

# Visual Analytics Assignment 2 Report

## Section 1: Data Indexing

### Data Cleaning

The raw CSV data was first cleaned using a Python script to ensure all fields are well-typed, missing values are handled, and city names are standardized. The main steps include:

- Converting numeric fields to the correct type (e.g., ratings, votes, cost)
- Filling missing values with defaults
- Standardizing date and location formats
- Removing duplicates
- Outputting a cleaned CSV file for indexing

#### Key code snippet:

```
df['AggregateRating'] = pd.to_numeric(df['AggregateRating'],
errors='coerce')
df['Votes'] = pd.to_numeric(df['Votes'], errors='coerce')
df['AverageCostForTwo'] = pd.to_numeric(df['AverageCostForTwo'],
errors='coerce')
df['RestaurantName'] = df['RestaurantName'].fillna('Unknown Restaurant')
df['Date'] = df['Date'].apply(validate_date)
df.to_csv('Assignment2/restaurants_cleaned.csv', sep=';', index=False)
```

### Indexing with Grok Pipeline

To ingest the cleaned data into Elasticsearch, we initially used a Grok pipeline but later switched to a more reliable direct JSON document approach. Here's our journey and lessons learned:

#### Initial Approach: Grok Pipeline

The initial approach used a Grok pipeline to parse raw CSV lines:

- Converting each row to a single raw string
- Using Grok patterns to parse fields
- Converting field types after parsing

However, this approach had limitations:

- Complex string parsing could lead to data loss
- Field type conversion was error-prone
- Debugging was difficult when fields were missing

## Improved Approach: Direct JSON Documents

We improved the indexing process by directly constructing JSON documents:

- Each field is explicitly typed and mapped
- No intermediate string parsing required
- Better control over data integrity

**Key code snippet of the improved approach:**

```
def prepare_bulk_data(csv_file):
    df = pd.read_csv(csv_file, sep=';')
    bulk_data = []
    for _, row in df.iterrows():
        doc = {
            "SerialNumber": int(row['SerialNumber']),
            "RestaurantName": str(row['RestaurantName']),
            "AverageCostForTwo": int(row['AverageCostForTwo']),
            "AggregateRating": float(row['AggregateRating']),
            "RatingText": str(row['RatingText']),
            "Votes": float(row['Votes']),
            "Date": str(row['Date']),
            "Coordinates": str(row['Coordinates']),
            "City": str(row['City']),
            "Country": str(row['Country']),
            "Continent": str(row['Continent']),
            "City/Country/Continent": f"
{str(row['City'])}/{str(row['Country'])}/{str(row['Continent'])}"
        }
        bulk_data.append(json.dumps({"index": {"_index": INDEX_NAME}}))
        bulk_data.append(json.dumps(doc))
    return '\n'.join(bulk_data) + '\n'
```

## Key Learnings

Through this process, we learned several important lessons:

- Direct data mapping is more reliable than string parsing

- Explicit type conversion prevents data loss
- Simpler processing pipelines are easier to debug
- Maintaining data integrity requires careful handling at each step

**Result:** All cleaned restaurant data is now correctly indexed in Elasticsearch as structured JSON documents, with all fields properly preserved and typed.

## Query 1: Find restaurants with 'bar' in the name but not 'barbecue(s)' or 'barbeque(s)'

**Query:**

```
GET restaurants/_search
{
  "query": {
    "bool": {
      "must": [{"wildcard": {"RestaurantName": "*bar*"}}],
      "must_not": [
        {"wildcard": {"RestaurantName": "*barbecue*"}},
        {"wildcard": {"RestaurantName": "*barbeque*"}},
        {"wildcard": {"RestaurantName": "*barbecues*"}},
        {"wildcard": {"RestaurantName": "*barbeques*"}},
        {"match": {"RestaurantName": "bar"}}
      ]
    }
  },
  "_source": ["RestaurantName", "City/Country/Continent", "Votes"]
}
```

**Results:**

```
{
  "took": 13,
  "timed_out": false,
  "_shards": {
    "total": 1,
    "successful": 1,
    "skipped": 0,
    "failed": 0
  },
  "hits": {
    "total": {
      "value": 89,
      "relation": "eq"
    }
  }
}
```

```
},
"max_score": 1,
"hits": [
  {
    "_index": "restaurants",
    "_id": "zdk0fpYBFfumWPjvvv5K",
    "_score": 1,
    "_source": {
      "RestaurantName": "Paribar",
      "City/Country/Continent": "S   Paulo/Brazil/South America",
      "Votes": 46
    }
  },
  {
    "_index": "restaurants",
    "_id": "E9k0fpYBFfumWPjvvv9L",
    "_score": 1,
    "_source": {
      "RestaurantName": "Bardenay",
      "City/Country/Continent": "Boise/United States/North America",
      "Votes": 879
    }
  },
  {
    "_index": "restaurants",
    "_id": "GNk0fpYBFfumWPjvvv9L",
    "_score": 1,
    "_source": {
      "RestaurantName": "Barbacoa Restaurant",
      "City/Country/Continent": "Boise/United States/North America",
      "Votes": 538
    }
  },
  {
    "_index": "restaurants",
    "_id": "F9k0fpYBFfumWPjvvzBL",
    "_score": 1,
    "_source": {
      "RestaurantName": "Barrett Junction Cafe",
      "City/Country/Continent": "Potrero/United States/North America",
      "Votes": 9
    }
  },
  {
    "_index": "restaurants",
    "_id": "Rtk0fpYBFfumWPjvvzBL",
```

```
    "_score": 1,
    "_source": {
      "RestaurantName": "Rhubarb Le Restaurant",
      "City/Country/Continent": "Singapore/Singapore/Asia",
      "Votes": 33
    }
  },
  {
    "_index": "restaurants",
    "_id": "DNkOfpYBFfumWPjvvzFM",
    "_score": 1,
    "_source": {
      "RestaurantName": "The Cafe Baraco",
      "City/Country/Continent": "Ahmedabad/India/Asia",
      "Votes": 317
    }
  },
  {
    "_index": "restaurants",
    "_id": "09kOfpYBFfumWPjvvzFM",
    "_score": 1,
    "_source": {
      "RestaurantName": "Doon Darbar",
      "City/Country/Continent": "Dehradun/India/Asia",
      "Votes": 121
    }
  },
  {
    "_index": "restaurants",
    "_id": "tNkOfpYBFfumWPjvvzLm",
    "_score": 1,
    "_source": {
      "RestaurantName": "The Grill Darbar",
      "City/Country/Continent": "Faridabad/India/Asia",
      "Votes": 17
    }
  },
  {
    "_index": "restaurants",
    "_id": "2tkOfpYBFfumWPjvvzLm",
    "_score": 1,
    "_source": {
      "RestaurantName": "Barista",
      "City/Country/Continent": "Ghaziabad/India/Asia",
      "Votes": 33
    }
  }
}
```

```

    },
    {
      "_index": "restaurants",
      "_id": "A9k0fpYBFfumWPjvvzPm",
      "_score": 1,
      "_source": {
        "RestaurantName": "Zambar",
        "City/Country/Continent": "Gurgaon/India/Asia",
        "Votes": 802
      }
    }
  ]
}
}

```

## Query 2: Find the 3 most expensive restaurants within 20km of Singapore in 2016

Query:

```

GET restaurants/_search
{
  "size": 3,
  "query": {
    "bool": {
      "must": [
        {"range": {"Date": {"gte": "2016-01-01", "lte": "2016-12-31"}}},
        {"geo_distance": {"distance": "20km", "Coordinates": {"lat": 1.3521,
"lon": 103.8198}}}
      ]
    }
  },
  "sort": [{"AverageCostForTwo": {"order": "desc"}}],
  "_source": ["RestaurantName", "AverageCostForTwo", "Date",
"City/Country/Continent", "Coordinates"]
}

```

Results:

```

{
  "took": 9,
  "timed_out": false,
  "_shards": {
    "total": 1,

```

```
"successful": 1,
"skipped": 0,
"failed": 0
},
"hits": {
  "total": {
    "value": 6,
    "relation": "eq"
  },
  "max_score": null,
  "hits": [
    {
      "_index": "restaurants",
      "_id": "Q9kOfpYBFfumWPjvvzBL",
      "_score": null,
      "_source": {
        "AverageCostForTwo": 500,
        "Coordinates": {
          "lon": 103.8403602,
          "lat": 1.279419756
        },
        "RestaurantName": "Restaurant Andre",
        "Date": "2016-01-28T14:00:04Z",
        "City/Country/Continent": "Singapore/Singapore/Asia"
      },
      "sort": [500]
    },
    {
      "_index": "restaurants",
      "_id": "TtkOfpYBFfumWPjvvzBL",
      "_score": null,
      "_source": {
        "AverageCostForTwo": 300,
        "Coordinates": {
          "lon": 103.8601766,
          "lat": 1.290800881
        },
        "RestaurantName": "Summer Pavilion",
        "Date": "2016-05-22T10:07:17Z",
        "City/Country/Continent": "Singapore/Singapore/Asia"
      },
      "sort": [300]
    },
    {
      "_index": "restaurants",
      "_id": "StkOfpYBFfumWPjvvzBL",
```

```

    "_score": null,
    "_source": {
      "AverageCostForTwo": 80,
      "Coordinates": {
        "lon": 103.8621195,
        "lat": 1.310668316
      },
      "RestaurantName": "The Refinery Singapore",
      "Date": "2016-05-26T10:08:13Z",
      "City/Country/Continent": "Singapore/Singapore/Asia"
    },
    "sort": [80]
  }
]
}

```

### Query 3: Find pizza restaurants with good ratings

Query:

```

GET restaurants/_search
{
  "query": {
    "bool": {
      "must": [
        {"wildcard": {"RestaurantName": "*pizza*"}},
        {"bool": {
          "should": [
            {"term": {"RatingText.keyword": "Good"}},
            {"term": {"RatingText.keyword": "Very Good"}},
            {"term": {"RatingText.keyword": "Excellent"}}
          ],
          "minimum_should_match": 1
        }}
      ],
      "must_not": [{"wildcard": {"RestaurantName": "*pasta*"}}]
    }
  },
  "_source": ["City/Country/Continent", "AggregateRating", "RestaurantName", "Votes"],
  "sort": [{"AggregateRating": {"order": "desc"}}],
  "size": 100
}

```



## Results:

```
{
  "took": 10,
  "timed_out": false,
  "_shards": {
    "total": 1,
    "successful": 1,
    "skipped": 0,
    "failed": 0
  },
  "hits": {
    "total": {
      "value": 61,
      "relation": "eq"
    },
    "max_score": null,
    "hits": [
      {
        "_index": "restaurants",
        "_id": "xNkOfpYBFfumWPjvv9L",
        "_score": null,
        "_source": {
          "RestaurantName": "Ingleside Village Pizza",
          "AggregateRating": 4.9,
          "City/Country/Continent": "Macon/United States/North America",
          "Votes": 478
        },
        "sort": [
          4.9
        ]
      },
      {
        "_index": "restaurants",
        "_id": "ldkOfpYBFfumWPjvwVPy",
        "_score": null,
        "_source": {
          "RestaurantName": "Pizza ÔÁl Forno",
          "AggregateRating": 4.7,
          "City/Country/Continent": "Ankara/Turkey/Asia",
          "Votes": 104
        },
        "sort": [
          4.7
        ]
      }
    ]
  },
}
```

```
{
  "_index": "restaurants",
  "_id": "gNk0fpYBFfumWPjvvy9L",
  "_score": null,
  "_source": {
    "RestaurantName": "Fong's Pizza",
    "AggregateRating": 4.6,
    "City/Country/Continent": "Des Moines/United States/North
America",
    "Votes": 728
  },
  "sort": [
    4.6
  ]
},
{
  "_index": "restaurants",
  "_id": "gtk0fpYBFfumWPjvwVLy",
  "_score": null,
  "_source": {
    "RestaurantName": "Pizza Hut",
    "AggregateRating": 4.6,
    "City/Country/Continent": "Vizag/India/Asia",
    "Votes": 289
  },
  "sort": [
    4.6
  ]
},
{
  "_index": "restaurants",
  "_id": "M9k0fpYBFfumWPjvwDgz",
  "_score": null,
  "_source": {
    "RestaurantName": "Joey's Pizza",
    "AggregateRating": 4.5,
    "City/Country/Continent": "Mumbai/India/Asia",
    "Votes": 2662
  },
  "sort": [
    4.5
  ]
},
{
  "_index": "restaurants",
  "_id": "r9k0fpYBFfumWPjvvzBM",
```

```
"_score": null,
"_source": {
  "RestaurantName": "Pizza Di Rocco",
  "AggregateRating": 4.4,
  "City/Country/Continent": "Abu Dhabi/United Arab Emirates/Asia",
  "Votes": 471
},
"sort": [
  4.4
]
},
{
  "_index": "restaurants",
  "_id": "-tkOfpYBFfumWPjvvzBM",
  "_score": null,
  "_source": {
    "RestaurantName": "Pizza Hut",
    "AggregateRating": 4.4,
    "City/Country/Continent": "Agra/India/Asia",
    "Votes": 134
  },
  "sort": [
    4.4
  ]
},
{
  "_index": "restaurants",
  "_id": "AtkOfpYBFfumWPjvvzFM",
  "_score": null,
  "_source": {
    "RestaurantName": "La Pino'z Pizza",
    "AggregateRating": 4.4,
    "City/Country/Continent": "Ahmedabad/India/Asia",
    "Votes": 113
  },
  "sort": [
    4.4
  ]
},
{
  "_index": "restaurants",
  "_id": "HtkOfpYBFfumWPjvvz9L",
  "_score": null,
  "_source": {
    "RestaurantName": "Guido's Original New York Style Pizza",
    "AggregateRating": 4.3,
```

```
    "City/Country/Continent": "Boise/United States/North America",
    "Votes": 410
  },
  "sort": [
    4.3
  ]
},
{
  "_index": "restaurants",
  "_id": "Ndk0fpYBFfumWPjvv9L",
  "_score": null,
  "_source": {
    "RestaurantName": "A & A Pagliai's Pizza",
    "AggregateRating": 4.3,
    "City/Country/Continent": "Cedar Rapids/Iowa City/United States",
    "Votes": 485
  },
  "sort": [
    4.3
  ]
},
{
  "_index": "restaurants",
  "_id": "Adk0fpYBFfumWPjvvzFM",
  "_score": null,
  "_source": {
    "RestaurantName": "Fozzie's Pizzaiolo",
    "AggregateRating": 4.3,
    "City/Country/Continent": "Ahmedabad/India/Asia",
    "Votes": 731
  },
  "sort": [
    4.3
  ]
},
{
  "_index": "restaurants",
  "_id": "g9k0fpYBFfumWPjvwVLy",
  "_score": null,
  "_source": {
    "RestaurantName": "Pizza Hut",
    "AggregateRating": 4.3,
    "City/Country/Continent": "Vizag/India/Asia",
    "Votes": 230
  },
  "sort": [
```

```

    4.3
  ]
},
{
  "_index": "restaurants",
  "_id": "BdkOfpYBFfumWPjvv9L",
  "_score": null,
  "_source": {
    "RestaurantName": "Giuseppe's Pizza & Italian Specialities",
    "AggregateRating": 4.1,
    "City/Country/Continent": "Augusta/United States/North America",
    "Votes": 430
  },
  "sort": [
    4.1
  ]
},
{
  "_index": "restaurants",
  "_id": "HNkOfpYBFfumWPjvv9L",
  "_score": null,
  "_source": {
    "RestaurantName": "Flying Pie Pizzeria",
    "AggregateRating": 4.1,
    "City/Country/Continent": "Boise/United States/North America",
    "Votes": 550
  },
  "sort": [
    4.1
  ]
},
{
  "_index": "restaurants",
  "_id": "atkOfpYBFfumWPjvvzFM",
  "_score": null,
  "_source": {
    "RestaurantName": "Pizza Hut",
    "AggregateRating": 4.1,
    "City/Country/Continent": "Bhopal/India/Asia",
    "Votes": 118
  },
  "sort": [
    4.1
  ]
},
{

```

```
"_index": "restaurants",
"_id": "BtkOfpYBFfumWPjvvzTm",
"_score": null,
"_source": {
  "RestaurantName": "Pizza Central",
  "AggregateRating": 4.1,
  "City/Country/Continent": "Gurgaon/India/Asia",
  "Votes": 151
},
"sort": [
  4.1
]
},
{
  "_index": "restaurants",
  "_id": "JtkOfpYBFfumWPjvwUk9",
  "_score": null,
  "_source": {
    "RestaurantName": "Tossin Pizza",
    "AggregateRating": 4.1,
    "City/Country/Continent": "New Delhi/India/Asia",
    "Votes": 647
  },
  "sort": [
    4.1
  ]
},
{
  "_index": "restaurants",
  "_id": "VNkOfpYBFfumWPjvwVLy",
  "_score": null,
  "_source": {
    "RestaurantName": "Pizza Hut",
    "AggregateRating": 4.1,
    "City/Country/Continent": "Surat/India/Asia",
    "Votes": 216
  },
  "sort": [
    4.1
  ]
},
{
  "_index": "restaurants",
  "_id": "bNkOfpYBFfumWPjvwVLy",
  "_score": null,
  "_source": {
```

```
    "RestaurantName": "Pizza Hut",
    "AggregateRating": 4.1,
    "City/Country/Continent": "Vadodara/India/Asia",
    "Votes": 269
  },
  "sort": [
    4.1
  ]
},
{
  "_index": "restaurants",
  "_id": "ZtkOfpYBFfumWPjvvzBL",
  "_score": null,
  "_source": {
    "RestaurantName": "El Fredo Pizza",
    "AggregateRating": 4,
    "City/Country/Continent": "Sioux City/United States/North
America",
    "Votes": 280
  },
  "sort": [
    4
  ]
},
{
  "_index": "restaurants",
  "_id": "XdkOfpYBFfumWPjvvzPm",
  "_score": null,
  "_source": {
    "RestaurantName": "Instapizza",
    "AggregateRating": 4,
    "City/Country/Continent": "Gurgaon/India/Asia",
    "Votes": 219
  },
  "sort": [
    4
  ]
},
{
  "_index": "restaurants",
  "_id": "TdkOfpYBFfumWPjvvzTm",
  "_score": null,
  "_source": {
    "RestaurantName": "La Pino'z Pizza",
    "AggregateRating": 4,
    "City/Country/Continent": "Gurgaon/India/Asia",
```

```
    "Votes": 259
  },
  "sort": [
    4
  ]
},
{
  "_index": "restaurants",
  "_id": "J9k0fpYBFfumWPjvwDgz",
  "_score": null,
  "_source": {
    "RestaurantName": "Joey's Pizza",
    "AggregateRating": 4,
    "City/Country/Continent": "Mumbai/India/Asia",
    "Votes": 5145
  },
  "sort": [
    4
  ]
},
{
  "_index": "restaurants",
  "_id": "Rdk0fpYBFfumWPjvwDgz",
  "_score": null,
  "_source": {
    "RestaurantName": "Pizza Hut",
    "AggregateRating": 4,
    "City/Country/Continent": "Mysore/India/Asia",
    "Votes": 304
  },
  "sort": [
    4
  ]
},
{
  "_index": "restaurants",
  "_id": "I9k0fpYBFfumWPjvwUk9",
  "_score": null,
  "_source": {
    "RestaurantName": "Napoli Pizza",
    "AggregateRating": 4,
    "City/Country/Continent": "New Delhi/India/Asia",
    "Votes": 41
  },
  "sort": [
    4
  ]
}
```



```
]
},
{
  "_index": "restaurants",
  "_id": "aNkOfpYBFfumWPjvvy9L",
  "_score": null,
  "_source": {
    "RestaurantName": "Crust Stone Oven Pizza",
    "AggregateRating": 3.9,
    "City/Country/Continent": "Davenport/United States/North America",
    "Votes": 136
  },
  "sort": [
    3.9
  ]
},
{
  "_index": "restaurants",
  "_id": "DtkOfpYBFfumWPjvwD-5",
  "_score": null,
  "_source": {
    "RestaurantName": "Instapizza",
    "AggregateRating": 3.9,
    "City/Country/Continent": "New Delhi/India/Asia",
    "Votes": 55
  },
  "sort": [
    3.9
  ]
},
{
  "_index": "restaurants",
  "_id": "ldkOfpYBFfumWPjvwEX_",
  "_score": null,
  "_source": {
    "RestaurantName": "Pizza Hut Delivery",
    "AggregateRating": 3.9,
    "City/Country/Continent": "New Delhi/India/Asia",
    "Votes": 189
  },
  "sort": [
    3.9
  ]
},
{
  "_index": "restaurants",
```

```
"_id": "2dk0fpYBFfumWPjvwU28",
"_score": null,
"_source": {
  "RestaurantName": "Pizza Hut",
  "AggregateRating": 3.9,
  "City/Country/Continent": "Noida/India/Asia",
  "Votes": 173
},
"sort": [
  3.9
]
},
{
  "_index": "restaurants",
  "_id": "2tk0fpYBFfumWPjvwU28",
  "_score": null,
  "_source": {
    "RestaurantName": "PizzaExpress",
    "AggregateRating": 3.9,
    "City/Country/Continent": "Noida/India/Asia",
    "Votes": 366
  },
  "sort": [
    3.9
  ]
},
{
  "_index": "restaurants",
  "_id": "XNk0fpYBFfumWPjvvzBL",
  "_score": null,
  "_source": {
    "RestaurantName": "Jerry's Pizza",
    "AggregateRating": 3.8,
    "City/Country/Continent": "Sioux City/United States/North
America",
    "Votes": 178
  },
  "sort": [
    3.8
  ]
},
{
  "_index": "restaurants",
  "_id": "z9k0fpYBFfumWPjvwD-5",
  "_score": null,
  "_source": {
```

```
    "RestaurantName": "Instapizza",
    "AggregateRating": 3.8,
    "City/Country/Continent": "New Delhi/India/Asia",
    "Votes": 492
  },
  "sort": [
    3.8
  ]
},
{
  "_index": "restaurants",
  "_id": "vdk0fpYBFfumWPjvwEG5",
  "_score": null,
  "_source": {
    "RestaurantName": "Play Pizza",
    "AggregateRating": 3.8,
    "City/Country/Continent": "New Delhi/India/Asia",
    "Votes": 270
  },
  "sort": [
    3.8
  ]
},
{
  "_index": "restaurants",
  "_id": "W9k0fpYBFfumWPjvv9L",
  "_score": null,
  "_source": {
    "RestaurantName": "Tony's Italian Restaurant & Pizza",
    "AggregateRating": 3.7,
    "City/Country/Continent": "Dalton/United States/North America",
    "Votes": 116
  },
  "sort": [
    3.7
  ]
},
{
  "_index": "restaurants",
  "_id": "QNk0fpYBFfumWPjvvzPm",
  "_score": null,
  "_source": {
    "RestaurantName": "California Pizza Kitchen",
    "AggregateRating": 3.7,
    "City/Country/Continent": "Gurgaon/India/Asia",
    "Votes": 980
  }
}
```

```
    },
    "sort": [
      3.7
    ]
  },
  {
    "_index": "restaurants",
    "_id": "zdkOfpYBFfumWPjvwDYz",
    "_score": null,
    "_source": {
      "RestaurantName": "Ovenstory Pizza",
      "AggregateRating": 3.7,
      "City/Country/Continent": "Gurgaon/India/Asia",
      "Votes": 23
    },
    "sort": [
      3.7
    ]
  },
  {
    "_index": "restaurants",
    "_id": "StkOfpYBFfumWPjvwDp6",
    "_score": null,
    "_source": {
      "RestaurantName": "Domino's Pizza",
      "AggregateRating": 3.7,
      "City/Country/Continent": "New Delhi/India/Asia",
      "Votes": 336
    },
    "sort": [
      3.7
    ]
  },
  {
    "_index": "restaurants",
    "_id": "1tkOfpYBFfumWPjvwU28",
    "_score": null,
    "_source": {
      "RestaurantName": "Instapizza",
      "AggregateRating": 3.7,
      "City/Country/Continent": "Noida/India/Asia",
      "Votes": 214
    },
    "sort": [
      3.7
    ]
  }
]
```

```

},
{
  "_index": "restaurants",
  "_id": "4NkOfpYBFfumWPjvwVHy",
  "_score": null,
  "_source": {
    "RestaurantName": "Pizza Hut",
    "AggregateRating": 3.7,
    "City/Country/Continent": "Noida/India/Asia",
    "Votes": 189
  },
  "sort": [
    3.7
  ]
},
{
  "_index": "restaurants",
  "_id": "VtkOfpYBFfumWPjvvzBL",
  "_score": null,
  "_source": {
    "RestaurantName": "Bob Roe's Pizza",
    "AggregateRating": 3.6,
    "City/Country/Continent": "Sioux City/United States/North
America",
    "Votes": 92
  },
  "sort": [
    3.6
  ]
},
{
  "_index": "restaurants",
  "_id": "cNkOfpYBFfumWPjvvzPm",
  "_score": null,
  "_source": {
    "RestaurantName": "Instapizza",
    "AggregateRating": 3.6,
    "City/Country/Continent": "Gurgaon/India/Asia",
    "Votes": 426
  },
  "sort": [
    3.6
  ]
},
{
  "_index": "restaurants",

```

```
"_id": "q9k0fpYBFfumWPjvvzXn",
"_score": null,
"_source": {
  "RestaurantName": "Domino's Pizza",
  "AggregateRating": 3.6,
  "City/Country/Continent": "Gurgaon/India/Asia",
  "Votes": 146
},
"sort": [
  3.6
]
},
{
  "_index": "restaurants",
  "_id": "B9k0fpYBFfumWPjvwDgz",
  "_score": null,
  "_source": {
    "RestaurantName": "Domino's Pizza",
    "AggregateRating": 3.6,
    "City/Country/Continent": "Ludhiana/India/Asia",
    "Votes": 86
  },
  "sort": [
    3.6
  ]
},
{
  "_index": "restaurants",
  "_id": "sNk0fpYBFfumWPjvwDt6",
  "_score": null,
  "_source": {
    "RestaurantName": "Pizza Hut Delivery",
    "AggregateRating": 3.6,
    "City/Country/Continent": "New Delhi/India/Asia",
    "Votes": 107
  },
  "sort": [
    3.6
  ]
},
{
  "_index": "restaurants",
  "_id": "09k0fpYBFfumWPjvwEX_",
  "_score": null,
  "_source": {
    "RestaurantName": "Pizza Hut Delivery",
```

```
    "AggregateRating": 3.6,
    "City/Country/Continent": "New Delhi/India/Asia",
    "Votes": 103
  },
  "sort": [
    3.6
  ]
},
{
  "_index": "restaurants",
  "_id": "ldkOfpYBFfumWPjvwUt-",
  "_score": null,
  "_source": {
    "RestaurantName": "Pizza Hut",
    "AggregateRating": 3.6,
    "City/Country/Continent": "New Delhi/India/Asia",
    "Votes": 141
  },
  "sort": [
    3.6
  ]
},
{
  "_index": "restaurants",
  "_id": "JNkOfpYBFfumWPjvwUx_",
  "_score": null,
  "_source": {
    "RestaurantName": "Domino's Pizza",
    "AggregateRating": 3.6,
    "City/Country/Continent": "New Delhi/India/Asia",
    "Votes": 24
  },
  "sort": [
    3.6
  ]
},
{
  "_index": "restaurants",
  "_id": "UdkOfpYBFfumWPjvwU-9",
  "_score": null,
  "_source": {
    "RestaurantName": "Domino's Pizza",
    "AggregateRating": 3.6,
    "City/Country/Continent": "Noida/India/Asia",
    "Votes": 547
  },
},
```

```
    "sort": [
      3.6
    ]
  },
  {
    "_index": "restaurants",
    "_id": "ldk0fpYBFfumWPjvv9L",
    "_score": null,
    "_source": {
      "RestaurantName": "Happy Joe's Pizza & Ice Cream",
      "AggregateRating": 3.5,
      "City/Country/Continent": "Dubuque/United States/North America",
      "Votes": 74
    },
    "sort": [
      3.5
    ]
  },
  {
    "_index": "restaurants",
    "_id": "V9k0fpYBFfumWPjvvzXn",
    "_score": null,
    "_source": {
      "RestaurantName": "Konetto Pizza",
      "AggregateRating": 3.5,
      "City/Country/Continent": "Gurgaon/India/Asia",
      "Votes": 32
    },
    "sort": [
      3.5
    ]
  },
  {
    "_index": "restaurants",
    "_id": "39k0fpYBFfumWPjvvzXn",
    "_score": null,
    "_source": {
      "RestaurantName": "La Pino'z Pizza",
      "AggregateRating": 3.5,
      "City/Country/Continent": "Gurgaon/India/Asia",
      "Votes": 22
    },
    "sort": [
      3.5
    ]
  },
}
```



```
{
  "_index": "restaurants",
  "_id": "adk0fpYBFfumWPjvwDp6",
  "_score": null,
  "_source": {
    "RestaurantName": "Pizza Hut",
    "AggregateRating": 3.5,
    "City/Country/Continent": "New Delhi/India/Asia",
    "Votes": 541
  },
  "sort": [
    3.5
  ]
},
{
  "_index": "restaurants",
  "_id": "T9k0fpYBFfumWPjvwDx6",
  "_score": null,
  "_source": {
    "RestaurantName": "Domino's Pizza",
    "AggregateRating": 3.5,
    "City/Country/Continent": "New Delhi/India/Asia",
    "Votes": 119
  },
  "sort": [
    3.5
  ]
},
{
  "_index": "restaurants",
  "_id": "Ftk0fpYBFfumWPjvwEG5",
  "_score": null,
  "_source": {
    "RestaurantName": "Papa Pizza",
    "AggregateRating": 3.5,
    "City/Country/Continent": "New Delhi/India/Asia",
    "Votes": 21
  },
  "sort": [
    3.5
  ]
},
{
  "_index": "restaurants",
  "_id": "xNk0fpYBFfumWPjvwEP_",
  "_score": null,
```

```
"_source": {
  "RestaurantName": "Pizza Station",
  "AggregateRating": 3.5,
  "City/Country/Continent": "New Delhi/India/Asia",
  "Votes": 20
},
"sort": [
  3.5
]
},
{
  "_index": "restaurants",
  "_id": "ddk0fpYBFfumWPjvwUg9",
  "_score": null,
  "_source": {
    "RestaurantName": "Domino's Pizza",
    "AggregateRating": 3.5,
    "City/Country/Continent": "New Delhi/India/Asia",
    "Votes": 194
  },
  "sort": [
    3.5
  ]
},
{
  "_index": "restaurants",
  "_id": "wNk0fpYBFfumWPjvwUg9",
  "_score": null,
  "_source": {
    "RestaurantName": "Domino's Pizza",
    "AggregateRating": 3.5,
    "City/Country/Continent": "New Delhi/India/Asia",
    "Votes": 123
  },
  "sort": [
    3.5
  ]
},
{
  "_index": "restaurants",
  "_id": "f9k0fpYBFfumWPjvwUt-",
  "_score": null,
  "_source": {
    "RestaurantName": "Ovenstory Pizza",
    "AggregateRating": 3.5,
    "City/Country/Continent": "New Delhi/India/Asia",
```

```
    "Votes": 12
  },
  "sort": [
    3.5
  ]
},
{
  "_index": "restaurants",
  "_id": "rtk0fpYBFfumWPjvwUx_",
  "_score": null,
  "_source": {
    "RestaurantName": "Instapizza After Hours",
    "AggregateRating": 3.5,
    "City/Country/Continent": "New Delhi/India/Asia",
    "Votes": 21
  },
  "sort": [
    3.5
  ]
},
{
  "_index": "restaurants",
  "_id": "F9k0fpYBFfumWPjvwU68",
  "_score": null,
  "_source": {
    "RestaurantName": "Domino's Pizza",
    "AggregateRating": 3.5,
    "City/Country/Continent": "Noida/India/Asia",
    "Votes": 56
  },
  "sort": [
    3.5
  ]
},
{
  "_index": "restaurants",
  "_id": "Qtk0fpYBFfumWPjvwVLy",
  "_score": null,
  "_source": {
    "RestaurantName": "Pizza Hut",
    "AggregateRating": 3.5,
    "City/Country/Continent": "Ranchi/India/Asia",
    "Votes": 66
  },
  "sort": [
    3.5
  ]
}
```

```
    ]
  }
]
}
```

## Aggregation Query: Find cities with highest average restaurant prices

Query:

```
GET restaurants/_search
{
  "size": 0,
  "query": {
    "range": {
      "Votes": {
        "gte": 100
      }
    }
  },
  "aggs": {
    "cities": {
      "terms": {
        "field": "City.keyword",
        "min_doc_count": 10,
        "size": 7,
        "shard_size": 100,
        "order": {
          "avg_price": "desc"
        }
      },
      "aggs": {
        "avg_price": {
          "avg": {
            "field": "AverageCostForTwo"
          }
        }
      }
    }
  }
}
```

Results:

```
{
  "took": 19,
  "timed_out": false,
  "_shards": {
    "total": 1,
    "successful": 1,
    "skipped": 0,
    "failed": 0
  },
  "hits": {
    "total": {
      "value": 2799,
      "relation": "eq"
    },
    "max_score": null,
    "hits": []
  },
  "aggregations": {
    "cities": {
      "doc_count_error_upper_bound": -1,
      "sum_other_doc_count": 2650,
      "buckets": [
        {
          "key": "Jakarta",
          "doc_count": 16,
          "avg_price": {
            "value": 308437.5
          }
        },
        {
          "key": "Colombo",
          "doc_count": 14,
          "avg_price": {
            "value": 2535.714285714286
          }
        },
        {
          "key": "Hyderabad",
          "doc_count": 17,
          "avg_price": {
            "value": 1358.8235294117646
          }
        },
        {
          "key": "Pune",
          "doc_count": 20,
```

```

      "avg_price": {
        "value": 1337.5
      }
    },
    {
      "key": "Jaipur",
      "doc_count": 18,
      "avg_price": {
        "value": 1316.6666666666667
      }
    },
    {
      "key": "Kolkata",
      "doc_count": 20,
      "avg_price": {
        "value": 1272.5
      }
    },
    {
      "key": "Bangalore",
      "doc_count": 20,
      "avg_price": {
        "value": 1232.5
      }
    }
  ]
}

```

### Analysis:

This query identifies the 7 cities with highest average restaurant prices, filtered by restaurants with  $\geq 100$  votes and cities with  $\geq 10$  restaurants. Results show Jakarta (308,437.5) and Colombo (2,535.71) as the most expensive, followed by Hyderabad, Pune, Jaipur, Kolkata, and Bangalore (1,200-1,400 range). Note: Jakarta's exceptionally high price suggests potential data anomalies.

## Aggregation Query 2: Calculate weighted average rating for restaurants within 10km of Noida

### Query:

```

GET restaurants/_search
{

```

```
"size": 0,
"_source": ["AggregateRating", "Votes", "City", "Country", "Coordinates"],
"query": {
  "bool": {
    "filter": [
      {
        "geo_distance": {
          "distance": "10km",
          "Coordinates": {
            "lat": 28.535517,
            "lon": 77.391029
          }
        }
      }
    ]
  }
},
"aggs": {
  "weighted_avg_rating": {
    "weighted_avg": {
      "value": {
        "field": "AggregateRating"
      },
      "weight": {
        "field": "Votes"
      }
    }
  },
  "cities": {
    "terms": {
      "field": "City.keyword",
      "size": 100
    },
    "aggs": {
      "weighted_avg_rating": {
        "weighted_avg": {
          "value": {
            "field": "AggregateRating"
          },
          "weight": {
            "field": "Votes"
          }
        }
      }
    },
    "country": {
      "terms": {
```

```

        "field": "Country.keyword",
        "size": 1
    }
}
}
}
}
}
}
}

```

## Results:

```

{
  "took": 45,
  "timed_out": false,
  "_shards": {
    "total": 1,
    "successful": 1,
    "skipped": 0,
    "failed": 0
  },
  "hits": {
    "total": {
      "value": 1094,
      "relation": "eq"
    },
    "max_score": null,
    "hits": []
  },
  "aggregations": {
    "weighted_avg_rating": {
      "value": 3.475335691819247
    },
    "cities": {
      "doc_count_error_upper_bound": 0,
      "sum_other_doc_count": 0,
      "buckets": [
        {
          "key": "Noida",
          "doc_count": 999,
          "weighted_avg_rating": {
            "value": 3.4814953434299833
          },
          "country": {
            "doc_count_error_upper_bound": 0,
            "sum_other_doc_count": 0,

```



```

        "buckets": [
            {
                "key": "India",
                "doc_count": 999
            }
        ]
    },
    {
        "key": "New Delhi",
        "doc_count": 73,
        "weighted_avg_rating": {
            "value": 3.351716305182232
        },
        "country": {
            "doc_count_error_upper_bound": 0,
            "sum_other_doc_count": 0,
            "buckets": [
                {
                    "key": "India",
                    "doc_count": 73
                }
            ]
        }
    },
    {
        "key": "Faridabad",
        "doc_count": 22,
        "weighted_avg_rating": {
            "value": 2.902033887475224
        },
        "country": {
            "doc_count_error_upper_bound": 0,
            "sum_other_doc_count": 0,
            "buckets": [
                {
                    "key": "India",
                    "doc_count": 22
                }
            ]
        }
    }
]
}

```

## Analysis:

This query calculates the weighted average rating for restaurants within 10km of Noida (28.535517, 77.391029). Results show:

- Overall weighted average rating: 3.48
  - Noida (999 restaurants): 3.48
  - New Delhi (73 restaurants): 3.35
  - Faridabad (22 restaurants): 2.90
- All restaurants are located in India.

## Aggregation Query 3: Analyze non-Asian restaurants with "Very Good" rating

### Query:

```
GET restaurants/_search
{
  "size": 0,
  "query": {
    "bool": {
      "must": [
        {
          "term": {
            "RatingText.keyword": "Very Good"
          }
        }
      ],
      "must_not": [
        {
          "term": {
            "Continent.keyword": "Asia"
          }
        }
      ]
    }
  },
  "aggs": {
    "votes_ranges": {
      "range": {
        "field": "Votes",
        "ranges": [
          {
            "from": 100,
            "to": 250,
```

```

        "key": "100-250 votes"
      },
      {
        "from": 250,
        "to": 400,
        "key": "250-400 votes"
      },
      {
        "from": 400,
        "to": 550,
        "key": "400-550 votes"
      }
    ]
  },
  "aggs": {
    "cost_stats": {
      "stats": {
        "field": "AverageCostForTwo"
      }
    }
  }
}
}
}
}
}

```

## Results:

```

{
  "took": 27,
  "timed_out": false,
  "_shards": {
    "total": 1,
    "successful": 1,
    "skipped": 0,
    "failed": 0
  },
  "hits": {
    "total": {
      "value": 291,
      "relation": "eq"
    },
    "max_score": null,
    "hits": []
  },
  "aggregations": {

```

```
"votes_ranges": {
  "buckets": [
    {
      "key": "100-250 votes",
      "from": 100,
      "to": 250,
      "doc_count": 75,
      "cost_stats": {
        "count": 75,
        "min": 0,
        "max": 535,
        "avg": 78.6,
        "sum": 5895
      }
    },
    {
      "key": "250-400 votes",
      "from": 250,
      "to": 400,
      "doc_count": 61,
      "cost_stats": {
        "count": 61,
        "min": 10,
        "max": 400,
        "avg": 57.62,
        "sum": 3515
      }
    },
    {
      "key": "400-550 votes",
      "from": 400,
      "to": 550,
      "doc_count": 37,
      "cost_stats": {
        "count": 37,
        "min": 10,
        "max": 570,
        "avg": 87.16,
        "sum": 3225
      }
    }
  ]
}
```

## Analysis:

This query analyzes non-Asian restaurants rated "Very Good" across different vote ranges:

- 100-250 votes: 75 restaurants, cost range 0-535, avg 78.6
- 250-400 votes: 61 restaurants, cost range 10-400, avg 57.62
- 400-550 votes: 37 restaurants, cost range 10-570, avg 87.16

Total of 173 restaurants analyzed.

## Shard Size Problem Discussion

### Query 1: Cities with Highest Average Restaurant Prices

#### Problem:

The original query may face the shard size problem because:

1. Each shard only returns its local top 7 results
2. When ordering by average price, some shards might contain high-priced restaurants but with fewer than 10 restaurants per city
3. This could lead to missing cities that should actually be in the top 7

#### Solution:

```
{
  "size": 0,
  "query": {
    "range": {
      "Votes": {
        "gte": 100
      }
    }
  },
  "aggs": {
    "cities": {
      "terms": {
        "field": "City.keyword",
        "min_doc_count": 10,
        "size": 7,
        "shard_size": 100,
        "order": {
          "avg_price": "desc"
        }
      }
    },
    "aggs": {
      "avg_price": {
```

```

        "avg": {
          "field": "AverageCostForTwo"
        }
      }
    }
  }
}

```

### Improvements:

1. Added `shard_size: 100` to ensure each shard returns more candidate cities
2. Maintained `size: 7` for final result count
3. This modification helps prevent missing potential high-price cities due to shard limitations

## Query 2: Weighted Average Rating Near Noida

### Analysis:

This query does not face the shard size problem because:

1. Uses `weighted_avg` aggregation which performs global calculations
2. City aggregation uses `size: 100`, sufficient for the limited geographic area
3. Geographic filter (10km radius) significantly reduces the data scope
4. Results show only 3 cities (Noida: 999, New Delhi: 73, Faridabad: 22)
5. No `sum_other_doc_count` indicates no missing cities

### Current Query Settings:

```

"aggs": {
  "weighted_avg_rating": {
    "weighted_avg": {
      "value": {
        "field": "AggregateRating"
      },
      "weight": {
        "field": "Votes"
      }
    }
  },
  "cities": {
    "terms": {
      "field": "City.keyword",
      "size": 100
    }
  }
}

```

```
}  
}
```

### Data Validation:

- Total hits: 1,094 documents
- City distribution:
  - Noida: 999 restaurants
  - New Delhi: 73 restaurants
  - Faridabad: 22 restaurants
- No missing cities in results

The current settings are adequate for accurate weighted average calculations without requiring shard size adjustments.

## Query 3: Non-Asian "Very Good" Restaurants by Vote Ranges

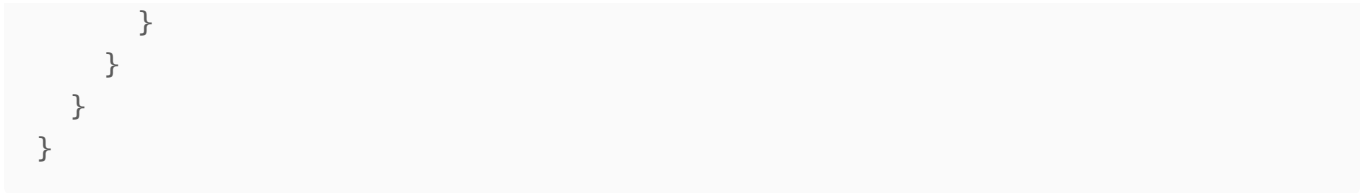
### Analysis:

This query does not face the shard size problem because:

1. Uses `range` aggregation with fixed numerical ranges
2. Each range has precise boundaries (100-250, 250-400, 400-550)
3. `stats` aggregation performs global calculations
4. Results show complete statistical information for each range
5. Relatively small dataset (291 total documents)

### Current Query Settings:

```
"aggs": {  
  "votes_ranges": {  
    "range": {  
      "field": "Votes",  
      "ranges": [  
        {"from": 100, "to": 250},  
        {"from": 250, "to": 400},  
        {"from": 400, "to": 550}  
      ]  
    },  
    "aggs": {  
      "cost_stats": {  
        "stats": {  
          "field": "AverageCostForTwo"  
        }  
      }  
    }  
  }  
}
```

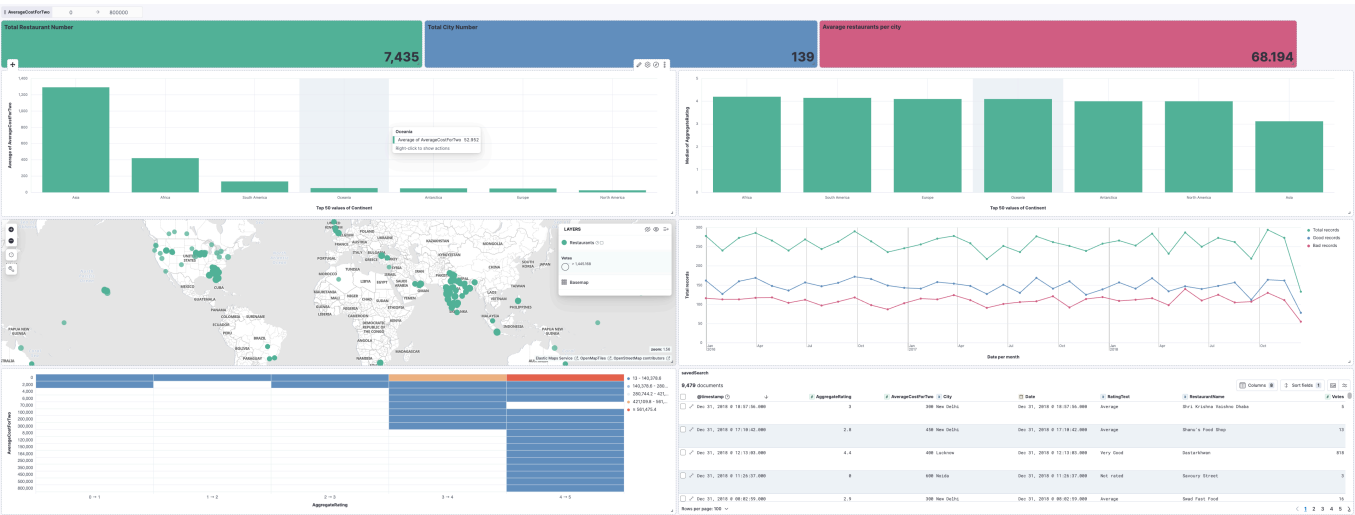


**Data Validation:**

- Total hits: 291 documents
- Range distribution:
  - 100-250 votes: 75 restaurants
  - 250-400 votes: 61 restaurants
  - 400-550 votes: 37 restaurants
- Each range has complete cost statistics (min, max, avg, sum)

The current settings are sufficient for accurate range-based aggregation without requiring shard size adjustments.

**Section 2: Data Visualization**



**Part 1: Dashboard Creation**

The dashboard was created to provide comprehensive insights into the restaurant dataset. Here's a detailed breakdown of each component:

**1. Key Metrics Panel**

The dashboard displays three essential metrics at the top:

- Total Restaurants: 7,435 restaurants reviewed
- Total Cities: 139 unique cities covered



- Average Restaurants per City: 68.194 restaurants

## 2. Price Range Filter

Implemented a dynamic filter that allows users to select specific price ranges, enabling focused analysis of restaurants within particular price segments.

## 3. Continental Analysis

Two visualizations show continental patterns:

- Average Rating by Continent: Bar chart showing aggregate ratings across continents
- Average Cost for Two by Continent: Bar chart displaying average costs across continents

## 4. Review Trends Over Time

Created a line chart showing:

- Total reviews trend (blue line)
- Positive reviews trend (green line, ratings > 3)
- Negative reviews trend (red line, ratings ≤ 3)

The chart effectively shows the evolution of restaurant reviews and sentiment over time.

## 5. Geographic Distribution

Implemented an interactive map where:

- Each restaurant is represented by a marker
- Marker size correlates with the number of votes
- Provides clear visualization of restaurant density and popularity across regions

## 6. Restaurant Details Table

Added a saved search showing detailed restaurant information including:

- Restaurant name
- Location details
- Ratings and reviews
- Cost information

This provides granular access to individual restaurant data.

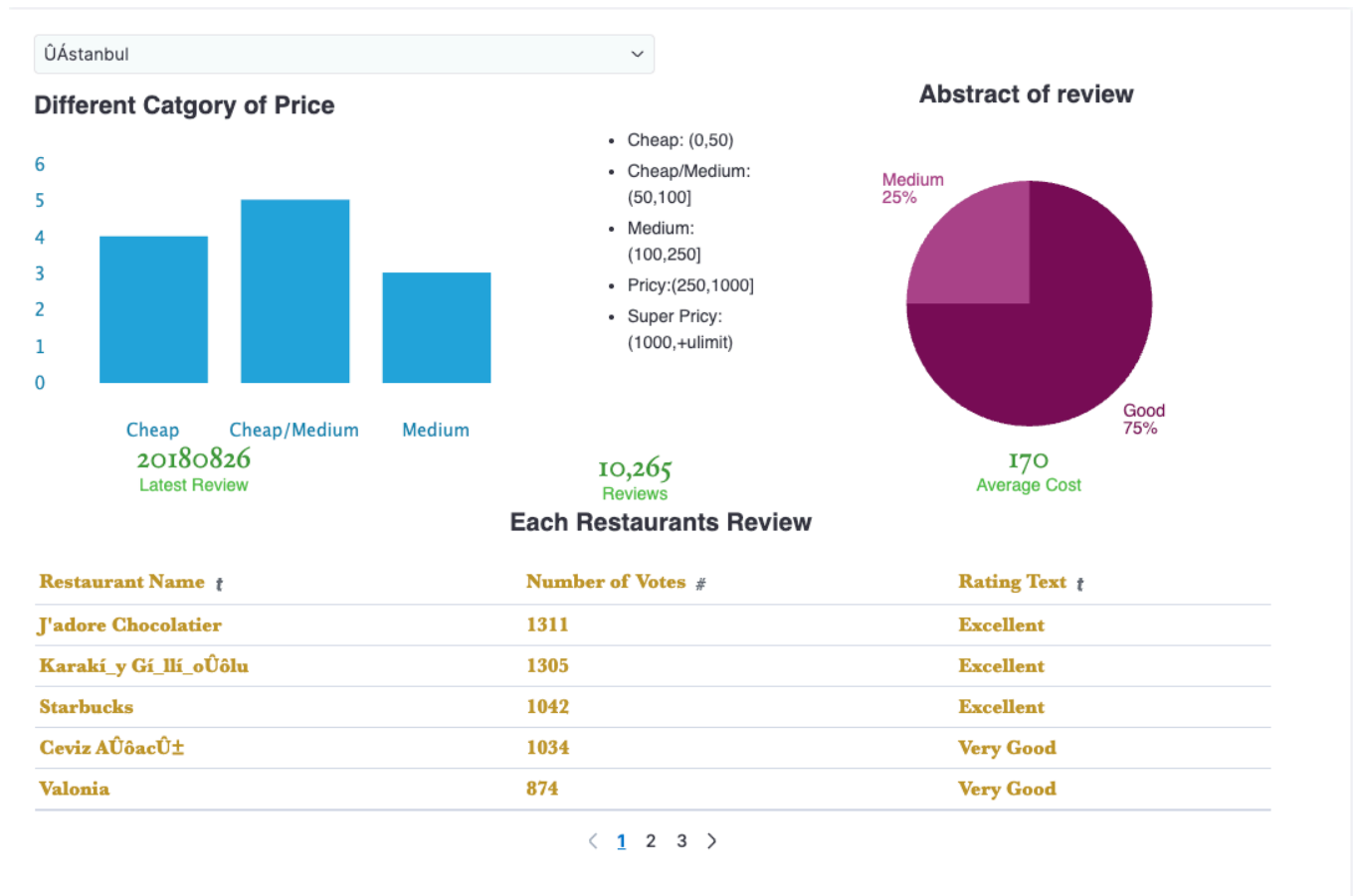
## 7. Rating-Price Heat Map

Created a heat map visualization that shows:

- Price ranges divided into 20 bins on one axis
- Rating ranges ([0-1), [1-2), [2-3), [3-4), [4-5]) on the other axis
- Color intensity indicating the number of votes

This effectively shows the relationship between price, rating, and popularity.

## Part 2: Canvas Creation



The canvas provides a city-focused view of restaurant ratings with the following components:

### 1. City Selection

Implemented a dropdown filter allowing users to select specific cities (currently showing "Istanbul" in the screenshot).

### 2. Price Category Distribution

Created a bar chart showing restaurant counts across price categories:

- Cheap ( $\leq 50$ ): 4 restaurants
- Cheap/Medium (50-100]: 5 restaurants

- Medium (100-250]: 3 restaurants
- Pricy (250-1000]: Not shown in current view
- Super Pricy (>1000): Not shown in current view

### 3. Rating Distribution

Added a pie chart showing the percentage distribution of restaurant ratings:

- Good (>4): 75%
- Medium (1.5-4]: 25%
- Bad ( $\leq 1.5$ ): Not present in current view

### 4. Restaurant Details Table

Implemented a paginated table (5 rows per page) showing:

- Restaurant Name
  - Number of Votes
  - Rating Text
- Current view shows top restaurants including "J'adore Chocolatier" and "Starbucks" with their respective ratings.

### 5. Key Metrics

Displayed important statistics:

- Total Reviews: 10,265
- Latest Review Date: 2018-08-26
- Average Cost: 170

The canvas provides an intuitive and interactive way to analyze restaurant data for each city, with clear visualizations and metrics that help users understand the dining landscape in their selected location.