POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

SMBUD (Systems and Methods for Big Unstructured Data) Project

MongoDB & Elasticsearch queries

Authors:
Acquadro Patrizio - 11087897
Drugman Tito Nicola - 10775899
Gadiaga David - 11069658

Course: Systems and Methods for Big Unstructured Data

Professor: Prof. Brambilla Marco

Academic Year: 2024-25

Deliverable:	SMBUD Project Document
Authors:	Acquadro Patrizio, Drugman Tito Nicola, David Gadiaga
Version:	1.0
Date:	December 1, 2024
Download page:	GitHub Repository
Copyright:	

Contents

1 | Introduction

In this project, we aim to tackle two distinct but equally critical problems through data-driven analysis, leveraging the power of MongoDB and Elasticsearch. These problems are centered around understanding user preferences and optimizing experiences in two domains: high-profile travel planning and anime streaming.

1.1. Problem 1: *Assisting a High-Profile Family in Planning Their Trip to the United States*

1.1.1. Description

Our objective is to address a common challenge faced by high-end travel agencies: providing impeccable service to demanding clients with unique requirements. We envisioned a specific case where a high-profile family (parents, two children, and possibly a Chinese breed dog) wants to plan a trip to the United States. For this family, money is not a constraint, but convenience and fulfillment of their special requests are crucial.

To meet this scenario effectively, travel agencies analyze several factors:

- **Performance of Commercial Activities:** Analyze which activities receive the highest customer satisfaction, identifying variations in positive reviews across different cities in the US. This supports the selection of the best options to meet the family's travel needs.

- **Segmentation of Customer Preferences:** Explore how consumers perceive different categories of activities, analyzing trends in popularity and ratings. This analysis allows for highly personalized recommendations aligned with the interests and priorities of individual family members.

- **Business Hours:** Assess trends in business hours and identify which days of the week or times of day are most common for businesses to be open. This helps in planning visits to maximize convenience and accessibility.

1.1.2. Rationale for Choosing MongoDB

MongoDB is particularly well-suited to address these challenges for several reasons:

- **Flexible Schema for Complex Data:** Yelp datasets contain complex and variable types of data, such as business attributes, reviews, and user information. MongoDB's flexible schema enables efficient storage and querying of unstructured and semi-structured data.
- **Support for Real-Time Behavioral Analysis:** MongoDB allows recording interactions and preferences of the family during the trip, providing updated suggestions based on real-time data.
- **Optimization for Hierarchical and Linked Data:** With its flexible structure, MongoDB easily manages hierarchical relationships, such as travel preferences of each family member, to personalize itineraries and recommendations.
- **Adaptability to Changes:** MongoDB supports the evolution of the dataset without requiring a complete schema refactor, enabling quick responses to new requests or changes in the travel plan.

1.2. Problem 2: *Analyzing Anime Performance and Trends for a Streaming Platform*

1.2.1. Description

The primary challenge we aim to address is understanding the dynamics of anime performance and trends to support a streaming platform in making strategic decisions to optimize content selection and promotion, enhance user experience, and increase audience engagement.

This includes the analysis of factors such as:

- **Performance Analysis by Studio and Production Status:** Understand how different studios perform based on the production status of their anime by calculating average scores. This helps identify studios producing high-quality content at various production stages.
- **Temporal Trends in Anime Ratings:** Analyze how anime ratings have changed over the years by calculating average scores for each premiere year. This helps identify periods with high-performing anime.
- **Identification of Top-Rated Anime:** Identify the top 10 anime for each year with scores above a certain threshold. This highlights the most popular and critically acclaimed anime, facilitating content selection and implementation of effective recommendation algorithms.

1.2.2. Rationale for Choosing Elasticsearch

Elasticsearch is particularly well-suited to address these challenges for several reasons:

- **Full-Text Search Capabilities:** The anime dataset includes extensive descriptions, synopses, and user reviews. Elasticsearch excels at full-text search, enabling quick and relevant searches within these complex textual fields.
- **Aggregations:** Elasticsearch's aggregation capabilities efficiently calculate complex metrics such as averages, counts, and distributions. This is essential for performance analyses by studio, temporal trends, and top-rated anime identification, providing detailed and timely insights for platform decisions.

2 | Dataset

2.0.1. Problem 1 - *Analysis of Yelp Dataset (MongoDB)*

Dataset Description and Structure

Figure 2.1: Yelp Dataset

This dataset is a subset of Yelp's businesses, reviews, and user interactions. It was initially curated for the **Yelp Dataset Challenge**, an initiative encouraging students to analyze Yelp's data and present their findings. The most recent dataset includes comprehensive details about businesses located in eight metropolitan areas across the USA and Canada, providing a rich foundation for research and analysis. The dataset is accessible on Kaggle, and it is divided into several JSON files, each representing distinct aspects of Yelp's ecosystem:

- `yelp_academic_dataset_business.json`: Offers detailed information on businesses, including their geographical location, operational attributes, and assigned categories.
- `yelp_academic_dataset_checkin.json`: Contains data about user check-ins across various businesses.
- `yelp_academic_dataset_review.json`: Includes written reviews by users for specific businesses.
- `yelp_academic_dataset_tip.json`: Contains brief tips left by users about businesses.
- `yelp_academic_dataset_user.json`: Documents user profiles, including connections, reviews, and other social interactions.

For the purpose of this study, the `yelp_academic_dataset_business.json` file has been selected. This file is essential for exploring business attributes, analyzing

their geographical distribution, and understanding consumer engagement, making it highly relevant to our problem objectives.

Each entry in the JSON file (whose total amount is 150346) represents a business, with attributes stored as key-value pairs. These attributes may vary across records, as the JSON format allows flexibility, meaning some fields may be present in one record but absent in another. Below is a list of these key fields, along with their corresponding MongoDB data types:

- `business_id`: (`string`) - A 22-character unique identifier for each business.
- `name`: (`string`) - The official name of the business.
- `address`: (`string`) - The physical address of the business.
- `city`: (`string`) - The city where the business operates.
- `state`: (`string`) - The state where the business is located.
- `postal_code`: (`string`) - The ZIP code of the business's address.
- `latitude`: (`float`) - Latitude coordinate of the business's location.
- `longitude`: (`float`) - Longitude coordinate of the business's location.
- `stars`: (`float`) - Average star rating for the business, rounded to half-stars (e.g., 4.5).
- `review_count`: (`integer`) - Total number of reviews the business has received.
- `is_open`: (`integer`) - A binary value indicating whether the business is currently operational (1 for open, 0 for closed).
- `attributes`: (`object`) - Key-value pairs describing various business features. Some values, such as `BusinessParking`, are nested objects with fields like `garage`, `street`, or `lot`.
- `categories`: (`array of strings`) - A list of categories assigned to the business, such as "Mexican," "Gastropubs," or "Bars."
- `hours`: (`object`) - Maps each day of the week to a string representing opening and closing times (e.g., `"Monday": "10:00-21:00"`).

Non-Relational Database Implementation

MongoDB was chosen to manage and query the dataset due to its flexibility, scalability, and schema-less structure. This makes it an excellent choice for handling JSON-based data with varying levels of detail across documents. Below is an example of how a single business entry is stored in MongoDB:

```
{
  "business_id": "xyz789",
```

```
  "name": "Sunrise Café",
  "address": "456 Elm St",
  "city": "Austin",
  "state": "TX",
  "postal_code": "73301",
  "latitude": 30.2672,
  "longitude": -97.7431,
  "stars": 4.0,
  "review_count": 85,
  "is_open": 1,
  "attributes": {
    "OutdoorSeating": true,
    "WiFi": "free",
    "BikeParking": true
  },
  "categories": ["Breakfast & Brunch", "Cafes"],
  "hours": {
    "Monday": "6:30-14:30",
    "Tuesday": "6:30-14:30",
    "Wednesday": "6:30-14:30",
    "Thursday": "6:30-14:30",
    "Friday": "6:30-15:00",
    "Saturday": "7:00-15:00",
    "Sunday": "7:00-15:00"
  }
}
```


2.0.2. Problem 2 - *Analysis of Anime Dataset (ElasticSearch)*

Dataset Description and Structure

Figure 2.2: Dataset of Anime Data

The dataset used in this project reflects the Japanese animation industry, providing

detailed information on 24,905 anime. Supplied as a CSV file named *anime-dataset-2023.csv*, it was downloaded from Kaggle, a platform widely used in undergraduate courses for academic projects. This dataset was suggested by the professor as a primary resource for exploring data related to features, ratings, popularity, and production details.

- **anime id** (`long`): Unique identifier for each anime, used to distinguish entries in the dataset.
- **Name** (`text`): Original name of the anime in its production language.
- **English name** (`text`): English name of the anime, useful for international analyses.
- **Other name** (`keyword`): Alternate titles or names by which the anime is known in different languages.
- **Score** (`double`): Numerical rating (from 1 to 10) based on user reviews, indicative of perceived quality.
- **Genres** (`text`): Genres associated with the anime, separated by commas (e.g., action, adventure).
- **Synopsis** (`text`): Brief summary of the plot, essential for understanding narrative content.
- **Type** (`keyword`): Type of anime (TV series, movie, OVA, etc.).
- **Episodes** (`keyword`): Number of episodes, indicative of overall runtime.
- **Aired** (`text`): The airing period of the anime, including start and end dates.
- **Premiered** (`keyword`): Season (e.g., Spring, Summer) and year of debut, useful for studying seasonality.
- **Status** (`keyword`): Current status of the anime (completed, ongoing, planned).
- **Producers** (`text`): Production companies responsible for the development of the anime.
- **Licensors** (`text`): Official licensors (often streaming platforms).
- **Studios** (`keyword`): Animation studios involved in the production.
- **Source** (`keyword`): Source material (manga, light novel, original).
- **Duration** (`keyword`): Average duration of each episode.
- **Rating** (`keyword`): Age rating (e.g., PG-13, R).
- **Rank** (`keyword`): Position based on quality or popularity criteria.
- **Popularity** (`long`): Popularity ranking based on user interactions.

- **Favourites** (`long`): Number of times the anime has been saved as a favorite by users.
- **Scored By** (`double`): Total number of users who rated the anime.
- **Members** (`long`): Number of users who added the anime to their personal lists.
- **Image URL** (`keyword`): Link to the anime's image or poster.

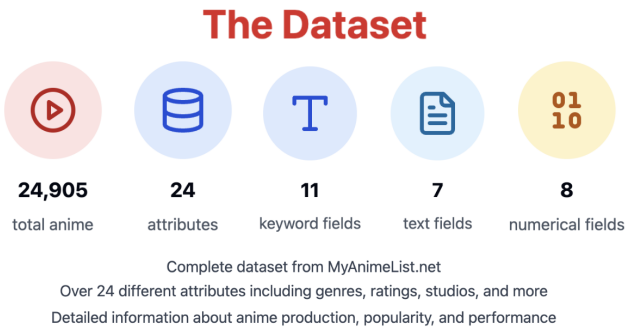

Figure 2.3: Recap of the Anime Dataset

Non-Relational Database Implementation

To represent this dataset was chosen Elasticsearch, where a mapping was configured to optimize the handling of textual and numerical data. Below is an example of the JSON mapping used for the index:

```
PUT /anime/_mapping
{
  "properties": {
    "anime_id": { "type": "long" },
    "name": { "type": "text" },
    "english_name": { "type": "text" },
    "other_name": { "type": "keyword" },
    "score": { "type": "double" },
    "genres": { "type": "text"},
    "synopsis": { "type": "text" },
    "type": { "type": "keyword" },
    "episodes": { "type": "keyword" },
    "aired": { "type": "text"},
    "premiered": { "type": "keyword" },
    "status": { "type": "keyword" },
    "producers": { "type": "text" },
    "licensors": { "type": "text" },
    "studios": { "type": "keyword" },
```

```
    "source": { "type": "keyword" },
    "duration": { "type": "keyword" },
    "rating": { "type": "keyword" },
    "rank": { "type": "keyword" },
    "popularity": { "type": "long" },
    "favourites": { "type": "long" },
    "scored_by": { "type": "double" },
    "members": { "type": "long" },
    "image_url": { "type": "keyword" }
  }
}
```

3 | Queries

3.1. Problem 1 - *Queries with MongoDB*

Query 1 - Find Open Dentists in Philadelphia on Sunday without Appointment

Description: This query searches for dentists in Philadelphia that are open on Sundays, do not require appointments, have a rating higher than 3 stars, and are currently open.

Context: The family has just arrived in Philadelphia from Rome, exhausted but excited for their adventure. Sadly, one of the kids has been complaining about severe tooth pain for the entire flight. It's Sunday, and they're unsure where to find help, so they urgently call their travel consultant to locate a dentist open on Sunday in Philadelphia with no appointment required. Since money is not a problem they will spend anything for their child, the starts of the dentist should be greater than 3.

Key Components:

- **Location and Category Filters**: Filters businesses in "Philadelphia" with categories matching "dentist" (case-insensitive).
- **Operating Hours and Status**: Ensures the dentist is open on Sunday and currently open (`is_open: 1`).
- **Rating and Appointment Requirements**: Filters dentists with stars greater than 3 and that do not require appointments.
- **Sorting**: Orders the results by `stars` in descending order.

Query:

```
db.getCollection('mycollection3').find(
  {
    city: "Philadelphia",
    categories: { $regex: /dentist/i },
    "hours.Sunday": { $exists: true, $ne: "" },
    is_open: 1,
    stars: { $gt: 3 },
```

```
    "attributes.ByAppointmentOnly": "False"
  },
  {
    name: 1,
    adress: 1,
    city: 1,
    stars: 1,
    is_open: 1,
    attributes: 1,
    categories: 1
  }
).sort({ stars: -1 });
```

Outcome (partial):

```
{
  _id: ObjectId('6762c0b837de02daf2cf9bbf'),
  name: 'Philadelphia Mills Dental Center',
  city: 'Philadelphia',
  stars: 5,
  is_open: 1,
  attributes: {
    AcceptsInsurance: 'True',
    ByAppointmentOnly: 'False',
    BusinessAcceptsCreditCards: 'True'
  },
  categories: 'Health & Medical, Dentists, General
     Dentistry'
}
```

```
{
  _id: ObjectId('6762c09737de02daf2cd5515'),
  name: 'Center City Emergency Dentist',
  city: 'Philadelphia',
  stars: 3.5,
  is_open: 1,
  attributes: {
    BusinessAcceptsCreditCards: 'True',
    WiFi: "u'free'",
    ByAppointmentOnly: 'False',
    AcceptsInsurance: 'True'
  },
  categories: 'Orthodontists, Health & Medical, Oral
     Surgeons, General Dentistry, Dentists, Cosmetic
```

```
    Dentists'
}
```

Extra: In the following plot we can see the distribution of Star Ratings for businesses. The query returned only three results, so the number of businesses in the plot is very limited. If we decided to relax one of the constraints we could expect more documents returned and thus a more overcrowded plot.

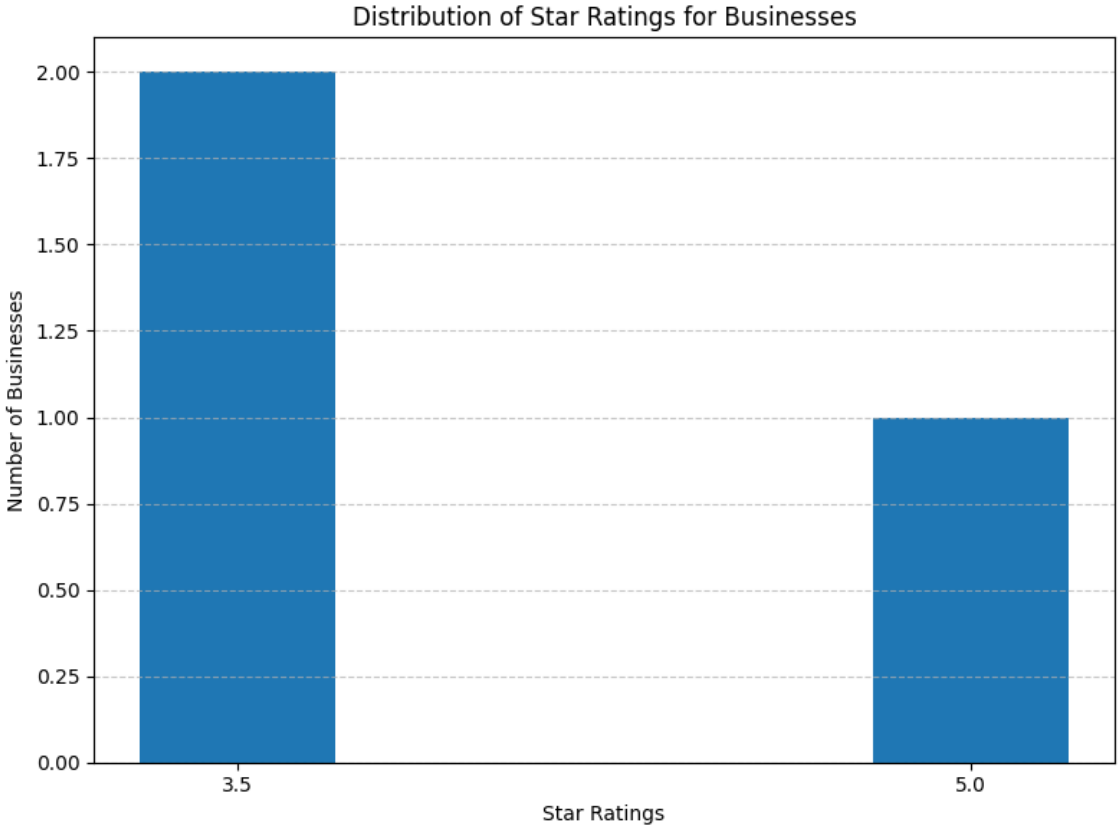

Figure 3.1: Star ratings for returned business.

Query 2 - Culinary Planning: Number of Restaurants by State

Description: This query aggregates data to identify the states with the highest number of open restaurants rated 4.5 stars or higher, while also calculating the average number of reviews to ensure quality.

Context: The family often seeks out the best dining experiences available. To help them enjoy their meals, they ask to the travel consultant to identify the states with the highest concentration of open restaurants rated 4.5 stars or higher, averaging the number of reviews to ensure quality.

Key Components:

- **Category and Rating Filters**: Filters businesses with categories matching "restaurant" (case-insensitive), open (`is_open: 1`), and stars ≥ 4.5.
- **Grouping and Aggregation**: Groups the results by `state`, counting the number of restaurants and calculating the average `review_count`.
- **Sorting and Limiting**: Sorts the states by `restaurant_count` in descending order and limits the results to the top 3 states.

Query:

```
db.getCollection('mycollection3').aggregate([
  {
    $match: {
      categories: { $regex: /restaurant/i },
      is_open: 1,
      stars: { $gte: 4.5 }
    }
  },
  {
    $group: {
      _id: "$state",
      restaurant_count: { $sum: 1 },
      average_reviews: { $avg: "$review_count" }
    }
  },
  {
    $sort: { restaurant_count: -1 }
  },
  {
    $limit: 3
  }
]);
```

Outcome (total):

```
{
  _id: 'PA',
  restaurant_count: 1633,
  average_reviews: 104.89957134109002
}
```

```
{
  _id: 'FL',
  restaurant_count: 1405,
  average_reviews: 120.067615658363
}
```

```
{
  _id: 'TN',
  restaurant_count: 641,
  average_reviews: 148.36817472698908
}
```

Extra In the query we limited ourself to only threes states. So on the x-axis we have the three states with the highest amount of restaurants that satisfies the constraints we imposed. On the y-axis we have the number of resturants that do indeed satisfy what the family asked.

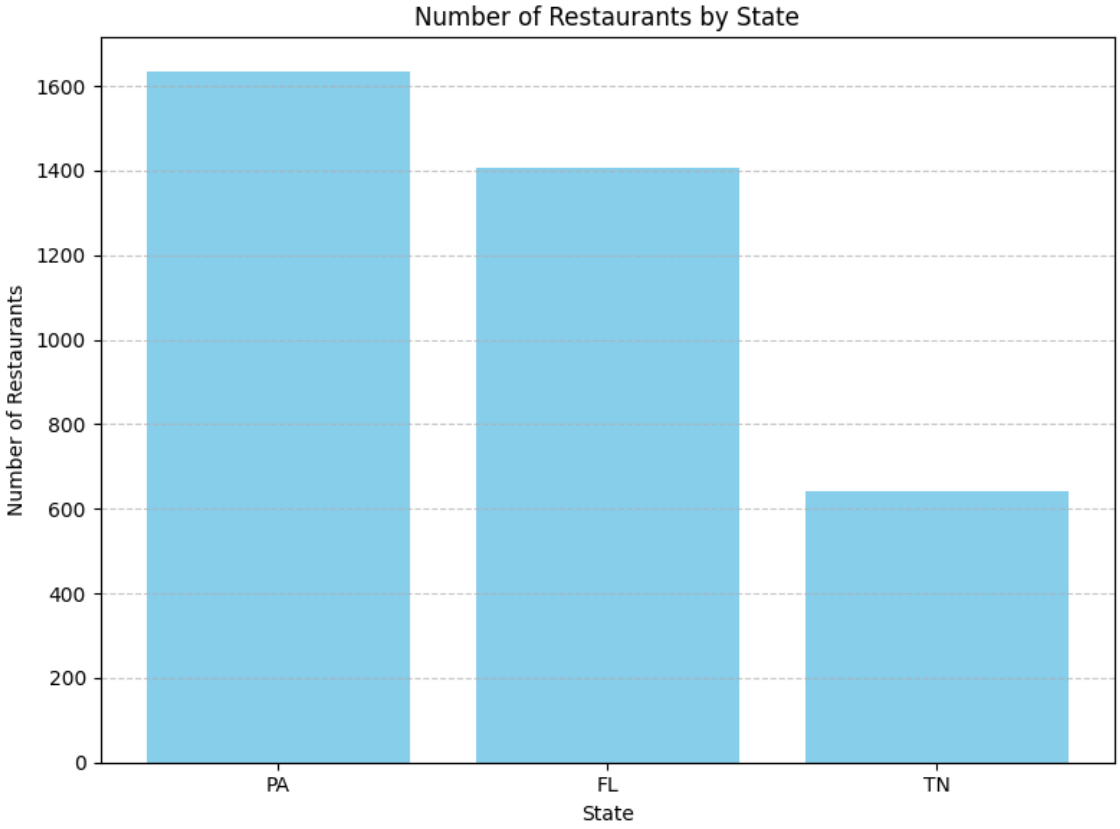

Figure 3.2: Number of Restaurant by State

Query 3 - Cities in California (Excluding *Goleta*)

Description: This query identifies the two most popular cities in California for shopping based on the number of businesses, excluding Goleta, which the family has already explored.

Context: The family plans to spend some days of their trip in California and they are looking for cities with plenty of shopping options. Since they visited *Goleta* for a wedding three months ago and they have already explored the city, they want to focus on new destinations. The call the travel consultant (who of course is always working for the family, no matters the time of the day) to find the two most popular cities in California, based on the number of businesses, excluding *Goleta*.

Key Components:

- **State and City Filters**: Filters businesses in the state of "CA" and excludes the city "Goleta".
- **Grouping and Counting**: Groups the results by `city` and counts the number of businesses in each city.
- **Sorting and Limiting**: Sorts the cities by business count in descending order and limits the results to the top 2 cities.

Query:

```
db.getCollection('mycollection3').aggregate([
  { $match: { state: "CA" } },
  {
    $group: {
      _id: "$city",
      count: { $sum: 1 }
    }
  },
  { $match: { _id: { $ne: "Goleta" } } },
  { $sort: { count: -1 } },
  { $limit: 2 }
]);
```

Outcome (total):

```
{
  _id: 'Santa Barbara',
  count: 3829
}
```

```
{
  _id: 'Carpinteria',
```

```
  count: 298
}
```

Extra In the plot we can see the two cities in California that have the highest number of businesses. It was our intention to not include *Goleta* in the query since it was the second most popular city for businesses.

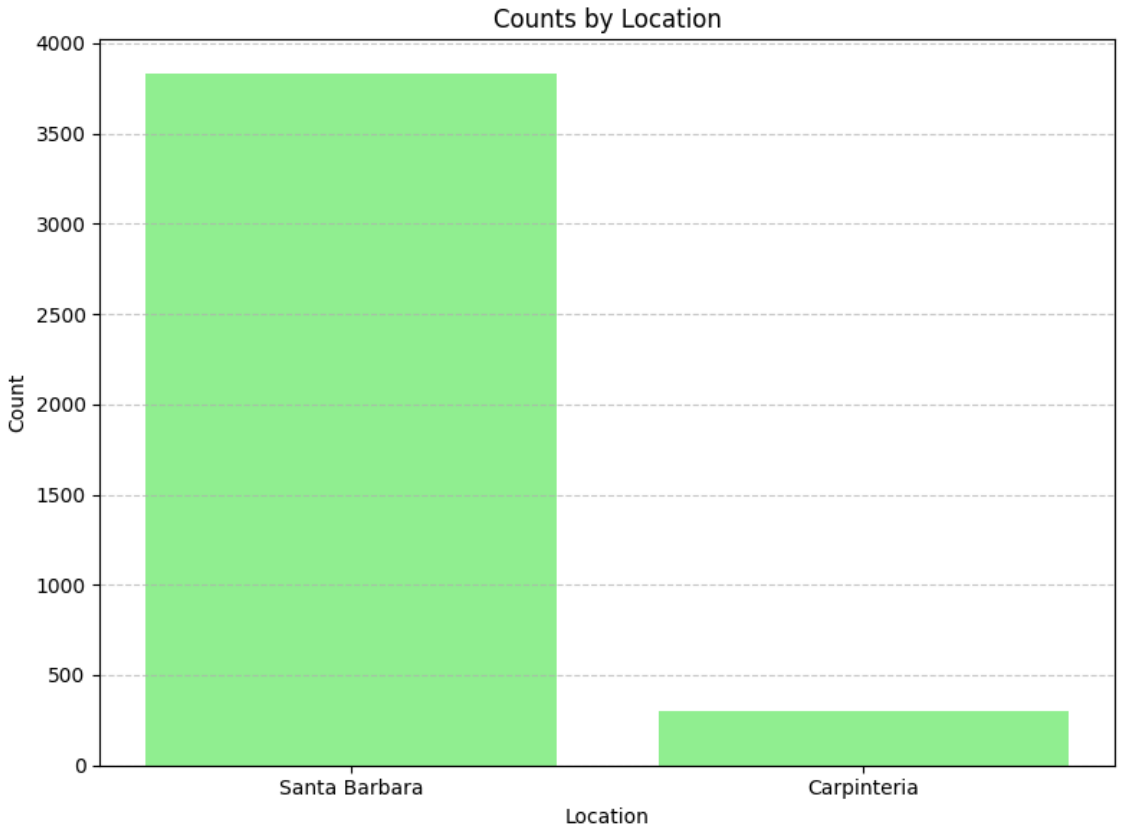

Figure 3.3: Count by Location.

Query 4 - Best Pub in Each State

Description: This query selects the best pub in each state based on the highest star rating and the number of reviews as a tiebreaker. It then lists these top pubs alphabetically by state, limiting the results to 10 pubs.

Context: The father really enjoys craft beer (we could almost talk about addiction). For each state, find the best pub based on rating (and review count as tiebreaker). Then list them alphabetically by state. He would really like to try all of them but he must limited to 10 since his personal flight jet has limited fuel.

Key Components:

- **Category Filter**: Filters businesses with categories matching "Pubs" (case-insensitive).
- **Sorting and Grouping**: Sorts the results by `stars` descending and `review_count` descending, then groups by `state` to select the top pub per state.
- **Projection and Limiting**: Projects relevant fields and limits the final results to 10 pubs, sorted alphabetically by state.

Query:

```
db.getCollection('mycollection3').aggregate([
  {
    $match: {
      categories: { $regex: /\bPubs\b/i }
    }
  },
  {
    $sort: { stars: -1, review_count: -1 }
  },
  {
    $group: {
      _id: "$state",
      bestPub: { $first: "$$ROOT" }
    }
  },
  {
    $replaceRoot: { newRoot: "$bestPub" }
  },
  {
    $project: {
      name: 1,
      address: 1,
      city: 1,
      state: 1,
```

```
      stars: 1,
      review_count: 1,
      categories: 1
    }
  },
  {
    $sort: { state: 1 }
  },
  {
    $limit: 10
  }
]);
```

Outcome (partial):

```
{
  _id: ObjectId('6762c0a437de02daf2ce40c3'),
  name: 'Sea Change Brewing',
  address: '5302 50 Street',
  city: 'Beaumont',
  state: 'AB',
  stars: 5,
  review_count: 5,
  categories: 'Food, Pubs, Breweries, Bars, Nightlife'
}
```

```
{
  _id: ObjectId('6762c0a337de02daf2ce3057'),
  name: 'Bar Pass ',
  address: '417 N 4th Ave',
  city: 'Tucson',
  state: 'AZ',
  stars: 5,
  review_count: 9,
  categories: 'Bars, Beer Gardens, Pubs, Cocktail Bars,
     Dance Clubs, Nightlife'
}
```

```
{
  _id: ObjectId('6762c0a537de02daf2ce4ede'),
  name: 'Brass Bear Brewing & Bistro',
  address: '28 Anacapa St, Ste E',
  city: 'Santa Barbara',
  state: 'CA',
```

```
  stars: 4.5,
  review_count: 269,
  categories: 'Wineries, Food, Breweries, Nightlife, Bars
     , Arts & Entertainment, Champagne Bars, Pubs,
     American (New), Brewpubs, Restaurants'
}
```

Extra: In this plot we can see on the x-axis the number of stars, while on the y-axis we have the number of reviews. We can see that the number of stars is either 5 or 4.5.

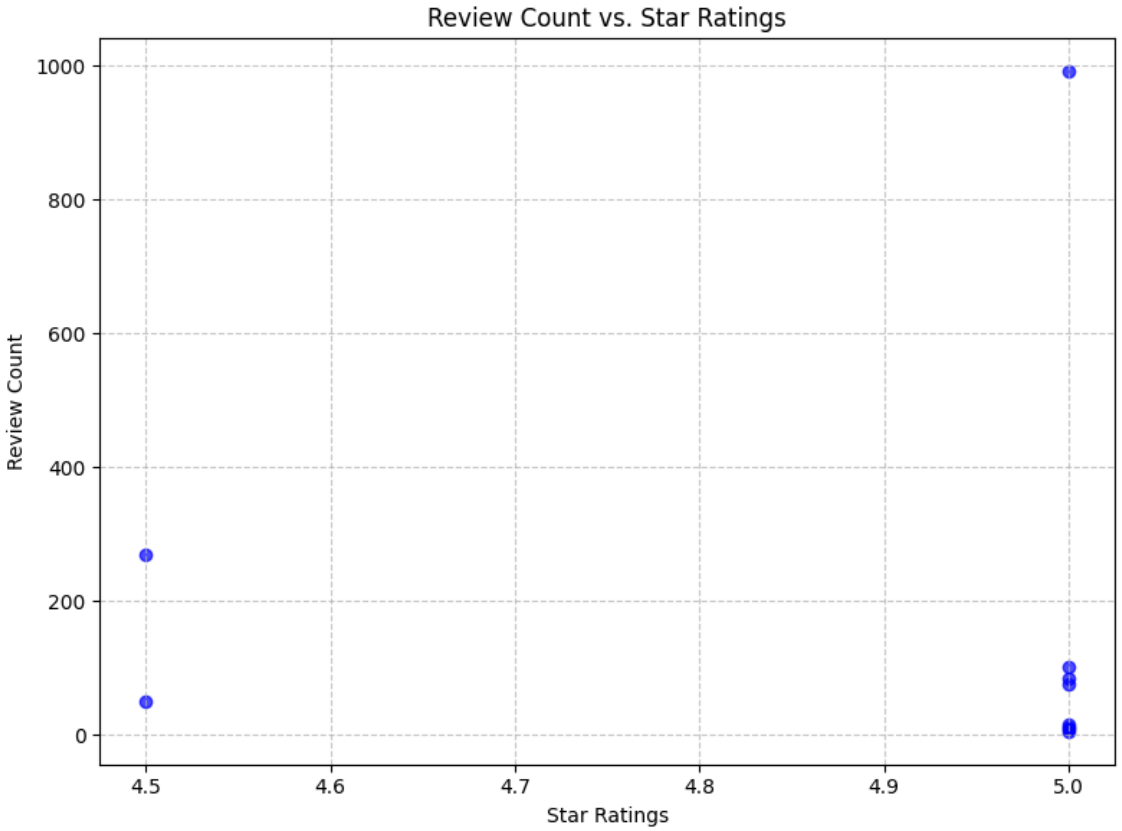

Figure 3.4: Review Count vs Star Ratings

Query 5 - Real Estate Averages in FL, TX, AZ

Description: This query calculates the average star ratings and average review counts for businesses categorized as "Real Estate" in the states of Florida (FL), Texas (TX), and Arizona (AZ).

Context: The couple might want to invest in real estate while traveling. They want average star ratings and review counts for businesses listed as “Real Estate” in Florida, Texas and Arizona.

Key Components:

- **Category and State Filters**: Filters businesses with categories matching "Real Estate" (case-insensitive) and located in "FL", "TX", or "AZ".
- **Grouping and Aggregation**: Groups the results by `state` and calculates the average `stars` and average `review_count`.
- **Projection and Sorting**: Projects the state along with the calculated averages and sorts the results by `averageStars` in descending order.

Query:

```
db.getCollection('mycollection3').aggregate([
  {
    $match: {
      categories: { $regex: /\bReal Estate\b/i },
      state: { $in: ["FL", "TX", "AZ"] }
    }
  },
  {
    $group: {
      _id: "$state",
      averageStars: { $avg: "$stars" },
      averageReviewCount: { $avg: "$review_count" }
    }
  },
  {
    $project: {
      _id: 0,
      state: "$_id",
      averageStars: 1,
      averageReviewCount: 1
    }
  },
  {
    $sort: { averageStars: -1 }
  }
```

```
]);
```

Outcome (total):

```
{
  averageStars: 2.776760563380282,
  averageReviewCount: 14.526760563380282,
  state: 'FL'
}
```

```
{
  averageStars: 2.678466076696165,
  averageReviewCount: 13.607669616519175,
  state: 'AZ'
}
```

```
{
  averageStars: 2.25,
  averageReviewCount: 8.5,
  state: 'TX'
}
```

Extra In this plot we have on the x-axis the average star, while on the y-axis the average reviews. Since we are limited only on three states (TX, AZ and FL) the plot is particularly empty. It is interesting to notice that possibly there is a linear correlation between the average stars and the number of reviews.

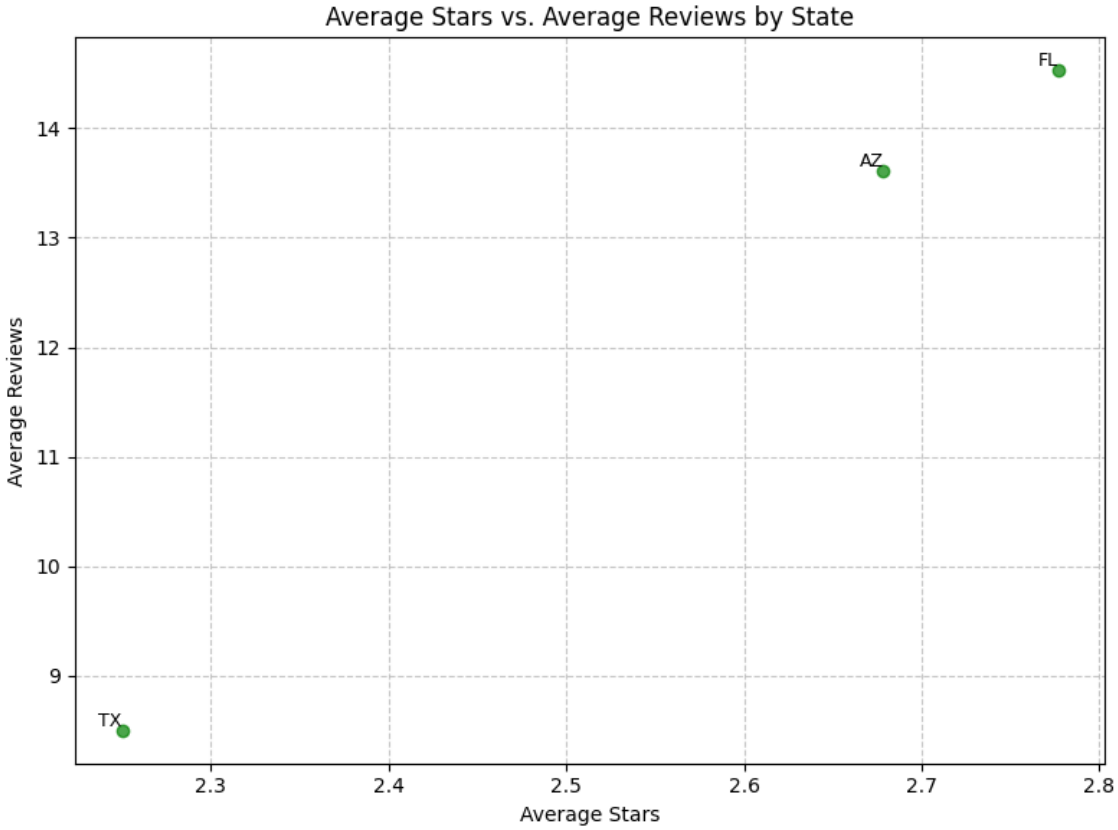

Figure 3.5: Average Stars vs Average Reviews by State

Query 6 - Shopping on bikes (but where?)

Description: This query searches for stylish shopping centers in Arizona that offer bike parking and are located in cities starting with the letter "T".

Context: The mother receives a call from a friend, recommending a chic shopping experience in Arizona. She decides to rent bikes and reach the shopping center (so it must have bike parking). However, she forgets the name of the place, only recalling that it's located in a city starting with "T". So she call the travel consultant (who as we said is 24h available and does not sleeps) to help her.

Key Components:

- **Category and Location Filters**: Filters businesses with categories matching "Shopping Centers" (case-insensitive), located in Arizona (`state: "AZ"`) and cities starting with "T".
- **Amenities and Open Status**: Ensures the shopping centers are open (`is_open: 1`) and have bike parking (`"attributes.BikeParking": "True"`).
- **Projection**: Selects relevant fields such as `name`, `city`, `state`, `categories`, and `attributes`.

Query:

```
db.getCollection('mycollection3').aggregate([
  {
    $match: {
      categories: { $regex: /\bShopping Centers\b/i },
      is_open: 1,
      "attributes.BikeParking": "True",
      city: { $regex: /^T/i },
      state: "AZ"
    }
  },
  {
    $project: {
      name: 1,
      city: 1,
      state: 1,
      categories: 1,
      attributes: 1
    }
  }
]);
```

Outcome (partial):

```
{
  name: 'St. Philips Plaza',
  city: 'Tucson',
  state: 'AZ',
  attributes: {
    BusinessAcceptsCreditCards: 'True',
    BikeParking: 'True',
    RestaurantsPriceRange2: '3',
    BusinessParking: "{'garage': False, 'street': False,
       'validated': False, 'lot': False, 'valet': False}"
  },
  categories: 'Shopping, Restaurants, Shopping Centers,
     Nightlife'
}
```

```
{
  name: 'Foothills Mall',
  city: 'Tucson',
  state: 'AZ',
  attributes: {
    BusinessParking: "{'garage': False, 'street': False,
       'validated': False, 'lot': True, 'valet': False}",
    RestaurantsPriceRange2: '2',
    BikeParking: 'True',
    BusinessAcceptsCreditCards: 'True',
    DogsAllowed: 'False'
  },
  categories: 'Shopping, Shopping Centers'
}
```

Extra: In this plot we can see the distribution of Parking types in Tucson (Arizona). We cans see that that the majority of business have parking lot, while only one have a garage.

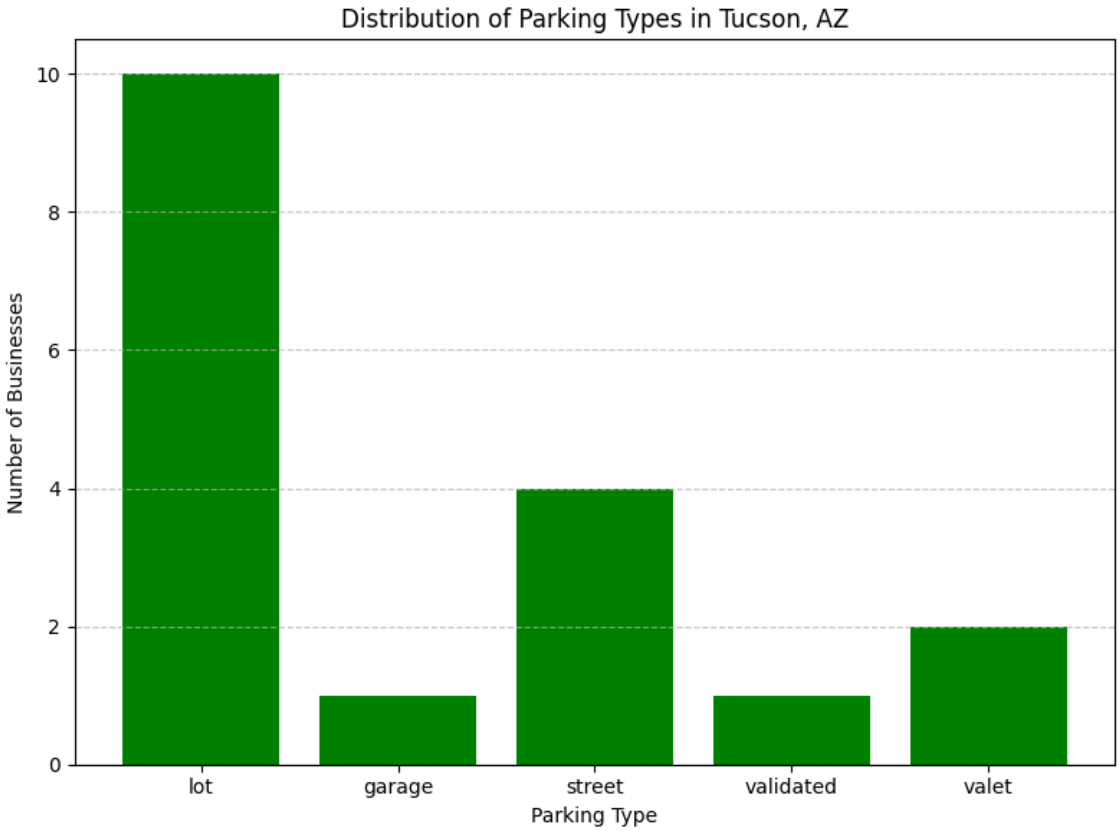

Figure 3.6: Distribution of Parking Types in Tuscon

Query 7 - A lunch with the kids and dog

Description: This query searches for open restaurants that are highly rated (4 stars or above), dog-friendly, suitable for children, and offer good meals during lunch.

Context: The family is planning a relaxed lunch and needs a restaurant that accommodates everyone, including their "wonderful" Chinese crested dog. They specifically look for highly-rated (4 stars or above) establishments that are dog-friendly, great for kids, and have good meals at lunch. The travel travel consultant is woke up at 4am (classical time-zone problems) to find the restaurants that accommodate the family requests.

Key Components:

- **Category and Rating Filters**: Filters businesses with categories matching "Restaurants" (case-insensitive), open (`is_open: 1`), and stars ≥ 4.
- **Attributes Filters**: Ensures the restaurants are dog-friendly (`"attributes.DogsAllowed": "True"`), suitable for kids (`"attributes.GoodForKids": "True"`), and offer good lunch options (`"attributes.GoodForMeal"` contains "lunch: True").
- **Sorting and Limiting**: Sorts the results by `stars` and `review_count` in descending order, and limits the results to the top 3 restaurants.

Query:

```
db.getCollection('mycollection3').aggregate([
  {
    $match: {
      is_open: 1,
      categories: { $regex: /\bRestaurants?\b/i },
      "attributes.GoodForMeal": { $regex: /'lunch':\s*
         True/i },
      "attributes.GoodForKids": "True",
      "attributes.DogsAllowed": "True",
      stars: { $gte: 4 }
    }
  },
  {
    $sort: {
      stars: -1,
      review_count: -1
    }
  },
  {
    $project: {
      _id: 0,
      name: 1,
```

```
      categories: 1,
      is_open: 1,
      stars: 1,
      review_count: 1,
      "attributes.DogsAllowed": 1,
      "attributes.GoodForMeal": 1,
      "attributes.GoodForKids": 1
    }
  },
  {
    $limit: 3
  }
]);
```

Outcome (partial):

```
{
  name: 'Blues City Deli',
  stars: 5,
  review_count: 991,
  is_open: 1,
  attributes: {
    GoodForKids: 'True',
    GoodForMeal: "{'dessert': False, 'latenight': False,
       'lunch': True, 'dinner': False, 'brunch': False, '
       breakfast': False}",
    DogsAllowed: 'True'
  },
  categories: 'Delis, Bars, Restaurants, Nightlife, Pubs,
      American (Traditional), Sandwiches'
}
```

```
{
  name: 'Tumerico',
  stars: 5,
  review_count: 705,
  is_open: 1,
  attributes: {
    DogsAllowed: 'True',
    GoodForMeal: "{u'breakfast': True, u'brunch': True, u
       'lunch': True, u'dinner': None, u'latenight':
       False, u'dessert': False}",
    GoodForKids: 'True'
  },
```

```
  categories: 'Mexican, Gluten-Free, Vegetarian,
     Restaurants, Vegan'
}
```

Extra On this plot we can see the star rating (not-interesting is always 5). On the y-axis we have the number of reviews. Among the documents returned *Blues City Deli* is the one that received the most amount of reviews (almost 1000) while *Cafe Soleil* had less than 500 reviews.

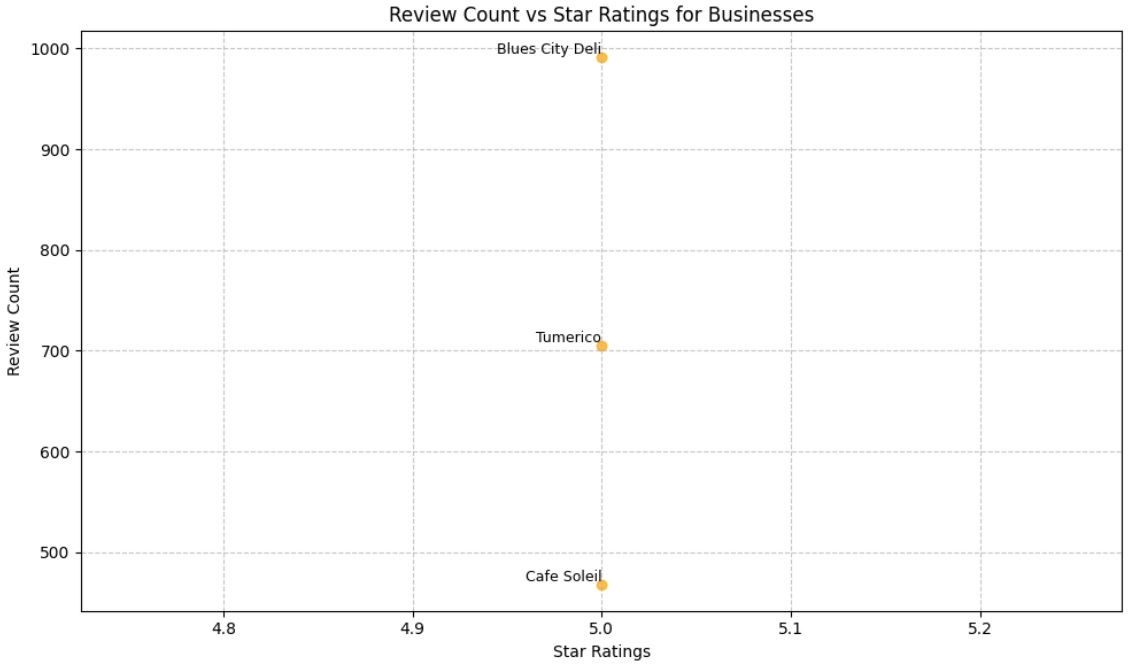

Figure 3.7: Review Count vs Star Rating for Businesses

Query 8 - Luxurious hotel in Florida

Description: This query searches for luxurious hotels in Florida that are open, have a rating above 4 stars, and are dog-friendly.

Context: During their trip to Florida, the family wants to stay in a luxurious hotel that warmly welcomes their Chinese crested dog. To ensure a comfortable and enjoyable experience, they specifically seek hotels with a rating above 4 stars that are dog-friendly. The travel needs to searches the database for open hotels in Florida meeting these preferences.

Key Components:

- **Category and Location Filters**: Filters businesses with categories matching "Hotel(s)" (case-insensitive), located in Florida (`state: "FL"`), and open (`is_open: 1`).
- **Rating and Amenities**: Ensures the hotels have stars > 4 and allow dogs (`"attributes.DogsAllowed": "True"`).
- **Projection**: Selects relevant fields such as `name`, `is_open`, `stars`, `review_count`, `address`, `city`, `state`, `categories`, and `"attributes.DogsAllowed"`.

Query:

```
db.getCollection('mycollection3').aggregate([
  {
    $match: {
      is_open: 1,
      categories: { $regex: /\bHotel(s)?\b/i },
      state: "FL",
      stars: { $gt: 4 },
      "attributes.DogsAllowed": "True"
    }
  },
  {
    $project: {
      name: 1,
      is_open: 1,
      stars: 1,
      review_count: 1,
      address: 1,
      city: 1,
      state: 1,
categories: 1,
      "attributes.DogsAllowed": 1
    }
  }
```

```
]);
```

Outcome (partial):

```
{
  _id: ObjectId('6762c0a037de02daf2cdec88'),
  name: 'AC Hotel by Marriott Tampa Airport',
  address: '4020 West Boy Scout Boulevard',
  city: 'Tampa',
  state: 'FL',
  stars: 4.5,
  review_count: 53,
  is_open: 1,
  attributes: {
    DogsAllowed: 'True'
  },
  categories: 'Venues & Event Spaces, Hotels, Event
     Planning & Services, Hotels & Travel'
}
```

```
{
  _id: ObjectId('6762c09a37de02daf2cd83bd'),
  name: 'Hollander Hotel',
  address: '421 4th Ave N',
  city: 'St. Petersburg',
  state: 'FL',
  stars: 4.5,
  review_count: 265,
  is_open: 1,
  attributes: {
    DogsAllowed: 'True'
  },
  categories: 'Event Planning & Services, Hotels & Travel
     , Hotels'
}
```

Extra in the plot we decided to show the total reviews for each document that was returned group by city. We can clearly see that Tampa had the most amount of reviews (this actually might reflect the real world: Tampa is indeed the biggest of those cities, followed by St. Petersburg).

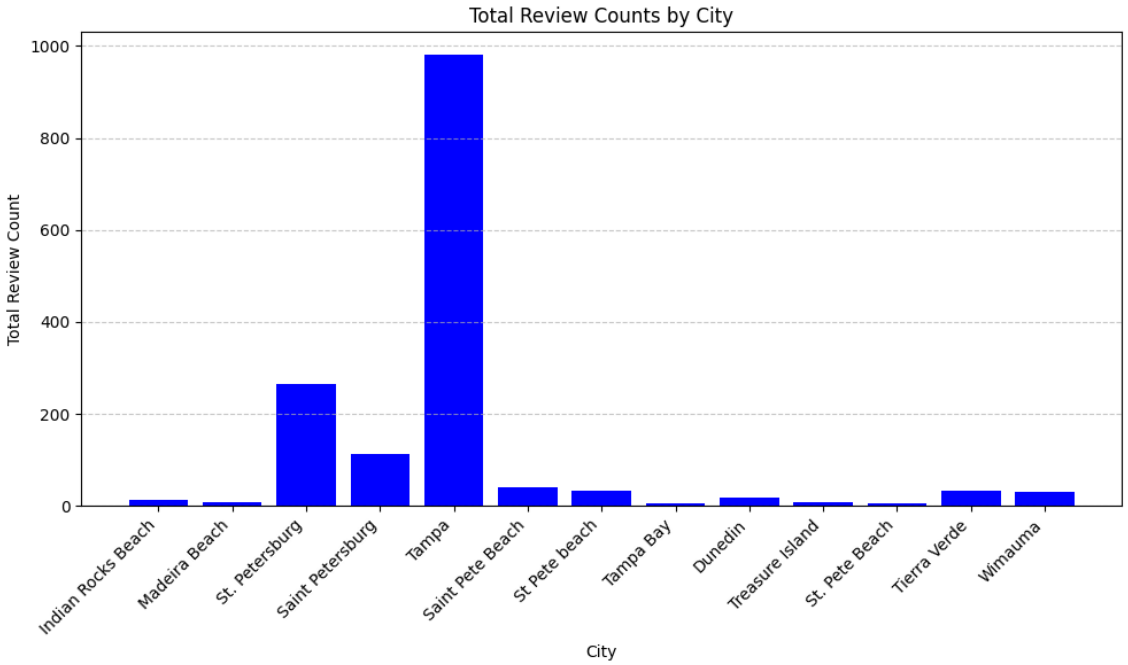

Figure 3.8: Total Review Counts by City

Query 9 - Family Fun in California

Description: This query searches for amusement parks or arcades in California that are open on Saturdays, have a rating of at least 4 stars, and have received at least 10 reviews.

Context: The kids want some weekend fun. The family is looking for an amusement parks or arcades in California, open on saturday, with a decent rating (at least 4 stars) and at least 10 reviews to ensure they're well-established.

Key Components:

- **Category and Location Filters**: Filters businesses with categories matching "Amusement" or "Arcade" (case-insensitive), located in California (`state: "CA"`), and open (`is_open: 1`).
- **Operating Hours and Ratings**: Ensures the businesses are open on Saturday (`"hours.Saturday"` exists and is not empty), have stars ≥ 4, and have `review_count` ≥ 10.
- **Sorting and Projection**: Sorts the results by `stars` and `review_count` in descending order, and projects relevant fields such as `name`, `city`, `stars`, `review_count`, `categories`, and `hours`.

Query:

```
db.getCollection('mycollection3').aggregate([
  {
    $match: {
      state: "CA",
      is_open: 1,
      categories: { $regex: /\b(Amusement|Arcade)\b/i },
      "hours.Saturday": { $exists: true, $ne: "" },
      stars: { $gte: 4 },
      review_count: { $gte: 10 }
    }
  },
  {
    $sort: { stars: -1, review_count: -1 }
  },
  {
    $project: {
      name: 1,
      city: 1,
      stars: 1,
      review_count: 1,
      categories: 1,
      hours: 1
```

```
    }
  }
]);
```

Outcome (total):

```
{
  _id: ObjectId('6762c0a037de02daf2cdef88'),
  name: 'Goleta Butterfly Grove',
  city: 'Goleta',
  stars: 4.5,
  review_count: 26,
  categories: 'Active Life, Pets, Hiking, Pet Services,
     Amusement Parks, Parks',
  hours: {
    Monday: '6:0-20:0',
    Tuesday: '6:0-20:0',
    Wednesday: '6:0-20:0',
    Thursday: '6:0-20:0',
    Friday: '6:0-20:0',
    Saturday: '6:0-20:0',
    Sunday: '6:0-20:0'
  }
}
```

Extra: No plot was possible since only one document was returned.

Query 10 - Emergency Auto Repair in Florida

Description: This query searches for auto repair shops in Florida that are either open 24 hours or close at 22:00 on at least one weekday (Monday to Friday).

Context: The father realizes he might need emergency auto repairs while driving in Florida. He specifically wants shops that advertise 24-hour availability or that close at 22:00 on at least one weekdays.

Key Components:

- **Category and Location Filters**: Filters businesses with categories matching "Auto" (case-insensitive), located in Florida (`state: "FL"`), and open (`is_open: 1`).
- **Operating Hours Conditions**: Uses to filter businesses that either are open 24 hours (`"attributes.Open24Hours": "True"`) or have closing times at 22:00 on any weekday (`"hours.Monday"` to `"hours.Friday"` matching regex `/-22:0{1,2}/i`).
- **Sorting and Projection**: Sorts the results by `stars` and `review_count` in descending order, and projects relevant fields such as `name`, `city`, `stars`, `review_count`, `categories`, specific weekday hours, and `"attributes.Open24Hours"`.

Query:

```
db.getCollection('mycollection3').aggregate([
  {
    $match: {
      state: "FL",
      is_open: 1,
      categories: { $regex: /\bAuto\b/i },
      $or: [
        { "attributes.Open24Hours": "True" },

        { "hours.Monday":     { $regex: /-22:0{1,2}$/i }
          },
        { "hours.Tuesday":    { $regex: /-22:0{1,2}$/i }
          },
        { "hours.Wednesday":  { $regex: /-22:0{1,2}$/i }
          },
        { "hours.Thursday":   { $regex: /-22:0{1,2}$/i }
          },
        { "hours.Friday":     { $regex: /-22:0{1,2}$/i }
          }
      ]
    }
  },
```

```
  {
    $project: {
      _id: 0,
      name: 1,
      city: 1,
      stars: 1,
      review_count: 1,
      categories: 1,
      "hours.Monday": 1,
      "hours.Tuesday": 1,
      "hours.Wednesday": 1,
      "hours.Thursday": 1,
      "hours.Friday": 1,
      "attributes.Open24Hours": 1
    }
  },
  {
    $sort: { stars: -1, review_count: -1 }
  }
]);
```

Outcome (partial):

```
{
  name: 'Deluxe Vehicle Detailing and Paint Correction',
  city: 'Belleair Bluffs',
  stars: 5,
  review_count: 6,
  attributes: {},
  categories: 'Auto Detailing, Automotive, Car Wash',
  hours: {
    Monday: '5:0-22:0',
    Tuesday: '5:0-22:0',
    Wednesday: '5:0-22:0',
    Thursday: '5:0-22:0',
    Friday: '5:0-22:0'
  }
}
```

```
{
  name: 'Power Up Auto Inc',
  city: 'Seminole',
  stars: 4.5,
  review_count: 7,
```

```
  attributes: {},
  categories: 'Auto Parts & Supplies, Automotive, Auto
     Repair, Auto Glass Services',
  hours: {
    Monday: '5:0-4:45',
    Tuesday: '5:0-4:45',
    Wednesday: '5:0-4:45',
    Thursday: '6:0-22:0',
    Friday: '5:0-4:45'
  }
}
```

Extra In this plot we show, for each business that was returned, the opening and closing hours day by day. Since we did not ask for Saturday and Sunday their columns are empty.

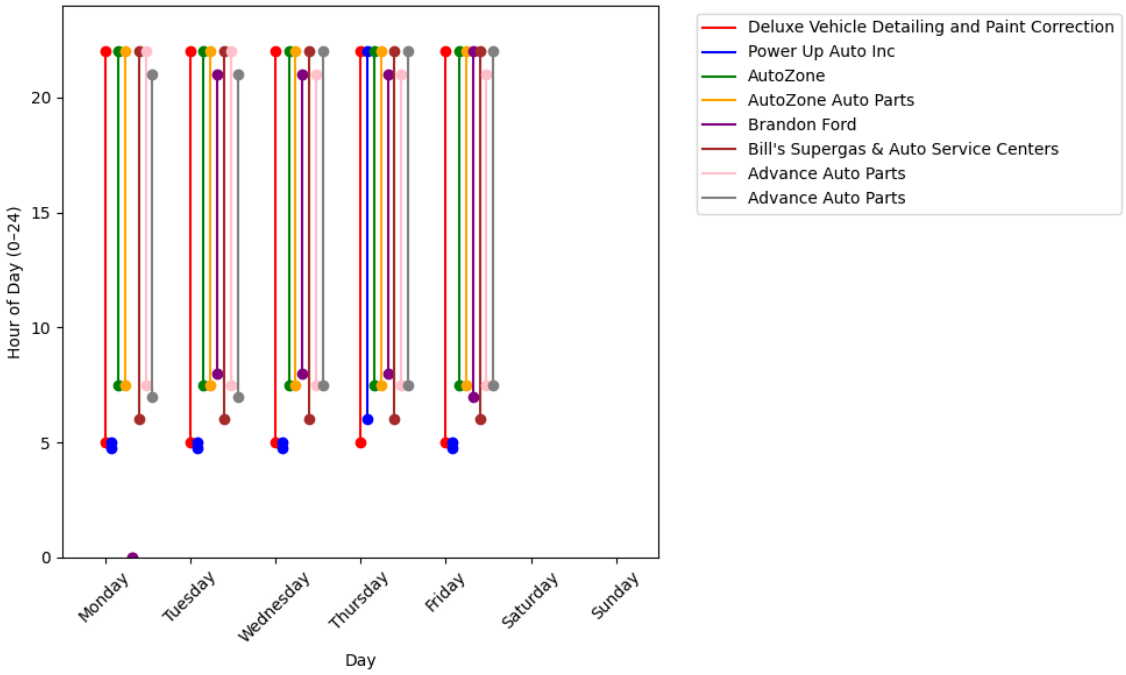

Figure 3.9: Opening Hours day by day

3.2. Problem 2 - *Queries with Elasticsearch*

0. GET /anime/_mapping

```
GET /anime/_mapping
```

0. PUT /anime/_mapping

```
PUT /anime/_mapping
{
  "properties": {
    "Genres": {
      "type": "text",
      "fielddata": true
    }
  }
}
```

Query 1: Calculating the Average Anime Score by Studio and Production Status

Description: This query aggregates anime data by studio and production status (e.g., *Finished Airing* or *Currently Airing*) using Elasticsearch aggregations. For each combination of studio and status, it calculates the average anime score. The goal is to analyze how different studios perform based on the status of their anime, providing an overview of their average performance in various production stages.

Key Components:

- **Terms Aggregation:** Groups anime by `Studios` and subsequently by `Status`.
- **Avg Aggregation:** Computes the average of the `Score` field for each group.

Query:

```
GET /anime/_search
{
    "size" : 0,
    "aggs" : {
      "anime_per_studios" : {
        "terms" : {"field" : "Studios"},
        "aggs" : {
          "anime_per_status" : {
            "terms" : {"field" : "Status"},
            "aggs" : {
```

```
            "avarage_score" : {
              "avg" : {
                "field" : "Score"
              }
            }
          }
        }
      }
    }
}
```

Outcome:

```
{
  "took": 1,
  "timed_out": false,
  "_shards": {
    "total": 1,
    "successful": 1,
    "skipped": 0,
    "failed": 0
  },
  "hits": {
    "total": {
      "value": 10000,
      "relation": "gte"
    },
    "max_score": null,
    "hits": []
  },
  "aggregations": {
    "anime_per_studios": {
      "doc_count_error_upper_bound": 0,
      "sum_other_doc_count": 9017,
      "buckets": [
        {
          "key": "UNKNOWN",
          "doc_count": 3527,
          "anime_per_status": {
            "doc_count_error_upper_bound": 0,
            "sum_other_doc_count": 0,
            "buckets": [
              {
```

```
                "key": "Finished Airing",
                "doc_count": 3518,
                "avarage_score": {
                  "value": 5.564070494599204
                }
              },
              {
                "key": "Currently Airing",
                "doc_count": 9,
                "avarage_score": {
                  "value": 6.448888888888889
                }
              }
            ]
          }
        },
        {
          "key": "Toei Animation",
          "doc_count": 693,
          "anime_per_status": {
            "doc_count_error_upper_bound": 0,
            "sum_other_doc_count": 0,
            "buckets": [
              {
                "key": "Finished Airing",
                "doc_count": 689,
                "avarage_score": {
                  "value": 6.6223512336719885
                }
              },
              {
                "key": "Currently Airing",
                "doc_count": 4,
                "avarage_score": {
                  "value": 6.8525
                }
              }
            ]
          }
        },
        ...
      ]
    }
  }
```

```
}
```

Query 2: Temporal Analysis of Anime Scores Using SQL in Elasticsearch

Description: This query leverages Elasticsearch's SQL capabilities to perform a temporal analysis of anime. It calculates the average scores of anime released each premier year, excluding those without assigned scores. Results are sorted in descending order based on the average scores, allowing identification of the best-performing anime years.

Key Components:

- **SELECT and AVG:** Selects the `Premiered` field and computes the average `Score`.
- **WHERE Clause:** Filters anime with non-null `Score`.
- **GROUP BY and ORDER BY:** Groups results by `Premiered` year and sorts by average score.

Query:

```
POST /_sql?format=txt
{
  "query": """
    SELECT
      Premiered,
      AVG(Score) AS average_score
    FROM
      anime
    WHERE
      Score IS NOT NULL
    GROUP BY
      Premiered
    ORDER BY
      average_score DESC
  """
}
```

Outcome:

Premiered	Average Score
summer 1991	7.83
winter 1985	7.67
summer 1996	7.6000000000000005

Premiered	Average Score
winter 1993	7.45
winter 1974	7.37
spring 1996	7.369230769230769
winter 1995	7.297999999999999
spring 1990	7.283333333333334
fall 1978	7.265000000000001
summer 1995	7.26
spring 1970	7.246666666666666
winter 2009	7.232083333333333
fall 2022	7.2063265306122455
spring 1993	7.191111111111111
fall 2012	7.173720930232559
fall 2008	7.17
winter 1996	7.165714285714286
summer 1985	7.165
fall 2005	7.148823529411764
spring 2007	7.132941176470588
spring 1989	7.1325
fall 1995	7.13111111111111
fall 2014	7.1230612244897955
fall 1996	7.109999999999999
spring 2008	7.105333333333333
fall 1993	7.101428571428571
spring 2005	7.099677419354839
spring 2023	7.097142857142857
spring 2010	7.096363636363637
winter 1986	7.0840000000000005
winter 1987	7.083333333333333
spring 2022	7.081
winter 1963	7.08
winter 1994	7.08
fall 2011	7.074117647058824
winter 2023	7.071666666666666
winter 2010	7.069999999999999
fall 2009	7.068648648648649
summer 2007	7.0672727272727265
spring 1995	7.064444444444444
fall 2007	7.059142857142857
summer 2021	7.047837837837838
spring 2009	7.04047619047619
fall 2006	7.0389795918367355
fall 1991	7.036363636363637
fall 2010	7.035483870967742

Premiered	Average Score
winter 2012	7.024333333333333
fall 2003	7.013939393939395
summer 2012	7.002758620689655
summer 2009	7.0009523809523815
fall 1982	6.9975
fall 1990	6.996666666666667
spring 2021	6.995306122448979
spring 2012	6.987551020408163
winter 2018	6.9868
winter 2008	6.9824
spring 2002	6.980769230769231
fall 2013	6.9746938775510205
spring 1968	6.963333333333334
spring 2006	6.963174603174603
summer 2013	6.957674418604651
fall 2004	6.952619047619048
fall 1999	6.950689655172414
fall 2000	6.946842105263158
fall 2020	6.9418
spring 1985	6.941111111111111
winter 1970	6.94
winter 1992	6.9383333333333335
winter 2011	6.9377272727272725
spring 1978	6.932857142857143
spring 1997	6.922941176470588
summer 2008	6.922857142857143
fall 1994	6.92125
fall 2002	6.92
spring 2004	6.916842105263157
fall 1980	6.90625
summer 1990	6.904999999999999
summer 1980	6.903333333333333
winter 2020	6.899545454545454
fall 2021	6.897826086956522
spring 2003	6.8832258064516125
fall 1992	6.88
summer 2011	6.879666666666667
spring 1986	6.877777777777777
spring 2015	6.876153846153846
summer 1983	6.871666666666667
summer 2022	6.8704347826086964
spring 2011	6.867192982456141
spring 1987	6.862222222222222

Premiered	Average Score
winter 2001	6.862
winter 2021	6.861929824561403
fall 2019	6.85530612244898
summer 2014	6.854090909090909
spring 1969	6.853333333333334
winter 2007	6.8508000000000004
winter 1997	6.85
spring 2000	6.848999999999999
winter 2015	6.848
summer 2010	6.845416666666666
winter 1978	6.843333333333334
spring 1998	6.840000000000001
summer 2005	6.837999999999999
fall 1986	6.837777777777777
fall 1983	6.833333333333333
summer 1977	6.833333333333333
summer 2018	6.8294
spring 1994	6.8244444444444445
fall 1989	6.820714285714286
spring 2014	6.8188524590163935
winter 2022	6.818250000000001
spring 2020	6.815434782608696
summer 2019	6.811842105263159
winter 1990	6.8100000000000005
summer 1998	6.797142857142857
summer 2016	6.791836734693878
fall 1977	6.790909090909091
fall 1997	6.783846153846154
summer 2000	6.782000000000001
spring 1992	6.7725
winter 1991	6.77
summer 1978	6.7675
spring 2019	6.76695652173913
summer 2015	6.764583333333333
spring 2018	6.75514705882353
spring 1981	6.750909090909091
summer 2004	6.746842105263158
winter 2002	6.743125
fall 1981	6.740714285714286
fall 2018	6.739090909090908
spring 2001	6.735294117647059
spring 1983	6.732857142857143
winter 1983	6.728

Premiered	Average Score
winter 2019	6.727560975609756
spring 2013	6.725531914893618
summer 2006	6.721739130434782
fall 1968	6.72
winter 2016	6.719333333333333
spring 1984	6.717272727272728
fall 1985	6.716666666666666
fall 1969	6.715999999999999
fall 1979	6.71375
fall 2001	6.7118518518518515
winter 2000	6.708181818181817
summer 2002	6.703636363636363
winter 2005	6.695416666666667
winter 1967	6.695
winter 1998	6.692727272727273
summer 1984	6.6899999999999995
winter 2004	6.689047619047619
spring 2016	6.689016393442622
winter 1981	6.683333333333334
spring 1999	6.6825
fall 2017	6.6798245614035086
fall 1975	6.676666666666667
spring 1991	6.671000000000001
fall 2015	6.668166666666667
fall 2016	6.667391304347826
winter 1979	6.6625
spring 1979	6.661
summer 2017	6.657659574468084
winter 2014	6.654565217391305
fall 1998	6.650833333333334
spring 1988	6.650666666666667
winter 2013	6.649210526315789
fall 1965	6.63
spring 1975	6.628333333333334
fall 1972	6.626666666666666
fall 1988	6.62625
winter 1999	6.625333333333333
fall 1987	6.613333333333333
winter 2006	6.601785714285714
winter 2017	6.600425531914894
summer 1982	6.6
fall 1984	6.597272727272728
summer 1999	6.59235294117647

Premiered	Average Score
summer 1981	6.59
summer 1989	6.59
spring 1980	6.583333333333333
spring 2017	6.58265625
winter 1976	6.573333333333333
fall 1974	6.57
winter 1988	6.568333333333334
winter 1982	6.5649999999999995
spring 1973	6.5600000000000005
spring 1967	6.55
winter 1977	6.55
summer 2001	6.526923076923077
winter 1980	6.50875
spring 1982	6.506363636363636
summer 1972	6.49
summer 1994	6.486666666666667
summer 2020	6.48
winter 1984	6.475
summer 2003	6.474705882352941
summer 1997	6.4725
winter 1968	6.470000000000001
winter 1972	6.456666666666667
fall 1976	6.455555555555556
winter 2003	6.453
summer 1992	6.445
winter 1969	6.4399999999999995
winter 1973	6.425
winter 1989	6.3925
winter 1966	6.39
fall 1973	6.362222222222223
spring 1972	6.3500000000000005
spring 1971	6.345000000000001
spring 1976	6.345
summer 1988	6.343333333333334
fall 1971	6.3425
fall 1970	6.34
summer 1986	6.32
spring 1974	6.3
winter 1975	6.293333333333333
summer 1987	6.234999999999999
UNKNOWN	6.193395001334163
summer 1975	6.19
summer 1976	6.109999999999999

Premiered	Average Score
summer 1964	6.0649999999999995
fall 1966	6.06
summer 1965	6.06
summer 1979	6.029999999999999
fall 1967	5.98
winter 1971	5.975
spring 1977	5.973333333333334
spring 1961	5.97
winter 1965	5.97
fall 1963	5.9625
summer 1974	5.96
summer 1962	5.86
spring 1966	5.85
winter 1964	5.8
spring 1965	5.77
summer 1966	5.615

Query 3: Identifying the Top 10 Anime per Year with Scores Greater Than 7

Description: This query filters anime with scores greater than or equal to 7 and, for each premier year, selects the top 10 highest-scoring anime. It uses aggregations to group anime by premier date and applies an internal ranking to extract the top performers annually, providing insights into the most successful yearly productions.

Key Components:

- **Range Query:** Filters anime with `Score` ≥ 7.
- **Terms Aggregation:** Groups by `Premiered` field.
- **Top Hits Aggregation:** Selects the top 10 anime sorted by descending `Score` in each group.

Query:

```
POST /anime/_search
{
  "size": 0,
  "query": {
    "range": {
      "Score": {
        "gte": 7.0
      }
    }
```

```
  },
  "aggs": {
    "aired_same_day": {
      "terms": {
        "field": "Premiered",
        "size": 1000
      },
      "aggs": {
        "anime_ranked": {
          "top_hits": {
            "size": 10,
            "_source": ["Name", "Score", "Premiered"],
            "sort": [
              {
                "Score": {
                  "order": "desc"
                }
              }
            ]
          }
        }
      }
    }
  }
}
```

Outcome:

```
{
  "took": 20,
  "timed_out": false,
  "_shards": {
    "total": 1,
    "successful": 1,
    "skipped": 0,
    "failed": 0
  },
  "hits": {
    "total": {
      "value": 10000,
      "relation": "gte"
    },
    "max_score": null,
    "hits": []
```

```
  },
  "aggregations": {
    "studios": {
      "doc_count_error_upper_bound": 0,
      "sum_other_doc_count": 353,
      "buckets": [
        {
          "key": "UNKNOWN",
          "doc_count": 3527,
          "finished_anime": {
            "doc_count": 3518
          },
          "not_finished_anime": {
            "doc_count": 9
          },
          "percentage_finished": {
            "value": 99.74482563084774
          }
        },
        {
          "key": "Toei Animation",
          "doc_count": 693,
          "finished_anime": {
            "doc_count": 689
          },
          "not_finished_anime": {
            "doc_count": 4
          },
          "percentage_finished": {
            "value": 99.42279942279943
          }
        },
        {
          "key": "Sunrise",
          "doc_count": 506,
          "finished_anime": {
            "doc_count": 505
          },
          "not_finished_anime": {
            "doc_count": 1
          },
          "percentage_finished": {
            "value": 99.80237154150198
          }
```

```
        },
        {
          "key": "J.C.Staff",
          "doc_count": 363,
          "finished_anime": {
            "doc_count": 361
          },
          "not_finished_anime": {
            "doc_count": 2
          },
          "percentage_finished": {
            "value": 99.44903581267218
          }
        },
        {
          "key": "Madhouse",
          "doc_count": 322,
          "finished_anime": {
            "doc_count": 321
          },
          "not_finished_anime": {
            "doc_count": 1
          },
          "percentage_finished": {
            "value": 99.68944099378882
          }
        },
        {
          "key": "TMS Entertainment",
          "doc_count": 269,
          "finished_anime": {
            "doc_count": 267
          },
          "not_finished_anime": {
            "doc_count": 2
          },
          "percentage_finished": {
            "value": 99.25650557620817
          }
        },
        {
          "key": "Studio Deen",
          "doc_count": 267,
          "finished_anime": {
```

```
            "doc_count": 266
          },
          "not_finished_anime": {
            "doc_count": 1
          },
          "percentage_finished": {
            "value": 99.625468164794
          }
        },
        {
          "key": "Production I.G",
          "doc_count": 252,
          "finished_anime": {
            "doc_count": 251
          },
          "not_finished_anime": {
            "doc_count": 1
          },
          "percentage_finished": {
            "value": 99.60317460317461
          }
        }
      ]
    }
  }
}
```


Query 4: Percentage of Completed vs. Ongoing Anime by Studio

Description: This query examines studio productivity by analyzing the percentage of completed (*Finished Airing*) versus ongoing (*Currently Airing*) anime. By limiting the analysis to 1,000 studios, it calculates the completion rate for each studio.

Key Components:

- **Terms Aggregation:** Groups anime by `Studios`, limited to 1,000 results.
- **Filter Aggregations:** Counts finished and ongoing anime for each studio.
- **Bucket Script Aggregation:** Calculates the percentage of completed anime relative to the total anime for each studio.

Query:

```
GET /anime/_search
{
  "size": 0,
  "aggs": {
    "studios": {
      "terms": {
        "field": "Studios",
        "size": 1000
      },
      "aggs": {
        "finished_anime": {
          "filter": {
            "term": {
              "Status": "Finished Airing"
            }
          }
        },
        "not_finished_anime": {
          "filter": {
            "term": {
              "Status": "Currently Airing"
            }
          }
        },
        "percentage_finished": {
          "bucket_script": {
            "buckets_path": {
              "finished_count": "finished_anime._count",
              "not_finished_count": "not_finished_anime.
                 _count"
            },
            "script": "params.finished_count / (params.
               finished_count + params.not_finished_count
               ) * 100"
          }
        }
      }
    }
  }
}
```

Outcome:

```
{
```

```
"took": 62,
"timed_out": false,
"_shards": {
  "total": 1,
  "successful": 1,
  "skipped": 0,
  "failed": 0
},
"hits": {
  "total": {
    "value": 4284,
    "relation": "eq"
  },
  "max_score": null,
  "hits": []
},
"aggregations": {
  "aired_same_day": {
    "doc_count_error_upper_bound": 0,
    "sum_other_doc_count": 0,
    "buckets": [
      {
        "key": "UNKNOWN",
        "doc_count": 2286,
        "anime_ranked": {
          "hits": {
            "total": {
              "value": 2286,
              "relation": "eq"
            },
            "max_score": null,
            "hits": [
              {
                "_index": "anime",
                "_id": "g-L85JMBSCkiqoTYc34J",
                "_score": null,
                "_source": {
                  "Name": "Shingeki no Kyojin: The
                     Final Season - Kanketsu-hen",
                  "Premiered": "UNKNOWN",
                  "Score": 9.05
                },
                "sort": [
                  9.05
```

```
      ]
    },
    {
      "_index": "anime",
      "_id": "2uL85JMBSCkiqoTYaXbs",
      "_score": null,
      "_source": {
        "Name": "Gintama: The Final",
        "Premiered": "UNKNOWN",
        "Score": 9.04
      },
      "sort": [
        9.04
      ]
    },
    {
      "_index": "anime",
      "_id": "PBH85JMBGbVWSHHuJJpR",
      "_score": null,
      "_source": {
        "Name": "Ginga Eiyuu Densetsu",
        "Premiered": "UNKNOWN",
        "Score": 9.02
      },
      "sort": [
        9.02
      ]
    },
    {
      "_index": "anime",
      "_id": "puL85JMBSCkiqoTYTG-R",
      "_score": null,
      "_source": {
        "Name": "Koe no Katachi",
        "Premiered": "UNKNOWN",
        "Score": 8.94
      },
      "sort": [
        8.94
      ]
    },
    {
      "_index": "anime",
      "_id": "YeL85JMBSCkiqoTYQmhW",
```

```
      "_score": null,
      "_source": {
        "Name": "Gintama Movie 2: Kanketsu-
           hen - Yorozuya yo Eien Nare",
        "Premiered": "UNKNOWN",
        "Score": 8.91
      },
      "sort": [
        8.91
      ]
    },
    {
      "_index": "anime",
      "_id": "SuL85JMBSCkiqoTYX3SR",
      "_score": null,
      "_source": {
        "Name": "Violet Evergarden Movie",
        "Premiered": "UNKNOWN",
        "Score": 8.9
      },
      "sort": [
        8.9
      ]
    },
    {
      "_index": "anime",
      "_id": "DuL85JMBSCkiqoTYc38J",
      "_score": null,
      "_source": {
        "Name": "Kaguya-sama wa Kokurasetai:
           First Kiss wa Owaranai",
        "Premiered": "UNKNOWN",
        "Score": 8.87
      },
      "sort": [
        8.87
      ]
    },
    {
      "_index": "anime",
      "_id": "ehH85JMBGbVWSHHuU6Ue",
      "_score": null,
      "_source": {
        "Name": "Kimi no Na wa.",
```

```
                    "Premiered": "UNKNOWN",
                    "Score": 8.85
                  },
                  "sort": [
                    8.85
                  ]
                },
                {
                  "_index": "anime",
                  "_id": "duL85JMBSCkiqoTYeoDg",
                  "_score": null,
                  "_source": {
                    "Name": "Idol",
                    "Premiered": "UNKNOWN",
                    "Score": 8.85
                  },
                  "sort": [
                    8.85
                  ]
                }
              ]
            }
          }
        },
        ...
      ]
    }
  }
}
```


Query 5: Percentage of High-Scoring Anime with Death Themes Produced by Bandai

Description: This query identifies anime with scores above 8 that mention "death" in their synopsis and calculates the percentage of such anime produced by Bandai Visual relative to the total. This evaluates Bandai's influence in a specific thematic niche.

Key Components:

- **Bool Query:** Filters for `Score` > 8 and presence of the word "death" in the `Synopsis`.
- **Filters Aggregation:** Creates separate filters for anime produced by Bandai

Visual and the total dataset.

- **Value Count and Bucket Script:** Counts anime in each filter and calculates the percentage of Bandai-produced anime within the selected criteria.

Query:

```
GET /anime/_search
{
  "size": 0,
  "query": {
    "bool": {
      "must": [
        { "range": { "Score": { "gt": 8 } } },
        { "match": { "Synopsis": "death" } }
      ]
    }
  },
  "aggs": {
    "all_animes": {
      "filters": {
        "filters": {
          "death_bandai": {
            "bool": {
              "must": [
                { "range": { "Score": { "gt": 8 } } },
                { "match": { "Synopsis": "death" } },
                { "match": { "Studios": "Bandai Visual" }
                    }
              ]
            }
          },
          "total_death": {
            "bool": {
              "must": [
                { "range": { "Score": { "gt": 8 } } },
                { "match": { "Synopsis": "death" } }
              ]
            }
          }
        }
      },
      "aggs": {
        "death_bandai_count": {
          "filter": {
```

```
            "bool": {
              "must": [
                { "range": { "Score": { "gt": 8 } } },
                { "match": { "Synopsis": "death" } },
                { "match": { "Studios": "Bandai Visual" }
                    }
              ]
            }
          },
          "aggs": {
            "count_bandai": {
              "value_count": {
                "field": "Score"
              }
            }
          }
        },
        "total_death_count": {
          "filter": {
            "bool": {
              "must": [
                { "range": { "Score": { "gt": 8 } } },
                { "match": { "Synopsis": "death" } }
              ]
            }
          },
          "aggs": {
            "count_total": {
              "value_count": {
                "field": "Score"
              }
            }
          }
        },
        "percentage": {
          "bucket_script": {
            "buckets_path": {
              "bandai_count": "death_bandai_count>
                 count_bandai",
              "total_count": "total_death_count>
                 count_total"
            },
            "script": "params.bandai_count / params.
               total_count * 100"
```

```
          }
        }
      }
    }
  }
}
```

Outcome:

```
{
  "took": 5,
  "timed_out": false,
  "_shards": {
    "total": 1,
    "successful": 1,
    "skipped": 0,
    "failed": 0
  },
  "hits": {
    "total": {
      "value": 45,
      "relation": "eq"
    },
    "max_score": null,
    "hits": []
  },
  "aggregations": {
    "all_animes": {
      "buckets": {
        "death_bandai": {
          "doc_count": 0,
          "death_bandai_count": {
            "doc_count": 0,
            "count_bandai": {
              "value": 0
            }
          },
          "total_death_count": {
            "doc_count": 0,
            "count_total": {
              "value": 0
            }
          }
        },
```

```
        "total_death": {
          "doc_count": 45,
          "death_bandai_count": {
            "doc_count": 0,
            "count_bandai": {
              "value": 0
            }
          },
          "total_death_count": {
            "doc_count": 45,
            "count_total": {
              "value": 45
            }
          },
          "percentage": {
            "value": 0
          }
        }
      }
    }
  }
}
```

Query 6: Categorizing Anime by Member Counts with Top 10 per Range

Description: This query segments anime into ranges based on their member counts (e.g., less than 100,000, between 100,001 and 500,000, etc.) and selects the top 10 highest-scoring anime within each range. The goal is to analyze how member engagement correlates with perceived anime quality.

Key Components:

- **Range Aggregation:** Defines distinct ranges for member counts.
- **Top Hits Aggregation:** Extracts the top 10 anime with the highest scores within each range.
- **Sort and Source Filtering:** Orders anime by descending `Score` and selects relevant fields like `Name` and `Score`.

Query:

```
GET /anime/_search
{
  "size": 0,
```

```
  "query": {
    "match_all": {}
  },
  "aggs": {
    "members_ranges": {
      "range": {
        "field": "Members",
        "ranges": [
          { "to": 100000 },
          { "from": 100001, "to": 500000 },
          { "from": 500001, "to": 1000000 },
          { "from": 1000001 }
        ]
      },
      "aggs": {
        "top_scores": {
          "top_hits": {
            "size": 10,
            "sort": [
              { "Score": { "order": "desc" } }
            ],
            "_source": ["Name", "Score"]
          }
        }
      }
    }
  }
}
```

Outcome:

```
{
  "took": 17,
  "timed_out": false,
  "_shards": {
    "total": 1,
    "successful": 1,
    "skipped": 0,
    "failed": 0
  },
  "hits": {
    "total": {
      "value": 10000,
      "relation": "gte"
```

```
    },
    "max_score": null,
    "hits": []
  },
  "aggregations": {
    "members_ranges": {
      "buckets": [
        {
          "key": "*-100000.0",
          "to": 100000,
          "doc_count": 13769,
          "top_scores": {
            "hits": {
              "total": {
                "value": 13769,
                "relation": "eq"
              },
              "max_score": null,
              "hits": [
                {
                  "_index": "anime",
                  "_id": "duL85JMBSCkiqoTYeoDg",
                  "_score": null,
                  "_source": {
                    "Name": "Idol",
                    "Score": 8.85
                  },
                  "sort": [
                    8.85
                  ]
                },
                {
                  "_index": "anime",
                  "_id": "1OL85JMBSCkiqoTYaXjs",
                  "_score": null,
                  "_source": {
                    "Name": "Kingdom 3rd Season",
                    "Score": 8.81
                  },
                  "sort": [
                    8.81
                  ]
                },
                {
```

```
      "_index": "anime",
      "_id": "F-L85JMBSCkiqoTYcH2r",
      "_score": null,
      "_source": {
        "Name": "Kingdom 4th Season",
        "Score": 8.75
      },
      "sort": [
        8.75
      ]
    },
    {
      "_index": "anime",
      "_id": "suL85JMBSCkiqoTYL13a",
      "_score": null,
      "_source": {
        "Name": "Ashita no Joe 2",
        "Score": 8.71
      },
      "sort": [
        8.71
      ]
    },
    {
      "_index": "anime",
      "_id": "rxH85JMBGbVWSHHubatD",
      "_score": null,
      "_source": {
        "Name": "The First Slam Dunk",
        "Score": 8.62
      },
      "sort": [
        8.62
      ]
    },
    {
      "_index": "anime",
      "_id": "buL85JMBSCkiqoTYaXjs",
      "_score": null,
      "_source": {
        "Name": "Mo Dao Zu Shi: Wanjie Pian",
        "Score": 8.57
      },
      "sort": [
```

```
            8.57
          ]
        },
        {
          "_index": "anime",
          "_id": "z-L85JMBSCkiqoTYaXjs",
          "_score": null,
          "_source": {
            "Name": " S h o u j o Kageki  Revue
               Starlight Movie",
            "Score": 8.55
          },
          "sort": [
            8.55
          ]
        },
        {
          "_index": "anime",
          "_id": "juL85JMBSCkiqoTYcHyr",
          "_score": null,
          "_source": {
            "Name": "Karakai Jouzu no Takagi-san
               Movie",
            "Score": 8.53
          },
          "sort": [
            8.53
          ]
        },
        {
          "_index": "anime",
          "_id": "xuL85JMBSCkiqoTYc38J",
          "_score": null,
          "_source": {
            "Name": "Road of Naruto",
            "Score": 8.49
          },
          "sort": [
            8.49
          ]
        },
        {
          "_index": "anime",
          "_id": "QhH85JMBGbVWSHHubatD",
```

```
                "_score": null,
                "_source": {
                  "Name": "Gintama: The Semi-Final",
                  "Score": 8.47
                },
                "sort": [
                  8.47
                ]
              }
            ]
          }
        }
      },
      {
        "key": "100001.0-500000.0",
        "from": 100001,
        "to": 500000,
        "doc_count": 1512,
        "top_scores": {
          "hits": {
            "total": {
              "value": 1512,
              "relation": "eq"
            },
            "max_score": null,
            "hits": [
              {
                "_index": "anime",
                "_id": "D-L85JMBSCkiqoTYaXrt",
                "_score": null,
                "_source": {
                  "Name": "Bleach: Sennen Kessen-hen",
                  "Score": 9.07
                },
                "sort": [
                  9.07
                ]
              },
              {
                "_index": "anime",
                "_id": "g-L85JMBSCkiqoTYc34J",
                "_score": null,
                "_source": {
                  "Name": "Shingeki no Kyojin: The
```

```
                    Final Season - Kanketsu-hen",
                  "Score": 9.05
                },
                "sort": [
                  9.05
                ]
              },
              {
                "_index": "anime",
                "_id": "2uL85JMBSCkiqoTYaXbs",
                "_score": null,
                "_source": {
                  "Name": "Gintama: The Final",
                  "Score": 9.04
                },
                "sort": [
                  9.04
                ]
              },
              {
                "_index": "anime",
                "_id": "a-L85JMBSCkiqoTYQmhW",
                "_score": null,
                "_source": {
                  "Name": "Gintama': Enchousen",
                  "Score": 9.03
                },
                "sort": [
                  9.03
                ]
              },
              {
                "_index": "anime",
                "_id": "PBH85JMBGbVWSHHuJJpR",
                "_score": null,
                "_source": {
                  "Name": "Ginga Eiyuu Densetsu",
                  "Score": 9.02
                },
                "sort": [
                  9.02
                ]
              },
              {
```

```
      "_index": "anime",
      "_id": "wBH85JMBGbVWSHHubapD",
      "_score": null,
      "_source": {
        "Name": "Fruits Basket: The Final",
        "Score": 9
      },
      "sort": [
        9
      ]
    },
    {
      "_index": "anime",
      "_id": "SxH85JMBGbVWSHHuWagR",
      "_score": null,
      "_source": {
        "Name": "Gintama.",
        "Score": 8.98
      },
      "sort": [
        8.98
      ]
    },
    {
      "_index": "anime",
      "_id": "3xH85JMBGbVWSHHuWakS",
      "_score": null,
      "_source": {
        "Name": "3-gatsu no Lion 2nd Season",
        "Score": 8.93
      },
      "sort": [
        8.93
      ]
    },
    {
      "_index": "anime",
      "_id": "YeL85JMBSCkiqoTYQmhW",
      "_score": null,
      "_source": {
        "Name": "Gintama Movie 2: Kanketsu-
          hen - Yorozuya yo Eien Nare",
        "Score": 8.91
      },
```

```
                "sort": [
                  8.91
                ]
              },
              {
                "_index": "anime",
                "_id": "9RH85JMBGbVWSHHuWakS",
                "_score": null,
                "_source": {
                  "Name": "Owarimonogatari 2nd Season",
                  "Score": 8.88
                },
                "sort": [
                  8.88
                ]
              }
            ]
          }
        }
      },
      {
        "key": "500001.0-1000000.0",
        "from": 500001,
        "to": 1000000,
        "doc_count": 265,
        "top_scores": {
          "hits": {
            "total": {
              "value": 265,
              "relation": "eq"
            },
            "max_score": null,
            "hits": [
              {
                "_index": "anime",
                "_id": "vuL85JMBSCkiqoTYTG-R",
                "_score": null,
                "_source": {
                  "Name": "Gintama ",
                  "Score": 9.06
                },
                "sort": [
                  9.06
                ]
```

```
    },
    {
      "_index": "anime",
      "_id": "BRH85JMBGbVWSHHubatD",
      "_score": null,
      "_source": {
        "Name": "Kaguya-sama wa Kokurasetai:
           Ultra Romantic",
        "Score": 9.05
      },
      "sort": [
        9.05
      ]
    },
    {
      "_index": "anime",
      "_id": "CuL85JMBSCkiqoTYO2Qs",
      "_score": null,
      "_source": {
        "Name": "Gintama'",
        "Score": 9.04
      },
      "sort": [
        9.04
      ]
    },
    {
      "_index": "anime",
      "_id": "9uL85JMBSCkiqoTYc34J",
      "_score": null,
      "_source": {
        "Name": "\"Oshi no Ko\"",
        "Score": 8.98
      },
      "sort": [
        8.98
      ]
    },
    {
      "_index": "anime",
      "_id": "SuL85JMBSCkiqoTYX3SR",
      "_score": null,
      "_source": {
        "Name": "Violet Evergarden Movie",
```

```
            "Score": 8.9
          },
          "sort": [
            8.9
          ]
        },
        {
          "_index": "anime",
          "_id": "ZuL85JMBSCkiqoTYQmlX",
          "_score": null,
          "_source": {
            "Name": "Monogatari Series: Second
               Season",
            "Score": 8.77
          },
          "sort": [
            8.77
          ]
        },
        {
          "_index": "anime",
          "_id": "RhH85JMBGbVWSHHuJJhO",
          "_score": null,
          "_source": {
            "Name": "Hajime no Ippo",
            "Score": 8.76
          },
          "sort": [
            8.76
          ]
        },
        {
          "_index": "anime",
          "_id": "GuL85JMBSCkiqoTYcH2r",
          "_score": null,
          "_source": {
            "Name": "Mob Psycho 100 III",
            "Score": 8.72
          },
          "sort": [
            8.72
          ]
        },
        {
```

```
                    "_index": "anime",
                    "_id": "4eL85JMBGbVWSHHubbqT",
                    "_score": null,
                    "_source": {
                      "Name": "Hikaru no Go",
                      "Score": 8.72
                    },
                    "sort": [
                      8.72
                    ]
                  },
                  {
                    "_index": "anime",
                    "_id": "W9H85JMBGbVWSHHuWoZT",
                    "_score": null,
                    "_source": {
                      "Name": "Gintama: Shirogane no
                         Tamashii-hen",
                      "Score": 8.7
                    },
                    "sort": [
                      8.7
                    ]
                  }
                ]
              }
            }
          }
        ]
      }
    }
}
```


Query 7: Total Members and Leading Studio by Genre

Description: This query groups anime by genre, calculates the total number of members for each genre, and identifies the studio producing the most anime within each genre. Studios marked as *UNKNOWN* are excluded, highlighting the most influential contributors.

Key Components:

- **Terms Aggregation:** Groups anime by `Genres`.
- **Sum Aggregation:** Calculates the total number of members per genre.

- **Filter Aggregation:** Excludes studios labeled as `UNKNOWN`.
- **Nested Terms Aggregation:** Identifies the studio with the most anime produced for each genre.

Query:

```
GET /anime/_search
{
  "size": 0,
  "aggregations": {
    "genres": {
      "terms": {
        "field": "Genres",
        "size": 30,
        "order": [
          {
            "total_members": "desc"
          }
        ]
      },
      "aggregations": {
        "total_members": {
          "sum": {
            "field": "Members"
          }
        },
        "studios": {
          "filter": {
            "bool": {
              "must_not": {
                "term": {
                  "Studios": "UNKNOWN"
                }
              }
            }
          },
          "aggregations": {
            "studio_count": {
              "terms": {
                "field": "Studios",
                "size": 1
              }
            }
          }
```

```
        }
      }
    }
  }
}
```

Outcome:


```
{
  "took": 102,
  "timed_out": false,
  "_shards": {
    "total": 1,
    "successful": 1,
    "skipped": 0,
    "failed": 0
  },
  "hits": {
    "total": {
      "value": 10000,
      "relation": "gte"
    },
    "max_score": null,
    "hits": []
  },
  "aggregations": {
    "genres": {
      "doc_count_error_upper_bound": 0,
      "sum_other_doc_count": 0,
      "buckets": [
        {
          "key": "action",
          "doc_count": 3956,
          "total_members": {
            "value": 416787879
          },
          "studios": {
            "doc_count": 3616,
            "studio_count": {
              "doc_count_error_upper_bound": 0,
              "sum_other_doc_count": 3326,
              "buckets": [
                {
                  "key": "Toei Animation",
```

```
                "doc_count": 290
              }
            ]
          }
        }
      },
      {
        "key": "comedy",
        "doc_count": 5069,
        "total_members": {
          "value": 308576900
        },
        "studios": {
          "doc_count": 4375,
          "studio_count": {
            "doc_count_error_upper_bound": 0,
            "sum_other_doc_count": 4135,
            "buckets": [
              {
                "key": "Toei Animation",
                "doc_count": 240
              }
            ]
          }
        }
      },
      {
        "key": "fantasy",
        "doc_count": 3290,
        "total_members": {
          "value": 304159927
        },
        "studios": {
          "doc_count": 2796,
          "studio_count": {
            "doc_count_error_upper_bound": 0,
            "sum_other_doc_count": 2489,
            "buckets": [
              {
                "key": "Toei Animation",
                "doc_count": 307
              }
            ]
          }
```

```
          }
        },
        {
          "key": "drama",
          "doc_count": 2198,
          "total_members": {
            "value": 224695262
          },
          "studios": {
            "doc_count": 1975,
            "studio_count": {
              "doc_count_error_upper_bound": 0,
              "sum_other_doc_count": 1853,
              "buckets": [
                {
                  "key": "Toei Animation",
                  "doc_count": 122
                }
              ]
            }
          }
        },
        {
          "key": "romance",
          "doc_count": 1831,
          "total_members": {
            "value": 221994257
          },
          "studios": {
            "doc_count": 1692,
            "studio_count": {
              "doc_count_error_upper_bound": 0,
              "sum_other_doc_count": 1598,
              "buckets": [
                {
                  "key": "J.C.Staff",
                  "doc_count": 94
                }
              ]
            }
          }
        },
        {
          "key": "adventure",
```

```
          "doc_count": 2672,
          "total_members": {
            "value": 192608191
          },
          "studios": {
            "doc_count": 2376,
            "studio_count": {
              "doc_count_error_upper_bound": 0,
              "sum_other_doc_count": 2096,
              "buckets": [
                {
                  "key": "Toei Animation",
                  "doc_count": 280
                }
              ]
            }
          }
        }
      ]
    }
  }
}
```

Query 8: Boosted Search for Finished Anime with Specific Themes

Description: This query searches for finished anime containing the words "crime" and "galaxy" in their synopsis. Additionally, it boosts scores for anime that include the words "humanity," "year," or "earth," prioritizing relevant themes in the search results.

Key Components:

- **Bool Query:** Combines must and should conditions.
- **Must Clauses:** Requires the presence of "crime" and "galaxy" in the synopsis and `Finished Airing` status.
- **Should Clause:** Boosts scores if "humanity," "year," or "earth" are present.
- **Boost:** Applies a boosting factor to enhance the relevance of specific themes.

Query:

```
GET /anime/_search
{
```

```
  "size": 100,
  "query": {
    "bool": {
      "must": [
        {
          "match": {
            "Synopsis": {
              "query": "crime galaxy",
              "operator": "and"
            }
          }
        },
        {
          "term": {
            "Status": "Finished Airing"
          }
        }
      ],
      "should": [
        {
          "match": {
            "Synopsis": {
              "query": "humanity year earth",
              "operator": "or"
            }
          }
        }
      ],
      "boost": 1.5
    }
  }
}
```

Outcome:

```
{
  "took": 3,
  "timed_out": false,
  "_shards": {
    "total": 1,
    "successful": 1,
    "skipped": 0,
    "failed": 0
  },
```

```
  "hits": {
    "total": {
      "value": 2,
      "relation": "eq"
    },
    "max_score": 18.772451,
    "hits": [
      {
        "_index": "anime",
        "_id": "VxH85JMBGbVWSHHuJJdM",
        "_score": 18.772451,
        "_source": {
          "Other name": "                              ",
          "Producers": "Bandai Visual",
          "Rating": "R - 17+ (violence & profanity)",
          "Studios": "Sunrise",
          "Aired": "Apr 3, 1998 to Apr 24, 1999",
          "Source": "Original",
          "Name": "Cowboy Bebop",
          "English name": "Cowboy Bebop",
          "Episodes": "26.0",
          "Genres": "Action, Award Winning, Sci-Fi",
          "anime_id": 1,
          "Premiered": "spring 1998",
          "Status": "Finished Airing",
          "Popularity": 43,
          "Image URL": "https://cdn.myanimelist.net/
             images/anime/4/19644.jpg",
          "Duration": "24 min per ep",
          "Rank": "41.0",
          "Synopsis": """Crime is timeless. By the year
             2071, humanity has expanded across the
             galaxy, filling the surface of other planets
              with settlements like those on Earth. These
              new societies are plagued by murder, drug
             use, and theft, and intergalactic outlaws
             are hunted by a growing number of tough
             bounty hunters.

Spike Spiegel and Jet Black pursue criminals throughout
   space to make a humble living. Beneath his goofy and
   aloof demeanor, Spike is haunted by the weight of his
   violent past. Meanwhile, Jet manages his own troubled
   memories while taking care of Spike and the Bebop,
```

```
   their ship. The duo is joined by the beautiful con
   artist Faye Valentine, odd child Edward Wong Hau
   Pepelu Tivrusky IV, and Ein, a bioengineered Welsh
   Corgi.

While developing bonds and working to catch a colorful
   cast of criminals, the Bebop crew's lives are
   disrupted by a menace from Spike's past. As a rival's
   maniacal plot continues to unravel, Spike must choose
   between life with his newfound family or revenge for
   his old wounds.""",
          "Score": 8.75,
          "Type": "TV",
          "Scored By": 914193,
          "Licensors": "Funimation, Bandai Entertainment
             ",
          "Favorites": 78525,
          "Members": 1771505
        }
      },
      {
        "_index": "anime",
        "_id": "3uL85JMBSCkiqoTYL1_b",
        "_score": 11.692469,
        "_source": {
          "Other name": "                        ",
          "Producers": "Pixy",
          "Rating": "Rx - Hentai",
          "Studios": "Anime Antenna Iinkai",
          "Aired": "Nov 24, 2007 to Sep 27, 2009",
          "Source": "Visual novel",
          "Name": "Soukou Kijo Iris",
          "English name": "UNKNOWN",
          "Episodes": "4.0",
          "Genres": "Hentai",
          "anime_id": 3707,
          "Premiered": "UNKNOWN",
          "Status": "Finished Airing",
          "Popularity": 7315,
          "Image URL": "https://cdn.myanimelist.net/
             images/anime/4/6481.jpg",
          "Duration": "27 min per ep",
          "Rank": "UNKNOWN",
          "Synopsis": """In the distant future, mankind
```

```
            has left Earth to reside among the vast
            reaches of space, joining the interspecies
            collective known as the Space Federation.
            Following this development, the Soviet
            Security Force (SSF) was formed to establish
             humanity's presence throughout the galaxy.
            Among the SSF's many divisions is the
            Special Armored Company (SAC) #101. Using
            powerful mobile humanoid weapons called "
            Armored Knights," these female enforcers
            dispense justice upon corrupt officials,
            crime lords, and other villainous forces
            threatening public order.

Soukou Kijo Iris follows Iris Rebel, a lieutenant of SAC
   #101, as her squad investigates reports of trafficking
    activities on the planet Left. After her Elven
   partner Mei Li Naceri suddenly goes missing in action,
    Iris narrowly escapes an ambush and loses
   consciousness. After being captured by Orcs and taken
   to an underground slave auction, she is thoroughly
   humiliated by gleeful bidders before being sold off
   for an exorbitant price to Bozuk, owner of the most
   renowned brothel in Dark Town. Now forced into a life
   of prostitution, Iris must attempt to seek out her
   partner before the pleasures brought by her clients
   drive her to the brink of insanity.""",
            "Score": 6.56,
            "Type": "OVA",
            "Scored By": 2210,
            "Licensors": "UNKNOWN",
            "Favorites": 42,
            "Members": 6246
          }
        }
      ]
    }
}
```

Query 9: Action/Drama Anime with Weighted Attributes

Description: This query selects up to 10 anime in the Action or Drama genres that contain "crime" or "galaxy" in their synopsis. It boosts scores for anime with

member counts between 10,000 and 800,000 and those that include the words "love" or "war." Results are grouped by studio, returning the top 50 studios.

Key Components:

- **Function Score Query:** Combines a base query with score-boosting functions based on specific conditions.
- **Bool Must Clauses:** Filters by genres and keywords in the synopsis.
- **Range and Match Filters:** Applies score increments based on `Favorites` count and additional keywords.
- **Aggregations:** Groups results by `Studios` and limits to the top 50 represented.

Query:

```
GET /anime/_search
{
  "size": 10,
  "query": {
    "function_score": {
      "query": {
        "bool": {
          "must": [
            {
              "match": {
                "Genres": {
                  "query": "Action, Drama",
                  "operator": "or"
                }
              }
            },
            {
              "match": {
                "Synopsis": {
                  "query": "crime galaxy",
                  "operator": "or"
                }
              }
            }
          ]
        }
      },
      "boost_mode": "sum",
      "functions": [
```

```
        {
          "filter": {
            "range": {
              "Favorites": {
                "gte": 10000,
                "lte": 800000
              }
            }
          },
          "weight": 2
        },
        {
          "filter": {
            "match": {
              "Synopsis": {
                "query": "love war",
                "operator": "or"
              }
            }
          },
          "weight": 1.5
        }
      ]
    }
  },
  "aggs": {
    "best_studio": {
      "terms": {
        "field": "Studios",
        "size": 50
      }
    }
  }
}
```

Outcome:

```
{
  "took": 4,
  "timed_out": false,
  "_shards": {
    "total": 1,
    "successful": 1,
    "skipped": 0,
```

```
    "failed": 0
  },
  "hits": {
    "total": {
      "value": 171,
      "relation": "eq"
    },
    "max_score": 11.100157,
    "hits": [
      {
        "_index": "anime",
        "_id": "TuL85JMBSCkiqoTYK1rt",
        "_score": 11.100157,
        "_source": {
          "Other name": "                ",
          "Producers": "Fuji TV, Tokuma Shoten",
          "Rating": "R - 17+ (violence & profanity)",
          "Studios": "Pierrot Plus",
          "Aired": "Dec 17, 1994 to Feb 21, 1995",
          "Source": "Manga",
          "Name": "Wild 7",
          "English name": "UNKNOWN",
          "Episodes": "2.0",
          "Genres": "Action, Drama, Mystery",
          "anime_id": 1959,
          "Premiered": "UNKNOWN",
          "Status": "Finished Airing",
          "Popularity": 9824,
          "Image URL": "https://cdn.myanimelist.net/
            images/anime/1164/134100.jpg",
          "Duration": "55 min per ep",
          "Rank": "7941.0",
          "Synopsis": """An elite police crime fighting
            team is formed using convicted criminals.
            Using their first hand knowledge of the
            criminal mindset, the Wild 7 are first sent
            after a gang of Bank robbers. Then in Wild
            Biker Knights they are sent after a huge
            crime syndicate that has many high powered
            government officials in its deep pockets.
(Source: ANN)""",
          "Score": 6.19,
          "Type": "OVA",
          "Scored By": 692,
```

```
          "Licensors": "Enoki Films, Urban Vision",
          "Favorites": 1,
          "Members": 2632
        }
      },
      {
        "_index": "anime",
        "_id": "PuL85JMBSCkiqoTYL2Dc",
        "_score": 10.5775385,
        "_source": {
          "Other name": "
                                                ",
          "Producers": "UNKNOWN",
          "Rating": "PG-13 - Teens 13 or older",
          "Studios": "Planet",
          "Aired": "Dec 30, 2006 to Apr 25, 2007",
          "Source": "Original",
          "Name": "Ginga Tetsudou Monogatari:
             Wasurerareta Toki no Wakusei",
          "English name": "The Galaxy Railways: A Letter
             from the Abandoned Planet",
          "Episodes": "4.0",
          "Genres": "Drama, Sci-Fi",
          "anime_id": 3854,
          "Premiered": "UNKNOWN",
          "Status": "Finished Airing",
          "Popularity": 9865,
          "Image URL": "https://cdn.myanimelist.net/
             images/anime/13/6905.jpg",
          "Duration": "30 min per ep",
          "Rank": "5824.0",
          "Synopsis": """This is the OVA of The Galaxy
             Railways.
The Galaxy Express 999 crashes on the off-limits planet
   of Herise and it's up to the SDF Sirius Platoon to
   assist.
(Source: Wikipedia)""",
          "Score": 6.6,
          "Type": "OVA",
          "Scored By": 720,
          "Licensors": "Discotek Media",
          "Favorites": 2,
          "Members": 2597
        }
```

```
    },
    {
      "_index": "anime",
      "_id": "TRH85JMBGbVWSHHuNJ8A",
      "_score": 10.20174,
      "_source": {
        "Other name": "                                ",
        "Producers": "Kadokawa Shoten",
        "Rating": "PG-13 - Teens 13 or older",
        "Studios": "asread.",
        "Aired": "May 25, 2007",
        "Source": "Original",
        "Name": "Kiddy GiRL-AND Pilot",
        "English name": "UNKNOWN",
        "Episodes": "1.0",
        "Genres": "Action, Sci-Fi",
        "anime_id": 5859,
        "Premiered": "UNKNOWN",
        "Status": "Finished Airing",
        "Popularity": 8120,
        "Image URL": "https://cdn.myanimelist.net/
          images/anime/6/12294.jpg",
        "Duration": "6 min",
        "Rank": "5613.0",
        "Synopsis": "ES members  clair  and Lumi re
          once were involved in a battle to decide the
           fate of the galaxy. They prevailed and
          peace was restored to the galaxy. Now,
          several years later, the galaxy once again
          stands on the brink of chaos. This conflict
          will see the dawn of two new heroines;
          Ascoeur and Q-feuille.",
        "Score": 6.64,
        "Type": "OVA",
        "Scored By": 1616,
        "Licensors": "UNKNOWN",
        "Favorites": 6,
        "Members": 4836
      }
    },
    {
      "_index": "anime",
      "_id": "oOL85JMBSCkiqoTYK1jr",
      "_score": 9.764661,
```

```
        "_source": {
          "Other name": "                    ",
          "Producers": "Ginga Ya, Studio Kyuuma, Trilogy
             Future Studio",
          "Rating": "PG-13 - Teens 13 or older",
          "Studios": "Planet",
          "Aired": "Oct 5, 2003 to Apr 11, 2004",
          "Source": "Original",
          "Name": "Ginga Tetsudou Monogatari",
          "English name": "The Galaxy Railways",
          "Episodes": "26.0",
          "Genres": "Action, Adventure, Drama, Sci-Fi",
          "anime_id": 1490,
          "Premiered": "fall 2003",
          "Status": "Finished Airing",
          "Popularity": 5607,
          "Image URL": "https://cdn.myanimelist.net/
             images/anime/6/39821.jpg",
          "Duration": "24 min per ep",
          "Rank": "3661.0",
          "Synopsis": """In the distant future, trains
             are no longer bound by their physical tracks
             . Instead, they take to the skies and travel
              across the universe on the Galaxy Railways,
              transporting mankind from planet to planet.
              However, the Galaxy Railways are no safer
             than traditional trains: criminals,
             terrorists, and vile aliens always find a
             way to stir up trouble.
Manabu Yuuki, a rash and hot-headed man, is the latest
   addition to the Galaxy Railways' elite Space Defence
   Force (SDF). These brave men and women are responsible
    for protecting the railways and responding to any
   unprecedented danger, risking their lives to protect
   the innocent from evil. But as this drama unfolds and
   the SDF's greatest crisis draws nearer, Manabu must
   truly learn what it means to be a member of the SDF
   before it is too late.""",
          "Score": 7.07,
          "Type": "TV",
          "Scored By": 3174,
          "Licensors": "Funimation",
          "Favorites": 44,
          "Members": 11995
```

```
        }
      }
    ]
  }
}
```

Query 10: Ranking Producers by Score Variability per Premiered Date

Description: This query groups anime by their premier date (`Premiered` field) and, for each group, calculates the maximum and minimum scores of the anime released on that date. Subsequently, it computes the difference between these two scores and ranks the premier dates based on this difference in descending order. Finally, it limits the results to the top 10 premier dates with the highest variability in anime scores. The goal is to identify premier dates with the greatest fluctuations in perceived anime quality.

Key Components:

- **Terms Aggregation on `Premiered`:** Groups anime by premier date.
- **Max and Min Aggregations:** Computes the maximum (`max_score`) and minimum (`min_score`) scores for each premier group.
- **Bucket Script Aggregation:** Calculates the difference between `max_score` and `min_score` for each group.
- **Bucket Sort Aggregation:** Ranks the groups by score difference in descending order and limits the results to the top 10 groups.

Query:

```
PUT /anime/_mapping
{
  "properties": {
    "Producers": {
      "type": "text",
      "fielddata": true
    }
  }
}

GET /anime/_search
{
  "size": 0,
  "aggs": {
```

```
    "producers_group": {
      "terms": {
        "field": "Premiered",
        "size": 50
      },
      "aggs": {
        "max_score": {
          "max": {
            "field": "Score"
          }
        },
        "min_score": {
          "min": {
            "field": "Score"
          }
        },
        "score_difference": {
          "bucket_script": {
            "buckets_path": {
              "maxScore": "max_score",
              "minScore": "min_score"
            },
            "script": "params.maxScore - params.minScore"
          }
        },
        "sort_by_difference": {
          "bucket_sort": {
            "sort": [
              {
                "score_difference": {
                  "order": "desc"
                }
              }
            ],
            "size": 10
          }
        }
      }
    }
  }
}
```

Outcome:

```
{
  "took": 17,
  "timed_out": false,
  "_shards": {
    "total": 1,
    "successful": 1,
    "skipped": 0,
    "failed": 0
  },
  "hits": {
    "total": {
      "value": 10000,
      "relation": "gte"
    },
    "max_score": null,
    "hits": []
  },
  "aggregations": {
    "producers_group": {
      "doc_count_error_upper_bound": 0,
      "sum_other_doc_count": 2011,
      "buckets": [
        {
          "key": "UNKNOWN",
          "doc_count": 11243,
          "max_score": {
            "value": 9.05
          },
          "min_score": {
            "value": 1.85
          },
          "score_difference": {
            "value": 7.200000000000001
          }
        },
        {
          "key": "winter 2021",
          "doc_count": 57,
          "max_score": {
            "value": 8.8
          },
          "min_score": {
            "value": 2.9
          },
```

```
          "score_difference": {
            "value": 5.9
          }
        },
        {
          "key": "spring 2015",
          "doc_count": 52,
          "max_score": {
            "value": 9.06
          },
          "min_score": {
            "value": 3.28
          },
          "score_difference": {
            "value": 5.780000000000001
          }
        },
        {
          "key": "winter 2017",
          "doc_count": 47,
          "max_score": {
            "value": 8.98
          },
          "min_score": {
            "value": 3.3
          },
          "score_difference": {
            "value": 5.680000000000001
          }
        },
        {
          "key": "spring 2018",
          "doc_count": 68,
          "max_score": {
            "value": 8.53
          },
          "min_score": {
            "value": 3.06
          },
          "score_difference": {
            "value": 5.469999999999999
          }
        },
        {
```

```
          "key": "summer 2017",
          "doc_count": 47,
          "max_score": {
            "value": 8.88
          },
          "min_score": {
            "value": 3.86
          },
          "score_difference": {
            "value": 5.020000000000001
          }
        },
        {
          "key": "winter 2022",
          "doc_count": 40,
          "max_score": {
            "value": 8.8
          },
          "min_score": {
            "value": 3.79
          },
          "score_difference": {
            "value": 5.010000000000001
          }
        },
        {
          "key": "winter 2014",
          "doc_count": 46,
          "max_score": {
            "value": 8.27
          },
          "min_score": {
            "value": 3.29
          },
          "score_difference": {
            "value": 4.9799999999999995
          }
        },
        {
          "key": "fall 2017",
          "doc_count": 57,
          "max_score": {
            "value": 8.93
          },
```

```
          "min_score": {
            "value": 4.23
          },
          "score_difference": {
            "value": 4.699999999999999
          }
        },
        {
          "key": "summer 2016",
          "doc_count": 49,
          "max_score": {
            "value": 8.49
          },
          "min_score": {
            "value": 3.83
          },
          "score_difference": {
            "value": 4.66
          }
        }
      ]
    }
  }
}
```