MongoDB Aggregation Examples

• Ensure you have pymongo installed before running cells in this notebook

```
import pymongo
from bson.json_util import dumps
import pprint

# --> Update the URI with your username and password <--
uri = "mongodb://sarahh:12345@localhost:27017"
client = pymongo.MongoClient(uri)
mflixdb = client.mflix
demodb = client.demodb</pre>
```

About Aggregates in PyMongo

- Aggregation uses pipelines.
- A **pipeline** is a sequence of stages through which documents proceed.
- Some of the different stages that can be used are:
 - match
 - project
 - sort
 - limit
 - unwind
 - group
 - lookup

\$match

```
"votes": 1759,
            "id": 139738
        },
        "type": "movie"
    },
        " id": {
            "$oid": "573a13a5f29313caabd13572"
        "plot": "Cartoon figures announce, via comic strip balloons,
that they will move - and move they do, in a wildly exaggerated
style.",

"genres": [
            "Animation",
            "Short",
            "Comedy"
        "runtime": 7,
        "cast": [
            "Winsor McCav"
        ],
        "poster":
"https://m.media-amazon.com/images/M/MV5BYzg2NjNhNTctMjUxMi00ZWU4LWI3Z
jYtNTIONTQxNThjZTk2XkEyXkFqcGdeQXVyNzg50Tk20A@@. V1 SY1000 SX677 AL .j
pg",
        "title": "Winsor McCay, the Famous Cartoonist of the N.Y.
Herald and His Moving Comics",
        "fullplot": "Cartoonist Winsor McCay agrees to create a large
set of drawings that will be photographed and made into a motion
picture. The job requires plenty of drawing supplies, and the
cartoonist must also overcome some mishaps caused by an assistant.
Finally, the work is done, and everyone can see the resulting animated
picture.",
        "languages": [
            "Enalish"
        "released": {
            "$date": {
                "$numberLong": "-1853539200000"
            }
        },
        "directors": [
            "Winsor McCay",
            "J. Stuart Blackton"
        "writers": [
            "Winsor McCay (comic strip \"Little Nemo in
Slumberland\")",
            "Winsor McCay (screenplay)"
```

```
"awards": {
            "wins": 1,
            "nominations": 0,
            "text": "1 win."
        "lastupdated": "2015-08-31 01:05:38.577000000",
        "year": 1911,
        "imdb": {
            "rating": 7.3,
            "votes": 1036,
            "id": 1737
        },
        "countries": [
            "USA"
        "type": "movie",
        "num_mflix_comments": 0
  }
]
```

match and project

```
c = mflixdb.movies.aggregate([
    {"$match": {"year": {"$lte": 1920}}},
    {"$project": {"_id":0, "title": 1, "cast": 1}},
])
print(dumps(c, indent=4))
[
    {
        "cast": [
            "A.C. Abadie",
            "Gilbert M. 'Broncho Billy' Anderson",
            "George Barnes",
            "Justus D. Barnes"
        "title": "The Great Train Robbery"
    },
        "cast": [
            "Frank Powell",
            "Grace Henderson",
            "James Kirkwood",
            "Linda Arvidson"
        "title": "A Corner in Wheat"
    },
{
```

```
"cast": [
            "Winsor McCay"
        "title": "Winsor McCay, the Famous Cartoonist of the N.Y.
Herald and His Moving Comics"
    },
    {
        "cast": [
             "Jane Gail",
            "Ethel Grandin",
            "William H. Turner",
            "Matt Moore"
        ],
"title": "Traffic in Souls"
    },
        "cast": [
            "Winsor McCay",
            "George McManus",
            "Roy L. McCardell"
        ],
"title": "Gertie the Dinosaur"
    },
{
        "cast": [
            "Stanley Hunt",
            "Sarah Constance Smith Hunt",
            "Mrs. George Walkus",
            "Paddy 'Malid"
        ],
"title": "In the Land of the Head Hunters"
    },
{
        "cast": [
            "Pearl White",
            "Crane Wilbur",
            "Paul Panzer",
            "Edward Jos\u00e8"
        "title": "The Perils of Pauline"
    },
        "cast": [
            "George Beban",
            "Clara Williams",
            "J. Frank Burke",
            "Leo Willis"
        "title": "The Italian"
```

```
},
{
    "cast": [
         "John McCann",
        "James A. Marcus",
        "Maggie Weston",
        "Harry McCoy"
    ],
"title": "Regeneration"
},
{
    "cast": [
         "Howard C. Hickman",
         "Enid Markey",
        "Lola May",
        "Kate Bruce"
    ],
"title": "Civilization"
},
{
    "cast": [
         "Tyrone Power Sr.",
         "Helen Riaume",
         "Marie Walcamp",
        "Cora Drew"
    ],
"title": "Where Are My Children?"
},
{
    "cast": [
         "Mary Pickford",
        "Madlaine Traverse",
        "Charles Wellesley",
        "Gladys Fairbanks"
    ],
"title": "The Poor Little Rich Girl"
},
{
    "cast": [
        "Douglas Fairbanks",
        "Eileen Percy",
         "Calvert Carter"
         "Charles Stevens"
    ],
"title": "Wild and Woolly"
},
{
    "cast": [
         "Tula Belle",
```

```
"Robin Macdougall",
        "Edwin E. Reed",
         "Emma Lowry"
    "title": "The Blue Bird"
},
{
    "cast": [
         "Harold Lloyd",
        "Mildred Davis",
        "'Snub' Pollard",
        "Peggy Cartwright"
    ],
"title": "From Hand to Mouth"
},
{
    "cast": [
         "Harold Lloyd",
         "Roy Brooks",
        "Mildred Davis",
        "Wallace Howe"
    ],
"title": "High and Dizzy"
},
{
    "cast": [
        "Buster Keaton",
        "Sybil Seely"
    ],
"title": "One Week"
},
{
    "cast": [
        "Edward Jobson",
        "Beulah Booker",
        "Edward Connelly",
        "Edward Alexander"
    ],
"title": "The Saphead"
},
    "title": "The Kiss",
    "cast": [
        "May Irwin",
         "John C. Rice"
    ]
},
{
    "cast": [
```

```
"Em-koy-e-tie",
            "Hunting Horse",
            "Esther LeBarre",
            "Wanada Parker"
        "title": "The Daughter of Dawn"
    },
        "title": "The Kiss",
        "cast": [
            "May Irwin",
            "John C. Rice"
        1
    },
        "cast": [
            "Winsor McCay"
        "title": "Winsor McCay, the Famous Cartoonist of the N.Y.
Herald and His Moving Comics"
  }
]
```

match project limit and sort

```
c = mflixdb.movies.aggregate([
    {"$match": {"year": {"$lte": 1920}}},
    {"$sort": {"title": 1}},
    {"$limit": 5},
    {"$project": {"_id":0, "title": 1, "cast": 1}},
])
print(dumps(c, indent=4))
[
    {
        "cast": [
            "Frank Powell",
            "Grace Henderson",
            "James Kirkwood",
            "Linda Arvidson"
        "title": "A Corner in Wheat"
    },
        "cast": [
            "Howard C. Hickman",
            "Enid Markey",
            "Lola May",
            "Kate Bruce"
```

```
],
"title": "Civilization"
    },
        "cast": [
            "Harold Lloyd",
            "Mildred Davis",
            "'Snub' Pollard",
            "Peggy Cartwright"
        ],
        "title": "From Hand to Mouth"
    },
{
        "cast": [
            "Winsor McCay",
            "George McManus",
            "Roy L. McCardell"
        "title": "Gertie the Dinosaur"
    },
        "cast": [
             "Harold Lloyd",
            "Roy Brooks",
            "Mildred Davis",
            "Wallace Howe"
        "title": "High and Dizzy"
  }
]
```

Unwind

```
{
    "cast": "Grace Henderson",
    "title": "A Corner in Wheat",
    "rating": 6.6
},
    "cast": "James Kirkwood",
    "title": "A Corner in Wheat",
    "rating": 6.6
},
    "cast": "Linda Arvidson",
    "title": "A Corner in Wheat",
    "rating": 6.6
},
    "cast": "Howard C. Hickman",
    "title": "Civilization",
    "rating": 6.3
},
    "cast": "Enid Markey",
    "title": "Civilization",
    "rating": 6.3
},
{
    "cast": "Lola May",
    "title": "Civilization",
    "rating": 6.3
},
    "cast": "Kate Bruce",
    "title": "Civilization",
    "rating": 6.3
},
    "cast": "Harold Lloyd",
    "title": "From Hand to Mouth",
    "rating": 7
},
    "cast": "Mildred Davis",
    "title": "From Hand to Mouth",
    "rating": 7
},
{
    "cast": "'Snub' Pollard",
    "title": "From Hand to Mouth",
    "rating": 7
```

```
},
        "cast": "Peggy Cartwright",
        "title": "From Hand to Mouth",
        "rating": 7
    },
        "cast": "Winsor McCay",
        "title": "Gertie the Dinosaur",
        "rating": 7.3
    },
        "cast": "George McManus",
        "title": "Gertie the Dinosaur",
        "rating": 7.3
    },
        "cast": "Roy L. McCardell",
        "title": "Gertie the Dinosaur",
        "rating": 7.3
    },
        "cast": "Harold Lloyd",
        "title": "High and Dizzy",
        "rating": 7
    },
        "cast": "Roy Brooks",
        "title": "High and Dizzy",
        "rating": 7
    },
        "cast": "Mildred Davis",
        "title": "High and Dizzy",
        "rating": 7
    },
        "cast": "Wallace Howe",
        "title": "High and Dizzy",
        "rating": 7
    }
]
```

Grouping

```
# What is the average IMDB rating of all movies by year? sort the data
by year.

c = mflixdb.movies.aggregate([
```

```
{"$group": {"_id": {"release year": "$year"}, "Avg Rating":
{"$avg": "$imdb.rating"}}},
    {"$sort" : {"_id": 1}}
])
print(dumps(c, indent = 2))
  {
    "_id": {
    "release year": 1896
    "Avg Rating": 5.9
    "_id": {
    "release year": 1903
    "Avg Rating": 7.4
  },
   "_id": {
    "release year": 1909
    "Avg Rating": 6.6
  },
   " id": {
    "release year": 1911
    "Avg Rating": 7.3
  },
   " id": {
    "release year": 1913
    "Avg Rating": 6.0
  },
    "_id": {
    "release year": 1914
    "Avg Rating": 6.89999999999995
  },
    " id": {
    "release year": 1915
    "Avg Rating": 6.6
  },
```

```
{
    "_id": {
    "-~lea
  "release year": 1916
  "Avg Rating": 6.1
},
  " id": {
  "release year": 1917
  "Avg Rating": 6.9
  "_id": {
    "release year": 1918
  },
"Avg Rating": 6.6
},
  "_id": {
    "release year": 1919
  "Avg Rating": 7.0
},
  "_id": {
  "release year": 1920
  "Avg Rating": 6.9750000000000005
},
{
  "_id": {
    "release year": 1921
  "Avg Rating": 7.4
},
  "_id": {
  "release year": 1922
  "Avg Rating": 7.2
  "_id": {
    "release year": 1923
  "Avg Rating": 7.15
},
{
```

```
" id": {
   "release year": 1924
  "Avg Rating": 7.333333333333333
},
  "_id": {
    "release year": 1925
  "Avg Rating": 7.39999999999995
},
{
 "_id": {
    "release year": 1926
  "Avg Rating": 7.3500000000000005
  "_id": {
  "release year": 1927
  "Avg Rating": 7.225
},
 "_id": {
    "release year": 1928
  "Avg Rating": 7.2375
},
  " id": {
  "release year": 1929
  "Avg Rating": 7.057142857142857
},
  "_id": {
   "release year": 1930
  "Avg Rating": 6.970000000000001
},
  " id": {
  "release year": 1931
  "Avg Rating": 7.215000000000001
},
 "_id": {
```

```
"release year": 1932
 "Avg Rating": 7.377777777778
},
 "_id": {
  "release year": 1933
 "Avg Rating": 7.185
},
 " id": {
  "release year": 1934
 "Avg Rating": 7.173913043478261
},
 "_id": {
    "release year": 1935
 "Avg Rating": 7.203225806451614
},
 " id": {
  "release year": 1936
 "Avg Rating": 7.183333333333334
},
 "_id": {
    "release year": 1937
  "Avg Rating": 7.261290322580645
},
 " id": {
  "release year": 1938
 "Avg Rating": 7.286842105263157
},
 "_id": {
   "release year": 1939
 },
 "_id": {
  "release year": 1940
```

```
"Avg Rating": 7.116666666666667
},
  " id": {
   "release year": 1941
  "Avg Rating": 7.25
},
  "_id": {
   "release year": 1942
  "Avg Rating": 7.31666666666666
},
  " id": {
   "release year": 1943
  "Avg Rating": 7.2625
},
 "_id": {
    "release year": 1944
  "Avg Rating": 7.452173913043478
},
  "_id": {
  "release year": 1945
  "Avg Rating": 7.131034482758621
},
 " id": {
   "release year": 1946
  "Avg Rating": 7.397058823529412
},
  "_id": {
    "release year": 1947
  "Avg Rating": 7.2250000000000005
},
  " id": {
  "release year": 1948
  },
```

```
"Avg Rating": 7.223214285714286
},
  "_id": {
   "release year": 1949
  "Avg Rating": 7.175000000000001
},
  " id": {
  "release year": 1950
  "Avg Rating": 7.0618181818181816
},
  " id": {
   "release year": 1951
  "Avg Rating": 7.2407407407407405
},
  "_id": {
  "release year": 1952
  "Avg Rating": 7.28666666666666
},
  "_id": {
    "release year": 1953
  "Avg Rating": 7.104615384615385
},
  " id": {
  "release year": 1954
  "Avg Rating": 7.21063829787234
},
  " id": {
   "release year": 1955
  "Avg Rating": 7.240298507462687
},
{
  "_id": {
    "release year": 1956
  "Avg Rating": 7.219402985074627
```

```
" id": {
  "release year": 1957
  "Avg Rating": 7.098591549295775
  "_id": {
   "release year": 1958
  "Avg Rating": 7.1160000000000005
},
 " id": {
  "release year": 1959
  "Avg Rating": 7.1281690140845075
},
 " id": {
  "release year": 1960
  "Avg Rating": 7.205479452054795
},
{
 "_id": {
    "release year": 1961
 },
"Avg Rating": 7.116176470588235
},
  "_id": {
  "release year": 1962
  "Avg Rating": 7.167142857142857
},
 " id": {
  "release year": 1963
  "Avg Rating": 6.994202898550725
},
{
  " id": {
  "release year": 1964
  "Avg Rating": 7.21860465116279
},
```

```
" id": {
  "release year": 1965
 "Avg Rating": 7.257142857142856
},
 " id": {
  "release year": 1966
  "Avg Rating": 6.950574712643679
 "_id": {
    "release year": 1967
  "Avg Rating": 7.009876543209876
},
 " id": {
  "release year": 1968
  "Avg Rating": 7.096629213483146
},
 "_id": {
  "release year": 1969
  "Avg Rating": 7.0607476635514015
},
 "_id": {
    "release year": 1970
  "Avg Rating": 7.07916666666667
},
 " id": {
  "release year": 1971
  "Avg Rating": 7.228301886792453
},
 "_id": {
   "release year": 1972
  "Avg Rating": 7.0719008264462815
},
{
```

```
" id": {
   "release year": 1973
  "Avg Rating": 7.074107142857143
},
  "_id": {
    "release year": 1974
  "Avg Rating": 7.218446601941747
},
{
 "_id": {
    "release year": 1975
  "Avg Rating": 7.043925233644861
  "_id": {
  "release year": 1976
  "Avg Rating": 6.931896551724138
},
 "_id": {
    "release year": 1977
  "Avg Rating": 7.05655737704918
},
{
  "_id": {
  "release year": 1978
  "Avg Rating": 6.775
},
  "_id": {
   "release year": 1979
  "Avg Rating": 6.901526717557252
},
  " id": {
  "release year": 1980
  "Avg Rating": 6.638922155688623
},
 " id": {
```

```
"release year": 1981
  "Avg Rating": 6.7005952380952385
},
 "_id": {
  "release year": 1982
  "Avg Rating": 6.777401129943502
},
 " id": {
  "release year": 1983
  "Avg Rating": 6.774534161490683
},
 "_id": {
    "release year": 1984
  "Avg Rating": 6.636683417085427
},
  " id": {
  "release year": 1985
  "Avg Rating": 6.722751322751322
},
 "_id": {
    "release year": 1986
  "Avg Rating": 6.617894736842106
},
 " id": {
  "release year": 1987
  "Avg Rating": 6.636036036036036
},
 " id": {
   "release year": 1988
  "Avg Rating": 6.511553784860558
},
 "_id": {
  "release year": 1989
```

```
"Avg Rating": 6.679741379310345
},
 " id": {
  "release year": 1990
  "Avg Rating": 6.652
},
 "_id": {
   "release year": 1991
  "Avg Rating": 6.610084033613446
},
 " id": {
  "release year": 1992
  "Avg Rating": 6.697777777777785
},
 "_id": {
    "release year": 1993
  "Avg Rating": 6.608394160583941
},
  "_id": {
  "release year": 1994
  "Avg Rating": 6.62327868852459
},
 " id": {
   "release year": 1995
  "Avg Rating": 6.600806451612903
},
 "_id": {
    "release year": 1996
  "Avg Rating": 6.569041769041768
},
  " id": {
  "release year": 1997
  "Avg Rating": 6.6029612756264235
```

```
" id": {
  "release year": 1998
  "Avg Rating": 6.638791423001949
 "_id": {
   "release year": 1999
  "Avg Rating": 6.620582524271844
},
 "_id": {
  "release year": 2000
  "Avg Rating": 6.518588640275388
},
 " id": {
  "release year": 2001
  "Avg Rating": 6.5663398692810455
},
{
 "_id": {
    "release year": 2002
 },
"Avg Rating": 6.547749196141479
},
 "_id": {
    "release year": 2003
  "Avg Rating": 6.5318407960199005
},
 " id": {
  "release year": 2004
  "Avg Rating": 6.656342182890856
},
  " id": {
  "release year": 2005
  "Avg Rating": 6.582748948106592
},
```

```
" id": {
  "release year": 2006
 "Avg Rating": 6.597286821705427
},
 " id": {
  "release year": 2007
  "Avg Rating": 6.58555555555556
 "_id": {
    "release year": 2008
  "Avg Rating": 6.552994350282486
},
 " id": {
  "release year": 2009
  "Avg Rating": 6.506543075245365
},
 "_id": {
  "release year": 2010
 "Avg Rating": 6.512832369942197
},
 "_id": {
    "release year": 2011
  "Avg Rating": 6.453415453527435
},
 " id": {
  "release year": 2012
  "Avg Rating": 6.484486373165618
},
 "_id": {
   "release year": 2013
  "Avg Rating": 6.491477787851315
},
{
```

```
" id": {
   "release year": 2014
  "Avg Rating": 6.57874531835206
},
  "_id": {
    "release year": 2015
  "Avg Rating": 6.951388888888889
},
{
  "_id": {
    "release year": 2016
  "Avg Rating": null
  "_id": {
  "release year": "1981\u00e8"
  "Avg Rating": 8.0
},
 "_id": {
    "release year": "1986\u00e8"
  "Avg Rating": 7.7
},
{
  "_id": {
  "release year": "1987\u00e8"
  "Avg Rating": 8.9
},
  "_id": {
   _"release year": "1988\u00e8"
  "Avg Rating": 7.7
},
  "_id": {
    "release year": "1994\u00e81998"
  "Avg Rating": 8.2
},
{
 "_id": {
```

```
"release year": "1995\u00e8"
  "Avg Rating": 9.0
  " id": {
  "release year": "1996\u00e8"
  "Avg Rating": 8.4
},
  "_id": {
    "release year": "1997\u00e8"
  "Avg Rating": 7.2
},
 "_id": {
    "release year": "1999\u00e8"
  "Avg Rating": 8.9
  " id": {
  "release year": "2000\u00e8"
  "Avg Rating": 7.5
},
 "_id": {
    "release year": "2002\u00e8"
  "Avg Rating": 7.5
},
  "_id": {
  "release year": "2003\u00e8"
  "Avg Rating": 8.4
},
  "_id": {
   "release year": "2005\u00e8"
  "Avg Rating": 8.4
},
 "_id": {
    "release year": "2006\u00e8"
```

```
"Avg Rating": 7.2
},
  " id": {
  "release year": "2006\u00e82007"
  "Avg Rating": 9.0
},
 "_id": {
    "release year": "2006\u00e82012"
  "Avg Rating": 8.5
},
 "_id": {
    "release year": "2007\u00e8"
  "Avg Rating": 8.5333333333333333
},
 "_id": {
    "release year": "2009\u00e8"
  "Avg Rating": 8.3
},
  "_id": {
  "release year": "2010\u00e8"
  "Avg Rating": 8.2
},
 "_id": {
   "release year": "2011\u00e8"
  "Avg Rating": 7.75
},
 "_id": {
    "release year": "2012\u00e8"
  "Avg Rating": 7.5
},
 "_id": {
    "release year": "2014\u00e8"
  "Avg Rating": 7.25
```

Lookup

Reformatting Queries

```
match = {"$match": {"year": {"$lte": 1920}}}
limit = {"$limit": 5}
project = {"$project": {"_id":0, "title": 1, "cast": 1, "rating":
"$imdb.rating"}}
agg = mflixdb.movies.aggregate([match, limit, project])
print(dumps(agg, indent=2))
```