgRating: 7.234375,
    count: 32
}
Type "it" for more
Atlas atlas-h4o7hc-shard-0 [primary] sample_mflix >
```

🔟 Calculate the Average Runtime by Genre

(See which genres tend to have longer runtimes)

```
db.movies.aggregate([
  { $unwind: "$genres" },
  { $group: { _id: "$genres", avgRuntime: { $avg: "$runtime" } } },
  { $sort: { avgRuntime: -1 } }
])
```

```
> db.movies.aggregate([
    { $unwind: "$genres" },
    { $group: { _id: "$genres", avgRuntime: { $avg: "$runtime" } } },
    { $sort: { avgRuntime: -1 } }
  ])
< {
    _id: 'History',
    avgRuntime: 129.19392523364485
  }
  {
    _id: 'War',
    avgRuntime: 117.21500721500722
  }
  {
    _id: 'Biography',
    avgRuntime: 114.31020733652312
  }
  {
    _id: 'Western',
    avgRuntime: 112.875
  }
  {
    _id: 'Musical',
    avgRuntime: 111.65967365967366
  }
  {
    _id: 'Action',
    avgRuntime: 110.48759623609924
```

```
  {
    _id: 'Sport',
    avgRuntime: 109.15934065934066
  }
  {
    _id: 'Mystery',
    avgRuntime: 108.95733333333334
  }
  {
    _id: 'Romance',
    avgRuntime: 108.94769042520649
  }
  {
    _id: 'Drama',
    avgRuntime: 108.82702392658145
  }
  {
    _id: 'Crime',
    avgRuntime: 108.21696171263895
  }
  {
    _id: 'Adventure',
    avgRuntime: 107.50053191489361
  }
```

```
{
    _id: 'Thriller',
    avgRuntime: 105.02031509121062
}
{
    _id: 'Music',
    avgRuntime: 104.35870967741936
}
{
    _id: 'Fantasy',
    avgRuntime: 104.15325670498085
}
{
    _id: 'Sci-Fi',
    avgRuntime: 102.83298097251586
}
{
    _id: 'Comedy',
    avgRuntime: 99.02730109204369
}
{
    _id: 'Horror',
    avgRuntime: 96.21247429746401
}
{
```

```
{
    _id: 'Talk-Show',
    avgRuntime: 95
}
{
    _id: 'Film-Noir',
    avgRuntime: 94.81818181818181
}
Type "it" for more
Atlas atlas-h4o7hc-shard-0 [primary] sample_mflix>
```

1️⃣1️⃣ Calculate the Correlation Between IMDB Rating and Number of Awards

(Understand the relationship between critical acclaim and audience ratings)

```
db.movies.aggregate([
  { $match: { "imdb.rating": { $exists: true }, "awards.wins": {
$exists: true } } },
  { $group: {
      _id: null,
      avgRating: { $avg: "$imdb.rating" },
      avgAwards: { $avg: "$awards.wins" },
      ratingStdDev: { $stdDevPop: "$imdb.rating" },
      awardsStdDev: { $stdDevPop: "$awards.wins" },
      covariance: {
        $avg: {
```

```
              $multiply: [
                { $subtract: ["$imdb.rating", { $avg: "$imdb.rating" }] },
                { $subtract: ["$awards.wins", { $avg: "$awards.wins" }] }
              ]
            }
          }
        }
      },
      { $project: {
          _id: 0,
          correlation: {
            $divide: [
              "$covariance",
              { $multiply: ["$ratingStdDev", "$awardsStdDev"] }
            ]
          }
        }
      }
    ])
```

```
> db.movies.aggregate([
    { $match: { "imdb.rating": { $exists: true }, "awards.wins": { $exists: true } } },
    { $group: {
        _id: null,
        avgRating: { $avg: "$imdb.rating" },
        avgAwards: { $avg: "$awards.wins" },
        ratingStdDev: { $stdDevPop: "$imdb.rating" },
        awardsStdDev: { $stdDevPop: "$awards.wins" },
        covariance: {
          $avg: {
            $multiply: [
              { $subtract: ["$imdb.rating", { $avg: "$imdb.rating" }] },
              { $subtract: ["$awards.wins", { $avg: "$awards.wins" }] }
            ]
          }
        }
      }
    },
    { $project: {
        _id: 0,
        correlation: {
          $divide: [
            "$covariance",
            { $multiply: ["$ratingStdDev", "$awardsStdDev"] }
          ]
        }
      }
    }
```

```
              "$covariance",
              { $multiply: ["$ratingStdDev", "$awardsStdDev"] }
          ]
        }
      }
    }
  }
])
{
    correlation: 0
}
```