



POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Systems and Methods for Big and Unstructured Data Project

Author(s): **Alessandro Cecchetto 10865806**

Christian Ferrareis 10725804

Group Number:

Academic Year: 2023-2024

Contents

Contents	i
1 Introduction	1
2 Data wrangling	3
3 Dataset	5
4 Queries	9

1 | Introduction

We are going to analyze the TMDB dataset containing the information about movies published since the creation of cinematography. In this dataset we have a lot of numerical information, as the budget and the revenues corresponding to each movie, as well as text field as the title and the overview, where a lot of information is hidden within patterns. Lastly, attributes like the genre and the production companies give us the opportunity of searching for interesting insights grouping by those fields. We decided to use Elasticsearch to conduct this analysis because it is perfect for in text search, allowing to search for insight and filter for example on title sagas, as well as it is powerful enough to conduct aggregations over specific keyword and extract information coming from set of data, as average revenues per film genre or total revenues from different production companies.

2 | Data wrangling

In order to be able to upload the dataset in the ElasticSearch engine, we have to reduce the dimension of the file under 100 Mb and be sure of the quality of our data as well as keep only the valuable information needed for our analysis. We used pandas to prepare our dataset. In a Colab Notebook we have extract information on the dataset such as the number of entries, the number and the name of the attributes, as well as how many missing data were present. First of all we have converted the data in the right datatype, such as the "release_date" field was converted from string to date. We then tried to add the missing values for the titles, recovering them, where possible, from the information present in the poster of the movie. We then decided to keep only the "title" field dropping the "original title" field to make easier the analysis. We then extracted the "language" property by merging the information present in the "original_language" field with a conversion chart, assigning manually the leftover values and dropping the unknown one. We finally decided to drop the adult movies from our analysis, as well as the one who didn't present the release date, didn't had information about production country and company and where genres were not indicated. We eventually kept only the "Realised" movies as well as we considered only the one with at least one vote and presented a runtime. At this point we have dropped the information we didn't needed dropping the relative attributes and finally we exported our dataframe into a csv ready to be analysed.

All the operations are included in the file "TMDBmovie.ipynb".

3 | Dataset

The TMDB dataset contains data collected from the TMDB API, which provides a comprehensive movie database. In order to properly use the dataset for our analysis, we decided to perform data wrangling on it, as described in the previous chapter.

The data schema employed for our analysis is detailed below. It includes all the essential attributes of each entity, their data types, and relationships:

Movie:

- id (integer): unique identifier for each movie.
- title (text): name of the movie.
- vote_average (float): average user rating for the movie.
- vote_count (integer): number of votes received by the movie.
- release_date (date): the release date of the movie.
- revenue (integer): revenue generated by the movie in USD.
- runtime (integer): the runtime of the movie in minutes.
- budget (integer): the budget of the movie in USD.
- overview (text): a brief description of the movie.
- popularity (float): a metric indicating the movie's popularity.
- tagline (text): a short, catchy phrase that summarizes the movie.
- genres (list of text): list of genres associated with the movie, which implements the relationship movie-genre.
- production_companies (list of text): list of production companies involved in the movie, which represents the relationship movie-production_company
- language (text): The primary spoken language(s) in the movie.

The entire dataset is stored in a single Elasticsearch index, named `tmdb_movies`, where movie documents are stored in an inverted index structure.

The mapping detailed below precisely defines the storage and indexing structure for each movie document and its associated fields:

```
{
  "properties": {
    "@timestamp": {
      "type": "date"
    },
    "budget": {
      "type": "long"
    },
    "genres": {
      "type": "text",
      "field_data": true
    },
    "id": {
      "type": "long"
    },
    "language": {
      "type": "keyword"
    },
    "overview": {
      "type": "text"
    },
    "popularity": {
      "type": "float"
    },
    "production_companies": {
      "type": "text"
    },
    "release_date": {
      "type": "date",
      "format": "iso8601"
    },
    "revenue": {
      "type": "long"
    },
    "runtime": {
      "type": "long"
    },
    "tagline": {
      "type": "text"
    },
    "title": {
      "type": "text"
    },
    "vote_average": {
      "type": "float"
    },
    "vote_count": {
      "type": "long"
    }
  }
}
```

It is important to note that in the configuration of the "genres" field mapping, the "field-data" attribute has been set to true. This adjustment is important given the nature of Elasticsearch's handling of text fields. By default, text fields are optimized for searching, but they are not available for other operations like aggregations, which we will be doing in our analysis.

We decided to define the fields "title", "genres", "overview", "production_companies" and "tagline" as "text" in our Elasticsearch schema so as to enable its full-text search functionalities and to make search more flexible.

4 | Queries

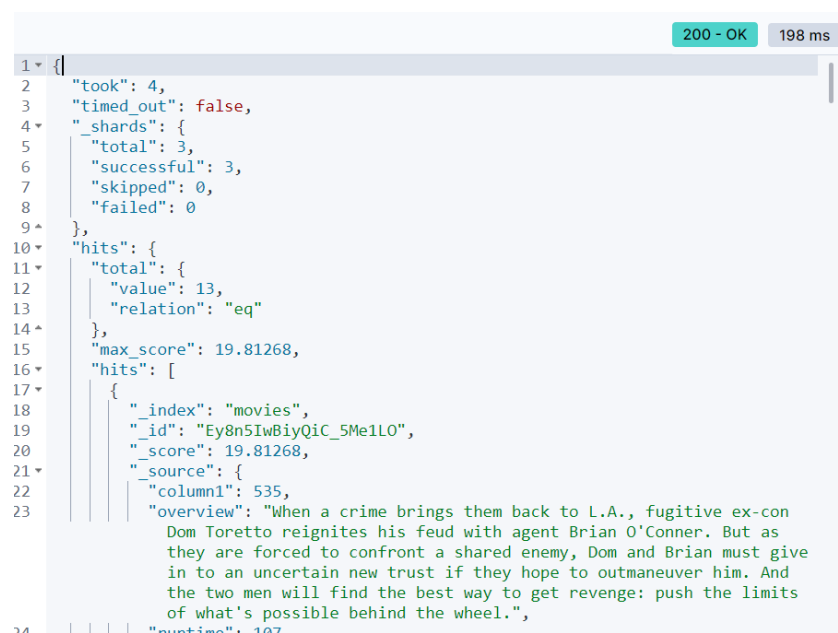
1. Fast & Furious

Search for movies titled "Fast & Furious", presenting the three word in the title:

GET /tmdb_movies/_search

```
{
  "query": {
    "match": {
      "title": {
        "query": "Fast & Furious",
        "operator": "and"
      }
    }
  }
}
```

(Partial) Query response:



```
1 {
2   "took": 4,
3   "timed_out": false,
4   "shards": {
5     "total": 3,
6     "successful": 3,
7     "skipped": 0,
8     "failed": 0
9   },
10  "hits": {
11    "total": {
12      "value": 13,
13      "relation": "eq"
14    },
15    "max_score": 19.81268,
16    "hits": [
17      {
18        "_index": "movies",
19        "_id": "Ey8n5IwBiyQic_5Me1LO",
20        "_score": 19.81268,
21        "_source": {
22          "column1": 535,
23          "overview": "When a crime brings them back to L.A., fugitive ex-con Dom Toretto reignites his feud with agent Brian O'Conner. But as they are forced to confront a shared enemy, Dom and Brian must give in to an uncertain new trust if they hope to outmaneuver him. And the two men will find the best way to get revenge: push the limits of what's possible behind the wheel.",
24          "runtime": 107
25        }
26      }
27    ]
28  }
29 }
```

(a) Partial query response.

2. Interstellar

Search for movies titled "Interstellar":

GET /tmdb_movies/_search

```
{
  "query": {
    "match": {
      "title": "Interstellar"
    }
  }
}
```

(Partial) Query response:



(b) Partial query response.

3. Comedies not about love

Search for comedy movies without "love" in the overview:

GET /tmdb_movies/_search

```
{
  "query": {
    "bool": {
      "must": [
```

```

        {"match": {"genres": "Comedy"}}
    ],
    "must_not": [
        {"match": {
            "overview": "love"
        }}
    ]
}
}
}

```

(Partial) Query response:

```

1 {
2   "took": 5,
3   "timed_out": false,
4   "_shards": {
5     "total": 3,
6     "successful": 3,
7     "skipped": 0,
8     "failed": 0
9   },
10  "hits": {
11    "total": {
12      "value": 10000,
13      "relation": "gte"
14    },
15    "max_score": 1.7581547,
16    "hits": [
17      {
18        "_index": "tmdb_movies",
19        "_id": "AzOd5YwB-336sOGYEGNi",
20        "_score": 1.7581547,
21        "_source": {
22          "overview": ""Def Comedy Jam played host to some of the greatest comedians of the 1990s, many of whom went on to greater fame, due in part to the exposure they received on the program. Here, Mark Curry and Steve Harvey both show off some comic chops that that could never air on their respective network TV shows, though Harvey has shown this side of himself in Spike Lee's excellent THE ORIGINAL KINGS OF COMEDY. Rounding out the performers are Don "D.C. " Curry (NEXT FRIDAY) and Arnez... Contains two episodes ""

```


(c) Partial query response.

4. Runtime 90-120

Search for movies with runtime between 90 and 120 minutes:

```
GET /tmdb_movies/_search
{
  "query": {
    "range": {
      "runtime": {
        "gte": 90,
        "lte": 120
      }
    }
  }
}
```

(Partial) Query response:



```
1 {
2   "took": 2,
3   "timed_out": false,
4   "_shards": {
5     "total": 1,
6     "successful": 1,
7     "skipped": 0,
8     "failed": 0
9   },
10  "hits": {
11    "total": {
12      "value": 10000,
13      "relation": "gte"
14    },
15    "max_score": 1,
16    "hits": [
17      {
18        "_index": "tmdb_movies",
19        "_id": "vKpw5YwButadmQypFMcL",
20        "_score": 1,
21        "_source": {
22          "overview": "The agent of an absentee landlord resorts to underhand means in order to evict
23            tenants from land that could be more profitably used for cattle. Adapted from the 1879 novel of
24            the same name.",
25          "runtime": 90,
26          "language": "English",
27          "title": "Knocknagow",
28          "revenue": 0,
29          "@timestamp": "1918-01-30T00:00:00.000+01:00",

```

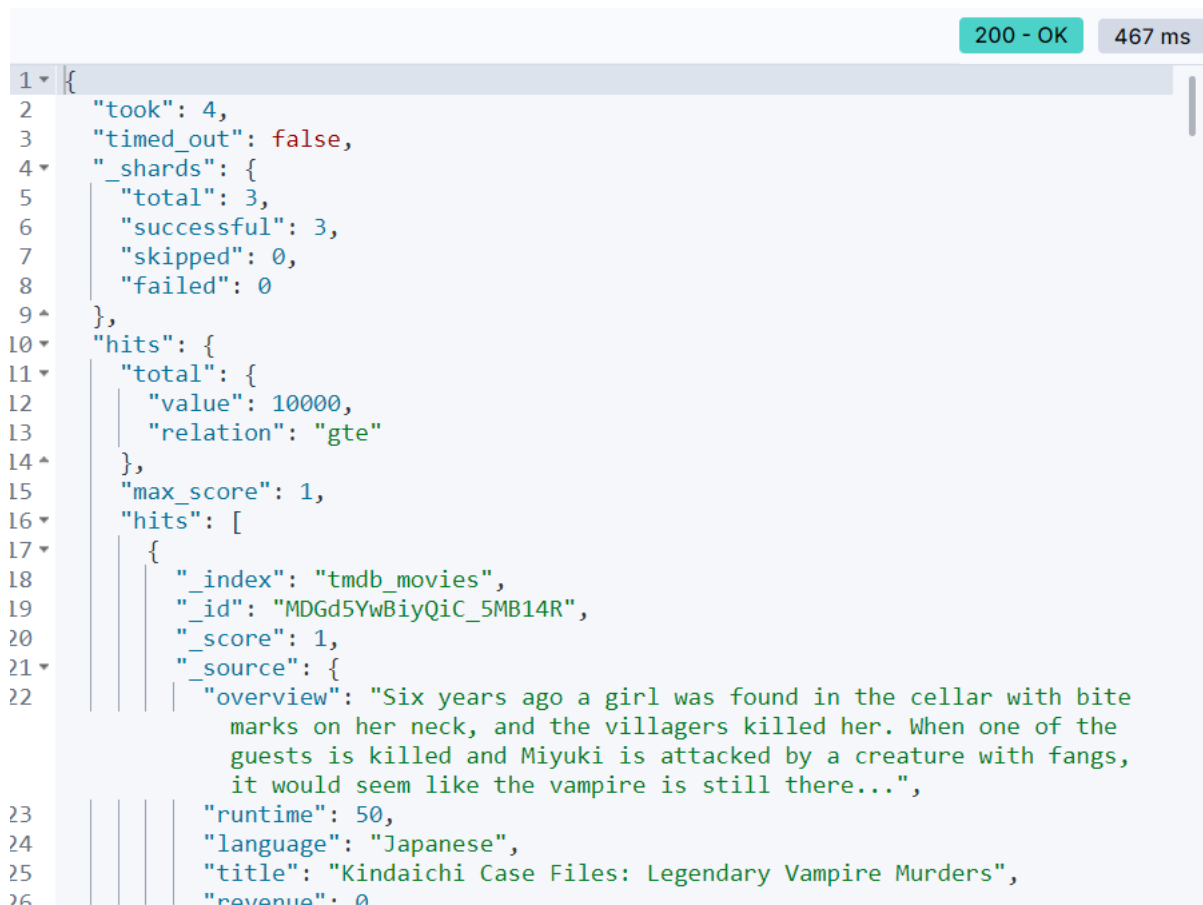
(d) Partial query response.

5. Vote higher than 8

Search for movies with rating greater than 8:

```
GET /tmdb_movies/_search
{
  "query": {
    "range": {
      "vote_average": {
        "gte": 8
      }
    }
  }
}
```

(Partial) Query response:



```
1 {
2   "took": 4,
3   "timed_out": false,
4   "_shards": {
5     "total": 3,
6     "successful": 3,
7     "skipped": 0,
8     "failed": 0
9   },
10  "hits": {
11    "total": {
12      "value": 10000,
13      "relation": "gte"
14    },
15    "max_score": 1,
16    "hits": [
17      {
18        "_index": "tmdb_movies",
19        "_id": "MDGd5YwBiyQiC_5MB14R",
20        "_score": 1,
21        "_source": {
22          "overview": "Six years ago a girl was found in the cellar with bite
23                      marks on her neck, and the villagers killed her. When one of the
24                      guests is killed and Miyuki is attacked by a creature with fangs,
25                      it would seem like the vampire is still there...",
26          "runtime": 50,
27          "language": "Japanese",
28          "title": "Kindaichi Case Files: Legendary Vampire Murders",
29          "revenue": 0
```

(e) Partial query response.

6. Big Budget, Low Returns

Search for movies with budget greater than 100 million dollars but revenue less than that amount.

```
GET /tmdb_movies/_search
{
  "query": {
    "bool": {
      "filter": [
        {
          "range": {
            "budget": {
              "gt": 100000000
            }
          }
        },
        {
          "range": {
            "revenue": {
              "lt": 100000000
            }
          }
        }
      ]
    }
  }
}
```

(Partial) Query response:



```

1 {
2   "took": 2,
3   "timed_out": false,
4   "shards": {
5     "total": 1,
6     "successful": 1,
7     "skipped": 0,
8     "failed": 0
9   },
10  "hits": {
11    "total": {
12      "value": 35,
13      "relation": "eq"
14    },
15    "max_score": 0,
16    "hits": [
17      {
18        "_index": "tmdb_movies",
19        "_id": "1apW5YwButadmQypJNw2",
20        "_score": 0,
21        "_source": {
22          "overview": "Dr. Sarah Chapel returns to the small town of Brighton Mills when her father Dr. Robert Chapel unexpectedly dies. His clients and friends miss him, and Sarah discloses that all of them have perfect health; they hear voices when they are sleeping; and they are addicted to the stimulant Tributol. Her further investigation discloses a dark secret about the haunting voices.",
23          "runtime": 96,
24          "language": "English",
25          "title": "Messengers",

```

(f) Partial query response.

7. Action, Comedy, Romance

Search for movies of genres "Action", "Comedy" and "Romance", giving higher score to the ones belonging to more genres

```

GET /tmdb_movies/_search
{
  "query": {
    "match": {
      "genres": {
        "query": "Action, Comedy, Romance",
        "operator": "or"
      }
    }
  }
}

```

(Partial) Query response:

```

1 {
2   "took": 6,
3   "timed_out": false,
4   "_shards": {
5     "total": 3,
6     "successful": 3,
7     "skipped": 0,
8     "failed": 0
9   },
10  "hits": {
11    "total": {
12      "value": 10000,
13      "relation": "gte"
14    },
15    "max_score": 4.6696367,
16    "hits": [
17      {
18        "_index": "tmdb_movies",
19        "_id": "EjGd5YwBiyQiC_5MB2ER",
20        "_score": 4.6696367,
21        "_source": {
22          "overview": "Ruby is a counter girl at the San Diego Soda Shop with a
23                      habit of being a girlfriend to Sailors stopping by. Things get a
24                      little zany when she sets her eyes on Bull's Eye McCoy a gunner who
25                      refuses to settle down.",
26          "runtime": 71,
27          "language": "English",
28          "title": "True to the Navy",
29          "revenue": 0

```

(g) Partial query response.

8. Alien Invasion

Search for movies released between 2010 and the current year that mention "Alien" in their title and "Invasion" in their overview:

GET /tmdb_movies/_search

```

{
  "query": {
    "bool": {
      "filter": [
        {
          "range": {
            "release_date": {
              "gte": "2010",
              "lte": "now/y"
            }
          }
        }
      ]
    }
  }
}

```

```

    }
  ],
  "must": [
    {
      "match": {
        "title": "Alien"
      }
    },
    {
      "match": {
        "overview": "Invasion"
      }
    }
  ]
}
}
}
}

```

(Partial) Query response:



```

1 {
2   "took": 2,
3   "timed_out": false,
4   "_shards": {
5     "total": 1,
6     "successful": 1,
7     "skipped": 0,
8     "failed": 0
9   },
10  "hits": {
11    "total": {
12      "value": 5,
13      "relation": "eq"
14    },
15    "max_score": 16.540644,
16    "hits": [
17      {
18        "_index": "tmdb_movies",
19        "_id": "VKIU5YwButadmQypS_Nz",
20        "_score": 16.540644,
21        "_source": {
22          "overview": "A documentary crew follows an elite unit of soldiers in the wake of an alien invasion",
23          "runtime": 93,
24          "language": "English",
25          "title": "Alien Outpost",
26          "revenue": 0,
27          "@timestamp": "2014-09-19T00:00:00.000+02:00",
28          "release_date": "2014-09-19",

```

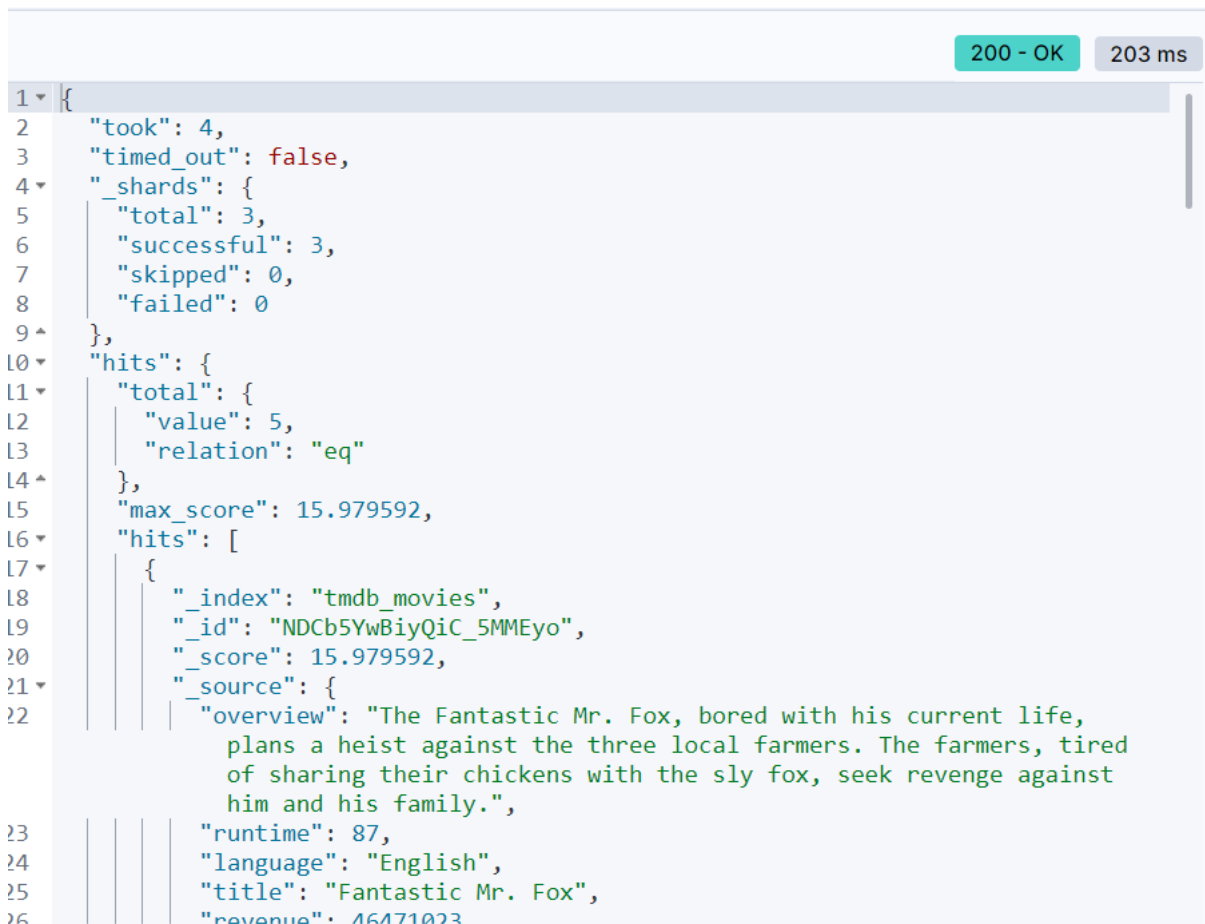
(h) Partial query response.

9. Very Fantastic

Search for movies with "fantastic" both in the overview and the tagline:

```
GET /tmdb_movies/_search
{
  "query": {
    "bool": {
      "must": [
        {"match": {"overview": "fantastic"}},
        {"match": {"tagline": "fantastic"}}
      ]
    }
  }
}
```

(Partial) Query response:



```
1 {
2   "took": 4,
3   "timed_out": false,
4   "_shards": {
5     "total": 3,
6     "successful": 3,
7     "skipped": 0,
8     "failed": 0
9   },
10  "hits": {
11    "total": {
12      "value": 5,
13      "relation": "eq"
14    },
15    "max_score": 15.979592,
16    "hits": [
17      {
18        "_index": "tmdb_movies",
19        "_id": "NDCb5YwBiyQiC_5MMEyo",
20        "_score": 15.979592,
21        "_source": {
22          "overview": "The Fantastic Mr. Fox, bored with his current life,
23                    plans a heist against the three local farmers. The farmers, tired
24                    of sharing their chickens with the sly fox, seek revenge against
25                    him and his family.",
26          "runtime": 87,
27          "language": "English",
28          "title": "Fantastic Mr. Fox",
29          "revenue": 16471023
30        }
31      }
32    ]
33  }
34 }
```

(i) Partial query response.

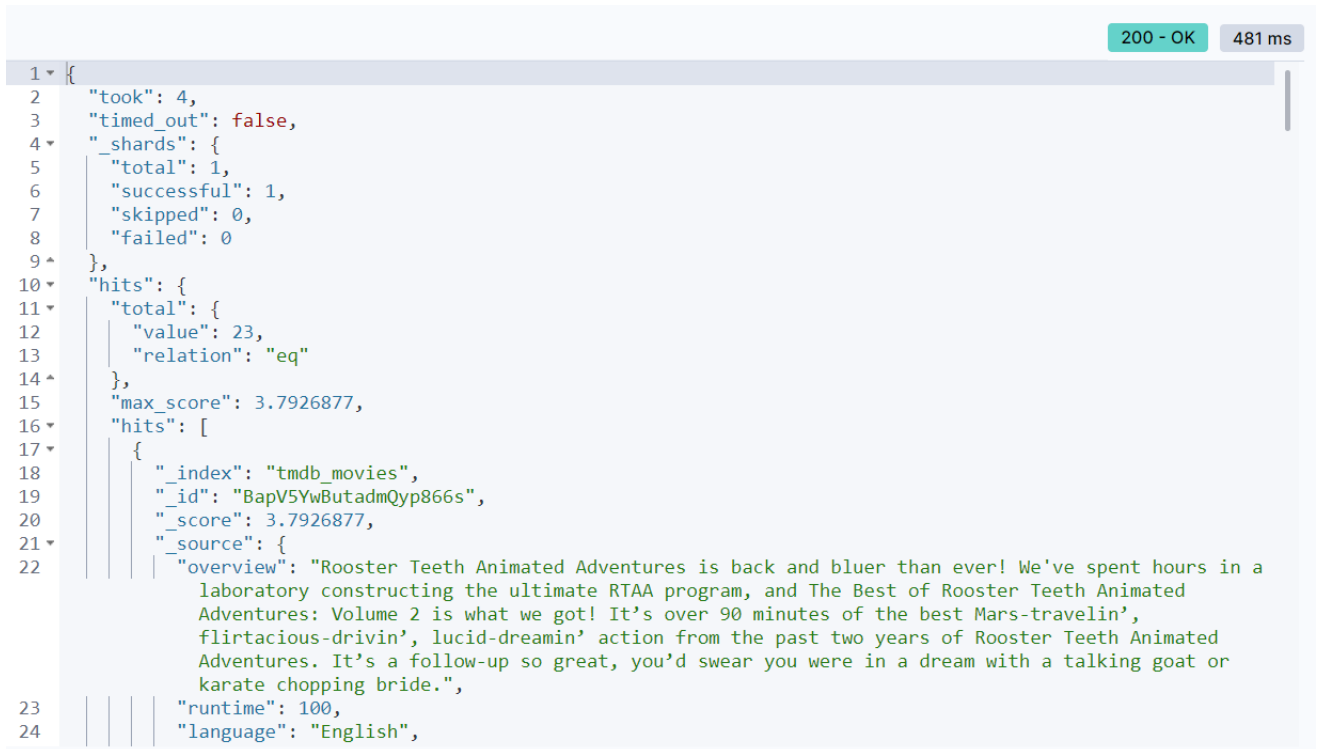
10. Action, Animation, Comedy

Search for action, animation or comedy movies produced in 2015 in English language. The language should not affect the final score:

GET /tmdb_movies/_search

```
{
  "query": {
    "bool": {
      "filter": [
        {
          "range": {
            "release_date": {
              "gte": "2015",
              "lte": "2015"
            }
          }
        },
        {
          "term": {
            "language": "English"
          }
        }
      ],
      "must": [
        {
          "match": {
            "genres": {
              "query": "Action Animation Comedy",
              "operator": "OR"
            }
          }
        }
      ]
    }
  }
}
```

(Partial) Query response:



```

1 {
2   "took": 4,
3   "timed_out": false,
4   "_shards": {
5     "total": 1,
6     "successful": 1,
7     "skipped": 0,
8     "failed": 0
9   },
10  "hits": {
11    "total": {
12      "value": 23,
13      "relation": "eq"
14    },
15    "max_score": 3.7926877,
16    "hits": [
17      {
18        "_index": "tmdb_movies",
19        "_id": "BapV5YwButadmQyp866s",
20        "_score": 3.7926877,
21        "_source": {
22          "overview": "Rooster Teeth Animated Adventures is back and bluer than ever! We've spent hours in a
                        laboratory constructing the ultimate RTAA program, and The Best of Rooster Teeth Animated
                        Adventures: Volume 2 is what we got! It's over 90 minutes of the best Mars-travelin',
                        flirtacious-drivin', lucid-dreamin' action from the past two years of Rooster Teeth Animated
                        Adventures. It's a follow-up so great, you'd swear you were in a dream with a talking goat or
                        karate chopping bride.",
23          "runtime": 100,
24          "language": "English",

```

(j) Partial query response.

11. Warner Bros films per genre

Count the movies published by Warner Bros based on their genres:

```

GET /tmdb_movies/_search
{
  "size": 0,
  "query": {
    "match": {
      "production_companies": {
        "query": "Warner Bros",
        "operator": "and"
      }
    }
  },
  "aggs": {
    "Warner_Bros_movies_per_genre": {
      "terms": {
        "field": "genres"
      }
    }
  }
}

```



```

    }
  }
}

```

(Partial) Query response:

```

2  "took": 19,
3  "timed_out": false,
4  "_shards": {
5    "total": 3,
6    "successful": 3,
7    "skipped": 0,
8    "failed": 0
9  },
10 "hits": {
11   "total": {
12     "value": 3506,
13     "relation": "eq"
14   },
15   "max_score": null,
16   "hits": []
17 },
18 "aggregations": {
19   "Warner_Bros_movies_per_genre": {
20     "doc_count_error_upper_bound": 0,
21     "sum_other_doc_count": 1755,
22     "buckets": [
23       {
24         "key": "comedy",
25         "doc_count": 1417
26       },
27       {
28         "key": "drama",
29         "doc_count": 1155
30     ]

```

(k) Partial query response.

12. Post 2000'

Search for movies released after 2000 with a budget less than 50 million, prioritizing those in the Drama genre or produced by Paramount:

GET /tmdb_movies/_search

```

{
  "query": {
    "bool": {
      "filter": [
        {

```

```

        "range": {
          "release_date": {
            "gte": "2000"
          }
        }
      },
      {
        "range": {
          "budget": {
            "lt": "500000000"
          }
        }
      }
    ],
    "should": [
      {
        "match": {
          "genres": "Drama"
        }
      },
      {
        "match": {
          "production_companies": "Paramount"
        }
      }
    ]
  }
}

```

(Partial) Query response:



```

1 {
2   "took": 14,
3   "timed_out": false,
4   "_shards": {
5     "total": 1,
6     "successful": 1,
7     "skipped": 0,
8     "failed": 0
9   },
10  "hits": {
11    "total": {
12      "value": 10000,
13      "relation": "gte"
14    },
15    "max_score": 7.1879907,
16    "hits": [
17      {
18        "_index": "tmdb_movies",
19        "_id": "aapU5YwButadmQyp7yT9",
20        "_score": 7.1879907,
21        "_source": {
22          "overview": "During a time of robber barons, child labor, and rising industry, two orphan girls
23            get a little help from their unexpected guardian angels who turn the tide of events and change
24            their lives forever.",
25          "runtime": 87,
26          "language": "English",
27          "title": "Buttons",
28          "revenue": 0,
29          "@timestamp": "2018-12-08T00:00:00.000+01:00",

```

(1) Partial query response.

13. Old films per genre

Return the number of movies released before 1930 for each genre:

```

GET /tmdb_movies/_search
{
  "size": 0,
  "query": {
    "range": {
      "release_date": {
        "lte": "1930-01-01"
      }
    }
  },
  "aggs": {
    "old_movies_per_genre": {
      "terms": {
        "field": "genres"
      }
    }
  }
}

```

(Partial) Query response:



```

2   "took": 3,
3   "timed_out": false,
4   "_shards": {
5     "total": 3,
6     "successful": 3,
7     "skipped": 0,
8     "failed": 0
9   },
10  "hits": {
11    "total": {
12      "value": 5380,
13      "relation": "eq"
14    },
15    "max_score": null,
16    "hits": []
17  },
18  "aggregations": {
19    "old_movies_per_genre": {
20      "doc_count_error_upper_bound": 0,
21      "sum_other_doc_count": 953,
22      "buckets": [
23        {
24          "key": "comedy",
25          "doc_count": 2131
26        },
27        {
28          "key": "drama",
29          "doc_count": 1923
30        }

```

(m) Partial query response.

14. Non-english

Average revenue for each genre of non-english movies:

GET /tmdb_movies/_search

```

{
  "size": 0,
  "aggs": {
    "non_eng_movies": {
      "filter": {
        "bool": {
          "must_not": {
            "term": {
              "language": "English"
            }
          }
        }
      }
    }
  }
}

```

```
    }
  }
},
"aggs": {
  "genres": {
    "terms": {
      "field": "genres"
    },
    "aggs": {
      "avg_revenue": {
        "avg": {
          "field": "revenue"
        }
      }
    }
  }
}
}
```

(Partial) Query response:

```

1 {
2   "took": 24,
3   "timed_out": false,
4   "_shards": {
5     "total": 1,
6     "successful": 1,
7     "skipped": 0,
8     "failed": 0
9   },
10  "hits": {
11    "total": {
12      "value": 10000,
13      "relation": "gte"
14    },
15    "max_score": null,
16    "hits": []
17  },
18  "aggregations": {
19    "non_eng_movies": {
20      "doc_count": 89812,
21      "genres": {
22        "doc_count_error_upper_bound": 0,
23        "sum_other_doc_count": 33261,
24        "buckets": [
25          {
26            "key": "drama",
27            "doc_count": 44417,
28            "avg_revenue": {
29              "value": 746046.3092959903
30            }
31          },
32          {
33            "key": "comedy",
34            "doc_count": 25806,
35            "avg_revenue": {
36              "value": 925699.1334185848
37            }
38          },
39          {
40            "key": "romance",
41            "doc_count": 11493,
42            "avg_revenue": {
43              "value": 856690.239798138
44            }
45          },
46          {
47            "key": "action",
48            "doc_count": 9356,
49            "avg_revenue": {
50              "value": 3000583.154446345
51            }
52          },
53          {
54            "key": "documentary",
55            "doc_count": 9343,

```

(n) Partial query response.

15. Cars and Family

Return movies with revenues between 50 and 100 millions preferring the ones with "cars" and "family" in the overview, excluding documentaries

```

GET /tmdb_movies/_search
{
  "query": {
    "bool": {
      "must": [

```

```
    {"range": {
      "revenue": {
        "gte": 500000000,
        "lte": 1000000000
      }
    }}
  ],
  "should": [
    {"match": {
      "overview": "cars, family"
    }}
  ],
  "must_not": [
    {"term": {"genres": "Documentary"}}
  ]
}
}
```

(Partial) Query response:



```

1 {
2   "took": 5,
3   "timed_out": false,
4   "_shards": {
5     "total": 3,
6     "successful": 3,
7     "skipped": 0,
8     "failed": 0
9   },
10  "hits": {
11    "total": {
12      "value": 1129,
13      "relation": "eq"
14    },
15    "max_score": 8.239176,
16    "hits": [
17      {
18        "_index": "tmdb_movies",
19        "_id": "ITCb5YwBiyQiC_5MOFSA",
20        "_score": 8.239176,
21        "_source": {
22          "overview": "In the 1930s, bored waitress Bonnie Parker falls in love
23            with an ex-con named Clyde Barrow and together they start a violent
24            crime spree through the country, stealing cars and robbing banks.",
25          "runtime": 111,
26          "language": "English",
27          "title": "Bonnie and Clyde",
28          "revenue": 50700000,
29          "@timestamp": "1967-08-14T00:00:00.000+02:00"
30        }
31      }
32    ]
33  }
34 }

```

(o) Partial query response.

16. Time Travel vs Love Story

The number of movies that mention "time travel" in their overview and the number of movies that mention "love story" in the same field:

GET /tmdb_movies/_search

```

{
  "size": 0,
  "aggs": {
    "mentioning_timetravel": {
      "filter": {
        "match": {
          "overview": {
            "query": "time travel",
            "operator": "and"
          }
        }
      }
    }
  }
}

```



```
GET /tmdb_movies/_search
{
  "size": 0,
  "query": {
    "bool": {
      "must": [
        {"match": {"genres": "Action"}}
      ],
      "filter": [
        {"range": {
          "release_date": {
            "gte": "1980-01-01",
            "lt": "1990-01-01"
          }
        }},
        {
          "range": {
            "budget": {
              "lt": 80000000
            }
          }
        }
      ],
      "must_not": [
        {"match": {"genres": "Comedy"}}
      ]
    }
  },
  "aggs": {
    "avg_budget": {
      "avg": {
        "field": "budget"
      }
    },
    "avg_revenue": {
      "avg": {
        "field": "revenue"
      }
    }
  }
}
```

```

    }
  }
}
}

```

(Partial) Query response:



```

1  {
2    "took": 4,
3    "timed_out": false,
4    "_shards": {
5      "total": 3,
6      "successful": 3,
7      "skipped": 0,
8      "failed": 0
9    },
10   "hits": {
11     "total": {
12       "value": 1742,
13       "relation": "eq"
14     },
15     "max_score": null,
16     "hits": []
17   },
18   "aggregations": {
19     "avg_budget": {
20       "value": 1384168.1360505167
21     },
22     "avg_revenue": {
23       "value": 5220467.83065442
24     }
25   }
26 }

```

(q) Partial query response.

18. Top hits

Count the number of movies for each genre having popularity greater than 70:

```

GET /tmdb_movies/_search
{
  "size": 0,
  "aggs": {
    "popular_films": {
      "filter": {
        "range": {
          "popularity": {
            "gt": 70

```

```

    }
  }
},
"aggs": {
  "genre": {
    "terms": {
      "field": "genres"
    }
  }
}
}
}
}

```

(Partial) Query response:

```

1 ▾ {
2   "took": 9,
3   "timed_out": false,
4 ▾  "shards": {
5     "total": 1,
6     "successful": 1,
7     "skipped": 0,
8     "failed": 0
9 ▸  },
10 ▾ "hits": {
11 ▾  "total": {
12     "value": 10000,
13     "relation": "gte"
14 ▸  },
15   "max_score": null,
16   "hits": []
17 ▸ },
18 ▾ "aggregations": {
19 ▾  "mentioning_timetravel": {
20     "doc_count": 240
21 ▸  },
22 ▾  "mentioning_lovestory": {
23     "doc_count": 2421
24 ▸  }
25 ▸ }
26 ▸ }

```

(r) Partial query response.

19. Marvel's revenues

Retrieve the sum coming from the revenues of action movies made by Marvel divided in 5 categories based on the average vote, considering the one with popularity greater than 60 and published since the '80s, considering only the one that received at least

100 votes:

```
GET /tmdb_movies/_search
{
  "size": 0,
  "query": {
    "bool": {
      "must": [
        {"range": {
          "release_date": {
            "gte": "1980-01-01"
          }
        }},
        {"range": {
          "popularity": {
            "gte": 60
          }
        }},
        {"range": {
          "vote_count": {
            "gte": 10
          }
        }},
        {"match": {"genres": "Action"}},
        {"match": {"production_companies": "Marvel"}}
      ]
    }
  },
  "aggs": {
    "rating_range": {
      "range": {
        "field": "vote_average",
        "ranges": [
          {"from": 5, "to": 6},
          {"from": 6, "to": 6.5},
          {"from": 6.5, "to": 7},
          {"from": 7, "to": 8},
        ]
      }
    }
  }
}
```

```
        {"from": 8, "to": 9},
        {"from": 9, "to": 10}
    ]
},
"aggs": {
  "revenues_per_avg_valuation": {
    "sum": {
      "field": "revenue"
    }
  }
}
}
```

(Partial) Query response:

```

1  {
2    "took": 4,
3    "timed_out": false,
4    "_shards": {
5      "total": 3,
6      "successful": 3,
7      "skipped": 0,
8      "failed": 0
9    },
10   "hits": {
11     "total": {
12       "value": 24,
13       "relation": "eq"
14     },
15     "max_score": null,
16     "hits": []
17   },
18   "aggregations": {
19     "rating_range": {
20       "buckets": [
21         {
22           "key": "5.0-6.0",
23           "from": 5,
24           "to": 6,
25           "doc_count": 0,
26           "revenues_per_avg_valuation": {
27             "value": 0
28           }
29         },
30         {
31           "key": "6.0-6.5",
32           "from": 6,
33           "to": 6.5,
34           "doc_count": 3,
35           "revenues_per_avg_valuation": {
36             "value": 2080016876
37           }
38         },
39         {
40           "key": "6.5-7.0",
41           "from": 6.5,
42           "to": 7,
43           "doc_count": 6,
44           "revenues_per_avg_valuation": {
45             "value": 4267297771
46           }
47         },
48         {
49           "key": "7.0-8.0",
50           "from": 7,
51           "to": 8,
52           "doc_count": 12,
53           "revenues_per_avg_valuation": {
54             "value": 12075382908
55           }
56         },
57         {
58           "key": "8.0-9.0",
59           "from": 8,
60           "to": 9,
61           "doc_count": 3,
62           "revenues_per_avg_valuation": {
63             "value": 5697845341
64           }
65         },
66         {
67           "key": "9.0-10.0",
68           "from": 9,
69           "to": 10,
70           "doc_count": 0,
71           "revenues_per_avg_valuation": {
72             "value": 0
73           }
74         }
75       ]
76     }
77   }
78 }

```

(s) Partial query response.

20. Popularity Profit by Runtime

Explore the relationship between a movie's runtime and its popularity and revenue: for each runtime category (low, medium high: -90, 90-120, 120+), calculate the average popularity and revenue:

```
GET /tmdb_movies/_search
{
  "size": 0,
  "aggs":{
    "runtime_ranges": {
      "range": {
        "field": "runtime",
        "ranges": [
          { "key": "low", "to": 90 },
          { "key": "medium", "from": 91, "to": 120 },
          { "key": "high", "from": 121 }
        ]
      },
      "aggs": {
        "avg_popularity": {
          "avg": {
            "field": "popularity"
          }
        },
        "avg_revenue": {
          "avg": {
            "field": "revenue"
          }
        }
      }
    }
  }
}
```

(Partial) Query response:


```

1  {
2    "took": 42,
3    "timed_out": false,
4    "_shards": {
5      "total": 1,
6      "successful": 1,
7      "skipped": 0,
8      "failed": 0
9    },
10   "hits": {
11     "total": {
12       "value": 10000,
13       "relation": "gte"
14     },
15     "max_score": null,
16     "hits": []
17   },
18   "aggregations": {
19     "runtime_ranges": {
20       "buckets": [
21         {
22           "key": "low",
23           "to": 90,
24           "doc_count": 85523,
25           "avg_revenue": {
26             "value": 601782.1697905826
27           },
28           "avg_popularity": {
29             "value": 2.3505774638260926
30         }
31       ],
32       {
33         "key": "medium",
34         "from": 91,
35         "to": 120,
36         "doc_count": 64075,
37         "avg_revenue": {
38           "value": 5402054.157175185
39         },
40         "avg_popularity": {
41           "value": 5.392992710696037
42         }
43       },
44       {
45         "key": "high",
46         "from": 121,
47         "doc_count": 15937,
48         "avg_revenue": {
49           "value": 18320067.96084583
50         },
51         "avg_popularity": {
52           "value": 7.126913532836304
53         }
54       }
55     ]
56   }
57 }

```

(t) Partial query response.