



# NO SQL Project Format

## Individual

Name of project Member:

ADAM GREEN

Dataset Link: [https://www.kaggle.com/datasets/tmdb/tmdb-movie-metadata?rv=1&select=tmdb\\_5000\\_movies.csv](https://www.kaggle.com/datasets/tmdb/tmdb-movie-metadata?rv=1&select=tmdb_5000_movies.csv)

//Note that I did some joins and removed some columns in Tableau Prep

### 1. **AGGREGATION query:** Highest Budget movies availability in Spanish

What is the business case of the query  
To see which of the highest revenue movies are available to the spanish market

Query below

Stage1: \$match

```
{
  "revenue":{"$gt":"200000000"}, "spoken_languages.name":"Español"
}
```

Stage2: \$project

```
{
  movie_id:1,title:1,budget:1,revenue:1,spoken_languages:1
}
```

stage3: \$group

```
{
  _id: "$spoken_languages",
  "occur": {"$sum":1
}
}
```

Stage4: \$sort

```
{
  revenue: -1
}
```

### 2. **AGGREGATION query:** Highest Revenue Action Movies

What is the business case of the query  
To see which action movies have the highest revenue  
//Hitman is no.1

Query below

Stage 1: \$unwind

```
{
```

```

    path: "$genres"
  }

  Stage 2: $match
  {
    "genres.name":{"$all":["Action"]}
  }

  Stage 3: $project
  {
    title:1,revenue:1,budget:1
  }

  Stage 4: $sort
  {
    revenue: -1
  }

```

### 3. **AGGREGATION query:** Top 5 Revenue Movies

What is the business case of the query  
 To see which top 5 movies have the highest revenue  
 //Hitman is still no.1

Query below

```

Stage1: $sort
{
  revenue: -1
}

Stage2: $limit
5

Stage3: $project
{
  title:1,revenue:1
}

```

### 4. **AGGREGATION query:** Average budget by Original language of movies

What is the business case of the query  
 To see which movies typically are the most expensive to make  
 //Its not English but Telugu, English is 2nd though

Query below

```

Stage1:$group
{
  _id: "$original_language", "averageBudget": {"$avg": "$budget"}
}

Stage2:$sort
{
  averageBudget: -1
}

```

### 5. **Comparison and Logical Operators query:** Expensive movies break even

What is the business case of the query  
 The purpose of this query is to look at the more expensive movies made and to see which ones break even.  
 //can change gt to lt to see which movies flopped

Query below

```
{
  "$and": [
    {"budget": {"$gt": 200000000}},
    {"revenue": {"$gt": "$budget"}}
  ]
}
```

## 6. Comparison and Logical Operators query: Data Exploration (See if revenue prediction is possible)

What is the business case of the query

The purpose is to see if the dataset has movies which are planned to be released i.e "Unreleased" but we see there's only 8 movies which are classed as "Post Production" or "Rumored".

If we had info for budget of movies and not revenue we could potentially predict the revenue success

Query below

```
{
  "status": {"$not": {"$eq": "Released"}}
}
```

## 7. \$expr query: Revenue exceeds Budget

What is the business case of the query

To see which movies made money. Could be useful to producers

Query below

```
{
  "$expr": { "$gt": ["$revenue", "$budget"] }
}
```

## 8. \$expr query: Above Average Movies

What is the business case of the query

To be able to see which movies are considered to be above average.  
Could be useful to DVD stores or streaming websites like Netflix.

Query below

```
{
  "$expr": {
    "$gt": ["$vote_average", { "$avg": "$vote_average" }]
  }
}
```

## 9. Arrays, sub documents and similar: James Cameron Movies

What is the business case of the query

To see which movies James Cameron directed

Query below

```
$filter
{
  "crew": {
    "$elemMatch": {
      "job": "Director",
      "name": "James Cameron"
    }
  }
}
```

```
$project
{title:1}
```

#### 10. Arrays, sub documents and similar: 20th Century Fox Movies

What is the business case of the query  
To see which movies were made by 20th Century Fox, can be used to calculate their profit per year based on movies

Query below

```
{"production_companies.name":"Twentieth Century Fox Film Corporation"}
```

#### 11. Arrays, sub documents and similar: Spy Movies

What is the business case of the query  
To find movies created in 2015 which contain the keyword spy. Spy is not a category but a popular genre

Query below

```
{
  "release_date": {"$regex": "2015"}, "keywords.name":"spy"
}
```

#### 12. Arrays, sub documents and similar: South African Movies

What is the business case of the query  
To see which movies had filming in my home country  
//Dredd was partly filmed in SA

Query below

```
{"production_countries.name":{"$eq":"South Africa"}}
```

//Remember that there's already array usage prior to the array queries.

#### Link for the Chart Dashboard

Link: (Keyword Word Chart)  
<https://charts.mongodb.com/charts-project-0-segeo/embed/charts?id=65675a05-e111-4b54-8cb1-6a31fcfb44a9&maxDataAge=60&theme=light&autoRefres>

Link: (Number of Movies Made at Given Budget)  
//turn on filter to take closer look at more expensive movies OR view next link  
<https://charts.mongodb.com/charts-project-0-segeo/embed/charts?id=65676b88-a121-4f1f-8029-9ffab6dc290c&maxDataAge=3600&theme=light&autoRefr>  
<https://charts.mongodb.com/charts-project-0-segeo/embed/charts?id=ddb9d149-8e9e-489f-bf9d-a08f785cc1dd&maxDataAge=3600&theme=light&autoRefr>

Link:(Highest Revenue Movies)  
<https://charts.mongodb.com/charts-project-0-segeo/embed/charts?id=65677517-12c2-4f6e-8e0f-2ce07b4c5a06&maxDataAge=3600&theme=light&autoRefr>

Link:(Highest Budget Movies)  
<https://charts.mongodb.com/charts-project-0-segeo/embed/charts?id=1bc6641f-3076-4215-a7dd-bee0d2b939b5&maxDataAge=3600&theme=light&autoRefr>